26th International Conference

ECONOMIC COMPETITIVENESS AND SUSTAINABILITY 2024

PROCEEDINGS

March 21st–22nd, 2024, Brno

Editors:
Petr David
Hana Vránová

MENDELU
Faculty
of Business
and Economics

2024

MENDELU
Faculty
of Business
and Economics

Mendel University in Brno
Faculty of Business and Economics

26TH ANNUAL INTERNATIONAL CONFERENCE

ECONOMIC COMPETITIVENESS AND SUSTAINABILITY 2024

PROCEEDINGS

Editors:
Petr David, Hana Vránová

March 21st–22nd, 2024
Mendel University in Brno
Czech Republic

Organizer:
International scientific conference Economic Competitiveness and Sustainability 2024 was organized by the Faculty of Business and Economics, Mendel University in Brno.

Editors:
Hana Vránová (https://orcid.org/ 0000-0001-8254-1422)
Petr David (https://orcid.org/ 0000-0002-5055-0989)

ISBN 978-80-7509-990-7 (online ; pdf)
https://doi.org/10.11118/978-80-7509-990-7

ABSTRACT

On March 21st and 22nd, 2024, the Faculty of Business and Economics of Mendel University in Brno organized already the 26th international conference "Economic Competitiveness and Sustainability" with over 70 onsite participants from the Czech Republic, Slovakia, Germany, Poland, Hungary, Ukraine, Norway, Serbia and USA. In the presented Proceedings you find 21 papers which were recommended by conference discussants and selected on the basis of a peer-review process. The presented research outputs contribute to and extend the current state of knowledge and will stimulate further debate not only in academia but also in other institutions of public and private sector. The submitted papers reacts on the current interdisciplinary problems arising in the areas of Economic Policy, Public Finance and Public Administration, Enterprise Information Systems and Technologies, Digital Transformation and Sustainability, Marketing and Management and Finance.

Keywords: Informatics, Economic Policy, Finance, Digitalization, Management, Artificial Intelligence

ABSTRAKT

Ve dnech 21. a 22. března 2024 uspořádala Provozně ekonomická fakulta Mendelovy univerzity v Brně již 26. ročník mezinárodní konference „Economic Competitiveness and Sustainability", kterého se zúčastnilo přes 70 účastníků z České republiky, Slovenska, Německa, Polska, Maďarska, Ukrajiny, Norska, Srbska a USA. V předloženém sborníku najdete 21 příspěvků, které byly doporučeny diskutanty konference a vybrány na základě recenzního řízení. Prezentované výstupy výzkumu přispívají a rozšiřují současný stav poznání a podněcují další diskusi nejen na akademické půdě, ale i v rámci veřejného a soukromého sektoru. Předložené příspěvky reagují na aktuální interdisciplinární problémy vznikající v oblastech hospodářské politiky, veřejných financí a veřejné správy, podnikových informačních systémů a technologií, digitální transformace a udržitelnosti, marketingu, managementu a financí.

Klíčová slova: Informatika, hospodářská politika, finance, digitalizace, management, umělá inteligence

LIST OF CONTENT

FOREWORD

Dear readers,

We are honoured to present you with the Conference Proceedings we are publishing as the one of the outputs of the 26th year of the International Conference Economic Competitiveness and Sustainability (ECOS 2024) that we organised in March 21–22, 2024. The conference traditionally takes place in the campus of the Mendel University in Brno, Czech Republic and provides a productive background for the development of new ideas not only for senior but also for young researchers.

We hosted two distinguished well-known speakers who contributed to the conference programme with their keynote speeches. Prof. Jarko Fidrmuc (Zeppelin University, Chair of International & Digital Economics, Friedrichshafen) gave a speech "Economic and Institutional Implications of Populism in CEECs". Prof. Daniel Münich (CERGE – EI, Prague) contributed to the discussion on the topic of "Scientific Publication Performance in Post-communist Countries: Still Lagging Far Behind".

We welcomed more than 70 researchers representing universities and research institutes from countries including Czech Republic, Slovakia, Germany, Poland, Hungary, Ukraine, Norway, Serbia and the United States of America, who also served as the discussants of the papers and helped to improve the quality of the research results presented during both conference days.

The purpose of the conference is to enhance academic debate on current problems in the global socio-economic environment and dynamic technological development. Last but not least it is a chance to meet our research counterparts in person and critically review our perspectives on contemporary research problems.

In the presented Proceedings you can find 21 papers which were recommended by conference discussants and selected on the basis of a peer-review process. The presented research outputs contribute to and extend the current state of knowledge and will stimulate further debate not only in academia but also in other institutions of public and private sector.

The submitted papers react to the current interdisciplinary problems arising in the areas of Economic Policy, Public Finance and Public Administration, Enterprise Information Systems and Technologies, Digital Transformation and Sustainability, Marketing and Management and Finance.

We would like to thank all the participants in the conference for their inspiring contributions. Furthermore, we are grateful to all the reviewers and the members of the scientific committee for their contribution to the organisation of this high-level scientific conference.

ECONOMIC COMPETITIVENESS AND SUSTAINABILITY 2024

CONTRIBUTIONS

OPEN-SOURCE COMPLIANCE: A TACTICAL APPROACH

Ashish Bakshi[1], Andreas Kotulla[2], Oldřich Faldík[1], Oldřich Trenz[1]

[1]Department of Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

[2]Bitsea GmbH, Schloßstraße 7, 53757 Sankt Augustin, Germany

ABSTRACT

As open source becomes increasingly prevalent, understanding the intricacies of various license types, including permissive and copyleft licenses, becomes essential for developers and organizations alike (Tourani, Adams and Serebrenik, 2017). This paper not only explores these license types but also examines the implications of copyright laws and Export Control Compliance (ECC) on open-source software.

A significant portion of the paper is dedicated to evaluating key tools used in open-source compliance, such as SW360, FOSSology, OSS Review Toolkit (ORT), and Software Bill of Materials (SBOM).

In this paper, a comprehensive analysis of open-source license compliance offers practical insights and recommendations for developers and organizations navigating the complexities of open-source software adoption. The specific contribution of this paper lies in providing a detailed comparative analysis of these tools, alongside a case study on their application in real-time audits.

Keywords: open-source compliance, sw360, fossology, ORT, SBOM

JEL: L8, O3

1 INTRODUCTION

In the evolving landscape of software development, open-source software (OSS) has emerged as a cornerstone, driving innovation and collaboration across industries. However, with the widespread adoption of OSS (Sherae, 2016), the complexity and importance of license compliance have become more pronounced. This paper aims to provide a comprehensive overview of the multifaceted aspects of open-source license compliance, Export Control Compliance (ECC), and the security vulnerabilities and risks associated with open-source software, which remain crucial yet challenging domains for many organizations and developers (He, Peters, Menzies and Yang, 2013).

Various tools have been developed to aid in navigating these complexities, such as SW360, FOSSology, OSS Review Toolkit (ORT), SPDX, and Software Bill of Materials (SBOM). This paper

https://doi.org/10.11118/978-80-7509-990-7-0008

will evaluate the efficacy of these tools in facilitating compliance with open-source licenses, alongside a practical case study of a real-time audit. It will aim to investigate the impact of complexities in open-source licensing, export controls, and security vulnerabilities on the sustainable development and utilization of open-source software. Additionally, it will seek to explore how specialized tools such as SW360, FOSSology, ORT, SPDX, and SBOM can contribute to the effective management of legal and security risks in open-source software projects.

2 UNDERSTANDING OPEN-SOURCE LICENSES

Copyleft licenses, in contrast to permissive licenses, are designed to ensure that derivative works of the software remain open source. They require that modifications and extensions of the original software be distributed under the same license terms (Lindman, Paajanen and Rossi, 2010). Copyleft licenses ensure that derivative works of the software remain open source, requiring any modifications to be distributed under the same license terms. GNU General Public License (GPL), GNU Lesser General Public License (LGPL), Mozilla Public License (MPL) are some examples of copyleft licenses (González-Barahona, 2009).

2.1 Permissive licenses

Permissive licenses are a category of open-source licenses that impose minimal restrictions on how software can be used, modified, and distributed (Maryka, 2015). These licenses are often preferred for their flexibility and ease of integration into proprietary projects. Due to their minimal restrictions, permissive licensed software is often used in a variety of applications, from open-source projects to commercial products. MIT License, Apache License 2.0, and BSD Licenses (2-clause and 3-clause) are some examples of permissive Licenses (Coleman, 2014).

2.2 Copyleft licenses

Copyleft licenses, in contrast to permissive licenses, are designed to ensure that derivative works of the software remain open source. They require that modifications and extensions of the original software be distributed under the same license terms. Copyleft licenses ensure that derivative works of the software remain open source, requiring any modifications to be distributed under the same license terms. GNU General Public License (GPL), GNU Lesser General Public License (LGPL), and Mozilla Public License (MPL) are some examples of copyleft licenses (Mathur, Choudhary, Vashist, Thies and Thilagam, 2012).

3 EXPORT CONTROL COMPLIANCE (ECC)

Export Control Compliance (ECC) is an integral part of distributing and developing open-source software in a global context. Developers and organizations need to be vigilant and proactive in understanding and adhering to relevant export control laws to avoid legal repercussions and ensure responsible software distribution (Kumar, 2022). This is especially crucial in the context of open-source software, which often transcends international borders. ECC refers to a set of laws and regulations imposed by countries to control the export of certain technologies, including software, for reasons of national security or foreign policy (Choi, 2008).

Different countries have varying regulations on software exports. For instance, the United States' Export Administration Regulations (EAR) and the International Traffic in Arms Regulations (ITAR) govern the export of software that could be used in military or strategic contexts. Developers and organizations must be aware of these regulations to ensure compliance.

Software is classified under different categories based on its potential application in sensitive areas, such as encryption technology. Open-source software that falls into certain categories may require specific export licenses or be subject to restrictions.

ECC can impact how open-source software is developed and shared. For example, developers might need to restrict access to certain code repositories or implement measures to prevent the transfer of controlled technology (Shim, 2011).

4 VULNERABILITIES IN OPEN-SOURCE SOFTWARE

Open-source software, known for its numerous benefits like accessibility and collaborative development, is not immune to security vulnerabilities and risks. These vulnerabilities can significantly impact not only the security of the software but also compliance with open-source licenses. The widespread use of open-source software introduces a range of vulnerabilities that can affect both security and compliance with licensing terms. Understanding these vulnerabilities and their implications is crucial for maintaining the integrity and legal standing of software projects.

4.1 Identifying Common Vulnerabilities

4.1.1 Code Quality and Complexity

Open-source projects, particularly large ones, can face issues related to code quality and complexity, making them susceptible to security vulnerabilities. For example, Apache Struts and OpenSSL have encountered significant vulnerabilities, such as the Struts remote code execution vulnerability and the infamous Heartbleed bug in OpenSSL.

4.1.2 Dependency Management

Many open-source projects depend on other libraries and frameworks, where vulnerabilities in these dependencies can compromise the entire project's security. An instance of this was seen in the event-stream incident, posing serious threats to projects like Node.js and other JavaScript frameworks.

4.1.3 Lack of Sustained Maintenance

Inconsistencies in maintaining and updating some open-source projects can lead to outdated code and unaddressed vulnerabilities. Older versions of WordPress plugins, for example, have been known to contain security flaws exploited in various attacks.

4.2 Interplay Between Vulnerabilities and Licensing Compliance

4.2.1 Security Updates and Licensing Terms

Licenses such as the GNU GPL mandate public disclosure of all modifications, including security patches. Failure to comply can result in legal disputes, evident in instances where organizations did not release modified source code back to the community.

4.2.2 Enforceability and Security Flaws

Vulnerabilities can challenge the enforceability of licenses. A security patch that alters the software's original functionality might inadvertently violate the original license terms.

4.2.3 Compliance Challenges in Addressing Vulnerabilities

Effectively managing vulnerabilities, as demonstrated by the Linux kernel (GPL-licensed), involves a delicate balance between regular security updates and adherence to licensing terms, a complex task for many organizations.

5 TOOLSETS FOR COMPREHENSIVE COMPLIANCE

5.1 SW360

SW360 excels in managing and documenting the licenses of all software components within a project. It facilitates an organized approach to maintaining compliance with various open--source licenses, thereby mitigating legal risks. The tool serves as a comprehensive catalog for software components, simplifying the process for organizations to track and manage the use of open-source software in their projects. This functionality is crucial for maintaining a clear overview of all software dependencies and their corresponding licenses.

SW360 can be seamlessly integrated into existing software development processes. This integration ensures that license compliance becomes a continuous and integral part of the development lifecycle, rather than an afterthought.

By providing a central platform for managing open-source components, SW360 fosters collaboration among development teams and enhances transparency in the usage and management of open-source software within an organization.

5.2 Open-Source Review Toolkit (ORT)

ORT offers an extensive suite of capabilities designed to streamline the review and analysis of open-source licenses. Its primary aim is to assist organizations in understanding and complying with open-source licenses, thereby reducing legal risks, and enhancing project integrity. ORT analyzes the licenses of software components within a project, providing a detailed overview of compliance requirements and potential risks.

The process begins with ORT scanning the files of a software project. It identifies and lists all the open-source components and dependencies used within the project.

To detect licenses associated with each component or dependency, ORT employs various scanners, such as *ScanCode*. It searches for license files, headers in source code, and other metadata to accurately identify the licenses.

5.3 FOSSology

FOSSology operates by scanning software code to detect open-source licenses, enabling users to review, curate, and report on the license information. Its support for SPDX and customizable scanning options make it a versatile tool for managing license compliance in diverse software projects. FOSSology scans the uploaded files to detect and identify open-source licenses, using advanced scanning techniques to examine file contents, including comments, headers, and documentation, for license information. Utilizing its comprehensive database of open-source licenses, FOSSology can identify a wide range of licenses in the scanned files. The tool highlights the exact text snippets where license terms are found, making it easier for users to review and confirm the license findings. Users can manually review the scan results to verify or correct the identified licenses. This step is essential, particularly in cases where the software contains custom licensing terms or dual-licensed files.

6 METHODOLOGY

The OSS Review Toolkit (ORT) is used for analyzing and reviewing dependencies in open-source software. It scans a project to identify all open-source components and their respective licenses. When an Ansible, *https://github.com/ansible/ansible*, project is run through ORT, it generates a detailed report of all the open-source components used in the project, along with their respective licenses.

After the ORT process, the same Ansible project is then fed into Fossology for a more thorough license analysis. This step is likely to provide a more granular view of the licenses and any potential issues or conflicts.

Following the analyses conducted by ORT and Fossology, the results are integrated into SW360. This tool serves as a centralized platform for managing these components, tracking license compliance, and documenting findings. SW360 allows for systematic management of open-source components, ensuring that all data is consistently updated and easily accessible for review. The aggregated data in SW360 can then be reviewed by audit and legal teams. These teams assess the compliance of the open-source components with the organization's policies and legal requirements.

The compliance of these open-source components is then rigorously evaluated by the organization's audit and legal teams. They assess each component against specific review criteria, including adherence to licensing terms, compatibility with internal policies, and legal risk management. This review ensures that all components meet the stringent standards required for organizational use and legal compliance, Finally, once the open-source components have been cleared by the audit and legal teams, they are approved for deployment in production environments. This approach effectively combines automated tooling with manual review, ensuring that open-source components are used responsibly and in compliance with legal requirements. This process not only aids in license compliance but also helps in managing security vulnerabilities that might be present in open-source components.

7 RESULT

The application of FOSSology to the Ansible codebase revealed intriguing results. The initial scan identified a significant number of files with GPL 3.0 licenses.

The report generated by FOSSology initially indicated that there were 616 files licensed under GPL 3.0 and another 377 files with GPL 2.0 licenses. Following the initial scan, a meticulous file-by-file review was conducted to pinpoint the presence of strict copyleft licenses and to identify the copyright holders for each component.

This granular examination was crucial to understand the extent and implications of copyleft licensing within the Ansible project.

The Ansible project was subjected to ORT's analysis, utilizing its analyzer and downloader modules. This process was aimed at meticulously scanning each dependency and transitive dependency within the project.

The analyzer module was particularly instrumental in breaking down and identifying the specifics of each dependency, offering a detailed look into the project's composition. ORT's comprehensive scanning covered not only the direct dependencies of Ansible but also the transitive dependencies, which are often overlooked yet crucial for a complete compliance picture.

This extensive scanning provided a deeper insight into the software stack, revealing the intricate network of dependencies within Ansible.

Upon completion of the ORT analysis, the SPDX (Software Package Data Exchange) report was seamlessly integrated into SW360 through its import functionality. This integration marked a pivotal step in centralizing and streamlining the compliance management process.

The imported SPDX report within SW360 became readily accessible to various teams, including the audit and legal teams. This accessibility facilitated a more efficient and transparent review process.

Such visibility is crucial for these teams to perform thorough compliance checks and to make informed decisions regarding the legal aspects of the project.

With the SPDX report in SW360, collaboration among different departments was significantly improved. The legal team, for instance, could easily cross-reference the report for any licensing issues, while the audit team could use the data for compliance verification.

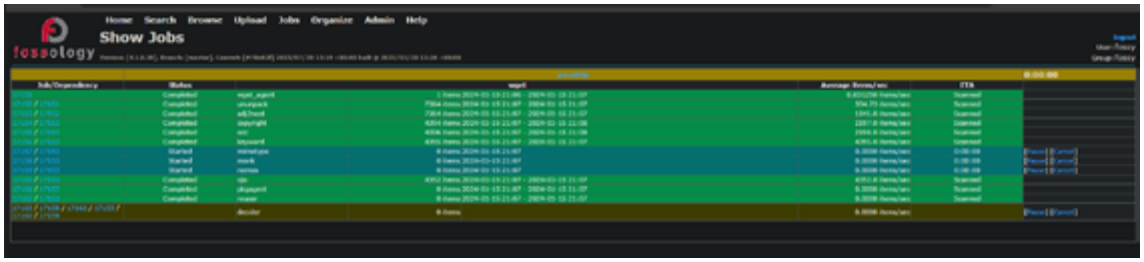


Fig. 1: Fossology scan Ansible

NUMBER	AUDITED	LICENSE
3	0	Apache-2.0
4	0	BSD
65	0	BSD-2-Clause
4	0	BSD-3-Clause
5	0	GPL
1	0	GPL-2.0
9	0	GPL-2.0+
616	0	GPL-3.0
377	0	GPL-3.0+
4	0	MIT
58	0	No_license_found
3	0	PSF-2.0
1	0	Public-domain
5	0	Python
3	0	See-URL
1	0	See-doc.OTHER

Tab. 1 Report of License Ansible

COPYRIGHT STATEMENT	FILE LOCATION
Copyright: Contributors to the Ansible project GNU General Public License v3.0+ (see COPYING or https://www.gnu.org/licenses/gpl-3.0.txt)	lib/ansible/compat/importlib_resources.py lib/ansible/modules/deb822_repository.py test/integration/targets/result_pickle_error/action_plugins/ result_pickle_error.py test/units/module_utils/urls/test_fetch_file.py test/units/module_utils/urls/test_split.py
Copyright: Contributors to the Ansible project	lib/ansible/module_utils/urls.py
Copyright: Ansible Team GNU General Public License v3.0+ (see COPYING or https://www.gnu.org/licenses/gpl-3.0.txt)	lib/ansible/modules/add_host.py lib/ansible/modules/group_by.py

Tab. 2 Report of Copyright Ansible

```
server-6 user (e) base … ort-native mime-types reporter-output-dir ls
bom.spdx.yml  gl-license-scanning-report.json  NOTICE_DEFAULT  scan-report.html  scan-report-web-app.html
server-6 user (e) base … ort-native mime-types reporter-output-dir cat gl-license-scanning-report.json
{
  "version": "2.1",
  "licenses": [
    {
      "id": "Apache-2.0",
      "name": "Apache License 2.0",
      "url": "https://spdx.org/licenses/Apache-2.0"
    },
    {
      "id": "BSD-2-Clause",
      "name": "BSD 2-Clause \"Simplified\" License",
      "url": "https://spdx.org/licenses/BSD-2-Clause"
    },
    {
      "id": "BSD-3-Clause",
      "name": "BSD 3-Clause \"New\" or \"Revised\" License",
      "url": "https://spdx.org/licenses/BSD-3-Clause"
    },
    {
      "id": "CC-BY-3.0",
      "name": "Creative Commons Attribution 3.0 Unported",
      "url": "https://spdx.org/licenses/CC-BY-3.0"
    },
    {
      "id": "CC0-1.0",
      "name": "Creative Commons Zero v1.0 Universal",
      "url": "https://spdx.org/licenses/CC0-1.0"
    },
    {
      "id": "ISC",
      "name": "ISC License",
      "url": "https://spdx.org/licenses/ISC"
    },
    {
      "id": "LicenseRef-scancode-public-domain-disclaimer",
      "name": "",
      "url": ""
    },
    {
      "id": "MIT",
      "name": "MIT License",
      "url": "https://spdx.org/licenses/MIT"
    },
    {
      "id": "WTFPL",
      "name": "Do What The F*ck You Want To Public License",
      "url": "https://spdx.org/licenses/WTFPL"
    },
    {
      "id": "X11",
      "name": "X11 License",
      "url": "https://spdx.org/licenses/X11"
    }
  ],
  "dependencies": [
    {
```

Fig. 2: ORT scan Ansible SPDX report

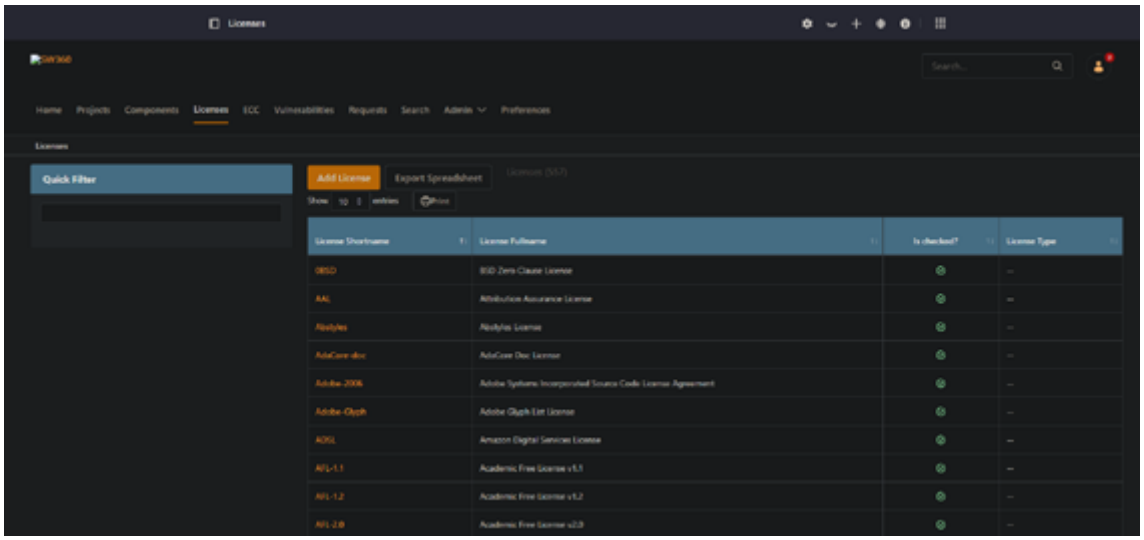


Fig. 3: SW360 Report imported

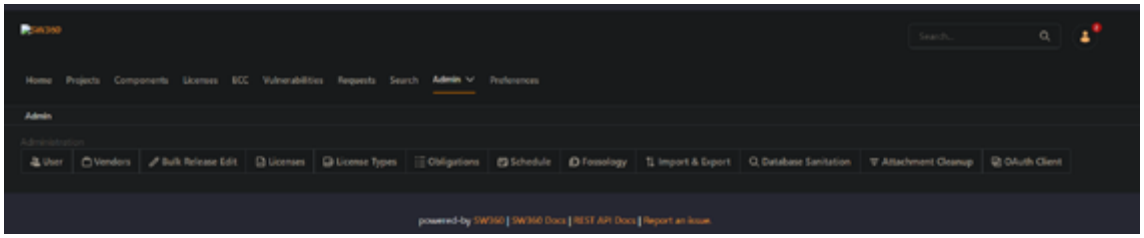


Fig. 4: SW360 ECC, Obligation, Vulnerability

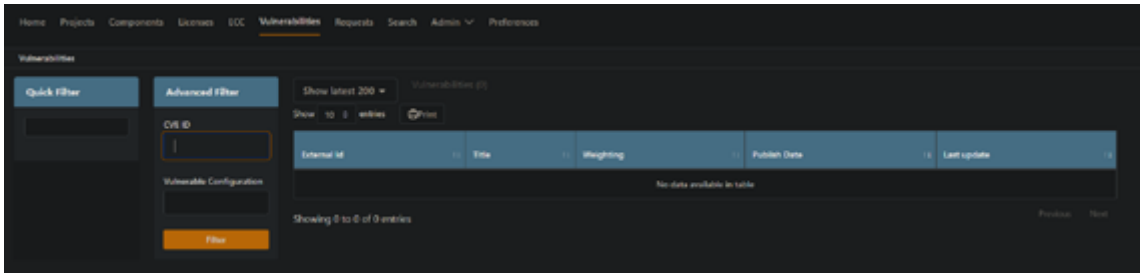


Fig. 5: SW360 CVE id

8 DISCUSSION AND CONCLUSIONS

The prevalence of GPL 3.0 licenses poses certain compliance requirements, especially considering the strict copyleft nature of this license. This finding necessitates a careful approach to how the Ansible software is used, modified, and distributed.

The identification of GPL 2.0 licenses also has significant implications, particularly in terms of compatibility with other licenses and the obligations it imposes on derivative works. The use of ORT and SW360 in this manner promotes responsible open-source software development practices. By ensuring thorough compliance, organizations can avoid legal pitfalls and maintain ethical standards.

These results highlight the complexity of managing open-source licenses in large projects. The mix of GPL 3.0 and GPL 2.0 licenses within Ansible underscores the need for thorough compliance checks and an understanding of license obligations.

Accurate license compliance management, facilitated by these tools, is essential for the long-term sustainability of open-source projects. It ensures that projects adhere to legal requirements, thereby securing their viability and reputation in the open-source community.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

CHOI, C.-H. 2008. A Study on Global Compliance of Global Companies under the Circumstance of Export Control. *The Korean Research Institute of International Commerce and Law.*

COLEMAN, M. A. 2014. *Freedom From Restriction, Freedom Of A Restriction: A Comparison Of Some Open Source Software Licenses*. https://arxiv.org/abs/1402.2079

GONZÁLEZ-BARAHONA, D. M. 2009. An Empirical Study of the Reuse of Software Licensed under the GNU General Public License. In: BOLDYREFF, C., CROWSTON, K., LUNDELL, B., WASSERMAN, A. I. (eds.). *Open Source Ecosystems: Diverse Communities Interacting. OSS 2009. IFIP Advances in Information and Communication Technology*. Vol 299. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02032-2_17

HE, Z., PETERS, F., MENZIES, T. and YANG, Y. 2013. Learning from Open-Source Projects: An Empirical Study on Defect Prediction. In: *2013 ACM / IEEE International Symposium on Empirical Software Engineering and Measurement.* Baltimore, MD, USA, 2013, pp. 45-54, doi: 10.1109/ESEM.2013.20

KUMAR, N. 2022. *Export Compliance as a Response To Export Control*. Rel. Guido Sassi. Politecnico di Torino, Corso di laurea magistrale in Ingegneria Gestionale (Engineering And Management).

LINDMAN, J., PAAJANEN, A. and ROSSI, M. 2010. Choosing an Open Source Software License in Commercial Context: A Managerial Perspective. In: *36th EUROMICRO Conference on Software Engineering and Advanced Applications. L*ille, France, 2010, pp. 237-244, doi: 10.1109/SEAA.2010.26

MATHUR, A., CHOUDHARY, H., VASHIST, P., THIES, W. and THILAGAM, S. 2012. An Empirical Study of License Violations in Open Source Projects. In: *35th Annual IEEE Software Engineering Workshop.* Heraklion, Greece, 2012, pp. 168-176, doi: 10.1109/SEW.2012.24

SHERAE, D. and STEWART, K. 2016. Open source project success: Resource access, flow, and integration. *The Journal of Strategic Information Systems*, 25(3), 159-176. https://doi.org/10.1016/j.jsis.2016.02.006

SHIM, S.-R. 2011. A Comparative Study on the Compliance Program(CP) of Strategic Export Control System between Korea and Japan. *International Commerce and Information Review*.

TOURANI, P., ADAMS, B. and SEREBRENIK, A. 2017. Code of conduct in open source projects. In: *IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER).* Klagenfurt, Austria, 2017, pp. 24-33, doi: 10.1109/SANER.2017.7884606

MARYKA, T., GERMAN, D. M. and POO-CAAMAÑO, G. 2015. On the Variability of the BSD and MIT Licenses. In: DAMIANI, E., FRATI, F., RIEHLE, D., WASSERMAN, A. (eds.). *Open Source Systems: Adoption and Impact. OSS 2015. IFIP Advances in Information and Communication Technology*. Vol 451. Springer, Cham. https://doi.org/10.1007/978-3-319-17837-0_14

Contact information
Ashish Bakshi: e-mail: xbakshi@mendelu.cz
Andreas Kotulla: e-mail: andreas.kotulla@bitsea.de
Oldřich Faldík: e-mail: oldrich.faldik@mendelu.cz
Oldřich Trenz: e-mail: oldrich.trenz@mendelu.cz

PERFORMANCE COMPARISON OF HTTP/3 SERVER IMPLEMENTATIONS

Jiří Balej[1], Tomáš Sochor[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

The paper opens the question of comparison published open-source webservers with support of HTTP/3. This is brand new protocol standardised in 2022 was developed with aim to speed up web communication and instead of TCP uses UDP with new QUIC protocol. There is summarized current state of HTTP protocols development and studies, where is compared performance of various HTTP versions. The aim of paper is to compare different open-source webserver implementations with HTTP/3 support in laboratory. Five different scenarios were presented to test ordinary real-life situations. The results of all three servers with different content and HTTP/1.1 or HTTP/3 protocols are presented. Main result would be better performance of Caddy and nginx server in bad connection conditions, but without speed limit, large delay or loss the OpenLiteSpeed was fastest.

Keywords: HTTP/3, QUIC, Caddy, Nginx, OpenLiteSpeed, webserver

JEL Code: L86, C88, O31

1 INTRODUCTION

Thirty-five years have passed since Tim Barnes-Lee at CERN proposed the first version of web communication. Since then, the HTTP (HyperText Transfer Protocol) has become the most used application protocol in the Internet. The first standardization by IETF (Internet Engineering Task Force) happened in 1996 in RFC 1945. Since this version many improvements like keep--alive connections, content negotiation and virtual hosts (in 1997 with HTTP/1.1, RFC 2068), binary coding, multiplexing and header compression (in 2015 with HTTP/2, RFC 7540) and using QUIC instead TCP protocol (in 2022 with HTTP/3, RFC 9114) were included to make HTTP complex application. The issue of security has been addressed by adding SSL/TLS layer under the basic HTTP communication (Lysenko, 2023), which provides authentication of server (and optionally a client), confidentiality and integrity of sent data.

HTTP communication also affect Cookies, which are capable of sharing session-related information between independent HTTP requests.

The main research question here is to check the difference among webservers implementations for HTTP/3 and to determine the dependence of performance on the network condition.

https://doi.org/10.11118/978-80-7509-990-7-0017

The key attribute for user's Quality of Service/Experience is the load time of the webpage, which will be tested as the "performance" parameter. Currently several webserver implementations support HTTP/3 and our experiment should give answer if those implementations perform similarly under the same conditions. As a reference, the well-known HTTP/1.1 protocol was tested to find out whether the performance of webservers with HTTP/1.1 and HTTP/3 protocols are similar.

After the introduction and stating the aim of article, the next part contain an actual state of HTTP protocol and other comparison studies of HTTP protocol versions are presented. Third section is devoted to experiment settings, topology and network conditions used in scenarios. Finally, the results are presented in graphs ontaining the measurement results of all prepared combinations. The findings are concluded and discussed in the last section.

2 ADVANCES IN HTTP PROTOCOL

One of the stable and even nowadays common version of HTTP protocol is HTTP/1.1 that has provided all necessary web functions for normal web operation and ruled the web for over 20 years. This changed around 2010, when Google started development of advanced HTTP protocol, which would speed up the web communication. The resulting SPDY should have been able to speed up loading time of webpage by 50% (Belshe, 2012). One of the key ideas was to use header compression and binary encoding. The multiplexing of packets for various content, its prioritization and server push have improved the speed greatly. All the functions were later included in the official HTTP/2 standard.

Right after that, work on development of HTTP/3 started, again driven by Google. All HTTP protocol were using TCP (Transmission Control Protocol) as transport-layer (L4) protocol. The TCP provide connection-oriented and reliable data transfer. TCP also uses sliding window parameter to dynamically utilize the maximum available bandwidth (congestion control). Biswal (2022) compares some of current TCP congestion control algorithms and states that BIC is default algorithm on Linux.

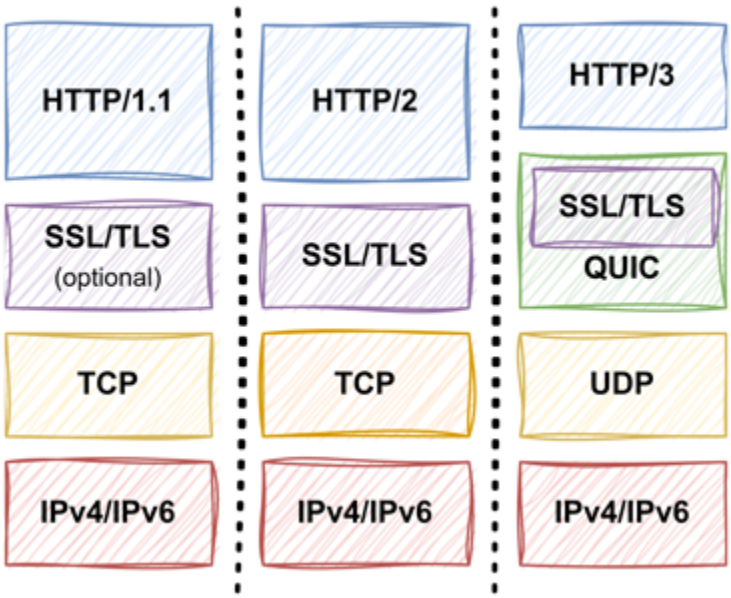


Fig. 1: Architecture of TCP/IP stack with HTTP
Source: Saadat, 2022

All the TCP congestion control algorithms rapidly slow down with loss of any packet and use numbered line of packets, which makes multiplexing of different streams complicated. That's why the new version of HTTP started to use UDP (User Datagram Protocol) as transport layer protocol, which offers connection-less best-effort send without any control of transmission. The architecture of TCP/IP stack with different versions of HTTP is on Figure 1, where is added the security layer (SSL/TLS) function.

2.1 QUIC

The QUIC (Quick UDP Internet Connections) protocol should outperform the drawbacks of TCP connections, with at the same time providing reliable data transfer necessary for HTTP. Moreover, QUIC contain the security layer (SSL/TLS) which is implied during initialization of communication and then protect whole communication.

Iyengar (2021) specifies in standard (RFC 9000) for QUIC that each communication starts with QUIC handshake, where are exchanged the supported cryptography algorithms and server is always authenticated by asymmetric cryptography (public key is in server certificate), optionally can be also the client authenticated. In case it is not the communication with server, QUIC can use so-called 0-RTT and because the server is known the client with the first packet send also encrypted data.

Other mechanisms used in QUIC specified by Lyengar (2021) are Connection Migration, Flow Control, Stream Multiplexing and Reliability. Packet numbering with positive acknowledgment and retransmission are common for reliability. QUIC moreover implies selective acknowledgment, fast retransmit and tail loss probes to achieve greater performance in case of data loss.

2.2 HTTP/3

The biggest difference of HTTP/2 is the usage of QUIC (and UDP) as transport layer, which could be a challenge for socket association, firewall permissions and packet inspection technologies. Bishop (2022) in specification of HTTP/3 (RFC 9114) also describes a new HTTP header "Alt-Svc" announcing its support and could ever redirect on different port number. Clients also open only one connection with the server, which is kept open until all requests are served. Loading of webpage could be also speeded up using server-push mode, where server sends required data without client's request.

Even the QUIC was standardized in 2022 the big companies like Cloudflare, Google, Microsoft, Meta and others uses it for several years. In January 2024 the average usage statistics (W3Techs, 2024) stated, that 28.4% of all websites use HTTP/3 and the number is increasing.

2.3 Studies and measurements over HTTP/3

Since the proposing idea of HTTP traffic over UDP till today, when HTTP/3 is standardized, there have been many survey comparing properties of HTTP/3 with its predecessors. One of the first papers was published by Carlucci (2015) and the compares TCP CUBIC with Google's public implementation of QUIC (v.21). Then Mefyesi (2016) published a study where the HTTP, SPDY and QUIC were compared for time needed to load a webpage. The question if QUIC would be better than others and for what cases is investigated by Cook (2017), all mentioned papers agree that older variants are better, at least in most common network conditions.

The state how QUIC is implemented was summarized by Rüth (2018), where 161 thousand of domain names support QUIC and about 3–9% of overall web traffic is QUIC. The better performance for video content and smaller files was HTTP/3 in the paper from Shreedhar (2022). Another performance measurement was done by Perna (2022), with conclusion of HTTP/3 advantages for large delays. Dubec (2023) compares not the browsers but client implementation.

Most of studies used public web services like Google to make comparison of performance, functions and adoption. Only few papers used any of open-source webserver with HTTP/3 support and tested the site in laboratory under various networks conditions.

3 METHODOLOGY

The aim of the paper is to test performance of open-source webservers with HTTP/3 support. Currently (beginning of 2024) is fully or experimentally supported by ASP.NET Core Kestrel, Caddy, Microsoft IIS, Nginx, OpenLiteSpeed and LiteSpeed. In web browsers HTTP/3 is supported more commonly and all major browsers currently support it.

From the list above, three open-source webservers – Caddy (version 2.7.6), Nginx (1.25.3) and OpenLiteSpeed (1.7.18-2) were chosen. As a client the line-based web browser "curl" (version 8.2.1) built with HTTP3 support was used and Google Chrome was used too to compare some results for which the line-based client output was not clear. For all devices (servers, client and traffic control) Ubuntu Server 22.04.3 LTS installed as VM in VirtualBox (7.0.10) was used. As the topology for the experiment shows in Figure 2, all devices are run on independent hardware (Intel Core i5-3470, 16 GiB RAM, SSD hard drive, 1 Gbit/s network card), all virtual machines have assigned 4 CPU cores and 8 GiB RAM. The only switch is a dedicated hardware (MikroTik RB4011GS+RM) with integrated 5 port 1 Gbit/s switching module.

On each webserver, there were three types of content website and two files (1 MiB and 10 MiB). The website was copy of homepage for Mendel University in Brno, which contain 30 files (HTML, CSS, JS, JPG,..) and its overall size is about 5 MiB. Based on general characteristics of Internet traffic (Araújo, 2019) four scenarios were set as shown in Table 1 with the worsened network conditions (speed, delay and loss), which were intentionally affected by Linux Traffic Control (tc). To ensure correctness of tc settings, we also tested the communication without any downgrade (speed up to 300 Mbps, latency 2 ms, loss 0%). Each scenario was tested 10 times against all 3 servers with protocols HTTP/1.1 and HTTP/3.

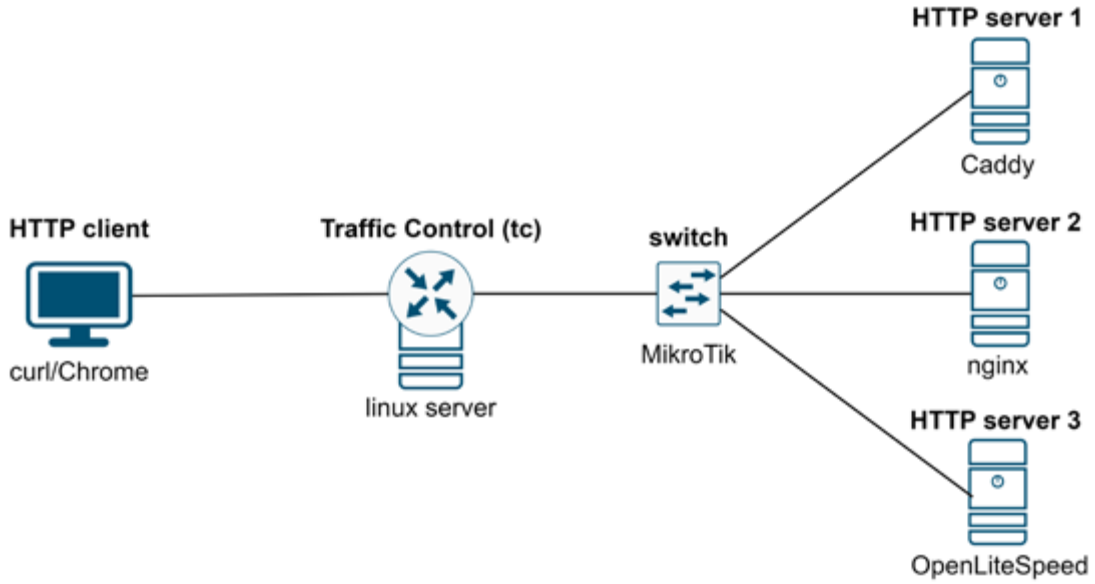


Fig. 2: Laboratory experiment topology

	Speed	Delay	Loss
High speed LAN connection (cable)	100 Mbps	5 ms	0.001%
Home WiFi	60 Mbps	25 ms	0.01%
WiFi in public place (like Café)	25 Mbps	70 ms	0.05%
Unstable connection in moving vehicle	10 Mbps	250 ms	0.1%

Tab. 1 Parameter of scenarios for HTTP/1.1 and HTTP/3 measurement

4 RESULTS

Collected measured values contained time needed to load whole file/webpage, all header values including status code of communication and the content. The example of result for one page load is shown in Figure 3, where are partial times for the process. In cast we got different status code, than 200 OK, the measurement was repeated. For each scenario there were 10 times of loading complete content and the median, average, minimum and maximum were used for comparisons.

For final comparison, we used the median values to avoid exceptional and random delays. The loading times in bar graphs for all scenarios are shown in Figure 4, mainly to compare performances between individual webservers. Each bar represents a time needed to load the whole content in milliseconds (ms), the lower number (lower bar) means better performance. The left side of graphs represents HTTP/1.1 protocol while the right one HTTP/3 protocol. Each of individual scenarios described in Table 1 are represented by a single graph as shown by the graph heading.

5 DISCUSSION AND CONCLUSIONS

The results prove that all three open-source webserver implementations are capable of handling HTTP/3 content. Of course, there are differences in implementations, for example webserver OpenLiteSpeed has best performance in scenario without any loss, delay or

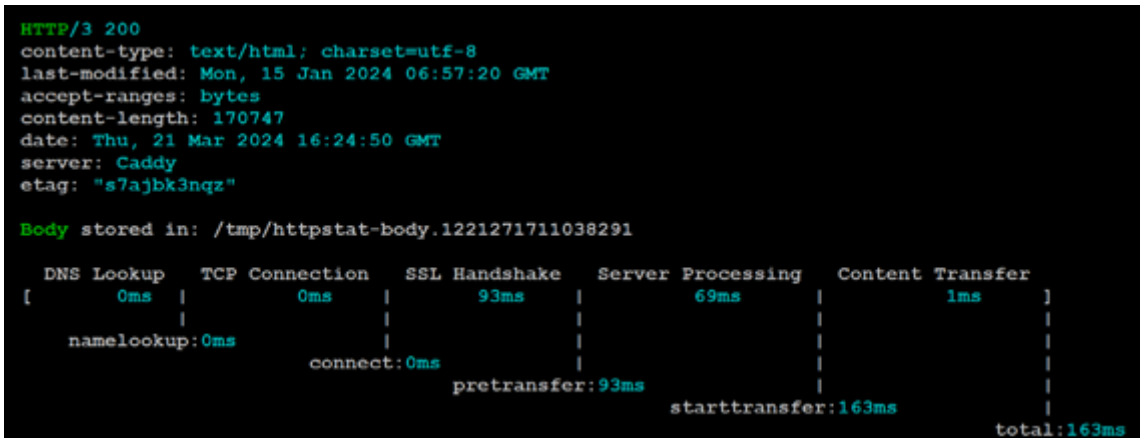


Fig. 3: Detail print of loading time result, where individual delay components are displayed.

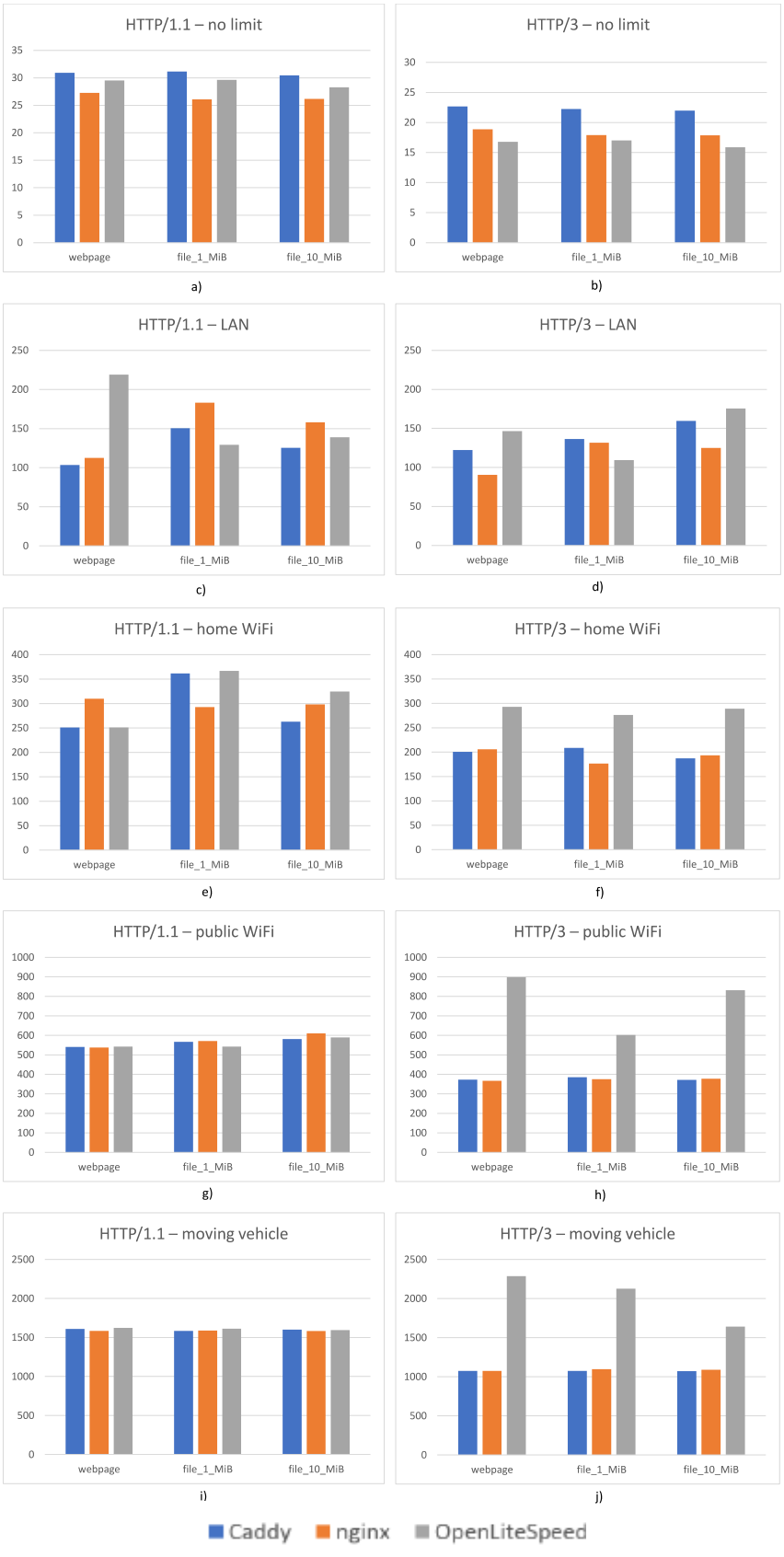


Fig. 4: Graph of median loading time for each scenario and type of content, webserver and HTTP protocol.

constrain of speed, on the other hand in most other scenarios is the one of worst. Webservers Caddy and nginx show similar performance in bad situations.

When one tries to compare HTTP/1.1. and HTTP/3 performance, there is no clear winner but it seems that our result confirms previously published results that HTTP/3 is better in good network conditions (as graphs on the left show) while the HTTP/3 dominance is not clear in worse transmission conditions.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ARAÚJO, J. 2019. Real-world latency and packet loss. *Codeavel blog* [online]. Available at: https://blog.codavel.com/performance-report-defining-use-cases [Accessed: 2023-12-17].

BELSHE, M., PEON, R., THOMSON, M. and MELNIKOV, A. 2012. SPDY Protocol draft-ietf-httpbis-http2-00. *Internet-Draft* [online]. Available at: https://datatracker.ietf.org/doc/html/draft-ietf-httpbis-http2-00

BISWAL, S. P., GUPTA, S., KUMAR, A. and PATEL, S. 2022. *Comparative Analysis of Compound TCP with Various End-to-End Congestion Control Mechanisms* [online]. Available at: http://dspace.nitrkl.ac.in/dspace/bitstream/2080/3879/1/PatelS_ICAC3N-22.pdf

CARLUCCI, G., DE CICCO, L. and MASCOLO, S. 2015. HTTP over UDP: an experimental investigation of QUIC. *Internet-Draft In Proceedings of the 30th Annual ACM Symposium on Applied Computing* (SAC '15). Association for Computing Machinery, New York, NY, USA, 609–614. https://doi.org/10.1145/2695664.2695706

COOK, S., MATHIEU, B., TRUONG, P. and HAMCHAUOI, I. 2017. QUIC: Better for what and for whom? *IEEE International Conference on Communications* (ICC). Paris, France, pp. 1-6. https://doi.org/10.1109/ICC.2017.7997281

DUBEC, J., BALAŽIA, J. and ČIČÁK, P. 2023. Performance evaluation of the HTTP/3 client implementations. *46th International Conference on Telecommunications and Signal Processing* (TSP), Prague, Czech Republic, pp. 260-263- doi: 10.1109/TSP59544.2023.10197834

KUTTER, M. and JAEGER, B. 2022. Comparison of Different QUIC Implementations, *Network Architectures and Services*- doi: 10.2313/NET-2022-07-1_10

LYSENKO, S. and SAVENKO, B. 2023. Distributed Discrete Malware Detection Systems Based on Partial Centralization and Self-Organization. *International Journal of Computing*, 22, 117-139.

MEGYESI, P., KRÄMER., Z. and MOLNÁR, S. 2016. How quick is QUIC?. *IEEE International Conference on Communications* (ICC), Kuala Lumpur, Malaysia, pp. 1-6, doi: 10.1109/ICC.2016.7510788

PERNA, G., TREVISAN, M., GIORDANO, D. and DRAGO, I. 2022. A first look at HTTP/3 adoption and performance, *Computer Communications*, Volume 187, pp. 115-124, ISSN 0140-3664. https://doi.org/10.1016/j.comcom.2022.02.005

RÜTH, J., POESE, I., DIETZEL, C. and HOHLFELD, O. 2018. A First Look at QUIC in the Wild. *Passive and Active Measurement* (PAM 2018). Springer, doi: 10.48550/arXiv.1801.05168

SHREEDHAR, R., PANDA, R., PODANEV, S. and BAJPAI, V. 2022. Evaluating QUIC Performance Over Web, Cloud Storage, and Video Workloads. In: *IEEE Transactions on Network and Service Management*, vol. 19, no. 2, pp. 1366-1381, June 2022. https://doi.org/10.1109/TNSM.2021.3134562.

SAADAT, R. 2022. Spirent Communications. In: *Building Next-Generation Web with HTTP/3*. https://www.spirent.com/blogs/building-next-generation-web-with-http-3 [Accessed: 2024-01-10].

Contact information

Jiří Balej: e-mail: jiri.balej@mendelu.cz
Tomáš Sochor: e-mail: tomas.sochor@mendelu.cz

NATURE OF INTERNAL COMMUNICATION AMONG E-WORKERS, CASE STUDY

Michal Beňo[1], Jan Kříž[2,] Dagmar Cagáňová[3], Zuzana Cagáňová[4]

[1]Prague City University, Business Faculty, Hybernská 24, 110 00 Praha 1, Czech Republic
[2]Institute of Technology and Business, Okružní 517, 370 01 České Budějovice, Czech Republic
[3]Comenius University Bratislava, Faculty of Management, Odbojárov 10, 820 05 Bratislava 25, Slovakia
[4]Mendel University in Brno, Faculty of Business and Economics, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

Research background: COVID-19 pandemic caused a rapid shift to full-time remote work environment. Workforce productivity, engagement, and success are affected by how the company interacts with its workforce. Human-to-human connection without clear communication is impossible. Communication is essential for a stable, engaged company. In e-work environment even more.

Purpose of the article: The purpose of this study is to provide recommendations for improving the flow of information and stregthen the sense of community among real e-employees in a remote work environment.

Methods: a quantitative study (web-based survey) in a Greek multicultural company has been used. Authors tried to find answers to following research questions: How do e-employees perceive internal communication methods in their organization? Which of the internal communication channels were used most effectively?

Findings & Value added: Based on the crea ted questionnaire, a total of 7 hypotheses were determined, all of which were statistically confirmed. Obtained data confirms the raising importance of internal communication, especially in times of crisis as employees considered as a homogenous. We can underline the sample based in Greece but located various European countries and in a context of unprecedented crisis. English as predominant spoken language. Finally, this research only explores the role of internal communication in an organizational context (formal, casual, organized, or unplanned).

The survey was conducted during pandemics and has some interesting implications for communication professionals and researchers in the field of internal communication.

The present study examines internal communication in e-working settings. Previous studies have excluded the situation for e-employees. The results provide new insights into internal communication behaviour in e-work.

Keywords: internal communication, e-work, methods, channels, multiculturality

JEL classification: D83, M12, M21, M54, J21

https://doi.org/10.11118/978-80-7509-990-7-0024

1 INTRODUCTION

E-working is occasionally promoted as having the potential to be the great equaliser of economic opportunity, competitiveness and sustainability.

According to EU statistics, in 2019, 5.5% employees worked from home, in 2020, due to the pandemic, this share increased to 12.3%; and in 2021, to 13.5% (Eurostat, 2022). Work from home (WFH) or e-commuting (Beno and Caganova, 2023) allows flexibility for individuals with personal and professional goals. However, it also comes with challenges (Ambikapathy and Ali, 2020; Kurkland and Bailey, 1999; Beno, 2021). Especially in relation to workplace engagement, including communication (Cagáňová et al., 2019). Erum et al. (2020, p. 16) stated that "when employees have a calling to their work they find themselves more adjusted for their job and are more likely to feel passion towards it, they are intrinsically motivated to perform it thus deriving happiness in their workplace". This disengagement causes various problems due to lowered productivity and decreased satisfaction. Lal et al. (2021) highlighted that social interactions (in the on-site workplace) are an essential feature of social life and shape the workforce's experience of work. However, organizations are able to work to intentionally address these issues. The potential tool may be internal communication (IC) (Cagáňová et al., 2015). It may help e-workers feel more socially connected.

IC is often interlinked with an organization's reputation and brand. But one of the most crucial links in the organization, which is easily overlooked, is the internal one, namely the workforce. They are the ones who create the brand. Organizations are able to convey the ambitions and values of the brands, but very often sufficient information flow is missing. IC increases trust in organizations, which is often associated with business success. (Cagáňová et al., 2010).

IC should be a deliberate, well-planned, and well-articulated element of a company's communication strategy. Many organizations are still operating remotely with employees scattered across the nation and around the world, like the authors' sample. Thus, it can be difficult to communicate in a way that keeps everyone engaged. Numerous studies have examined the main issues of IC (Flynn and Mazor, 2017; Oak, 2022; Robert Half, 2013; Queens University, 2023) and have linked IC and the degree to which employees are informed to job satisfaction and performance. The competitive benefits of IC come not only from satisfaction, productivity and sense of community but also from the positivity of external public relations. The purpose of this study is to provide recommendations for improving the flow of information and stregthen the sense of community among real e-employees in a remote work environment.

Simply identify and describe IC according to e-employees. A quantitative IC study has been chosen to identify the issues faced by e-employees in order to answer these questions. The following research questions were outlined:

How do e-employees perceive IC methods in their organization?

Which of the internal communication channels were used most effectively?

The company under investigation required anonymity. Its internal communication system was launched several years ago. In the questionnaire survey, employees and basic managers made up the largest sample, but middle and top managers also participated. The survey was conducted in a Greek multicultural company, with employees from Poland, Germany, the Czech Republic, Greece, Portugal, and Romania. All interviewees (total 182) met the condition of remote working mode.

2 METHODS

According to McDonough and McDonough (1997), the authors of this study implemented both interpretive and evaluative case study with the aim to interpret the data by developing conceptual categories and add judgement to the phenomena in the data. It allows a deeper understanding of the specifics of IC, providing a close understanding of the e-employees on IC and its meaning.

Because of the nature of a case-study, purposeful sampling was conducted with the emphasis on selecting an information rich group of respondents. The study is conducted on the entire target population which consists of 182 e-employees (57.7% of females, 40.7% of males and 1.6% diverse) with an average age 32, median 30 of a global Greece organization in operating in sector service. This confirms that service sector is the area with the highest female labour rate (Worldbank, 2021). Effective IC can be expected to be an important variable among surveyed e-workforce. All respondents worked for an Greek company, but they work in a remote working mode, i.e. they meet the conditions for the research of this work. At the same time, these are respondents from European countries, mainly from Poland, the Czech Republic, Greece, Germany, Portugal, and Romania.

The data from this research was collected using an online Google survey that was distributed through email accompanied by a brief explanation (with a link that directed to the compilation of a Google form approved by the head of surveyed company) from the 18th May till 9th June 2022. An online survey was chosen (on a voluntary base) as a means for collecting the data due to the company remote work environment and various locations internationally. As Evans and Mathur (2005) stated the suitable reason to utilize an online survey is when a large sample size is desired over a wide geographical area. It included 46 questions (7 open ended and 39 closed questions) asking IC constructs and remote work experience. Questions were modeled specifically on the areas highlighted in this study that may either mitigate or compound the effects of IC in remote work environment. The compilation time was estimated in approximately 15 minutes. Questionnaire was pre-tested with 3 respondents and redesigned in accordance with it. 5 point-Likert scale to measure satisfaction, positivity or negativity and 3 point-Likert and 5 point-Likert scale to measure agreement has been used.

In order to test seven hypotheses, the authors' set seven dimensions central to IC based on Ruck and Welch (2012) six dimensions. In this study, Pearson's chi-squared test has been used to analyse whether the characteristic values of the first variable are influenced by the characteristic values of the second variable and vice versa. The observed and expected frequency of 0.05 has been given. Statistical analyses were performed using Excel and SPSS.

3 RESULTS

As stated previously, this study will focus on the results of those surveyed who expressed some increase or decrease in their levels of internal communication, with discussion of neutral responses where necessary or of relevance to the study. These questions appeared at Questions 15 and 19 (identification), 33 and 35, (workplace relationships), 37 (channel) 42 and 46 (role), 41 and 46 (voice), 10 and 16 (performance), 28 and 37 (support).

IC is receiving focus and organizations are investing in the function (Chalmers, 2008; Helsby, 2009). The main role of IC is to improve corporate values by communicating effectively with the workforce, an important asset, and a sense of belonging.

19) How is your relationship with your colleagues?	15) My feeling of connectedness with my employer / team members				
	Neutral	Positive	Rather negative	Rather positive	Total
Average	42	24	0	0	66
%	63.64	36.36	0.00	0.00	
Extremely satisfied	0	59	7	45	111
%	0.00	53.15	6.31	40.54	
Total	42	83	7	45	177

Tab. 1 A sense of connection between employees and colleagues.
Source: Authors' work

3.1 Identification

- **H1: Relationships with colleagues related to feeling of connectedness with my employer / team members.**

In the question "How is your relationship with your colleagues?" we can only compare the values "Average" and "Extremely satisfied", since the value "dissatisfied" did not reach a sufficient frequency to participate in the test. Only 1.1% of employees were dissatisfied with their colleagues. Similarly, the choice "Negative" in the question "My sense of connection with the employer/colleagues" did not reach sufficient representation (0.6%), therefore this option cannot be tested.

It is also possible to read from Table No. 1 that we obtained a total of 177 answers to the question, although 182 respondents participated in the questionnaire, and 5 employees refused to answer the question. The test criterion value is 104.043, and the p-value of the test is 0.000. The p-value is lower than the chosen significance level of 0.05, so the dependence between relationships with colleagues and the feeling of connection with employees or team members has been confirmed. According to the table, those who are extremely satisfied with their relationships with colleagues have a statistically significantly higher likelihood of having a positive or rather positive relationship with their colleagues. From this, we can say that when employees feel connected to the company, either because they share the same vision as the company's leaders or because they feel that their co-workers have become like family, they are much less likely to want to look for employment elsewhere and thus are more productive.

Based on this data, fulfilling relationships within the work environment are interlinked with meaningful connections with co-workers. When employees feel connected to one another, they feel better. Connection seems to be the most crucial component because connected teams drive collaboration.

3.2 Workplace relationships

- **H2: Daily contact with your team leaders is related to being satisfied with the frequency of communication from leadership.**

More than 4/5 employees confirmed that they communicate with the team leader at least once a day. This majority is also satisfied with the frequency of communication they receive from management. The test criterion value is 156.865 and the p-value of the test is 0.000. The p-value is lower than the chosen significance level of 0.05, so the dependence between daily

contact with team leaders and satisfaction with the frequency of communication with management was confirmed. Those who do not have daily contact with team leaders are also less satisfied with the frequency of communication with management. This means that employees would like to be in contact with management more often (see the Table 2).

Corporate communication is vital element. Generally, communication with team leaders is a big challenge. Leaders typically hold the key to advancement within the organization. The highest managerial goal is to create a pleasant working atmosphere and find the most successful technique of motivating each employee, which is very difficult due to the diversity of employees and different personalities, but in this company this goal would be met.

3.3 Channel

- H**3: More than 40% of employees prefer to use Zoom as communication tool.**

Employees' channel preferences vary based on the activity and who they are interacting with. As audio and video communications have become more common in daily lives, people often start collaborations with the question: which video communications platform do you know and like to use? We will test this single hypothesis using the proportion test. The software provides only the p-value in the output. It is equal to 0.0379. It is therefore again smaller than the chosen level of significance, and we have thus managed to confirm that Zoom is the most popular communication channel for more than 40% of employees. In second place, e-mail is preferred by 30.77% of respondents, and the last chat is preferred by 5.49% of employees.

3.4 Role

- **H4: Chitchatting related to clarity of objectives.**

More than half (61.7%) of the research participants perceive Zoom as the majority source of information, another 21.7% slightly disagree. The remaining 16.6% of respondents reject this,

33) Do you have daily contact with your team leaders?	**35) How satisfied are you with the frequency of communication of leadership?**			
	Extremely satisfied	Neutral	Slightly / very satisfied	Total
Mostly disagree	0	0	14	14
%	0.00	0.00	100.00	
Slightly agree	0	0	17	17
%	0.00	0.00	100.00	
Mostly agree	0	25	52	77
%	0.00	32.47	67.53	
Completely agree	56	13	0	69
%	81.16	18.84	0.00	
Total	**56**	**38**	**83**	**177**

Tab. 2 Frequency of H2
Source: Authors' work

42) Most of the information is through chitchatting.	46) Are your work objectives clear each day? Each week?			
	Completely / mostly agree	Slightly / mostly agree	Completely agree	Total
Completely/mostly agree	0	21	5	26
%	0.00	80.77	19.23	
Slightly agree	5	34	0	39
%	12.82	87.18	0.00	
Mostly agree	3	18	51	72
%	4.17	25.00	70.83	
Completely agree	0	0	43	43
%	0.00	0.00	100.00	
Total	8	73	99	180

Tab. 3 Frequency of H4
Source: Authors' work

thus they receive information in another way. A more important finding is that 89% of employees are clear about their assigned daily/weekly tasks. The test criterion value is 108.854 and the p-value of the test is 0.000. The p-value is lower than the chosen significance level of 0.05, so the dependence between communication using chitchatting and understanding the goals every day / week was confirmed. Those who are not satisfied with the course of communication using normal calls are significantly more often not completely clear about the assignment of the task (clarity of goals). On the contrary, those who are very satisfied with informal conversations are clear about their daily goals. The table of observed frequencies are shown in Table 3.

41) Most work-related information from teammates and team leaders to me is transmitted through informal written channels (i.e., instant messages, chat).	46) Are your work objectives clear each day? Each week?			
	Completely / mostly agree	Slightly / mostly agree	Completely agree	Total
Completely / mostly agree	0	27	15	42
%	0.00	64.29	35.71	
Slightly / mostly agree	8	48	53	109
7.34	44.04	48.62		
Completely agree	0	0	31	31
%	0.00	0.00	100	
Total	8	75	99	182

Tab. 4 Frequency of H5
Source: Authors' work

3.5 Voice

- **H5: Communication using informal written channels is related to understanding daily/weekly goals.**

More than half (54.5%) of the participants confirmed that they receive work messages through informal written channels. 22.5% of respondents reject this statement only slightly. 23% mostly and completely disagree. This means that informal communication is accepted as casual communication between co-workers in the workplace. It is unofficial in nature and is based on informal social relationships that are formed in the workplace outside the normal hierarchy of the business structure. The test criterion value is 38.986 and the p-value of the test is 0.000. The p-value is lower than the chosen significance level of 0.05, so the dependence between communication using informal written channels (messages, chats) and the understanding of goals every day/week was confirmed. This type of communication is important in the workplace because it can help with employee morale and encourage a sense of belonging for both employees and clients or customers. Those who are not satisfied with the course of communication using informal written channels are significantly more often not completely clear about the assignment of the task (clarity of goals). On the contrary, those who are very satisfied with informal written communication are clear about their daily goals (see Table 4).

3.6 Performance

- **H6: The change in work productivity is related to the fact that employees enjoy remote work.**

With all the modern comforts of home vying for our attention, it would be understandable if employers saw a drop in productivity with remote workers. However, the opposite is true. In fact, 88% of employees seem to enjoy working remotely, with another 6.6% slightly disagreeing with this statement and 5.4% rejecting it. Almost half (49.5%) of respondents say their productivity has increased as a result of working from home. Only 7.1% of employees feel a decrease in work performance. The test criterion value is 137.556 and the p-value of the test is 0.000. The p-value is lower than the chosen significance level of 0.05, so the dependence between productivity and whether employees enjoy remote work has been confirmed. From Table 5, it can be said that those who enjoy working remotely increased their productivity significantly more often than those who did not enjoy working remotely.

3.7 Support

- **H7: The choice of the preferred communication channel is related to the number of requests for help.**

A somewhat disturbing finding emerged for the last hypothesis, where 27.1% of employees have difficulty asking for help. So, it seems that the open-door policy is only partially working. Workplaces reflect norms and standards set by management, which are then accepted (or not). Much of this depends on how these norms are enforced and demonstrated. Building a culture of helping is about the art of practicing what you preach. In this hypothesis, we tested whether these concerns are related to the communication channel. The test criterion value is 64.086 and the p-value of the test is 0.000. The p-value is lower than the chosen significance level of 0.05, thus the dependence between communication channel and concern about asking for help is confirmed. From Table 6, it can be seen that those who prefer chat as their communication channel are significantly more often not afraid to ask for help than those who use Zoom or e-mail.

10) I enjoy working remotely.	16) Has your work productivity changed when you have started work remotely?			
	Decreased	Increased	Same	Total
Completely / mostly disagree	0	5	5	10
%	0.00	0.00	50.00	
Slightly / mostly agree	0	10	74	84
7.34	0.00	11.90	88.10	
Completely agree	13	75	0	88
%	14.77	85.23	0.00	
Total	13	90	79	182

Tab. 5 Frequency of H6
Source: Authors' work

The authors' survey results confirm that it is increasingly important for IC professionals to consider the diversity of the internal team in the organization, as highlighted in previous research (Suh and Lee, 2016; Yeomans and FitzPatrick, 2017). Additionally, as highlighted by McQuail (2010) effective communication requires delivering the right channel, the right message, to the right audience, through the right channel at the right time.

The received data from the first hypothesis is in the vein of past studies that highlight that overall organizational efficacy and efficiency depend on its IC (Grimshaw and Mike, 2008; Hayase, 2009). Communication plays an essential role (Levinson, 1980) as further demonstrated. Interestingly, those who do not have daily contact with team leaders are also less satisfied with the frequency of communication with management. This means that employees would like to be in contact with management more often as confirmed by H2. Because as Men and

37) What communication channels work best for you.	28) I feel uncomfortable asking other for help?					
	Completely disagree	Mostly disagree	Slightly agree	Mostly agree	Completely agree	Total
Zoom	14	21	21	29	29	114
%	12.28	18.42	18.42	25.44	25.44	
E-mail	12	31	1	8	14	66
7.34	18.18	46.97	1.52	12.12	21.21	
Chat	12	20	3	0	1	36
%	33.33	55.56	8.33	0.00	2.78	
Undefined	3	12	8	0	8	31
%	9.68	38.71	25.81	0.00	25.81	
Total	41	84	33	37	52	247

Tab. 6 Frequency of H7
Source: Authors' work

Yue (2019) explained communication increases stakeholder awareness and satisfaction with what leaders are making and whether their decisions will mitigate tension and improve interpersonal relationships. Almost identical to Stauss and Hoffmann (2000) who stated that IC is practiced through top-down, bottom-up or horizontal communication. IC without internal communication digital channels does not work (Smith et al., 2015). Zoom is the most popular communication channel for more than 40% of employees. In second place, e-mail is preferred by 30.77% of respondents, and the last chat is preferred by 5.49% of employees. Almost similar to Friedl and Verčič (2011) who noted that millennial workforce prefer traditional media. This means that without these channels, it is impossible to match employees to the company priorities, goals and promote creativity. The difference between formal and informal communication lies in the material, not the instrument (Rodrigues et al., 2016; Welch and Jackson 2007). But, based on H4 and H5 data, those who are very satisfied with informal conversations and informal written communications are clear about their daily goals. Identical to Men (2014) who highlighted that the choice of medium for communication is important. According to Bočková and Lajčin (2021) job satisfaction, productivity, engagement and, with mandatory remote working during this pandemic, is more of a challenge for employers than ever before. Thus, H6 confirmed that those who enjoy working remotely increased their productivity significantly more often than those who did not enjoy working remotely. Similar to Dhanesh and Picherit-Duthler (2021) who highlight that social connection mediated the relationship between new ways of working and employee engagement. A somewhat disturbing finding emerged for the last hypothesis, where 27.1% of employees have difficulty asking for help. Probably, because as Zhao and Epley (2022) indicated those needing help consistently underestimated others' willingness to help, underestimated how positively helpers would feel, and overestimated how inconvenienced helpers would feel. Possible tool seems to be the modern technology because those who prefer chat as their communication channel are significantly more often not afraid to ask for help than those who use Zoom or e-mail. This result is in the vein of Jämsen et al. (2022) statement that communication technology played a crucial role in experiences. Obtained results confirm that "the most important driving force behind the modern economy today is creative, innovative and motivated employees" (Pilukienė, 2017, p. 223).

4 CONCLUSION

IC in the surveyed organization is viewed as important and as a specific segment within each element or task. There are several target groups. Firstly, managers, leaders and executives who are responsible for shaping organizational culture and strategies, especially policies for remote work. Secondly, human resource professionals as well as external communication specialists who design IC strategies and implement them within the company. Further, there are researchers and academics who study work dynamics and IC in remote environment. Finally, policymakers suggesting regulations for e-workers at the level of governments or the European Union.

Obtained data confirms the raising importance of IC, especially in times of crisis as employees considered as a homogenous.

The following research questions were outlined:

- How e-employees perceisve IC methods in their organization?
- They perceived IC as being tailored to their specific needs. But based on the data, it still invites us to re-consider the specific needs to adapt the content of the messages, their formats, and the best communication channels to these needs. E-employees who perceive communication as transparent are satisfied by the communication in their organization. It means that organization partly created an effective IC strategy. It has to be prepared and ready to react to sudden changes and circumstances. Creating an effective IC strategy

involves evaluating and reviewing the current strategy and identifying the strengths and weaknesses as demonstrated in the research data. Especially when employees are reluctant to ask for assistance.
- Which of the internal communication channels were used most effectively?
- Technology tools affect how e-workers feel about their organization and work. Communication is the most important predictor of feelings of value and engagement. Both formal and informal IC channels are the way by which e-employees communicate and connect. Zoom is the most popular communication channel for more than 40% of employees. In second place, e-mail is preferred by 30.77% of respondents, and the last chat is preferred by 5.49% of employees. E-mail seems not to be the most common way of connecting in the office anymore.

The study has found that the largest proportion of staff was positively affected in the areas of IC. There are recommendations to support IC within the surveyed company:

1. Enhance relationships among e-colleagues – to enccourage meaningful connections among e-workers through regular team-building activities and collaborative projects. Strengthening their relationships will enhance their sense of community with the company, reducing turnover rates and increasing productivity. They will feel more connected and more likely to cooperative effectively, which lead to higher performance and job satisfaction.
2. Increase daily contact with team leaders – to ensure that team leaders have daily check-ins with their teams through brief meetings or casual conversations to keep communication lines open. Regular interaction helps to osolve issues promptly, employee feel valued, which influences overall engagement and productivity.
3. Clarify objectives – support informal communication methods like casual chats or informal meetings to discuss daily and weekly goals because when employees understand their tasks, their productivity is higher and reducing errors.
4. Build a culture of support – promote a company culture where asking for help is encouraged and normalized. It can be achieved through training programs and clear communication from leadership side about the importance of seeking support. Asking for help can prevent burnout. Employees who feel support they are more engaged and loyal to the company.

The difference pre- and post-COVID-19 is that e-working shows that, if this ground work can be laid, value-enhancing, high skilled work can be done in terms of behaviours that matter for sustainability.

Regarding the limitations of this research, we can underline that the sample of our survey only concerned employees from organization based in Greece but located various European countries and in a context of unprecedented crisis. Authors were limited in studying certain fields on more detail. Additionally, the spoken predominant language in organization was English and it may cause problems of understanding and interpretation of sets of questions. This research only explores the role of IC in an organizational context (formal, casual, organized, or unplanned).

For future research, it would be interesting to extend this research by conducting a qualitative study. Other studies could aim to add more companies in different sectors. Future research may focus on the implications of feedback in IC.

Acknowledgements

The scientific paper was written within the framework of the projects KEGA 123PU-4/2023 and KEGA 065UK-4/2024 with title The Use of Artificial Intelligence in Human Resource Management.

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

AMBIKAPATHY, M. and ALI, A. 2020. Impact and Challenges towards employees work from home during Covid-19 (MCO) Period. *International Journal of Social Science Research*, 2(4), 97-107.

BEŇO, M. 2021. The Advantages and Disadvantages of E-working: An Examination using an ALDINE Analysis. *Emerging Science Journal*, 5, 11-20. https://doi.org/10.28991/esj-2021-sper-02

BENO, M. and CAGANOVA, D. 2023. Austrian future cubicle: commuting, e-commuting or both. *Acta Logistica*, 10(1), 135-140. https://doi.org/10.22306/al.v10i1.368

BOČKOVÁ, K. and LAJČIN, D. 2021. Home Office and Its Influence on Employee Motivation. *Journal of Management and Marketing Review*, 6(2) 94-109. https://doi.org/10.35609/gcbssproceeding.2021.12(3)

CAGÁŇOVÁ, D., ČAMBÁL, M. and WEIDLICHOVÁ LUPTÁKOVÁ, S. 2010. Intercultural Management – Trend of Contemporary Globalized World. *Electronics and electrical engineering*, 6(102), 51-54.

CAGÁŇOVÁ, D., PECHANOVÁ, Ľ. and HORŇÁKOVÁ, N. 2019. *Communication in Multicultural Teams in Industrial Enterprises*. Plzeň, Vydavatelství a nakladatelství Aleš Čeněk.

CAGÁŇOVÁ, D., ŠUJANOVÁ, J. and WOOLLISCROFT, P. 2015. *Multicultural and intercultural management within a global context*. Bratislava, Nakladateľstvo STU.

CHALMERS, S. 2008 (October). *The Changing Role of Internal Communications. Are Required Practitioner Competencies Keeping Pace with the Opportunities and Challenges?* https://eresearch.qmu.ac.uk/handle/20.500.12289/568

DHANESH, G. S. and PICHERIT-DUTHLER, G. 2021. Remote internal crisis communication (RICC) – role of internal communication in predicting employee engagement during remote work in a crisis. *Journal of Public Relations Research*, 33(5), 292-313. DOI:10.1080/1062726x.2021.2011286

EUROSTAT. 2022 (November 8). *Rise in EU population working from home*. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20221108-1#:~:text=The%20share%20of%20employed%20people,points%20between%202019%20and%202021

ERUM, H., ABID, G. and CONTRERAS, F. 2020. The calling of employees and work engagement: the role of flourishing at work. *Business, Management and Economics Engineering*, 18(1), 14-32. https://doi.org/10.3846/bme.2020.11430

EVANS, J. R. and MATHUR, A. 2005. The value of online surveys. *Internet Research*, 15(2), 195-219. https://doi/10.1108/10662240510590360/full/html

FLYNN, J. and MAZOR, A. H. 2017. T*he employee experience: Culture, engagement, and beyond, Global Human Capital Trends*. https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2017/improving-the-employee-experience-culture-engagement.html/#endnote-sup-6

FRIEDL, J. and VERČIČ, A. T. 2011. Media preferences of digital natives' internal communication: A pilot study. *Public Relations Review*, 37(1), 84-86. https://doi.org/10.1016/j.pubrev.2010.12.004

GRIMSHAW, J. and MIKE, B. 2008. How mature is your internal communication function? *Strategic Communication Management*, 12(3), 28-31.

HAYASE, L. K. T. 2009. Internal communication in organizations and employee engagement. *Digital Scholarship@UNLV*. https://digitalscholarship.unlv.edu/thesesdissertations/1176/

HELSBY, N. and CROTON, M. *2009. Internal communications – more to deliver? London: Watson Helsby* Consulting.

JÄMSEN, R., SIVUNEN, A. and BLOMQVIST, K. 2022. Employees' perceptions of relational communication in full-time remote work in the public sector. *Computers in Human Behavior*, 132, 107240. https://doi.org/10.1016/j.chb.2022.107240

KURKLAND, N. B. and BAILEY, D. E. 1999. Telework: The advantages and challenges of working here, there, anywhere, and anytime. *Organizational Dynamics,* 28(2), 53–68. https://doi.org/10.1016/S0090-2616(00)80016-9

LAL, B., DWIVEDI, Y. K. and HAAG, M. 2021. Working from Home During Covid-19: Doing and Managing Technology-enabled Social Interaction With Colleagues at a Distance. *Information Systems Frontiers,* 25, 1333-1350. https://doi.org/10.1007/s10796-021-10182-0

LEVINSON, S. C. 1980. Speech act theory: The state of the art. *Language teaching and linguistics: Abstracts*, 5-24.

MCDONOUGH, J. and MCDONOUGH, S. 1997. *Research Methods for English Language Teachers*. Routledge.

MCQUAIL, D. 2010. *McQuail's Mass Communication Theory*. Sage.

MEN, L. R. 2014. Strategic Internal Communication. *Management Communication Quarterly*, 28(2), 264–284. https://doi.org/10.1177/0893318914524536

MEN, L. R. and YUE, C. A. 2019. Creating a positive emotional culture: Effect of internal communication and impact on employee supportive behaviors. *Public Relations Review*, 45(3), 101764. https://doi.org/10.1016/j.pubrev.2019.03.001

OAK. 2022. (May 02). *16 Benefits of an Intranet You Can't Ignore 2021*. https://www.oak.com/blog/best-intranet-benefits/

QUEENSUNIVERSITY. 2023. *Communicating in the modern Workplace*. https://online.queens.edu/resources/infographic/communicating-in-the-workplace/

PILUKIENĖ, L. 2017. Importance of motivation and work pay of young employees in the value creation chain of a business company: assessment of changes in and formation of expectations. *Business, Management and Economics Engineering*, 15(2), 211-226. https://doi.org/10.3846/bme.2017.375

ROBERHALF. 2013. (October 22). *Press Releases. Survey: Human Resources Managers Cite Lack of Communication as Main Source of Low Employee Morale*. https://press.roberthalf.com/2013-10-22-Survey-Human-Resources-Managers-Cite-Lack-of-Communication-as-Main-Source-of-Low-Employee-Morale

RODRIGUES, A., AZEVEDO, C. and CALVO, V. 2016. Internal communication in organizations: practical instruments to help the shift change. *Millenium - Journal of Education, Technologies, and Health*, 01, 105–114. https://doi.org/10.29352/mill0201.09.00004

RUCK, K. and WELCH, M. 2012. Valuing internal communication; management and employee perspectives. *Public Relations Review*, 38(2), 294–302. https://doi.org/10.1016/j.pubrev.2011.12.016

SMITH, S. A., PATMOS, A. and PITTS, M. J. 2015. Communication and Teleworking: A Study of Communication Channel Satisfaction, Personality, and Job Satisfaction for Teleworking Employees. *International Journal of Business Communication*, 55(1), 44–68. https://doi.org/10.1177/2329488415589101

STAUSS, B. and HOFFMAN, F. 2000. Using business television. In: VAREY, R. J., LEWIS, B. R. (Eds.). *Internal Marketing: Directions for Management*. London, Routledge. 139-59.

SUH, T. and LEE, J. 2016. Internal audiences segmentation and diversity in internal communication. *Corporate Communications: An International Journal*, 21(4), 450–464.

WELCH, M. and JACKSON, P. R. 2007. Rethinking internal communication: a stakeholder approach. *Corporate Communications: An International Journal*, 12(2), 177–198. https://doi.org/10.1108/13563280710744847

WORLDBANK. 2021 (January). *Employment in services, female (% of female employment) (modeled ILO estimate)*. https://data.worldbank.org/indicator/SL.SRV.EMPL.FE.ZS

YEOMANS, L. and FITZPATRICK, L. 2017. Internal Communication. In: Exploring Public Relations. Pearson Education. https://eprints.leedsbeckett.ac.uk/id/eprint/3499/3/Internal%20Communication.pdf

ZHAO, X. and EPLEY, N. 2022. Surprisingly Happy to Have Helped: Underestimating Prosociality Creates a Misplaced Barrier to Asking for Help. *Psychological Science*, 33(10), 1708–1731. https://doi.org/10.1177/09567976221097615

Contact information

Michal Beno: email: michal.b@praguecityuniversity.cz

FOREIGN TRADE OF WINE AND FRESH GRAPES IN THE VISEGRAD COUNTRIES

Katarína Bírová[1], Patrik Rovný[1]

[1]Institute of Marketing, Trade and Social Studies, Faculty of Economics and Management, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 76 Nitra, Slovakia

ABSTRACT

The Visegrad countired, as countries of a single grouping that have pledged to help each other and to deepen their trade with each other. These are the countries of the Slovak Republic, the Czech Republic, Hungary and Poland. The main objective of the article is to identify the overall level of trade of the Visegrad countries in the commodity wine and fresh grapes concerning the World, the Visegrad countries as a whole and the individual countries of the Visegrad country. The analysis shall take into account the monetary clarification of trade volumes in millions of euros. The commodity structure is based on the international HS system. The main analysis is carried out through the basic indicators of turnover, trade balance and RCA (Revealed Comparative Advantage). Hungary has a comparative advantage in the wine trade at the World level but also within the Visegrad countries. The Slovak Republic has both comparative advantages and comparative disadvantages within the World and the Visegrad countries, but most of all with the Czech Republic and Poland. The Czech Republic showed comparative advantages only with Poland. Poland has no comparative advantages with any of the Visegrad countries.

Keywords: wine sector, Visegrad countries, foreign trade, export, import, comparative advantages, Revealed Comparative Advantage

JEL Code: Q170

1 INTRODUCTION

Wine is a globally popular beverage, but production is located in very few places. It is estimated that wine is produced in around sixty countries. Of these, many countries produce a negligible amount of wine. Countries that do not have the land area or climatic conditions to do so produce regional high-quality wines, but the quantity does not cover foreign markets and exports alone (Ayuda et al., 2020; Anderson and Pinilla, 2018; Rocchi and Stefani, 2002). The largest European traditional wine producers include coun- tries such as Italy, France, Spain and countries such as the USA, Australia, Chile, South Africa and New Zealand, which produce around 80–90% of the World volume. Interestingly (Dal Bianco et al., 2015). Claim

https://doi.org/10.11118/978-80-7509-990-7-0036

that almost half of the wines produced in wine producing countries are consumed outside their country of production (Holmes and Anderson, 2017). The biggest turning point in wine trade took place in the 1980s, when the potential for wine demand in foreign markets became apparent to countries due to changes in wine policy that affected wine-producing countries in Europe (Cembalo et al., 2014; (Morrison and Rabellotti, 2014; Aizenman and Brooks, 2008; Dion, 1952). However, according to BALOGH et al. (2018), the global share of European wine production is declining. For more efficient foreign trade, grape growers and wine producers need to share uniform and tailored rules and policies (Fleury and Fleury, 2003; Ring and Van de Ven, 1994). The role of these policies is to improve international relations with this commodity (Borák, J. and Vacek, T. 2018; Kuemmerle, 1999; Woodward and Rolfe, 1993). The Visegrad countries are the Slovak Republic, the Czech Republic, Poland and Hungary. It is a specific group of countries located in the centre of Europe and character- ized by their historical links and political-economic relations. The countries have undergone changes in recent years which have had an impact on the structure of agriculture, particularly after the collapse of the former socialist system or the Eastern bloc. Restructuring has also affected foreign trade and, consequently, EU accession has brought new opportunities for agrarian foreign trade. However, even without the countries' accession to the EU, the Visegrad countries represent important foreign trade partners for each other in all economic areas (Svatoš and Smutka, 2014; Smutka, 2014; Smutka et al., 2013; Bojnec and Ferto, 2009; Basek and Kraus 2009; Bussiereet al., 2005). The agrarian foreign trade between the Slovak Republic and the Czech Republic is specific in that for many years these countries formed a single state and have existed separately for a relatively short time (Bielik et al., 2013; Urban 1991). For both countries, wine is a popular alcoholic beverage and annual per capita consumption averages 18 liters. However, Slovakia is a small wine producer and from a European perspective represents 0.2% of the total European production overall production is still decreasing and thus the negative balance in this sector is deepening (Judinová, 2018) The Czech Republic is no better off in the long term the country shows a negative balance, although in volume terms exports and imports are increasing every year, but imports still far exceed exports (Borák and Vácek, 2018). Hungary has been one of the top wine-producing countries since almost the 18th century and the largest wine producer, which is why it boasts an active trade deficit. Hungary has favourable conditions for viticulture although all countries are currently struggling with vineyard decline (State of the World Vine and Wine Sector in 2022, 2022; Bozzay, 2021; Pallás, E. I. et al., 2016). Poland does not naturally have good climatic conditions but is trying to develop and noble vineyards so that it can be grown there as well. Foreign trade is therefore more a matter of importing wine from various countries around the World and thus has a strongly passive trade balance. However, we can say that Poland is also a country where wine as an alcoholic beverage is very popular (Exporting wine to the Polish market, 2016; The alcoholic beverages market in Poland, 2014). The main objective of this article is to compare the foreign trade and comparative advantages of the Visegrad countries in the commodity group wine and fresh grapes.

2 METHODOLOGY AND DATA

The reference period, changes and developments in research is the span of years from 2000 to 2022. The main commodity structure has been defined according to the internationally recognized HS system, namely HS code number 2204 in full: Wine of fresh grapes. incl. fortified wines; grape must. partly fermented and of an actual alcohol- ic strength of > 0.5% vol or grape must with added alcohol of an actual alcoholic strength of > 0.5% vol (Database - Eurostat, 2024). In the first phase of the analysis, we analyzed the foreign trade of the individual countries with the World and within the Visegrad countries. In the second phase, we analyzed the basic foreign trade and the comparative advantages revealed in the trade of the Visegrad

countries with each other as follows: SK-CZ, SK-HU, SK-PL, HU-CZ, HU-PL, CZ-PL. The data for the individual analyses were taken from Eurostat public databases. Values are expressed in millions of euros. The following methods and calculations were used in the paper:

Base index – this is the evolution of a number over time. The method was used in the analysis of the development of foreign trade, i.e. Imports and Exports in individual countries with the World and the Visegrad countries, where we used the following relationship (1) (Matejková et al., 2013).

(1)

$$IB(t) = Qt / Q0$$

Where:

IB(t).....the base index of the selected current period, Qt – the value of the current period, Q0 – the value of the base period

Turnover and trade balance – are the basic indicators of foreign trade. The trade balance can take two forms, namely active and passive. Active, if it is a state when a country's exports are higher than imports and passive if imports are higher than exports. The indicators are calculated according to the following relationships.

(2)

$$Turnover = Export + Import$$

(3)

$$Trade\ balance = Export – Import$$

Revealed comparative advantage (RCA) index – is used to determine the relative advantage or disadvantage in a commodity group in a given country (Revealed Comparative Advantage, 2021).

(4)

$$RCA = \ln\left(\frac{\frac{x_{ij}}{m_{ij}}}{\frac{X_j}{M_j}}\right)$$

Where:

x_{ij} .....the value of exports of the i-th product group of the analysed industry of country j, in our case it was the commodity group HS 2204;

m_{ij} .....the value of imports of the i-th product group of the analysed industry of country j, in our case it was the com- modity group HS 2204;

X_j.....the value of total exports from country j, in our case the total exports consisted of agri-food exports of the country, which were compiled from the commodity groups HS 01-24.

M_j.....value of total imports into country j, in our case total imports were agri-food imports of the country, which were compiled from HS commodity groups 01-24.

For the RCA indicator: RCA < 0 - induces comparative disadvantage, RCA > - induces revealed comparative advantages,

3 RESULTS

3.1 Foreign trade with Wine of fresh grapes of the Visegrad countries.

3.1.1 Foreign trade with Wine of fresh grapes in Slovak Republic

Slovakia is the smallest of the Visegrad countries. The overview of foreign trade (Table 1) reflects a fluctuating situation over the period under review. It is evident that the country is dominated by total imports of wine and fresh grapes, resulting in a passive trade balance for almost the entire period. The only exceptions are 2002 and 2003, when the country had an active trade balance. Within the framework of the Visegrad countries relations, the country trades extensively with the Visegrad countries. In this respect, the active trade balance has been more frequent from 2000 to 2003. This was followed by a passive balance until 2017 when the active and passive balances alternated. The curves show us that the country trades to a high extent in exports only with the countries of the Visegrad countries, but most importantly with the Czech Republic and Hungary. In terms of imports, the most important countries are Italy, Hungary and Germany.

PRODUCT	Wine of fresh grapes. incl. fortified wines; grape must. partly fermented and of an actual alcoholic strength of > 0.5% vol. or grape must with added alcohol of an actual alcoholic strength of > 0.5% vol.							
CODE	2204							
FLOW	IMPORT	EXPORT	Turnover	Balance	IMPORT	EXPORT	Turnover	Balance
INDICATOR	Value in million €				Value in million €			
PERIOD	Foreign trade Slovakia between World				Foreign trade Slovakia between V4			
2000	6.153	4.313	10.47	-1.84	2.247	3.994	6.24	1.75
2005	19.364	5.435	24.80	-13.93	8.151	5.180	13.33	-2.97
2010	67.963	10.901	78.86	-57.06	41.551	10.043	51.59	-31.51
2015	63.450	16.231	79.68	-47.22	29.738	14.988	44.73	-14.75
2022	69.266	25.840	95.11	-43.43	29.143	23.384	52.53	-5.76
Basic index 2000/2005	3.15	1.26	2.37	7.57	3.63	1.30	2.14	-1.70
2005/2010	3.51	2.01	3.18	4.10	5.10	1.94	3.87	10.60
2010/2015	0.93	1.49	1.01	0.83	0.72	1.49	0.87	0.26
2015/2022	1.09	1.59	1.19	0.92	0.98	1.56	1.17	0.39
2000/2022	11.26	5.99	9.09	23.60	12.97	5.85	8.42	-3.30

Tab. 1 Perview table of Slovak Republic foreign trade in million €
Source: Eurostat, own calculations (2024)

PRODUCT	Wine of fresh grapes. incl. fortified wines; grape must. partly fermented and of an actual alcoholic strength of > 0.5% vol. or grape must with added alcohol of an actual alcoholic strength of > 0.5% vol.							
CODE	2204							
FLOW	IMPORT	EXPORT	Turnover	Balance	IMPORT	EXPORT	Turnover	Balance
INDICATORS	Value in million €				Value in million €			
PERIOD	Foreing trade Czechia between World				Foreing trade Czechia between V4			
2000	29.068	2.566	31.63	-26.50	11.706	1.310	13.02	-10.40
2005	86.923	4.960	91.88	-81.96	13.998	4.409	18.41	-9.59
2010	142.357	21.137	163.49	-121.22	142.357	21.137	48.95	-9.94
2015	175.850	26.529	202.38	-149.32	175.850	26.529	46.94	-6.18
2022	270.663	38.879	309.54	-231.78	50.013	10.536	60.55	-39.48
Basic index 2000/2005	2.99	1.93	2.90	3.09	1.20	3.37	1.41	0.92
2005/2010	1.64	4.26	1.78	1.48	10.17	4.79	2.66	1.04
2010/2015	1.24	1.26	1.24	1.23	1.24	1.26	0.96	0.62
2015/2022	1.54	2.40	1.53	1.55	0.28	0.70	1.29	6.39
2000/2022	9.31	9.01	9.79	8.75	4.27	2.64	4.65	3.80

Tab. 2 Overview table of Czech Republic foreign trade in million €
Source: Eurostat, own calculations (2024)

The year 2004 became an important milestone for Slovakia's foreign trade when the country joined the European Union. Although the active balance is still recorded in the years 2000–2003, the total volume of imports increased significantly by up to 11 times over the whole period under review. Exports over the whole period increased only 6 times, which is almost half as much. Trade with the Visegrad coutries also had a rising character, with imports increasing 13 times and exports only 6 times. The rapidly increase in trade volumes took place in the first 5 years after EU accession.

3.1.2 Foreign trade with Wine of fresh grapes in the Czech Republic

The Czech Republic is a very important partner in foreign trade with Slovakia when in the past they formed one country. (Table 2) shows the development of trade in wine and fresh grapes for the period under review both at the level of foreign trade with the World and with the Visegrad countries. Looking at the trends, the total imports to the Czech Republic differ significantly, starting at around 29 million euros in 2000 and increasing up to 9.31 times by 2022. Currently, imports of the commodity from all over the World amount to 270 million euros. Paradoxically, the Czech Republic has huge imports from the World, but if we compare this with the volume of imports from the Visegrad countries, this volume represents only 18%. The highest growth in imports from the Visegrad countries was recorded between 2005 and 2010 when imports increased 10 times. The biggest breakthrough came in 2013 when imports started to decline rapidly. Total exports are again very low and have not exceeded 50 million euros over the whole period under review. However, we can state that the trend has a rising character. And during the whole period under review, it has

increased 9 times, most significantly between 2005 and 2010. The worst is the export to the Visegrad countries, which has increased 2.6 times over the whole period, but since 2013 it has had a significantly de- creasing character. Overall, the Czech Republic shows a passive trade balance.

3.1.3 Foreign trade with Wine of fresh grapes in Hungary

Hungary is the second largest of the Visegrad countries in terms of area, but Hungary is considered a wine-growing and wine-producing country, as evidenced by its foreign trade. The country is the only one that has long been exporting a commodity in excess of its imports (Table 3). Exports of the commodity have shown an increasing trend over the long term. From the early 2000s to 2022, exports have increased almost 2 times. From this perspective, it may appear as if exports are stagnating or increasing only modestly. This is not the case, but rather a long-term stable situation, due to the fact that the average volume of Hungarian exports over this period is EUR 77 million. However, if we look at exports from the point of view of the Visegrad countries, we can state that exports to the Visegrad countries account for only 29% of the country's total exports, and the country, therefore, exports 71% of its production to other countries in the World. Thus, since the Country is sufficiently self-sufficient in this sector and has long been showing an active trade balance which has increased 1.77 times in the last years 2015 to 2022 this growth has been the steepest of the whole period under review. Total imports into the country have shown a volatile pattern. With commodity imports being very low in the first five years, growing 7.7 times over the entire period under review, a faster growth rate than exports. The highest change in imports

PRODUCT	Wine of fresh grapes. incl. fortified wines; grape must. partly fermented and of an actual alcoholic strength of > 0.5% vol. or grape must with added alcohol of an ac- tual alcoholic strength of > 0.5% vol.							
CODE	2204							
FLOW	IMPORT	EXPORT	Turnover	Balance	IMPORT	EXPORT	Turnover	Balance
INDICATORS	Value in million €				Value in million €			
PERIOD	Foreing trade Hungary between World				Foreing trade Hungary between World			
2000	2.812	69.400	72.21	66.59	0.000	14.573	14.57	14.57
2005	10.903	57.929	68.83	47.03	0.025	12.315	12.34	12.29
2010	15.127	73.447	88.57	58.32	0.210	21.956	22.17	21.75
2015	19.144	74.051	93.20	54.91	0.323	21.266	21.59	20.94
2022	21.570	118.757	140.33	97.19	1.113	38.388	39.50	37.27
Basic index 2000/2005	3.88	0.83	0.95	0.71	0.00	0.85	0.85	0.84
2005/2010	1.39	1.27	1.29	1.24	8.37	1.78	1.80	1.77
2010/2015	1.27	1.01	1.05	0.94	1.54	0.97	0.97	0.96
2015/2022	1.13	1.60	1.51	1.77	3.45	1.81	1.83	1.78
2000/2022	7.67	1.71	1.94	1.46	0.00	2.63	2.71	2.56

Tab. 3 Overview table of Hungary foreign trade in million €
Source: Eurostat, own calculations (2024)

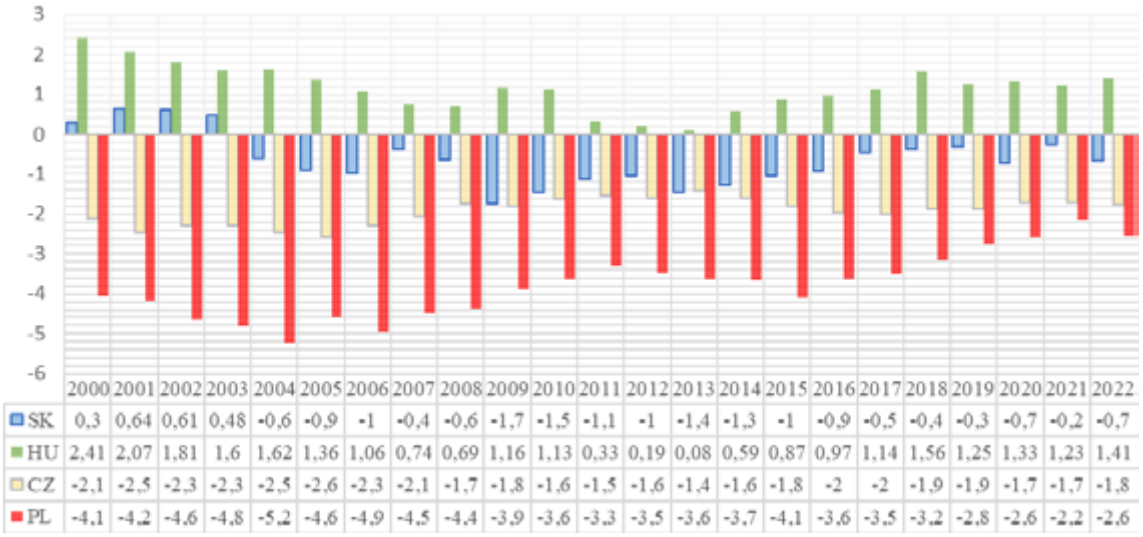
PRODUCT	Wine of fresh grapes. incl. fortified wines; grape must. partly fermented and of an actual alcoholic strength of > 0.5% vol. or grape must with added alcohol of an actual alcoholic strength of > 0.5% vol.							
CODE	2204							
FLOW	IMPORT	EXPORT	Turnover	Balance	IMPORT	EXPORT	Turnover	Balance
INDICATORS	Value in million €				Value in million €			
PERIOD	Foreing trade Poland between World				Foreing trade Poland between V4			
2000	53.885	0.765	54.651	-53.120	5.619	0.032	5.65	-5.59
2005	90.576	1.196	91.772	-89.381	5.557	0.284	5.84	-5.27
2010	153.705	5.034	158.739	-148.671	5.009	1.739	6.75	-3.27
2015	222.258	5.551	227.809	-216.707	8.022	0.068	8.09	-7.95
2022	380.584	43.374	423.958	-337.211	6.748	0.001	6.75	-6.75
Basic index 2000/2005	1.68	1.56	1.68	1.68	0.99	8.82	1.03	0.94
2005/2010	1.70	4.21	1.73	1.66	0.90	6.13	1.16	0.62
2010/2015	1.45	1.10	1.44	1.46	1.60	0.04	1.20	2.43
2015/2022	1.71	7.81	1.86	1.56	0.84	0.01	0.83	0.85
2000/2022	7.06	56.67	7.76	6.35	1.20	0.02	1.19	1.21

Tab. 4 Overview table of Poland foreign trade in million €
Source: Eurostat, own calculations (2024)

occurred between 2000–2005 and the highest import in value terms was in 2013 when the import volume reached 33 million euros. The country's self-sufficiency in this sector is also reflected in the country's imports from the Visegrad countries, which are very low. Ie no imports in 2000 and 2001.

3.1.4 Foreign trade with Wine of fresh grapes in Poland

Poland boasts the largest land area of all the countries of the Visegrad coutries, but in terms of its geographical location on the planet, it does not have very good climatic conditions for the production of wine and fresh grapes. The country has therefore had a long-standing passive balance of trade, and each year the passive balance is widening. (Table 4) The country's total imports of the commodity have increased 7 times from the beginning to the current year 2022. Compared to the Visegrad countries, Poland has the highest imports in 2022 at 380 million euros. This may point to an increase in wine con- sumption in Poland. In terms of imports from the Visegrad countries, we can state that Poland imports the commodity to the country only to a small extent, with an average of EUR 7 million over the whole period under review. Compared to total imports, imports from the Visegrad represent 4% for the whole period. Within the total exports, there has been a long-term growth, and it has increased up to 57 times over the whole period and the significant years for growth were from 2015 to the present. In terms of commodityexports to the Visegrad countries, this is not such a progression, and we can state that the highest growth was recorded from 2000 to 2010. Thus, Poland is looking for other countries to export its commodity.

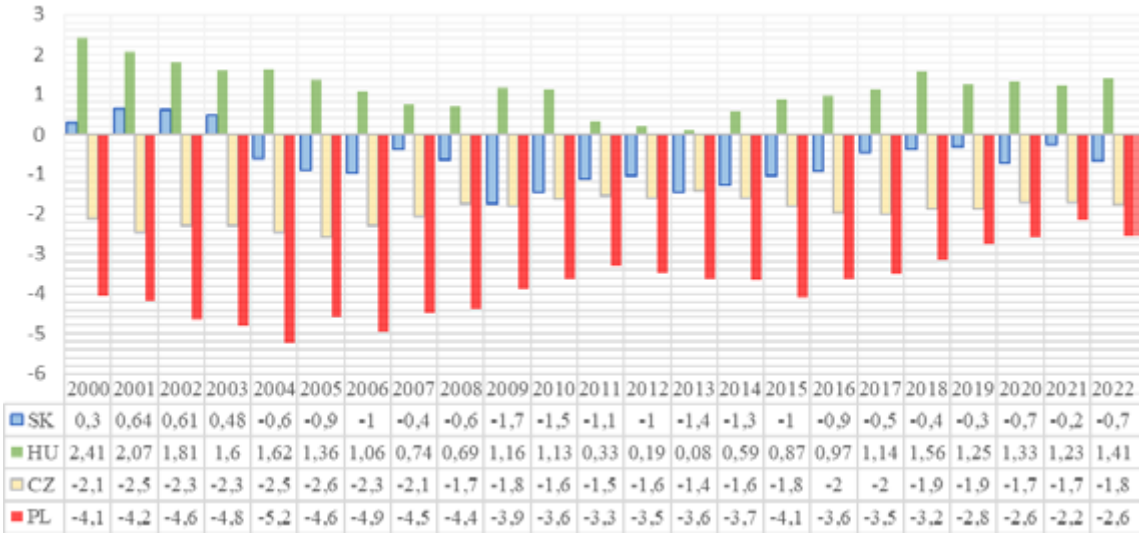


	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
SK	0,3	0,64	0,61	0,48	-0,6	-0,9	-1	-0,4	-0,6	-1,7	-1,5	-1,1	-1	-1,4	-1,3	-1	-0,9	-0,5	-0,4	-0,3	-0,7	-0,2	-0,7
HU	2,41	2,07	1,81	1,6	1,62	1,36	1,06	0,74	0,69	1,16	1,13	0,33	0,19	0,08	0,59	0,87	0,97	1,14	1,56	1,25	1,33	1,23	1,41
CZ	-2,1	-2,5	-2,3	-2,3	-2,5	-2,6	-2,3	-2,1	-1,7	-1,8	-1,6	-1,5	-1,6	-1,4	-1,6	-1,8	-2	-2	-1,9	-1,9	-1,7	-1,7	-1,8
PL	-4,1	-4,2	-4,6	-4,8	-5,2	-4,6	-4,9	-4,5	-4,4	-3,9	-3,6	-3,3	-3,5	-3,6	-3,7	-4,1	-3,6	-3,5	-3,2	-2,8	-2,6	-2,2	-2,6

Fig. 1: Revealed Comparative Advantage All Visegrad countries with World
Source: Eurostat, own calculations (2024)

3.2 Revealed Comparative Advantage Wine of fresh grapes on Visegrad countries

Like every country, the individual Visegrad countries have their stronger and weaker sectors. For a better understanding of the position of each country's foreign trade within the Visegrad countries wine and fresh grape industry with the World and also the Visegrad countries, the paper highlights the comparative advantages within the RCA index analysis. The analysis of the wine and fresh grape trade reveals comparative ad- vantages in this sector in Hungary (Figure 1). This points to the findings that Hungary is indeed a strong country in this sector and during the period considered it only had a comparative advantage, while in 2000 and



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
SK	0,3	0,64	0,61	0,48	-0,6	-0,9	-1	-0,4	-0,6	-1,7	-1,5	-1,1	-1	-1,4	-1,3	-1	-0,9	-0,5	-0,4	-0,3	-0,7	-0,2	-0,7
HU	2,41	2,07	1,81	1,6	1,62	1,36	1,06	0,74	0,69	1,16	1,13	0,33	0,19	0,08	0,59	0,87	0,97	1,14	1,56	1,25	1,33	1,23	1,41
CZ	-2,1	-2,5	-2,3	-2,3	-2,5	-2,6	-2,3	-2,1	-1,7	-1,8	-1,6	-1,5	-1,6	-1,4	-1,6	-1,8	-2	-2	-1,9	-1,9	-1,7	-1,7	-1,8
PL	-4,1	-4,2	-4,6	-4,8	-5,2	-4,6	-4,9	-4,5	-4,4	-3,9	-3,6	-3,3	-3,5	-3,6	-3,7	-4,1	-3,6	-3,5	-3,2	-2,8	-2,6	-2,2	-2,6

Fig. 2: Revealed Comparative Advantage All Visegrad countries with Visegrad countries
Source: Eurostat, own calculations (2024)

	RCA					
	SK and CZ	SK and HU	SK and PL	HU and CZ	HU and PL	CZ and PL
2000	1.7047	-	-	-	-	1.1562
2001	2.4442	-3.8055	-	-	-	2.3654
2002	1.8152	-3.5791	3.3989	4.7909	6.6321	2.5337
2003	1.5800	-5.4325	4.8679	10.1738	-	3.6741
2004	0.7016	-7.7189	1.0632	3.1699	13.2652	0.9677
2005	0.6726	-8.7172	2.2488	5.5373	-	-1.4680
2006	0.3074	-2.8330	-	7.4489	8.0576	6.3893
2007	0.6826	-1.3864	2.8107	2.3512	-	1.1028
2008	0.5056	-2.7608	-	2.9141	-	2.4203
2009	-0.9103	-4.5449	-	3.6465	-	0.4477
2010	-0.4001	-6.7600	-	4.0287	-	0.5421
2011	0.2997	-5.7980	-1.6205	3.1950	-	0.9653
2012	0.3323	-3.2083	-1.6104	4.3892	7.5038	0.7712
2013	0.4188	-7.0361	-7.3034	3.7707	-	1.1660
2014	0.0589	-5.1556	1.6011	4.3488	-	1.2187
2015	0.5183	-6.4441	-0.5837	3.7499	-	3.0708
2016	1.3426	-6.3517	0.3497	3.7477	5.9141	5.1214
2017	1.9268	-6.7412	-2.6252	2.6152	7.0671	4.8269
2018	2.0394	-9.3880	-0.5006	2.7850	6.8813	5.3273
2019	2.1605	-5.2989	0.5863	3.0332	6.6372	3.5333
2020	2.0437	-4.8968	3.1126	5.0975	6.8119	6.5194
2021	2.5801	-5.5398	2.5114	7.2606	-	4.8766
2022	1.8525	-10.6514	-2.2190	6.3940	-	4.4866

Tab. 5 Summary table of the RCA indices of the Visegrad countries against each other
Source: Eurostat, own calculations (2024)

2001 the country had no imports of the commodity into the country. Slovakia, despite being the smallest country in terms of size in this sector, had a comparative advantage in this sector until 2003. The second weak- est country in terms of comparative advantage is the Czech Republic. Paradoxically, the Czech Republic is well known for its wine regions and high production in terms of wine and has very high imports, which may indicate a high consumption of wine in the Czech Republic. However, the RCA index has a long-term com The RCA index reaches a nega- tive trend while. This indicates that Imports largely exceed exports. Therefore, the coun- try will never achieve a comparative advantage in this sector. parative disadvantage with the highest index value reaching -2.48 in 2021. The worst-performing

country in the survey is Poland. Poland is generally a large country but with different climatic con- ditions where grapes and wine do not fare well. The RCA index is negative here, with Imports far exceeding Exports, and this country will never have a comparative ad- vantage in this sector. The lowest value of the RCA index in Poland was achieved in 2004 (volume -5.23).

In (Figure 2) the next part of the RCA index research, we focused on a narrower selection of countries with which the Visegrad countries trade. We compare foreign trade between individual Visegrad countries with the foreign trade of the Visegrad countries as a whole. Closer analysis reveals that Hungary has high values in the uncovered comparative advantage in comparison the Visegrad countries, with the highest value of comparative advantage being 11.48 in 2003. However, the RCA index itself indicates a downward trend from the early 2000s to the present. The lowest value is in 2009 at -1.35, which could be related to adverse climate changes or the impact of the economic crisis in 2008. Otherwise, Slovakia maintains a very conservative line.

The Czech Republic is slightly worse off, showing negative values throughout the period under review, which represents a long-term comparative disadvantage, and the range of values is -2.73. Between 2007 and 2017, the performance was closer to 0 comparative advantage. Again, the weakest region is Poland with long-term and significantly negative RCA index values reaching as low as -10.20.

3.3 Revealed Comparative Advantage Wine of fresh grapes between the Visegrad countries

A closer look at the trade of the Visegrad countries with each other yields some interesting findings, which are presented in (Table 5). In the table, the yellow color represents zero imports and the blue color represents zero exports in the foreign trade of each country. The main objective was to find out the comparative advantages between the individual Visegrad countries. Significant and at the same time positive values of comparative advantages were shown by the Slovak Republic with the Czech Republic. It represents the largest share of the Slovak Republic exports to the Czech Republic. The only exceptions were the years 2009 and 2010 where a comparative disadvantage was demonstrated. However, the values largely exceed the level of 1. Therefore, we can assess that we are an important supplier of this commodity to the Czech Republic. In the second comparison between Slovakia and Hungary, we reach the opposite values, i.e. we reach a comparative disadvantage. As in the previous analyses, it emerges that Hungary is the dominant player in the Visegrad countries in the foreign trade in wine and fresh grapes. In 2022, Slovakia achieved the highest comparative disadvantage in comparison with Hungary at -10.65.

Poland is in third place in the Slovak Republic comparison. In this analysis, trade is very volatile and there were no exports of the commodity to Poland between 2008 and 2010.

Conversely, in 2000–2001 and 2006, there were no imports from Poland. All in all, Slovakia has a high comparative advantage vis-à-vis Poland. The relationship between Hungary and the Czech Republic has the same character as with Slovakia. Hungary's comparative advantage over the Czech Republic goes well beyond Hungary and Hungary is an important supplier of commodities to the Czech Republic. The relationship between Hungary and Poland presents only two possibilities in 2002, 2004, 2012 and 2016–2020 Hungary had a comparative advantage vis-à-vis Poland and in the other years, there were no imports from Poland to Hungary between the countries. In this case, Hungary achieved the highest RCA index value of 13.27. The last of the relationships is the Czech Republic and Poland. The RCA indices for the whole period show a comparative advantage on the side of the Czech Republic with the highest value in 2019 namely 5.33. The only exception was in 2005 when, on the contrary, Poland had a com- parative advantage.

4 DISCUSSION AND CONCLUSIONS

The Visegrad countries show different behavior of foreign trade in wine and fresh grapes. This is due to the fact that each country has a different size and agricultural con- ditions for grape and wine production. As evidenced by the State of the World Wine and Wine Sector in 2022 by Bozzay (2021) and Pallás, E. I. et al, (2016) Hungary is a country characterized by its wine-growing regions, and this is evidenced by its foreign trade in this commodity. We have observed a long-term active trade balance and high exports of the country, which have been moving more slowly over the whole period under review, with an increasing character, which was only 1.71 times over the twelve years. However, this is indicative of the stability of foreign trade in this commodity. As far as total exports to the Visegrad countries are concerned, this dynamic has been slightly more pronounced at 2.63 times from the initial year to the current year 2022. Imports are much lower as compared to exports although a more dynamic growth of imports has been recorded over the period under review at 7.67 times. Within the Visegrad countries, Hungary did not import anything into the country from 200 to 2001 and imports took off especially in 2005–2010. The RCA index indicates that the country has high compara- tive advantages both at the World level and for the Visegrad countries. Slovakia, as the smallest country in the analysis, surprised and came out of all the countries, that it trades mainly with the Visegrad countries. The trade balance turned out positive especially in the first years 2000–2003. The dynamics of total imports was at the level of 11.26 times from the beginning of 2000 to the present, and so were the imports of the Visegrad countries, which increased 12.97 times. Exports have increased approximately 6 times over the whole period, both to the World and within the Visegrad countries. Slovakia also surprised in comparative advantages in relation to the Visegrad countries, namely with the Czech Republic and Poland. The same situation is described by Judinová, (2018) and Rogovská (2018). The Czech Republic, also a major wine region, has long shown a passive trade balance. Comparative advantages have been revealed only with the country Exporting wine to the Polish market (2016) The alcoholic beverages market in Poland (2014) Poland and towards the other Visegrad countries and the World only a comparative disadvantage in this commodity has been revealed as evidenced by Borák and Vácek (2018). The steepest trend was in the Czech Republic's total imports of wine and fresh grapes into the country which increased 9.31 times over the period, as well as exports themselves increased 9 times. Exports to the Visegrad countries were less dynamic and increased only 2.64 times, but this is once as much as imports which increased 4.27 times. Poland, as the largest country in terms of area, has slightly unfavorable climatic conditions to produce grapes and wine, which is why in the analysis the balance of trade has long been negative and deeply in negative figures. This is why the country's imports are so high. The analysis showed that the country trades more with other World wine-producing countries than with the Visegrad countries. Negative values were also found for the comparative advantages and disadvantages with the Visegrad countries. However, the dynamics of foreign trade shows that the country is interested in consuming wine because imports have increased 7 times in the period under review from a World perspective. In terms of the Visegrad countries, imports have increased only by a marginal 1.20 times. However, Poland is keen to develop its wine industry, and this is particularly evidenced by exports, which have increased by a factor of 56.67 Worldwide. From the point of view of the Visegrad countries, exports are more negative, and this is because the dynamics were shown from 2000 until 2010, after which exports to these countries began to decline and, as a result, exports to these countries increased by a factor of only 0.02.

REFERENCES

AIZENMAN, J., and BROOKS, E. 2008. Globalization and taste convergence: the cases of wine and beer. *Review of International Economics*, 16(2): 217-233. https://doi.org/10.1111/j.1467-9396.2007.00659.x

ANDERSON, Y., and PINILLA, V. 2018. *Wine Globalization: A New Comparative History*. Cambridge University Press New York. 13(1): 566.

AYUDA, M., FERRER-PÉREZ, H., and PINILLA, V. 2020. Explaining World wine exports in the first wave of globalization, 1848–1938, *Journal of Wine Economics*, 15(3): 263– 283.

BALOGH, J. M., and JÁMBOR, A. 2018 The Role of Culture, Language and Trade Agree- ments in Global Wine Trade. *AGRIS on-line Papers in Economics and Informatics*, 10(3), 17-29. https://doi.org/10.7160/aol.2018.100302

BAŠEK, V. and KRAUS J. 2009. Czech foreign agricultural trade after joining the European Union. *Agric. Econ. –Czech.,* 55(3): 583–595.

BIELIK, P., SMUTKA, L., SVATOŠ, M., and HUPKOVÁ, D., 2013. Czech and Slovak agricul- tural foreign trade - two decades after the dissolution, *Zemědělská Ekonomika*, 59(10): 441–453. https://doi.org/10.17221/26/2013-agricecon

BOJNEC, Š. and FERTO I. 2009. Agro-food trade competitiveness of Central European and Balkan countries. *FoodPolicy*, 34(5): 417–425.

BORÁK, J. and VACEK, T. 2018. Czech Foreign Wine Trade – Comparative Advantages Distribution in Relation to the European Union, *AGRIS on-line Papers in Economics and Informatics*, 10 (3): 31-43. https://doi.org/10.7160/aol.2018.100303

BOZZAY, R., 2021. Hungarian Wine Export in the Late Middle Ages and Early Modern Time to the Low Countries. *Czech and Slovak Journalof Humanities*, 2(1): 12-18.

BUSSIÈRE, M., FIDRMUC, J., and SCHNATZ, B. 2008. EU Enlargement and Trade Integra- tion: Lessons from a Gravity Model. *Review of Development Economics*, 12(3): 562– 576.

CBI. 2016. Exporting wine to the Polish market. *CBI.eu* [Online]. Available: https://www.cbi.eu/sites/default/files/pdf/research/1052.pdf [Accessed: 2024-01-05].

CEMBALO, L., CARACCIOLO, F., and POMARICI, E. 2014. Drinking cheaply: the demand for basic wine in Italy. *Australian Journal of Agricultural and Resource Economics*, 58(3): 374-391. https://doi.org/ 10.1111/1467-8489.12059

DAL BIANCO, A., BOATTO, V. L., CARACCIOLO, F. and SANTERAMO, F. G. 2015 Tariffs and non-tariff frictions in the World wine trade. *European Review of Agricultural Economics*, 49(1): 31-57. https://doi.org/10.1093/erae/jbv008

DATABASE - EUROSTAT. 2024. *Eurostat*. [Online]. https://ec.europa.eu/eurostat/web/main/data/database [Accessed: 2023-08-08].

DION, R. 1952. Querelles des anciens et des modernes sur le facteurs de qualité du vin, *Annales De Géographie*, 61 (328): 417-431. http://www.jstor.org/ stable/23442544 [Accessed: 2023-12-20].

FLEURY, A. and FLEURY, M. T. 2003. Competitive strategies and core competencies: Perspective for the internationalization of industry in Brazil. *Integrated Manufacturing Systems*, 14(1): 16-25. https://doi.org/10.1108/09576060310453317

HOLMES, A. and ANDERSON, K. 2017. Convergence in National Alcohol Consumption Patterns: new global indicators,' Journal of Wine Economics, 12(2): 117–148. https://doi.org/10.1017/jwe.2017.15

JUDINOVÁ, E. 2018. Slovak wine exports - Determinants and competitiveness, *International Scientific Days 2018*. Towards Productive, Sustainable and Resilient Global Agriculture and Food Systems: 2451-2464. https://doi.org/10.15414/isd2018.s13.06

KPMG IN POLAND. 2014. The alcoholic beverages market in Poland. *assets.kpmg.com* [Online]. https://assets.kpmg.com/content/dam/kpmg/pdf/2016/02/The-alcoholic-beverages-market-in-Poland-2014-online.pdf [Accessed: 2024-01-05].

KUEMMERLE, W. 1999. The drivers of foreign direct investment into research and development: An empirical investigation. *Journal of International Business Studies*, 30(1): 1-24. https://doi.org/10.1057/palgrave.jibs.8490058

MATEJKOVÁ, E., et al. 2013. *Praktikum zo štatistiky*. Nitra SPU, 2013. 200s. ISBN 978-80- 552-1050-6.

MORRISON, A. AND RABELLOTTI, R. 2014. *Gradual catch up and enduring leadership in the global wine industry*, American Association of Wine Economists, Working paper 148. [Online]. http://www.wine-economics.org/dt_catalog/aawe- working-paper-no-148-business/ [Accessed: 2023-12-20].

OIV. 2022. *State of the World Vine in Wine Sector in 2022*. [Online] Available at: https://www.oiv.int/sites/default/files/documents/2023_SWVWS_report_EN.pdf [Accessed: 2024-50-01].PALLÁS, E. I. et al., 2016. Sustainable wine and grape production, the example of Hungary. *Visegrad Journal on Bioeconomy and Sustainable Development*, 5(2): 53-57.

REVEALED COMPARATIVE ADVANTAGE. 2022. In unctadstat.unctad.org [Online]. https://unctadstat.unctad.org/en/RcaRadar.html [Accessed: 2023-25-12].

RING, P. S. and VAN DE VEN, A. 1994. Developmental processes of cooperative interor- ganizational relationship. *Academy of Management Review*, 19(1): 90-118. [Online]. Available: DOI 10.2307/258836. [Accessed: 2023-25-12].

ROCCHI, B. and STEFANI, G. 2001. *Aspetti tecnologici, organizzativi e istituzionali nei processi di sviluppo rurale: alcune riflessioni*, Basile, Milan: Franco Angeli. pp. 332–352. ISBN 88-464-7961-0.

ROGOVSKÁ, V. 2018. Competitiveness of the Slovak wine industry in the European economic area with a focus on the Visegrad countries. In: *XXI IGWT SYMPOSIUM - Sustainability, Quality and Innovation: A Global View of Commodity Sciences*, 1(1):100–112. [Online]. Available: https://www.cabdirect.org/cabdirect/abstract/20183326621. [Accessed: 2023-12-28].

SMUTKA L. 2014. Mutual trade in agricultural products among Visegrad countries – basic development trends. *Studia Negotia*, 59(1): 1 – 18.

SMUTKA L., SVATOŠ M., QINETI A., and SELBY R. 2013. Visegrad countries' agricultural foreign trade development(transformation process). *Scientia Agriculturae Bohemica*, 44(1): 38-46

SVATOŠ, M. and SMUTKA, Ľ. 2014. Visegrad countries - agrarian foreign trade develop- ment in relation to their total merchandise trade performance. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego W Warszawie. Problemy Rolnictwa Świa- towego*, 14(29): 158–172. https://doi.org/10.22004/ag.econ.198880

URBAN, O. 1991. *České a slovenské dějiny do roku 1918*. (*Czech and Slovakian History till 1918*.) Svoboda, Praha. ISBN 80-205-0193-2.

WOODWARD, D. and ROLFE, R. 1993. The location of export oriented foreign direct investment in the Caribbean basin. *Journal of International Business Studies*, 24(1): 121.

Contact information

Katarína Bírová: email: xbirovak@uniag.sk
Patrik Rovný: email: patrik.rovny@uniag.sk

PROTECT YOURSELF FROM AI HALLUCINATIONS: EXPLORING ORIGINS AND BEST PRACTICES

Jana Dannhoferová[1], Petr Jedlička[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

Although AI-powered chat systems like ChatGPT can be trusted, we shouldn't rely on them completely. They can sometimes produce irrelevant, misleading or even false responses, known as hallucination effects. The causes can be both systemic and user related. User behavior, particularly in the area of prompt engineering, has an impact on the quality and accuracy of the result provided. Based on the literature review, we have identified the most common types of hallucination effects and provided examples in created categories. Finally, we have highlighted what users should consider when writing prompts and given recommendations for them to minimize hallucination effects in responses obtained from AI systems. Understanding how hallucinations occur can help ensure that these powerful tools are used responsibly and effectively. However, the quality of responses is always a matter of judgment, and the user's level of expertise and critical thinking is an important factor.

Keywords: artificial intelligence, AI systems, large language models, hallucination effect, text-to-text prompt engineering

JEL Code: O33, J24

1 INTRODUCTION

Artificial Intelligence (AI) is rapidly penetrating various areas of human activity. Despite the strengths of AI, generative AI models such as GPT have limitations and weaknesses. One of these is the fact that they can produce seemingly credible but incorrect, irrelevant, misleading or even false answers (Shen et al., 2023), (Xiao and Wang, 2021). This phenomenon, known as the hallucination effect, is a common problem in many large language models (LLMs) (Athaluri et al., 2023), (Bang et al., 2023), (Rohrbach et al., 2018). According to Bernard Marr (2023), hallucination in AI refers to the generation of output that may sound plausible but is either factually incorrect or unrelated to the given context. An AI hallucination is when an

https://doi.org/10.11118/978-80-7509-990-7-0049

AI makes up false information or facts that aren't based on real data or events (Keary, 2024). Hallucinations are so common that OpenAI even warns users within ChatGPT that "ChatGPT may produce inaccurate information about people, places, or facts".

1.1 Why is AI hallucination a problem?

The hallucination effect is an issue that can hinder user trust in AI systems. If the misinformation hallucinated by AI spreads like wildfire on the internet, making it appear authoritative and written by humans, it will undermine user confidence – making it difficult for users to trust information on the internet (Sushir, 2024). If AI systems produce incorrect or misleading information, users may lose trust in the technology, hindering its adoption in various sectors (Marr, 2023).

They are also ethically problematic, as inconsistencies in training data can lead to the mass dissemination of misinformation (Sushir, 2024). LLMs increase the ability to create realistic AI-generated fake content, which plays an important role in the misinformation phenomenon that our world is currently facing (Bontridder and Poullet, 2021). AI systems can expose users to legal threats if their responses are inaccurate or misleading. This is confirmed by Athaluri et al. (2023), who state that hallucinations can raise a number of ethical and legal issues and negatively affect decision-making. Poor decisions in areas such as health care can have serious consequences.

Although several studies have already described the weaknesses of AI systems, no one has yet addressed the behavior of the users of these systems, which can also affect the accuracy of the result provided. There is also a lack of categorization of types of errors, including corrective measures.

The aim of this paper is to identify the most common types of hallucination effects that can occur in responses generated by AI systems, provide examples and identify the main causes. Based on the results of the literature review, appropriate recommendations will be provided that will lead to the minimization of hallucination effects in the generated responses.

Research questions:

- What are the most common types of hallucination?
- What can users do to minimize the hallucination effect?

Categorizing the types of hallucination errors will help users of AI systems to understand what types of errors to look out for and how to modify their behavior to minimize these errors in their results.

2 LITERATURE REVIEW

2.1 LLMs

Large Language Models (LLMs) are deep learning models trained to understand and generate natural language. They use a two-stage training pipeline to learn efficiently from data. In the initial pre-training stage, LLMs use a self-supervised learning approach, which allows them to learn from large amounts of unannotated data without the need for manual annotation. In the subsequent fine-tuning phase, LLMs are trained on small, task-specific, annotated datasets to use the knowledge gained in the pre-training phase to perform specific tasks as intended by end users. As a result, LLMs achieve high accuracy on various tasks with minimal human-provided labels (Shen et al., 2023). The conversational artificial intelligence (AI) systems such as ChatGPT simulate a conversation with a human (Gupta et al., 2020). They can answer questions and provide information. The use of LLMs helps them to learn the grammar, syntax and context of different languages or subjects.

2.2 Limits of LLMs

Research has revealed significant gender and racial bias in AI systems. Some facial analysis software couldn't recognize a dark-skinned face until a person put on a white mask. When given the task of guessing the gender of a face, some systems performed significantly better on male faces than female faces (Buolamwini, 2019). Another major problem with the LLM is political bias. A team of researchers from the Technical University of Munich and the University of Hamburg provided evidence that ChatGPT has a "pro-environmental, left-libertarian orientation" (Hartmann et al., 2023). The same conclusion was reached by Fujimoto and Takemoto (2023). These results often arise from the AI model's inherent biases, lack of understanding of the real world, or limitations of the training data. In other words, the AI system 'hallucinates' information on which it has not been explicitly trained, leading to unreliable or misleading responses (Marr, 2023).

According to Athaluri et al. (2023) AI hallucination usually occurs due to adversarial examples such as varied input data that confuse the AI systems into misclassifying and misinterpreting them, resulting in inappropriate and hallucinatory output. Bang et al. (2023) they concluded that the ChatGPT suffers from hallucination problems like other AI systems and it generates more extrinsic hallucinations from its parametric memory as it does not have access to an external knowledge base. T. Sushir (2024) describes four basic types of hallucination:

- Sentence contradiction: This occurs when an LLM model produces a sentence that completely contradicts its previously asserted sentence.
- Factual contradiction: This type of hallucination occurs when the AI model presents false or fictitious information as fact.
- Prompt contradiction: This type of hallucination occurs when the output contradicts the prompt for which it generated an output.
- Random or irrelevant hallucinations: This hallucination occurs when the model produces output that is completely irrelevant to the given prompt.

Hallucinatory errors and weaknesses in AI models are usually caused by the following:

Cause	Description of cause
Misinterpretation of ambiguous input	LLMs may misinterpret ambiguous statements, leading to inaccurate or unintended responses.
Lack of contextual understanding	Chatbots may struggle to maintain the context of a conversation, resulting in responses that seem unrelated or inappropriate.
Overconfidence	An AI model that makes overly confident predictions even when faced with uncertain or ambiguous input, leading to inaccurate responses.)
Lack of common sense reasoning	AI models may lack common sense reasoning, leading to responses that seem illogical or impractical in certain situations.
Failure to recognize sarcasm or irony	AI systems may struggle to recognize sarcasm or irony in text, leading to literal interpretations and potentially incorrect responses.

Tab. 1 Causes of hallucinatory errors

Kenny Lee (2023) identified three main factors that cause LLMs to hallucinate: training data, lack of objective orientation, and inappropriately worded sentences. First, large language models have been developed through unsupervised training on large and heterogeneous datasets. These datasets come from many sources, making it difficult to ensure their impartiality and factual accuracy. The language model alone is not capable of distinguishing between truth and falsehood. Moreover, the inclusion of diverse and subjective perspectives within the training data further hampers the model's ability to discern objective truths. As a result, the model generates outputs that are likely, based on the patterns it has learned during the training process. Secondly, LLMs are susceptible to producing incorrect output when tasked with functions outside their training scope. Models such as GPT, Palm and Cohere are designed for broad natural language processing tasks. As a result, they may struggle to make accurate judgements when dealing with queries that require specialist knowledge in areas such as medicine, law and finance. Thirdly, to operate a LLM, users enter text as prompts. These prompts guide the LLM to perform certain tasks, similar to programming, but using natural language rather than programming languages. It is therefore essential that users write these prompts with the utmost precision. If the prompt is out of context, the LLM may produce an incorrect or completely unrelated response to what the user intended.

2.3 Prompt Engineering

It follows from the above that there is only one way for the user to suppress the hallucinatory effects, and that is to write the prompts appropriately. However, the quality of the prompts we give to generative AI models plays a critical role in determining the quality and relevance of the output. Clear, detailed and well-structured prompts are more likely to produce desirable results, while vague or irrelevant prompts can lead to unsatisfactory results. The methodology of designing effective requests or queries (prompts) to large language models is called prompt engineering. There are many strategies and guidelines on how to write prompts correctly when communicating with a generative AI model (Korzynski et al., 2023).

2.3.1 Multi-turn prompting

Based on the responses generated by the model. Prompts are thus structured as a series of turns or exchanges between the user and the model. This structure allows the model to consider the context of the entire conversation when generating a response, rather than just the most recent user input. This optimizes the quality and relevance of the generated responses (Bang et al., 2023). In this approach users provide input prompts and refine them over multiple turns or iterations.

2.3.2 Few-shot prompting

Few-shot prompting refers to a technique in which the user is provided with a small number of desired output examples prior to entering the task (Song et al., 2022). This approach aims to enable AI models to perform a given task with only a small number of labeled examples or prompts. In few-shot prompting, the AI model is trained or fine-tuned on a limited set of examples that demonstrate the desired task. These examples are typically provided as prompts to the model, which then learns to generalize from them to perform the task on new inputs. The focus of few-shot prompting is to adapt to a specific task or domain with minimal supervision, using the model's pre-trained knowledge to achieve task performance with few examples (Reynolds and McDonell, 2021).

2.3.3 Chain-of-Thought Prompting

Chain-of-thought (CoT) prompting focuses on facilitating multi-turn interactions or conversations between the user and the AI model, allowing for more coherent and contextually relevant responses. In chain-of-thought prompting, the prompts provided to the AI model are

structured to guide the flow of the conversation over multiple turns. The prompts are designed to build on the context established in previous turns, creating a coherent chain of thought throughout the interaction. The key feature of chain-of-thought prompting is its emphasis on maintaining context and coherence across multiple turns of the interaction, allowing for more natural and engaging conversations between the user and the AI model (Wei et al., 2022).

2.3.4 Self-Consistency

Self-consistency is an approach that simply asks a model the same prompt multiple times and takes the majority result as the final answer. It is a follow-up to CoT prompting and is more powerful when used in conjunction with it (Cheng et al., 2023), (Wang et al., 2022).

2.3.5 Temperature

The temperature parameter is a key feature in controlling the creativity and randomness of the responses generated by language models such as GPT-4. It essentially determines how conservative or adventurous the model is in its predictions. When writing prompts, users can specify the temperature parameter to adjust the level of randomness in the generated responses. The temperature value typically ranges between 0 and 1, with lower values producing more conservative and deterministic responses, and higher values producing more creative and varied outputs. Users can experiment with different temperature values to explore the creativity and variety of responses generated by the model. Lower temperatures (e.g. 0.1 or 0.2) tend to produce more predictable and coherent text that closely matches the training data. In contrast, higher temperatures (e.g. 0.7 or 1.0) introduce more randomness and variability, potentially leading to more imaginative but less reliable outputs (Mishra, 2023).

3 METHODOLOGY AND DATA

The basis of this study is a critical review of the findings that deal with the hallucination effects that can occur in the text responses provided by LLMs such as ChatGPT, Bard and Bing. By analyzing the existing research and reviewing the literature, we will attempt to map the various examples of hallucination effects and then organize them into categories. We characterize each category with a brief description and give a typical example from prompt engineering practice.

Next, we look at the causes of hallucination effects. A review of the literature shows that the causes of response errors can be both LLM and user related, particularly in the area of prompt engineering. In our study we will not look at model-side causes because we cannot control for them. However, we will focus on errors that occur on the user side. This is mainly because the user can change his or her behavior in the future, but has no way of influencing the behavior of the system itself.

The literature review also shows that there are currently a large number of different recommendations regarding the accuracy of prompts. We will try to select and provide those recommendations that will help to minimize hallucination effects. We hope that this paper will be a methodological guide for users, helping them to minimize errors in the responses generated by LLMs.

4 RESULTS

4.1 What are the most common types of hallucination?

By examining current research and reviewing the literature, we aimed to answer the research question: What are the most common types of hallucination effects? We looked for specific examples of hallucinations to identify different types and forms of hallucination effects that can manifest in responses. AI hallucinations can range from minor inconsistencies to completely false or fabricated responses, and could potentially mislead users who rely solely on the model's responses without independently verifying the facts.

4.1.1 Unrealistic Cause and Effect

However, AI systems can produce results that express unrealistic cause-and-effect relationships if the training data on which they have been trained contains biases or inaccuracies. For example, if an AI system trained on flawed data incorrectly concludes that "eating ice cream prevents sunburn", this would be a hypothetical example of expressing an unrealistic cause--and-effect relationship. Srinivasan and Chander (2021) give an example of unrealistic cause and effect when "a child wearing sunglasses is labeled as a failure, a loser, a nonstarter, an unsuccessful person".

4.1.2 Historical Revisions

Bernard Marr (2023) gives examples of LLMs inadvertently revising historical events. To the question, "When did Leonardo da Vinci paint the Mona Lisa?" he received the answer: "Leonardo da Vinci painted the Mona Lisa in 1815." This was incorrect because the Mona Lisa was painted between 1503 and 1506, or perhaps as late as 1517. When asked, "Tell me a fact about George Washington," he received the answer: "George Washington is known for inventing the cotton gin". These claims are unrelated, as Eli Whitney, not George Washington, invented the cotton gin.

4.1.3 Unsupported Claims

LLMs could make unsubstantiated claims without providing evidence. In a promotional video released by Google in February 2023, its AI chatbot Bard made a false claim. It incorrectly stated that "the James Webb Space Telescope has captured the first image of a planet outside our solar system" (Sushir, 2024), (Keary, 2024). However, this claim was inaccurate. LLMs can produce information that doesn't match the temporal sequences. For example, in the current conflict between Israel and Gaza, both the Bard and Bing systems incorrectly claimed that a ceasefire had been declared, probably based on news from May 2023. Bard then backtracked and said: "No, I am not sure that is correct. I apologize for my earlier response," but also made up casualty figures for two days into the future (Gillham, 2023).

4.1.4 Racial and Gender Bias

AI systems have been found to perpetuate racial bias in predicting recidivism rates. One notable example is the COMPAS system, which predicts that black defendants are at higher risk of reoffending than they actually are, while the opposite is true for white defendants (Cossins, 2018).

4.1.5 Misleading Information

Users may also receive information that is misleading or leads to incorrect conclusions. AI-generated writing was suspected when the Microsoft Start travel pages published a guide to places to visit in the Canadian capital, Ottawa. While there were errors in the details of some locations, most of the comments about the article were about how it included the Ottawa Food Bank as a tourist hotspot, encouraging readers to visit on an empty stomach (Gillham, 2023).

Amazon's Kindle Direct Publishing sold what appeared to be AI-written guides to foraging for edible mushrooms. One e-book encouraged the collection and consumption of legally protected species. Another mushroom guide included instructions that contradicted accepted best practices for identifying mushrooms that are safe to eat (Gillham, 2023).

4.1.6 Geographical Errors

AI-generated text can sometimes contain geographical errors, highlighting the importance of critically evaluating and fact-checking information generated by such models, especially when it comes to issues of geographical accuracy. Answers may include instances where places are incorrectly associated with different regions or countries, incorrect geographical boundaries or relationships between places, fictitious or non-existent places, or confusion between similar names. An example is the prompt "Name three cities in the United States" and the response "New York, Los Angeles, Toronto" (Lutkevich, 2023).

4.1.7 Random Output

Sometimes the output contradicts the prompt for which it generates output. LLMs can produce output that is completely irrelevant to the prompt given. For example, if the prompt is "Write an invitation to my friends for my birthday party". The model might generate output such as "Happy anniversary, Mum and Dad". (Sushir, 2024). A. Riaz (2023) confirms that AI systems can generate stories or narratives based on given prompts or data. However, due to limitations in understanding context or logical coherence, the stories generated may have nonsensical or illogical plots, resembling hallucinatory narratives.

4.1.8 Reasoning Errors

This type of error occurs when an AI system fails to apply correct logical reasoning or common sense to a problem. Reasoning errors are a significant challenge in AI, particularly for LLMs, which often struggle with tasks that require an understanding of the world that humans take for granted (Richardson and Heck., 2023).

We have divided the types of AI hallucinations into several categories, as shown in the Table 2.

It is important to recognize that all these examples illustrate the inherent limitations and potential risks associated with LLMs, particularly in terms of their ability to generate accurate and contextually appropriate information. AI hallucinations are a significant barrier to the reliability and accuracy of AI-generated content. Users are advised to approach the output of such models with caution and critical evaluation. Mitigating these challenges requires a comprehensive strategy that includes improved context awareness and user education.

4.2 What can users do to minimize the hallucination effect?

While the AI systems themselves cannot be influenced by users, the input they provide as prompts can. It's therefore crucial to provide clear and specific prompts, while unclear, inaccurate, inconsistent or contradictory prompts should be avoided (Sushir, 2024), (Lutkevich, 2023).

4.2.1 How to prevent AI hallucinations generally

While artificial intelligence has notable strengths, generative AI models such as ChatGPT have limitations and vulnerabilities. Hallucinations are considered an inherent part of LLMs. However, there are ways to reduce hallucinations. First, company owners must ensure that the AI model's training datasets are regularly updated and expanded to account for and keep up with cultural, political, and other evolving events (Sushir, 2024). In addition, AI hallucination can certainly be minimized by improving training inputs through the inclusion of diverse, accurate, and contextually relevant datasets, as well as frequent user feedback and the involvement of human reviewers to evaluate the outputs generated by an AI system Athaluri et al.

Name of category	Short description	Example of response
Unrealistic Cause and Effect	LLMs might suggest unrealistic cause-and-effect relationships.	"A child wearing sunglasses is labeled as a failure, loser, nonstarter, unsuccessful person." (Srinivasan and Chander, 2021)
Historical Revisions	LLMs may inadvertently revise historical events.	"Leonardo da Vinci painted the Mona Lisa in 1815." (Marr, 2023)
Unsupported Claims	LLMs could make unsupported assertions without providing evidence.	"The James Webb Space Telescope had taken the first image of a planet outside our solar system." (Sushir, 2024)
Racial and Gender Bias	LLMs perpetuate racial biases in predicting recidivism rates.	"Black defendants pose a higher risk of recidivism." (Cossins, 2018)
Misleading Information	LLMs might provide information that is misleading or leads to incorrect conclusions.	"The Ottawa Food Bank is a tourist hotspot" (Gillham, 2023)
Geographical Errors	LLMs may provide inaccurate information about locations.	"New York, Los Angeles and Toronto are three cities in the United States." (Lutkevich, 2023)
Random Output	LLMs generate completely irrelevant output to the given prompt.	"Happy anniversary, Mom and Dad." (Sushir, 2024)
Reasoning Errors	LLMs produce outputs that defy common sense.	Prompt: "When I was 6, my sister was half my age. Now I'm 70. How old is my sister?" Output: "35" (Richardson and Heck, 2023)

Tab. 2 Types of errors in the LLM responses

(2023). Providing users with information about how the AI model works and its limitations can help them understand when to trust the system and when to seek additional verification (Marr, 2023). However, apart from solutions to this problem on the part of AI systems, there are several ways in which users can avoid or minimize the occurrence of hallucinations when communicating with LLMs.

4.2.2 Best practices for prompt writing to minimize hallucinations

Providing clear and specific prompts, along with the relevant context, is essential when interacting with an AI system. This clarity helps to guide the system towards the intended output, thereby increasing the accuracy and relevance of the response. Including detailed context in prompts allows the AI to better understand the nuances of the request, minimizing the likelihood of generating irrelevant or incorrect information. Therefore, users should carefully consider the information they provide to ensure it is comprehensive and relevant. By doing so, they can significantly improve the efficiency and effectiveness of the AI's performance. Korzynski et al. (2023) state that effective prompting may include:

- Context – includes information about the role the model is to play in the task, or any necessary information about the situation that may justify it. Example: "You are a human resources manager in a trading company."
- Instruction – the task to be performed. Example: "Write an email to the customer offering new products".
- Input data – data and facts that the model should use to complete the task. Example: keywords to include in the response.

- Expected output format – information about the format and type of output in which the answer is to be provided. Examples: “Generate a CSV file. Generate Python code”.

The use of idiomatic expressions, colloquial slang, or overly technical terminology can also obscure contextual understanding when interacting with AI systems. Users can adjust the temperature parameter, which controls the degree of randomness in the output. A higher temperature setting increases the variety and creativity of the text, allowing outputs that may exceed the expectations set by the input. However, this increase in randomness also increases the likelihood of producing responses that do not strictly follow the input patterns, potentially resulting in outputs that could be perceived as incorrect or misleading in certain situations. Conversely, a lower temperature setting will produce more consistent and predictable outputs that closely match the input patterns. This reduces the likelihood of producing outputs that are incorrect or misleading.

A notable limitation of Large Language Models (LLMs) is their tendency to produce inaccurate results in tasks requiring multi-step reasoning, such as arithmetic or logic problems. However, the accuracy of these models improves significantly when they are given multiple examples (few-shot learning), instructed to break down tasks into sequential steps (chain of thought) and to aggregate their results. This can significantly reduce the hallucinatory effects. Another effective strategy for improving accuracy is the self-consistency method. This technique is based on the premise that complex reasoning problems can often be approached in different ways but still lead to the same correct answer. Introduced by Wang et al. (2022), this method has been shown to significantly improve performance on arithmetic and common sense reasoning tasks across several large language models of different scales. In addition, the self-evaluation technique allows users to distinguish between correct and incorrect answers. By asking a model to generate responses along with probabilities of their correctness, users can use these well-calibrated probabilities to filter out likely incorrect responses. This technique assumes that the model is generally aware of its own knowledge limitations, thus allowing for more reliable filtering of the generated output (Kadavath et al., 2022).

5 CONCLUSION

AI systems have made significant progress, but they are not error-free. It is crucial for users to recognise these limitations in order to accurately assess AI-generated responses. Ultimately, the evaluation of such results must rely on human judgment. The user’s expertise and robust critical thinking skills are particularly important. Our research aimed to identify the most common risks associated with the use of AI systems. We identified the most common types of hallucinatory effects in text responses and provided guidelines for users to reduce errors in the responses they receive from these systems. An open question is whether it is possible to completely eliminate or correct hallucinations in AI.

Acknowledgements
This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, which is co-founded by the Operational Programme Research, Development and Education.

REFERENCES

ATHALURI, S., MANTHENA, S., KESAPRAGADA, V., YARLAGADDA, V., DAVE, T. and DUDDUMPUDI, R. T. S. 2023. Exploring the boundaries of reality: Investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus*, 15(4), e37432. DOI: 10.7759/cureus.37432

BANG, Y., CAHYAWIJAYA, S., LEE, N., DAI, W., SU, D., WILIE, B., LOVENIA, H., JI, Z., YU, T., CHUNG, W., DO, Q. V., XU, Y. and FUNG, P. 2023. A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. *arXiv preprint arXiv*: 2302.04023. DOI: 10.48550/arXiv.2302.04023

BONTRIDDER, N. and POULLET, Y. 2021. The role of artificial intelligence in disinformation. *Data & Policy*, 3, e32. DOI: 10.1017/dap.2021.20.

BUOLAMWINI, J. 2019. Artificial intelligence has a problem with gender and racial bias: Here's how to solve it. *Time* [Online]. Available at: https://time.com/5520558/artificial-intelligence-racial-gender-bias/ [Accessed 2024, February 19].

CHENG, F., ZOUHAR, V., ARORA, S., SACHAN, M., STROBELT, H. and EL-ASSADY, M. 2023. RELIC: Investigating large language model responses using self-consistency. *arXiv preprint arXiv*: 2311.16842. DOI: 10.48550/arXiv.2311.16842

COSSINS, D. 2018. Discriminating algorithms: 5 times AI showed prejudice. *Newscientist* [Online]. Available at: https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/ [Accessed 2024, February 19].

FUJIMOTO, S. and TAKEMOTO, K. 2023. Revisiting the political biases of ChatGPT. *Frontiers in Artificial Intelligence*, 6, 2023. DOI: 10.3389/frai.2023.1232003

GILLHAM, J. 2023. *AI hallucination factual error problems* [Online]. Available at: https://originality.ai/blog/ai-hallucination-factual-error-problems [Accessed 2024, February 19].

GUPTA, A., HATHWAR, D. and VIJAYAKUMAR, A. 2020. Introduction to AI chatbots. *International Journal of Engineering Research and Technology,* 9, 255–258. DOI: 10.17577/IJERTV9IS070143

HARTMANN, J., SCHWENYOW, J. and WITTE, M. 2023. The political ideology of conversational AI: Converging evidence on ChaGPT's pro-environmental, left-libertarian orientation. *arXiv preprint arXiv*: 2301.01768. DOI: 10.48550/arXiv.2301.01768

KADAVATH, S., CONERLY, T., ASKELL, A., HENIGHAN, T., DRAIN, D. et al. 2022. Language models (mostly) know what they know. *arXiv preprint arXiv:* 2207.05221. DOI: 10.48550/arXiv.2207.05221.

KEARY, T. 2024. *AI Hallutinations* [Online]. Available at: https://www.techopedia.com/definition/ai-hallucination [Accessed 2024, February 19].

KORZYNSKI, P., MAZUREK, G., KRZYPKOWSKA, P. and KURASINSKI, A. 2023. Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review,* 11(3), 25–37. DOI: 10.15678/EBER.2023.110302

LEE, K. 2023. *Understanding LLM Hallucinations and how to mitigate them* [Online]. Available at: https://kili-technology.com/large-language-models-llms/understanding-llm-hallucinations-and-how-to-mitigate-them [Accessed 2024, February 19].

LUTKEVICH, B. 2023. *AI hallucination* [Online]. Available at: https://www.techtarget.com/whatis/definition/AI-hallucination. [Accessed 2024, February 19].

MARR, B. 2023. *ChatGPT: What Are Hallucinations And Why Are They A Problem For AI Systems* [Online]. Available at: https://bernardmarr.com/chatgpt-what-are-hallucinations-and-why-are-they-a-problem-for-ai-systems/ [Accessed 2024, February 19].

MISHRA, A. N. 2023. *Hallucination in Large Language Models* [Online]. Available at: https://medium.com/@asheshnathmishra/hallucination-in-large-language-models-2023-f7b4e77855ae [Accessed 2024, February 19].

REYNOLDS, L. and MCDONELL, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. Paper presented at: *the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. DOI: arXiv:2102.07350

RIAZ A. 2023. *29 Mind-Blowing Examples of AI Hallucinations* [Online]. Available at: https://vividexamples.com/examples-of-ai-hallucinations/ [Accessed 2024, February 19].

RICHARDSON, C. and HECK, L. 2023. Commonsense Reasoning for Conversational AI: A Survey of the State of the Art. arXiv preprint arXiv: 2302.07926. DOI: 10.48550/arXiv.2302.07926

ROHRBACH, A, HENDRICKS, L. A., BURNS, K., DARRELL, T. and SAENKO, K. 2018. Object Hallucination in Image Captioning. *arXiv preprint arXiv*: 1809.02156. DOI: 10.48550/arXiv.1809.02156

SHEN, Y., HEACOCK, L., ELIAS, J., HENTEL, K. D., REIG, B., SHIH, G. and MOY, L. 2023. ChatGPT and Other Large Language Models Are Double-edged Swords. *Radiology*, 307(2). DOI: 10.1148/radiol.230163

SONG, W., LIYAN, T., AKASH, M., JUSTIN, F. R., GEORGE, S., YING, D. and YIFAN, P. 2022. Trustworthy assertion classification through prompting. *Journal of Biomedical Informatics*, 132, 104139. ISSN 1532-0464. DOI: 10.1016/j.jbi.2022.104139

SRINIVASAN, R. and CHANDER, A. 2021. Biases in AI Systems. *Communications of the ACM*, 64(8), 44–49. DOI: 10.1145/3464903

SUSHIR T. 2024. What is AI Hallucination, and Can it Be Fixed? *Geekflare* [Online]. Available at: https://geekflare.com/ai-hallucination/ [Accessed 2024, February 19].

WANG, X., WEI, J., SCHUURMANS, D., LE, Q., CHI, E., NARANG, S., CHOWDHERY, A. and ZHOU, D. 2022. Self-Consistency Improves Chain of Thought Reasoning in Language Models. *arXiv preprint arXiv*: 2203.11171. DOI: 10.48550/arXiv.2203.11171

WEI, J., WANG, X., SCHUURMANS, D., BOSMA, M., ICHTER, B., XIA, F., CHI, E. H., LE, Q. V. and ZHOU, D. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *arXiv preprint arXiv*: 2201.11903v6. DOI: 10.48550/arXiv.2201.11903

XIAO, Y. and WANG, W. Y. 2021. On Hallucination and Predictive Uncertainty in Conditional Language Generation. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*. Main Volume, 2734–2744. Association for Computational Linguistics.

Contact information
Jana Dannhoferová: e-mail: jana.dannhoferova@mendelu.cz
Petr Jedlička: e-mail: petr.jedlicka@mendelu.cz

THE USABILITY OF ALGORITHMS FROM GRAPH THEORY IN THE FIELD OF MULTICRITERIA ANALYSIS

Radim Farana[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

The main ideas, on how to use algorithms from the Graph Theory to improve the process of Multicriteria Analysis were presented in the paper (Farana, 2016). This paper presents practical results obtained by the use of these algorithms in practical decision-making procedures when multicriteria analysis has been used.

Graph algorithms were used in two situations when determining the values of the weights of decision criteria. First was the checking, if the Fuller's triangle, filled in by an expert, is filled in correctly. For this verification, a complete graph is used in which the vertices represent the criteria the orientation of the edges their mutual significance. A method of gluing vertices could be used for criteria with the same significance. The resulting graph must be acyclic. Twenty-five decision tasks with seven or more criteria were analyzed and the obtained results will be presented in the paper.

The second application was the elimination of the overdetermination of the assessment in Saaty's method. A spanning tree describing dependencies between criteria has been used according to the algorithm in (Farana, 2016). Obtained results were compared with the full Saaty's matrix when the number of compared pairs of criteria is k – 1 for k criteria, compared to the number of k(k – 1)/2 in the classic Saaty's method. Fifteen decision tasks with seven or more criteria were analyzed and the obtained results will be presented in the paper. The paper presents the differences between the assessment given directly by experts and the assessment obtained using the spanning tree and shows that the described method is applicable in practice. The experience of experts using the proposed procedure, obtained through a guided interview, was mostly positive.

Keywords: decision making, multicriteria analysis, graph theory, Fuller triangle, Saaty's method, acyclic graph, spanning tree

JEL Code: D81 Criteria for Decision-Making under Risk and Uncertainty

https://doi.org/10.11118/978-80-7509-990-7-0060

1 INTRODUCTION

Methods of multi-criteria analysis of variants (multi-criteria decision-making) are described in several publications, e.g. (Multi-criteria analysis, 2009; Triantaphyllou, 2000; Jablonský, 2007; Malakooti, 2013) and are still the source of several applications in solving complicated decision-making tasks, see e.g. (Borovcová, 2010). There are also many publications focused on the quality of value estimation and comparison of different approaches, see e.g. (Agarski, 2019).

Number of criteria	Number of experts	Expert experience	Number of problems	Longer cycle length
7	7	Experienced	0	0
9	8	Very experienced	0	0
8	6	Inexperienced	1	3
11	10	Experienced	1	3
7	8	Inexperienced	0	0
9	7	Inexperienced	0	0
8	5	Experienced	0	0
8	7	Very experienced	0	0
9	7	Inexperienced	1	3
10	12	Experienced	1	4
12	11	Experienced	0	0
7	7	Very experienced	0	0
8	7	Very experienced	0	0
7	9	Inexperienced	1	3
9	7	Experienced	0	0
9	9	Experienced	0	0
10	8	Inexperienced	0	0
8	6	Very experienced	0	0
7	7	Experienced	0	0
8	5	Experienced	0	0
9	7	Very experienced	0	0
12	8	Experienced	2	3
8	7	Experienced	0	0
7	5	Very experienced	0	0
9	7	Inexperienced	0	0

Tab. 1 Results of using Fuller's triangle in implemented multi-criteria decision-making tasks

We use the knowledge of experts to choose the best option while respecting a large number of, often conflicting, criteria. Above all, in the phase of determining the importance of individual criteria. Several methods have been gradually developed that try to help experts in determining the importance of criteria, especially when there are more of them. The Fuller's triangle and the Saaty's method are particularly effective methods. Unfortunately, the facilitation of expert decision-making is balanced by the danger of conflicting evaluation, which can negatively affect the result of the entire multi-criteria analysis, because we know from practice that these tasks are often very sensitive to the significance of the criteria. In other words, even a small change in the significance of the criteria can cause a significant change in the resulting evaluation of the variants.

The paper (Farana, 2016) presented how we can effectively use methods known from Graph Theory, e.g. (Gross, 2006), to detect these inconsistencies or even avoid them. This paper presents practical experiences from the use of these methods.

2 FULLER'S TRIANGLE

Fuller's triangle (also called the Pairwise comparison method), is a way to compare and determine the significance of a large number of evaluation criteria. The expert is presented with a set of all pairs of criteria with a request to mark which of the pair is more important, or may also mark both as equally important. This greatly simplifies his decision-making on the one hand, but at the same time, there is a danger that his opinion will be inconsistent.

This contradiction cannot be resolved and in more complex cases it is possible that it will not even be detected. For this, we can advantageously use a graph (complete), in which the vertices represent the criteria and the edges of their mutual evaluation. Firstly, we use the method of gluing for vertices with the same importance, next we orient the edges towards a more significant criterion. The resulting graph must be acyclic. If the graph contains cycles, the evaluation is inconsistent.

Table 1 presents the obtained results of using Fuller's triangle in implemented multi-criteria decision-making tasks with 7 or more criteria. It is evident that a very small number of problems have been identified. The typical length of the cycle is 3. Experts with little experience made the most mistakes.

A structured interview method was used to identify the source of faults. Most often, the experts stated that they found the criteria to be very similarly important, but for some reason, they did not mark them as equally important. This also shows the possibility of removing the cycle by marking all criteria in the cycle as equally important.

An interesting fact was discovered during the discussion with very experienced experts. A number of them stated that they were aware of the danger of inconsistent evaluation. And they prevent it by ranking the criteria in order of importance before they begin to fill in Fuller's triangle. Then the question is whether to use directly the method of order to determine the significance of the criteria.

3 SAATY'S METHOD

Saaty's method is a well-known method that enables a more sensitive evaluation of criteria (Saaty, 1977). Expert is evaluating every pair of criteria using values from 1 to 9 to determine the strength of preferences. Since all pairs of criteria are compared, the same problem as in Fuller's triangle can arise, i.e. the creation of a cycle. Then also the same method to identify the cycle, described above, can be used. However, the possibility to express the importance of preferences brings additional risks of inconsistency in evaluation. To eliminate overdetermination of evaluations, it is possible to use a graphic interpretation of the relationships

	A	B	C	D	E
A	1	1	1/5	1/3	5
B	1	1	1/5	1/3	5
C	5	5	1	3	9
D	3	3	1/3	1	7
E	1/5	1/5	1/9	1/7	1

Tab. 2 Saaty's table calculated from the expert's evaluations

between the criteria in the form of a graph. In this graph, we will find the spanning tree (any of them), we will evaluate criteria pairs at the spanning tree only, and calculate the rest of the evaluations (Farana, 2016).

Table 2 presents an example of five criteria (A to E) evaluation when only pairs A-B, B-C, A-D, and A-E were given by an expert and all rest values were calculated. It is evident, that the main diagonal contains values of 1, and the reciprocal values we obtain according to the relation (B-A) = 1/(A-B).

The next procedure already respects Saaty's method, so we determine the weights of the criteria and we can check the consistency of Saaty's matrix by calculating the variance estimate. The resulting value for the example from Table 2 is 0.155, and for five criteria a variance estimate of less than 0.2 is required.

Fifteen decision tasks with seven or more criteria were analyzed and experts were asked about their opinion on this method. The obtained results were significantly dependent on the experience of the experts:

Very experienced – neutral opinion, they do not see a problem in the overdetermination of the assessment, and they are not concerned about inconsistent assessment.

Experienced – predominantly positive opinion, they see the danger in the overdetermination of the assessment, and they are concerned about inconsistent assessment.

Inexperienced – a completely positive opinion, they see a danger in the overdetermination of the assessment, and they are very concerned about inconsistent assessment. They prefer to provide as few ratings as possible.

4 RESULTS

The paper presented two possible applications of algorithms from graph theory using classical methods for determining the significance of criteria. Sophisticated methods of evaluating the importance of criteria such as Fuller's triangle or Saaty's method carry the risk of inconsistent evaluation. The presented applications of algorithms from graph theory allow inconsistencies to be detected or even eliminated.

5 DISCUSSION AND CONCLUSIONS

The use of graph algorithms gives new options to support decision-making, many publications are focused on this area, see e.g. (Nordeman 2020, Nestrenko 2022). But classical methods are still used and the use of presented procedures and applications from graph theory can significantly support them. The results achieved and the opinions of experts, presented in this paper, prove it.

Acknowledgements
This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed by Operational Programme Research, Development and Education.

REFERENCES

AGARSKI, B., HADZISTEVIC, M., BUDAK, I., MORACA, S. and VUKELIC, D. 2019. Comparison of approaches to weighting of multiple criteria for selecting equipment to optimize performance and safety. Online. *International Journal of Occupational Safety and Ergonomics*. 25(2), 228-240. ISSN 1080-3548. https://doi.org/10.1080/10803548.2017.1341126 [cit. 2024-01-15].

BOROVCOVÁ M. 2010. Methods of multi-criteria evaluation of variants and their use in the selection of a financial institution's product (in Czech). In: *5th International conference Management and modeling of financial risks*, VŠB-TUO EkF, 2010, 8 pp. https://docplayer.cz/8392238-Metody-vicekriterialniho-hodnoceni-variant-a-jejich-vyuziti-pri-vyberu-produktu-financni-instituce.html

FARANA, R. 2016. Using graphs in multi-criteria analysis tasks (in Czech). In: *XLI. Seminar ASR '2016 "Instruments and Control"*. Ostrava: VŠB-TU Ostrava, 22. 4. 2016, s. 47 – 53. ISBN 978-80-248-3910-3. http://akce.fs.vsb.cz/2016/asr2016/Sbornik-ASR2016.pdf

GROSS, J. L. and YELLEN, J. 2006. *Graph theory and its applications*. 2nd Edition. New York: Chapman and Hall. ISBN 1-58488-505-x

JABLONSKÝ, J. 2007 *Operations research: quantitative methods for economic decision-making* (in Czech). 3. ed. Praha: Professional Publishing, 323 s. ISBN 978-80-86946-44-3

MALAKOOTI, B. 2013. *Production and operation systems with multiple objectives*. John Wiley & Sons, , 1114 pp. ISBN 9781118585375

DEPARTMENT FOR COMMUNITIES AND LOCAL GOVERNMENT. 2009. *Multi-criteria analysis: a manual*. 2009. London: Department for Communities and Local Government, 68 pp. ISBN: 978-1-4098-1023-0. http://eprints.lse.ac.uk/12761/1/Multi-criteria_Analysis.pdf

NESTERENKO, O., NETESIN, I., POLISCHUK, V. and SELIN, Y. 2022. Graph-based decision making for varying complexity multicriteria problems. *Computer Science Journal of Moldova*, 30, 3(90). https://doi.org/10.56415/csjm.v30.21

NORDEMANN, F., TÖNJES, R., PULVERMÜLLER, E. and TAPKEN, H. 2020. Graph-based Multi-Criteria Optimization for Business Processes. In: *10th International Symposium on Business Modeling and Software Design (BMSD)*, Lecture Notes in Business Information Processing, Springer, 2020. https://doi.org/10.1007/978-3-030-52306-0_5

SAATY, T. L. 1977. A scaling method for priorities in hierarchical structures. Online. *Journal of Mathematical Psychology*. 5, 3, 234-281. ISSN 00222496. https://doi.org/10.1016/0022-2496(77)90033-5

SAATY, T. L. 1988. What is the Analytic Hierarchy Process? Online. In: MITRA, G., GREENBERG, H. J., LOOTSMA, F. A., RIJKAERT, M. J. and ZIMMERMANN, H. J. (eds.). *Mathematical Models for Decision Support*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 109-121. ISBN 978-3-642-83557-5 https://doi.org/10.1007/978-3-642-83555-1_5

TRIANTAPHYLLOU, E. 2000. *Multi-criteria decision making methods: a comparative study*. Boston, Mass.: Kluwer Academic Publishers, 2000, 288 pp. ISBN 0-7923-6607-7

Contact information
Radim Farana: e-mail: radim.farana@mendelu.cz

DIFFERENT INSIGHT INTO THE VAT GAP USING MIMIC MODEL

Iva Hasíková[1]

[1]Department of Finance and Accounting, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

The paper focuses on the estimation of the MIMIC model for quantification of VAT gap. MIMIC model is a specific type of structural equation models, which treats the VAT gap as a latent unmeasurable variable whose emergence and size are influenced by causes and whose presence is reflected in indicators; causes and indicators must be measurable. The contribution of this model is identification of causes of VAT gap, that are potential sources of VAT collection inefficiencies. The MIMIC model was built on data from selected European countries and according to the model VAT gap has these significant causes: openness of economy, corruption perception index, general government expenditure, final consumption and e-government development index. Developed using data from European countries, the model can be applied on each of these countries for quantification of VAT gap. These outputs can support the recommendations leading to improved efficiency of VAT collection.

Keywords: VAT gap, determinants of VAT gap, efficiency of VAT collection, MIMIC model, structural equation model

JEL Code: C39, C51, H26

1 INTRODUCTION

Value added tax ("VAT") is one of the most important tax sources in most European countries. However, it is often connected with the risk of tax evasion. According to the estimates of the European Union ("EU") in 2020, the total VAT evasion loss was 93 billion euro, this means that every second of the year 2020 cost the public budgets of the EU member states around 3.000 euro (Baert, 2023). Tax evasion generally poses threats to the economic environment, the tax system and society. According to Novysedlák and Palkovičová (2012), tax evasion and tax avoidance ruins the economic environment because some entities gain an advantage that may in the long run encourage other entities to similar practices, even if tax evasion was not their first intention. Moreover, tax evasion is a major source of inequality, regardless of a country's redistribution objective (Argentiero et al., 2021).

When dealing with tax evasion and examining tax collection efficiency, the tax gap can provide valuable information. In the case of VAT, the VAT gap has become an indicator of

https://doi.org/10.11118/978-80-7509-990-7-0065

VAT fraud (Moravec et al., 2021). VAT gap is defined as the difference between the amount of VAT that should be collected under the legislation and the VAT actually collected (Carfora et al., 2020). The VAT gap covers revenues lost due to tax frauds and also due to bankruptcies, taxpayer insolvency or tax liability miscalculations (Jonski & Gajewski, 2022).

When quantifying the VAT gap, econometric methods make it possible to consider more aspects, for example the quality of public sector institutions or the level of digitalization of the public administration. This paper while building multiple indicator multiple causes (hereinafter 'MIMIC') model considers variables such as unemployment, openness of economy, index of economy freedom, corruption perception index (hereinafter 'CPI'), general government expenditure, final consumption expenditure, e-government development index (hereinafter 'EGDI') as causes of the VAT gap. The existence of a VAT gap is evidenced by the growth of GDP (gross domestic product) per capita and VAT revenue (indicators).

Contribution of this paper is the MIMIC model with data from selected European countries, that model can be applicable to these countries.

2 LITERATURE REVIEW

For estimating the VAT gap and the tax gap in general, there are three approaches in the literature: the bottom-up approach, the top-down approach and methods based on econometric modelling (Alm, 2012; Kasnauskienė and Krimisieraitė, 2015; Poniatowski et al., 2020). The bottom-up approach is based on tax audits and direct interviews with taxpayers (Alm, 2012). According to Novysedlák and Palkovičová (2012) into bottom-up approach belong these methods: selection of a random sample of taxpayers and targeted controls based on risk analysis. Gajewski and Jonski (2022) alert, that tax authorities using the second mentioned method target taxpayers with the highest expected value of unreported tax liabilities, so such data source could lead to overestimation of the gap.

The top-down approach processes macroeconomic data and international accounts data that cover the entire national economy (Poniatowski et al. 2020). As data sources can be used data from statistical offices on the production of gross domestic product or supply and use tables which provide information on the production of individual industries but also on the consumption by these industries and sectors of the national economy.

In contrast to the other two approaches, the approach based on econometric modelling not only provides information about the size of tax evasion, but also identifies the factors and determinants that have influence on its size. On the other hand, the disadvantage is that the econometric model can only determine the development (year-on-year changes) of tax evasion over time; its value in the base period needs to be determined by another method (Kasnauskinė and Krimisieraitė, 2015; Schneider, 2005).

The MIMIC model was pioneered by Frey and Weck-Hanneman (1984), who used it to estimate the extent of the shadow economy in member states of Organization for Economic Co-operation and Development (hereinafter 'OECD'). Schneider et al. (2010) and Tedds (2005) also used it in connection with calculation of share of the shadow economy on GDP. Frey and Weck-Hanneman (1984) chose this method in response to the fact that all approaches used until that time assessed the extent of the shadow economy on the basis of just one indicator depending on the method used (the currency in circulation – the demand for currency approach) and moreover, they did not consider almost any other causes of the shadow economy. The MIMIC model was also used for research into the determinants of the VAT gap in Lithuania (Kasnauskienė and Krimisieraitė, 2015).

Using MIMIC model it is possible to include into the calculation of VAT gap variables such as e-government development index (hereinafter 'EGDI'), this variables have not been considered regards VAT gap calculation in several European countries yet.

3 METHODOLOGY

3.1 Variable selection

The following section introduces variables for the MIMIC model. Two types of variables are needed for this model, causes and indicators. Table 1 consists of candidate variables as causes and indicators, short explanation, source of the data and authors, who have already considered these variables regarding VAT gap.

Center for social and economic research (hereinafter 'CASE') included among the possible determinants of the VAT gap the unemployment rate as an indicator of taxpayers' liquidity difficulties (Poniatowski et al. 2018). The unemployment rate also expresses income inequality or poverty (Reckon, 2009). In this research, unemployment will be understood as an index of the economic cycle.

The openness of an economy is expressed as the share of the sum of imports and exports on a country's GDP. This variable was examined by Aizenman and Jinjarak (2008) and also more recently by Carfora et al. (2020). According to these authors, the openness of an economy has a positive effect on the efficiency of VAT collection, i.e., it reduces the VAT gap. According to the research by Pluskota (2022), the factor of the share of foreign trade (exports and imports) in GDP is significant for the whole EU.

Variable	Explanation	Author	Source
Causes			
Unemployment rate	Poverty, income inequality, index of the economic cycle	Poniatowski et al. (2018); Reckon (2009)	Eurostat
Openness of economy	Risk of carousel fraud, the openness of the economy	Zídková and Pavel (2016)	Eurostat
Index of economic freedom	Government quality, tax burden, open market	Godin and Hindriks (2015)	The Heritage Foundation
CPI	Government quality, corruption perceived	Reckon (2009); CASE (2013)	Transparency International
General government expenditure	Size of the public sector	Reckon (2009); Zídková and Pavel (2016)	Eurostat
Final consumption	Purchases of final consumers, potential VAT base	Zídková and Pavel (2016)	Eurostat
EGDI	Information technologies in government and tax offices	Poniatowski et al. (2020)	United Nations
Indicators			
Growth of GDP per capita	Reflection of tax evasion between taxpayers	Kasnauskienė and Krimisieraitė (2015)	Eurostat
VAT revenue on GDP	Level of VAT revenue in each country	Kasnauskienė and Krimisieraitė (2015)	Eurostat

Tab. 1 Candidate variables as causes and indicators

Openness of economy also brings an opportunity for missing trade, intra community fraud and carousel fraud (Frunza, 2019). In this case openness of economy has a negative effect on the efficiency of VAT collection.

The efficiency of the tax system is strongly influenced by the quality of the government which means mainly the formulation and implementation of various regulations but also the degree of independence of tax administration of political pressure (Godin & Hindriks, 2015). According to the findings of Chan and Ramly (2018), the redistributive effect of the VAT system also depends on the quality of the government structure, otherwise the VAT system could be highly regressive and widens income inequality.

The index of economic freedom has 12 principles for sustained progress and prosperity, many of which involve the quality of government, as described above (The Heritage Foundation, 2023).

CPI is another variable entering the model. CPI measures, how corrupt each country's public sector is perceived to be according to experts and businesspeople, higher CPI means lower perceived corruption in country (Transparency International, 2023). CPI is next to the Index of economy freedom another sign of quality of public institution. Reckon (2009) included CPI into its econometric analysis of VAT gap and it showed up as the variable with the strongest relationship with the size of the VAT gap, lower perceived corruption is connected to the lower VAT gap.

CASE also included CPI into its regression analysis of VAT gaps' determinants, but its results showed positive, however insignificant, relationship between VAT gap and CPI, it indicates that improvement in perception of corruption within a country is connected to the higher VAT gap (Barbone et al., 2013). CASE (Barbone et al., 2013) and Reckon (2009) reached opposite results about relationship between CPI and VAT gap.

For the purpose of the model, government expenditure is expressed as a share of GDP in each country. Reckon (2009) included government expenditure in his research into the causes of the VAT gap because it reflects both the total tax burden and also the size of the public sector with tax audits and other types of regulation. Zídková and Pavel (2016) included government spending as a share of GDP in their study on the causes of the VAT gap because it reflects the size of the public sector. They argue that a larger size of the public sector will lead to a reduction in the VAT gap.

Final consumption expenditure is the part of the expenditure that is spent mainly by households on goods and services that will be used to directly satisfy individual needs (Eurostat, 2016).

There are purchases by final consumers which can proxy potential VAT base (Zídková and Pavel, 2016). These purchases can be made in cash and as such they do not come under the scrutiny of the tax authority. A study carried out by Immordino and Russo (2018) shows that cashless payments negatively affect the VAT gap. Even for this reason, cash final consumption is much more problematic for VAT collection than intermediate consumption by manufacturing businesses, which can claim VAT deductions on their purchases (Zídková and Pavel, 2016).

EGDI is an index produced by the United Nations for its member states and it is to reflect how countries use information technology (United Nations, 2023). The use of this index helps to take into account advancing digitalization whose purpose is to eliminate tax evasion. CASE used Information Technology expenditures related to GDP in order to capture the effect of implementing innovative processes into tax administration, result of their research show a statistically significant negative effect of Information Technology expenditures on GDP on VAT gap (Poniatowski et al., 2020). Digitalization will allow the government to access and analyse the necessary information, which will increase the efficiency of tax collection (Alm, 2021).

GDP per capita will be used as the first indicator to reveal VAT collection inefficiency, in particular VAT evasion. Schneider et al. (2013) used this variable as an indicator when examining the shadow economy, arguing that the informal economy must necessarily be reflected in the formal economy that is captured by the statistical offices. Kasnauskienė and Krimisieraitė

(2015) examined the determinants of the VAT gap using the MIMIC model and used real GDP per capita as one of the indicators to eliminate the effect of inflation. For the purpose of the MIMIC model, the VAT revenue is expressed as the share of VAT collection in the GDP of each country, so it can be comparable with each other. This variable was chosen as a second indicator as tax noncompliance must necessarily be reflected in a decrease in VAT revenue (Kasnauskienė and Krimisieraitė, 2015).

3.2 MIMIC model

MIMIC model is based on the statistical theory of a latent (unobserved) variable which is measured using multiple measurable causes and indicators.

The MIMIC model is a specific type of structural equations model, consisting of two models: a structural model and a measurement model (Schneider et al., 2010). To estimate the variance of a latent variable, the MIMIC model uses unstandardized estimates, which means that the first indicator is always fixed at level 1 and is called the **reference indicator**. All the other estimates change by a given coefficient, if the **reference indicator** changes by 1 (Acock, 2013).

In addition to several measurable (observed) variables, there is also a latent variable in these two models, but in each model, it has a different "role". In the structural model, the latent variable is the dependent variable that is influenced by the measurable variables entering the model. The equation can be expressed as follows:

(1)

$$\eta_t = \gamma' x_t + \varsigma_t$$

Where x_t' is a (1*q) vector of time series x_{it}, $i=1,....,q$ containing potential causes of the hidden variable η_t and γ' is a vector of coefficients expressing the relationship between the hidden variable and their causes. ς_t expresses the error term.

In the measurement model it is true that the latent variable is independent, whereas the measurable variables entering the model are dependent on it.

(2)

$$y_t = \lambda\eta_t + \varepsilon_t$$

Where y' is a (1*p) vector of time series of indicators of the hidden variable, λ is a vector of regression coefficients. ε' is a vector of white noise.

The latent unobserved variable, the VAT gap in this case, is first linked to the observed indicator variables, all within the measurement model. And then the relationships between the latent unobserved variable and the observed explanatory variables (causes) in the structural model are examined.

Using Equation 1 in Equation 2, we obtain a multiple regression model where the explanatory endogenous variables y_{jt}, $j = 1,...,p$ are indicators of the latent variable η and the explanatory exogenous variables x_{it}, $i = 1,....,q$. The model can be expressed by the equation:

(3)

$$y_t = \Pi x_t + z$$

Where $\Pi=\lambda\gamma'$ is the matrix and $z = \lambda\varsigma + \varepsilon$. The error term zis a (p*1) vector of linear combinations of white noise ς and ε from the structural model and the measurement model.

In short, the first step is to confirm or reject the predicted relationship between the VAT gap (latent variable) and its causes and indicators. Once the relationship is confirmed by the MIMIC model, the MIMIC index will be calculated using the structural model equation, Equation 1. Equation 4 expresses the calculation; it is a modified Equation 1.

(4)

$$\tilde{\eta}_t=\gamma x_{1t}+\gamma x_{2t}..+\gamma x_{qt}$$

Where x_{1t} to x_{qt} express the variables of causes at the level of at least 5%. The MIMIC index only expresses the relative development of the latent variable, the VAT gap in this case. In order to calibrate the relative values into absolute values, a baseline variable obtained by another method must be used.

For conversion is used the following equation:

(5)

$$\hat{\eta}_t = \frac{\tilde{\eta}_t}{\tilde{\eta}_{base}} * \dot{\eta}_{base}$$

Where $\tilde{\eta}_t$ expresses the value of the MIMIC index at time t according to Equation 4. $\tilde{\eta}_{base}$ is the value of the MIMIC index in the base period and $\dot{\eta}_{base}$ is an estimate of the latent variable obtained by another method.

3.3 Data

The MIMIC model is built on panel data from selected European countries (a total of 26 countries) between the years 2002 and 2020. Due to data availability, the maximum time series was chosen to provide a sufficient database for the estimation of the MIMIC model. With regard to the availability of EGDI, which represents digitalization, the time series could not start earlier than 2002 and the aim was to get an overview of the VAT gap development up to the most recent possible year.

4 RESULTS

Table 1 presents the MIMIC models. In Model 1 all variables are included. The aim is to retain in the structural model only those variables that will be significant at least at the 5% level of significance. The stepwise selection eliminates statistically insignificant variables step by step, the whole process is shown in Table 1.

When deciding which model is the best possible, information criteria can be of help. In this research Akaike information criterion (hereinafter 'AIC') was used. The most appropriate model is the one with the lowest value of this criterion. Some other indicators of model quality

are also best in the final model. The comparative fit index (CFI) expresses how close a given model is to a perfect fit for the data used. It takes values from 0 to 1 and the higher number means the better model.

The interpretation of coefficients in structural model of MIMIC model is very similar to the interpretation of coefficients in regression analysis. Their value shows the resulting change in the VAT gap for a unit change in the cause variable, under ceteris paribus condition. In the following paragraphs the coefficients from structural model are interpreted according to the model 1, which consist all of the variable, significant and also insignificant. Final consumption is the strongest driver of the VAT gap. According to the model 1, if **final consumption** increases by 1 percent, then the VAT gap also increases by about 5.71 percent. **Final consumption** is significant at level 5%, so it remain in model 2 and also in final model, value of its coefficient fluctuates around 6 to 7%. ¨

Higher **unemployment** increases VAT gap, according to the model 1 if **unemployment** increases by 1 percent, then the VAT gap also increases by 0.01 percent. In model 2 coefficient of variable **unemployment** stays stable, however this variable had to be remove from the final model due to too high p-value.

Higher **openness of economy** also increases VAT gap, if **openness of economy** increases by 1 percent, then the VAT gap also increases by 0.02 percent. **Openness of economy** is significant on needed level of significance, value of its coefficient remain stable in model 2 and also in final model. There is no unified conclusion about effect of **openness of economy** on VAT gap, according to the Aizenman and Jinjarak (2008) its effect on VAT gap is negative, on contrary according to Frunza (2019) open economy create opportunities for carousel frauds and

	Model 1	Model 2	Model 3 *final model*
Structural	Coefficient (P-value)	Coefficient (P-value)	Coefficient (P-value)
Unemployment	0.01	0.01	
Openness of economy	0.02 ***	0.02 ***	0.02 ***
Index of economy freedom	-0.01		
CPI	-0.27 ***	-0.32 ***	-0.14 ***
General government expenditure	-4.44 ***	-5.3 ***	-4.53 ***
Final consumption	5.71 ***	7.08 ***	6.05 ***
EGDI	-2.01 ***	-2.44 ***	-2.20 ***
Measurement			
VAT revenue on GDP	1	1	1
Growth of GDP per capita	13.01 ***	10.89 ***	12.97 ***
Statistics			
CFI	0.851	0.80	0.86
Information criteria			
AIC	850.57	-838.84	-2517.93

Tab. 2 Model MIMIC for Europe-26 (author calculations in the STATA program), VAT revenue on GDP is used as a reference indicator

other VAT frauds, which have positive effect on VAT gap. This research confirms conclusion about positive effect of openness of economy on VAT gap.

Index of economy freedom has a negative effect on VAT gap, if **Index of economy freedom** increases by 1 percentage point, then VAT gap decreases by 0.01 percent. This variable is not statistically significant on needed level of significance. Due to too high p-value this variable had to be remove from model 2 and final model.

CPI has also negative effect on VAT gap. If **CPI** increases by 1 percentage point (it means less perception of corruption), then VAT gap decreases by 0.27 percent. **CPI** is also significant on needed level of significance, so it remains in model 2 and final model. Value of the coefficient remains stable. CASE (Barbone et al., 2013) and Reckon (2009) also examined the influence of CPI on VAT gap and they reached opposite results. This research confirmed results of Reckon (2009) about negative effect of CPI on VAT gap.

General government expenditure has also a negative effect on VAT gap, but this variable is much stronger. If **general government expenditure** or rather its share of general government expenditure on GDP increases by 1 percent, then VAT gap decreases by 4.44 percent. This variable is significant on needed value of significance, so it remains in model 2 and in final model, value of its coefficients fluctuates around 4 to 5.5 percent. As in the case of **openness of economy** there is no unified conclusion about effect of **general government expenditure** on VAT gap, this research confirms conclusion from research of Zídková and Pavel (2016) about negative effect of general government expenditure on VAT gap.

EGDI decreases VAT gap, compared to previous one it is not very strong variable. If **EGDI** increases by 1 percentage point, then VAT gap decreases by 2.01 percent. **EGDI** is significant variable on needed level of significance, so it remains in model 2 and final model. Value of the coefficient remains stable.

Table 2 Model MIMIC for Europe-26 (author calculations in the STATA program), VAT revenue on GDP is used as a reference indicator.

5 DISCUSSION AND CONCLUSIONS

In this research the MIMIC model for quantification of VAT gap was estimated. This model is applicable for 26 European countries. MIMIC index determines year-on-year changes in VAT gap according to the year-on-year changes in significant causes of VAT gap. For second step baseline variable is needed for conversion year-on-year changes in VAT gap into absolute values of VAT gap.

Contribution of the MIMIC model is identification of significant causes of VAT gap, these causes means potential sources of VAT collection inefficiencies. This contribution also hides a huge limitation of the research conducted, which is the omission of an important input variable. Such omission would lead to biased results of the MIMIC model. In order to avoid such omission, a literature search was made of studies on the causes or determinants of the VAT gap. The search included studies by foreign and domestic authors and also authors under the auspices of the European Union, such as CASE or Reckon.

Significant causes include final consumption, general government expenditure, openness of economy and EGDI. Recommendations for the further development of indirect tax policy should be based on statistically significant causes of VAT evasion. A recommendation arising from this research is to focus on the digitalization of tax offices, which can increase the efficiency of VAT collection. Digitalization can also simplify cooperation between tax offices at home and abroad, which could help to eliminate the economy openness factor as well. The paper shows a new insight on the VAT gap through modelling.

Acknowledgements
This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ACOCK, A. C. 2013. *Discovering Sturctural Equation Modeling Using Stata.* Vol. 1. Stata Press. ISBN: 9781597181396.

AIZENMAN, J. and JINJARAK, Y. 2008. The collection efficiency of the Value Added Tax: Theory and international evidence. *Journal of International Trade and Economic Development*, 17(3), 391-410.

ALM, J. 2012. Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance*, 19(1), 54-77.

ALM, J. 2021. Tax evasion, technology, and inequality. *Economics of Governance*, 22(4), 321-343.

ARGENTIERO, A., CASAL, S., MITTONE, L. and MORREALE, A. (2021). Tax evasion and inequality: some theoretical and empirical insights. *Economics of Governance*, 22(4), 309-320.

BAERT, P. 2023. *VAT in the digital age*. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739371/EPRS_BRI(2023)739371_EN.pdf

BARBONE, L., BELKINDAS, M., BETTENDORF, L., BIRD, R. M., BONCH-OSMOLOVSKIY, M. and SMART, M. 2013. Study to Quantify and Analyse the VAT Gap in the EU-27 Member States. *SSRN Electronic Journal*. DOI:10.2139/ssrn.2365189

CARFORA, A., DONGIOVANNI, S., MARABUCCI, A. and PISANI, S. 2020. *The Impact of Domestic Factors and Spillover Effects on EU Countries VAT Gap*. In Argomenti di Discussione.

CHAN, S. G. and RAMLY, Z. 2018. The role of country governance on value-added tax and inequality. *E & M Ekonomie a Management*, 21(4), 79-93.

EUROSTAT. 2016. *Glossary: Final consumption expenditure*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Final_consumption_expenditure

FREY, B. S. and WECK-HANNEMAN, H. 1984. The hidden economy as an 'unobserved' variable. *European Economic Review*, 26(1), 33-53.

FRUNZA, M.-C. 2019. *Value added tax fraud*. New York: Routledge. ISBN: 9781138298293.

GODIN, M. and HINDRIKS, J. 2015. *A Review of critical issues on tax design and tax administration in a global economy and developing countries*. CRED-UNamur, Working Paper, 7.

IMMORDINO, G. and RUSSO, F. F. 2018. Cashless payments and tax evasion. *European Journal of Political Economy*, 55, 36-43.

JONSKI, K. and GAJEWSKI, D. J. 2022. 'VAT Gap'Estimation: Distinguishing Between Informality and Fraud. *EC Tax Review*, 31(3).

KASNAUSKIENĖ, G. and KRIMISIERAITĖ, J. 2015. Using MIMIC models to examine determinants of VAT Gap in Lithuania. *Organizations and markets in emerging economies*, 6(1), 107-126.

MORAVEC, L., HINKE, J. and BORSICZKÁ, M. 2021. VAT Gap Estimation and Influence of Selected Methods of Financial Administration on its Reduction. *DANUBE*, 12(1), 45-60.

NOVYSEDLAK, V. and PALKOVICOVA, J. 2012. *The Estimate of the Value Added Tax Revenue Loss*. Bratislava: Institute for Financial Policy, The Ministry of Finance of the Slovak Republic.

PLUSKOTA, A. 2022. VAT gap determinants in the European Union and Poland. *Ekonomia I Prawo-Economics and Law*, 21(3), 623-633.

PONIATOWSKI, G., BONCH-OSMOLOVSKIY, M., DURAN-CABRÉ, J. M., ESTELLER-MORÉ, A. and ŚMIETANKA, A. 2018. *Study and reports on the VAT Gap in the EU-28 member states: 2018 Final report*. Available at SSRN 3272816.

PONIATOWSKI, G., ŚMIETANKA, A. and BONCH-OSMOLOVSKIY, M. 2020. *Study and reports on the VAT gap in the EU-28 member states: 2020 final report*. CASE Research Paper (503).

RECKON, L. 2009. *Study to quantify and analyse the VAT gap in the EU-25 member states*. Report for DG Taxation and Customs Union, 20, 5-109.

SCHNEIDER, F. 2005. Shadow economies around the world: what do we really know? *European Journal of Political Economy*, 21(3), 598-642.

SCHNEIDER, F., BUEHN, A. and MONTENEGRO, C. E. 2010. *Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007*. World Bank policy research working paper (5356).

TEDDS, L. M. 2005. Revisiting the Latent Variable/MIMIC Model Approach to Measuring the Size of the Canadian Underground Economy. In: SCHNEIDER, F. and BAJADA, C. *Size, causes and consequences of the underground economy*. New York: Routledge. ISBN: 9781138620575

TRANSPARENCY INTERNATIONAL. 2023. *About CPI*. https://www.transparency.org/en/about

THE HERITAGE FOUNDATION. 2023. *2023 index of economy freedom: About the index*. https://www.heritage.org/index/about

UNITED NATIONS. 2023. *E-Government Development Index (EGDI)*. ttps://publicadministration.un.org/egovkb/en-us/About/Overview/-E-Government-Development-Index

ZÍDKOVÁ, H. and PAVEL, J. 2016. What Causes the VAT Gap? 1. *Ekonomicky casopis*, 64(9), 811.

Contact information

Iva Hasíková: e-mail: xhasikov@mendelu.cz

SMART OCCUPANCY DETECTOR FOR MODEL RAILROAD

Jan Horáček[1], Jiří Rybička[2]

[1]Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 13 00 Brno, Czech Republic
[2]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 13 00 Brno, Czech Republic

ABSTRACT

For control of a model railway, simple track occupancy detectors are in use to detect the presence and absence of a vehicle on the track. However, these detectors do not use the full potential of a digital command control (DCC) system. A RailCom technology allows information to be obtained from the ve- hicle's DCC decoder, thus the detector can receive operationally important data from the rolling stock vehicle. The subject of the paper is the description of requirements and design of an own RailCom detector MTB-RC, which presents an alternative to commercially available model railway RailCom detectors. MTB-RC is an open-source and open-hardware project, which is also compatible with the rest of the trackside hardware used in the Track Vehicle Control Laboratory FBE MENDELU. MTB-RC can read the addresses of the DCC decoders present on the track and transmit them to the railway control software via MTBbus.

Keywords: railroad, model, interlocking, RailCom, occupancy detector, Digital Command Control, DCC, economic sustainability

JEL Code: L920

1 INTRODUCTION

The model railway in the Track Vehicle Control Laboratory is a complex computer-controlled electronic system. One of the key components required for the safe operation of a (model) rail- way is a track occupancy detector. The detector detects the presence of a vehicle – an engine or a car – in a section. The railway is then divided into multiple sections, each section is connected to one detector. A small model station with 3 tracks can contain circa 10 sections.

In the model, the track occupancy detector detects whether the current flows through the rails – from one rail to another. In Digital Command Control (DCC) (National Model Railroad Association, 2024b) system, each engine contains a decoder (a small electronic module) which draws some small current continuously. Each axle of a car contains an embedded resistor (15 kOhm) to draw the current as well. Thus any present vehicle in the model track draws current and this current is detected by the track occupancy detector as a voltage drop on a diode (other methods to detect current exist too).

https://doi.org/10.11118/978-80-7509-990-7-0075

In some situations, however, a different approach to occupancy detection is beneficial. An extension of a DCC protocol called *RailCom* allows the information to be read from the vehicle's decoder by the detector (DCCWiki, 2024a). The decoder must support RailCom, but most of the decoders nowadays have the support implemented and even enabled by default. Also, the RailCom detector only reads information from the decoders, so cars are not detected by the detector. Thus, the RailCom detector cannot simply replace the current-based detector mentioned previously. The RailCom detector must be perceived as an extension of the current-based detection. A technology with different usages, different implementation, and different price.

2 DCC

The Digital Command Control (DCC) protocol was created in 1992 by National Model Railroad Association (NMRA) (DCCWiki, 2024b). The association in cooperation with companies producing model rail- road hardware created a DCC Working group, which then released standards describing DCC from electrical and protocol point of view.

The DCC was designed as a one-way protocol. The data flows from the *command station* to the vehicle's decoders via tracks. The command station is controlled by throttles or computers. The main purpose of the command station is to generate a DCC signal. The signal is sometimes amplified by a *DCC booster*. The decoder is usually a small electronic module (e.g. 1×3 cm) with a powerful microcontroller that decodes DCC and controls the vehicle. Each decoder has its address, so different engines can be distinguished. The decoder controls mainly (1) motor, (2) lights, (3) sounds.

DCC transmits data in packets using voltage between rails (National Model Railroad Association, 2024a). We call wires from the com- mand station or a booster to track lines J and K. At each instant of a time, one rail compared to another has either positive or negative voltage. Length of pulses distinguishes logical 0 (100 us pulse) and 1 (58 us pulse). Each packet can contain a different number of bytes (see Figure 1).

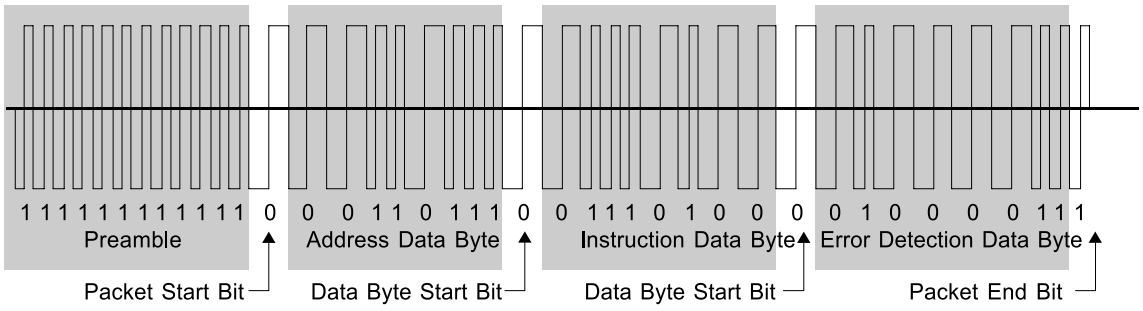


Fig. 1: An example of a DCC packet (National Model Railroad Association, 2024a)

The command station transmits packets one by another, cyclically addressing all decoders it wants to command. It is important to send the data (e.g. speed command) cyclically, because the command station does not know whether the decoder received the command successfully. The decoder could be temporarily busy, the engine could temporarily lose electrical contact with rails, etc.

Detailed specification of a DCC protocol is available in National Model Railroad Association (https://www.nmra.org/sites/default/files/s-92-2004-07.pdf).

3 RAILCOM

In 2006, an NMRA standard (NMRA, 2024) describing an extension of a DCC protocol was released. The standard describes so-called *RailCom protocol*, which allows the mobile decoders to transmit data back to the command station while normally operational.

The principle is as follows: in the space between DCC packets (right side of a Figure 1), the decoder transmits data. The space is called a *cutout*. The decoder contains a capacitor which gives it enough power for transmission. An overview of components participating in a RailCom transmission is shown in Figure 2.

RailCom transmission is current-based. To transmit a 0, the decoder must supply a current of 30+4/-6 mA with a voltage drop at the track of up to 2.2 V. To transmit a 1, the current must be at the most ±0.1 mA (NMRA, 2024). Compared to a voltage-based design, current-based transmission is much more resistant to interferences. However, a special switch in the command station or a booster is required. The switch just connects output J and K wires, so the current loop is closed. There are command stations and boosters with and without RailCom support. Thus, RailCom-supported booster or command station is required.

The last part in 2 is a *detector*. The detector is a device, which reads data from the decoder. The detector senses the current generated by the decoder and reads the data. Note there may be multiple detectors in the current loop, which is beneficial in some situations.

From the data point of view, RailCom transmission is a 250 kbis/s single-direction UART (NMRA, 2024). This is a large gap from DCC communication, which has an average speed of 6 kbit/s.

The decoder transmits data in 2 channels serially. Channel 1 usually contains an address of a decoder transmitting the data, channel 2 contains a response to a DCC command requesting the decoder to respond (e.g. read configuration from the decoder or acknowledgment).

4 REQUIREMENTS

Based on the possibilities of the RailCom, we infer use-cases in which the protocol is beneficial and the requirements for the solution.

We will focus on the address-reading capability of RailCom. There are other important usages – e.g. the command station does not have to resend data to the decoder cyclically, if the decoder confirmed (via RailCom) that it received the DCC packet correctly. However, this paper aims to focus on a decoder-address-reading capability.

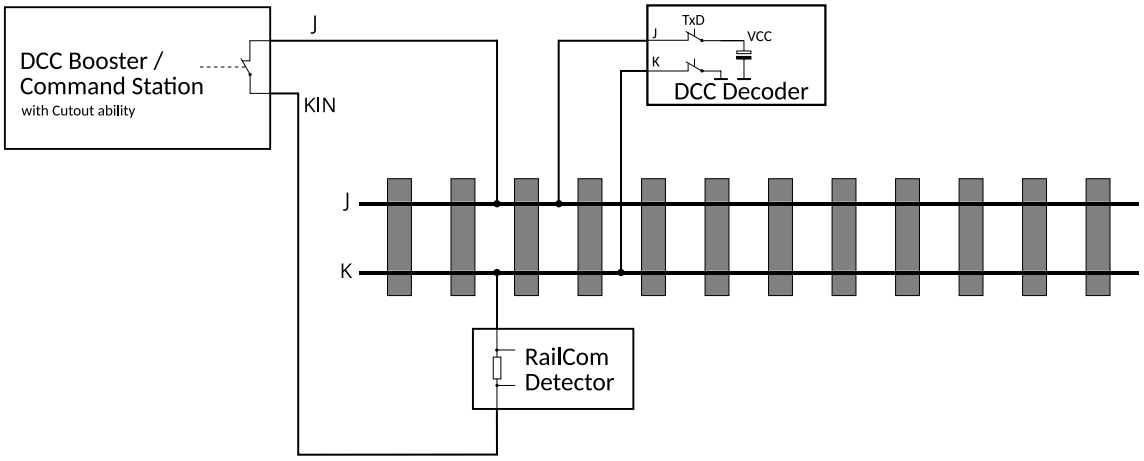


Fig. 2: Components participating in a RailCom transmission

Reading a decoder's address is beneficial in situations when the address is unknown. Usually, when the train is in movement on a computer-controlled railroad like the one in Track Vehicle Control Laboratory, the software already knows its address, because it must have com- manded it to move. We identified the following situations, in which RailCom technology is useful.

1. **Shunting in a train depot.** In the depot, engines are usually changed between rails fast and often. They are even sometimes packed or unpacked by the owner. Moving the responsibility to track vehicle addresses from the operator to a device seems to be beneficial.
2. **Track for new engines.** There is usually a track or tracks at the model railroad specifically designed for mounting new vehicles. The detector may automate the creation of a new train by automatically detecting the new engine's DCC address.
3. **Handover track.** When part of the track is managed by another railroad control system, the detector may be used on the entering track from a foreign system to detect incoming vehicle addresses.

Based on the use-cases, the detector should support several track circuits. The detector should be able to read the decoder's address. If more vehicles are present (e.g. multitraction, multiple engines on a track in a depot, etc.) the detector should be able to read the addresses of all of them. The detector should transmit detected addresses to the *train control system*. The detector should be financially available, customizable, and fixable/replaceable for long-term sustainability.

5 STATE OF THE ART

There are several industrial solutions for obtaining the address of a vehicle on the track.

5.1 LRC120 by Lenz Elektronik

Lenz Elektronik manufactures the *LRC120* RailCom Display module. The module can display the address of the decoder in the connected track section on the four-digit display. The module is available from many vendors, for example, AMB Models (AM Modely, 2024).

The main disadvantage of the module is the absence of an interface for reporting the address.

5.2 10808 detector by Roco

One of the largest companies producing vehicles and equipment for model railways offers the *Z21 10808 detector*. The detector can be connected to the Roco Z21 command station using R- Bus or CAN. It supports 8 sections and allows various settings. To use RailCom, the CAN bus must be used together with Roco Z21 command station (Roco, 2024).

From our point of view, a disadvantage of this solution is a vendor-lock requiring the use of the Roco command station. Deployment would also require adding cabling.

5.3 YD6016LN-RC by YaMoRC (Digikeijs)

The company Digikeijs, which offered a widespread and popular DCC hardware recently announced the end of its activities. The successor company is called YaMoRC, the products are similar.

YaMoRC offers a YD6016LN-RC RailCom track occupancy detector (similar to the former Digikeijs DR5088RC) with 16 track sections. The detector supports per-track channel-1 address reporting and one global detector for reading other data from vehicles. All information from the YD6016LN-RC is transported via the *LocoNet* bus, with firmware updates possible over a USB connection. The module is configurable over USB (YaMoRC, 2024).

Detector	N.o. tracks	Buses	Price
LRC120 *	1	N/A	1482 CZK
Roco 10808 **	8	R-Bus, CAN	2980 CZK*****
YD6016LN-RC ***	16	LocoNet	3200 CZK
RailComDisplay ****	1	N/A	DYI

Tab. 1 Comparison of commercially-available RailCom detectors.
* AMB Modely (2024), ** Roco (2024), *** YaMoRC (2024), **** F. M. Cañada (2024), ***** tvlaky.cz (2024)

However, the module is not available now and its deployment would require additional cabling to connect it to the command station.

5.4 RailComDisplay by F. M. Cañada

A well-known enthusiast for DCC automation F. M. Cañada designed a RailCom module with display, which shows the address of a DCC decoder similar to LRC120 (F. M. Cañada, 2024).

This module is local-only (without any bus), however the big advantage of the RailComDisplay is the availability of schematics. The module is a source of useful inspiration.

5.5 Conclusion

Several commercial RailCom feedback modules exist, and each has its advantages and disadvantages. Overall features are summarized in a Table 1.

For our requirements, we decided to design our own RailCom detector MTB-RC. The detector will be compatible with existing track hardware (no need to add more cabling), and it will be fully customizable as the whole design is upon us. Moreover, we will provide an open-source and open-hardware alternative to the existing commercial products.

6 MTB-RC

An own MTB-RC module was designed. Schematic and PCB design were created in an open software KiCad and are available online[1]. MTB-RC supports the following features.

1. Support for 8 track circuits. This number is a good compromise between scalability and price.
2. Connection to the MTBbus (Jan Horáček, 2024). MTBbus is currently used for stationary hardware control in Track Vehicle Control Laboratory.
3. Open-source and open-hardware design.

6.1 Principles of design

Now we describe important principles of design of the MTB-RC RailCom detector.

The detector is connected to the MTBbus and DCC. These two buses are generally galvanic separated, so the design is composed of two top-level galvanic separated parts.

1. DCC part,
2. MTB part.

[1] https://github.com/kmzbrnoI/mtb-rc

The detector contains a microcontroller (MCU) to process the data. It is not desired for the MCU to be powered off when the DCC to the track is disconnected (e.g. because of a short-circuit on the track). So the MCU lays in the MTB part powered by MTBbus power supply.

The MTB part of the detector contains standard MTB module components like on e.g. MTB-UNI module (Jan Horáček, 2024).

- Jumpers to select the address
- Indication LEDs
- RS485 driver (an integrated circuit)
- Button

An important question is how the data shall be transferred from the DCC part to the MTB part (and vice versa if required). To answer the question we need to understand how RailCom detection works in general.

6.2 RailCom detection principle

The RailCom detection principle is described in NMRA (2024) in detail. Here, we briefly describe key principles.

As we discussed in chapter 3, one of the J and K wires from the command station (or the booster) to the track is simply cut and both ends are connected to the detector. Let's assume the K wire is cut. A Figure 3 shows how these two ends are connected inside the detector.

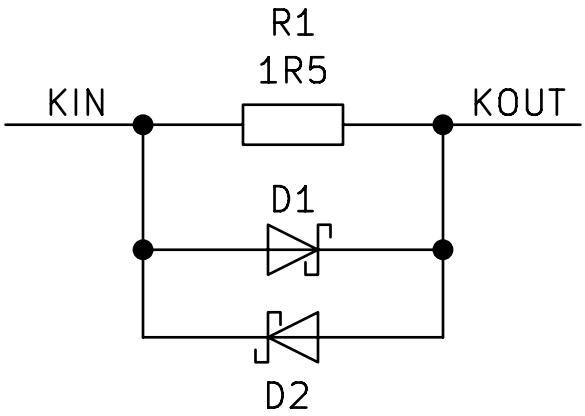


Fig. 3: RailCom current detection principle.

The current flow causes a voltage drop $U_{\text{KIN–KOUT}}$ on R1, which is then measured. In addi- tion, protective Schottky diodes D1 and D2 are added to avoid high temperatures of R1 while the maximum continuous current of 3 A flows. The detector uses adequate parts and adequate resistor size to dissipate the heat. Schottky diodes are used so the maximum voltage between KIN and KOUT is never more than cca 0.5 V.

The voltage $U_{\text{KIN–KOUT}}$ is then compared using window comparator circuit. See Figure 4.

The comparators are powered from e.g. -5 V to +5 V with 0 = KIN. Voltage references ±18.5 mV are defined in (NMRA, 2024).

The comparators have an open-collector output. If $U_{\text{KIN–KOUT}}$ (-18.5, +18.5) mV, (log- ical 0 of RailCom), DATA is connected to -5 V, otherwise (logical 1 of RailCom), DATA is in high-impedance state. Output data are directly in UART format readable by MCU.

MTB-RC contains 8 instances of circuits presented in Figures 3 and 4 to support 8 separate track circuits.

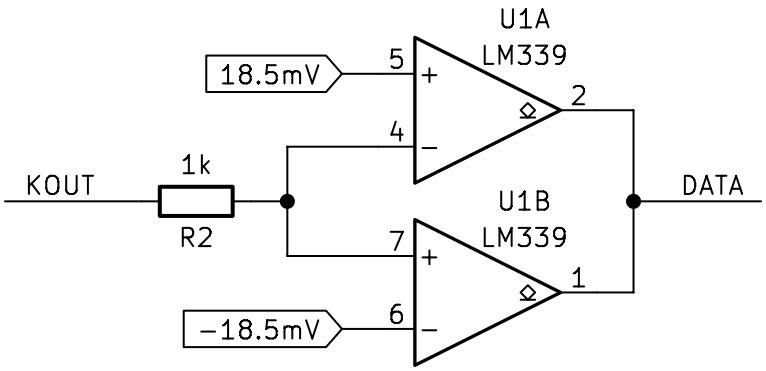


Fig. 4: A window comparator circuit

6.3 MTB-DCC interface

As small MCUs with 8 separate UARTs basically do not exist, multiplexing is used. STM32F103 ARM Cortex-M3 MCU was selected as a main MCU, because of its availability, relatively low price, powerful core, and advanced peripherals (STMicroelectronics, 2024). STM32F103 has 3 UARTs, out of which one must be used for MTBbus communication. The remaining two UARTs shall be used for RailCom reading. This implies 4-way multiplexing. The proposed solution is not ideal, as some data are always lost by design, but the loss should not break address detection.

The multiplexer is present in the DCC part. This implies no MCU in the DCC part is required. The interface between the DCC part and the MTB part consists of:

- 2 UARTs (DCC → MTB)
- 2 multiplexer-control wires (MTB → DCC)
- DCC signal (DCC → MTB)
- Cutout signal (DCC → MTB)

Cutout signal is used to inform MCU that RailCom cutout occurred, so the MCU starts reading RailCom data (via UARTs). DCC signal is transmitted so the processor can read the DCC from the command station, so it knows to which command the decoder answers.

6.4 Multiple decoders detection problem

One of the key problems of RailCom detection is multiple decoders in a section. As discussed in cahpter 3, each decoder sends its address in channel 1, which results in unreadable data when multiple decoders are present in a single track section.

The idea of multiple address detection is to read DCC from the command station and check for RailCom data in channel 2. If the command station sends any command to a decoder present in the track, the decoder (and only the decoder!) responds in channel 2. This allows us to detect decoders, even if there are more of them on the track. A necessary condition is the decoder being addressed by the command station. This issue must be solved separately and is out of the scope of this paper.

6.5 Firmware

The firmware for STM32F103 MCU is implemented in C language and is available online[2]. STM32 HAL library was used. The firmware consists of RailCom decoding, DCC decoding, and MTB communication.

7 CONCLUSION

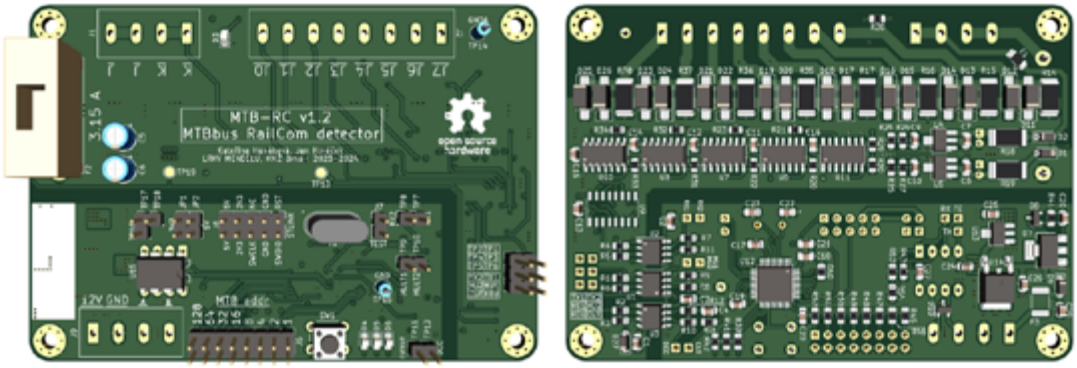


Fig. 5: MTB-RC PCB

Design of the MTB-RC RailCom detector has been described, the module has been manufactured, firmware was implemented and the module passed basic laboratory tests.

Overall parameters of the MTB-RC are summarized in a Table 2. Compare the table to the Table 1.

Detector	N.o. tracks	Buses	Price
MTB-RC	8	MTBbus	~ 600 CZK

Tab. 2 Overall features of MTB-RC.

The detector makes it possible to use additional functionalities of the track operation. One of the functions that we expect to make operations more efficient is the automation of deploying another vehicle in a designated section (most often a locomotive depot). This minimizes the time spent on engine changing and avoids errors caused by entering the wrong address by an operator.

Putting the RailCom detector in operation is the first step to a more reliable and efficient track control system equipped with the ability to read acknowledgments of DCC commands sent to vehicles.

Acknowledgments

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

NATIONAL MODEL RAILROAD ASSOCIATION. 2024a. Communications Standards For Digital Command Control. *NMRA* [online]. https://www.nmra.org/sites/default/files/s-92-2004-07.pdf [visited on 01/11/2024].

NATIONAL MODEL RAILROAD ASSOCIATION. 2024b. DCC Working Group – DCC Home Page. *NMRA* [online]. https://www.nmra.org/dcc-working-group [visited on 01/11/2024].

CAÑADA, F. M. 2024. RailComDisplay – Local display of RailCom data. *RailCom* [online]. https://usuaris.tinet.cat/fmco/railcom_en.html [visited on 01/11/2024].

DCCWIKI. 2024a. History of Digital Command Control. *DCCWiki* [online]. https://dccwiki.com/DCC _ History [visited on 01/11/2024].

DCCWIKIB. 2024b. RailCom. *DCCWiki* [online]. https://dccwiki.com/RailCom [visited on 01/06/2024].

Jan Horáček. MTB – IO bus for model railroad control. https://mtb.kmz-brno.cz/ [visited on 01/13/2024].

ITVLAKY.CZ. 2024 Z21 detector Roco 10808. *itvlaky* [online]. https://www.itvlaky.cz/dcc-vazby/10808/ [visited on 06/24/2024].

AMB MODELY. 2024 LE15120 LRC120 zobrazovací modul adresy RailCom. *AMB Modely* [online]. https://www. ambmodely.cz/le15120-lrc120-zobrazovaci-modul-adresy-railcom/d-8149 [visited on 01/06/2024].

NMRA. 2024. Communications Standard for Digital Command Control Basic Decoder Transmission. *NMRA* [online]. https://www.nmra.org/sites/default/files/s-9.3.2_2012_12_10.pdf [visited on 01/06/2024].

ROCO. 2024. Z21 detector. *Z21* [online]. https://www.z21.eu/en/products/z21-detector-railcom [visited on 01/06/2024].

STMICROELECTRONICS. 2024. *STM32F103x8 Datasheet*. https://www.st.com/resource/en/datasheet/stm32f103c8.pdf [visited on 01/13/2024].

YAMORC. 2024. YD6016LN-RC RailCom enabled feedback module. *YaMoRC* [online]. http://yamorc.de/products/?singleproduct=1246 [visited on 01/11/2024].

Contact information
Jan Horáček: e-mail: jan.horacek@mendelu.cz
Jiří Rybička: e-mail: jiri.rybicka@mendelu.cz

USE OF ANNOTATED IMAGE DATA FOR FRUIT DIVERSITY ANALYSIS

Miroslav Jaroš[1], Jiří Podivín[1], Petr Pernes[1], Oldřich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

This paper deals with a method of development of an annotated image dataset for the detection and classification of plant tissues, aimed at supporting automation in agriculture. The work includes a collection of high-definition image data, their annotation and utility scripts, with the aim of creating a universally accessible dataset for the scientific community. The method is designed to be compatible with off-the-shelf hardware, in order to better support research and development in the field of automated plant identification and plant disease diagnostics. This approach has the potential to significantly improve the efficiency of cultivation processes and support the implementation of advanced technologies in the agricultural sector, along with the automation of this sector.

Keywords: image analysis, plant classification, dataset, learning, annotation, image data

JEL Code: Q16, C55, C88

1 INTRODUCTION

In agricultural engineering, proper fruit detection and counting using image analysis and computer vision algorithms is critical for automating and streamlining processes from measuring phenotypic characteristics to harvesting. The difficulty of this solution lies not only in the availability of the required dataset for learning the model used, but also in the ability of this model to perform its own detection and classification at the level of plants and their tissues. This is quite a challenging discipline, especially due to the diverse structure, color, and considerable number of fruits available, as reported by Ukwuoma, et al. (2022). Until 2018, various machine learning methods were used for detection and classification, and applications using deep learning methods began to gain ground (Zhang, et al., 2014; Then, Kim, 2017), alternatively algorithms published in (Stastny, Skorpil, 2007; Stastny et al., 2021).

Availability of a suitable image dataset is an indispensable prerequisite for training of a model. That is, the availability of a sufficiently comprehensive dataset, consisting of a diverse collection of samples. Considerable number of images is a necessary condition for accurate representation of the actual environment, for example, in orchards and greenhouses, where the fruits may be occluded by the leaves of the plant, overlap, or they may have the same color

https://doi.org/10.11118/978-80-7509-990-7-0084

as the rest of the plant. When an image dataset is available, it is then possible to train a model, or compare existing models and select one that has performance for the selected application deployment (Ukwuoma et al., 2022). Some methods may also include further data augmentation, for example, for better extraction of image features, i.e. the ability to better identify the object being evaluated (Jia et al., 2023), or to increase the number of samples.

Image datasets are created by a sequence of the following steps, which include: image data collection (can involve taking images of selected agricultural crops at different stages of development, at different angles and under different lighting conditions); data labeling and annotation (manual, semi-automatic, or automatic annotation of images, i.e. labeling selected parts of a plant and assigning them to specified classes in order to classify individual crops and plants); splitting the dataset (for purposes of cross validation); optimization and pre-processing of data (adjustment of images with regard to changing the resolution of images in order to train models effectively); incorporation of diversity (extension of the dataset to include a wider range of plants, growing conditions, rotation of images, or use of augmentation to ensure robustness of the model), (Chiu et al., 2020; Wspanialy et al., 2020; Grinblat, et al. 2016). It is worth mentioning that beyond specialized datasets of agricultural crops, there are also general image datasets. Created in order to enable the training of models for the recognition of various objects, and thus also plants. However, there is a limit to the applicability of these resources for training of models for specific use in agriculture (Deng et al., 2009).

In general, the number of available samples (cardinality) and its division into individual classification classes are important for datasets. It is obvious that as the number of classification classes increases, it is necessary to have a larger total number of images. For example, a dataset of weeds, consisting of 14.035 images, was necessary to train a model capable of identifying 25 weed species, with achieved accuracy in the range of 91.8% - 92.4% (Wang et all, 2022). Currently, datasets of various plants and their fruits are available, but the differences are in their robustness, the number of classification classes and the classification ability itself. Other datasets include, for example, the tomato leaf dataset (Chang, 2020) or the tomato fruit dataset (Afonso et al., 2021).

1.1 Description of the created dataset (research content)

As part of our work, we focused on the development of a unique semantic segmentation dataset that is characterized by image resolution and partially automated mask creation. The dataset is characterized by images in 4K resolution, allowing to capture the detailed structure of plants, from leaves to individual fruits (Sapoukhina et al., 2022). This provides a considerable level of detail and allows for deeper analysis and understanding of plant structures.

Unlike other available datasets (Gajjar et al., 2022), the dataset created contains diverse image types that include a wide range of agricultural environment scenarios, including complex backgrounds and various stages of plant growth. The dataset was created by a group of students and is freely available to all interested researchers and companies. This approach makes it possible to provide a unique dataset that is not only more extensive than commonly available, but also provides more detailed information for research and applied purposes. Application deployment of such a dataset is possible at the level of learning models for the identification of plants, plant elements, and diseases. (Fenu et al., 2021). The dataset is published under a “Community Data License Agreement – Permissive, Version 2.0” for free download for all for further research and commercial use.

1.2 Aim of the paper

The aim of the paper is to describe a method of dataset creation and provide a comprehensive dataset of agricultural crops for learning classification and detection models usable in the field of agriculture. Emphasis is also placed on the usability of common available off-the-shelf

hardware for image data collection, such as mobile phones, and the availability of these image sources, with an intended effect of increasing the efficiency of the cultivation process, identifying plant diseases, strengthening local production, deploying automation in the cultivation process while maintaining sustainability (Hughes et al., 2015).

2 METHODOLOGY AND DATA

The methodology describes the methodological framework for image data collection and the creation of a dataset for learning models for classification and detection, including possible extensions and limitations of this approach.

LabelStudio's annotation tool builds on previous toolchains and methods used to label images and includes several complex annotation workflows. Because agricultural imagery has a large number of objects and with more complex shapes than those in many general data sets, we focused on this tool for this, which allowed us to annotate multiple types of annotations and combine them in diverse ways.

LabelStudio was chosen for its flexibility, support for a variety of data formats, and customizability of annotation processes. To ensure consistency of annotations across the dataset, a manual has been created describing the specifications for tagging individual elements such as Fruit, Stem and Leaf with an XML file definition for LabelStudio, see Figure 1. This definition ensures that all annotators act consistently and with respect to the required quality standards.

2.1 Methodology

The annotation methods used by this tool can be applied separately or in combination with other tools to identify each object in an image. These annotated objects can contain several separate segments, allowing for adaptation to occlusion situations. In other words, even if a part of an object is obscured or not fully visible, the system can identify and annotate those parts of the object that are visible. In this way, it is possible to obtain the most complete and accurate annotation of the object despite the presence of occlusal situations. In addition to the main label, each object can also carry additional metadata that provides further contextual information. As an example, certain fruit may have associated metadata describing its ripeness, a text description, or the ID of the plant in question.

For our purposes we have augmented LabelStudio tools with an integrated model for generating segmentation masks, Segment Anything Model (SAM), further referred to as SAM. This integration allowed us to generate accurate segmentation masks with minimum effort compared to manual approach.

The following is a description of the main tools implemented by LabelStudio and used for our methodology:

2.1.1 Keypoint

Keypoint Annotation in LabelStudio, supported by the SAM, is designed to generate masks for objects using a single point, chosen manually by the user of the LabelStudio. The process begins with the user's selection of a point, to which the SAM model responds by assessing and suggesting the optimal location and shape of the mask. The user can approve this mask or add additional points to increase the annotation's accuracy. The system also allows you to define negative points to exclude some areas from the annotation. Although there is an automatic point confirmation feature, we have found that in order to maintain high accuracy of the annotations it is advisable to disable it. This approach makes LabelStudio augmented with SAM a potent tool for users looking for a combination of efficiency and accuracy in image data annotation.

2.1.2 Rectangle

Rectangle annotation in LabelStudio, supported by the SAM model, is an effective solution for identifying and delineating objects in image data. The process begins with the user's selection of the rectangular box that encloses the object. The SAM model then generates a single mask for the object enclosed. This annotation is useful for cases where the objects are already known and SAM is familiar with them, as it helps the user to better determine the boundaries of the object and be more likely to correctly identify the object. This approach is ideal for projects where accurate localization of known objects is key, and SAM provides important support within the annotation process.

Rectangle annotation can be used with great effect to quickly label large objects within the processed image. But it is not well suited to situations when objects are small relative to the image size, or when their boundaries wouldn't fit within a rectangle selection.

2.1.3 Generating Greyscale masks for semantic segmentation

Machine learning's methodology for generating semantic segmentation masks often uses greyscale images for representing class identification of individual pixels.

In practice, multiple annotations of objects in the same class, such as a leafs, stems, fruits, etc., are merged into a single mask, representing all pixels belonging to a specific class in the original image. (Öztürk, B. and Özkar, M. 2022; Guo, S. et all 2020)

2.2 Data

2.2.1 Data Collection

Data was collected in 4K resolution from a variety of different devices, from mobile phones and digital cameras to drones. This multidisciplinary approach allowed us to capture the image material from different angles and in different lighting conditions, which contributes to greater variability and richness of the dataset. Each image in the dataset contains at least ten annotations, which ensures sufficient diversity of objects to be used for further processing.

2.2.2 Uniqueness of the data

Great emphasis was placed on the uniqueness of the data in order to avoid redundancy and ensure broad representativeness of the samples. Each image is unique and brings new value for research purposes. By carefully selecting and structuring the data, we are able to differentiate between different states and types of plant material, which is crucial for machine learning and computer vision applications in the field of agriculture.

The presented methodology and data collection represent an innovative approach in the field of agricultural research. Detailed resolution and complex annotations allow us to analyze and understand the visual characteristics of plants in depth. The results of this project have the potential to significantly contribute to progress in the automation of agricultural processes and the improvement of decision-making mechanisms in this area.

3 RESULTS

The result of our method is a dataset allowing us to both train new models or existing models using the fine-tuning method, for purposes of semantic segmentation of plant tissues. Its uniqueness lies, firstly, in the size of the sample, and secondly, in the way the input data is acquired using consumer hardware such as mobile phone cameras.

Due to the use of the SAM model with the specific annotation method, it became necessary to modify the original export routine used by LabelStudio in order to allow for local export of annotations on demand. Which can be impossible in Label Studio if the dataset is sufficiently large. Furthermore, Label Studio does not natively allow you to merge masks into one, but

generates masks one object at a time, which is unsatisfactory for training purposes. Therefore it was necessary to adjust individual steps of the export process and write a complex script solving the issue. The reason is the use of the SAM model and the annotation method, which the LabelStudio did not expect, but with the advent of SAM, this has fundamentally changed.

Result is a unique dataset of plant organs, with labels for fruit, foliage and stems. With fruit consisting of a mix of apples, rose hips, tomatoes. It is prepared in the expected format and always described in the form of the original file name and then the mask separately for leaves, fruits and stems. Thus, merged masks are the output of multiple annotations of one class by multiple annotators for an individual image.

The use of greyscale mask, combined with advanced machine learning algorithms such as Mask R-CNN, makes it possible to achieve high accuracy in object detection and segmentation in scenarios where it is critical, such as in automated harvesting in agriculture.

This code serves as a template specifying how the SAM model will be integrated into the annotation workflow and LabelStudio GUI. It describes in detail what parameters and settings are used for the correct functioning of the model within the annotation tool, including some specific instructions and rules that determine the behavior of the model when identifying and segmenting objects in images and representation of created masks in LabelStudio GUI.

```
<View>

<Image name="image" value="$image" zoom="true" zoomControl="true" rotateControl="false"

<BrushLabels name="tag" toName="image">

<Label value="Leaf" background="#66ff00"/>

<Label value="Fruit" background="#0000ff"/>

<Label value="Stem" background="#ffff00"/>

<Label value="Flower" background="#ff00d0"/>

</BrushLabels>

<Header value="Please select the KeyPoint"/>
<KeyPointLabels name="tag2" toName="image" smart="true">

<Label value="Leaf" background="#000000" showInline="true" smart="true"/>

<Label value="Fruit" background="#000000" showInline="true" smart="true"/>

<Label value="Stem" background="#000000" showInline="true" smart="true"/>

<Label value="Flower" background="#000000" showInline="true" smart="true"/>

</KeyPointLabels>

<Header value="Please select the Rectangle"/>
<View><Filter toName="tag3" minlength="0" name="filter"/><RectangleLabels name="tag3" t
```

Fig. 1: The image illustrates the XML configuration code used in Label Studio to set up and define SAM for automatic annotation.

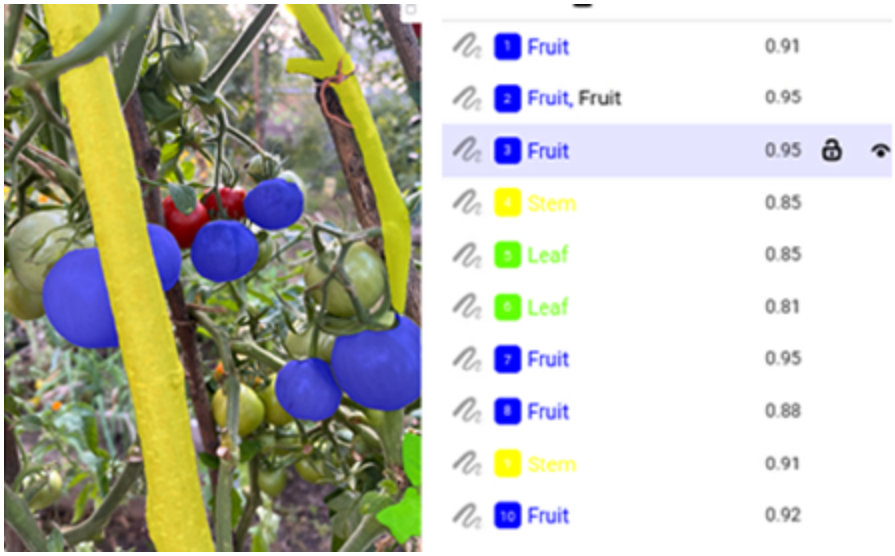


Fig. 2: An example of the Label Studio interface used to create segmentation masks – the masks were generated by the SAM model. On the right is the estimated accuracy of each mask.

For a deeper understanding of how this XML configuration and definition for the SAM model are used in LabelStudio, it is useful to look at the relevant documentation or examples that may be available online. You can visit the LabelStudio website or browse the SAM documentation directly for more information on exactly how this configuration and definition is applied in practice.

Fig. 3: Overlayed with masks. The colors of each mask class have been changed for better visibility.

The use of greyscale images to indicate the affiliation of pixels to individual classes is one of the basic methods of computer vision, specifically in solving visual data segmentation tasks. The advantages of the method include easy coding of the resulting masks, and unambiguous interpretation of the model.

The disadvantage is the possibility of assigning each pixel to exactly one class. This makes it impossible to hierarchically classify objects or recognize individual instances of objects of the same class.

4 DISCUSSION AND CONCLUSIONS

In our efforts to optimize data handling and enhance machine learning applications in agriculture, we evaluated various data export formats offered by Label Studio. The internal representation of segmentation masks in Label Studio uses Run-Length Encoding (RLE) to minimize storage demands. Because of this choice, export of the data, particularly in formats like COCO, places significant computational demands when handling high-resolution images. This challenge has sometimes prevented the complete export of annotated datasets. To address this, we developed utility software that allows for direct data export from the internal format used by Label Studio, thus circumventing the less efficient default export mechanisms.

Creating segmentation masks in Label Studio has demonstrated considerable advantages for agricultural machine learning and computer vision. The interface is intuitively designed, accessible even to non-experts, and is highly adaptable to specific project requirements. However, the substantial volume and high resolution of the processed data pose challenges. Despite these, the potential for significant automation and process enhancement in agriculture through our methodology is vast. Future work will aim to expand this approach to additional plant types and tissues and integrate other technologies such as spectral analysis to further enhance object detection and classification capabilities.

Our dataset presents a significant evolution over traditional datasets like plant village (Hughes et al., 2015), which primarily focus on segmenting specific plant parts or tomato image dataset (Afonso et al., 2021) In contrast, our dataset encompasses a comprehensive scope, capturing entire plant structures – leaves, fruits, and stems–in a single high-resolution image. This extensive segmentation capability facilitates a more detailed and holistic analysis of plant features, crucial for advancing agricultural research and applications.

The diversity of our dataset is also augmented by the variety of data collection devices employed, ranging from mobile phones and digital cameras to drones. This multidisciplinary approach not only diversifies the perspectives and lighting conditions under which the images are captured but also greatly enhances the dataset's variability. This is crucial for developing robust machine learning models as each image is annotated with at least ten distinct labels, ensuring a rich diversity conducive to intricate object detection and segmentation tasks.

The uniqueness of our dataset is underscored by its high resolution and comprehensive annotations. We have meticulously curated our dataset to eliminate redundancy and ensure a wide representativeness of agricultural samples. Each unique image enriches agricultural research, distinguishing between various states and types of plant materials effectively.

Our dataset supports both the development of new models and the fine-tuning of existing ones for the semantic segmentation of plant tissues. Modifications were made to the standard export routines in Label Studio to facilitate on-demand local annotation export, addressing the limitations of handling extensive datasets and merging annotations for effective training.

This innovative dataset not only propels the automation of agricultural processes but also enhances decision-making mechanisms within the sector. Using a greyscale mask in conjunction with advanced algorithms like Mask R-CNN achieves high precision in object detection and segmentation, critical for applications such as automated harvesting. Prepared in expected formats with clearly identified and separated masks for different plant parts, our dataset

is poised for immediate application in cutting-edge agricultural research, marking a significant contribution to the field and promising to drive substantial advancements in agricultural technology and plant science.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

Thanks to the students from the course Fundamentals of Artificial Intelligence for data collection and annotation of dataset data, specifically the following students: Belek Omurzakov, Daniil Astapenko, Jakub Hanuška, Laura Molnárová, Martin Navrátil, Matěj Pour, Jakub Malák, Ondřej Duda, Jakub Ossendorf, Libor Randula, Radek Růžička, Lukáš Kleveta, Martin Weiss, Lucie Suchánková, Filip Kosar, Filip Adámek, Adam Lněnička, Filip Jahn, Josef Pavlíček, Adam Smekal, Tomáš Zatloukal, Martin Kellner, Maxim Kipila, David Vlastník.

Published dataset: https://huggingface.co/datasets/farmaieu/plantorgans (FarmAi u.z., 2024)

REFERENCES

AFONSO, M., FONTEIJN, H., POLDER, G., WEHRENS, R., LENSINK, D., FABER, N. and MOOIJ, M. 2021. *Rob2Pheno Annotated Tomato Image Dataset (Version 3)* [online]. Available at: https://data.4tu.nl/articles/dataset/Rob2Pheno_Annotated_Tomato_Image_Dataset/13173422/3

CHANG, Y.-H. 2020. *Dataset of Tomato Leaves* [dataset online]. https://data.mendeley.com/datasets/ngdgg79rzb/1

CHIU, M. T., XU, X., WEI, Y., HUANG, Z., SCHWING, A., BRUNNER, R., KHACHATRIAN, H., KARAPETYAN, H., DOZIER, I., ROSE, G., WILSON, D., TUDOR, A., HOVAKIMYAN, N., HUANG, T. S. and SHI, H. 2020. *Agriculture-Vision: A Large Aerial Image Database for Agricultural Pattern Analysis (Version 2)* [online]. https://arxiv.org/abs/2001.01306

DENG, J., DONG, W., SOCHER, R., LI, L.-J., KAI, L. and LI, F.-F. 2009. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition* [online]. https://ieeexplore.ieee.org/document/5206848

FARMAI, U. Z. 2024. plantorgans (Version 6bdf588) [dataset]. *Hugging Face*. https://doi.org/10.57967/HF/2023

FENU, G. and MALLOCI, F. M. 2021. DiaMOS Plant: A Dataset for Diagnosis and Monitoring Plant Disease. *Agronomy*, 11(11), p. 2107. MDPI AG. https://www.mdpi.com/2073-4395/11/11/2107

GAJJAR, V. K., NAMBISAN, A. K. and KOSBAR, K. L. 2022. Plant Identification in a Combined-Imbalanced Leaf Dataset. In: I*EEE Access*. Vol. 10, pp. 37882–37891. Institute of Electrical and Electronics Engineers (IEEE). https://ieeexplore.ieee.org/document/9751066

GRINBLAT, G. L., UZAL, L. C., LARESE, M. G. and GRANITTO, P. M. 2016. Deep learning for plant identification using vein morphological patterns. Computers and Electronics in Agriculture, 127, pp. 418–424. *Elsevier BV* [online]. https://www.sciencedirect.com/science/article/pii/S0168169916302242

GUO, P., XUE, Z., LONG, L. R. and ANTANI, S. 2020. Cross-Dataset Evaluation of Deep Learning Networks for Uterine Cervix Segmentation. *Diagnostics*, 10(1), p. 44. MDPI AG [online]. https://www.mdpi.com/2075-4418/10/1/44

HUGHES, D. P. and SALATHE, M. 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics (Version 2). *arXiv* [online]. https://arxiv.org/abs/1511.08060

JIA, W., XU, Y., LU, Y., YIN, X., PAN, N., JIANG, R. and GE, X. 2023. An accurate green fruits detection method based on optimized YOLOX-m. *Frontiers in Plant Science*, 14. Frontiers Media SA [online]. https://www.frontiersin.org/articles/10.3389/fpls.2023.1187734/full

ÖZTÜRK, B. and ÖZKAR, M. 2022. Creating and Using Mask Images for Segmentation in Point Cloud Data. In: The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences: Vol. XLIII-B2-2022 (pp. 1133–1138). *Copernicus GmbH* [online]. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLIII-B2-2022/1133/2022/

PAK, M. and KIM, S. 2017. A review of deep learning in image recognition. In: 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT). *IEEE* [online]. https://ieeexplore.ieee.org/document/8320684

SAPUKHINA, N., BOUREAU, T. and ROUSSEAU, D. 2022. Plant disease symptom segmentation in chlorophyll fluorescence imaging with a synthetic dataset. *Frontiers in Plant Science*, Vol. 13. Frontiers Media SA [online]. https://www.frontiersin.org/articles/10.3389/fpls.2022.969205/full

STASTNY, J. and SKORPIL, V. 2007. Analysis of Algorithms for Radial Basis Function Neural Network. In: *Personal Wireless Communications*. Springer New York, vol. 245, pp. 54-62, ISSN 1571-5736, ISBN 978-0-387-74158-1, WOS:000250717300005

STASTNY, J., SKORPIL. V., BALOGH, Z. and KLEIN, R. 2021. Job shop scheduling problem optimization by means of graph-based algorithm. Applied Sciences, 11(4), 1921. ISSN 2076-3417. https://doi.org/10.3390/app11041921

UKWUOMA, C. C., ZHIGUANG, Q., BIN HEYAT, M. B., ALI, L., ALMASPOOR, Z. and MONDAY, H. N. 2022. Recent Advancements in Fruit Detection and Classification Using Deep Learning Techniques. In: KHAN, D. M. (Ed.). *Mathematical Problems in Engineering*. pp. 1–29. Hindawi Limited [online]. https://www.hindawi.com/journals/mpe/2022/9210947/.

WANG, P., TANG, Y., LUO, F., WANG, L., LI, C., NIU, Q., and LI, H. 2022. Weed25: A deep learning dataset for weed identification. *Frontiers in Plant Science*, 13, Frontiers Media SA [online]. https://www.frontiersin.org/articles/10.3389/fpls.2022.1053329/full

WSPANIALY, P., BROOKS, J. and MOUSSA, M. 2020. An Image Labeling Tool and Agricultural Dataset for Deep Learning (Version 1). *arXiv* [online]. https://arxiv.org/abs/2004.03351

ZHANG, Y., WANG, S., JI, G. and PHILLIPS, P. 2014. Fruit classification using computer vision and feedforward neural network. J*ournal of Food Engineering*, 143, 167–177. Elsevier BV. https://doi.org/10.1016/j.jfoodeng.2014.07.001

Contact information

Miroslav Jaroš: e-mail: qqjaros7@mendelu.cz
Jiří Podivín:e-mail: qqpodivin@mendelu.cz
Petr Pernes: e-mail: qqpernes@mendelu.cz
Oldřich Trenz: e-mail: oldrich.trenz@mendelu.cz

ADAPTIVE DATACENTER MONITORING BASED ON THE LORAWAN NETWORK INFRASTRUCTURE

Andrej Juríčka[1], Jiří Balej[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

High availability and quick response to abnormal situations are the key aspects for a reliable datacenter. Cooperation between physical environment monitoring and high-level cluster / container orchestration could increase the overall durability of the entire system. This paper describes the proposal of an entry-level monitoring system based on the LoRaWAN network infrastructure from a physical point of view to the application point of view. All components are open-source use, without any additional license cost. Compared to typical monitoring applications, the cost-effective and main advantage lies in the interconnection solution for a large datacenter environment. The entire system consists of well-known technologies and applications interconnected via reliable protocols, with the addition of environment-specific rulesets. Based on these preferences, the management of systems such as virtualization or container orchestration systems can be more flawless and energy efficient.

Keywords: datacenter monitoring, LoRaWAN, MQTT, IoT, Prometheus Alertmanager, virtual machine management, container management

JEL Code: C88, L63, L86, L96

1 INTRODUCTION

The implementation of proper monitoring strategy could be a very difficult task to achieve. With the increasing importance and complexity of IT systems, monitoring and management tools become difficult to maintain. Datacenter monitoring is a very complex and wide topic, so the actual possibilities of tools and systems are tremendous. There are already very helpful and practical tools that can add another point of view to the enterprise system. Often, these tools are built on well-known technologies and protocols and are ready to deploy without any special effort.

With the increase of IoT systems in daily life, there is a great possibility to use this technology in datacenters. One of the possibilities is monitoring.

https://doi.org/10.11118/978-80-7509-990-7-0093

The main focus of this paper is the datacenter environment. According to Mehta et al. (2018), the key component in the datacenter environment is power consumption. Power consists of approximately 35% of the operating cost of a data center. IT devices consumes only about 30% to 60% of the overall electric bill (Polonelli et al., 2019). One of the partial goals will be to reduce the costs related to the operation of the data center. This aspect will be discussed further.

Modern datacenter architecture is based on horizontal computing scaling, especially when the entire space belongs to one tenant. This is achieved by technologies such as virtualization or containerization. The main goal is to centrally orchestrate or manage the entire compute power with the ability to effectively use resources. Orchestration tools such as Proxmox or Kubernetes can manage these resources, but with their design patterns they are not capable of reacting to events like temperature, or power feed failure. The interconnection between an IoT ecosystem and API-based orchestration functionality could increase the overall durability and resilience of critical infrastructure services.

The main aim of this work is to propose and implement a flexible monitoring solution that is scalable, robust, cost-effective, and can be easily integrated into existing infrastructure. The system architecture will be discussed in section 3.

2 METHODOLOGY

Before defining the entire system architecture, there were some necessary steps to mitigate future complications during the system design process. One of the biggest crucial steps was the definition of exact system functionality based on current deficiency and demand in datacenter monitoring. This work is focused on one of the main datacenter facilities at Mendel University in Brno.

→ The research between scientific papers was done in order to discover the current state of the art in this area. A few of academic papers were identified, and their benefits will be discussed in the next chapter together with our own design.

→ After an analysis of current academic work, open source, and commercial solutions, a decision was made to create the own system architecture, which should implement all the specified functionality.

→ The prototype of the sensor module was built with all required inputs and outputs. Key functionality was implemented and tested. Consequently, all the required system parts, starting with the IoT sensor module and ending with API calls, were also implemented.

→ Currently, the system is under active testing and further development.

3 SYSTEM ARCHITECTURE

Architecture defines the key aspects of the entire system. The whole system can be divided into the following main sections:

- IoT rack sensor module,
- LoRaWAN network infrastructure,
- Data transmission and processing,
- Rulesets and orchestration management.

The general architecture of the system is shown in Figure 1. Each section of the system is part of a separate topic, following the IoT module.

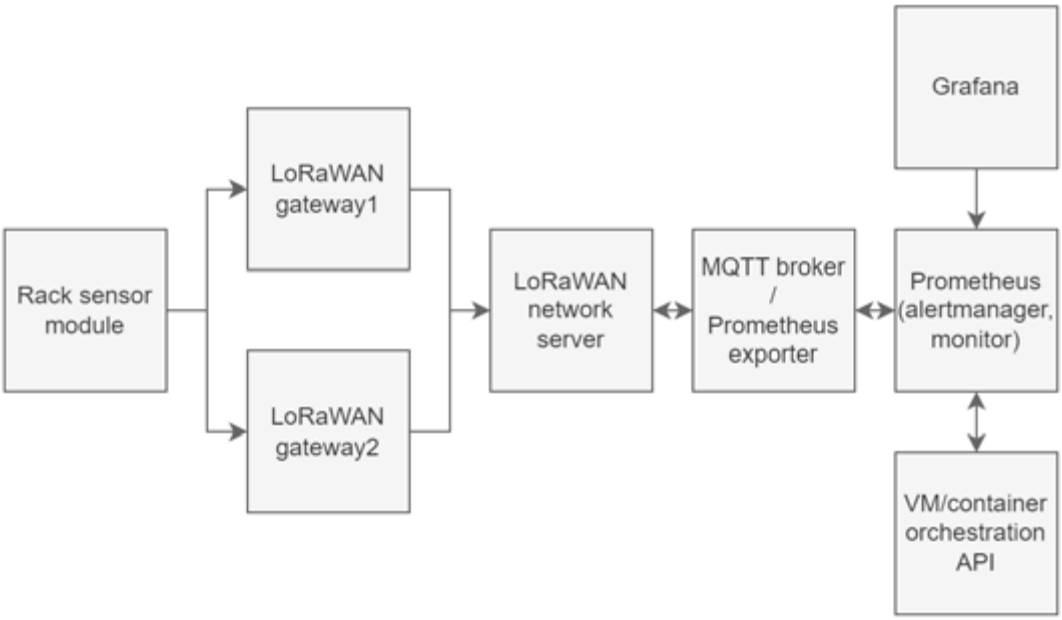


Fig. 1: Monitoring system architecture diagram.

3.1 IoT rack sensor module

Based on the ESP32 SoC (System-on-a-chip), the sensor module is responsible for gathering raw measurements from various components, processing these measurements and send it to the LoRaWAN network server via LoRa interface. A sensor module is intended to be installed in every physical rack enclosure in the datacenter. The power source is taken from both redundant PDU (Power Distribution Unit) to identify potential outages of power feeds.

The sensor module is equipped with the following sensors:

- rack inlet temperature sensor,
- rack outlet temperature sensor,
- humidity sensor,
- rack front door magnetic sensor,
- rack rear door magnetic sensor,
- dual power source sensor.

On the other site, the sensor module implements two output components, a 20x4 character LCD display for quick status observability, and a LoRa transmission module (RFM96) for communication via LoRaWAN network. The prototype of the module is shown in Figure 2.

3.2 LoRaWAN network infrastructure

With the idea of monitoring the whole datacenter, the final number of sensor modules will be close to 50. This is very challenging in terms of communication between those devices and the monitoring server.

According to Polonelli et al. (2019), the LoRa communication protocol is reliable in datacenter environments. The paper mentioned discusses the solution with up to 20 wireless sensors in one room and comparison with various alternative communication protocols, such as Zigbee.

The key components within LoRaWAN infrastructure are device, gateway, and network server (LoRa Alliance, 2024). A device role is accomplished by the IoT rack sensor module, while the LoRa gateway is accomplished by MikroTik routerboard with R11e-LR2 module (Fig. 2). For better fault-tolerance and reliability, there are 2 gateways in the system.

Gateway device forwards all upstream and downstream data between devices and network server. For this design, the LoRaWAN network server on ChirpStack implementation was chosen. ChirpStack is an open-source LoRaWAN Network Server which can be used to set up LoRaWAN networks. ChirpStack provides a web-interface for the management of gateways, devices and tenants as well to set up data integrations with the major cloud providers, databases and services commonly used for handling device data. (ChirpStack, 2024)

An alternative to on-premises solution is one of cloud IoT network providers. The most common and widely used is The Things Network (TTN).

The security between the IoT device and the network server is based on standard LoRaWAN encryption features. The Network Session Key (provides integrity feature) and the Application Session Key (provides confidentiality) are used in the communication process. The entire communication is encrypted by the AES-128 standard. (The Things Network, 2024)

3.3 Data transmission and processing

For the purposes of communication between the LoRaWAN network server and the following Prometheus ecosystem, the MQTT standard was used. MQTT allows fast message transmission between publisher and subscriber agent (OASIS, 2024).

To properly deliver data message from IoT devices to Prometheus monitoring system, there is need for some additional processing. The message is encoded by default; hence additional manipulation is required. After message decoding and translation is done, message data fields are published via Prometheus exporter client. This allows to scrape target metrics directly from the Prometheus monitoring system.

3.4 Rulesets and orchestration management

3.4.1 Prometheus Alertmanager

Upon receiving metrics values, these are stored in the Prometheus time series database. The time series database, which is a crucial part of Prometheus, is also used as a data visualization source. Measured data are visualized using Grafana software and queried via the PromQL language.

Evaluation and monitoring of collected metrics is performed via another Prometheus component, Alertmanager. Alertmanager is responsible for metrics evaluation and the following actions. Examples of implemented alerts: deviating outlet air temperature, state change of rack doors, power feed fail, and so on. An action is the result of active alerts. It is possible to claim that the main monitoring system logic is based on defined alerts.

3.4.2 Orchestration automation of datacenter clusters

The typical control of complex clusters solutions is done via web interface or special GUI clients. Among control methods, controlling the systems via API calls is available too.

This API functionality is used by the monitoring system to perform automation tasks. Currently, covered implementations for virtual machines (VM) and containers management are Proxmox Virtual Environment and Kubernetes. The monitoring system controls the simple behavior of these two solutions.

The typical monitoring use cases are:

- relocate virtual compute resources to another physical location in case of power fail (partial or complete power fail of rack frame),

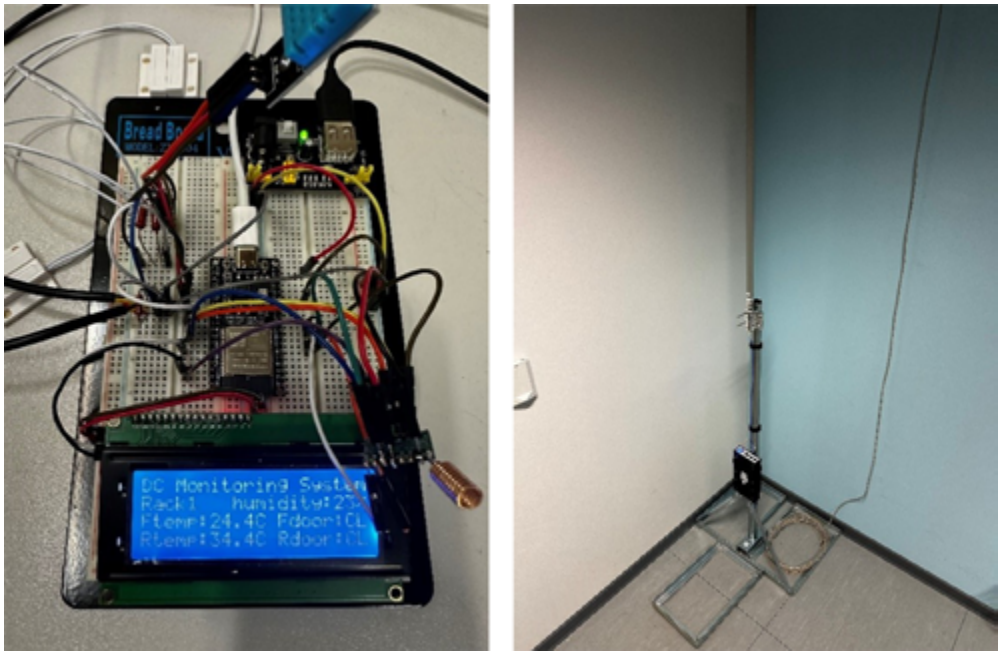


Fig. 2: Prototype of rack sensor module (left), LoRaWAN network gateway (right).

- balance virtual compute resources across multiple racks to accomplish desired temperature and power distribution,
- move physical servers into “maintenance” state, to perform service tasks,
- basic security monitoring (physical rack access).

4 RESULTS

The proposed system was successfully implemented and deployed in a testing environment. All requested features have been implemented. Besides the environment metrics, there is also the possibility to monitor physical security such as rack door state.

The IoT sensors and network server communicate through the LoRaWAN protocol, which is extremely suitable for this type of environment with many devices. Also, communication between network server and Prometheus collector is accomplished by widely used MQTT standard.

All the server-site components (Chirpstack LoRaWAN Network Server, Prometheus monitoring system, MQTT client and Grafana instance) are deployed in high-available virtual environment to achieve greater reliability and robustness.

As mentioned, LoRaWAN communication is secure by default using AES-128 encryption standard. This is secure enough in the context of transferred data.

During the test operation, in comparison with typical and current monitoring workflow, the use of the proposed system has significantly improved reaction time to incident.

5 DISCUSSION AND CONCLUSIONS

Datacenter monitoring is a very crucial topic within the IT industry. This paper describes the proposed monitoring solution based on the LoRaWAN IoT infrastructure. According to academic papers, it is worth using LoRa protocol within this kind of environment.

Selection of the physical and software components was performed based on its compatibility, available documentation, and difficulty of implementation. The final solution consists exclusively with only open-source components, which are typical for their low purchase cost and general availability.

The solution in this paper consists exclusively of on-prem deployment. Typical IoT applications serve as cloud applications, with remote data processing in cloud. One of the possible implementations is edge processing right beside gateways, as mentioned by Truong, (2018).

Possible discussion could be held on the topic of IoT module microprocessor. A similar datacenter monitoring project uses STM32L4 (Polonelli et al. 2019), or Arduino Uno (Santiago et al. 2019) microcontroller. The selection of the ESP32 model was made on its universality and compatibility with all software and hardware components.

This communication topology could be extended not only to one datacenter, but also to the whole university campus to monitor every single network node, due to the LoRa great distance coverage and flexible expansion.

Further work will be focused on the additional level of automatization and possible extension to the ability to monitor another location or monitor building-based rack enclosures. A no less important task will be to implement a more precise reliable protection against the occurrence of unexpected behavior that could have a negative impact on the entire infrastructure of the data center.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

CHIRPSTACK. 2024. Introduction – ChirpStack open-source LoRaWAN® Network Server. *Chirpstarks*. https://www.chirpstack.io/docs/. [Accessed: 4 January 2024].

LORA ALLIANCE. 2024. What is LoRaWAN® S*pecification – LoRa Alliance*®. https://lora-alliance.org/about-lorawan/ [Accessed: 10 January 2024].

MEHTA, G., MITTRA, G. and YADAV, V. 2018. Application of IoT to optimize Data Center operations. *2018 International Conference on Computing, Power and Communication Technologies (GUCON).* https://doi.org/10.1109/GUCON.2018.8674939

OASIS. 2024. OASIS Message Queuing Telemetry Transport (MQTT) TC. *Oasis-open*. https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=mqtt [Accessed: 05 January 2024].

POLONELLI, T., BRUNELLI, D., BARTOLINI, A. and BENINI, L. 2019. A LoRaWAN Wireless Sensor Network for Data Center Temperature Monitoring. *Applications in Electronics Pervading Industry, Environment and Society,* 169-177. https://doi.org/10.1007/978-3-030-11973-7 20

SANTIAGO, A. M., PANO-AZUCENA, A., GOMEZ-ZEA, J. et al. 2019. Adaptive model IoT for Monitoring in Data Centers. *IEEE Access*. https://doi.org/10.1109/ACCESS.2019.2963061

THE THINGS NETWORK. 2024. Learn. *The Things Network*. Available at: https://www.thethingsnetwork.org/docs/. [Accessed: 8 January 2024].

TRUONG H., 2018. Enabling Edge Analytics of IoT Data: the Case of LoRaWAN. *2018 Global Internet of Things Summit (GIoTS).* DOI: 10.1109/GIOTS.2018.8534429.

Contact information

Andrej Juríčka: e-mail: andrej.juricka@mendelu.cz
Jiří Balej: e-mail: jiri.balej@mendelu.cz

EVOLVING LANDSCAPE OF ARTIFICIAL INTELLIGENCE IN GEORGIA

Nadia Mtchedlidze[1], Zuzana Papulová[1]

[1]Department of Strategy and Business, Faculty of Management, Comenius University, Odbojárov 10, 820 05, Bratislava, Slovak Republic

ABSTRACT

The current paper provides a detailed evaluation of Artificial Intelligence (AI) adoption in Georgia, identifying the opportunities and challenges within political, economic, social, technological, legal, and environmental contexts. We developed a novel theoretical framework to characterize AI stakeholders and used an Autoregressive Distributed Lag (ARDL) model to investigate how AI influences macroeconomic indicators like high-technology exports. The findings indicate significant positive short-term and long-term impacts of R&D expenditure on high-technology exports, with ICT goods exports also contributing positively over time. In contrast, real GDP negatively affects these exports, suggesting the need for policy adjustments to support AI implementation. The study highlights the importance of strengthening policy frameworks and promoting digital education to enhance AI integration in Georgia's digital strategy.

Keywords: Artificial Intelligence, Developing Economy, Georgia, ARDL

JEL Code: O33, O14, O10

1 INTRODUCTION

In the modern digital world, artificial intelligence (AI) is developing dynamically. In many businesses it is identified as a source of increasing efficiency and generating insights for better decision making.

In a highly diverse developing world, positive relationship is observed between the digital transformation index and economic development, labour productivity and job employment (Ali, 2022). According to Dahlman, Mealy, and Wermelinger (2016), "the digital economy fosters growth and productivity and supports inclusive development".

Considering the international experiences, the AI may be an important element in accelerating development and economic growth in developing countries, but its major impacts are still not fully investigated.

Our objective is to analyse the impact of AI on the example of small-scale developing economies. For this purpose, the Caucasus region and in particular Georgia was chosen.

https://doi.org/10.11118/978-80-7509-990-7-0099

The purpose of this paper is to explore AI current state in Georgia and identify the factors that would influence AI future adoption in the country. Therefore, the main research question is, “What challenges is Georgia facing in adopting AI?” This study endeavors not only to outline these challenges but also to propose strategic directions for policy and implementation that could facilitate a smoother integration of AI technologies in Georgia.

2 LITERATURE REVIEW

Review of the literature indicates a scarcity of academic sources regarding the utilization of artificial intelligence in Georgia. Notably, scholarly works on this subject have only recently emerged.

Researchers have raised questions about the establishment of an AI national strategy and an appropriate strategic framework (Eristavi, D., Davituri, G. 2021). According to the PMC research paper, “having a clear strategy could help to coordinate several governmental policies and ensure that there is no contradiction between the AI strategic goals and certain sectoral goals” (Parulava, 2021).

Gigiuashvili and Makasarashvili (2021) also reviewed the potential of artificial intelligence for Georgia in their work – “possibilities of using artificial intelligence in post-pandemic Georgia”. In their article, authors speak about the necessity of development of artificial intelligence technologies in Georgia. They also note that a possible risk of AI adoption can be rising unemployment.

Abuselidze and Mamaladze (2021) also raised the issue of unemployment and possibility of a reduction in labour demand. The authors recommend that the state should pay special attention to the creation and development of educational programs on artificial intelligence.

Napetvaridze (2022) in his work “Artificial intelligence in Georgia and in the world” notes that the development of digital technologies is an irreversible process. The author believes that artificial intelligence can be a source of increasing efficiency in the provision of public services and ensuring the participation of citizens in the decision-making process.

In summary, the literature review highlights a growing interest in the strategic development and implications of artificial intelligence in Georgia, yet reveals a critical lack of comprehensive research on the subject.

3 METHODOLOGY

The research begins with a literature review to identify what previous studies have reported about AI technology specifically in Georgia. We examine academic articles, corporate reports, and policy documents to spot gaps.

Following the literature review, we introduce a novel framework designed specifically for this research to describe and analyze the AI ecosystem in Georgia. This framework details the components and stakeholders within the ecosystem and explores how they interact and affect AI development.

Prior to conducting the quantitative analysis, a PESTLE analysis is performed. This analysis assesses the Political, Economic, Social, Technological, Legal, and Environmental factors that could influence the AI ecosystem in Georgia. The PESTLE analysis helps to identify external factors that might affect the development and deployment of AI technologies, thereby providing a macro-environmental backdrop against which the AI ecosystem operates.

The main quantitative analysis uses an Autoregressive Distributed Lag (ARDL) model to investigate how R&D expenditure, ICT goods exports, and real GDP influence high-technology exports. The ARDL model is suitable for our study as it can analyze both the immediate and delayed effects of these economic variables. We test the following hypotheses:

- H1: Increased R&D expenditure can boost high-technology exports in the short term.
- H2: Over the long term, a higher share of ICT goods exports significantly enhances high--technology export growth.
- H3: Real GDP negatively impacts high-technology exports in the short term, requiring careful economic policy adjustments.

The findings will not only fill existing gaps in the current research but also offer valuable insights and frameworks that can be adapted by other researchers and policymakers in similar emerging markets.

4 GEORGIAN CONTEXT

According to the "Government AI Readiness Index 2023" of Oxford Insight, Georgia ranks 99th in the global context out of 193 countries worldwide. It positions 10th out of 18 countries on a regional scale (South and Central Asia) and falls behind its neighbour countries: Turkey, Azerbaijan and Armenia (AI Readiness Index 2023).

According to the Index indicators, the country has no AI vision, maturity and lacks infrastructure. However, Georgia scored the highest in terms of Data Availability and Data Representation. The state has made significant improvements in governance, ethics, and digital Capacity. With insights gleaned from AI Readiness Index, it becomes imperative to delve further into Georgia's artificial intelligence landscape. Consequently, we have created a novel theoretical framework to analyze the state of AI ecosystem in Georgia.

In Figure 1, we present the Georgian Artificial Intelligence Ecosystem, composed of the following groups: local and foreign investors, government, businesses, and local community. In the center of the model, we placed the business association AI Georgia. Its aim is to promote and raise awareness of artificial intelligence throughout the country. The organization also promotes the introduction of artificial intelligence into the private sector and fosters the opening and maintenance of dialogue between enterprises, government, and state legislatures (AI Georgia).

If all bodies would strengthen cooperation between each other, we would observe AI enhancement, opportunities could be maximized, and risks could eliminate. However, the limitations of this framework are that it does not account the differences in the influence and impact of stakeholders. These are complex issues and more in-depth research is required.

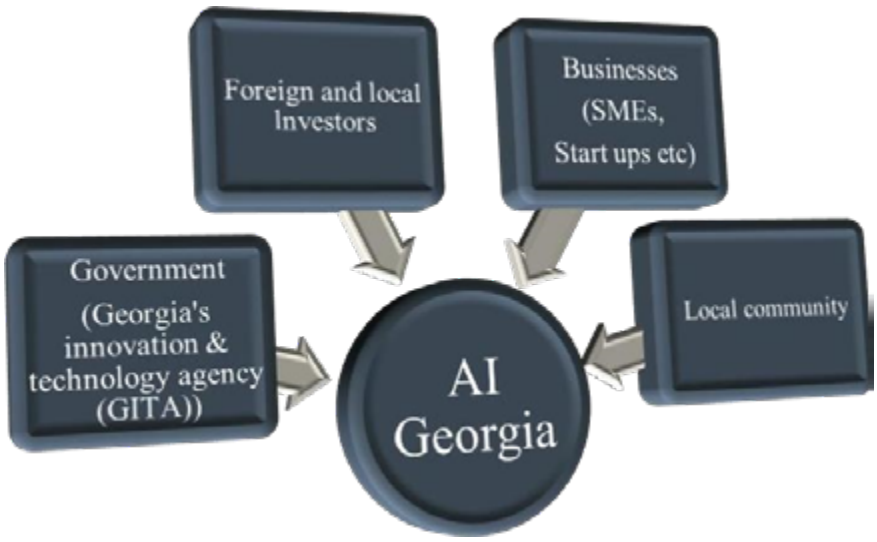


Fig. 1: A description model for assessing the AI ecosystem and its current agents in Georgia.

4.1 PESTLE Analysis addressing AI adoption in Georgia.

The aim of the PESTLE analysis of this research is to develop a deep understanding of the external environment affecting the stages of artificial intelligence development in Georgia.

Political Factors can play a crucial role in the AI adoption and development in Georgia. Georgian legislation doesn't define a concept of artificial intelligence, neither special legislation regulating artificial intelligence software services do exist (Eristavi, D., Davituri, G. 2021). We also believe that artificial intelligence should be given a special importance in Georgia's legal framework and that its application to other legislative requirements should be transparent. According to the OECD (2021), the country has undertaken efforts to develop an institutional and policy framework for digital transformation. The adoption of a comprehensive national digital strategy would help define clear objectives and measures and strengthen cooperation between stakeholders. Furthermore, the central point of AI policy implementation should be investment in R&D (OECD 2022).

Economic factors to consider are the cost of adopting AI and the costs of maintaining AI systems. As mentioned in the report of McKinsey Global Institute (2018), Initial investment, the continuous improvement of technology and applications and significant transitional costs could limit the adoption of AI by smaller companies. Unsurprisingly, the transition to AI is likely to cause number of costs for Georgian business environment since it is mostly composed of small firms. Other economic factors include the impact on employment rate, economic growth, and employee skills. These are sensitive indicators for the country.

AI adoption leads to human capital development and fosters new skills development. Wilson et al. (2017) conducted research on jobs in AI-driven businesses and technology.

They confirmed that companies that use advanced artificial intelligence systems need employees who can explain to non-technical professionals the internal workings of complex algorithms and introduced three new roles. Trainers are human workers who can teach AI systems how to operate. Explainers bridge the gap between technology and business leaders. Sustainers — will help ensure that the AI system works as designed. In Georgian AI driven business environment, we can expect the same scenarios. Higher AI adoption could lead to the development of ICT skills.

Social factors, particularly how society views and interacts with AI, are another important aspect to consider. Therefore, it is crucial to determine if any public fears exist, as cultural values and technological literacy may influence how Georgian society perceives and adopts AI. The Georgian population currently has limited exposure to artificial intelligence, and as a result, they are not well-informed about this phenomenon. It is undeniable that understanding the extent of misunderstanding among the general public regarding AI is of paramount importance. Another research question that arises here is to what extent the hopes and expectations of the Georgian people align with reality. We believe that conducting a survey of respondents can be an ideal way to explore this subject and should be carried out accordingly.

Technological factors, related to the development and use of AI, can significantly impact the success of AI projects. According to the National Statistics office, In Georgia, the ICT sector grows and the number of employees in the sector has grown as well. There were 2.86 million internet users in Georgia and internet penetration rate stood at 76.4 percent of the total population at the start of 2023 (Digital Georgia 2023). The penetration rate of the Internet in Georgia is estimated at 73.5 percent by 2026. The growth is expected in the mobile phone subscriptions and the number of fixed-broadband subscriptions per 100 inhabitants. This proves that technological environment is dynamically developing in the country and artificial intelligence development can be based on solid foundations. Moreover, the availability of data is another vital technological factor influencing AI development. Although Georgia established the Open Data Portal in 2015, challenges persist in terms of data availability (Eristavi, D., Davituri, G. 2021). The Global Data Barometer reveals that Georgia faces obstacles in open

Political	Economic	Social
• Governmental policies • Investments in AI • Grants and fundings • Impact on National Security • Impact on labor laws	• Cost of AI adoption Potential economic impact on AI • Impact on employment rate and job markets Economic barriers • Impact on skill market	• AI public perception • Positive / Negative impacts of AI solutions on society • Technological literacy of the society
Technological	**Legal**	**Environmental**
• Data availability • Technological infrastructure • Technology changes	• Local laws and regulations • Intellectual property protection for AI-related patents, copyrights etc. • Date privacy laws	• High energy consumption of AI systems • AI equipment recycling and disposal. • Sustainable development

Tab. 1 PESTLE Analysis addressing AI adoption in Georgia

data governance due to the absence of a framework for exchanging public data. However, advancements in cloud computing technologies offer promising prospects for accelerating AI development, reducing costs, and enhancing flexibility, particularly for small and medium-sized enterprises (ITU 2021).

Legal factors include local laws and regulations that can affect AI further adoption.

The most relevant legal instrument in the field of Artificial Intelligence is the Georgian Personal Data Protection Act, which sets the standards for the collection and possession of data (Parulava 2021). The Georgian National Bank has developed regulatory measures for the risk management of data-based statistical, artificial intelligence and machine learning models to encourage the appropriate use of models and reduce potential risks, aimed at establishing a framework for effective risk management. Model Risk Management Standards consider the current practices and challenges of the Georgian financial sector and modern international supervision experiences (NBG 2022). In parallel, Georgia's oversight of intellectual property rights is facilitated by the government agency Sakpatenti, with the country holding longstanding membership in both the World Intellectual Property Organization (WIPO) and the World Trade Organization (WTO).

Environmental factors can have a crucial importance. The energy consumption associated with AI compute has direct implications for production, transportation, and operations, contributing to factors such as carbon footprints and water consumption (Wynsberghe 2021). However, there is a growing acknowledgment of the need for sustainable approaches to AI development. Scholars have begun exploring the concept of "sustainable artificial intelligence," which prioritizes the compatibility of AI development with environmental conservation efforts (Yu, Zhang, et al. 2021).

Additionally, the disposal and recycling of artificial intelligence equipment present environmental challenges that need consideration, particularly as AI technology becomes more prevalent in Georgia. Ensuring responsible environmental management of electronic waste is crucial for the long-term sustainability of artificial intelligence. One potential solution could be implementing frameworks such as the Artificial Intelligence-based Hybridized Intelligent Framework (AIHIF) for automated recycling, which optimizes waste management processes (Wisskirchen et al. 2017). By addressing these environmental factors, Georgia can foster sustainable AI development while minimizing its ecological footprint.

4.2 Exploring the Economic Dynamics of AI Development

We aim to investigate the dynamics of AI development in Georgia, focusing on its technological export capabilities. Our study employs an Autoregressive Distributed Lag (ARDL) model to analyze both short-term and long-term relationships between key economic variable.

The dependent variable is high-technology exports (% of manufactured exports). Independent variables include research and development expenditure (% of GDP), ICT goods exports (% of total goods exports), and real GDP.

R&D expenditure drives innovation, while ICT goods exports reflect the digital economy's role in technological advancement. Real GDP captures the broader economic environment's influence on technological exports.

4.2.1 4.2.1 Model Specification and data preparation

The general form of the ARDL model for our study is specified as follows:

(1)

$$\mathrm{HTE}_t = \alpha_0 + \sum_{i=1}^{p} \beta_i \, \mathrm{HTE}_{t-i} + \sum_{i=0}^{q} \gamma_i \, \mathrm{RD}_{t-i} + \sum_{i=0}^{s} \lambda_i \, \mathrm{ICT}_{t-i} + \sum_{i=0}^{t} \phi_i \, \mathrm{GDP}_{t-i} + \epsilon_t$$

Our analysis uses data from the National Statistics Office of Georgia and the World Bank, covering the period from 2013 to 2021. We disaggregate these annually reported variables into quarterly data using the Denton-Cholette method (Sax C, et. al 2013). We perform log transformations on all variables to stabilize variance and normalize data distribution. For real GDP, we additionally adjust for seasonality using the X-11 method, removing seasonal effects to better analyze underlying trends and cyclical patterns.

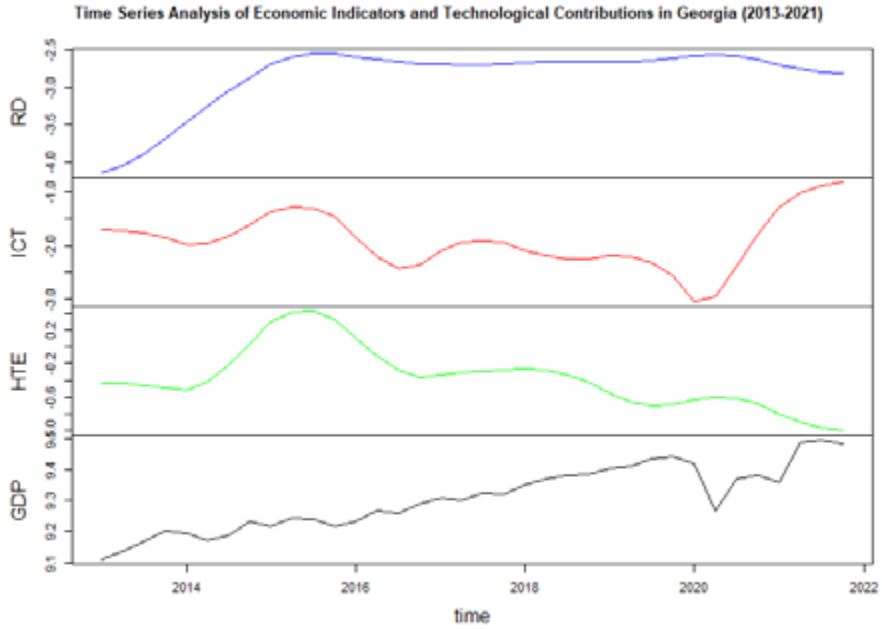


Fig. 2: Time Series Analysis of Economic Indicators and Technological Contributions in Georgia (2013-2021)

Source: Analysis conducted by the authors using R software

Based on our analysis of the time series graphs for economic indicators and technological contributions in Georgia from 2013–2021, as illustrated in Figure 2, the data exhibits non-stationary behavior overall. The variables display level shifts, trends, and lack of mean reversion, indicating that the time series are not stationary. Further tests, such as the Augmented Dickey-Fuller (ADF) test, could be employed to formally assess the stationarity of each series.

We first perform the ADF test to determine the integration order of each variable. If our variables are either I(0) or I(1), the ARDL model is appropriate. Given this, we proceed with the ARDL approach, treating all variables as I(0) for model estimation.

Level	1pct	5pct	10pct	Statistic
l.HTE.tau3	-4.15	-3.5	-3.18	-5.7215978502874
l.HTE.phi2	7.02	5.13	4.31	18.8835095224786
l.HTE.phi3	9.31	6.73	5.61	16.9827413362574
First difference				
d.l.HTE.tau3	-4.15	-3.5	-3.18	-3.99947469703212
d.l.HTE.phi2	7.02	5.13	4.31	7.99599027254634
d.l.HTE.phi3	9.31	6.73	5.61	11.3677229951813

Tab. 2 ADF Test results for HTE
Source: Analysis conducted by the authors using R software

Level	1pct	5pct	10pct	Statistic
l.RD.tau3	-4.15	-3.5	-3.18	-1.54654201198179
l.RD.phi2	7.02	5.13	4.31	1.56622449067009
l.RD.phi3	9.31	6.73	5.61	2.27086637597985
First difference				
d.l.RD.tau3	-4.15	-3.5	-3.18	-3.48547315494067
d.l.RD.phi2	7.02	5.13	4.31	4.36457208490618
d.l.RD.phi3	9.31	6.73	5.61	6.35389321107155

Tab. 3 ADF Test results for R&D
Source: Analysis conducted by the authors using R software

Level	1pct	5pct	10pct	Statistic
l.ICT.tau3	-4.15	-3.5	-3.18	1.67755429917397
l.ICT.phi2	7.02	5.13	4.31	1.67755429917397
l.ICT.phi3	9.31	6.73	5.61	2.46524873313422
First difference				
d.l.ICT.tau3	-4.15	-3.5	-3.18	-4.1725977887524
d.l.ICT.phi2	7.02	5.13	4.31	5.87290319016142
d.l.ICT.phi3	9.31	6.73	5.61	8.78550198282987

Tab. 4 ADF Test results for ICT
Source: Analysis conducted by the authors using R software

Level	1pct	5pct	10pct	Statistic
l.GDP.tau3	-4.15	-3.5	-3.18	-3.14531048765588
l.GDP.phi2	7.02	5.13	4.31	4.10584732956181
l.GDP.phi3	9.31	6.73	5.61	5.07095660316153
First Difference				
d.l.GDP.tau3	-4.15	-3.5	-3.18	-4.76590956323406
d.l.GDP.phi2	7.02	5.13	4.31	7.6552314322281
d.l.GDP.phi3	9.31	6.73	5.61	11.4794755757278

Tab. 5 ADF Test results for GDP
Source: Analysis conducted by the authors using R software

The results of the ADF test indicate the integration orders of the variables as follows:

- High-technology exports (HTE) is integrated of order 0 (I(0)).
- Research and development expenditure (RD) is integrated of order 1 (I(1)).
- ICT goods exports (ICT) is integrated of order 1 (I(1)).
- Real GDP (GDP) is integrated of order 1 (I(1)).

These results confirm that the ARDL model is suitable for our analysis, as it can handle a mix of I(0) and I(1) variables.

4.2.2 Model Identification

Before estimating the ARDL model, we transformed our variables and ensured their stationarity. We used the function in R to determine the optimal number of lags based on the Bayesian Information Criterion (BIC), which balances model fit and complexity.

Estimation of the ARDL model was performed using Ordinary Least Squares (OLS). The function results indicated that the optimal lag structure, based on the Bayesian Information Criterion (BIC), is with lags of 2 for HTE, 2 for RD, 0 for ICT, and 1 for GDP, achieving the lowest BIC value of -86.621. This model captures the necessary past information most efficiently for our analysis.

We used the Variance Inflation Factor (VIF) to detect multicollinearity, which occurs when independent variables are highly correlated, which can cause inflated standard errors, making it difficult to determine the individual effect of predictor variables and leading to unreliable statistical inference.

	VIF
L(l.HTE, 1)	49.2794
L(l.HTE, 2)	36.8579
l.RD	436.4411
L(l.RD, 1)	2121.4321
L(l.RD, 2)	789.4687
l.ICT	1.7782
l.GDP	8.2569
L(l.GDP, 1)	6.4074

Tab. 6 Multicollinearity Assessment of Economic Indicators
Source: Analysis conducted by the authors using R software

Term	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.70770	1.33755	5.7626	0.00001
l.RD	3.34820	0.38190	8.7672	0.00000
l.ICT	0.04004	0.01178	3.3993	0.00236
l.GDP	-0.42750	0.14367	-2.9756	0.00658

Tab. 7 Short-run Multipliers
Source: Analysis conducted by the authors using R software

The VIF results indicated potential multicollinearity issues among the variables in the ARDL model. High VIF values for lagged high-technology exports (HTE) and research and development expenditure (RD) suggest significant correlation with other variables. Lower VIF values for ICT goods exports (ICT) and real GDP (GDP) imply less multicollinearity for these variables.

We conducted diagnostic tests on our regression model's residuals. The Breusch-Godfrey test showed no significant autocorrelation ($p > 0.05$). The Breusch-Pagan test indicated constant variance of residuals, with no significant heteroscedasticity (BP = 5.6, df-8, $p = 0.69$). The Jarque-Bera test confirmed that the residuals follow a normal distribution ($p = 0.24$).

The validation tests confirm that our ARDL model's residuals show no significant autocorrelation, heteroscedasticity, or deviation from normality, demonstrating that key assumptions are satisfied. However, despite the model's robustness, we observed potential multicollinearity among some variables, which should be considered in parameter interpretation. These findings support the overall reliability of our model.

The regression results show a highly significant model (p-value < 2e-16) with an adjusted R-squared of 0.995, indicating an excellent fit. Most coefficients are statistically significant, suggesting strong relationships between high-technology exports and the independent variables.

Key findings include:

- Lagged high-technology exports (HTE) have a significant positive effect at lag 1 and a negative effect at lag 2.
- Research and development expenditure (RD) has a significant positive immediate effect and significant negative and positive effects at lags 1 and 2, respectively.
- ICT goods exports (ICT) and real GDP (GDP) show significant effects, with ICT positively and GDP negatively influencing high-technology exports.

These results suggest that past values of HTE, RD, and GDP significantly impact current high-technology exports, highlighting the importance of these factors in AI development in Georgia.

Term	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	39.89188	4.27595	9.3294	0.00000
l.RD	2.12219	0.27776	7.6404	0.00000
l.ICT	0.20725	0.05893	3.5170	0.00177
l.GDP	-3.66500	0.43692	-8.3883	0.00000

Tab. 8 Long-run Multipliers
Source: Analysis conducted by the authors using R software

The short-run multipliers from the regression results show the immediate impact of each independent variable on high-technology exports:

- A 1% increase in R&D expenditure boosts high-technology exports immediately by approximately 3.35% ($p < 0.00001$).
- 1% rise in ICT goods exports results in an approximate 0.04% increase in high-technology exports ($p = 0.00236$).
- A 1% increase in real GDP decreases high-technology exports by about 0.43% ($p = 0.00658$).

The long-run multipliers from the regression results show the sustained impact of each independent variable on high-technology exports:

Over the long term, a 1% increase in R&D expenditure correlates with a 2.12% rise in high-technology exports ($p < 0.00001$).

A 1% rise in ICT goods exports leads to a 0.21% increase ($p = 0.00177$).

A 1% increase leads to a long-term decrease of about 3.67% in high-technology exports ($p < 0.00001$).

The findings from our ARDL model analysis provide substantial evidence supporting the formulated hypotheses regarding the impact of R&D expenditure, ICT goods exports, and real GDP on high-technology exports in Georgia. Specifically:

- Hypothesis 1 was clearly supported, as R&D expenditure showed a significant positive impact on high-technology exports in the short term, highlighting its immediate benefits to technological export capabilities.
- Hypothesis 2 received partial support, indicating a modest yet positive effect of ICT goods exports on long-term high-technology export growth.
- Hypothesis 3 was also confirmed, with real GDP negatively affecting high-technology exports in the short term, emphasizing the necessity for strategic economic adjustments.

The robustness of the ARDL model, confirmed by diagnostic tests showing no issues with autocorrelation, heteroscedasticity, or normality deviations, lends further credibility to these results. However, the presence of multicollinearity suggests caution in the interpretation of some effects, particularly those related to overlapping economic variables.

Overall, these insights not only validate our hypotheses but also highlight the necessity for continued R&D investment and the strategic support of the digital economy. For policymakers, these findings emphasize the importance of addressing the complexities introduced by real GDP dynamics to bolster Georgia‘s position in the global high-technology market.

5 DISCUSSION AND CONCLUSIONS

This study evaluated the adoption of Artificial Intelligence (AI) in Georgia, highlighting both opportunities and challenges. We introduced a new framework for analyzing AI stakeholders and employed the Autoregressive Distributed Lag (ARDL) model to examine AI‘s impact on macroeconomic indicators like high-technology exports. Our results showed significant relationships between these exports and key economic factors including R&D expenditure, ICT goods exports, and real GDP. The ARDL analysis confirmed that R&D expenditure significantly boosted high-technology exports both in the short-term and long-term, while ICT goods exports also contributed positively over time. However, real GDP negatively impacted these exports, underscoring the need for specific economic policies to support AI adoption.

The study also stressed the importance of strengthening policy frameworks and enhancing digital education to advance AI implementation nationwide. Despite relying on publicly available data and not extensively exploring AI’s direct economic impacts or its application in specific business environments, the research provided valuable insights into Georgia‘s AI ecosystem and suggested areas for further investigation.

REFERENCES

ABUSELIDZE, G. and MAMALADZE, L. 2021. The impact of artificial intelligence on employment before and during pandemic: A comparative analysis. *Journal of Physics: Conference Series*, 1840(1), 012040. IOP Publishing.

AI GEORGIA. 2023. *AI Georgia* [online]. https://www.aigeorgia.ge/ [Accessed: 2023-06-05].

AI READINNESS INDEX 2023. Government AI Readiness Index. *Oxford Insight* [online]. https://oxfordinsights.com/ai-readiness/ai-readiness-index/ [Accessed: 2024-02-07].

ALY, H. 2022. Digital transformation, development, and productivity in developing countries: is artificial intelligence a curse or a blessing? *Review of Economics and Political Science*, Vol. 7 No. 4, pp. 238-256.

DAHLMAN, C., MEALY, S., and WERMELINGER, M. 2016. *Harnessing the digital economy for developing countries*. OECD Development Centre Working Papers, No. 334, p. 8.

KEMP, S. 2023. Digital 2023: Georgia. *Datareportal* [online]. https://datareportal.com/reports/digital-2023-georgia [Accessed: 2024-02-07].

ERISTAVI, D. and DAVITURI, G. 2021. The use of Artificial intelligence systems in Georgia, Tbilisi. *IDFI* [online]. https://idfi.ge/en/artificial%20intelligence_international_tendencies_and_georgia [Accessed: 2024-05-07].

EUROFAST. 2020. Tax incentives in Georgia for international IT and maritime companies. *Eurofast* [online]. https://euro-fast.eu/2020/10/16/tax-incentives-in-georgia-for-international-it-and-maritime- companies/ [Accessed: 2023-06-12].

FREY, C. B. and OSBORNE, M. A. 2017. The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*. Jan 1; 114:254-80.

GEOSTAT - National Statistics office of Georgia. 2023. *Geostat* [Online]. Available at: https://www.geostat.ge/en [Accessed: 2023-06-12].

GIGUASHVILII, G. and MAKASARASHVILI, T. 2021. Possibilities of Using Artificial Intelligence in Post-Pandemic Georgia. *Globalization and Business*. 12, 83-87.

GITA. 2023. *Georgia's innovation & technology agency* [online]. https://gita.gov.ge/en [Accessed: 2023-05-12].

IDFI. 2023. Article: Access to Open Data in Georgia. *IDFI* [online]. https://idfi.ge/en/access-to-opena-data-in-georgia [Accessed: 2023-07-05].

ITU. 2021. Georgia digital development country profile. *ITU* [online]. https://www.itu.int/en/ITU

MCKINSEY GLOBAL INSTITUTE. 2018. Notes from the AI frontier modelling: The impact of AI on the world economy, p 3. Mckinsey Global Institute [online]. www.mckinsey.com/mgi [Accessed: 2023-05-10].

NAPETVARIDZE, V. 2022. Artificial Intelligence in Georgia and in the World. *Politics/*პოლიტიკა. Dec 31(6).

NATIONAL STATISTICS OFFICE OF GEORGIA. 2024. Real Quarterly GDP. *Datareportal* [online]. https://geostat.ge/media/61254/04_mshp-mudmiv-fasebshi.xlsx [Accessed: 2024-06-05].

NATSIOSPOULOS, K. and TZEREMES, N. G. 2022. ARDL bounds test for cointegration: Replicating the Pesaran et al. (2001) results for the UK earnings equation using R. *Journal of Applied Econometrics*, 37(5), 1079-1090. https://doi.org/10.1002/jae.2919

OECD. 2021. State of implementation of the OECD AI Principles: Insights from national AI policies,' OECD Digital Economy Papers 311, OECD Publishing.

OECD. 2022. *Fostering Business Development and Digitalisation in Georgia*. OECD Publishing, Paris.

PARULAVA, G. 2021. *Georgia - Fit for the Age of Artificial Intelligence?* PMC research center Policy paper, Tbilisi.

PFAFF, B. 2008. *Analysis of Integrated and Cointegrated Time Series with R*. Second Edition. Springer, New York. ISBN 0-387-27960-1

RYAN, J. A. and ULRICH, J. M. 2024. *xts: eXtensible Time Series*. R package version 0.13.2, https://CRAN.R-project.org/package=xts

SAX, C. and EDDELBUEITTEL, D. 2018. Seasonal Adjustment by X-13ARIMA-SEATS in R. *Journal of Statistical Software*, 87(11), 1-17. https://doi.org/10.18637/jss.v087.i11

SAX, C. and STEINER, P. 2013. Temporal Disaggregation of Time Series. *The R Journal*, 5(2), 80-87. https://doi.org/10.32614/RJ-2013-028

TRAPLETTI, A. and HORNIK, K. 2023. *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-55. https://CRAN.R-project.org/package=tseries

WICKHAM, H. and BRYAN, J. 2023. *readxl: Read Excel Files*. R package version 1.4.3, https://CRAN.R-project.org/package=readxl.

WICKHAM, H., FRANCOIS, R., HENRY, L., MULLER, K. and VAUGHAN, D. 2023. *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. https://CRAN.R-project.org/package=dplyr

WILSON, H. J., DAUGHERTY, P. and BIANZINO, N. 2017. The jobs that artificial intelligence will create. *MIT Sloan Management Review*. Jul 1; 58(4): 14.

WIPO. 2023. World intellectual property Organization. *WIPO* [online]. https://www.wipo.int/directory/en/details.jsp?country_code=GE, [Accessed: 2023-06-13].

WISSKIRCHEN, G., BIACABE, B. T., BORMAN, U., MUNTZ, A., NIEHAU, G., SOLER, G. J. and VON BRAUCHITSCH, B. 2017. Artificial intelligence and robotics and their impact on the workplace. *IBA Global Employment Institute*. Apr; 11(5): 49-67, p 14.

WORLD BANK. 2024. *Research and Development Expenditure (% of GDP)* [online]. https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS [Accessed: 2024-03-07].

WYNSBERGHE, V. A. 2021. Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics 1*, no. 3: 213-218.

XIE, Y. 2023. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.45. https://yihui.org/knitr/

YU, K. H., ZHANG, Y., LI, D., MONTENEGRO-MARIN, C. E., and KUMAR, P. M. 2021. Environmental planning based on reduce, reuse, recycle and recover using artificial intelligence. *Environmental Impact Assessment Review*, 86, p.106492.

ZEILEIS, A. and GROTHENDIECK, G. 2005. zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*, 14(6), 1-27. https://doi.org/10.18637/jss.v014.i06

ZEILEIS, A. and HOTHORN, T. 2002. Diagnostic Checking in Regression Relationships. *R News*. 2(3), 7-10. https://CRAN.R-project.org/doc/Rnews/

Contact information

Nadia Mtchedlidze: e-mail: mtchelidze1@uniba.sk
Zuzana Papulova: e-mail: zuzana.papulova@fm.uniba.sk

BEHAVIORAL EXPERIMENTS IN PUBLIC SECTOR

Mária Murray Svidroňová[1], Nikoleta Jakuš Muthová[1]

[1] Department of Public Economics and Regional Development, Faculty of Economics, Matej Bel University in Banská Bystrica, Tajovského 10, 975 90 Banská Bystrica, Slovakia

ABSTRACT

In this paper we present preliminary results of a research project aimed at mapping the preferences of young people in the areas of public sector such as housing, employment and commuting. To do so we will employ methods of behavioral economics and experiments. First of all, we did a bibliometric analysis of studies dealing with behavioral experiments in the selected areas to reveal a gap for future research. We focused on a time period of 2010–2023. For mapping we used VosViewer and data from the Web of Science database. Results indicate that even there are some experiments done in the areas of housing, employment or commuting, there is no methodology based on behavioral economics to reveal the preferences of young people in these areas.

Keywords: behavioral economics, experimental research, preferences of young people, housing preferences, employment preferences, transportation preferences

JEL Code: H41, C91, C93

1 INTRODUCTION

American-style suburbs are popping up in all Central and Eastern Europe countries and have become a popular option for many. It is assumed that a better definition of the spatial relations of cities and towns will be based on 1) an understanding of movement patterns in a defined area and 2) home ownership as a dominant factor defining social status – home ownership becomes more important than employment (Conn, 2014). The project will therefore map preferred commuting patterns based on the so-called sleep-space /work-space concepts – where people sleep (that is, live) versus where they work. The COVID-19 pandemic made it possible to operate in the so-called home office mode, working from home became a common thing and commuting to work changed to 3–4 days a week from the original 5. Commuting to work is not about distance (number of km), but about time (hours spent commuting), that is home office fundamentally changes preferences about housing, employment and mobility.

Giddens (1991) claims that everyone is constantly searching for his own life story, in which the living space plays an important role. He states that "a person's identity should not be found in their behavior, nor in the reactions of others – however important that may seem,

https://doi.org/10.11118/978-80-7509-990-7-0111

but in their ability to keep a particular 'story' going." He refers to this as an "ongoing story of the self." Individuals try to live their "life story" based on choices. These possibilities allow them to achieve material satisfaction, but also social status. This life story can also be referred to as the so-called "diary of self". Taking this point of view into account, it is precisely rural areas that can provide various details for the life fulfillment of individuals and influence their preferences also in terms of housing and employment. A Gallop (Newport, 2018) survey of 1.499 respondents representing randomly selected adults found a gap between their desires and reality, as only 15% of Americans actually live in small rural towns, while 27% of respondents said they would like to live there to live.

A survey by Eurofound (2019) in the European Union found that a higher proportion of rural residents than urban residents feel a strong sense of belonging in their community. The survey, on the other hand, pointed to the negatives of rural life, the financial difficulties of rural areas were more pronounced in southern and eastern Europe, especially compared to their urban counterparts. However, Noronha et al. (2013) found that the old concept of the rural environment as a place with insufficient amenities and lower education of people is gradually disappearing. In addition to population density and environmental amenities, the gap between urban and rural areas is rapidly closing in many countries, as the gap in lifestyle, education, availability of services and access to information and communication technologies is also shrinking significantly. Cole et al. (2020) during a 10-year study of Slovak students from the Matej Bel University in Banská Bystrica found that generation Z (Gen Z) indicated the place in which they want to live, to a greater extent, is outside a big city. The previous generation of millennials often described their need to travel abroad, at least for short period of time, but this is no longer applied to Gen Z to such an extent. It should be noted that the unemployment rate in Slovakia dropped from 15% at the beginning of 2010 (during the 2008 financial crisis) to 4.9% in December 2019 (before COVID-19) i.e., travel for Gen Z was no longer necessary to earn money to bring home. Many students describe their jobs in the city center, but with their own houses in a small town (suburbs), When it is necessary to own a car – commuting by car is not perceived negatively. A very small number of students (around 5%) described their preferences for living in the center of a big city with a description of luxury or loft apartment.

From the above, it is clear that it is important to map young people's preferences in areas such as housing, employment and transport mobility, including new points of view to the usual definitions of urban and rural environments or social status. Therefore, a research project has started in co-operation of four Slovak universities to create and test a methodology based on the principles of behavioral economics to determine the preferences of young people in the areas of housing and employment and based on these findings, to formulate recommendations for the creation of strategic documents of housing policy, employment policy and regional development. The first step of the research is to conduct bibliometric and systematic analysis to map the current status of BE use in given public sector areas, mostly on behavioral experiments and methodologies for revealing youth preferences. Bibliometric analysis has become a popular method for researching individual issues during the last few years (Slyvkanyč and Glova, 2023).

This paper's objective is to systematically examine the conceptual approach to use of behavioral economics in public sector, namely housing, employment and commuting (public transport) in the period 2010 to 2023.

2 BEHAVIORAL ECONOMICS IN PUBLIC SECTOR

No preference can be precisely quantified when assessing young people's ideas about their future lives. In such a situation, it is difficult to create a test that measures preference based on an action. To circumvent this problem, the knowledge of behavioral economics (BE) can be used. BE which focuses on the individuals with limited rationality, revealing what really

influences their decisions and actions. The very importance of BE lies in its contribution to understanding the causes of irrational behavior of individuals and finding ways to push individuals to make better decisions. Behavioral economics tries to bring a psychological dimension to economics.

Behavioral economics knowledge can be perceived in advertisements for various goods and services. However, the use of behavioral and experimental approaches is also increasingly popular in public administration, and their use can be found not only in the context of tax and fee collection (Hallsworth et al., 2017; Cranor et al., 2020; Gillitzer and Sinning, 2020; Sloboda et al. 2020) or saving energy and water (Chabé-Ferret et al., 2019), but also in the context of performance evaluation, effectiveness of audits (Avis et al., 2018; Belle, 2015; Engel et al., 2017), motivation of managers in the public sector (Belle & Cantarelli, 2014) or perceptions of the quality of public services (Chetty, 2015). Behavioral experiments aimed at young people found application, e.g. in employee recruitment and work habits (Kadric, 2015; MINDWORX, 2017; Pawar, 2016), financial behaviour (Thaler, Sunstein, 2010; Harris, 2017), lifestyle (Goldstein, 2011) and health care (Linkenbach and Perkins, 2003); Thornton, 2008). Based on the study of almost 1.000 studies and researches, we can conclude that regarding housing preferences, we found only publications aimed at revealing the degree of discrimination, e.g. whether a person is willing to live near marginalized communities (it was mostly about the context of the USA, that is, it mainly concerned the issue of race, e.g. Kuklinski et al., 1997), or the publications focused on migration policy (Hainmuller et al., 2014) but not on housing issues as such, not to mention the complexity of the connection to young people's preferences regarding employment and commuting. At the same time, it can be stated that this type of data regarding the preferences of young people is absent in Slovakia, data collection is currently underway at the employment, social affairs and family offices, but again this only concerns rental housing, not a complex issue.

Chetty (2015) claims that BE is also related to the use of the methodological apparatus of behavioral science (e.g., experiments), which allow better estimation of the effects of interventions (public policies) not only in terms of effectiveness, but also in terms of the effects on the well-being of individuals. More precisely estimating effects naturally has the potential to improve the ability of policymakers to identify optimal public policy. Experimental testing therefore appears to be the most appropriate method.

According to Aab (2005), consumer preferences can be determined in two ways. The first, indirect method, is based on the observation of the consumer's market behavior. Economic models are used to explain the relationship between consumer behavior (revealed preferences) and the value of non-market goods, and thus also to determine the value of the good itself. Indirect methods work with revealed preferences of individuals. The second, direct way of finding out the preferences of individuals is the method of direct questioning of a certain set of individuals. With this method, the individual states, or contingently (conditionally) values its preferences (stated preferences) on the hypothetical market, i.e., the willingness to pay for the good (determining the maximum price that one is willing to pay for the good – willingness to pay – WTP) or the willingness to accept compensation for the loss resulting from the production of a certain good (determines the minimum compensation due to the impossibility of consuming the given good – willingness to accept – WTA) . Unlike the indirect method, which reflects only explicit values, the direct method reveals the total value of the collective good, including its explicit and implicit value of benefits and costs (Stejskal et al., 2013).

Truc (2022) using bibliometric analysis has documented that while individual BE articles have become less intensely related to psychology, the growing number of BE articles in economics as a whole has intensified the overall interdisciplinarity between economics and psychology. There has been a rise in the importance of management studies, as well as a variety of other disciplines in the social and natural sciences, as behavioral economists have diversified their interdisciplinary relationships since the 2000s. In 2008, behavioral experiments started to use incentives for individuals' choices with foreseeable outcomes but without exclusion of alternative choices or reliance on financial stimuli. These so-called nudges have been applied in many fields to influence decision-making of individuals.

3 METHODOLOGY AND DATA

This paper aims to systematically examine the conceptual approach to use of behavioral economics in housing, employment and commuting (as fields of public sector) in the period 2010 to 2023. To fulfill the goal, we formulated following research questions (RQ):

- RQ1: What is the number of articles published in the given topic?
- RQ2: What are the research areas and categories in the given topic?
- RQ3: What is the keyword occurrence in the analyzed articles?

The main database used in this study is Clarivate's Web of Science (WoS). WoS is a popular database used by researchers for its extensive coverage of scientific literature in various fields. Using WoS, we filtered the submissions to ensure a relevant and representative sample. The use of WoS and filters ensured reliable and relevant sources for our analysis. The data used was collected on January 5, 2024. In Table 1, we present the data collection process according to individual criteria.

VOSviewer software was used to perform bibliographic analyses. VOSviewer is free software for constructing and visualizing bibliometric networks from journals or individual publications. Data can be constructed based on citation relationships, bibliographic links, co-citations or co-authorship. The software offers a text mining function that can be used to visualize co-occurrence networks of important information from the scientific literature. This software connects keywords using the power of association (by default). Association strength is used to normalize the strength of links between items. (DeGroote, 2023).

Criteria	Results found
Keywords in the Title of the article	"Behavioral economics" or "behavioral economics methods (methodology) and laboratory experiment" or "experimental research" or "preferences of young people" or "housing (living) preferences" or "employment preferences" or "transport (mobility) preferences" or "preferences of generation" or "generation nudge" or "nudge" or "generation behaviours" or "generation preferences" or "gender differences" or "generations differences" **Total number of documents: 116,387**
Criterium 1	WoS categories: "Economics", "Management", "Public Administration", "Regional Urban Planning", "Urban Studies", "Political Science", "Family Studies", "Psychology Social", "Psychology experimental", "Behavioral Sciences", "Social Issues", "Demography", "Transportation", "Transportation Science Technology", "Social Science Interdisciplinary", "Multidisciplinary Sciences", "Social Sciences Mathematical Methods", "Mathematics Interdisciplinary Applications" AND exclude all others **Total number of documents: 23,806**
Criterium 2	Years 2010-2023 **Total number of documents: 17,756**
Criterium 3	Language: "English" **Total number of documents: 17,453**
Criterium 4	Type of the document: "Article", "Open access" **Total number of documents: 7,500**
Criterium 5	WoS Citation topics: "Economics", "Management", "Social Psychology", "Economic Theory", "Political Science", "Transportation", Public Administration", "Regional Urban Planning", "Urban Studies", "Behavioral Sciences", "Transportation Science Technology", "Social Science Interdisciplinary", "Multidisciplinary Sciences", "Social Sciences Mathematical Methods", "Mathematics Interdisciplinary Applications" **Total number of documents: 2,389 = final sample**

Tab. 1 Sequence of data collection

4 RESULTS

In the monitored period of 2010–2023 we can observe an increasing number of publications dealing with the topic of behavioral experiments in public sector. The drop in 2023 can be explained by a simple fact that not all articles published in 2023, have been indexed in WoS yet.

The following Figure 2 captures the overview of the most active authors in the analyzed field of research.

Based on the research area, the analyzed publications cover the following topics (Figure 3).

To compare, we also present topics based on WoS categories (Figure 4). The most overlapping topics are (Business) Economics, Psychology (Social), Transportation, Public Administration, Behavioral Sciences and Operations Research Management Science. Based on this results,

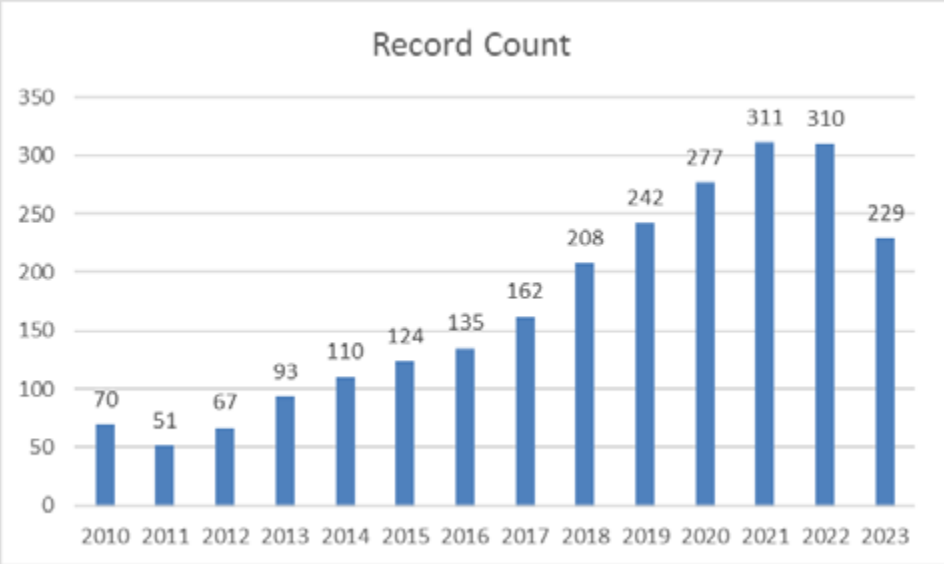


Fig. 1: Number of publications on behavioral experiments in public sector

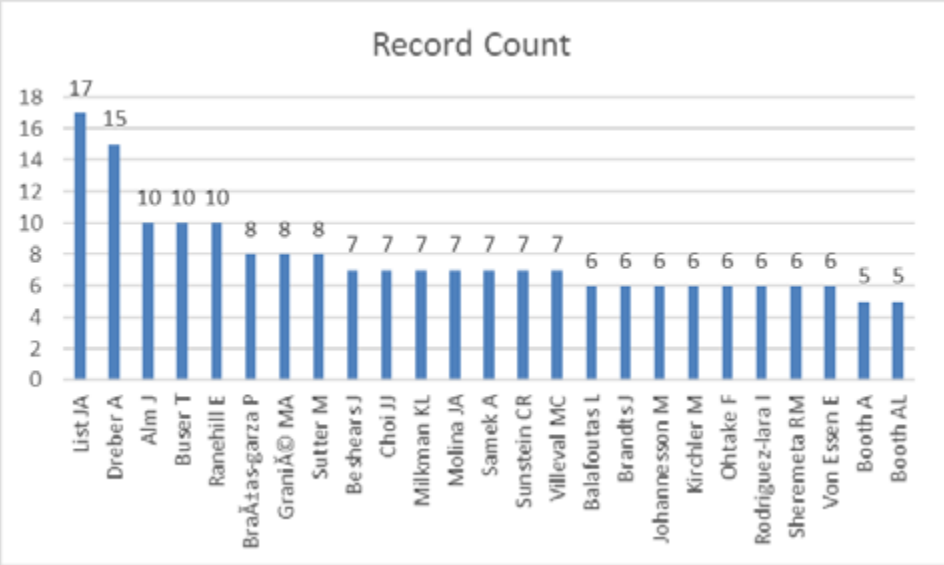


Fig. 2: Top 25 most active authors in the field of behavioral experiments in public sector

Fig. 3: Top 15 researched areas

we can already see that research criterium 1 should have contained also WoS category "Operations Research Management Science". According to the results of the co-occurrence analysis, the keywords were divided into eight different clusters, which represented 213 links and a total link strength of 372. This cluster is shown in Figure 5.

From the occurrence analysis it is clear that the analyzed studies using behavioral experiments include risk preferences, but not living preferences or methodology for preferences of (young) people. Some studies focus on life satisfaction and/or job satisfaction but from the Figure 5 it is not clear whether these studies directly use a behavioral experiment. The issue for further investigation shall focus on the use of behavioral experiments in identifying preferences of (young) people in the areas of transportation, employment and living preferences. To do so, a scoping review can be applied, i.e. based on the bibliometric analysis, the relevant articles will be downloaded, read and categorized using PRISMA method.

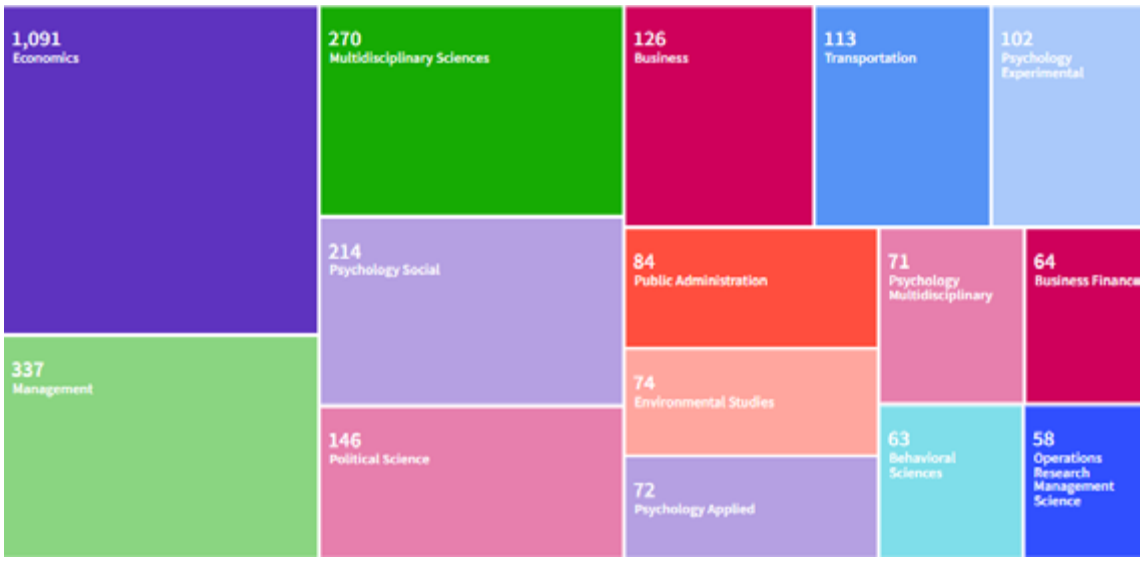


Fig. 4: Top 15 Web of Science Categories

Fig. 5: Keyword occurrence analysis

5 DISCUSSION AND CONCLUSIONS

Jia and Mustafa (2022) conducted a bibliometric analysis of the existing literature related to the investigation and application of nudge by analyzing 1.706 publications retrieved from Web of Science. The results indicated that (a) being a relatively newly developed theory, interest in nudge in academia has expanded both in volume and disciplines, mostly by Western scholars and behavioral economists; (b) future studies in nudge-related fields are expected to consolidate its current frontiers in individual behaviors while shedding light on new territories such as the digitalized environment. Rawat (2019) focused his research on behavioral studies using behavioral experiments in public policy. He stated that so far behavioral interventions linked to public policy have withheld any overhauling changes to the existing setup. Instead, they have concentrated on how information is presented to the actors (individuals), its salience, the convenience of different options and what actors know about others´ decisions. This is also reflected in academic writing on the subject, i.e. no methodology to reveal preferences of individuals in any area of public sector or public services has been documented based on the bibliometric analysis.

This paper presents a bibliometric overview of studies using behavioral experiments in the selected areas of public sector, namely in the areas of housing, employment or commuting (public transport). For mapping we used VosViewer and data from the Web of Science database. Results indicate that even there are some experiments done in the areas of housing, employment or commuting, there is no methodology based on behavioral economics, that is using behavioral experiments, to reveal the preferences of young people in these areas.

This study also has its limitations. The Web of Science database is a comprehensive database, but it does not cover all existing works on behavioral experiments in given areas of public sector. Our results apply for Web of Science documents but may differ for other databases. Also, one must not generalize the results because the bibliometric analysis analyzes titles, keywords and abstracts of published studies, this method does not analyze entire documents.

Despite the limitations, bibliometric analysis offers a comprehensive overview of the field of interests (in our case it is the area of behavioral experiments in selected areas of public sector). Study points out current trends in research that can serve as inspiration for other researchers. Based on bibliographic analysis, we have identified gaps that we will try to cover in future research. In particular, we shall use also other databases, e.g. SCOPUS and Science Direct, which could be completed with databases of publishing houses like Wiley, Emerald and Sage. Next, we shall include also WoS category/research area "Operations Research Management Science". Most of all, the research shall continue with so-called scoping review, where not only abstracts but full-texts are analyzed.

Acknowledgements

This work was supported by the Slovak Research and Development Agency [Project nr. APVV-22-0090].

REFERENCES

ALM, J. and JACOBSON, S. 2007. Using laboratory experiments in public economics. *National Tax Journal*, 60(1): 129- 152.

AVIS, E., FERRAZ, C. and FINAN, F. 2018. Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians. *Journal of Political Economy*, 126(5), 1912–1964.

BELLE, N. 2015. Performance-Related Pay and the Crowding Out of Motivation in the Public Sector: A Randomized Field Experiment. *Public Administration Review*, 75 (2), 230-241.

BELLE, N. and CANTARELLI, P. 2014. Monetary Incentives, Motivation, and Job Effort in the Public Sector. *Review of Public Personnel Administration*, 35(2), 99-123.

COLE, D. A., MURRAY SVIDROŇOVÁ, M., GUBALOVÁ, J. and KOŽIAK, R. 2020. *Opustené budovy v kontexte rozvoja samospráv na Slovensku*. Belianum.

CONN, S. 2014. *Americans Against the City: Anti-urbanism in the Twentieth Century*. Oxford University Press.

CRANOR, T., GOLDIN, J., HOMONOFF, T. and MOORE, L. 2020. Communicating Tax Penalties to Delinquent Taxpayers: Evidence from a Field Experiment. *National Tax Journal*, 73(2), 331–360.

DEGROOTE, S. 2023. *Subject and Course Guides: Bibliometric Analysis and Visualization: Bibliometrics with VOSViewer*. [Online]. https://researchguides.uic.edu/c.php?g=1233392&p=9025986 [Accessed: 20240201].

ENGEL, E., JORDAN, F., RAU, T. and REPETTO, A. 2017. Supreme Audit Institutions and Deterrence: Experimental Evidence from Chile. *Economic Behavior & Organization,* (2017): 1-23.

EUROFOUND, 2019. I*s rural Europe being left behind?* European Quality of Life Survey 2016, Publications Office of the European Union, Luxembourg.

GIDDENS, A. 1991. *Modernity and Self-Identity*. Stanford, California: Stanford University Press.

GILLITZER, C. and SINNING, M. 2020. Nudging businesses to pay their taxes: Does timing matter? *Journal of Economic Behavior & Organization*, 169, 284–300.

HALLSWORTH, M., LIST, J. A., METCALFE, R. D. and VLAEV, I. 2017. The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148, 14–31.

HARRIS, E. 2017. Generation Nudge: What A Massive Experiment Is Revealing About Millennial Money Habits. *Forbes* [Online]. https://www.forbes.com/sites/elizabethharris/2017/04/21/generation-nudge-what-a-massive-experiment-is-revealing-about-millennial-money-habits/?sh=3a1835432652 [Accessed: 20231210].

HAINMULLER, J., HOPKINS, D. and YAMAMOTO, T. 2014. Casual inference in xonjoint analysis: understanding multidimensional choice via stated preference experiments. *Political analysis*, 22, 1-30.

CHABÉ-FERRET, S., LE COENT, P., REYNAUD, A., SUBERVIE, J. and LEPERCQ, D. 2019. Can we nudge farmers into saving water? Evidence from a randomised experiment. *European Review of Agricultural Economics*, 46(3), 393–416.

CHETTY, R. 2015. Behavioral Economics and Public Policy: A Pragmatic Perspective. *American Economic Review*, 105(5), 1–33. https://doi.org/10.1257/aer.p20151108

JIA, C. and MUSTAFA, H., 2022. A Bibliometric analysis and review of nudge research using VOSviewer. *Behavioral Sciences*, 13(1), p.19.

KADRIC, A. 2015. *Nudging digital buttons of the Millennials with your digital projects* [online]. https://www.slideshare.net/annakadric/nudging-digital-buttons-of-the-millennials-with-your-digital-projects [Accessed: 20231210].

KUKLINSKI, J., COBB, M. and GILENS, M. 1997. Racial attitudes and the "new South". *Journal of Politics*, 59(2), 323-349.

NEWPORT, F. 2018. *Americans Big on Idea of Living in the Country, Gallup* [online]. https://news.gallup.com/poll/245249/americans-big-idea-living-country.aspx [Accessed: 20231210].

NORONHA, V. E., LEEUWEN, E. and NIJKAMP, P. 2013. *Towns in a Rural World*. Ashgate, Aldershot.

PAWAR, A. 2016. *How to nudge employees towards company goals* [online]. https://www.peoplematters.in/article/benefits-and-rewards/how-to-nudge-employees-towards-company-goals-14066 [Accessed: 20231210].

RAWAT, S. 2019. A bibliometric analysis of behavioural studies in economics and public policy journals. *Handbook of behavioural change and public policy*, pp.49-62.

SLOBODA, M., PAVLOVSKÝ, P. and SIČÁKOVÁ-BEBLAVÁ, E. 2020. The effectiveness of behavioral interventions on increasing revenue from local fee. *Review of Behavioral Finance*, 14(1), 1–15.

SLYVKANYČ, N. and GLOVA, J. 2023. Bibliometric study of the conceptual approach of intellectual capital. *Journal of Economics and Social Research*, 24(2) 137-156.

STEJSKAL, J. et al. 2013. *Meření hodnoty veřejných služeb (na příkladu veřejných knihoven)*. Praha: Wolter Kluwer. 208 pp.

TRUC, A., 2022. Interdisciplinary influences in behavioral economics: a bibliometric analysis of cross-disciplinary citations. *Journal of Economic Methodology*, 29(3), pp.217-251.

Contact information

Mária Murray Svidroňová: e-mail: maria.murraysvidronova@umb.sk
Nikoleta Jakuš Muthová: e-mail: nikoleta.jakusmuthova@umb.sk

MICROCONTROLLERS SUITABLE FOR ARTIFICIAL INTELLIGENCE

Petr Pernes[1], Miroslav Jaroš[1], Jiří Podivín[1], Oldřich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics Mendel University in Brno, Zemědělská 1, 613 00 Brno-sever, Czech republic

ABSTRACT

Artificial intelligence (AI) has become increasingly prevalent in various applications, from self-driving cars to facial recognition. However, the implementation of AI on resource-constrained devices such as microcontrollers has been a challenge due to the limited computational power and memory. In recent years, advances in AI technology and the development of specialized hardware have enabled the realization of AI on microcontrollers. This opens new opportunities for AI applications in domains such as embedded systems, the Internet of Things (IoT), and wearable devices. This article provides an overview of microcontrollers suitable for AI, discusses their benefits and challenges, presents a methodology for selecting suitable microcontrollers for AI applications, and highlights the criteria essential for effective implementation. Additionally, initial results from applying this methodology, including a comparative analysis of various microcontrollers, are discussed. Key findings emphasize the potential of specific microcontrollers like ARM Cortex-M7, Arm Ethos-U55, STMicroelectronics STM32F429, and Espressif ESP32-S3/C3 in AI applications. Future directions for the evolution of AI-enabled microcontrollers are also explored.

Keywords: Embedded systems, Internet of Things (IoT), Real-time processing, Power consumption, Hardware acceleration, Machine learning, Neural networks

JEL Code: C45

1 INTRODUCTION

The growing demand for AI-powered devices has driven the development of microcontrollers with enhanced processing capabilities and AI accelerators. These devices are capable of executing AI algorithms, such as neural networks, on-board, enabling real-time processing and decision-making. This is particularly beneficial for applications where latency is critical, such as robotics and industrial automation (DE VITA, 2022).

IoT devices are used for surveillance, environment monitoring, healthcare, and many more applications. They are used to respond to an external variable but also in more complex settings where there is a need for fast and autonomous reactions. The combination of IoT devices and artificial intelligence brought about a new type of intelligent device capable of

https://doi.org/10.11118/978-80-7509-990-7-0120

learning and interacting with its surroundings using past information. This combination is possible due to the improved capabilities of microcontrollers and the effort to reduce deep learning algorithms' memory and energy footprint (Muhoza, 2023)

To use AI algorithms effectively in the real world, it is necessary to have sufficiently powerful computing resources. Microcontrollers are one way to provide this performance. Microcontrollers are small, energy-efficient and affordable computers that are designed for use in embedded systems (Zhang, 2023).

The aim of the article is the overview and search of possible usable microcontroller (MCU) suitable for (AI), MCU has the potential to revolutionize various fields. Discuss the potential benefits, address any challenges, and explore future directions for this evolving technology.

2 BENEFITS OF AI IN MICROCONTROLLERS

In recent years, microcontrollers have also started to be used for AI applications. This is because AI algorithms are becoming more efficient and can be implemented on smaller and less powerful computing platforms.

One of the most prominent applications of artificial neural networks in the consumer space is automatic speech recognition in digital assistants (Bushur, 2023).

The integration of AI into microcontrollers brings several advantages, including:

- Reduced latency and improved responsiveness: AI algorithms can be executed locally on the microcontroller, eliminating the need for data transmission to a central server. This reduces latency and ensures real-time decision-making (Novac, 2021).
- Enhanced data privacy and security: by processing data locally, the need for data transmission is minimized, which protects sensitive data from unauthorized access.
- Increased autonomy and adaptability: AI-powered microcontrollers can adapt to their environment and make decisions autonomously, without the need for constant communication with a central server.
- Reduced power consumption: AI algorithms can be optimized for microcontrollers, leading to lower power consumption.
- Possible running on battery power: run without mains power supply, extended battery life.
- Low cost - microcontrollers are relatively inexpensive, energy efficient, making them affordable for a wide range of applications.
- Small size and weight: microcontrollers are very small and lightweight, making them ideal for use in mobile applications.
- AI close to data: processing data on site, applications are less dependent on internet connectivity.
- Possible mash functionality: Microcontrollers with mash functions can include various connectivity options such as Wi-Fi, Bluetooth and Ethernet. Using for computing, connectivity etc (Qi, 2020).

Disadvantages of microcontrollers for AI applications include:

- Limited computing power: Microcontrollers typically have lower computing power than, for example, personal computers or servers. This can be a problem for applications that require heavy computation.
- Limited memory: microcontrollers typically have less memory than, for example, personal computers or servers. This can be a problem for applications that need to store large amounts of data.
- Energy efficiency focus: microprocessors are designed for a specific purpose, exceeding these limits is problematic.
- Software development complexity: programming microprocessors requires a deep understanding of the architecture and principles of operation. Optimising software for low power consumption can be time consuming and complex.

- Limited AI support or non-existent: existence of available library like TensorFlow Lite for Microcontrollers.
- Limited peripheral support: some microprocessors lack peripherals that are necessary for specific applications. This may require the use of external components that increase power consumption and cost.
- Long inference time: for some AI tasks, the inference time, i.e. the time required to perform a computation, may be too long for microprocessors. This can lead to unacceptable latency and performance loss (Novac, 2021).

3 METHODOLOGY

The methodology for selecting a suitable microcontroller to work with AI should include the following steps:

- Mapping of available MCUs: Use online tools and databases, performance, memory, peripherals, build a comparison table of MCUs.
- Define Application Requirements: The second step is to clearly define the requirements of your AI application. Aspects like: The specific types of AI algorithms the application will use.
- Define evaluation metrics: Performance, benchmarks (CoreMark, DMIPS, EEMBC CoreMark) with respect to specific AI application needs. Factors: CPU architecture, performance, memory, peripherals, AI algorithm support, energy efficiency.
- Evaluate Available Microcontrollers: Once you understand your application requirements, you can start researching and evaluating available microcontrollers.
- Calculating the score: for each MCU, calculate the score according to the criteria combination of metrics.
- Select the Appropriate Microcontroller: select the one that best aligns with your application's requirements. This will ensure your AI application has the necessary processing power, memory, and features to function effectively. Consider other factors not included in the metrics.
- Presentation of results: In the research paper, provide a table of MCU comparisons and justify the selection of the most appropriate candidate.

4 CRITERIA FOR SELECTING A MICROCONTROLLER FOR USE IN AI

When selecting a microcontroller for use in AI, it is important to consider the following criteria:

- CPU architecture: microcontrollers with 32-bit CPU architecture generally have higher performance and efficiency than microcontrollers with 8-bit, 16-bit, 32-bit architecture.
- Performance: the performance of a microcontroller is an important factor in deter-mining what AI algorithms can be run on it. For example, SPEC2017 can be used to measure performance.
- Memory: microcontrollers with sufficient memory can store larger and more complex AI models.
- Peripherals: I/O pins, microcontrollers with appropriate peripherals (connectivity), such as Wi-Fi or Bluetooth, can facilitate communication with other devices.
- AI algorithm support - some microcontrollers offer support for specific AI algorithms, which can make them easier to implement.
- Energy efficiency: Microcontrollers with high energy efficiency are important for battery-powered devices.
- Price: low-cost microcontrollers are preferred.

Other recommendations:

- For a comprehensive assessment of energy performance, consider analysing energy consumption in different operating modes (active, sleep, etc.).
- For specific AI tasks (e.g. image processing, speech recognition), consider using dedicated MCUs with integrated accelerators.
- For testing and prototyping AI applications on selected MCUs, consider using development kits and platforms.
- Socket configuration (available pins): BGA, LQFP, etc.

In addition to general microcontroller parameters, it is also important to consider the specifics of the particular AI algorithms that will be executed on the microcontroller. For example, machine learning algorithms for image classification may require more computational resources than speech recognition algorithms. alternatively, algorithms published in (Stastny, Skorpil, 2007; Stastny et al., 2021).

Special hardware accelerators can improve the performance and efficiency of AI algorithms on microcontrollers. These accelerators often target specific types of AI algorithms, such as neural networks.

5 PARTIAL RESULTS OF METHODOLOGICAL PROCEDURE

The initial stages of the methodological procedure for selecting a suitable microcontroller for AI applications involve mapping the available microcontrollers and defining their key parameters. Here are the partial results based on the attached paper:

a) Mapping Available Microcontrollers:
The following microcontrollers were considered: ARM Cortex-M7, Arm Ethos-U55, STMicroelectronics STM32F429, Espressif ESP32-S3, and Espressif ESP32-C3 (Loukatos, 2023; Yoon, 2020)

b) Defining Key Parameters:
The key parameters identified for comparison include clock speed, flash memory, SRAM, peripherals, applications, and cost.

Parameter	ARM Cortex-M7	Arm Ethos-U55	STMicroelectronics STM32F429	Espressif ESP32-S3	Espressif ESP32-C3
Processor	32-bit RISC	32-bit RISC	32-bit RISC	32-bit RISC	32-bit RISC
Core	Cortex-M7	Ethos-U55	Cortex-M4	Xtensa LX7	Xtensa LX7
Clock speed	Up to 400 MHz	Up to 500 MHz	Up to 180 MHz	Up to 240 MHz	Up to 160 MHz
Flash memory	Up to 2 MB	Up to 16 MB	Up to 2 MB	Up to 4 MB	Up to 4 MB
SRAM	Up to 512 KB	Up to 4 MB	Up to 256 KB	Up to 512 KB	Up to 384 KB
Peripherals	USB, Ethernet, CAN, SPI, I^2C, UART, etc.	USB, Ethernet, CAN, SPI, I^2C, UART, etc.	USB, Ethernet, CAN, SPI, I^2C, UART, etc.	USB, Ethernet, CAN, SPI, I^2C, UART, etc.	USB, Ethernet, CAN, SPI, I^2C, UART, etc.
Applications	Industrial control, IoT, wearables, etc.	Machine learning, image processing, etc.	Industrial control, IoT, automotive, etc.	IoT, wearables, home appliances, etc.	IoT, wearables, home appliances, etc.
Cost	$10-$20	$5-$10	$5-$10	$5-$10	$2-$5

Tab. 1 MCU characteristics

Their operating frequencies are in the order of hundreds of MHz, working memory is in the order of 0.5 MB and the price is under 20 dollars. The final choice of microcontroller would depend on the specific needs of the application and the available budget.

c) Define evaluation metrics:
For each microcontroller, calculate a score based on the combination of the evaluation metrics. This step is generally detailed and involves assigning weights to each metric and then computing a composite score for comparison.
Example Scoring Approach:
- CPU Performance (30%)
- Memory (25%)
- Peripherals/Applications (15%)
- Energy Efficiency (20%)
- Cost (10%)

d) Initial analysis:
- Performance: The Arm Ethos-U55 has the highest clock speed (500 MHz) and the most significant amount of SRAM (4000 KB), making it potentially the best option for more demanding AI applications such as machine learning and image processing. The ARM Cortex-M7, with a clock speed of 400 MHz and 512 KB of SRAM, is also a strong candidate for industrial control and IoT applications.
- Memory: Flash memory is critical for storing AI models. The Arm Ethos-U55 stands out with 16 MB, significantly more than the other microcontrollers.
- Cost: The Espressif ESP32-C3 is the most cost-effective option at $3, suitable for budget-conscious projects or applications with less demanding AI requirements.
- Peripherals/Applications: Different microcontrollers are tailored to various applications, with the ARM Cortex-M7 and STMicroelectronics STM32F429 being suitable for industrial control and IoT, while the Espressif models are ideal for IoT and wearable applications.

Generally, embedded systems are limited in two aspects: processing power and memory. For example, the ESP32-WROOM-32E, a popular system on a-chip (SoC) for embedded systems, possesses at most two cores operating at 240 MHz with 520 kB of SRAM and 16 MB of SPI flash (Bushur, 2023).

6 RESULTS

Microcontrollers are increasingly suitable for use in AI. Their small size, low power consumption, and relatively low cost make them ideal for use in mobile applications and other applications where a low power profile is important.

Currently, AI developers are working on developing new algorithms that are optimized for microcontrollers. These algorithms will be able to make better use of the limited resources of microcontrollers.

Initial results show that microcontrollers such as ARM Cortex-M7, Arm Ethos-U55, STMicroelectronics STM32F429, Espressif ESP32-S3 or ESP32-C3 are increasingly suitable for AI applications.

7 DISCUSSION AND CONCLUSIONS

Microcontrollers are expected to play an increasingly significant role in AI in the future. This is because AI algorithms are becoming increasingly efficient and can be implemented on smaller and less powerful computing platforms. The development of microcontrollers with AI is a rapidly developing field with significant potential. The ability to execute AI algorithms on

board microcontrollers opens up new possibilities for AI applications in a variety of fields. As technology evolves, we can expect to see even more sophisticated and innovative AI applications powered by microcontrollers. (Muhoza, 2023) This opens up new opportunities for AI applications in domains such as embedded systems, Internet of Things (IoT), and wearable devices. (Campero-Jurado, 2020)

While the potential of AI-powered microcontrollers is vast, there are some challenges to consider. The processing power of microcontrollers is generally lower compared to traditional computers. This may limit the complexity of AI models that can be implemented. Additionally, training and deploying AI models on microcontrollers can be more complex compared to traditional platforms (Zhang, 2020). We have added additional MCUs to be considered for use with AI.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education and by the project Internal grant agencies at Mendel University in Brno, Faculty of Business and Economics, Brno, Czech republic, with number IGA24-PEF-DP-020.

REFERENCES

BUSHUR, Jacob and CHEN, Chao. 2023. Neural Network Exploration for Keyword Spotting on Edge Devices. *Future Internet*, 15, n. 6. ISSN 1999-5903. Available at: https://doi.org/10.3390/fi15060219 [Accessed. 2023-12-02].

BRUNO, C., LICCIARDELLO, A., NASTASI, G. A. M., PASSANITI, F., BRIGANTE, C., SUDANO, F., FAULISI, A. and ALESSI, E. 2021. Embedded Artificial Intelligence Approach for Gas Recognition in Smart Agriculture Applications Using Low Cost MOX Gas Sensors. In: *2021 Smart Systems Integration (SSI)*. 2021 Smart Systems Integration (SSI). IEEE. https://doi.org/10.1109/ssi52265.2021.9467029

CAMPERO-JURADO, I., MÁRQUEZ-SÁNCHEZ, S., QUINTANAR-GÓMEZ, J., RODRÍGUEZ, S. and CORCHADO, J. 2020. Smart Helmet 5.0 for Industrial Internet of Things Using Artificial Intelligence. *In Sensors*, 20(21), p. 6241. MDPI AG. https://doi.org/10.3390/s20216241

DE VITA, Fabrizio, NOCERA, Giorgio, BRUNEO, Dario, TOMASELLI, Valeria and FALCHETTO, Mirko. 2022. On-Device Training of Deep Learning Models on Edge Microcontrollers. Online. In: *2022 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE*

Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics). IEEE, 2022, p. 62– 69. ISBN 978-1-6654-5417-9

LOUKATOS, D., KONDOYANNI, M., ALEXOPOULOS, G., MARAVEAS, C. and ARVANITIS, K. G. 2023. *On-Device Intelligence for Malfunction Detection of Water Pump Equipment in Agricultural Premises: Feasibility and Experimentation*. Sensors, MDPI AG.

MA, Q. and WANG, Y. 2021. RETRACTED ARTICLE: Application of embedded system and artificial intelligence platform in Taekwondo image feature recognition. *Journal of Ambient Intelligence and Humanized Computing*, 13(S1), p. 23–23. Springer Science and Business Media LLC. https://doi.org/10.1007/s12652-021-03222-9

MUHOZA, Aimé Cedric, BERGERET, Emmanuel, BRDYS, Corinne and GARY, Francis. 2023. Power consumption reduction for IoT devices thanks to Edge-AI: Application to human activity recognition. *Online. Internet of Things*, 24. ISSN 25426605. Available at: https://doi.org/10.1016/j.iot.2023.100930 [Accessed. 2023-12-02].

NOVAC, P.-E., BOUKLI HACENE, G., PEGATOQUET, A., MIRAMOND, B. and GRIPON, V. 2021. Quantization and Deployment of Deep Neural Networks on Microcontrollers. *Sensors*, 21(9), p. 2984. MDPI AG. https://doi.org/10.3390/s21092984

OSUWA, A. A., EKHORAGBON, E. B. and FAT, L. T. 2017. Application of artificial intelligence in Internet of Things. In: *9th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE. https://doi.org/10.1109/cicn.2017.8319379.

QI, H. 2020. Fuzzy logic hybridized artificial intelligence for computing and networking on internet of things platform. *Peer-to-Peer Networking and Applications*, Vol. 13, Issue 6, pp. 2078–2088. Springer Science and Business Media LLC. https://doi.org/10.1007/s12083-019-00827-y

SAKR, Fouad, BELLOTTI, Francesco, BERTA, Riccardo and DE GLORIA, Alessandro. 2020. Machine Learning on Mainstream Microcontrollers. *Sensors*, 20(9). ISSN 1424-8220

STASTNY, J., SKORPIL, V. 2007. Analysis of Algorithms for Radial Basis Function Neural Network. *Personal Wireless Communications*, Springer New York, 245, p. 54–62, ISSN 1571-5736, ISBN 978-0-387-74158-1, WOS:000250717300005.

STASTNY, J., SKORPIL. V., BALOGH, Z. and KLEIN, R. 2021. Job shop scheduling problem optimization by means of graph-based algorithm. *Applied Sciences*, 11(4), ISSN 2076-3417. URL: https://doi.org/10.3390/app11041921

YOON, Y. H., HWANG, D. H., YANG, J. H. and LEE, S. E. 2020. Intellino: Processor for Embedded Artificial Intelligence. *Electronics*, 9(7), p. 1169. MDPI AG. https://doi.org/10.3390/electronics9071169

ZHANG, Z. and LI, J. 2023. A Review of Artificial Intelligence in Embedded Systems. *Micromachines*, 14(5), p. 897. MDPI AG. https://doi.org/10.3390/mi14050897

ZHANG, D., LIU, Y., CHENG, L. and YANG, H. 2020. A survey of machine learning for resource-constrained devices. *Journal of Network and Computer Applications*, 167, 102738. https://doi.org/10.1016/j.neucom.2023.02.006

Contact information

Petr Pernes: e-mail: petr.pernes@mendelu.cz
Miroslav Jaroš: e-mail: qqjaros7@mendelu.cz
Jiří Podivín: e-mail: qqpodivi@mendelu.cz
Oldřich Trenz: e-mail: oldrich.trenz@mendelu.cz

ANALYTICAL PLATFORM

Jan Přichystal[1], Roman Valovič[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

The Analytical Platform is an online platform that simplifies work of analytical department, front-office, back-office and compliance for investment firms. It is using an artificial intelligence, data analytics and machine learning in finance and investments. Analytical Platform helps investment funds, securities traders and professional investors gain higher alpha, and lower beta and get flawless records of investment decisions. It offers various specialized tools and services in a flexible and personalized way to fit investor's needs. This article presents the components of the platform, the technologies used for the development and approaches for sentiment analysis and summarization of financial articles as a source of key information for decision support.

Keywords: data mining, financial markets, natural language processing, machine learning, sentiment analysis, text summarization, web application

JEL Code: L86

1 INTRODUCTION

Analytical Platform is a multifunctional platform combining sets of functions for performing data analysis and text analysis. The goal of the platform is to remove information asymmetry and bring added value to users in the form of reducing transaction costs, streamlining work processes, and uncovering new business opportunities using unique software combining data and text analytics capabilities.

The logic of the platform is built on a multi-layered architecture. The first layer is the development environment (DPS Lab), which allows you to combine tools in the field of data and text analytics and build your own scenarios to obtain the desired outputs. Verified combinations with market potential for the selected domain are then displayed within the end applications (DPS Applications). A communication interface (DPS Result sharing) is also available.

The project aimed to create a platform that can be used through microservices as part of larger units and information flows of the enterprise architecture of communication and computing systems. The goal was achieved through the creation of a modular system that:

1. provides functionality for data import and basic data/text analytics (DPS Lab),
2. allows you to connect and combine individual tools to achieve the desired result (DPS Lab),
3. provides end applications for use in the selected field (DPS Applications),

https://doi.org/10.11118/978-80-7509-990-7-0127

4. enables sharing and discussion of results within working analytical teams (DPS Result sharing) and at the same time enables differentiated access to applications (API, web application, analyses).

Users can work with the platform and independently define the functions they will use to process the input data. Alternatively, they will use an industry package prepared for their work role within a specific market segment and thus use the industry experience implemented directly in the platform. The aim of this article is to introduce the Analytical platform as a new decision support tool in the field of securities investment, which uses artificial intelligence and machine learning tools to identify key parameters when building individual investors' portfolios.

2 SYSTEM COMPONENTS

The platform consists of several relatively separate components that the user can use independently or share the results from these individual components and use them for complex analyses.

2.1 DPS Lab

The data platform (DPS Lab) is composed of several basic technological objects that enable the creation of advanced data processing scenarios. The input object is a set of functions for importing data of various formats. Data input accepts machine-readable files such as XML, CSV, JSON and other data formats (for example, links to YouTube videos from which information is extracted). The range of data used varies from open data of the public sector (US Statements/EDI), through data downloaded by web crawlers (for example from the information websites Yahoo Finance or Seeking Alpha), through paid APIs (for example RapidAPI), to private data that users they upload to the platform themselves.

The platform enables the input of diverse data from different fields, it is possible to apply functions and methods specifically developed for a specific field cross-field. For example, the sentiment analysis of financial market articles. This makes it possible to achieve innovative, experimental data results that can be repeatedly compared with each other. This methodical procedure of systematic work with data finds application in data-intensive fields in which understanding data is a key competence for gaining or maintaining a competitive advantage.

The application allows, for example, to analyze videos devoted to product reviews. In the domain case of investment research, the analyst will use them for decision support, when he expects that the high-quality products of the company being evaluated will likely contribute to the growth of the stock price. In the same way, the information obtained helps buyers in the supply chain for stockpiling and other strategic decisions. In the field of construction and project preparation, testing is taking place, for example, with crowdfunding platforms for scoring developers, i.e. companies coming in with a need to finance their projects. Separate article (Přichystal et al., 2023) deals with this module in more detail.

2.2 Portfolio Manager

This component enables the management of stock portfolios according to the expected return, volatility or Sharpe ratio, according to the procedures of Harry Markowitz, who won the Nobel Prize for this theory.

The application combines a theory that investment company analysts have not yet been able to apply to robust portfolios, at the same time they did not have sufficiently high-quality pre-processed input data for it, which are built on DPS Lab (obtaining price data from the

Rapid API or scraping from Yahoo Finance, preprocessing data and subsequent determination restrictions for choosing the length of the time series) and subsequently enriched in the application with specific functions that interest end users. This is, for example, the possibility of exchanging a specific number of shares, or the choice of a specific stock pool, so that both effective results are achieved according to theory and at the same time the principles of the specific analyst who uses the tool are observed.

2.3 Stock Picking Lab

The application provides the management of stock portfolios according to multi-factor analysis in combination with the use of machine learning for the selection and operation of models. Multifactor analysis means that individual indicators that are important for analysts are calculated based on the input data. SP Lab builds on the activities and knowledge acquired before the start of the project implementation, when the consortium already had a functional model in the Matlab program generating predictions on a monthly basis and a linear model for evaluating shares, including the basic decomposition of scores, these outputs were displayed in a simple web application.

As part of the follow-up activities of the project, knowledge was significantly expanded and the DPS Lab and StockPicking Lab applications now provide users with the possibility of training their own model for valuing stocks, a view of the decomposition of individual indicators (displaying the results of individual indicators + weights that the model selected and trained for them in a given prediction assigned), which provides analysts with new opportunities to understand data and make better investment decisions.

2.4 Summary of Financial Articles

Summary of Financial Articles is a system designed to speed up the process of obtaining and processing information for these experts and thus facilitate their work. This system obtains articles, news or tweets from selected sources, especially in the field of financial markets, and analyzes their content. The aim is to offer a summarized view of individual documents in the form of several sentences summarizing the main ideas, to determine their sentiment and to provide aggregated data for specified time periods. A more detailed description of this module is available in Jakúbek et al. (2023).

2.5 DPS Result sharing

The output of the project is implemented in the form of a web application, which allows the use of the platform without the need for installation and complicated configuration via a web browser. The advantage of this solution is independence from the used technical equipment (personal computer, tablet, mobile phone) and operating system (MS Windows, MacOS, GNU Linux, ...).

The advantage of the web application is also the easy possibility of teamwork of several users on joint projects, sharing of solutions in progress and team communication provided by the DPS Result sharing application, within which users can share results and work in progress.

3 SYSTEM DESIGN

3.1 Frontend implementation

The frontend and backend of the system are two standalone applications that communicate via thin REST API on the web server. Heavy frontend is delivered to client browser just once and the content of the website is then partially updated, in places that require change, directly in the browser. This concept is called Single Page Application (SPA) and can be implemented in various JavaScript libraries or frameworks. In our case, to deal with view layer in MVC software architecture, we chose ReactJS library.

3.1.1 Used technologies

ReactJS is still a popular open-source library that is used for building web (and mobile) composable user interfaces. In ReactJS, the user interface is broken down in a large number of reusable components. These components manage their states to handle both the logic and the view itself. Their hierarchical structure clearly defines unidirectional flow of data in the application and their APIs and lifecycle hooks describe the behavior.

The state of the component is changed after data mutation or by executing a user action and this update triggers a smart and controlled rendering process. In the background, ReactJS engine tries to improve performance by minimizing expensive manipulation with real DOM as much as possible, so it first renders the changes to the virtual DOM represented in JavaScript. Then its job is to reconcile the recent changes against the old one and only the necessary ones applied to the real DOM.

To get the best of ReactJS world, we used JSX (syntax extensions to JavaScript), TypeScript, functional components and React hooks. To manage application-level state we injected Redux into application. Redux is a predictable state container which plays a helping hand while developing large scale JavaScript application. To keep a similar look and feel to our previous apps, we selected Ant Design as the UI framework. This framework is known mostly in the East and does not follow ubiquitous and shabby material design.

3.2 System backend

The technological complexity and computing power requirements vary dramatically between applications; for example, the SFA Rest API, which responds to a client request with a preprocessed result stored in a database, has only minimal performance requirements on the machine on which the application runs. SFA Sentiment Analysis, on the other hand, provides outputs based on the computations of a complex BERT neural network.

However, the workload of each application varies throughout the day. While the Rest API is queried virtually constantly during normal business hours (8AM–4PM), Sentiment Analysis classifies a new batch of articles every hour. Not only from the point of view of application sustainability and robustness, but also from the point of view of cost-effectiveness, we did not implement the SFA system as a single monolith but opted for a “micro services” architecture. Although there is no strict definition of the term micro service, but we can talk about the framework that this architecture should fulfill:

- services communicate via HTTP (Fowler, 2015),
- services are independent of each other (Nadareishvili, 2016),
- services can be implemented using different programming languages, databases, environments and technologies, depending on the task they are supposed to solve (Chen, 2018),
- services are small, communicate only in a bounded context, autonomous, built and deployed by an automated process (Nadareishvili, 2016).

We implemented the fulfillment of these requirements using Docker technology. In Docker jargon, each service represents a so-called container. The container is a small simple standalone software package and as a monolith contains everything needed to run the application – runtime environment code, system libraries and configuration. Containers isolate software and the environment, ensuring platform independence. Containers can be shut down, restarted and changed independently without directly affecting the running of any other service.

It also allows us to run each service individually and thus efficiently optimize costs – e.g. by scaling performance or shutting down servers completely when the service is not in use. This would not be possible with a monolith architecture.

We provided the automated build requirement for the deployment using Github CI/CD, which provides tools to automate this process in addition to versioning the code. Thus, the developer's only concern is to upload code changes to the main repository. To minimize errors in the code, automated tests are run before each deployment to prevent changes from being introduced to the code if they fail.

3.3 Communication interface

The Flask framework[1] was used for development, which implements the usual best-practices in the field of web application development. It also enables fast deployment thanks to the WSGI communication interface, which is supported by an Apache web server. The framework works as a REST-API, i.e. that, based on the URI and HTTP method of the request performs action and returns the appropriate response, usually structured in JSON format. As a result, a single backend implementation can handle both web and mobile client requests.

During the development we put emphasis on the creation of documentation, which is often neglected resulting in a mismatch between the actual and documented functionality. To avoid this, we used tools that document individual API endpoints automatically. The result is the Swagger framework user interface, which clearly specifies how one can interact with our REST API.

3.4 Data storage

Two different repositories are used for the storage of persistent data. Articles are automatically retrieved by scraping from various websites. Unfortunately, each site has a different structure, which also changes over time, so we cannot rely on the consistency of the data obtained. Therefore, in the first phase, the data is stored in the non-relational database MongoDB. Non-relational databases do not store data in tables, but as so-called documents. Each document then stores the data in key-value form, but the internal structure of the documents may not be uniform (i.e., each document may contain different keys).

In the second phase, the data is normalized and stored in the traditional PostgreSQL relational database. In this database, we store structured information about each article, such as the title, date of publication, author abstract and, of course, the textual content of the article.

To avoid having to modify the scraper every time the site structure changes, we implemented an intermediate step where we transform data from a non-relational MongoDB database into a relational PostgreSQL database. In addition to changing the structure, this intermediate step allows us to keep track of changes to the article content itself that may have occurred over time.

[1] https://flask.palletsprojects.com/

3.5 Text processing

An important part of the Analytical platform is the analysis of published articles and news on portals dealing with the area of financial markets. This information is used in the Analytical platform as a basis for decision support when building an investment portfolio. When analyzing the content of articles, we mainly focus on:

- Summarization – the original long text can be shortened by the extractive or abstractive summarization technique to a paragraph containing a few sentences. The extractive technique is based on identifying and selecting the most important parts from the original text, the abstractive will generate a text containing the main topics of the article.
- Sentiment analysis – for the summarized content of the article, the sentiment is evaluated into one of 3 classes: positive, negative or neutral. The sentiment of each news can be evaluated from different perspectives, for example, information about the decline in electric car sales will not please Tesla shareholders, but on the other hand, oil company shares will probably rise.

3.5.1 Text summarization

We analyzed several different approaches from the point of view of both abstractive and extractive summary creation. We compared outputs of our own solutions and commonly available libraries (e.g. Sumy[2], Python Text Summarizer[3]) in Chochula (2021).

The summary, which contains essential information from the whole article, was created in an extractive way. The Gensim library from the Python programming language was used to create the summary. The extraction process is based on the TextRank algorithm (Mihalcea and Tarau, 2004), which is a variant of the PageRank (Brin and Page, 1998) algorithm used by Google to evaluate the relevance of websites.

Due to dissatisfaction with the outputs obtained from the above-mentioned approaches, we decided to focus on newer algorithms based on large language models (LLM). As Deng (2023), Zhang (2023a), Zhang (2023b) or Sun (2023) state, the use of LLM for sentiment analysis in news and articles from the field of financial markets is possible. The same approach can be applied in the case of summarizing the content of reports and articles from this area. The advantage of using LLM is the possibility of generating abstract summaries, which allows us to achieve a more readable level of content compared to the extractive approach while maintaining the same information value. The use of LLM in the field of summarization is mentioned, for example, by the Laban (2023), Sarode (2023) or Zhang (2023c).

The goal is to get information about the content of the article summarized in a few sentences and structured in JSON format. This is important for further processing and presentation of the content. In order to get output from LLM in the required format, we need to create a suitable prompt. Preparing a prompt is not always simple and straightforward. We arrived at the desired result after several iterations and modifications. The key requirements that had to be included in the prompt were a list of individual parts that should appear in JSON format – a summary of the entire article and an array of strings in the form of bullet points. It was also important to state that the output should be formatted in accordance with the RFC8259 specification defining the structure of the JSON format. The last key point made in the prompt included a demonstration of how the result should be formatted. Based on these requirements, we subsequently obtained exactly the output we required.

[2] https://pypi.org/project/sumy/

[3] https://www.geeksforgeeks.org/python-text-summarizer/

```
  {
  “summary”: “Apple is rumored to bring the Action button, a capacitive and
pressure-sensitive feature, to all iPhone models in the iPhone 16 lineup in
2024. Previously, this feature was only planned for high-end Pro models. Sources
suggest Apple intends to provide consistency for developers across its entire
product line.”,
  “bulletpoints”: [“Apple is planning to make Action button capacitive and
pressure sensitive”, “Action button will be available on all iPhone models in
iPhone 16 lineup”, “Consistency is the reason behind bringing Action button to
entire lineup”]
  }
```


3.5.2 Sentiment analysis

For sentiment analysis, we used several deep learning algorithms according to the state of the art in this field, such as BERT (and its derivatives) or LSTM (Šťastná, 2022). In general, the drawback of machine learning is the need for a large amount of labeled data to train and build the model. Sentiment analysis is a domain-oriented task; a model trained with data from a certain domain (e.g., movie reviews) usually performs poorly in another domain (e.g., financial news). On the other hand, data from the financial domain is limited and manual labeling is expensive and time-consuming.

Therefore, we collect data from available sources on the Internet. Specifically, we use the web scraper tool to obtain data from seekingalpha.com. The articles on this site are categorized into “Bearish” (market declines) and “Bullish” (market rising) (see Fig. 1). In this terminology, we consider bullish news to be positive and bearish news to be negative. In total, we have collected more than 8.5 thousand articles that we can use to train the model.

We compared several algorithms and inputs for this task and tried to uncover how article length affects the accuracy of sentiment prediction. We attempted to classify sentiment for three types of inputs – article title only, article summary, and full article text. As shown in Fig. 2, all methods achieved the highest F1 score for article summaries, which suggests that the article title does not contain enough information; on the other hand, the raw article text is likely to contain unusable features that are not relevant for sentiment analysis.

Due to the excellent results, we obtained when using LLM in the area of summarizing the content of articles, we decided to use this modern approach in the area of sentiment identification as well. A big advantage is that it is possible to determine the sentiment not only for the entire article, but also for individual tickers that appear in the text, including the justification for why the rating is positive or negative. The output is again formatted in JSON. Part of

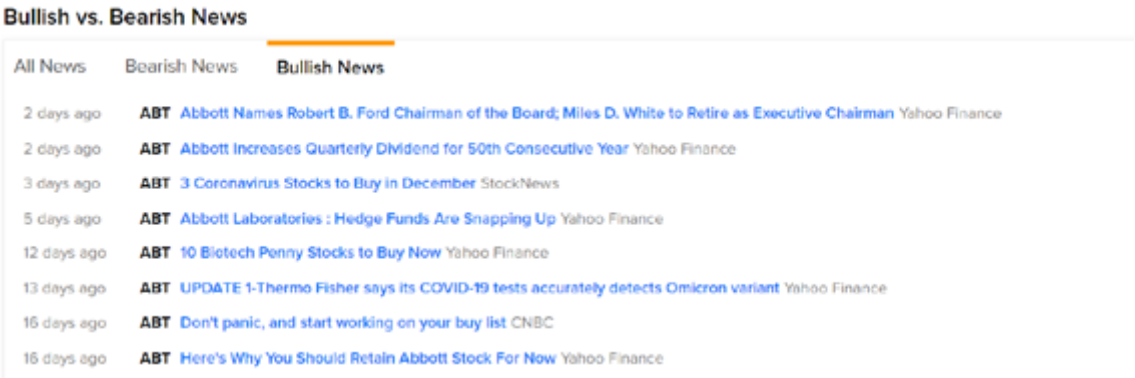


Fig. 1: Bearish and bullish news on Seeking alpha website

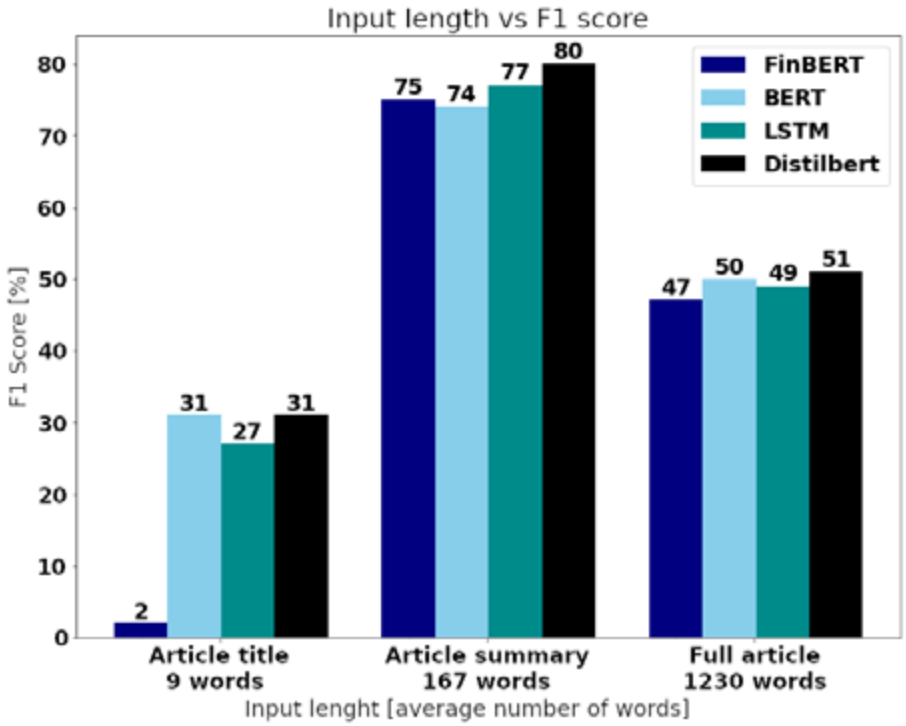


Fig. 2: Performance of methods depending on input length

the prompt again, as in the case of summarization, is the exact specification of the required structure of the JSON output format and the presentation of an example of how the output should look. In the prompt, there is a request to determine the sentiment for the entire article and individual tickets separately.

```
  { “general”: {
          “sentiment”: “negative”,
          “reasoning”: “The text mentions that multiple companies are
struggling, with Tesla reporting a miss on both top and bottom lines,
IBM’s sales coming in just shy of expectations, and Alcoa experiencing a decline
in sales.”
    }
    “companies”: [
          {
          “id”: 5011, “symbol”: “TSLA”, “sentiment”: “negative”,
          “reasoning”: “The text mentions that Tesla is struggling with
declining sales and profit margins, and the company has cut prices multiple times
this year, indicating a negative sentiment.”
          },
          {
          “id”: 4684, “symbol”: “IBM”, “sentiment”: “neutral”,
          “reasoning”: “The text mentions IBM’s mixed results with profit
beating expectations but sales coming in just a bit short of what analysts were
looking for.” } ] }
```

For the reasons stated above, we decided to replace the summarization of articles using the Gensim tool as well as the sentiment classification implemented until now using the BERT tool with a language model that handles both of these tasks.

4 DISCUSSION AND CONCLUSIONS

The presented system is currently used by Cyrrus Corporate Finance, a.s. for their daily usage. Based on client requirements, additional information sources are regularly added and the portfolio of analyzed articles is still expanding. The list of long-term analyzes presented in the form of graphs is also under development. The goal is also to integrate the created API into a complex application. This will allow regular users to create their own business strategies, where one of the key attributes may be the sentiment of published articles

There are a number of similar projects available on the Internet that offer analysis of articles in the area of financial markets and securities trading. Sentimentrader provides daily sentiment report and ad hoc reports if there is anything especially timely or unusual. The Daily Sentiment Report includes an overview of where short- and intermediate-term sentiment is each day, along with updates on indicator extremes or studies focused primarily on sentiment, breadth and price action (Sentimentrader, 2021).

Sentdex (2021) is an application providing sentiment analysis in Finance, Politics and Geographical area. The Financial analysis shows the sentiment for selected companies and long-term analysis of S&P 500 Index. The analysis is available for 7 days, 30 days, 6 months and 1 year interval. The stock price is available in the presented chart too.

Stockgeist is an interactive platform for monitoring the current popularity of 2200+ publicly traded companies. It provides ranking based on the number of published messages about companies, stock overview, aggregated news from various websites, stock sentiment analysis based upon social media messages, wordcloud with the most popular topics the crowd is talking about, fundamentals such as market capitalization and P/E ratio, short story on the selected company to give a general idea of what they do and charts that display historical data about social media messages and positivity. Stockgeist (2021)

Stocktwits is the platform focusing on the same problem. It shows what actual investors and traders are saying in real time about stocks, crypto, futures and forex. (Stocktwits, 2021)

We can see from this overview the sentiment analysis is a very important task in natural language processing and the financial analysts are really interested in information published in articles and news. We know from a discussion with our customer that even a summarized view of published texts is very important. However, other applications do not provide this functionality, which we consider an important advantage of our solution. Our application also offers sentiment analysis related to individual tickers mentioned in the article, which other applications do not allow. The disadvantage of Analytical platforms is the absence of information related to cryptocurrency trading

Our research and experiments show that using LLM for content summarization and sentiment analysis provides better results than procedures based on traditional approaches. LLMs enable not only the creation of higher-quality outputs, for example in the form of abstract summarization of the text, but also more advanced analyzes that were complicated using traditional approaches, such as aspect-based sentiment analysis. Therefore, the tools offered by the Analytical Platform are now based on this technology. The software is available at https://test.analyticalplatform.com/dps/login.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education and by the project EG17_176/0015722 – Data processing software (2020–2023, MPO/EG).

REFERENCES

BRIN, S. and PAGE, L. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7), 107–117.

CHEN, L. 2018. Microservices: architecting for continuous delivery and DevOps. In: *2018 IEEE International conference on software architecture (ICSA),* pp. 39–397. IEEE.

CHOCHULA, P. 2021. Sumarizace dokumentů z oblasti finančních trhů, Final thesis. Mendel University in Brno. https://is.mendelu.cz/lide/clovek.pl?id=71901;zalozka=7;zp=69217;studium=107183

DENG, X., BASHLOVKINA, V., HAN, F., BAUMGARTNER, S. and BENDERSKY, M. 2023. What do llms know about financial markets? A case study on reddit market sentiment analysis. In: *Companion Proceedings of the ACM Web Conference 2023*, pp. 107–110.

FOWLER, M. and LEWIS, J. 2015. Microservices: Nur ein weiteres Konzept in der Softwarearchitektur oder mehr. *Objektspektrum*, 1(2015), 14–20.

JAKÚBEK, T., VALOVIČ, R. and PŘICHYSTAL, J. 2022. Summary of Financial Articles. In: *24th International Conference Economic Competitiveness and Sustainability: Abstracts. 2022*. Brno: Mendelova univerzita v Brně, 18. ISBN 978-80-7509-826-9

LABAN, P., KRYŚCIŃSKI, W., AGARWAL, D., FABBRI, A. R., XIONG, C., JOTY, S. and WU, C. S. 2023. SummEdits: Measuring LLM Ability at Factual Reasoning Through the Lens of Summarization. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 9662-9676.

MIHALCEA, R. and TARAU, P. 2004. Textrank: Bringing order into text. In: *Proceedings of the 2004 conference on empirical methods in natural language processing*, pp. 404–411.

NADAREISHVILI, I., MITRA, R., MCLARTY, M. and AMUNDSEN, M. 2016. *Microservice architecture: aligning principles, practices, and culture*. O'Reilly Media, Inc.

PŘICHYSTAL, J., JAKÚBEK, T. and VALOVIČ, R. 2023. NLP Studio. Working paper. In: *25th International Conference Economic Competitiveness and Sustainability: Abstracts.* Brno: Mendelova univerzita v Brně, 2023. 45. ISBN 978-80-7509-912-9. https://akela.mendelu.cz/~jprich/prichystal-NLPstudio.pdf

SARODE, K., JAVAJI, S. R. and KALAKONNAVAR, V. 2023. Question-focused Summarization by Decomposing Articles into Facts and Opinions and Retrieving Entities. arXiv preprint. *arXiv*: 2310.04880.

SENTDEX. 2021. *Sentdex* [online]. http://sentdex.com/.

SENTIMENTRADER. 2021. Sentimentrader [online]. https://sentimentrader.com/.

STOCKGEIST. 2021. *Stockgeist* [online]. https://www.stockgeist.ai/.

STOCKTWITS. 2021. *Stocktwits* [online]. https://stocktwits.com/.

SUN, X., LI, X., ZHANG, S., WANG, S., WU, F., LI, J., ... and WANG, G. 2023. Sentiment analysis through llm negotiations. arXiv preprint. *arXiv*: 2311.01876.

ŠŤASTNÁ, P. 2020. *Analýza sentimentu v oblasti financí a obchodování s cennými papíry*. Final thesis. Mendel University in Brno. https://is.mendelu.cz/auth/lide/clovek.pl?id=68274;zalozka=7;zp=71400;studium=107655

ZHANG, B., YANG, H., ZHOU, T., ALI BABAR, M. and LIU, X. Y. 2023a. Enhancing financial sentiment analysis via retrieval augmented large language models. In: *Proceedings of the Fourth ACM International Conference on AI in Finance*, pp. 349-356.

ZHANG, W., DENG, Y., LIU, B., PAN, S. J. and BING, L. 2023b. Sentiment Analysis in the Era of Large Language Models: A Reality Check. arXiv preprint. *arXiv*: 2305.15005.

ZHANG, B., YANG, H., ZHOU, T., ALI BABAR, M. and LIU, X. Y. 2023c. Enhancing financial sentiment analysis via retrieval augmented large language models. In: *Proceedings of the Fourth ACM International Conference on AI in Finance,* pp. 349–356.

Contact information

Jan Přichystal: email: jan.prichystal@mendelu.cz

Roman Valovič: email: roman.valovic@mendelu.cz

A LITERATURE REVIEW OF BUSINESS PERFORMANCE MEASUREMENT

Nikola Sobotková[1]

[1]Department of Finances, Faculty of Business and Management, Brno University of Technology, Kolejní 2906/4, 612 00 Brno, Czech Republic

ABSTRACT

The paper is aimed at a critical review of the literature dealing with the measurement of business performance.

Because, nowadays the importance of implementing modern and effective management methods to maintain competitive advantage in almost all business sectors is emphasised, given the increasing competitive pressure. The measurement of business performance is also an important aspect of management and decision-making in organisations. Various indicators are currently being investigated to show the importance of modern approaches and effective measurement systems. This paper aims to identify a list of these modern methods, their bottlenecks and point out the possibility of introducing new and better indicators for performance measurement.

The aim of this work is thus to create a critical review of the literature, especially about the latest findings of research articles on the selected topic. The purpose of this article is then to point out the limits of the current state of literature in the field of modern methods to measure business performance and highlight possible research gaps arising from the review in this area.

Keywords: literature review, business performance, performance measurement methods

JEL Code: G30, M21

1 INTRODUCTION

Nowadays, many factors influence the performance of a company. In the past, performance was measured only by accounting or financial ratios. While these financial indicators can be used to measure the performance of a company and to assess its current situation, nowadays it is no longer possible to view the performance of a company solely from this financial perspective. However, this does not mean that this approach is no longer used. Financial ratios are still important for business management today, but they are insufficient if taken in isolation. It is also necessary to look at the company from many other perspectives, including, for example, customer or employee satisfaction. The interconnectedness of the company's strategy with all activities in the company is also an important aspect. Strategy can be seen as the

https://doi.org/10.11118/978-80-7509-990-7-0137

readiness of the business for the future and the basis for future decision-making. It must be based on the needs of the business and respond to external changes.

There is also huge competition across sectors in the market today. Therefore, it is imperative that every business keeps an eye on the needs of its customers and tries to offer them a different product to improve its position in the market. So one of the key factors is the needs of the customer. However, before a business can reach the customer with a service, the employee is also at the forefront and has a significant role in creating the product that leads to satisfying the customer's needs. A company should therefore also be concerned about the satisfaction of its employees because there is also a lot of competition in the labour market today. If a company has a satisfied employee, this individual can contribute to a satisfied customer. Other factors are technology and improving processes during production and its efficiency.

There are now many modern methods for monitoring and measuring business performance that look at performance from perspectives other than financial, as was previously the case. Each of these methods looks at business performance in a slightly different way and uses different indicators to measure and manage the business. However, the basic point remains the same.

"If you can´t measure it, you can´t manage it."
Peter. F. Drucker

2 A LITERATURE REVIEW

This chapter aims to demonstrate in one place the definition of the basic knowledge for a comprehensive picture of the current situation of literary mapping in the given topic.

1.1 Business Performance

In the context of business performance and its subsequent measurement, it is important to analyse the different terms used.

In economics, there is a drive towards a unified definition of the term enterprise. Wöhe (2013) defines an enterprise as an economic unit engaged in the production and sale of goods and services to make a profit. The concept of the enterprise was also defined in the Commercial Code (Act No. 513/1991 Coll.), which was replaced in 2014 by the Business Corporations Act (Act No. 90/2012 Coll.) and the New Civil Code (Act No. 89/2012 Coll.), where the enterprise is defined as a business corporation, in other words, a form of company or cooperative

The word performance has recently been used in almost everyday operations, across all disciplines. In this article, the definition of business performance will be important.

Many authors agree that corporate performance has long been a highly debated topic in finance and also represents an area suitable for further research. As a result of the multiplicity of authors dealing with the topic, there is a diversity of definitions of the very concept of enterprise performance. For this reason, business performance can be assessed from many perspectives and can therefore be considered a multidimensional concept. The large number of authors also causes disagreements in the determination of these parameters. Allen and Tommasi (2001) distinguish five dimensions of performance: efficiency, effectiveness, economy, consistency, and quality. Other authors include fairness and stability among the dimensions. However, most authors agree on the importance of the three core dimensions, namely efficiency, effectiveness, and economy. According to Bouckaert and Halligan (2008), these three dimensions are also known as the 3E model. Again, in the scientific literature, these three concepts are perceived in different ways. King (2006) sees efficiency as the ratio of inputs to outputs of a firm at which maximum volume and quality are achieved using low costs. Efficiency also generally seeks to answer the question of whether we are doing the right

things. Effectiveness, on the other hand, expresses the degree to which effects are achieved in line with objectives. Here the question of whether we are doing things right is addressed. According to Bouckaert and Halligan (2008), efficiency and effectiveness are key determinants of overall organizational performance.

Now to the actual definitions of business performance. Drucker (1992) perceives business performance as "the final test of any organisation", Hindls, Holman and Hronová (2003) as "the ability of a business to capitalise its capital", Nenadál (2004) as "the degree of results achieved by individuals, groups, organisations and other processes" and Kislinger (2011) perceives business performance as "the ability of a business entity to achieve results over a certain period that are comparable based on certain given criteria with the results of other entities".

Business performance can be simplistically described as all the activities in a business that must be interrelated for the business to be efficient and prosperous, to have a tendency for future development and to be able to respond seamlessly to constantly changing business sub--conditions. This is the only way the enterprise can maintain its competitive advantage. Every enterprise should strive to continuously improve the level of performance of its business and constantly monitor, measure and evaluate this level. Because nowadays every enterprise is in a very strong competitive environment. According to Taouab and Issor (2019), firm performance is a key factor in the economic, social and political development of countries.

"do the right things" & "do the things right"
Peter. F. Drucker

Due to the variety of perceptions of individual authors, firm performance can be categorised according to different aspects of functioning within the firm. Thus, according to Kaplan and Norton (1992), Slack, Chambers, and Johnston (2001), Becker, Huselid, and Ulrich (2001), and Kotler and Caslione (2009), we distinguish, for example: financial performance, economic performance, operational performance, human resource performance, social performance, environmental performance, or sustainable performance.

Some authors view the financial and economic performance of the firm as one concept. In the following section, we will deal with the economic performance of the enterprise perceived in this way. However, the current turbulent competitive environment shows that evaluating businesses only in terms of financial and economic performance does not lead to long-term sustainability. It is these other important aspects of business that should be included in the evaluation of overall business performance.

Brooks and Oikonomou (2018) reviewed, for example, the literature on the impact of ESG (environmental, social, and governance) on business performance, reporting a statistically moderately significant positive relationship between CSR (Corporate Social Performance) and business performance. Drempetic, Klein, and Zwergel (2020) add that ESG scores are also influenced by the size of the firm, which is rather due to the better ability of these enterprises to report their data.

According to Marler and Boudreau (2017), HR Analytics has also been developed in recent years, using technology to describe and analyse data related to HR, human capital management, and showing a positive relationship in the context of business performance

2.1 Business Performance Measurement

Business performance can be measured by various methods. For example, as Synek (2008) points out, in practice, non-financial measures of performance are increasingly used alongside financial measures. It is no longer sufficient to look only at hard financial measures quantifiable from mandatorily published financial statements. Nowadays, there is an increasing emphasis on soft metrics - human potential, customer and employee satisfaction, or

innovation and quality. The methods can therefore be divided into traditional and modern. In traditional approaches, measurement focuses only on the past or present of the company. Traditional approaches rely mainly on financial indicators. The conventional approach is therefore to perform a complete financial analysis of the enterprise. However, this approach currently distorts the telling power in assessing the long-term competitiveness of the enterprise. At the same time, they cannot be omitted as they form the basic framework for assessing the financial situation of an enterprise. Modern approaches to business management through the use of non-financial indicators eliminate the non-deficiencies of financial measures. In recent years, they have brought a completely new perspective to the measurement and evaluation of enterprise performance. However, it is essential to select meaningful indicators that reflect the needs of the enterprise. These modern approaches include benchmarking, the EFQM Model of Excellence, Six Sigma or the comprehensive approach using the Balanced Scorecard concept. Solař, Bartoš (2006)

In the following, these selected methods will be briefly characterised and presented.

2.1.1 Benchmarking

Benchmarking is the process by which a company measures its performance against companies that represent either the world's top performers or the company's main competitors. A natural part of benchmarking is benchlearning, or learning from better practice. Nenadál (2004)

Benchmarking can be divided into two basic types, namely internal and external, depending on where it is used.

Internal benchmarking is implemented within an organization and compares the different organizational units with each other. The advantage of this benchmarking is the effective reduction of differences in the performance of these individual centres or organisational units. The disadvantage is that it is more likely to be used in large enterprises where individual centres are more likely to perform similar activities but in different locations. This is almost unattainable for small or medium-sized enterprises.

External benchmarking is a situation where an enterprise is compared with another organisation. In contrast, this type of benchmarking is largely an option for SMEs. However, the problem is to find a suitable organisation to benchmark against, if it is usually a direct competitor. The advantage, however, is that there is an opportunity to learn and improve your processes based on information from a truly best-in-class company. Nenadál, Vykydal and Halfarová (2011)

In terms of the nature of the chosen object, benchmarking can be further divided into several types, which are executive, functional, process or strategic benchmarking.

2.1.2 EFQM Excellence Model

This model represents a comprehensive analysis of the company within all its processes. Through the EFQM model, a company can identify areas for improvement or weaknesses in the business. The EFQM model is internationally recognised as the most comprehensive management tool. Today, it is the most demanding but at the same time the most sophisticated methodological material for the development of management systems in the world. The model covers a total of 9 areas of business management, where five criteria create the prerequisite for good business results and the other four criteria assess the results and objectives of the company. These criteria are leadership, strategy, people, partnerships and resources, processes, customers, people results, company results and key economic results. Each criterion is scored to give the business a view of the results it should be achieving.

The model is used as a means of self-assessment for the company and is used as a basis for strategic planning. The EFQM model is based on an assessment of the tendency to improve in terms of profit, customer orientation, a clearly defined strategy, the development of human potential, innovation, improving relationships with partners and access to employees or customers. Nenadál (2004)

2.1.3 Balanced Scorecard

The most well-known representative of complex performance measures today is the Balanced Scorecard method, introduced in the 1990s by Robert Kaplan and David Norton. It was founded precisely as a modern approach to measuring business performance, in response to criticism of one-sided indicators that were more oriented towards the past and the financial concept of these indicators in the form of hard factors. In today's information era, these financial measures are no longer sufficient. It is now important for a company to focus on soft factors along with hard data, which can include, for example, employee qualifications or customer relations. This concept therefore replaces traditional financial measures focusing on the past but adds new measures. The Balanced Scorecard is therefore a system of balanced scorecards that focuses on measuring performance, but at the same time anchoring it within the entire corporate management system and also seeks to link corporate strategy with operational activities. Kaplan and Norton (2005)

The Balanced Scorecard is nowadays often used, among other things, as a strategic business management system. Thus, it is used in enterprises to manage long-term strategy and provides managers with some framework for implementing critical managerial processes. These processes can include transforming vision and strategy into their goals, communicating these plans and metrics, or improving feedback and learning capabilities. Kaplan and Norton (2005)

The Balanced Scorecard concept adds new measures of financial performance called drivers that can inform changes before they are reflected in lagging indicators. Therefore, these momentum measures are referred to as future performance measures. They can be used by a business to measure how it is creating value for its customers, or the extent to which it needs to improve the skills of its workforce or the quality of its systems. Kaplan and Norton (2005)

2.1.4 Six sigma

This model focuses on the analysis of qualitative problems and their elimination. It is mainly concerned with increasing customer benefit and improving business results. The main aspect here is the approach to improving product quality through teamwork.

Six Sigma aims at identifying and eliminating the causes of errors in the production and business process and then meeting all customer requirements in all important processes. The process of this method starts with the customers and the aim is to satisfy all their requirements to the maximum. There is a strong emphasis on documenting the execution of the work and examining the work among employees on the job site. Thus, this model can be seen as a comprehensive methodology for measuring the performance of processes in a company and their post-intentional improvement. To implement the model in the management of an enterprise, the enterprise needs to meet the following prerequisites: continuous improvement of processes, all activities composed of processes, support of top management, setting priorities, expanding the competence of employees, promoting teamwork, continuous assessment of process performance, determining the subject and method of measurement, and determining work teams

In this method, we encounter the concept of the "magic triangle", which represents the central requirements of competitors, which are quality, time, cost and innovation. If a business can master this magic triangle, it could become better and faster than its competitors. Töpfer (2008)

2.1.5 EVA – Economic Value Added

One of these modern approaches works with economic value added, which shows the efficiency of a company using its capital and the excess of the company's profit over the weighted average cost of capital. The higher its value, the higher the value is for the efficiency of the company's use of capital. The high values of this indicator indicate a high level of capital gain.

Based on the Economic Value Added, a corporate Value Based Management system is developed. This enterprise management system is based on maximizing the economic value

added. The goal of all decisions in the enterprise is then to increase value for shareholders and owners. This management system was developed in the context of the increasing globalisation of markets and is different from the other approaches. It can be seen as one of the latest trends in approaches to assessing the economic efficiency of companies. The globalisation of markets is linked to the new attractiveness of the importance of international investors, whose pressure to defend their interests is greater than that of private investors. Voříšek (2008)

2.1.6 Other methods

However, many other approaches are found in practice. Selecting the appropriate indicator to measure business performance is therefore not straightforward. New concepts for measuring and evaluating non-financial indicators are emerging, which is gradually influencing the traditional methods of measuring business performance. In practice, there is not yet a uniform approach in the field of implementation of these methods and their application in individual enterprises is different. This subsequently leads to problems in the field of inter-company comparison.

Chenhall and Langfield-Smith (2007) also state the necessity of choosing the appropriate method of performance measurement in relation to different disciplines in the company, such as marketing or human resource management, where it is necessary to use non-financial indicators rather than traditional financial indicators. Within marketing or human resources, customer loyalty or satisfaction are important aspects. Thus, managers from the functions in question are trying to develop metrics that are more relevant to their areas of management. In marketing, there are several different areas of research on performance measurement. Customer loyalty then emerges as a more influential factor than, for example, market share.

In conclusion, it is useful to add a final perspective to the methods given. It was mentioned in the introduction that the concept of performance is to some extent also linked to the concept of efficiency. It is therefore important to mention the method of measuring efficiency, which is a key task for management. Kao (2014) mentions the DEA (Data Envelopment Analysis) or modified NDEA method for measuring the effectiveness of systems, which takes into account the effectiveness of the individual processes within the system. This approach

Methods	Concepts	Tools
profit-based - EBIT, EBITDA	BSC - Balanced Scorecard	Financial analysis
cash flow based	EFQM Model excellence	PESTLE analysis
market data based - EPS	Six Sigma	SWOT analysis
based on sub-areas - profitability, activity, liquidity, debt, productivity	TQM - total quality management	Benchmarking
EVA - economic value added	VBM - Value-based management	Controlling
SVA - shareholder value added	JIT - just in time	Financial planning
DCF - discounted cash flow	ABC - activity based cost	KPIs
MVA - market value added	Lean management	Activity-based costing
Other..	Performance Prism	Risk management
	Other..	Other..

Tab. 1 Methods, concepts and tools used to measure business performance
Source: own processing

allows identifying the causes of inefficiencies and provides a detailed view of system performance. DEA analysis also allows the comparison of the performance of different enterprises from an efficiency perspective and identifies those that perform best. It therefore also serves in the context of benchmarking

The following is a listing of the existing range of concepts, methods and tools used to measure enterprise performance.

3 BARRIERS TO CURRENTLY USED METHODS

Measuring business performance is a key part of business management, but there are various shortcomings and limitations of the methods used for this purpose:

Financial indicators versus non-financial indicators: traditional performance measurement methods often depend on financial indicators such as profit, sales, return on investment (ROI), and others. These indicators may inaccurately reflect the true state and potential of the business, as they ignore important non-financial aspects such as customer satisfaction, product quality, innovation capabilities, or company culture.

Short-term versus long-term performance: many methods focus on short-term results, which can lead to ignoring long-term goals and strategies. This focus can encourage decisions that are beneficial in the short term but detrimental to the long-term sustainability and growth of the business.

Relativity and context dependency: Performance measurement often depends on comparisons with competitors or industry standards. This approach can be problematic because it ignores the unique context and strategy of each business. It also does not take into account different market conditions or economic cycles.

Complexity and clarity: Some advanced methods may be too complex and difficult to understand for all employees. This can lead to problems with internal communication and clarity of corporate objectives.

Subjectivity and manipulation: Methods that involve subjective evaluations, such as employee performance evaluations, may be prone to bias and manipulation. In addition, financial ratios can be influenced by accounting practices and creative interpretations.

Change and adaptability: The corporate environment is constantly changing, requiring flexible and adaptable methods of measuring performance. However, many traditional methods may be too rigid and inadequate for today‘s dynamic conditions.

As mentioned above, the Balanced Scorecard concept is considered to be the most comprehensive method. Hoque and James (2000) also confirmed a positive correlation between the use of BSC and firm performance. But this method also has its barriers, which have been mentioned by its authors. These barriers are the unfeasible vision and strategy, the non-connection of the strategy with the objectives of individuals, the non-connection of the strategy with the allocation of resources and only tactical feedback. However, the non-sufficiency of human resources can be considered as the main factor limiting the implementation and use of this method. The problem in most cases is more accurately a non-sufficiency of qualified people. Another serious shortcoming is the poor weighting of indicators and metrics. Kaplan and Norton (2005)

Even though the BSC method is the most widely used, mainly because of its simplicity, 80% of companies make mistakes with this method. An interesting perspective on these mistakes is provided by, for example, Mr Bernard Marr, one of the world‘s most respected experts in the areas of strategy, business performance and AI in business, who has worked with many global organisations including Amazon, Google, Microsoft, Astra Zeneca, NATO, Toyota, The Royal Air Force, Shell, the United Nations, Walmart and many others.

These errors include, for example, the flaw in the development of the balanced scorecard, where in most corporate enterprises only two or three people compile the scorecard. The

system is then biased and has no real reporting capacity. The management should sit down with each key person in the company and talk to them about their strategic objectives, and then develop the system together so that everyone contributes equally. Another mistake occurs when the enterprise chooses what indicators to measure instead of what it should do. The business then measures everything that moves, everything that the competition measures, and not what matters. It therefore does not have well-established KPIs. It measures everything easy to measure and easy to get data for, instead of what it needs to track. The business must start by defining its strategic objectives and designing a clear strategy map from which it will know the purpose of the business and derive the right metrics. If the business has a strategic goal, it should develop a key performance question and then by answering that question it will get an insight into how well it is meeting that goal, and only then will the information need to be formulated according to that question. Another error is caused by looking at the Balanced Scorecard method as a four-perspective framework, where the company starts filling in the individual squares instead of thinking about what its business strategy is and what it needs, what it wants to achieve in terms of financial perspective, what it wants to sell to customers, what its value proposition is and what are the basic factors that will allow us to cover the entire strategy of the company. With this perspective, the framework can have five or six perspectives. Following just looking at the method as a four-perspective framework, we need to first identify a strategic map to understand the cause-and-effect relationship as mentioned earlier. It is often stated that authors Kaplan and Norton would start with this third book first, as this understanding is essential for the next steps. This visual map then gives a clearer overview of the whole enterprise and goes well with communication. It is a complete mistake to adopt this balanced system from someone else who has already worked out their structure. This is tempting, but every business is different and focuses on different clients etc. Thus, it is logical that it will have other appropriate indicators. Many businesses also create a strategy that looks good. However, if a business does not make sure that it has a clear action plan and projects that will enable it to meet its strategic objectives, it will have non-existent strategic plans and fail to implement them in the long run. Balanced scorecards should propose actions and initiatives for each strategic objective. A clear catch is if the business does not regularly communicate internally. If a business measures KPIs, has strategic goals and tracks measured metrics and indicators, it has the ideal foundation and data to understand, analyze and communicate all of this. It should engage people across the business to get information on how well the targets are being met throughout the year. Further, these metrics should be regularly reviewed and monitored monthly. Discuss together what the business is doing well, why things need to change and how to do things in the future, for example using dashboards. Instead of management just being overwhelmed with data that no one else understands and tracking the past instead of focusing on how to support future developments. The final shortcoming is that the business does not revise its strategy even though its business, environment, competitors or products change. It is necessary to check strategic goals to see if they need updating and are still true (Marr, 2020).

Even the simple benchmarking method has its reasons for not being extended. The negative side of this method is, for example, overestimation of the company‘s results, management‘s rejection of external information and the fact that management often fails to admit its shortcomings.

Another significant barrier to all modern indicators for measuring the performance of an enterprise is the fact that these methods were developed mainly in 1986–1993, so they can no longer be considered completely modern, given the turbulent development of the economic environment.

Ittner and Larcker (2003) also point out the numerous errors that companies commit when trying to measure non-financial performance indicators. These errors again include incorrect setting of objectives, misidentification of appropriate non-financial indicators for the firm, or erroneous measurements.

Tangen (2004) points out that many businesses still rely on traditional financial performance indicators and fail to recognise the importance of non-financial measures.

It is important to note that no one method can perfectly measure the performance of an enterprise, and therefore it is often recommended to use a combination of them for a more comprehensive and balanced assessment. According to Taouab and Issor (2019), finding the ideal concept for managing and measuring business performance is a complex problem.

4 RESULTS

Measuring business performance is a complex process that includes various aspects such as financial indicators, resource efficiency and market competitiveness. A key element is the financial health and stability of the business. Financial analysis tools have long been used to assess financial performance. These traditional financial ratios are still widely used, but their limitations in capturing the complexity of modern business operations are increasingly being discussed in the context of the growing tendency towards a competitive environment. Thus, non-financial indicators have become important in recent years, seeking to eliminate the shortcomings of financial ones.

The weakness in modern indicators for measuring corporate performance lies in the need for more comprehensive and modern measures to match the evolving business environment.

5 DISCUSSION AND CONCLUSIONS

This paper has attempted to review the literature on measuring sub-enterprise performance.

Enterprise performance measurement is a critical aspect of management and decision-making in any enterprise. Various publications have examined different dimensions and indicators of measuring the performance of a sub-enterprise and have shed light on the importance of modern approaches and effective measurement systems. At the same time, their barriers have been pointed out.

The current state of the literature provides valuable insights into the development and importance of modern indicators for measuring enterprise performance. Overall, the literature reflects the developments in the field of enterprise performance measurement and highlights the need for modern indicators, effective measurement systems and linking the necessary aspects to enhance organizational performance.

A multifaceted approach is needed to address the shortcomings in modern indicators for measuring enterprise performance. The current literature review, within the framework of the obsolescence of the so-called “modern” approaches, constantly opens up space for further investigation and research in the field.

Acknowledgements

This article contributes to the framework of the specific research project “Development Trends of Selected Aspects of Corporate Financial Management”, under the registration number FP-S-22-7935, Faculty of Business, Brno University of Technology.

REFERENCES

ALLEN, R. and TOMASSI, D. 2001. *Managing Public Expenditure. A Reference Book for Transition Countries*. Francie: OECD. 656 p. ISBN 92-64-18653-0. https://read.oecd-ilibrary.org/governance/managing-public-expenditure_9789264192607-en#page358

BECKER, B. E., HUSELID, M. A. and ULRICH, D. 2001. *The HR Scorecard: Linking People, Strategy, and Performance*. Harvard Business School Press.

BOUCKAERT, G. and HALLIGAN, J. 2008. *Managing Performance: International Comparisons*. Routledge. 272 p. ISBN 978-0415414953

BOUCKAERT, G., HALLIGAN, J. 2008. Managing Performance: International Comparisons. New York: Routledge. 276 p. ISBN 978-0415464544

BROOKS, C. and OIKONOMOU, I. 2018. The effects of environmental, social and governance disclosures and performance on firm value: A review of the literature in accounting and finance. *The British Accounting Review*, 50(1), 1-15

CHENHALL, R. H. and LANGFIELD-SMITH, K. 2007. Multiple perspectives of performance measures. *European Management Journal*, 25(4), 266-282.

DREMPETIC, S., KLEIN, C. and ZWERGEL, B. 2020. The influence of firm size on the ESG score: Corporate sustainability ratings under review. *Journal of Business Ethics*, 167(2), 333-360.

DRUCKER, P 1992. *The Age of Discontinuity: Guidelines to Our Changing Society*. paperback.

HINDLS, R., HOLMAN, R. and HRONOVÁ, S. 2003. *Ekonomický slovník*. 1. vydání. Praha: C. H. Beck, ISBN 80-7179-819-3

HOQUE, Z. and JAMES, W. 2000. Propojení měřítek vyvážených ukazatelů s velikostí a tržními faktory: Vliv na výkonnost organizace. *Journal of Management Accounting Research*, 12, 1-170.

ITTNER, C. D. and LARCKER, D. F. 2003. Coming up short on nonfinancial performance measurement (Vycházení vstříc nefinančnímu měření výkonnosti). *Harvard Business Review*, 81, 11.

KAO, C. 2014. Network data envelopment analysis: A review. *European Journal of Operational Research*, 239(1), 1-16.

KAPLAN, R. S. and NORTON, D. P. 1992. The balanced scorecard: Measures that drive performance. *Harvard Business Review*, 70(1), 71-79

KAPLAN, R. S. and NORTON, D.P. 2005. *Balanced Scorecard: strategický systém měření výkonnosti podniku*. 4. vydání Praha: Management Press, 267 p. ISBN 80-7261-124-0

KISLINGEROVÁ, E. a kol. 2011. *Nová ekonomika – nové příležitosti?* 1. vydání, Praha : Nakladatelství C. H. Beck, 322 p., ISBN 97-880-7400-403-2

KOTLER, P. and CASLIONE, J. A. 2009. *Chaotics: The Business of Managing and Marketing in the Age of Turbulence*. New York: American Management Association. 206 p. ISBN 0-8144-1521-0

KRÁL, B. 2006. *Manažerské účetnictví*. 2. vydání. Praha: Management Press. 622 p. ISBN 80-7261-141-0

MARLER, J. H. and BOUDREAU, J. W. 2017. An evidence-based review of HR Analytics. *The International Journal of Human Resource Management*, 28(1), 3-26

MARR, Bernard. 2020. The Top 10 Balanced Scorecard Mistakes You Must Avoid. *Youtube* [online]. https://www.youtube.com/watch?v=J_iygjqEKdA [Accessed: 2024-01-15].

NENADÁL, J. 2004. *Měření v systémech managementu jakosti*. 2. doplněné vydání. Praha: Management Press, 2004. 335 p. ISBN 80-7261-110-0

NENADÁL,J., VYKYDAL, D. and HALFAROVÁ, P. 2011. *Benchmarking: mýty a skutečnost: model efektivního učení se a zlepšování*. 1 vydání. Praha: Management Press, 265 p. ISBN 978-80-7261-224-6

SLACK, N., CHAMBERS, S. and JOHNSTON, R. 2001. *Operations Management*. 3rd Edition. Pearson Education.

SOLAŘ, J. and BARTOŠ, V. 2006. *Rozbor výkonnosti firmy*. 3. vydání. Brno: VUT Brno, Fakulta podnikatelská, 163 p. ISBN 80-214-3325-6

SYNEK, M. 2008. *Podniková ekonomika*. 4. vydání. Praha: CH Beck. 526 p.

TANGEN, S. 2004. Měření výkonnosti: od filozofie k praxi. *International Journal of Productivity and Performance Management*, 53(8), 726-737. ISSN 1741-0401

TAOUAB, O. and ISSOR, Z. 2019. Firm performance: Definition and measurement models. *European Scientific Journal*, 15(1), 93-106.

TÖPFER, A. a kol. 2008. *Six Sigma, Koncepce a příklady pro řízení bez chyb*. 1. vydání. Brno: Computer Press, 508 p. ISBN 978-80-251-1766-8

VOŘÍŠEK, J. a kol. 2008. *Principy a modely řízení podnikové informatiky*. Praha: Economia, 441 p. ISBN 978-80-245-1440-6

WÖHE, G. 2013. *Einführung in die Allgemeine Betriebswirtschaftslehre*. Vahlen Verlag. 950 p. ISBN 978-3800617289

Contact information
Nikola Sobotková: e-mail: Nikola.Sobotkova@vutbr.cz

HOW DOES A COUNTRY'S LEVEL OF ECONOMIC DEVELOPMENT INFLUENCE DIGITAL ADVANCEMENT? EVIDENCE FROM EUROPEAN COUNTRIES

Anđelka Stojanović[1], Isidora Milošević[1], Sanela Arsić[1]

[1]Technical Faculty in Bor, University of Belgrade, VJ 12, 19210 Bor, Serbia

ABSTRACT

Recent years have brought numerous challenges to Europe and the world in the form of a complex geopolitical situation, threats to the safety and health of the population, deepening economic differences, and a lack of natural resources. Dealing with some of the challenges mentioned above is carried out through digitization and the development of the single market. The European Commission enacted numerous documents and implemented a series of activities aimed at strengthening the European digital market. Actions and policies aim to support digitization to increase economic activity and achieve other social benefits such as empowerment of people, solidarity and sustainability. Due to unequal initial levels of development and the readiness of states, companies and individuals for the changes brought by information technologies, a digital divide emerged. The main research question in this paper is whether the achieved level of economic development conditions the level of digitization. It also examines which elements in the digital transformation can be most influenced to reduce the digital divide. Structural equation modeling (SEM) is applied to answer the research question. The importance of the results is reflected in the fact that by confirming the connections between the digital level and economic parameters, the direction of influence on the reduction of digital divides and the fulfillment of globally set goals related to social equality can be defined.

Keywords: digital development, economic development, European countries, digital society and economy index

JEL Code: F63, O33, O52

https://doi.org/10.11118/978-80-7509-990-7-0148

1 INTRODUCTION

Digital technologies such as the Internet, electronic devices, and applications that collect, store, analyze, and share information are changing the world by altering how people go about their lives and businesses. These technologies provide new opportunities for people to establish their lives and businesses globally. However, in the context of the pandemic, wars, and environmental issues, the ongoing global crisis is becoming more severe and worsening the world's economic situation.

Digital development in developed countries happened rapidly with the help of governments and financial institutions but also with the support of international institutions. Developing countries, on the other hand, faced a lack of investment funds and a weak ITC infrastructure. Due to unequal initial levels of development and the readiness of states, companies and individuals for changes brought by the expansion of information technologies, the emergence of a digital divide has occurred, especially in cases of poorly developed and underdeveloped economies (Mubaraket al., 2020; Habibi and Zabardast, 2020). Even within the EU, not all member states have the same understanding and acceptance of digital transformation (Balcerzak and Pietrzak, 2017; Traşcă et al., 2019). Accordingly, the European Commission adopted numerous documents and implemented a series of activities aimed at strengthening the European Digital Single Market (European Commission, 2021a; European Commission, 2021b). The goals of the activities and policies are not only to support digitization to increase economic activity, but other social benefits should be achieved. The new vision of the European Union named 2030 Digital Compass states that digitization is a new force and imperative for achieving solidarity, sustainability and empowerment of people and businesses. Looking from a national point of view, it is necessary to build the basis for an efficient digital economy to influence the reduction of digital differences at the regional level and improve the country's competitive position.

Previous research on the connections between the level of digitalization and its results mainly refers to the impact of digitalization on economic performance, environmental performance, sustainability and social consequences (Li et al., 2020; Popkova et al., 2022). In light of the significant deepening of economic differences, the dilemma remains whether countries with a lower level of economic development and the ability to invest in ICT can actually initiate digital transformation and achieve benefits. There appears to be a lack of studies contributing to understanding and explaining the impact of a country's economic strength on digitalization. There is still space in the literature for answering questions: 1) What are the key variables defining the structure of digital development in European countries? 2) Does a spatial pattern of economic development affect specific aspects of digital development levels?

The novelty of this research lies in asking questions that are rarely found in the literature. Specifically, it focuses on questioning the relationship between economic and digital levels, opposing the main body of literature, which presupposes the influence of digital on the economic level. Therefore, the main research question in this paper is whether the achieved level of economic development conditions the level of digitization. It also examines which elements in the digital transformation can be most influenced to reduce the digital divide. Data on the digital society and economy and the level of economic development are used to define the research model, and structural equation modeling is applied to answer the research question. The importance of the results is reflected in the fact that by confirming the connections between the digital level and economic parameters, the direction of influence on the reduction of digital divides and the fulfillment of globally set goals related to social equality can be defined.

2 THEORETICAL FRAMEWORK

It is believed that the digitalization of society and the economy can significantly help in overcoming the crises that have arisen in recent years by removing physical borders, unhindered communication, and acquiring certain rights, as well as creating numerous opportunities for innovation and new markets. However, attempts to respond to crises have also shown the weaknesses of the digital space and created strong divides.

When analyzing research dealing with the level of digital development and the digital divide, the availability of certain information and communication technologies is first of all examined. In fact, the digital development of a country is measured by the connectivity of households and individuals. ICTs then imply the existence of an infrastructure consisting of electronic equipment, networks and software that can collect, exchange, process and store data (Lucendo-Monedero et al., 2019). When considering the impact of telecommunication infrastructure on economic growth, the conclusion is that basic telecommunication technologies (broadband technology) have a more significant impact in less developed countries, while in developed countries, the impact of mobile technology is more noticeable (Habibi and Zabardast, 2020; Myovella et al., 2020). The explanation of the results lies in the fact that underdeveloped countries still base their ICT level on broadband technology, while the mobile infrastructure is less developed.

The increased digitization and the inclusion of information and communication technologies in all spheres of life bring numerous business and social challenges. With the inclusion of Industry 4.0 technologies in business, traditional chains of value creation and the dynamics of the modern business environment are changing. The integration of digital technologies into business facilitates the collection and analysis of large amounts of data that are generated in modern operations, as well as speeding up the production process and responding to market demands, producing higher quality products with increased savings and achieving circular and sustainable goals (Traşcă et al., 2019). The main barrier to introducing digitalization in business processes is high costs and lack of funds, as well as the possibility of specific processes being optimized following new technological requirements (Pech and Vaněček, 2022). Economics is particularly highlighted as a critical obstacle in developing countries where companies do not have sufficient funds for investments in new technologies and the means to ensure returns from those investments (Kyobe, 2011).

Integrating ICT in the entire value chain can be complex, especially for small and medium-sized enterprises (SMEs). This is because it requires the use of networks and various IT systems. As a result, many SMEs find it challenging to achieve the necessary technological level required to use Industry 4.0 solutions and adapt their organizational structure to the new way of doing business (Sevinç et al., 2018).

The complexity of the organizational structure of companies and the readiness of management and employees to accept a new way of thinking and working have been identified as significant barriers to digitization and the application of Industry 4.0 technologies (Pech and Vaněček, 2022). Digitization brings disruptive changes in the working environment and working conditions, and new skills and specific knowledge are required, while, on the other hand, employees react much more slowly. Digital competencies include multiple disciplines related to using digital tools and applications, finding and understanding information, critical thinking and problem-solving, and the ability to communicate through ICT (Picatoste et al., 2018). Given that digitization reduces the need for low-income, low-skilled employees, digital competencies become crucial for securing a job and, consequently, higher employee productivity. The widespread adoption of Industry 4.0 has led to a deficit of competent professionals with the necessary technical skills and knowledge to facilitate the transition to the new methods of production, which involves a real-virtual working space. This shortage is particularly prevalent in specific technical occupations (Liboni et al., 2019). According to a study by Schröder in 2017, the shortage of human resources seen initially will be resolved in time with the

increasing number of programs and students pursuing education in the MINT subjects, i.e., mathematics, informatics, natural sciences, and technology (Schröder, 2017).

Mubarak et al. (2020) test the impact of socioeconomic factors on ICT acceptance. The results indicate a strong positive association of income and education with levels of ICT dispersion worldwide. Research by Heinz (2016) speaks about the influence of socioeconomic background on the availability of ICT technologies as an essential factor for developing digital competencies, highlighting the occurrence of inequality conditioned by social inequality. Kwilinski et al (2020) considered that countries at a higher level of digitization have a lower percentage of the population at risk of poverty and social exclusion. However, the results showed that a high level of digitization does not reduce the risk of poverty and social exclusion, especially in poorer EU members, due to low digital competencies, and the economic impact should be directed not only at ICT infrastructure but also people.

The ability to adopt and use new technologies largely depends on the environment created by the government by enacting specific regulations that help digitization, but also by investing in infrastructure and developing its digital competencies. Therefore, with the digitization of the business and personal spheres of life, the expected and required changes also apply to government services. In accordance with all the changes, digital government is being introduced, the task of which is to create the conditions for certain social demands of the population to be met through digital government systems. This means, on the one hand, the construction of a central functional infrastructure internally, but also the promotion of these services and harmonization with the needs and digital competencies of the population so that the built system functions externally. Yifan and Bei (2022), when examining the factors that influence the creation of digital government concluded that better economic conditions and financial support for digital transformations provide a better basis for creating digital government and e-services.

Following the research question and analysis of digitalization aspects, the proposed hypotheses are:

- **Hypothesis 1.** The economic level has a positive influence on ICT accessibility
- **Hypothesis 2.** The economic level has a positive influence on the integration of ICT in business
- **Hypothesis 3.** The economic level has a positive influence on the level of ICT skills of the population
- **Hypothesis 4.** The economic level has a positive influence on the digitization of public services.

3 METHODOLOGY AND DATA

The mentioned literary sources point to strong digital divides at different levels, whether they are about more economically developed and less developed countries or at the level of certain population groups. It can also be concluded that the digital divide is not only a consequence of the availability of digital technologies but also of the level of ability to use these technologies. Therefore, the problem of digital development is multidimensional (Balcerzak & Pietrzak, 2017), and its solution requires a complex analysis.

The problem is approached using the Digital Economy and Society Index (DESI) indicator. Namely, since 2014, the European Commission has been monitoring the digital progress of EU member states. To this end, the Eurostat Database, which provides uniform and high-quality statistical data on various topics in Europe, is used. First, five domains were used to define DESI, while the DESI 2021 report contained four areas under consideration: human capital, broadband connectivity, the integration of digital technologies by businesses and digital public services. DESI was adapted in 2021 to reflect two important policy initiatives: the Recovery and Resilience Facility and the Digital Decade Compass (European Commission,

2021b). The Eurostat Database also contains data for other European countries, making it possible to assess digital performance at the European level. DESI is also used in other research as a valuable tool for the unique identification of the level of digital competencies of countries (Kwilinski, 2020). In this research, data from the Eurostat database was mostly used for the indicators of digital development factors, while some indicators used OECD data.

Three socioeconomic indicators were chosen and combined into one factor to determine economic growth. First is the Gross Domestic Product (GDP) per capita, whose values were obtained from World Development Indicators and represent gross domestic product divided by midyear population. Another important indicator is the Employment Rate, which is the percentage of employed persons in relation to the total population. The indicator is based on the European Labor Force Survey (EU-LFS) results. The third indicator of economic development is the Final Consumption Expenditure of the General Government. General government final consumption expenditure consists of expenditures for collective consumption (defense, justice, etc.), which benefit society as a whole and expenditures for individual consumption (health care, housing, education, etc.), which reflect spending incurred by the government on behalf of an individual household (OECD, 2014). Figure 1 displays a general model of the relationship between Economic level and Connectivity, Integration of digital technology, Human Capital and Digital public services. The list of indicators used for assessing the proposed constructs in the research model is presented in Table 1.

Construct	Item	Description	Source
Connectivity	a1	Overall fixed broadband take-up	Eurostat
	a2	At least 100 Mbps fixed broadband take-up	OECD
	a3	Fast broadband (NGA) coverage	OECD
	a4	Mobile broadband take-up	OECD
Integration of digital technology	b1	SMEs with at least a basic level of digital intensity	Eurostat
	b2	Social media	Eurostat
	b3	Cloud	Eurostat
	b4	SMEs selling online	Eurostat
	b5	e-Commerce turnover	Eurostat
Human capital	c1	Above basic digital skills	Eurostat
	c2	ICT specialists	Eurostat
	c3	Enterprises providing ICT training	Eurostat
Digital public services	d1	e-Government Index	UN
	d2	e-Participation Index	UN
Economic growth	GDP	GDP per capita	World Development Indicators
	Empl	Employment	Eurostat
	Fin_cons	Final consumption expenditure of general government	Eurostat

Tab. 1 Items used to validate the hypothesis in the proposed model

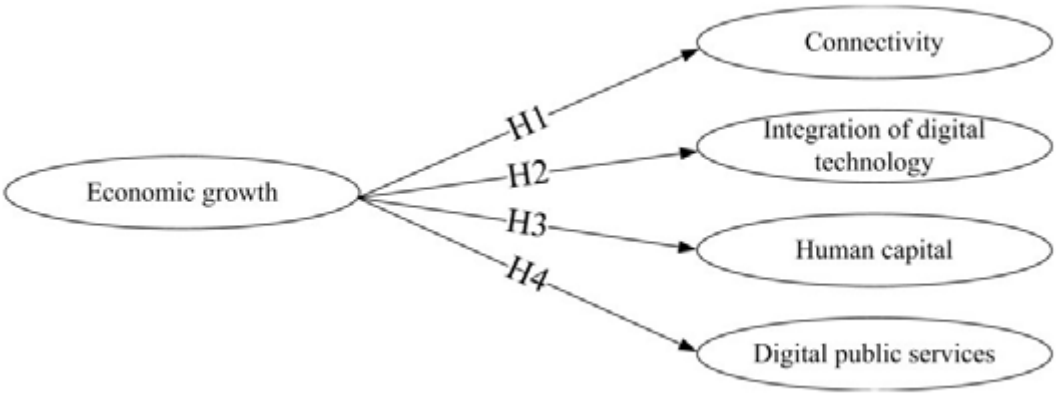


Fig. 1: Proposed research model

The analysis of the proposed research model, which assumes the impact of economic factors on digital economy factors, is based on Structural Equation Modelling (SEM). SEM is a widely used technique for testing theoretical hypotheses. It consists of two main components: a measurement model, which is essentially a confirmatory factor analysis (CFA), and a structural model that examines the hypothesized relationships between latent constructs. Confirmatory factor analysis (CFA) is a method for confirming assumptions based on theory. It involves defining the variables that represent the factors and creating a measurement model to show how the observed variables describe the latent variables in the proposed model. The key aspect of CFA is to assess the reliability and validity of the observed variables and their interconnections. The structural model examines the effects of independent variables on dependent variables. Once the measurement model is validated, the structural model is defined by establishing the relationships between the constructs. Through successive regression equations, the structural model determines the characteristics of the relationships between variables. Additionally, the assessment of the explained variability of the dependent variables (R^2) is significant during the evaluation of the structural model. The advantage of structural modelling compared to other statistical techniques is that it enables the simultaneous examination of mutual dependencies of a series of connections between variables (Hair et al., 2014), which is the need of this research.

4 RESULTS

Statistical data in the period 2015–2019 for 28[1] European countries were considered, resulting in a sample of 140 lines of data used in analysis.

Five variables that make up the research model were used to evaluate the impact of the economic level on the level of the digital economy. SmartPLS software (Ringle et al., 2015) was used to determine the validity of the measurement and structural levels. The convergent validity of the constructs was confirmed by Cronbach's Alpha coefficient with values higher than the recommended value of .70 (Cronbach, 1951) as well as Average Variance Extracted (AVE) values above the recommended value of .50. Construct reliability (CR) is also used to check convergent validity and high values (between 0.60 and 0.70 and above) mean that all

[1] The data were available for 28 countries, comprising the 26 EU member states (Malta's data were missing) plus the UK and Norway. Time spane depended on data availability.

	Cronbach's alpha	Composite reliability (CR)	Average variance extracted (AVE)
Connectivity	0.871	0.911	0.718
Integration of digital technology	0.889	0.919	0.695
Human capital	0.839	0.903	0.757
Digital public services	0.907	0.953	0.909
Economic level	0.785	0.88	0.715

Tab. 2 Convergent validity of the constructs

	Connectivity	Digital public services	Economic level	Human capital	Integration of digital technology
Connectivity	0.848				
Digital public services	0.519	0.954			
Economic level	0.74	0.592	0.845		
Human capital	0.681	0.648	0.854	0.87	
Integration of digital technology	0.659	0.591	0.702	0.806	0.834

Tab. 3 Discriminant validity of the constructs

	Original sample	Sample mean	Standard deviation	T statistics	P values
Economic level -> Connectivity	0.74	0.741	0.032	23.26	0.000
Economic level -> Human capital	0.854	0.856	0.019	45.014	0.000
Economic level -> Integration of digital technology	0.702	0.707	0.05	14.084	0.000
Economic level -> Digital public services	0.592	0.596	0.039	15.344	0.000

Tab. 4 The results of hypothesis testing

indicators consistently represent the same latent constructs (Hair et al., 2014) (Table 2). Also, the discriminant validity of the measurement model is met, given that the correlations between any two constructs are lower than the average variance estimate (AVE) (Table 3). The statistics presented indicate the acceptability of the measurement model, given that unidimensionality, convergent, and discriminant validity are shown.

For testing the structural model, i.e., the hypothesis, a bootstrapping module with 5000 re-samples was used to test the significance of the proposed relationships. Analysing the *t*-test statistics, whose value for all proposed paths is greater than 1.96, it can be concluded that all proposed hypotheses are accepted. The connections between the variables were observed through the β coefficients (values presented on arrows in the model), which means a stronger predictive connection between the variables if the absolute value is higher. High values of β coefficients indicate a strong influence of economic factors on all tested variables, thus

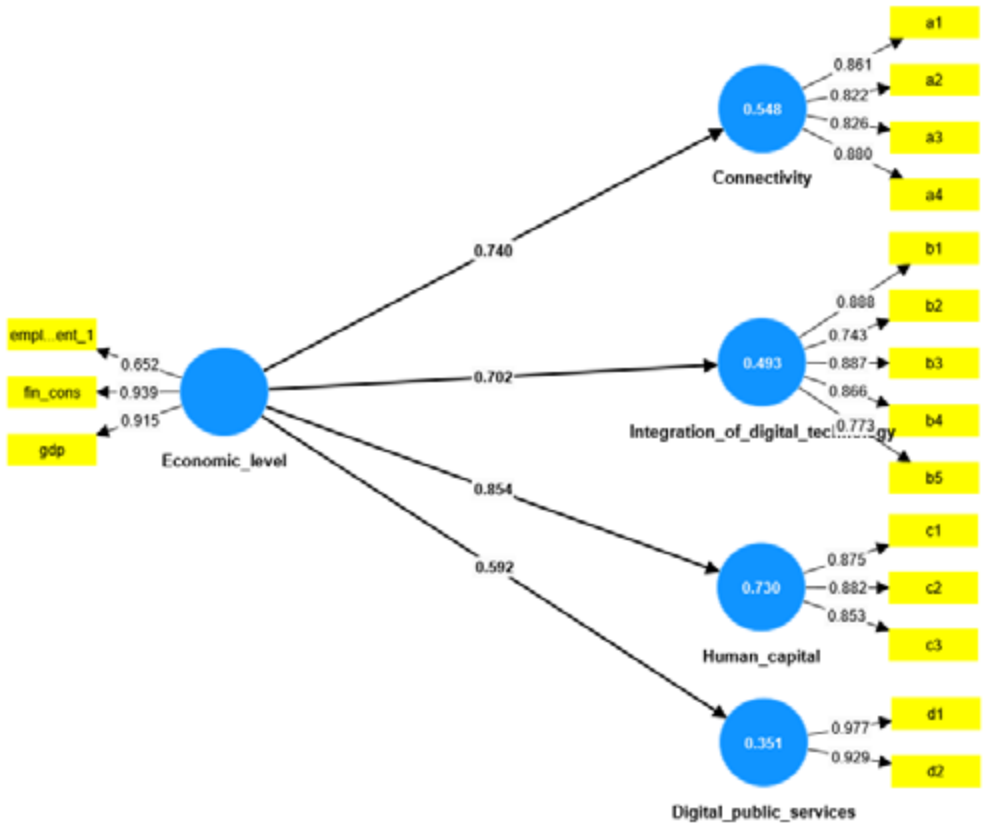


Fig. 2: Summary of the results for measurement and structural model

confirming all four proposed hypotheses. The statistical significance of the obtained results was confirmed with P-values at a significance level of 0.000 (Table 4). The value of the explained variations ranges from 35.1% for the Public services variable to 73% for the Human capital variable, which is considered a significant effect (Ringle et al., 2014) (Figure 2).

5 DISCUSSION AND CONCLUSIONS

By considering the four key points in the process of successful digital transformation, which have been set through the strategy of the European Commission called the 2030 Digital Compass, it is possible to determine the places where digital divides occur and the economic condition of their appearance. The model proposed in the research tests the assumption that the influence of economic factors on the level of digital development is positive and significant. The proposed indicators were selected based on a review of previous research and represent the most significant elements of the DESI index. The quality and validity of the model and the collected data are determined by the statistical criteria used to measure reliability and validity, which are above the recommended thresholds together with high coefficients of determination R^2 for the dependent variables.

Analyzing the results makes it possible to confirm strong positive links between the economic and digital levels. Cruz-Jesus et al. (2017) found that the relationship between digital and economic development is not linear. Thus, by looking at certain digital-level factors, it can be concluded that economic factors strongly influence human capital. Digital transformation

must be based on the ability of the population to be digitally literate and on highly qualified experts, and this is possible through the economic impact on the development of the digital education system and effective programs for the development of digital competencies of the workforce.

Another factor describing the digital level, which is strongly influenced by economics, is the availability of digital infrastructure. The literature suggests a strong link between digitization acceptance patterns and GDP per capita (Mubarak et al., 2020). This research supports the idea that increased income and higher economic levels lead to greater adoption of ICT. In this segment, additional efforts should be made towards increasing investment to deliver the benefits for society by having an accessible and reliable ICT infrastructure.

It has been evidenced that companies can benefit from digital technologies by increasing productivity directly and indirectly, as the effect extends throughout the supply chain and affects its position among competitors. This conditions constant investments in the digital transformation of business and the application of Industry 4.0 technologies, which represent the core of the development of new products, new production processes and the creation of networked chains of new value creation (Dalenogare et al., 2018). In order to realize the intention of creating a Single Market, businesses that lag in the digital transformation should be strongly economically and institutionally supported in achieving these goals.

Providing digital government services is a multi-layered process that requires the coordination of several elements (Twizeyimana and Andersson, 2019). First, an adequate digital infrastructure that enables the functioning of digital services aimed at people at the local and national levels is necessary. Services that are created must be developed in such a way as to enable their use despite certain individual limitations of the user. On the other hand, digital administration and direct channels of communication with the population enable governments to understand the population‘s needs more clearly and to direct additional efforts and funds to the development of programs and procedures to increase the efficiency of e-government and ensure further digital development.

The results of this study should be considered in the context of developing strategies to improve the digitization process in countries whose degree of digitization is at a lower level, as well as the increase of digitization in companies to survive in the global competitive race.

The limitation of the paper is that it does not compare individual countries or groups of countries in order to determine measures of disproportion in the level of digitization. For further research, it‘s crucial to consider how different countries or regions may have varying characteristics that could impact the model‘s results. Conducting subgroup analyses and exploring alternative model specifications should ensure the validity of the model‘s conclusions.

Acknowledgement

The research presented in this paper was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, within the funding of scientific and research work at the University of Belgrade, Technical Faculty in Bor, according to contract with registration number 451-03-65/2024-03 /200131.

REFERENCES

BALCERZAK, P. A. and PIETRZAK, B. M. 2017. Digital Economy in Visegrad Countries. Multiple-criteria Decision Analysis at Regional Level in The Years 2012 and 2015. *Journal of Competitiveness*, 9(2), 5-18.

CRONBACH, L. J. 1951. Coefficient alpha and the internal structure of test. *Psychometrica*, 16: 297-334.

CRUZ-JESUS, F., OLIVEIRA, T., BACAO, F. and IRANI, Z. 2017. Assessing the pattern between economic and digital development of countries. *Information Systems Frontiers*, 19: 835–854.

DALENOGARE, L. S., BENITEZ, G. B., AYALAN. F. and FRANK, A. G. 2018. The expected contribution of Industry 4.0 technologies for industrial performance. *International Journal of Production Economics*, 204: 383-394.

EUROPEAN COMMISSION. 2021a. Communication From the Commission to the European Parliament, the Council, the European Economic and Socialcommittee and the Committee of the Regions. *2030 Digital Compass: the European way for the Digital Decade*. Brussels: COM(2021).

EUROPEAN COMMISSION. 2021b. *Digital Economy and Society Index 2021: overall progress in digital transition but need for new EU-wide efforts*. Brussels.

HABIBI, F. and ZABARDAST, M. 2020. Digitalization, education and economic growth: A comparative analysis of Middle East and OECD countries (101370, Ed.). *Technology in Society*, 63.

HAIR, J. F., SARSTEDT, M., HOPKINS, L. and KUPPELWIESER, V. G. 2014. Partial least squares structural equation modeling (PLS-SEM) An emerging tool in business research. *European Business Review*, 26(2).

HEINZ, J. 2016. Digital Skills and the Influence of Students' Socio-Economic Background. An Exploratory Study in German Elementary Schools. *Italian Journal of Sociology of Education*, 8(2), 186-212.

KWILINSKI, A., VYSHNEVSKYI, O. and DZWIGOL, H. 2020. Digitalization of the EU Economies and People at Risk of Poverty or Social Exclusion. *Journal of Risk and Financial Management*, 13: 142.

KYOBE, M. 2011. Investigating the key factors influencing ICT adoption in South Africa. *Journal of Systems and Information Technology*, 13(3), 255-267.

LI,Y., DAI, J. and CUI, L. 2020. The impact of digital technologies on economic and environmental performance in the context of industry 4.0: A moderated mediation model. *International Journal of Production Economics*, 229: 107777.

LIBONI, L. B., CEZARINO, L. O., JABBOUR, C. J. C., OLIVEIRA, B. G. and STEFANELLI, N. O. 2019. Smart industry and the pathways to HRM 4.0: implications for SCM. *Supply Chain Management*, 24 (1), 124-146.

LUCENDO-MONEDERO, A. L., RUIZ-RODRÍGUEZ, F. and GONZÁLEZ-RELAÑO, R. 2019. Measuring the digital divide at regional level. A spatial analysis of the inequalities in digital development of households and individuals in Europe. *Telematics and Informatics*, 41: 197–217.

MUBARAK, F., SUOMI, R. and KANTOLA, S.-P. 2020. Confirming the links between socio-economic variables and digitalization worldwide: the unsettled debate on digital divide. J*ournal of Information, Communication and Ethics in Society*, 18(3), 415-430.

MYOVELLA, G., KARACUKA, M.,and HAUCAP, J. 2020. Digitalization and economic growth: A comparative analysis of Sub-Saharan Africa and OECD economies. *Telecommunications Policy*, 44, 101856.

OECD 2014. *"General government final consumption", in National Accounts at a Glance 2014*. OECD Publishing, Paris.

PECH, M. and VANĚČEK, D. 2022. Barriers of New Technology Introduction and Disadvantages of Industry 4.0 for Industrial Enterprises. *Serbian Journal of Management*, 17(1), 197–206.

PICATOSTE, J., PÉREZ-ORTIZ, L. and RUESGA-BENITO, S. M. 2018. A new educational pattern in response to new technologies and sustainable development. Enlightening ICT skills for youth employability in the European Union. *Telematics and Informatics*, 35(4), 1031-1038.

POPKOVA, E. G., DE BERNARDI, P., TYURINA, Y. G. and SERGI, B. S. 2022. A theory of digital technology advancement to address the grand challenges of sustainable development. *Technology in Society*, 68: 101831.

RINGLE, C. M., DA SILVA, D. and BIDO, D. 2014. Structural Equation Modeling with the SmartPLS. *Brazilian Journal of Marketing – BJM*, 13(2), 56-73.

RINGLE, C. M., WENDE, S. and BECKER, J.-M. 2015. *SmartPLS 3*. Boenningstedt: SmartPLS GmbH. http://www.smartpls.com

SCHRÖDER, C. 2017. *The Challenges of Industry 4.0 for Small and Medium-sized Enterprises*. FRIEDRICH-EBERT-STIFTUNG, Germany

SEVINÇ, A., GÜR, S. and EREN, T. 2018. Analysis of the Difficulties of SMEs inIndustry 4.0 Applications by Analytical Hierarchy Process and Analytical Network Process. *Processes*, 6(12), 264.

TRAŞCĂ, D.L., ŞTEFAN, G.M., SAHLIAN, D.N., HOINARU, R., and ŞERBAN-OPRESCU, G.-L. 2019. Digitalization and Business Activity. The Struggle to Catch Up in CEE Countries. Sustainability, 11: 2204.

TWIZEYIMANA, J. D. and ANDERSSON, A. 2019. The public value of E-Government – A literature review. *Government Information Quarterly*, 36(2), 167-178.

YIFAN, G. and BEI, L. 2022. Influencing Factors and Multiple Paths of Construction Ability of Digital Government: Qualitative Comparative Analysis Based on 31 Chinese Provinces. *Procedia Computer Science*, 199: 1213-1220.

Contact information

Anđelka Stojanović: e-mail: anstojanovic@tfbor.bg.ac.rs
Isidora Milošević: e-mail: imilosevic@tfbor.bg.ac.rs
Sanela Arsić: e-mail: saarsic@tfbor.bg.ac.rs

MOBILE AUGMENTED REALITY OBJECT DETECTION APPLICATION

Jan Strnad[1], Jaromír Landa[1]

[1]Computer science department, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

This article proposes a Mobile Augmented Reality (MAR) application for object detection. The application can detect predefined objects in the camera stream and display information about them. Object detection poses many challenges, and a common approach is to perform it remotely on a server. However, this requires an active internet connection. Alternatively,detection can be performed locally using a model stored on the device. How- ever, not all devices have the capability to perform real-time detection. We have created a Mobile Augmented Reality app that can detect objects in the camera stream. The app can perform detection locally or remotely, depending on the device's configuration. Sec- ondly, the app's ability to perform detection locally or remotely makes it versatile. The paper has two main contributions. Firstly, the proposed application architecture can be applied to any similar MAR app. The application was tested on multiple Android devices to determine the minimum configuration required for local object detection.

Keywords: object detection, Mobile Augmented Reality, Tensor Flow

JEL Code: L63, L86

1 INTRODUCTION

Machine vision, including artificial intelligence, plays a crucial role in various mobile ap- plications. One of the most common groups of applications is Mobile Augmented Reality (MAR), which allows virtual objects to be visualised in real-world environments. MAR ap- plications are typically operated on users' mobile devices, such as their phones (Zhou and Zhao, 2022). MAR applications utilise object detection to identify real-world objects in a camera stream and the augmented reality principle to visualise information specific to the detected object(s). However, object detection presents several challenges. There are three main approaches to performing object detection with mobile devices: a) on the server, b) on the device, or c) a combination of both (Ghasemi et al., 2022). Ghasemi et al. (2022) published a review that clearly demonstrates the prevalence of processing on a remote server. In this method, the camera image from a device is sent to the server, where object detection is performed. This approach has a clear advantage in that it utilises the compu- ting power of the server rather

https://doi.org/10.11118/978-80-7509-990-7-0159

than relying on the device (Ghasemi et al., 2022). However, it also presents several challenges, with communication latency being the most signifi- cant. Edge computing can partially eliminate this issue by performing detection on a local server connected to the same Wi-Fi network as the device (Liu and Han, 2018). However, as the computing power of mobile devices increases every year, many applications now use local models to perform necessary tasks. Some applications use a combination of server and device, with the device used to find the ROI and the server used to recognise specific objects (Knez and Šajn, 2020).

This paper presents a system for object detection that can be performed locally or remotely on a server. However, it is important to note that object detection is just one aspect of the entire MAR process. The model is first trained and then prepared for local or remote object detection using the TensorFlow framework. Once the object is detected, information about it is visualised using augmented reality. The Android platform is used to test the proposed methods.

The contribution of this work is twofold. First, we have proposed and developed an object detection and AR visualisation system. The system decides whether the device is capable of fast local object detection. If so, the detection is performed locally. Otherwise, the image is sent to a remote processing server and the object location information is sent back to the device. Secondly, we carried out performance tests. We tested object detection on several Android devices and compared the results with remote object detection using two different client-server communication technologies: REST API and Web Socket.

The rest of the paper is organised as follows. In the next chapter, we review recent articles on object detection. The ‚Methods and Materials' section outlines the proposed system and includes the test methodology. The results of the tests are presented and discussed in the ‚Results' section. Finally, concluding remarks are given in the ‚Discussion and Conclusions' section.

2 LITERATURE REVIEW

Object detection is the process of identifying a specific object or class of objects in a com- puter image. There are various algorithms available for object detection, with the most common being Convolutional Neural Networks (CNN). These algorithms can be classified into two main categories: two-stage and one-stage detectors (Martinez-Alpiste et al., 2022). Two-stage detection involves dividing the image into separate parts. The algorithm takes each input part and, after passing through convolution and pooling layers, outputs the object classes (Ghasemi et al., 2022). In contrast, one-stage object detection algo- rithms like SSD (one-stage single shot detectors) or YOLO (You Only Look Once) identify objects with just one pass through the image. The output of all detectors is the location of the object's bounding box in the image (Xiong et al., 2021). Two crucial requirements for object detection are speed and accuracy. Typically, there is a trade-off between the two, where higher speed results in lower accuracy. Object detection can be performed in three ways: on the server, on the client device, or a combination of the two (Ghasemi et al., 2022).

2.1 Server-side object detection

Server-side object detection is the most common type. Images are sent to a remote server where detection takes place. This type of detection requires high accuracy and low end- to-end network latency. However, low latency can significantly reduce accuracy due to changes in the user's view (Liu, Li and Gruteser, 2019).To optimize this process, several techniques can be employed, such as edge computing, federated learning, or software- defined networks (Xiang, Seeling and Fitzek, 2021). Edge computing reduces latency by performing computations closer to the source device on a local network (Liu and Han, 2018). It is important to note that training can also be demanding on device or server performance, not just object recognition.

Zhou and Zhao (2022) propose the use of Fed- erated Learning, which allows each device to train a shared model collaboratively without sharing local data with others.

2.2 Client-side object detection

The primary benefit of client-side object detection is its offline capability and privacy. The model is stored locally on the device where the detection is performed, eliminating the need for a remote server. As the processing is done locally, privacy is also a significant advantage. Personal images are not sent to a remote service (Savchenko, Demochkinb and Grechikhinb, 2022). Several frameworks are used for local object detection, including TensorFlow Mobile (TFM), TensorFlow Lite (TFL), OpenCV, and Qualcomm Snapdragon (Martinez-Alpiste et al., 2022). However, local detection can be energy-intensive (Apicharttrisorn et al., 2019) and demanding on devices with limited computational power (Martinez-Alpiste et al., 2022). There are several ways to optimize the process. Cai et al. (2020) propose a cooperative scheme between the GPU and the CPU for processing.

2.3 Combination of server-side and client-side object detection

Object detection requires a tradeoff between speed and accuracy. To partially solve this issue, Li et al. (2022) proposed a parallel offloading scheme that combines server-side and client-side processing. The mobile device is used to detect large objects and regions of interest (ROIs) containing small objects, while small objects are detected on a remote server. Wang et al. (2022) also employ this principle for autonomous mobile vision. Small object detection is delegated to the edge. Both Li et al. (2022) and Wang et al. (2022) demonstrate significant improvements in accuracy.

3 METHODOLOGY AND DATA

This section outlines the proposed object detection system, which aims to provide fast and accurate object detection on all devices, regardless of their processing power. The system has been designed to detect objects either on a remote server or on a local device. The minimum requirements for Android devices to perform local detection are specified. If a device is not powerful enough to perform local detection, the remote server is used. The second objective is to evaluate the effectiveness of REST API and Web Sockets for real-time object detection. To test the proposed system, both a remote server and an An- droid application were developed. The application captures camera feed, performs object detection either on the server or on the device, and presents information about the de- tected objects in augmented reality using the SceneForm framework.

3.1 Remote server-based object detection

The goal of the server-based approach is to offload the performance-intensive object de- tection from the device to a remote server. The device captures an image from the camera and sends it to the remote server. The server receives the image, processes it and per- forms the detection. Objects are detected using the TensorFlow library. This library re- quires trained models in the tflite format. This format was chosen for its known compati- bility with mobile devices and can also be used for server-side detection. The tflite format is used by many projects, such as Chilukuri, Yi and Seong (2022) and Azzo, Taqi and Mila- nova (2018).

Once the detection process is complete, the server sends a list of detected objects (label, confidence and bounding box) back to the device in JSON format. The device pro- cesses the list and then renders the detected objects in augmented reality directly into the camera stream.

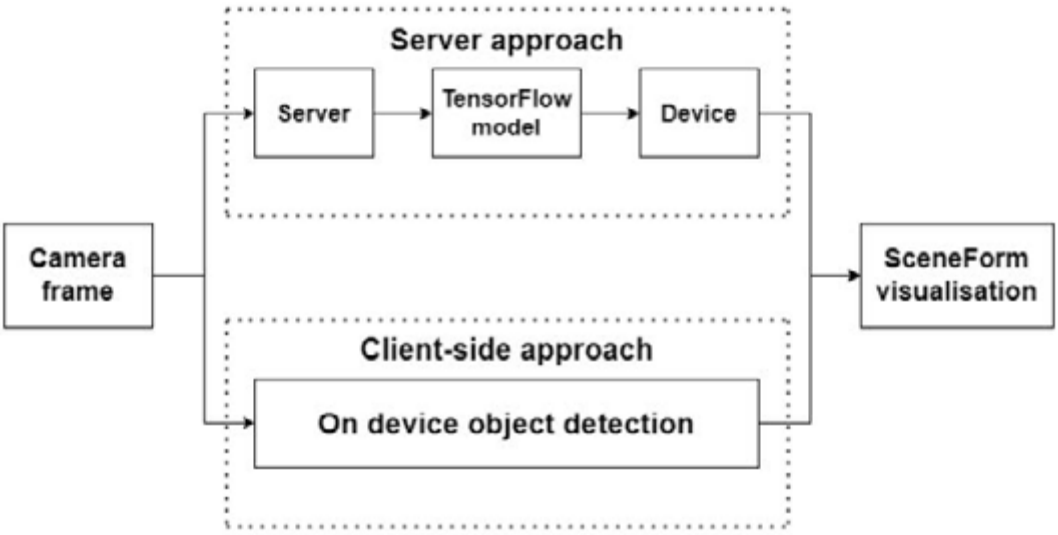


Fig. 1: Object detection process

This process illustrates the drawback of server-based detection. The image from the camera has to be sent to the server, which can take a long time. The two most common communication schemes today are REST API and WebSockets. Both REST API and Web- Sockets have been implemented and tested.

3.2 Client-side object recognition

This approach removes a major disadvantage of the server approach, which is the need for a permanent Internet connection. The whole process can be done completely offline. Google‘s ML Kit library was used for detection. The ML Kit uses the same tflite model for- mat as the previous server-based approach. In the case of client-side detection, an image from the camera is passed directly to the detector, which returns a list of the resulting detected objects.

4 RESULTS

To test the functionality and performance of each approach, we conducted tests on both. In terms of server-side object detection, we compared the time it took to communicate using REST API and Web Sockets. Table 1 clearly shows that Web Sockets have a signifi- cant advantage over REST API. Regarding the REST API, the time it takes for requests and responses to travel between the server and device is slower due to the need to send mul- tiple images per second to the server and receive the resulting objects in a timely manner. The detection speed on the server side remains the same when using the same model and detection method to identify objects.

To compare client-side object detection, we compared seven different mobile phones with varying configurations (refer to Table 2). The tested mobile phone devices vary with different configurations and Android version to test the detection on commonly used de- vices. As this is an offline process, we did not consider the network load of the phones. We established several metrics for comparison, as described above. The results in Table 2 show that the RAM parameter is one of the main attributes affecting the resulting detec- tion time. Moreover, newer processors exhibit significantly improved performance in ob- ject detection, as confirmed by the CPU usage metric. Moreover, newer processors exhibit significantly improved performance

Type	Request	Detection time	Response	Totaltime
RESTAPI	228.451 ms	99.075ms	206.258 ms	533.784 ms
Websocket	34.364ms	99.075ms	25.397ms	158.836ms

Tab. 1 The results of server-side object detection

in object detection, as confirmed by the CPU usage metric. Moreover, newer processors exhibit significantly improved performance in object detection, as confirmed by the CPU usage metric. The more advanced processors are no- ticeably less burdened.

The findings indicate that newer and more powerful devices have a distinct ad- vantage in client-based object detection, as it saves time and eliminates the need for a constant internet connection. However, for older phones, detection time is longer com- pared to the server approach that uses web sockets. In such cases, it is recommended to perform object detection on the server. Real-time use can omit detection using REST API communication.

As previously stated, our system adjusts to current conditions. Through testing, we have determined the necessary requirements for a mobile device to perform local object recognition: a) a minimum of 6 GB of RAM, b) new octa-core processors, and c) Android version 10 or higher. The results show the advantage of using newer mobile phones over older models in both hardware configuration and available Android OS version.

Refer to Fig. 2 for the final visualization of the detection.

Device name	CPU	RAM	Detection time	RAM usage	Android version	CPU usage
Samsung S21 FE	Snap- dragon 888	6 GB	14.37 ms	444 MB	13	29%
Pixel 5	Snap- dragon 765G	8 GB	71.31 ms	600 MB	13	46%
Samsung S8	Snap- dragon 835	4 GB	160 ms	500 MB	9	49%
Pixel XL	Snap-dragon 821	4 GB	213 ms	700 MB	10	44%
Xiaomi Mi A3	Snap-dragon 665	4 GB	236 ms	600 MB	11	52%
OnePlus 7 Pro	Snap-dragon 855	8 GB	25 ms	437 MB	11	25%
Pixel 6	Google tensor	8 GB	24.7 ms	600 MB	12	38%

Tab. 2 Object detection process

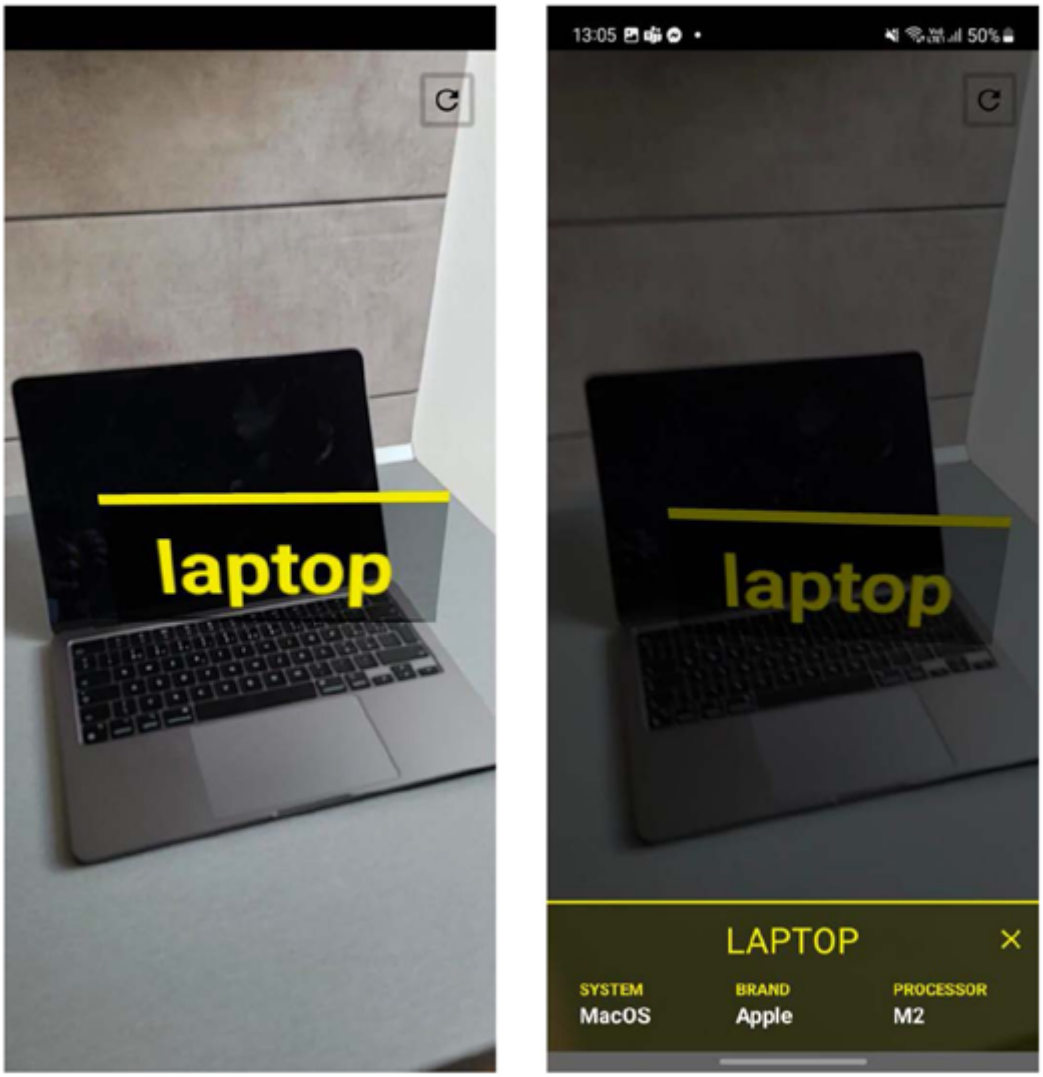


Fig. 2: Visualization of the detection process using Scene Form framework.

5 DISCUSSION AND CONCLUSIONS

This paper proposes a Mobile Augmented Reality application for object detection. Object detection can be performed either server-side or client-side, each with its own advantages and disadvantages. Server-side detection offers the advantage of a central processing point without the need to download the trained model to the device, making model up- dates more efficient. However, it has the disadvantage of slower processing time. Even when using web sockets, the time required for detection is much longer than that of local, powerful, and new devices. The use of standard REST API communication is only useful for applications that do not require real-time object detection. Over time, communication using Web Sockets is much faster and more efficient due to its parallel capabilities. The Web Socket allows for simultaneous request reception and response transmission. This means that when the server sends detected objects, it can already be receiving the next image to perform a new detection. In contrast, with the classic REST API, communication is only possible in one direction, which leads to increased detection and response time over time.

Based on the tests, client-side object detection was better in most cases. However, this was not the case with older devices. However, client-side detection can lead to increased memory and CPU usage due to the number of detected objects stored. As mentioned in Ghasemi et al. (2022) there is a trade-off between speed and accuracy of object detection. In our case, the accuracy did not change, since the object were always detected, however the speed of the detection can vary significantly based on the device configuration. Our results prove that with newer devices, this trade-off becomes less significant as the speed of the detection significantly increases. The results show the clear advantage of client-side detection over server-side.

Therefore, we have set minimal requirements for Android mobile devices to perform object recognition. If these requirements are not met, it is recommended to perform ob- ject detection server-side. The minimal requirements are a) at least 6 GB of RAM, b) new octa-core processors, and c) Android version 10 or higher. If the requirements are not met, the system will switch from local detection to remote server-side detection.

Acknowledgements

Supported by grant No. IGA-PEF-TP-22-006 (Opportunities to use metaverse technology to support business processes) of the Internal Grant Agency FBE MENDELU.

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

APICHARTTRISORN, K., RAN, X., CHEN, J., KRISHNAMURTHY, S. V. and ROY-CHOW-DHURY, A. K. 2019. Frugal following: power thrifty object detection and tracking for mobile augmented reality. In: *Proceedings of the 17th Conference on Embedded Net- worked Sensor Systems (SenSys ,19)*. Association for Computing Machinery, New York, NY, USA, 96–109. https://doi.org/10.1145/3356250.3360044

AZZO, F., TAQI, A. M. and MILANOVA, M. 2018. Human Related-Health Actions Detection using Android Camera based on TensorFlow yObject Detection API. *International Journal of Advanced Computer Science and Applications,* 9(10). https://doi.org/10.14569/IJACSA.2018.091002

CAI, Y., Li, H., YUAN, G., NIU, W., Li, Y., TANG, X., REN, B. and WANG, Y. 2020. YOLObile: Real-Time Object Detection on Mobile Devices via Compression-Compilation CoDesign. *arXiv*. https://doi.org/10.48550/arXiv.2009.05697

GHASEMI, Y., JEONG, H. CHOI, S. H., PARK, K. and LEE, J. Y. 2022. Deep learning-based object

detection in augmented reality: A systematic review. *Computers in Industry*, 139, 103661. ISSN 0166-3615. https://doi.org/10.1016/j.compind.2022.103661

CHILUKURI, D. M., YI, S. and SEONG, Y. A robust object detection system with occlusion han- dling for mobile devices. *Computational Intelligence*, 38(4): 1338-1364. https://doi.org/10.1111/coin.12511

LI, X., QIN, Y., LIU, Z., ZOMAYA, A. and LIAO, X. 2022. Towards efficient and robust intelligent mobile vision system via small object aware parallel offloading. *Journal of Systems Architecture*, 129, 102595. ISSN 1383-7621. https://doi.org/10.1016/j.sysarc.2022.102595

LIU, Q. and HAN, T. 2018. DARE: Dynamic Adaptive Mobile Augmented Reality with Edge Computing. In: *IEEE 26th International Conference on Network Protocols (ICNP)*. Cambridge, UK, 2018, pp. 1-11. https://doi.org/10.1109/ICNP.2018.00011

LIU, L., LI, H. and GRUTESER, M. 2019. Edge Assisted Real-time Object Detection for Mo- bile Augmented Reality. In: *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom ,19)*. Association for Computing Machinery, New York, NY, USA, Article 25, 1–16. https://doi.org/10.1145/3300061.3300116

KNEZ, S. and ŠAJN, L. 2020. Food object recognition using a mobile device: Evaluation of currently implemented systems. *Trends in Food Science & Technology*, 99, 460-471. ISSN 0924-2244. http://doi.org/10.1016/j.tifs.2020.03.017

MARTINEZ-ALPISTE, I. et al. 2022. Smartphone-based real-time object recognition archi- tecture for portable and constrained systems. *J Real-Time Image Proc.,* 19. https://doi.org/10.1007/s11554-021-01164-1

SAVCHENKO, A. V., DEMOCHKIN, K. V. and GRECHIKHIN, I. S. 2022. Preference prediction based on a photo gallery analysis with scene recognition and object detection. *Pattern Recognition*, 121, 108248. ISSN 0031-3203. https://doi.org/10.1016/j.patcog.2021.108248

THIEN, H., PHAM, Q., PHAM, X., NGUYEN, T. T., HAN, Z. and KIM, D. 2023. Artificial intelligence for the metaverse: A survey. *Engineering Applications of Artificial Intelligence*, 117(Part A), 105581. ISSN 0952-1976. https://doi.org/10.1016/j.en- gappai.2022.105581

WANG, X., YANG, Z., WU, J., ZHAO, Y. and ZHOU, Z. 2021. EdgeDuet: Tiling Small Object Detection for Edge-Assisted Autonomous Mobile Vision. In: *IEEE INFOCOM 2021 – IEEE Conference on Computer Communications*. Vancouver, BC, Canada, pp. 1-10. https://doi.org/10.1109/INFOCOM42981.2021.9488843

XIONG, Y. et al. 2021. MobileDets: Searching for Object Detection Architectures for Mo- bile Accelerators. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3824-3833. https://doi.org/10.1109/CVPR46437.2021.00382

ZHOU, X. and ZHAO, J. 2022. Mobile Augmented Reality with Federated Learning in the Metaverse. 2022. *arXiv*. https://doi.org/10.48550/arXiv.2212.08324

Contact information

Jan Strnad: e-mail: strnad.hon@gmail.com
Jaromír Landa: e-mail: jaromir.landa@mendelu.cz

COMPARATIVE ANALYSIS OF SELECTED TIME SERIES FORECASTING APPROACHES FOR INDIAN MARKETS

Ankit Tripathi[1], Arpit Tripathi[1], Oldřich Trenz[1], Pawan Kumar Mishra[1]

[1]Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

Financial market analysis and prediction have been topics of interest to traders and investors for decades. This study assesses the performance of selected time series prediction methods like deep learning algorithms (Long short-term memory model (LSTM)), traditional statistical models (Seasonal Auto Regressive Integrated Moving Approach with eXogenous regressors (SARIMAX)), and advanced ensemble learning algorithms (XGBoost and FB-Prophet) using real-world data from the Indian financial market. The stock prices of Reliance Company serve as a case study, enabling a thorough evaluation of predictive accuracy and errors of the models. A pre-processing approach has been proposed and implemented, integrating significant economic factors (Gold Price, USD to INR conversion, Consumer Price Index (CPI), Wholesale Price Index (WPI) and Indian 10-year yield bond) and evaluated with technical metrics (Mean squared error, Mean Absolute Error and R2 Score). The study investigates how the inclusion of these factors impacts prediction accuracy across the selected time series prediction methods. The comparative evaluation of models before and after the pre-processing method sheds light on the evolving predictive accuracy of LSTM, SARIMAX, FB-Prophet, and XGBoost. The study showed that the SARIMAX (extension of ARIMA with seasonality and exogenous factors) and XGBOOST performed relatively well with the proposed approach while LSTM and FB prophet (though advanced) did not perform as expected in Indian financial markets. This research contributes to advancing the understanding of time series forecasting in the financial market of India, offering practical insights for decision-makers and researchers.

Keywords: Financial Time Series, Stock Market Prediction, Deep Learning in Finance, Ensemble Learning in Economics, ARIMA and XGBoost Analysis

JEL Classification: C22, C53, G17

https://doi.org/10.11118/978-80-7509-990-7-0167

1 INTRODUCTION

Predicting the trends of financial markets has always been a challenge for people from various fields such as investors, academic researchers, financial analysts, traders, and brokers. Despite professionals using different strategies for short-term investment, long-term investment, or sheer academic pursuit, forecasting results play a vital role. In the studies of Obthong et al. in 2020 and Sonkavde et al. in 2023, financial markets comprised different segments such as bond markets, derivates markets, stock markets and commodity markets (Obthong et al., 2020; Sonkavde et al., 2023). A wide variety of techniques have been used in the past by traders such as swing trading, event-driven trading, momentum trading, trend-following trading, scalping, position trading and state-of-the-art algorithmic trading. However, all these trading methods have inherent limitations due to the nature of stocks. The inherent non-parametric, nonlinear, and non-stationary nature of stock prices, coupled with the noise in time series data, adds complexity to the prediction task (Abu-Mostafa and Atiya, 1996; Farid et al., 2023). Furthermore, this coupled with the influences from the political and economic environment around the stock market gives it further complexity to forecast the price. As traditional statistical methods, including fundamental and technical analyses, grapple with limitations such as lagging indicators and prediction inaccuracies, the spotlight has shifted to cutting-edge machine learning and deep learning models in algorithmic trading. These artificial intelligence models based on historical observations of variable forms a relationship which is further used to extrapolate the data as its future predictions. The calculations and estimations previously done via statistics are being done by methods of machine learning such as the Long Short Term Memory method (LSTM) (Chen et al., 2015), Auto Regressive integrated Moving Approach (ARIMA) (Khashei and Bijari, 2011), SARIMA (Seasonal Auto-Regressive integrated Moving Approach) (Lee et al., 2008), Holt-Winters Exponential Smoothing (Dassanayake et al., 2021), FB-Prophet (Sharma et al., 2022) introduced by Facebook and XGBoost (Gumelar et al., 2020). While some of these methods have been around since the 1970s, others have been developed in the past decade with many hybrid models having the foundation of these basic models.

It is also important to know that majority of models known works on Artificial neural networks (ANN) as it is one of the most important type of non-parametric, non-linear time series model which have been used for time series forecasting (Khashei and Bijari, 2011). While the implementation of these models had been examined on various markets such as Hong Kong Stock Markets, Shanghai Stock Index, A-share market stocks in China, S&P 500 Future, Shenzhen Component Index (SZI) and Chinese Securities Index (CSI) but a relatively smaller number of research has been done on Indian markets (Liu et al., 2023, 2024; Nasiri and Ebadzadeh, 2023). The studies by authors Sapre, Gori and Seah in 2023 quotes that "India is the fifth largest economy in the world and is 3rd in purchasing power parity according to International Monetary fund (IMF)" (Sapre and Gori, 2023; Seah, 2022). This can also be interpreted as there is an immense potential in the Indian market that investors, traders, and brokers can take advantage of and multiply their investment several times. But to get maximum returns, it is necessary to have more data regarding the performance of various algorithmic models in Indian markets, the factors influencing the markets and its segments, the relation between these factors and performance of models and results of integration of different factors in different segments of markets. To gain insights into the methods, approaches or techniques that have been used or are currently being developed for stock price forecasting especially in the context of Indian market, studies such as Mehtab et al. in 2021 which focuses on utilizing LSTM-based models for stock price prediction, Singh and Borah in 2014, concentrating on forecasting State Bank of India (SBI) stock index prices through particle swarm optimization and fuzzy time series (FTS-PSO), and Mehtab and Sen in 2021, centred on predictions using Convolutional Neural Networks (CNN) and LSTMs, provide valuable insights in understanding the background and current trends in literature (Mehtab et al., 2021; Singh and Borah, 2014). The findings show that majority of publications have focused their research

on development of new hybrid models and their application in different domains. In the domain of financial markets, especially stock price forecasting, majority of the studies have also focused in establishing the relation between macroeconomic variables and stock prices but have not focused on the change in performance it brings to these forecasting models (Dhingra and Kapil, 2021; Subburayan et al., 2021). The methodologies are also majorly focused on the implementation of a single approach than the comparative evaluation of multiple models.

Another relevant thing to note is that the time span, nature of economy and sector (macroeconomic variables and industry type) affects the nature of data used for algorithm training. Therefore, to the best of author's knowledge, this study, provides the missing knowledge for Indian markets and reliance industry for the decade of 2nd December 2013 to 30th November 2023. This studies focuses on comparative analysis of selected models (LSTM, SARIMAX, XGBoost and FB-prophet) AND introducing the preprocessing approach of integration of macroeconomic variables and assessing the impact on selected models AND the implementation of data on Reliance Industry Ltd. (2nd December 2013 to 30th November 2023) which is a multinational company holding stocks for different businesses. This combination gives the study a unique perspective and offers the insights for newer age models to be built in Indian markets or segments.

2 CURRENT STATE

Investing in stock markets has been one of the approaches that has been employed by various traders, investors, businesses and dealers over the past decades to earn a significant sum of money over a period. In the past, fundamental techniques like candle charts analysis, analysing earning reports, dividend yield, quantitative analysis using statistics or mathematics model were used to forecast the prices and get better investment benefit ratio. After the implementation of Regulation National Market System in 2005, people started to focus on machine learning models due to its advantages which is speed, anonymity and certainty of execution with predefined criteria and strategies which can also be seen from the rise in number of publications in 2005 till date (Litzenberger et al., 2012). This has come today in form of various methods being developed to predict the stock prices considering all the possibilities.

The implementation of basics models like LSTM, ARIMA, GRU and others to various stock markets have provided the foundation to newer models. This has provided the data required in terms of how the market shifts when facing a crisis, or when a government change, or when a new economic policy is implemented. This data is the very precursor with which researchers try to compare the newly hybrid models. When looking with a cursory point of view, most of the models have been deployed or tested on American markets, Chinese markets, or European markets but not many models are tested in Indian markets (Liu et al., 2023, 2024; Nasiri and Ebadzadeh, 2023). India being the world fifth largest economy in terms of GDP in 2023 still does not have the amount of data, there should have been. When searched with Scopus with the query "time series forecasting in India" refined with the keyword "financial market" only 137 documents come to the view with being the study on 'Bombay Stock Exchange', 'CNX Nifty' and 'S and P' as oldest using vanilla GARCH model (Karmakar, 2005). The number of search results can easily be interpreted as 'low' compared to the larger economies in the world. This study focuses on filling that literature gap and provide basic supporting evidence that will encourage the state of art hybrid model to be tested and deployed in Indian stock markets.

The study has been focused on implementation of machine learning models mainly LSTM, ARIMA, FB Prophet and XG-Boost on Indian markets. The study by Kumaria et al. in 2023, that have focuses on the LSTM to analyse how the neural network anticipates to stock's price (Kumaria et al., 2023). In the study by Chatterjee et al. in 2022, GARCH based hybrid models are employed to analyse the impact of these models on banking, IT and pharma sector

(Chatterjee et al., 2022). In the study by Srivastava et al. in 2023. The author used NIFTY 50 dataset along with financial and social indicators to predict the stock price with LSTM, SVM, KNN, Random Forest and gradient boosting regressor models (Srivastava et al., 2023). Another study by J.P.S.Kumar et al. in 2023 compares the ARIMA and RNN-LSTM on Sensex and nifty dataset (J.P.S. Kumar et al., 2023). The study by Sharma et al. in 2022, employed the Fb-prophet model in comparison to ARIMA on NSE and BSE dataset (Sharma et al., 2022). Most of the studies have utilised either one or two basic models but not four including FB prophet and comparing them to evaluate the results. The evaluation metrics (MSE, MAE and R2 score) has also been same to evaluate the efficiency of models throughout various studies, making them the priority in this study. Another differentiating factor is that most of these studies have employed the models on basic features of any stock price data (Open, high, low, close, adj. close and date) but have not included the major economic factors or variables.

The studies by Baranidharan et al. in 2021, Dhingra and Kapil in 2021 and Hussain et al. in 2012 have supported the relationship between stock prices exchanges and macroeconomic variables but have not implemented them to the selected models and have not performed a comparative analysis (Dhingra and Kapil, 2021; Hussain et al., 2012; Subburayan et al., 2021). The study by Agarwal et al. in 2019, have utilised the technical indicators to support the hypothesis using O-LSTM model only (Agrawal et al., 2019).

The factor that these economic factors have been utilised with the integration of four different models make our study distinguished.

2.1 Selection of Macro-economic factors

Economic indicators, macroeconomic variable or interest rates play a significant role in influencing the financial markets especially stock prices. The economic factors chosen for integration are gold prices (Gold futures historical prices, 2023), Indian Rupees to United States Dollar (USD INR historical data, 2023), Indian 10-Year Bond yield price (India 10-year Bond Historical Data, 2023), Wholesale Price Index (India wholesale price index (WPI), 2023) and Consumer Price Index (India consumer price index (CPI), 2023).

Over the years, a lot of studies have indicated the relationship between gold prices and stock prices. In the study by Smith in 2001, the authors have provided the empirical evidence between gold price and stock prices by using four gold prices and six stock prices for the United States indices over the period 1991 to 2001 (Smith, 2001). The correlation was also unveiled in India by Bhunia in 2012, in which they which took stock prices of the National Stock Exchange (NIFTY) and the gold prices over April 2001 to March 2011 and by Patel in 2013, in which they took the Gold prices, Sensex , BSE 100 and Nifty for January 1991 to December, 2011 and performed ADF Test (Dickey and Fuller, 1981), Johansen's cointegration test (Johansen, 1995) and Granger Causality test (Granger, 1969) to find the underlying pattern in the relationship (Bhunia and Das, 2012; Patel, 2013). Thus, we have used this indicator to provide more information about the market to the models and to increase their efficiency.

The exchange rate has also been chosen to be one of the indicators as the various literatures have focused on the forex returns and stock prices. In the study by Batra et al. in 2020, the author concluded a negative relationship between the variables through Granger Causality test by using the data for 20 years i.e. from 25th January 2000 to 25th January 2020 (Batra et al., 2020). In the study by Lakshmanasamy in 2021, the author concluded that there is a positive relationship between BSE SENSEX return and Euro/Rupee and there is a negative relationship between USD/INR, GBP/INR and the stock prices (Lakshmanasamy, 2021). Also, the prices of Gold in the dataset are in USD and after the incorporation of the price conversion, the analysis of result became much simpler.

The Indian 10-Year Bond yield price is a relatively new factor is incorporated in the study to see the impact in the stock prices forecasting results. In the study by Panigrahi in 2022, the authors have quoted that it is an economic indicator which acts as a benchmark in equity

market too (Panigrahi, 2022). They have also focused to determine a long-term relationship between stock market and long-term interest rate (Indian 10-year bond yield) using the Johansen's cointegration test and vector error correction mechanism. In the study 'Identifying significant macroeconomic Indicators for Indian Stock Markets' by Aithal et al. in 2019, the authors have taken 44 indicators among which the variable 'government bond 10-year yield' is an indicator and they have used this with the features of NSE Nifty and BSE SENSEX indices and have performed a regression analysis (Aithal et al., 2019). Thus, the factor becomes highly important to assess the impact of economic variables on stock market.

The other factors are WPI (Wholesale Price Index) and Consumer Price Index (CPI) which have been added to the economic variables list. The Wholesale Price Index usually represents the change in the price of goods sold by wholesale across India and consumer price index represents the change in price of goods or services from the point of view of consumers. They both are crucial factors impacting the interest rate or currency rates and other macroeconomic variables which in turn impact the stock prices. Both of these indicators are used to measure inflation rate in India. In the studies by Hussain et al. in 2012 and Raheem Ahmed et al. in 2017, they have tried to assess the change in KSE 100 index through the variables that includes CPI and WPI and concluded that CPI and returns have a long-term relationship (Hussain et al., 2012; Raheem Ahmed et al., 2017).

3 METHODS

3.1 Selecting the models

As the domain of time series forecasting is significantly vast, various models have been used already for the implementation of stock price forecasting with newer models being developed every day. Thus, choosing any four models will be followed by the question why was that model selected. The bibliometrics offers a solution for this problem with the help of various analyses through quantifiable metrics as depicted in Figure 1. This helps in understanding

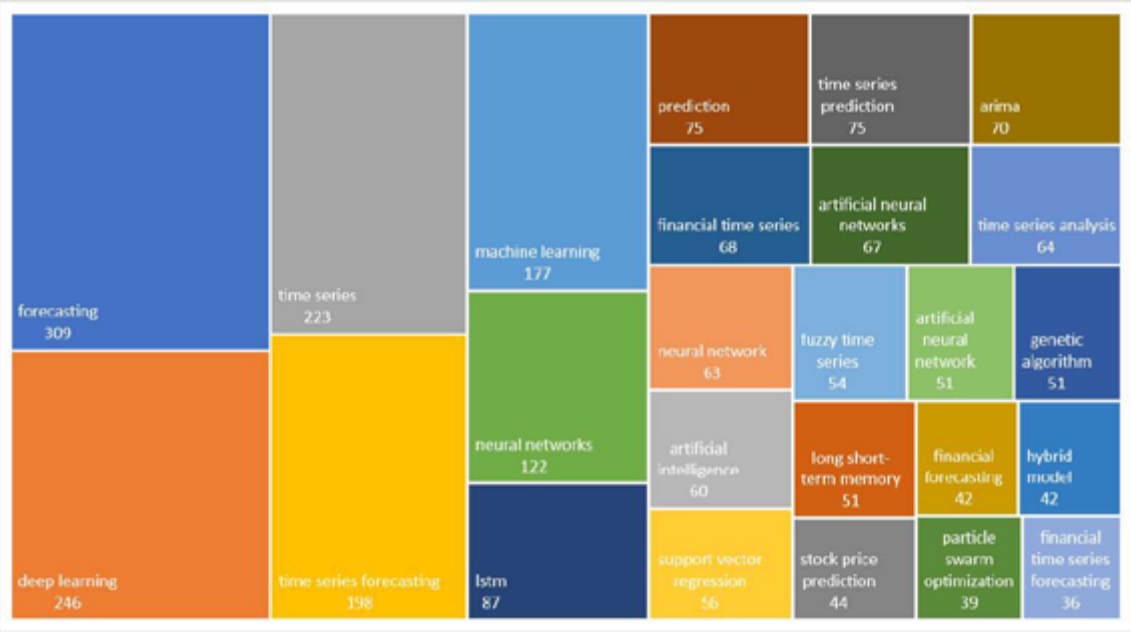


Fig. 1: Tree map of most frequent keywords.
Source: compiled by the author

that with respect to time series forecasting, financial markets and artificial intelligence, the most used models or approaches are LSTM and ARIMA.

Based on it, some models have been selected among which, two are traditional or foundational that have been used the highest number of times or have highest number of occurrence and two which are relatively state of art models. The selected state of art model is ensemble learning models like XG-Boost that also had a relation with time series forecasting. Another model like FB prophet was selected based on literatures. This will also be used to fill the literature gap in the domain and to understand the newer age model that have been classified as hybrid models. Thus, the models selected for the comparative analysis were LSTM, SARIMAX, FB-Prophet, XG-Boost.

3.2 Database and Search Strategy

While analyzing multiple time series could provide more comprehensive findings by encompassing a broader spectrum of the market, this study focuses on Reliance Industries Limited due to its prominence and influence in the Indian stock market. Reliance Industries is not only one of the largest companies in India by market capitalization but also plays a crucial role in various sectors, including energy, petrochemicals, textiles, natural resources, retail, and telecommunications. Its diverse business operations and significant market presence make it an ideal candidate for understanding broader market trends and economic influences (Biswas, 2018). By focusing on a key player like Reliance Industries, insights that are likely reflective of larger market behaviors and economic conditions can be gained, while maintaining a manageable scope for detailed analysis. The database chosen to extract the information about the stock prices of Reliance Industries Limited was Yahoo Finance (Reliance Industries Limited,2023). The domain offers various features including the historical data which have been used from December 02, 2013, to November 30, 2023. This data has been pre-processed and extracted in an CSV file with the daily frequency and as shown in Table 1.

3.3 Data Cleaning and Preprocessing

The necessary libraries were imported for the implementation of selected models. Standard Scalar has been to scale the values and to make model more efficient. Sequential, Adam (optimiser) and tensor flow keras have been imported to build and train the LSTM model. For Fb-Prophet, 'prophet' which is facebook's time series forecasting library for time series forecasting. For XGBoost, 'XGBoost' module which is used for creating and training the model. The

Date	Open	High	Low	Close	Adj Close	Volume	Gold Price	USD INR conversion	Indian 10 year bond price	CPI	WPI
02-12-2013	422.4952	427.4235	422.1485	423.6344	397.6064	3332837	1221.9	62.325	9.048	0.1124	7.52%
03-12-2013	423.4858	428.6865	421.3312	426.9034	400.6745	4238703	1220.8	62.37	9.069	0.1124	7.52%
04-12-2013	426.4081	429.6771	422.5942	424.0059	397.955	4043235	1247.2	62.06	9.09	0.1124	7.52%
05-12-2013	431.906	434.3825	429.9495	431.4106	404.9049	5784625	1231.9	61.763	9.108	0.1124	7.52%
06-12-2013	430.9897	433.2185	426.5567	429.2809	402.9059	3369441	1229	61.435	9.165	0.1124	7.52%

Tab. 1 Dataset merged with economical indicators
Source: compiled by the author

Fig. 2: Plotting thes tock prices
Source: compiled by the author

data that have been extracted in the 'csv' file was imported in the google colaboratory to work. *Adjusted Close price* was chosen with work with as it reflects the price of stock after inclusion of other factors. Some basics steps of data preprocessing were implemented to prepare the data for building the models and to make it more efficient such as:

1. Conversion of date to datetime format and setting 'Date' as index.
2. Dropping the rows with missing values.
3. Dropping the null values and dealing with NaN values.
4. The variables with high correlation coefficient were selected.
5. In case of Fb-prophet, Renaming the columns like date: 'ds' and Adj Close as 'y' to be compatible with Prophet model

After the merging of macro-economic variables:

1. As the WPI had '%' after values, so removing the percentage sign form WPI column and converting it to float.
2. Extracting the features and target and normalizing the features using Standard Scalar.

The selected evaluation metrics were Mean squared error, Mean absolute error and R2 score.

3.4 Implementation of machine learning algorithms

After the basic steps of data cleaning and preprocessing, the visualisation of stock prices was plotted with '*Date on x-axis*' and '*Adjusted close price on y-axis*' as depicted in Figure 2.

The dataset was split into two parts i.e., features (independent variables) and target (dependent variable). A standard scaler was applied to normalize the features by removing the mean and scaling to unit variance, ensuring that all features contribute equally to the model. For the models (LSTM, SARIMAX, XG-Boost and FB-prophet), training-testing splits were employed. 80% of the data was allocated for training, with 20% reserved for testing, following a standard 80:20 ratio. This approach was kept common for all the selected models to make comparative analysis general and efficient.

For LSTM, the sequential model was built using neural networks with input as features.shape (Obthong at al., 2020) which had 5 features. An input layer was provided with the number of input features. Two dense layers with 64 and 32 neurons were provided each followed by a dropout layer with a dropout rate of 20% to prevent overfitting. The activation function used was Rectified Linear Unit (ReLU) due to its ability to mitigate the vanishing gradient problem, its computational efficiency, and its tendency to converge faster compared to other activation functions like sigmoid or tanh. The output layer had a single neuron to predict the target variable. The model was compiled with the 'Mean Squared Error (MSE)' loss function, the 'Adam' optimizer (learning rate of 0.001), and 'Mean Absolute Error (MAE)' as the evaluation metric. The model was trained for 100 epochs. The forecasting plot is shown in Figure 3a). The study have also used an alternate extension of ARIMA i.e., SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous Factors). This type of model can take other variables into account while predicting the stock prices.

The best SARIMAX order was found to be (0,1,2) and no seasonal order was chosen. The autocorrelation and partial autocorrelation function (ACF and PACF) graphs were also plotted to determine order in the data. Augmented dickey-fuller test was also used to ensure staionarity in the data and to determine appropriate diffrencing order‘d’ but as the datasize was small

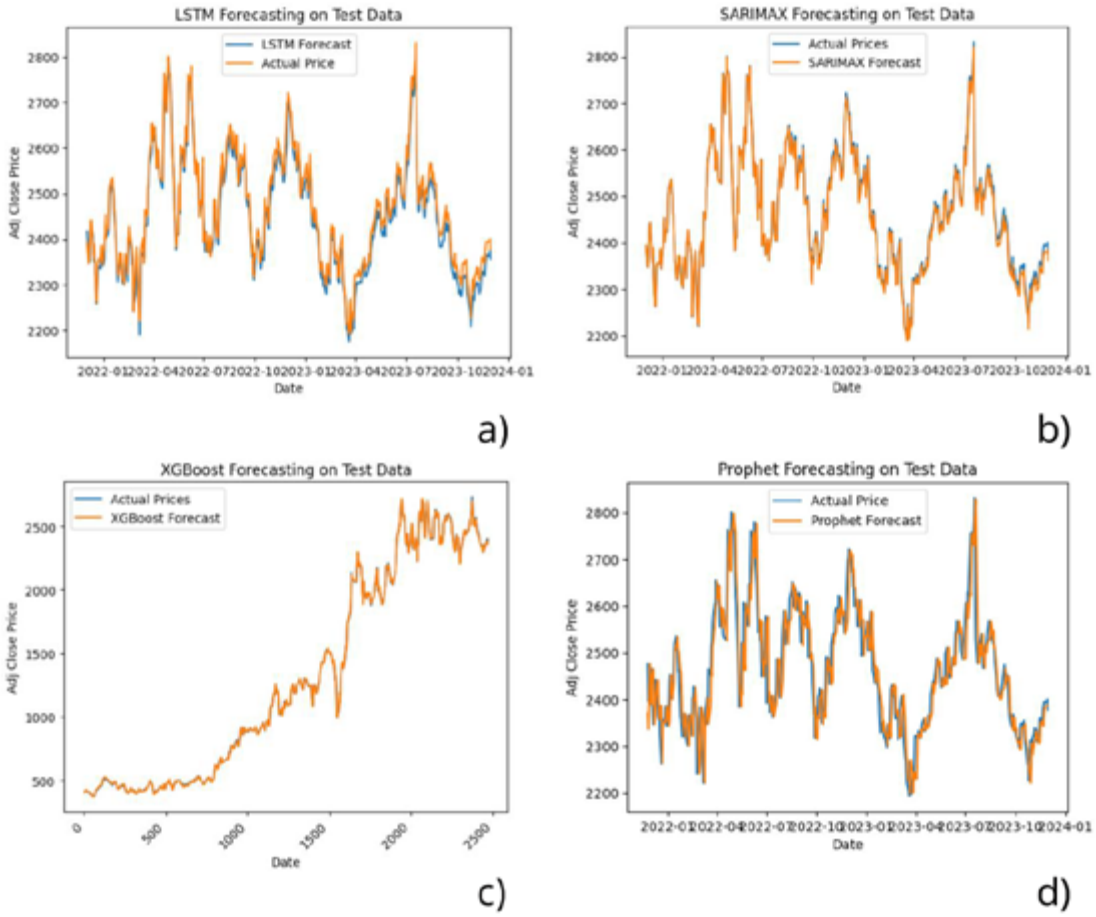


Fig. 3: a) Forecasting results by LSTM. b) Forecasting plot for SARIMAX. c) Forecasting plot for XGBoost. d) Forecasting plot for FB prophet.

Source: compiled by the author using (Google Colaboratory- Fb prophet before integration with economic variables, 2023.; Google Colaboratory- LSTM before integration with economic variables, 2023.; Google Colaboratory- SARIMAX before integration with economic variables, 2023.; Google Colaboratory- XGBoost before integration with economic variables, 2023.

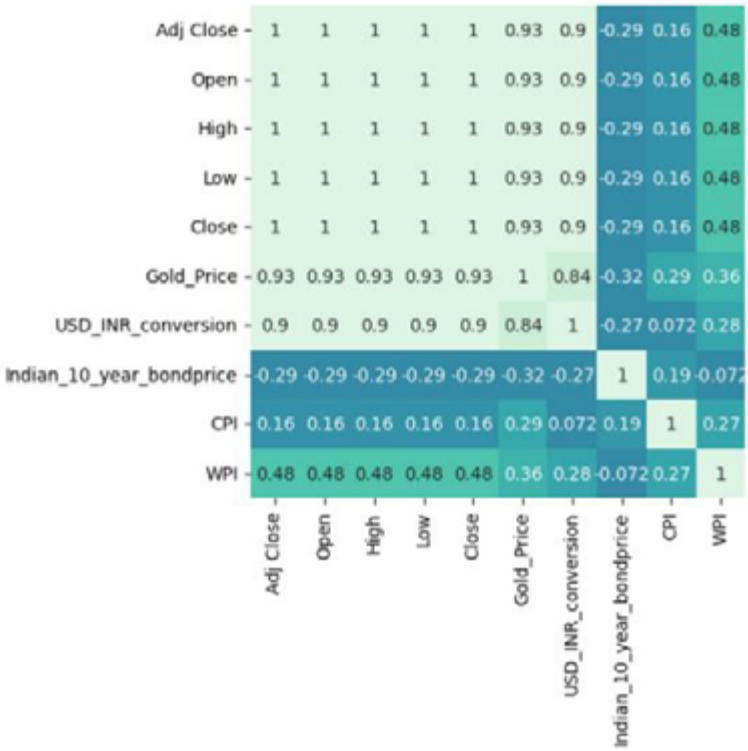


Fig. 4: Figure4: Correlation matrix.
Source: compiled by the author using Google Colaboratory- XGBoost after integration with economic variables, 2023

thus to prevent overfitting th best order was not chosen initally. The model was run with various orders to obtain the best result. The exogenous variables were the other features like open, high, close, and low. The evaluation metrics were root mean squared error, mean absolute error and r squared score. The results of forecasting on test dataset are given in Figure 3b).

For XGBoost, the model was prepared using XGBRegressor (extreme Gradient Boosting) from the XGboost library, which has been used for optimized and efficient implementation, high-speed and performance. The hyperparameters involved were learning rate which controls the contribution of each tree and was set to deafult (0.1), max_depth which represents the maximum depth of a tree and the default of 6 was used, n_estimators which represents number of trees to be fit (deault value of 100) and subsample fraction of samples to be used for fiiting the individual trees (default value of 1). The plot has been shown in Figure 3c).

For FB prophet, the model was initialised with daily seasonality. The other hyperparameters like yearly seasonality, seasonality mode and changepoint prior scale were set to default. The prediction of the model has been showed in Figure 3d).

3.5 Implementation of machine learning algorithms on pre-processed dataset merged with economic variables

The preparation and processing of the dataset (consisting of stock prices and economic variables) followed similar approach of pre-processing earlier. The basic pre-processing steps were taken from the starting like to drop the null values and renaming of the columns to avoid conflict between features. The values provided in the dataset of WPI and CPI were monthly rather than daily, thus we had to duplicate the value throughout the month to ensure smooth application of dataset on models. Figure 4 shows the correlation among all variables

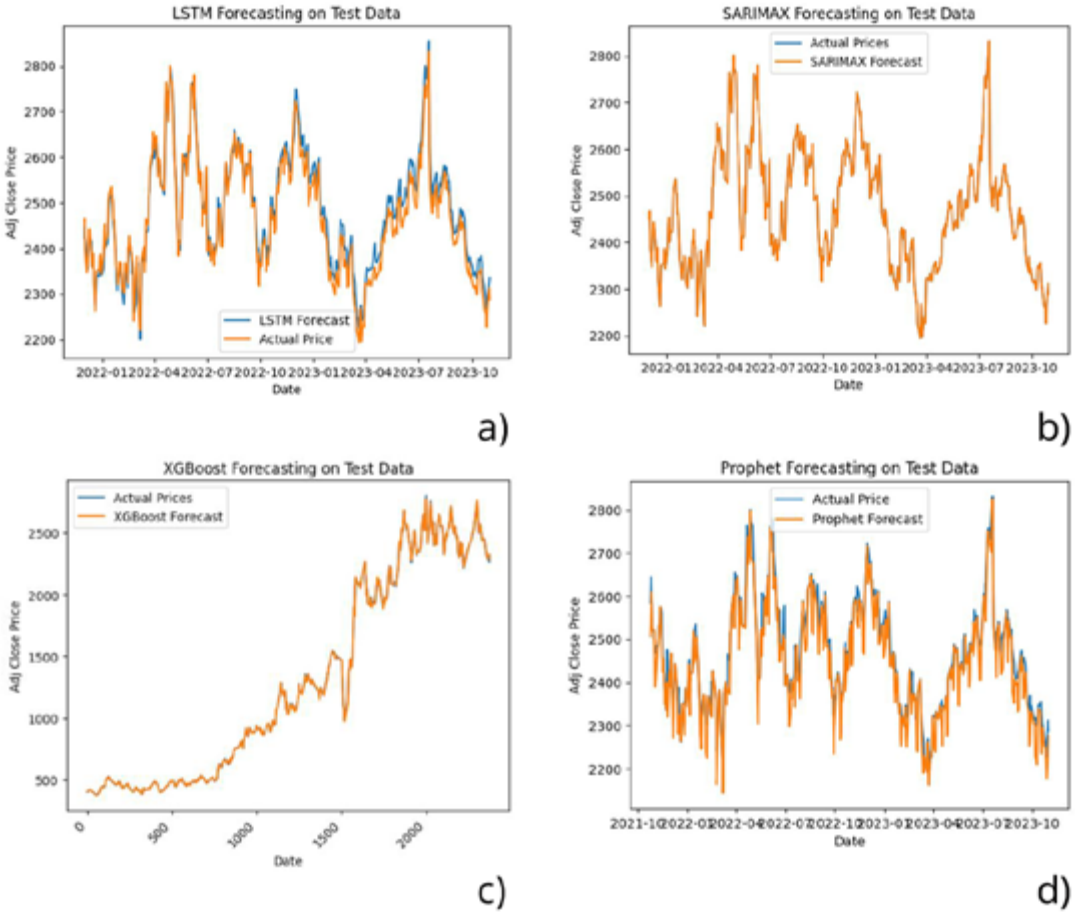


Fig. 5: a) Forecasting plot for LSTM – merged data set b) Forecasting plot of SARIMAX – merged with economical indicators c) Forecasting results of XG Boost d) Forecasting with FB prophet

Source: compiled by the author using Google Colaboratory- FBProphet after integration with economic variables, 2023.; Google Colaboratory- LSTM after integration with economic variables, 2023; Google Colaboratory- SARIMAX after integration with economic variables, 2023; Google Colaboratory- XGBoost after integration with economic variables, 2023.

using a heatmap or correlation matrix. The correlation coefficients indicate the strength of the relationships between the independent variables and the target variable. The correlation coefficient between Gold Price and US dollar to INR conversion is the highest among the economic factors included, at 0.93 and 0.90 respectively. Following that, WPI has a moderate correlation with the ‚Adj Close' at 0.48, while CPI has the lowest correlation coefficient at 0.16. Additionally, the factor Indian 10-year bond price shows an inverse relationship with the target variable, with a correlation coefficient of -0.29.

For LSTM, initial step of importing the necessary libraries was kept similar like previous model. After the pre-processing the data was split again in two ratios to analyse the result efficiently and compare it with previous LSTM model. The model was built using a similar process of developing a sequential neural network. The plot has been depicted in Figure 5a). For SARIMAX, importing the necessary libraries were kept like SARIMAX model before. The model was built from 'pmdarima' to select the best SARIMAX order based on Akaike Information Criterion which was (0,1,2) but again the order provided was (0,0,0) to observe the changes. The seasonality was kept at none. The forecasting has been depicted in Figure 5b). For XGBoost, the initial process of importing the necessary libraries and data pre-processing have been kept like the previous XGBoost model. The train test split has been kept 80 percent of training data and 20 percent of test data. The plotting of the results is depicted in Figure 5c).

Model	Trainnig %	Metric	Before integration with economic variables	After integration with economic variables	Interpretation	Possible Reasons for Change	Considerations for Improvement
LSTM	80%	MSE	569.8537	1366.4344	Increase in MSE indicates reduced prediction accuracy	Overfitting, Complexity from Economic Factors	Further hyperparameter tuning, regularization, feature analysis
		MAE	20.2353	32.2994	Rise in MAE suggests decreased accuracy	Overfitting, Influence of Economic Factors	Model simplification, feature selection, regularization
		R2	0.9639	0.9133	Drop in R2 implies reduced explanatory power	Overfitting, Complexity from Economic Factors	Feature engineering, model architecture adjustments, hyperparameter tuning

Tab. 2 Result analysis of LSTM

Source: compiled by the author using Google Colaboratory- LSTM after integration with economic variables, 2023; Google Colaboratory- LSTM before integration with economic variables, 2023

Model	Metric	Before integration with economic variables	After integration with economic variables	Interpretation	Possible Reasons for Change	Considerations for Improvement
SARIMAX	MSE	691.53	158.38	Significant decrease in MSE	Inclusion of economic factors	N/A (No change needed)
	MAE	25.71	11.26	Significant decrease in MAE	Inclusion of economic factors	N/A (No change needed)
	R2	0.96	0.99	R2 remained perfect after training	Inclusion of economic factors	N/A (No change needed)

Tab. 3 Table 3: Result analysis of SARIMAX

Source: compiled by the author using Google Colaboratory- SARIMAX after integration with economic variables, 2023; Google Colaboratory- SARIMAX before integration with economic variables, 2023

Model	Metric	Before integration with economic variables	After integration with economic variables	Interpretation	Possible Reasons for Change	Considerations for Improvement
FB PROPETH	MSE	2961.07	1912.35	Significant decrease in MSE	Influence of economic factors	Review economic factors, consider model adjustments
	MAE	41.12	25.87	Significant decrease in MAE	Influence of economic factors	Review economic factors, consider model adjustments
	R2	0.81	0.88	Decrease in R2 indicates reduced explanatory power	Influence of economic factors	Review economic factors, consider model adjustments

Tab. 4 Result analysis of Fb prophet

Source: compiled by the author using Google Colaboratory- Fb prophet before integration with economic variables, 2023; Google Colaboratory- FBProphet after integration with economic variables, 2023

Model	Metric	Before integration with economic variables	After integration with economic variables	Interpretation	Possible Reasons for Change	Considerations for Improvement
XGBOOST	MSE	31.35440	15.96573	Decrease in MSE suggests improved prediction accuracy	Influence of exogenous factors	Better preprocessing, inclusion of more factors
	MAE	4.15593	2.54016	Decrease in MAE suggests improved accuracy	Influence of exogenous factors	Better preprocessing, inclusion of more factors
	R2	0.99994	0.99997	Slight increase in R2	Influence of exogenous factors	Better preprocessing, inclusion of more factors

Tab. 5 Result analysis of XGBoost

Source: compiled by the author using Google Colaboratory- XGBoost after integration with economic variables, 2023; Google Colaboratory- XGBoost before integration with economic variables, 2023

For Fb-prophet, the initial steps taken were same as the previous model like importing the libraries and pre-processing approach. plotting of the actual price in comparison with predicted price have been showed in Figure 5d)

After the implementation of phase 1, the ranking of models based on MSE were as follows: LSTM (Rank 2), SARIMAX (Rank 3), FB-Prophet (Rank 4) and XGBoost (Rank 1). Based on R2 Score: LSTM (Rank 2), SARIMAX (Rank 3), FB-Prophet (Rank 4) and XGBoost (Rank 1). After the selection of economic variables and implementation on same models. The ranks of models shifted and is as follows based on MSE: LSTM (Rank 3), SARIMAX (Rank 2), FB-Prophet (Rank 4) and XGBoost (Rank 1). This showed that LSTM was not able to deal with the extra information and the model was overfitting. The analysis of the implementation of LSTM, SARIMAX, FB-prophet and XGBoost has been depicted in the Table 2, 3, 4 and 5 before and after merging the economic factors (gold prices, USD to INR conversion, WPI, CPI and Indian 10-year bond price). The performance of the model has been measured based on mean square error, mean absolute error and R2 score.

4 RESULTS

The LSTM results from Table 2 reveal a notable decline in predictive performance when the model was trained with 80% of the data, despite the inclusion of economic factors. This decline, reflected in the Mean Squared Error (MSE) increasing from 569.8537 to 1366.4344, Mean absolute error increasing from 20.2353 to 32.994 and decrease in R2 score from 0.9636 to 0.9133 prompts a closer examination of the inner workings of the LSTM architecture, particularly focusing on the input gate, forget gate, and output gate.

The input gate, responsible for determining the relevance of new information to be stored in memory, played a crucial role in this scenario. After integration, the LSTM might not have effectively utilized the economic indicators, potentially due to insufficient exposure to diverse instances in the training data. This limited exposure could have hindered the model‘s ability to generalize complex economic patterns, leading to a suboptimal decision-making process by the input gate. Furthermore, the forget gate, responsible for deciding what information from the previous state should be discarded, likely faced challenges in discerning crucial patterns in the absence of a sufficiently rich dataset. Consequently, the LSTM might have unintentionally discarded relevant economic information, resulting in a degradation of predictive accuracy. This shortfall is indicative of the model‘s struggle to capture long-term dependencies in economic trends, a core strength of LSTM architecture. On the other hand, it could also have been due to overfitting of data, as the sample size of data is relatively small and the

number of features has increased thus providing excess information which evidently resulted in decline of model's performance.

The performance evaluation of the SARIMAX model in Table 3 indicates a significant improvement when economic indicators are incorporated. Initially trained on 80% of the data without economic factors, the model relied solely on historical patterns for predictions. However, the inclusion of economic indicators such as gold price, Wholesale Price Index (WPI), Consumer Price Index (CPI), US to INR exchange rate, and India's 10-year yield bond price resulted in notable enhancements. The decrease in Mean Squared Error (MSE) from

691.53 to 158.38 and Mean Absolute Error (MAE) from 25.71 to 11.26, accompanied by an increase in the R-squared (R2) score from 0.96 to 0.99, underscores the effectiveness of incorporating economic factors. These improvements suggest that the SARIMAX model, characterized by its order parameters (p, d, q), benefitted from the inclusion of economic predictors. The choice of order parameters (0,0,0) signifies no autoregressive, differencing, or moving average components considered in the initial model. While the suggested parameters (p=0, d=1, q=2) based on manual and automatic inspection may lead to a perfect fit with an R2 score of 1, it's crucial to consider the implications of overfitting, especially with a small sample size. By opting for (0,0,0) parameters, the SARIMAX model remains parsimonious, avoiding unnecessary complexity that could potentially lead to overfitting. Moreover, it's important to note that stock prices are inherently volatile and influenced by numerous unpredictable factors. In such cases, overly complex models may capture noise rather than genuine patterns, leading to poorer generalization performance. The choice of (0,0,0) parameters indicates that SARIMAX primarily accounts for short-term dependencies in the time series data and adjusts to external factors through the inclusion of economic predictors. This approach strikes a balance between model complexity and performance, ensuring robustness in forecasting stock prices.

The results from Table 4 showcase an increase in the predictive performance of Facebook Prophet (FB Prophet) after training with 80% of the data, notably influenced by the inclusion of economic factors. FB Prophet, a renowned time series forecasting model, is adept at handling business problems through the utilization of historical data. Thus, the addition of economic indicators, such as gold price, Wholesale Price Index (WPI), Consumer Price Index (CPI), US to INR exchange rate, and India's 10-year yield bond price, led to a decrease in Mean Squared Error (MSE) from 2961.07 to 1912.35 and Mean Absolute Error (MAE) from 41.12 to 25.87. The perfect R-squared (R2) increased to 0.88, indicating a rise in the model's explanatory power. FB Prophet operates through the workflow involving the calculation of logistic trend functions, capacity values, yearly seasonality, and the identification of changepoints. The model's core equation comprises a trend component (g(t)), seasonality component (s(t)), holiday effect (h(t)), and an error term (εt). It leverages a piecewise function for trend modelling with linear trends and user-specified changepoints. Additionally, FB Prophet utilizes Fourier series for flexible modelling of seasonality patterns. The observed increase in FB Prophet's performance suggests that the model, sensitive to the influence of economic factors, may have adapted to the additional complexity introduced by these external variables. To enhance predictive accuracy furthermore, considerations for improvement could be a detailed review of the specific economic factors, potential adjustments to model parameters such as trend components and seasonality, and a strategic handling of anomalies within the economic indicators. This highlights the importance of a nuanced understanding and careful integration of external factors for effective time series forecasting using FB Prophet.

The results of XGBoost from Table 5 depicts that the model demonstrated a significant improvement in predictive accuracy after training with 80% of the data, marked by a decrease in Mean Squared Error (MSE) from 31.3544 to 15.96573, a reduction in Mean Absolute Error (MAE) from 4.15593 to 2.54016, and a slight increase in R-squared (R2) from 0.99994 to 0.99997. This enhancement can be attributed to XGBoost's ensemble learning approach, which sequentially trains decision trees and combines their outputs. The model effectively utilized exogenous factors, refining its predictions by compensating for errors made by

previous trees and mitigating overfitting through pruning. The mathematical framework, guided by an objective function comprising a Loss function and Regularization, facilitated optimal parameter tuning.

SARIMAX and XGBoost stands out as the most suitable model for forecasting in the provided Indian market data when compared to LSTM and FB Prophet. The models demonstrated a substantial improvement in predictive accuracy, showcasing its adaptability to the influence of exogenous factors. In contrast, LSTM showed sensitivity to dataset size, FB Prophet faced challenges even after the integration of economic factors and remained at R2 score of 0.88. Considering the flexibility, accuracy, and adaptability shown by XGBoost throughout, it emerges as the preferred choice for forecasting in the Indian market.

5 DISCUSSION

The studies Siami-Namini and Namin in 2018 and Fischer and Krauss in 2018 recognise the ability of LSTM to capture long term dependencies in time-series data (Fischer and Krauss, 2018; Siami-Namini and Namin, 2018). These studies accentuate the model's superiority in financial markets. However, performance of LSTM depends heavily on size and quality of training data. During the course of this study, MSE changed from 569.8537 to 1366.4344 highlighting the importance of hyperparameter tuning as emphasised by Gorgolis et. al in 2019 (Gorgolis et al., 2019). The model FB Prophet praised by Taylor and Letham in 2017 for its intuitively adjustable parameters shows a significant rise during this study upon inclusion of economic factors with MSE decreasing from 2961.07 to 1912.35 too (Taylor and Letham, 2017). The decline in performance resonates with findings by Taylor and Letham which indicates model's struggle with the integration of complex exogenous variables and this suggests the need for refined parameter adjustments as highlighted by Hamdani et al. in 2023 (Hamdani et al., 2023).

The study by authors Hyndman and Athanasopoulos in 2018, highlights the robustness of SARIMAX in various scenarios (Hyndman, R.J. and Athanasopoulos, G., 2018). The best order parameters (p=0, d=1, q=2) in this study resonates with recommendations from authors Hyndman and Athanasopoulos but the implementation has been performed on the order (0,0,0). In SARIMAX model the reduction of MSE and MAE is a tantamount to model's better performance. Although SARIMAX is highly effective in incorporating exogenous variables during the course of this study it shows potential for overfitting as indicated by reduction in MSE from 25.71 to 11.26. XGBoost as documented by Chen and Guestrin in 2016, as a powerful tree ensemble model and applied by study from Zhang and Chen in 2021 displays remarkable performance in financial domain (Chen and Guestrin, 2016; Zhang and Chen, 2021). The model's ability to handle exogenous variables is also supported by Zhang et al in 2018 (Zhang et al., 2018). The results of this study echo with Chen and Guestrin's findings as shown by the reduction in MSE from 31.35440 to 15.96573. During this study SARIMAX and XGBOOST emerge as reliable options for forecasting in Indian markets. Considering SARIMAX displays results that are prone to overfitting XGBoost's ability to effectively incorporate exogenous variables and ensemble approach suggests it as preferred model.

6 VALIDITY OF CONCLUSIONS / FUTUREWORK

The study demonstrates that among the selected models, XGBoost and SARIMAX have shown better efficiency and performance for Indian stock market predictions. On the other hand, LSTM and FB-Prophet had higher errors and lower R2 score which generally represents the extent of fitting of data in the model. The models made have been tuned for a single dataset/company in the whole financial market of India which holds different segments of corporations. Thus, it induces a further question like whether these conclusions or the results can be

generalized for another company in India or for another country? Or is there any robustness in the models against market anomality's or external shocks?

As mentioned, this study encompasses only one of the components affecting the financial market i.e., economic variables. A study by Sharif et al. in 2015 discusses the firm specific factors like firm size, return on equity and other variables and how it impacts the stock prices in Bahrin (Sharif et al., 2015). Another study by King et al. in 2012 discusses the factors like political events and economic news impacting the oil prices movements (King et al., 2012). In the study by Al Tamimi et al. in 2011, they discussed internal and external factors affecting the stock prices in UAE. Some of these internal factors were performance of company, change in board of directors etc. and external factors included government rules and regulations, market conditions, behaviour of participants and other uncontrollable variables (Al-Tamimi et al., 2011).

In most of the studies, authors have concluded that there is more than one factor involved that is impacting the stock prices in any country. Thus, to generalize the model for any country or for any other segment in India and getting it secure against external shocks, a study will need to incorporate all above-mentioned parameters into one. Even then, further questioning can be done like on what basis will one assign the weights to each factor so that the model can learn without getting biased to any one factor. The models made in this study are based on two factors i.e., model characteristics to deal with error (for instance XGBoost uses previous tree to deal with errors) and other factor impacting Indian financial market. If the future work wants to generalize the results of this study, they will have to take caution as every financial market has their own characteristics, regulations, and impact from economic conditions. For example, share of reliance industries are dependent on economic conditions as it is multi-national company. On the other hand, shares prices of a cosmetic company, food related companies will show different behaviour (for instance, food related firms will be more sensitive towards consumers, thus have a relation with CPI). The development of these models can be generalized only if, the future work includes implementation of these models on a company which is affected by economic indicators with change in data of these variables with each country. Also, the robustness of the model is heavily dependent on accuracy of data involved. Thus, it is a necessity to consider that. The current conclusion of the study is heavily dependent upon the characteristics of models. Thus, if there is external shock involved for instance, a new policy implemented by government that will affect the company's share price, the model will face difficulties in predicting the stock prices as they have done in this study. To make the results more generalized and model more robust further study is needed to be done in future.

The study also has other few shortcomings whilst delivering the subject matter which can be filled in future by another research. Some of these are:

Factors	Rank of countries			
	China	United States	United Kingdom	India
Internet Affordability	34	32	22	28
Internet Quality	10	6	32	16
Electronic Infrastructure	49	5	13	91
Electronic Security	79	43	23	66
Electronic Government	16	2	8	35

Tab. 6 Digital Quality Index 2023
Source: compiled by the author (using Digital Quality Index, 2023)

- The study included basic preprocessing approaches rather than complex processes to get the optimal performance of the desired models.
- The study also interprets the results based on the author's understanding on subject.
- The study focuses on particularly 5 economic indicators from India, some of which are only on monthly basis available like CPI and WPI. This shortcoming can be improved by taking daily prices to improve the model's accuracy and reducing the error further via paid websites or dealers.
- The study also focuses on several of the economic indicators which provides a general understanding of India's economic conditions. The research can be improved further by incorporating more economic indicators and implementing them to the same models.

An additional constraint to the hypothesis lies in the assumption that the limited availability of data and published materials is a primary factor contributing to the scarcity of new hybrid models developed for the Indian market. However, an alternative factor may also be influential, as exemplified by India's placement at the 52nd position among 121 countries in the Digital Quality Index 2023 as assessed by surfshark.com (DQI index, 2023). The evaluation criteria encompass Internet Affordability, Internet Quality, Electronic Infrastructure, Electronic Security, and Electronic Government. In Table 6, the comparison of the top four countries including India in terms of publication is grounded in these criteria. The findings suggest that the paucity of data, particularly regarding the performance of models in the Indian market, may be correlated with India's ranking of 91st out of 121 countries in electronic infrastructure. Despite being the fifth-largest economy and possessing the potential to yield substantial returns for investors, India's digital landscape does not adequately support technological advancement. Furthermore, the decline in Internet affordability from the 8th rank in 2020 to the 28th rank in 2023 raises concerns. This trend implies that even if novel applications or software are developed using new hybrid models, accessibility for users may be constrained. Consequently, this scenario contributes to a diminished interest in the field within the Indian market.

Even after considering these variables, according to annual growth rate of documents, domain of automated trading system is just going to get bigger in future ahead and one of the leading potential economies will be India and other developing countries.

7 CONCLUSION

This research contributes to the understanding of the performance of LSTM, SARIMAX, FB Prophet, and XGBoost models, emphasizing the impact of economic factors on predictive accuracy. The comprehensive exploration into time series forecasting within the dynamic landscape of the Indian financial market has yielded valuable insights, offering practical implications for deploying various algorithms and integrating economic indicators with conventional features for investors, decision makers, traders or academicians. The LSTM model demonstrated a superior performance initially but a notable declined was observed after the integration of economic variables. XGBoost exhibited efficient performance throughout, displaying a decrease in errors, and highlighting its robust ensemble learning capabilities. SARIMAX after incorporating economic indicators, delivered better results with a substantial decrease in both MSE and MAE. FB Prophet also performed relatively well with the new dataset, emphasizing the importance of careful consideration when incorporating economic factors. The findings can be applied in the domain of research, as it provides basic supplementary information on which newer age models for Indian financial market can be developed. The knowledge of the performance of selected models and the visualisation of the established relation between macro-economic variables and time series forecasting models can be used by potential investors, potential traders or decision makers who wants to create a strategy for buying and selling or multiply their investment through algorithmic trading.

Acknowledgements

The publication was supported by the IGA doctoral project (Internal grant agency PEF), namely project IGA-PEF-TP-23-013.

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ABU-MOSTAFA, Y. S. and ATIYA, A. F. 1996. Introduction to financial forecasting. *Appl Intell,* 6, 205–213. https://doi.org/10.1007/BF00126626

AGRAWAL, M., KHAN, A. U. and SHUKLA, P. K. 2019. Stock price prediction using technical indicators: a predictive model using optimal deep learning. *Learning,* 6, 7.

AITHAL, P., ACHARYA, D. and GEETHA, M. 2019. Identifying Significant Macroeconomic Indicators for Indian Stock Markets. *IEEE Access*, 7, 143829-143840. https://doi.org/10.1109/ACCESS.2019.2945603

AL-TAMIMI, H. A. H., ALWAN, A. A. and ABDEL RAHMAN, A. A. 2011. Factors Affecting Stock Prices in the UAE Financial Markets. *Journal of Transnational Management,* 16, 3–19. https://doi.org/10.1080/15475778.2011.549441

BATRA, V., KANDPAL, D. and SINHA, R. 2020. Relationship between exchange rate (usd/inr) and stock market indices in India (sensex). *Asian Journal of Research in Banking and Finance,* 10, 1–14.

BHUNIA, A. and DAS, A. 2012. Association between gold prices and stock market returns: Empirical evidence from NSE. *Journal of Exclusive Management Science,* 1, 1–7.

BISWAS, A. 2018. Impact of Reliance Industry Stock Price on NIFTY 50-Granger Causality Test.

RESEARCH REVIEW International Journal of Multidisciplinary, 3(11), 367-376. https://doi.org/10.5281/zenodo.1490556

CHATTERJEE, A., BHOWMICK, H. and SEN, J. 2022. Stock Volatility Prediction using Time Series and Deep Learning Approach. *arXiv*: 2210.02126. https://doi.org/10.48550/arXiv.2210.02126

CHEN, K., ZHOU, Y. and DAI, F. 2015. A LSTM-based method for stock returns prediction: A case study of China stock market. In: *IEEE International Conference on Big Data (Big Data)*. Santa Clara, CA, USA, pp. 2823-2824. https://doi.org/10.1109/BigData.2015.7364089

CHEN, T. and GUESTRIN, C. 2016. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 785–794. https://doi.org/10.1145/2939672.2939785

DASSANAYAKE, W., ARDEKANI, I., JAYAWARDENA, C., SHARIFZADEH, H. and GAMAGE, N. 2021. Forecasting accuracy of Holt-Winters exponential smoothing: Evidence from New Zealand. *New Zealand Journal of Applied Business Research,* 17, 11–30. https://doi.org/10.3316/informit.391329680991168

DHINGRA, K. and KAPIL, S., 2021. Impact of Macroeconomic Variables on Stock Market—An Empirical Study. In: LAKHANPAL, P., MUKHERJEE, J., NAG, B. and TUTEJA, D. (Eds.). *Trade, Investment and Economic Growth: Issues for India and Emerging Economies*. Springer, Singapore, pp. 177–194. https://doi.org/10.1007/978-981-33-6973-3_12

DICKEY, D. A. and FULLER, W. A. 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49, 1057–1072. https://doi.org/10.2307/1912517

FARID, S., TASHFEEN, R., MOHSAN, T. and BURHAN, A., 2023. Forecasting stock prices using a data mining method: Evidence from emerging market. *International Journal of Finance & Economics,* 28, 1911–1917. https://doi.org/10.1002/ijfe.2516

FISCHER, T. and KRAUSS, C. 2018. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research,* 270, 654– 669. https://doi.org/10.1016/j.ejor.2017.11.054

GOOGLE COLABORATORY. 2023. *Fb prophet before integration with economic variables* [online]. https://colab.research.google.com/drive/1MbtF5EsiczjPpseNTRPur0CNRiL3KCwk#scrollTo=6zNcy9mP-djZ [accessed 6.10.24].

GOOGLE COLABORATORY. 2023. *FB Prophet after integration with economic variables* [online]. https://colab.research.google.com/drive/1T1Wfgbbdzv1o0xIrIk3Cqd-ZeYvt8rol#scrollTo=CVTm37ixq6MX [accessed 6.10.24].

GOOGLE COLABORATORY. 2023. *LSTM after integration with economic variables* [online]. https://colab.research.google.com/drive/10geD107lGWAusxI4VJzA-06ywM_Xe0C4#scrollTo=gtjgxvaXbikH [accessed 6.10.24].

GOOGLE COLABORATORY. 2023. *LSTM before integration with economic variables* [online]. https://colab.research.google.com/drive/1KnpEyH-SOnyCGDWH2m9U3oDB4evqcxBm#scrollTo=aZhdNWwNNLCr [accessed 6.10.24].

GOOGLE COLABORATORY. 2023. *SARIMAX after integration with economic variables* [online]. https://colab.research.google.com/drive/1ZKlhC1P67NdeyysWn7Ml_rDouC4yvWEB#scrollTo=SYDozCaHrPuK [accessed 6.10.24].

GOOGLE COLABORATORY. 2023. *SARIMAX before integration with economic variables* [online]. https://colab.research.google.com/drive/1tBVT3L24QseeGO7VjBK3TuZvK44uCSCJ#scrollTo=ctZInzRBWxcK [accessed 6.10.24].

GOOGLE COLABORATORY. 2023. *XGBoost after integration with economic variables* [online]. https://colab.research.google.com/drive/1jadYL8RGEokaIWYo6IhQliYG0R3UR8wc#scrollTo=UJuoBi5FhGFl [accessed 6.10.24].

GOOGLE COLABORATORY. 2023. *XGBoost before integration with economic variables* [online]. https://colab.research.google.com/drive/1jCwnYGKrDhdghVySmDL5mf6gO7odMp#scrollTo=WJe9w8A8lP18 [accessed 6.10.24].

GORGOLIS, N., HATZILYGEROUDIS, I., ISTENES, Z. and GRAD-GYENGE, L. 2019. Hyperparameter Optimization of LSTM Network Models through Genetic Algorithm. In: *10th International Conference on Information, Intelligence, Systems and Applications (IISA)*. Patras, Greece, pp. 1-4, https://doi.org/10.1109/IISA.2019.8900675

GRANGER, C. W. J. 1969. Investigating Causal Relations by Econometric Models and Cross- spectral Methods. *Econometrica,* 37, 424–438. https://doi.org/10.2307/1912791

GUMELAR, A., SETYORINI, H., ADI, D., NILOWARDONO, S., LATIPAH, WIDODO, A., TEGUH WINOWO, A., SULISTYONO, M. and CHRISTINE, E. 2020. Boosting the Accuracy of Stock Market Prediction using XGBoost and Long Short-Term Memory. In: *International Seminar on Application for Technology of Information and Communication (iSemantic)*. Semarang, Indonesia pp. 609–613. https://doi.org/10.1109/iSemantic50169.2020.9234256

HAMDANI, A. F., SWANJAYA, D. and HELILINTAR, R. 2023. Facebook Prophet Model with Bayesian Optimization for USD Index Prediction. *JUITA: Jurnal Informatika*, 11, 293–300. https://doi.org/10.30595/juita.v11i2.17880

HUSSAIN, M., MALIK, A., RASOOL, N., FAYYAZ, M. and MUMTAZ, M. 2012. The Impact of Macroeconomic Variables on Stock Prices: An Empirical Analysis of Karachi Stock Exchange. *Mediterranean Journal of Social Sciences,* 3, 295–312. https://doi.org/10.5901/mjss.2012.v3n3p295

HYNDMAN, R. J. and ATHANASOPOULOS, G. 2018. *Forecasting: principles and practice*. OTexts.
https://books.google.cz/books?hl=en&lr=&id=_bBhDwAAQBAJ&oi=fnd&pg=PA7 &dq=forecasting+principles+and+practice+2018&ots=Tjg-wfWPGK&sig=lFdCll7vOVMtzQ5hShUpgM1KLc0&redir_esc=y#v=onepage&q=sa rima&f=false

JOHANSEN, S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press. New York.

KARMAKAR, M. 2005. Modeling Conditional Volatility of the Indian Stock Markets. *Vikalpa,* 30, 21–38. https://doi.org/10.1177/0256090920050303

KHASHEI, M. and BIJARI, M. 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing, The Impact of Soft Computing for the Progress of Artificial Intelligence*, 11, 2664–2675. https://doi.org/10.1016/j.asoc.2010.10.015

KING, K., DENG, A. and METZ, D. 2012. An econometric analysis of oil price movements: the role of political events and economic news, financial trading, and market fundamentals. *Bates White Economic Consulting*, 1, 53.

KUMAR, J. P. S., SUNDAR, R. and RAVI, A. 2023. Comparison of stock market prediction performance of ARIMA and RNN-LSTM model - A case study on Indian stock exchange. *Presented at the AIP Conference Proceedings*, 2875, 020010. https://doi.org/10.1063/5.0154124

KUMARIA, A., RAJKAR, A., RAUT, A. and NAIR, R. S. 2023. Forecasting the Indian Financial Markets with LSTM and Price Indicators. In: RAY, K. P., DIXIT, A., ADHIKARI, D. and MATHEW, R. (Eds.). *Proceedings of the 2nd International Conference on Signal and Data Processing*. Springer Nature, Singapore, pp. 395–403. https://doi.org/10.1007/978- 981-99-1410-4_33

LAKSHMANASAMY, T. 2021. The Relationship Between Exchange Rate and Stock Market Volatilities in India: ARCH-GARCH Estimation of the Causal Effects. *International Journal of Finance Research,* 2, 244–259.

LEE, K. J., CHI, A. Y., YOO, S. and JIN, J. J. 2008. Forecasting korean stock price index (kospi) using backpropagationn neural network model bayesian chiao's model and sarima model. *Journal of Management Information and Decision Sciences*, 11.

LITZENBERGER, R., CASTURA, J. and GORELICK, R. 2012. The impacts of automation and high frequency trading on market quality. *Annu. Rev. Financ. Econ.,* 4, 59–98.

LIU, L., PEI, Z., CHEN, P., LUO, H., GAO, Z., FENG, K. and GAN, Z. 2023. An Efficient GAN-Based Multi-classification Approach for Financial Time Series Volatility Trend Prediction. *Int J Comput Intell Syst,* 16, 40. https://doi.org/10.1007/s44196-023-00212-x

LIU, Y., HUANG, S., TIAN, X., ZHANG, F., ZHAO, F. and ZHANG, C. 2024. A stock series prediction model based on variational mode decomposition and dual-channel attention network. *Expert Systems with Applications,* 238, 121708.

MEHTAB, S., SEN, J. and DUTTA, A. 2021. Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models. In: THAMPI, S. M., PIRAMUTHU, S., LI, K.-C., BERRETTI, S., WOZNIAK, M. and SINGH, D. (Eds.). *Machine Learning and Metaheuristics Algorithms, and Applications*. Springer, Singapore, pp. 88–106. https://doi.org/10.1007/978-981-16-0419-5_8

NASIRI, H. and EBADZADEH, M. M. 2023. Multi-step-ahead stock price prediction using recurrent fuzzy neural network and variational mode decomposition. *Applied Soft Computing,* 148, 110867.

OBTHONG, M., TANTISANTIWONG, N., JEAMWATTHANACHAI, W. and WILLS, G. 2020. A survey on machine learning for stock price prediction: algorithms and techniques. In: *Presented at the 2nd International Conference on Finance, Economics, Management and IT Business.* (05/05/20 - 06/05/20), pp. 63–71. https://doi.org/10.5220/0009340700630071

PANIGRAHI, A. et al. 2022. Impact of Global and Domestic Economic Variables on 10-Year Indian Government Bond Yield: An Empirical Study. *IUP Journal of Applied Finance; Hyderabad*, 28(2), 5-23.

PATEL, S. A. 2013. Causal Relationship Between Stock Market Indices and Gold Price: Evidence from India. *IUP Journal of Applied Finance,* 19(1), 99-109.

RAHEEM AHMED, R., VVEINHARDT, J., ŠTREIMIKIENĖ, D., GHAURI, S. P. and AHMAD, N. 2017. Estimation of long-run relationship of inflation (CPI & WPI), and oil prices with KSE-100 index: evidence from Johansen multivariate cointegration approach. *Technological and Economic Development of Economy*, 23(4), 567-588. https://doi.org/10.3846/20294913.2017.1289422

SAPRE, A. A. and GORI, S. 2023. The Predicament of Land Acquisition, Displacement and Resettlement: An Analysis of Indian Scenario. *Journal of Asian and African Studies,* 00219096231179651. https://doi.org/10.1177/00219096231179651

SEAH, S. 2022. Untapped Potential in the ASEAN-India Relationship: Climate Change and Green Recovery, In: *ASEAN and India. WORLD SCIENTIFIC*, pp. 227–233. https://doi.org/10.1142/9789811262906_0027

SHARIF, T., PUROHIT, H. and PILLAI, R. 2015. Analysis of Factors Affecting Share Prices: The Case of Bahrain Stock Exchange. *International Journal of Economics and Finance,* 7(3), 207. https://doi.org/10.5539/ijef.v7n3p207

SHARMA, K., BHALLA, R. and GANESAN, G. 2022. *Time Series Forecasting Using FB-Prophet*. Presented at the ACM.

SIAMI-NAMINI, S. and NAMIN, A. S. 2018. Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. *arXiv*: 1803.06386. https://doi.org/10.48550/arXiv.1803.06386

SINGH, P. and BORAH, B. 2014. Forecasting stock index price based on M-factors fuzzy time series and particle swarm optimization. *International Journal of Approximate Reasoning,* 55, 812–833. https://doi.org/10.1016/j.ijar.2013.09.014

Smith, G. n. d. *The Price of Gold and Stock Price Indices for The United States*.

SONKAVDE, G., DHARRAO, D., BONGALE, A., DEOKATE, S., DORESWAMY, D. and BHAT, S. 2023. Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. *International Journal of Financial Studies,* 11, 94. https://doi.org/10.3390/ijfs11030094

SRIVASTAVA, S., PANT, M. and GUPTA, V. 2023. Analysis and prediction of Indian stock market: a machine-learning approach. *International Journal of System Assurance Engineering and Management,* 14, 1567–1585.

SUBBURAYAN, B., DHIVYA, N. and ALEX, A. 2021. Empirical Relationship of Macroeconomic Variables and Stock Prices : Indian Stock Market and Japanese Stock Market.

TAYLOR, S. and LETHAM, B. 2017. Forecasting at Scale. *The American Statistician*, 72(1), 37-45. https://doi.org/10.1080/00031305.2017.1380080

ZHANG, D., QIAN, L., MAO, B., HUANG, C. and SI, Y. 2018. A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGBoost. *IEEE Access*, 6, 21020-21031,. https://doi.org/10.1109/ACCESS.2018.2818678

ZHANG, Y. and CHEN, L. 2021. A Study on Forecasting the Default Risk of Bond Based on XGboost Algorithm and Over-Sampling Method. *Theoretical Economics Letters,* 11, 258–267. https://doi.org/10.4236/tel.2021.112019

Reliance Industries Limited (reliance.ns) stock historical prices & data. 2023. *Yahoo! Finance* [online]. https://finance.yahoo.com/quote/RELIANCE.NS/history?p=RELIANCE.NS [Accessed: 01 December 2023].

Digital quality of life index. 2023. *Surfshark.com* [online]. https://surfshark.com/dql2023 [Accessed: 20 December 2023].

Gold futures historical prices. 2023. *Investing.com India* [online]. https://in.investing.com/commodities/gold-historical-data [Accessed: 20 December 2023].

India 10-year Bond Historical Data. 2023. *Investing.com India* [online]. https://in.investing.com/rates-bonds/india-10-year-bond-yield-historical-data [Accessed: 20 December 2023].

India wholesale price index (WPI). 2023. *Investing.com India* [online]. https://in.investing.com/economic-calendar/indian-wpi-inflation-564 [Accessed: 20 December 2023].

India consumer price index (CPI). 2023. *Investing.com India* [online].. Available at: https://in.investing.com/economic-calendar/indian-cpi-973 [Accessed: 20 December 2023].

USD INR historical data. 2023. *Investing.com India* [online]. https://in.investing.com/currencies/usd-inr-historical-data%20 [Accessed: 20 December 2023].

Contact information

Ankit Tripathi: email: xtripat1@.mendelu.cz
Arpit Tripathi: email: xtripat2@.mendelu.cz
Oldrich Trenz: email: oldrich.trenz@mendelu.cz
Pawan Kumar Mishra: email: pawan.mishra@mendelu.cz

DIGITIZATION OF THE METHODOLOGY FOR ASSESSING THE SUSTAINABILITY OF PLANT PRODUCTION SYSTEMS

Pavel Turčínek[1], Vojtěch Krejsa[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

This contribution describes how a web application was created based on the given methodology for assessing the sustain-ability of plant production systems. The key indicators of this methodology are introduced. The way of transformation into digital form is shown. The outputs of the created application are presented.

Keywords: sustainability of plant production systems, digitization, web application

JEL Code: C88, Q59

1 INTRODUCTION

Sustainable agriculture is an issue that is important in many ways. El Chami et al. (2020) ask the question "How Can Sustainable Agriculture Increase Climate Resilience?" Celicourt et al. (2021) deal with sustainable water management in agrosystems. Arlauskienė and Šarūnaitė (2023) address the area of cover crop yield, nutrient storage, and release under different cropping technologies within sustainable agrosystems. A key factor for increasing soil fertility and promoting sustainable soil use in fruit orchard agrosystems is presented by Sofo et al. (2020). Bioinspired nanomodification strategies for sustainable agriculture are introduced by Xu et al. (2021).

Already in 2011, Professor Křen's team (2011) presented a methodology dealing with a comprehensive assessment of the sustainability of agricultural enterprises focused on plant production. The methodology contains three key groups of indicators, namely ecological, economic, and social indicators. Each of these groups contains even more specific and detailed indicators. Each of the indicators is calculated based on one or more input data. This methodology is described using a text document describing the sources of input data and calculations leading to individual indicators. The absence of a tool that would facilitate the calculation and automate it to a certain extent is almost incomprehensible in today's digitized age. Therefore, there was a request to facilitate the calculation process using a web application. This application was created as part of Krejsa's diploma thesis (2022).

https://doi.org/10.11118/978-80-7509-990-7-0187

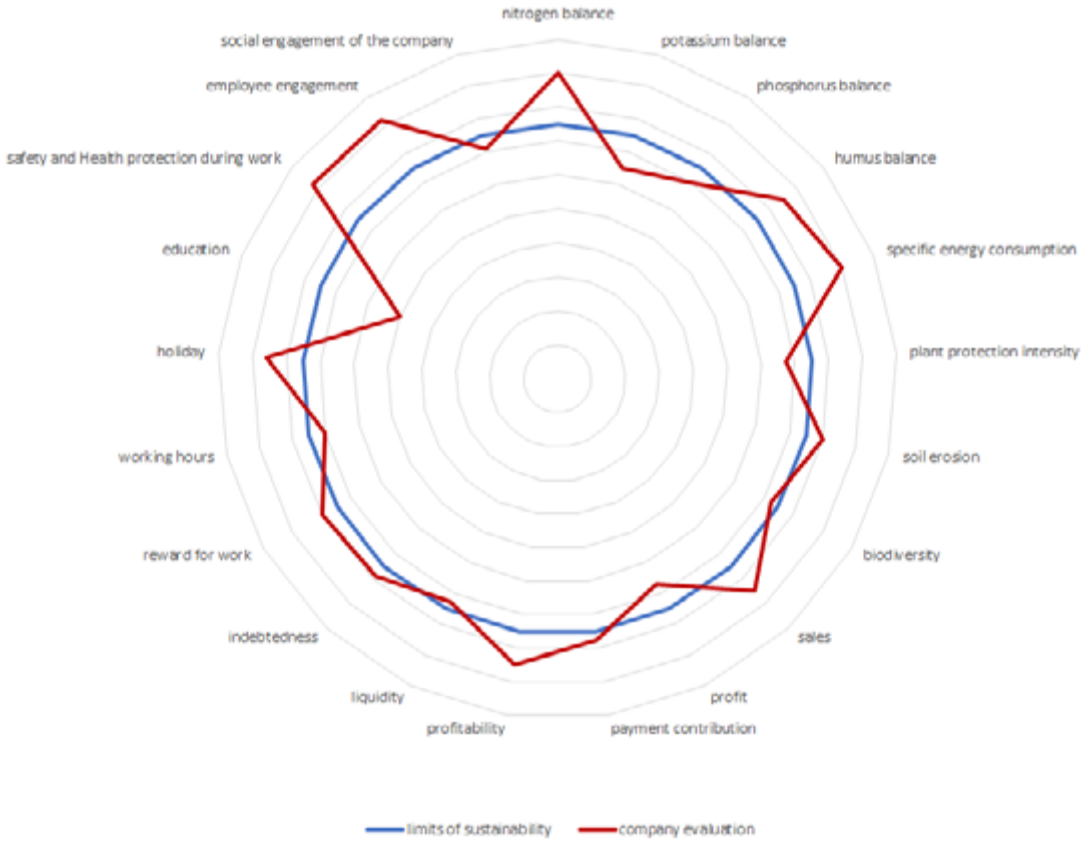


Fig. 1: Figure 1: Web graph for presentation of results
Source: Křen a kol., 2011

The purpose of this paper is to describe the created application that enables the mentioned evaluation of the sustainability of the agricultural enterprise to be carried out. The key is above all the clarity of the presented results. There is also a general emphasis on user-friendliness and speeding up the acquisition of inputs needed to achieve a rating.

2 METHODOLOGY FOR EVALUATING THE PERMANENT SUSTAINABILITY OF SYSTEMS FOR THE CONDITIONS OF THE CZECH REPUBLIC

The methodology dealing with the long-term sustainability of agricultural enterprises, according to which the evaluation will take place in the application, was described by Prof. Křen and his research team (2011). In the methodology, the name SAGROS system (sustainable agrosystems) appears, which was also used when creating the name of the web application. The output of the assessment according to the methodology is 21 indicators. These indicators can be divided into three main groups (ecological, economic, and social). Each of the indicators has its own evaluation function, which, based on the input value, can be used to obtain an output ranging from 0 to 1. A value of 0.75 was set as the limit of sustainability.

There is also a second document on the evaluation methodology, containing the methods of obtaining the data needed for the evaluation. This methodology was also created by the team of Prof. Křen (2012). Diverse data sources are mentioned here. This includes, for example, data from publicly available systems and authorities, but also data obtained from internal sources of the company.

As part of the methodology, the resulting evaluation of the company is usually presented using a graph. Due to the number of indicators and clarity, according to the methodology, the spider chart, shown in Figure 1, proved to be the most effective.

2.1 Ecological indicators

Within the ecological indicators, 8 output values are calculated:
- nitrogen balance,
- potassium balance,
- phosphorus balance,
- humus balance,
- specific energy consumption,
- plant protection intensity,
- soil erosion,
- biodiversity.

For the first six mentioned indicators, several scalar values are obtained from the user, with which basic mathematical operations are performed. The value obtained in this way is then transformed into the output value of the indicator using the evaluation function. For the last two indicators, in addition to loading other scalar inputs, data obtained from the public administration system is also needed. The last mentioned biodiversity indicator is the most complex of all indicators. It consists of 11 parts describing the land on which and in what way the farm operates.

2.2 Economic indicators

This group includes 6 output indicators. Particularly speaking about:
- sales,
- profit,
- payment contribution,
- profitability,
- liquidity,
- indebtedness.

It is possible to calculate the values of these indicators after collecting ten basic input values. As already mentioned, each of the indicators has its own evaluation function. An example evaluation function for the indebtedness indicator is illustrated in Figure 2.

2.3 Social indicators

As part of the social indicators, 7 resulting indicators will be calculated. It is about:
- reward for work,
- working hours,
- holiday,
- education,
- safety and Health protection during work,
- employee engagement,
- social engagement of the company.

Inputs for the evaluation of these indicators are obtained on the basis of interviews and questionnaires with company executives.

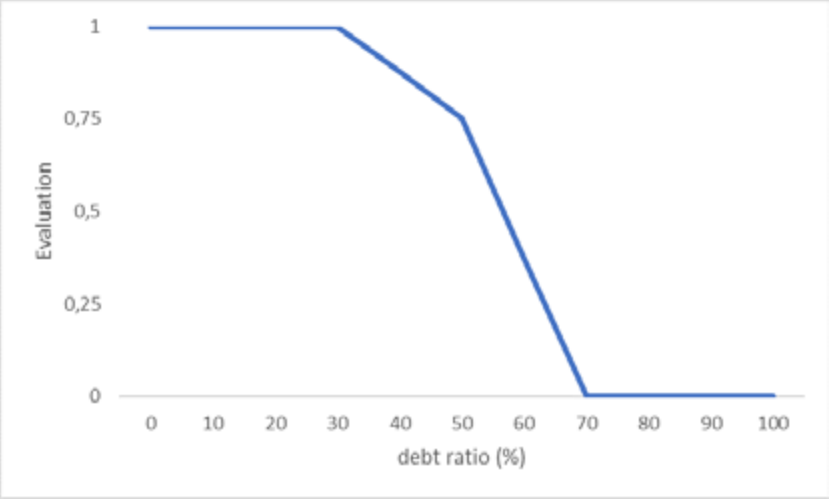


Fig. 2: Figure 2: Evaluation function of debt ratio
Source: Křen a kol., 2011

3 METHODOLOGY AND DATA

Currently, there is no software on the Czech market that would comprehensively assess the sustainability of an agricultural enterprise. According to an article on the Živá univerzita[1] (2022) website, the SustainAgri mobile application was recently published. It also deals with the sustainability of agricultural enterprises. In the application, it is possible to try out decision-making in an agricultural company within the scenarios. However, it is aimed at educating young or budding entrepreneurs in the field of agriculture and is therefore not suitable for the purpose of evaluating an existing agricultural enterprise.

Due to the absence of a suitable software solution for evaluating the sustainability of an agricultural enterprise, there was a request to create a web application that would enable this.

3.1 External data sources

As part of the evaluation of several indicators described in the methodology, data from external systems are used. As part of the design and creation of the application, this data will be downloaded and prepared in such a way that the user does not have to worry about obtaining it, thereby facilitating and speeding up the evaluation of the company.

3.1.1 LPIS

The first major data source used within the ecological indicators of soil erosion and biodiversity is data on parcels from the LPIS[2] system. As part of these indicators, the acreage and perimeter of the land block are used for evaluation. The parcels displayed in this system are publicly available. The procedure for downloading such data is described on the eAGRI (2022) web portal.

The obtained data can then be imported into a database with an extension to support geographic data. Parcels can be queried from the database for display in the interactive map in the web application. The downloaded information about the land blocks already contains the required area. The other characteristics of a given parcel can be calculated using the functions included in the database superstructure to support geographic objects.

1 Living University
2 LPIS is a geographic information system that primarily consists of records of agricultural land use.

3.1.2 Ministry of Labour and Social Affairs

The second external source in the application is data obtained from the website of Ministry of Labour and Social Affairs. These data are used to evaluate the social indicator remuneration for work. To evaluate the indicator, it is necessary to compare the entered wages with the average wage in the regional labor price statistics data set.

All data are available for download in JSON format on the Ministry‘s website. The obtained data can be serialized into objects and then inserted into the database.

3.2 Used technology

The application will be implemented according to the MVC design pattern. Technologies from three basic categories will be needed to implement the proposed application. These are frontend, backend, and database.

3.2.1 Frontend

The user interfaces of websites are usually handled using the HTML markup language. This page will be handled by a component of ASP.NET technology called Razor, which enables the creation of templates. A big advantage is the possibility of simple substitution of values from the backend. Another advantage is the possibility of linking to other pages registered in the project.

The frontend in the HTML markup language is often supplemented with styles in the CSS language. The interactivity of the front-end part can then be ensured using code written in JavaScript. The application will use a set of Bootstrap cascading styles. It is one of the most used and popular in this industry. Its main advantage is that the styles used in the application are consistent and clear. Another advantage is, from the point of view of implementation, ease of use and large quantity functionality.

It will also be necessary to use JavaScript libraries for working with the map, on which you can select soil blocks. The Leaflet library is suitable for this. This will allow the display and selection of loaded parcels. Another component needed for the application to run will be the jQuery library.

3.2.2 Backend

The C# language was chosen for the implementation of the proposed application. The latter is in second place in the presented graph under the name ASP.NET, which is a framework used in this language to implement web applications. The reason for this decision is the clarity of the created solutions, a large number of preset functions for the creation of the application, and, in addition, also quality documentation. In addition, the ASP.NET platform also directly offers the creation of a project structured according to the MVC architecture.

3.2.3 Database

As part of the implemented application, one of the most popular variants will be used, namely Microsoft SQL Server. The main reason was the possibility of using the SqlDataAdapter class in the C# language, which was chosen for the implementation of the backend. The class enables the intuitive assembly, execution, and processing of the results of database commands. The advantage is that when it is used correctly, user inputs are treated so that SQL injection cannot occur. Other reasons for choosing SQL Server are the clarity of the development environment, quality documentation, and extensive options for tuning the performance of database queries. This database will be used for storing downloaded or created code lists and also for storing data entered within the framework of individual solved projects.

A separate PostgreSQL database with the PostGIS extension will be used to store geographic objects. It will also contain a list of cadastral territories that will be used when downloading

geodata. The reason for this choice was primarily the quality of PostGIS geographical extension. Compared to the geographic part of Microsoft SQL Server, this offers a greater number of intuitive and useful functions that facilitate work with geodata.

4 RESULTS

The resulting application significantly facilitates the evaluation of the sustainability of crop production systems. Using forms, users can easily enter the necessary values. Some data are obtained automatically or based on the selection of a specific plot of land.

Thanks to the import of data from publicly available sources, the user is offered a map from which he can select the corresponding plot, as can be seen in Figure 3.

In this way, all plots can be easily added. The indicators resulting from the information about the land are then automatically added, which saves a lot of work.

After entering any input value, a page presenting the evaluation of the company is made available to the user. Since the input values can be entered continuously, it is possible to display partial results on the page. All available obtained indicator values and a list of indicators for which the rating could not be calculated are presented.

For the presentation of the results of the indicators, a spider chart is recommended as part of the methodology. This display method was also used within the created application. A screenshot of the page with the resulting graph is in Figure 4.

Within the presented graph, two data sets are used, this is the limit of sustainability (0.75 for all indicators) and the calculated values of the indicators. The color and size of the displayed

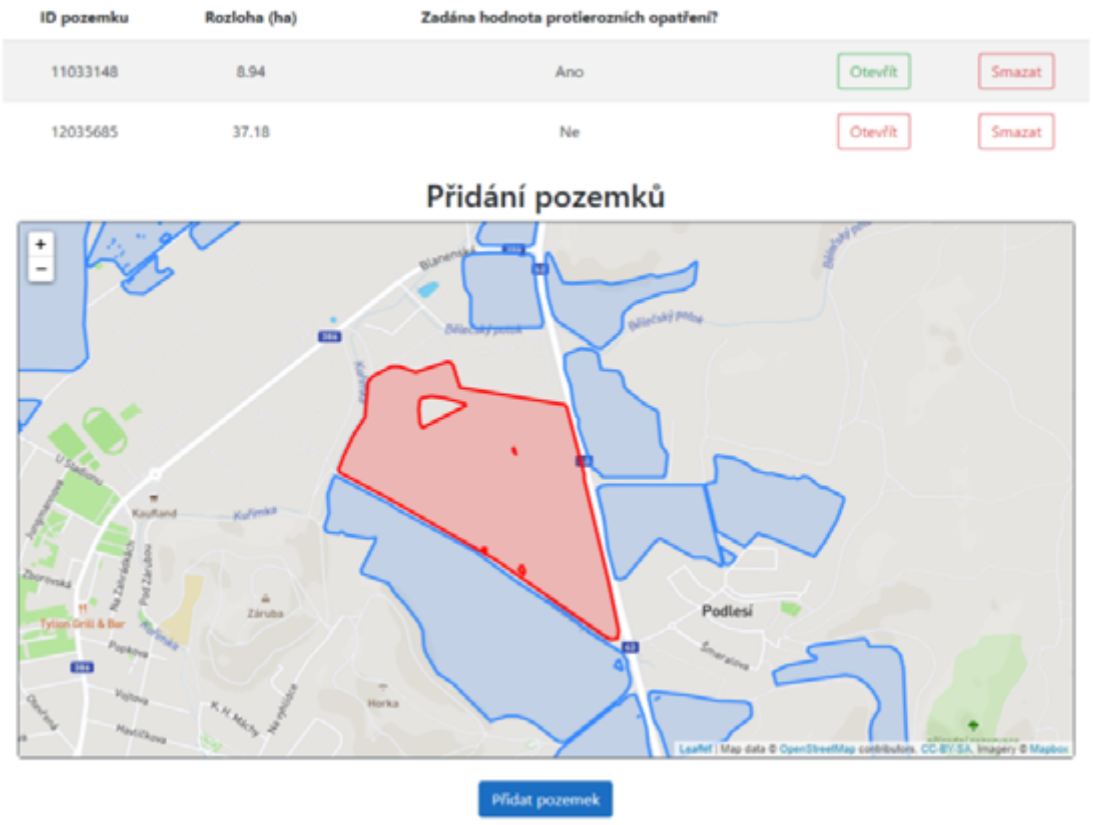


Fig. 3: Figure 3: Choosing plots in the web application
Source: Krejsa, 2022

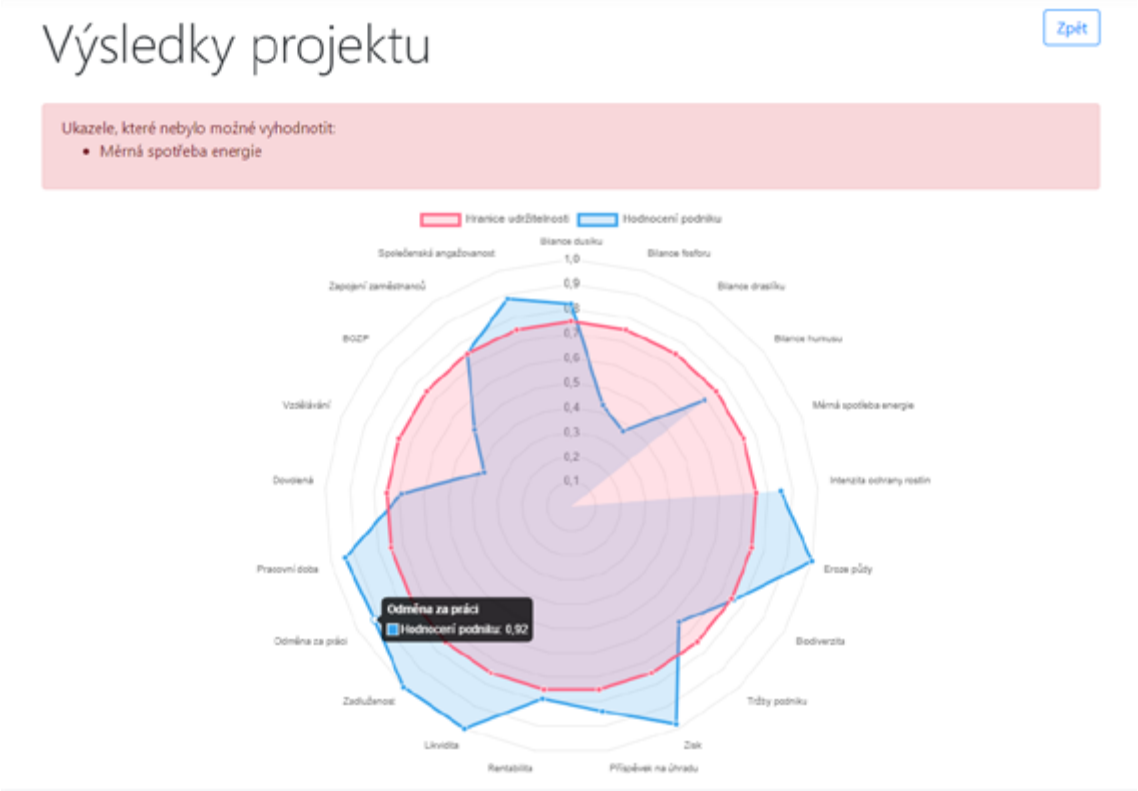


Fig. 4: Figure 4: Evaluation results
Source: Krejsa, 2022

elements in the chart can also be set within the data sets. A scale from 0 to 1, in increments of 0.1, is suitable for presenting the results. Incorrect values are omitted when plotting the business rating. The graph offers a degree of interactivity. When the cursor is placed on a point, the name of the indicator, the group of values, and the specific calculated value will be displayed.

5 DISCUSSION AND CONCLUSIONS

As part of the work, a web application was created that will enable the continuous loading of inputs for the calculation of the sustainability assessment of agricultural enterprises. The application can present results in a very similar format as mentioned in the methodology. The outputs of the application could be improved in the future. Specifically, this is the export of the specified input parameters. Whether it is an export in text form, which will be easy to read for ordinary users, or an export in JSON or XML format, which can be easily machine processed. The application is currently used mainly in education, if the application proves itself, it may also eventually be used commercially. This again could mean another set of feedback and possible modifications, according to the requirements of future users. It could be also beneficial to create a mobile application such as Prichystal (2016).

Acknowledgements
This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ARLAUSKIENĖ, A. and ŠARŪNAITĖ, L. 2023. Cover Crop Yield, Nutrient Storage and Release under Different Cropping Technologies in the Sustainable Agrosystems. *Plants*, 12(16), 2966. https://doi.org/10.3390/plants12162966

CELICOURT, P., ROUSSEAU, A. N., GUMIERE, S. J. and CAMPORESE, M. 2021. Hydro-informatics for sustainable water management in agrosystems. *Frontiers in Water*, 3, 758634. https://doi.org/10.3389/frwa.2021.758634

EAGRI - RESORTNÍ PORTÁL MINISTERSTVA ZEMĚDĚLSTVÍ. 2020. Veřejný export dat LPIS. *eAgri* [online]. https://eagri.cz/public/web/mze/farmar/LPIS/uzivatelske-prirucky/prirucky-pro-verejny-lpis/export-dat-lpis.html [cit. 2020-04-18].

EL CHAMI, D., DACCACHE, A. and EL MOUJABBER, M. 2020. How can sustainable agriculture increase climate resilience? A systematic review. *Sustainability*, 12(8), 3119. https://doi.org/10.3390/su12083119.

KREJSA V. 2022. *Webová aplikace na výpočet ukazatelů udržitelnosti systémů rostlinné produkce*. Diplomová práce. Mendelova univerzita v Brně.

KŘEN J., MARADA P., VALTÝNIOVÁ S., LIPAVSKÝ J., MÍŠA P. and SMUTNÝ V. 2011. *Metodika hodnocení trvalé udržitelnosti systémů rostlinné produkce pro podmínky ČR*. 1. vyd. Brno: Mendelova univerzita v Brně. 48 p. ISBN 978-80-7375-588-1

KŘEN J., MARADA P., VALTÝNIOVÁ S., LIPAVSKÝ J., MÍŠA P. and SMUTNÝ V. 2012. *Metodika získávání dat pro komplexní analýzy systémů rostlinné produkce*. 1. vyd. Brno: Mendelova univerzita v Brně. 46 p. ISBN 978-80-7375-588-1

PŘICHYSTAL, J. 2016. Mobile application for customers' reviews opinion mining. *Procedia-Social and Behavioral Sciences*, 220, 373-381. https://doi.org/10.1016/j.sbspro.2016.05.511

SOFO, A., MININNI, A. N. and RICCIUTI, P. 2020. Soil macrofauna: A key factor for increasing soil fertility and promoting sustainable soil use in fruit orchard agrosystems. *Agronomy*, 10(4), 456. https://doi.org/10.3390/agronomy10040456

XU, L., ZHU, Z. and SUN, D. W. 2021. Bioinspired nanomodification strategies: Moving from chemical-Based agrosystems to sustainable agriculture. ACS nano, 15(8), 12655-12686. https://doi.org/10.1021/acsnano.1c03948

ŽIVÁ UNIVERZITA - ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE. 2022. Jak podnikat trvale udržitelně? Pomůže nová mobilní aplikace. *Živá univerzita* [online]. https://zivauni.cz/jak-podnikat-trvale-udrzitelne-pomuze-nova-mobilni-aplikace/ [cit. 2023-12-18].

Contact information
Pavel Turčínek: e-mail: pavel.turcinek@mendelu.cz
Vojtěch Krejsa: e-mail: 72696@node.mendelu.cz

SURVEY OF LARGE LANGUAGE MODELS ON THE TEXT GENERATION TASK

Michaela Veselá[1], Oldřich Trenz[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

This paper focuses on the comparison of GPT, GPT-2, XLNet, T5 models on text generation tasks. None of the autoencoder models are included in the comparison ranking due to their unsuitability for text generation tasks. The comparison of the models was performed using the BERT-score metric, which calculates precision, recall and F1 values for each sentence. The median was used to obtain the final results from this metric. A preprocessed dataset of empathetic dialogues was used to test the models, which is presented in this paper and compared with other datasets containing dialogues in English. The tested models were only pre-trained and there was no fine-tune on the dataset used for testing. The transformers library from Hugging face and the Python language were used to test the models. The research showed on the pre-trained dataset empathic dialogues has the highest precision model T5, recall and F1 has the highest precision model GPT-2.

Keywords: natural language processing, auto-regressive transformers, large-scale model, natural language generation, decoder transformer, auto-encoding transformers, sequence to sequence model

JEL Code: C45, L86

1 INTRODUCTION

There has been a lot of progress in the area of large language models in the last few years. According to Tunstall et al. (2022), a major breakthrough in this area is taken to be the development of the Transformers model in 2017, which performed better than the previously used recurrent neural networks (RNNs) and algorithms such as (Stastny and Skorpil, 2007) or hybrid algorithms such as (Stastny et al., 2021), both in terms of machine translation tasks and even in training costs.

Another turning point in the development of large language models was the development of the BERT and GPT models, in 2018 (Tunstall et al., 2022). The BERT model falls into the group of models containing auto-encoding transformers (Devlin et al., 2018) and the GPT model contains auto-regressive transformers (Radford et al., 2018). The group of large language

https://doi.org/10.11118/978-80-7509-990-7-0195

models still includes models containing sequence to sequence transformers. Models containing this transformer are suitable for machine translation. While models containing auto-encoding transformer are suitable for text classification and models containing auto-regressive transformers are suitable for text generation tasks (Rahali and Akhloufi, 2023).

With the passage of time, improvements have been made to both the BERT model and the GPT model. The BERT model has been further modified under the names RoBERTa (Liu et al., 2019), ALBERT (Lan et al., 2019) or ELECTRA (Clark, 2020). The GPT model has retained its name and resorted to referring to versions of the models using numbers, such as GPT-2 (Radford et al., 2019), GPT-3 (Brown et al., 2020) and GPT-4 (OpenAI, 2023). The GPT model has also come to the attention of the general public due to the fact that OpenAI has made the model freely available to the public, who immediately took a great liking to it.

Currently, the biggest challenge in the field of large-scale language models, and especially for models designed for text generation tasks, is how to perform proper evaluation of the generated language (Thoppilan et al., 2022). The difficulty in evaluating models lies in the fact that for a properly generated language, a large number of aspects need to be addressed, ranging from grammar to coherence, and hence there is no simple way to perform natural language evaluation. One possible way is human annotation, but this is very expensive. Thus, the aim of this paper is to focus on the possible evaluation of language models designed for natural text generation.

2 METHODOLOGY

The methodology section introduces the datasets that contain the dialogues, the metrics used within this paper, and finally the models and their suitability for certain tasks.

2.1 Datasets

To evaluate models in the domain of conversation, datasets containing dialogues can be used. Since the goal is to explore models suitable for implementing a conversational client, datasets containing dialogs need to be selected. Due to the small number of datasets with dialogues in Czech, datasets in English will be used. These include the Ubuntu Dialogue Corpus dataset (Lowe et al, 2015), ProsocialDialog (Kim et al., 2022), SODA (Kim and Hessel, 2022), DailyDialog (Li, 2017) or DialogCC, which also contains images (Lee, 2022). In this case, the empathetic dialogues dataset will be used. The empathetic dialogues dataset contains 76.7 thousand rows of training data, 12 thousand rows of validation data and 10.9 thousand rows of test data (Li et al., 2020) This is a dataset that contains a transformers library designed for the python programming language. The dataset can be used for both text emotion classification and evaluation of generated text. For this reason, it was also selected because it contains emotions.

2.2 Metrics

To evaluate the generated text, it is also possible to use a large number of metrics. To evaluate the generated text, word-based metrics and contextual metrics can be used. Word-based metrics compare the words contained in the generated text and the reference text. While context-based metrics are mostly created as trained models for evaluation and their accuracy has been verified by human annotation (Sai, 2020). A lot of computational power is required when using context-based metrics. One of the possible metrics is BERT-score which is word based. The metric looks for a match using cosine similarity and takes into account the frequency of the document. Its great advantage is that it allows to take into account the importance of words (Zhang et al., 2019). It is a metric that is a good compromise between computational

power and the results it provides, so it will be used in the following paper. In order to use the BERT-score metric, the empathetic dialogues dataset needs to be modified and the reference text needs to be determined.

2.3 Models

The large language models that contain transformers can be divided into three groups, namely Autoregressive models which are suitable for generation tasks, sequence2sequence models which are suitable for machine translation and auto-encoder models which are suitable for text classification (Rahali and Akhloufi, 2023). Therefore, in order to compare the difference between the different types of models, two models that are not designed for text generation were also included in the comparison. These are the BERT model (Devlin, 2018) and the T5 model (Raffel et al., 2019). The models suitable for text generation tasks are XLNet (Yang et al., 2019), GPT (Radford et al., 2018), and GPT-2 (Radford et al., 2019). The models were selected based on the research conducted. Similar to the dataset, the Transformers library from Hugging Face can be used for the models. The library contains already pre-trained models that do not need to be trained further. If the user would still like to retrain the model for a specific type of task, fine-tuned functions are available.

3 RESULTS

Before testing the large language models, it is necessary to prepare the dataset. The empathetic dialogues dataset contains the conversation id (conv_id), utterance id (utterance_idx), what is the context of the dialogue (context), prompt, speaker id (speaker_idx), utterance (utterance), selfeval and tags. A sample dataset is shown in Table 1.

Because the BERT-score metric needs both the generated text and the reference text. It is possible to take every other answer as a reference within a dialogue. It is possible to notice that it is not possible to divide speakers into reference and those whose utterance is marked as the basis for the generated text, for the reason that the same spec may appear in multiple dialogues. Another of the snags in using this dataset can occur if there are an odd number of utterances in a conversation, then it is possible that the model will start generating a response even to an utterance where no response is desired. Should this case occur, it is possible to set the reference response to an empty string. The last situation that could occur is that the model would answer correctly, however it does not match the reference text. Since the Bert-score metric is used, which also takes into account the importance weight of each word, this problem should be avoided by the metric used. Because the crucial words are always repeated in the answers. In the original article, the values obtained from the BERT-score are set as the median (Zhang et al., 2019). This article uses the median to determine precision, recall and f1 median because the BERT-score metric at the point when the dialogue did not continue and thus the person would not respond to the message, but the model generated the text, the metric calculated all the resulting values as zero. So, what happened was that the resulting values contain the extreme outlier of zero, so it is more convenient to use the median.

4 DISCUSSION AND CONCLUSIONS

Testing has shown that T5 has the best results in precision. Recall and F1 has the best GPT-2 though. Sivarajkumar and Wang (2022) also tested GPT-2 and T5 but on classification task and they came out with precision for GPT-2 ranging from 0.73 to 0.50 and recall for T5 ranging from 0.64 to 0.71. 57 to 0.75 and for the T5 ranged from 0.61 to 0.72, for the F1 model the GPT-2 ranged from 0.55 to 0.74 and for the T5 ranged from 0.65 to 0.71. From the above

conv_id	utterance_idx	context	prompt	speaker_ idx	utterance	selfeval	tags
hit:1_conv:2	1	afraid	i used to scare for darkness	2	it feels like hitting to blank wall when i see the darkness	4|3|4_3|5|5	
hit:1_conv:2	2	afraid	i used to scare for darkness	3	Oh ya? I don't really see how	4|3|4_3|5|5	
hit:1_conv:2	3	afraid	i used to scare for darkness	2	dont you feel so.. its a wonder	4|3|4_3|5|5	
hit:1_conv:2	4	afraid	i used to scare for darkness	3	I do actually hit blank walls a lot of times but i get by	4|3|4_3|5|5	
hit:1_conv:2	5	afraid	i used to scare for darkness	2	i virtually thought so.. and i used to get sweatings	4|3|4_3|5|5	
hit:1_conv:2	6	afraid	i used to scare for darkness	3	Wait what are sweatings	4|3|4_3|5|5	
hit:1_conv:3	1	proud	I showed a guy how to run a good bead in welding class and he caught on quick.	3	Hi how are you doing today	3|5|5_4|3|4	<HI>
hit:1_conv:3	2	proud	I showed a guy how to run a good bead in welding class and he caught on quick.	2	doing good.. how about you	3|5|5_4|3|4	
hit:1_conv:2	1	afraid	i used to scare for darkness	2	it feels like hitting to blank wall when i see the darkness	4|3|4_3|5|5	

Tab. 1 Empathetic dialogues dataset pattern
Source: Li, Q. et al., 2020

results, it can be seen that Sivarajkumar and Wang (2022) also came out that the T5 model has better precision than GPT-2 in one case, but otherwise it can be generally said that GPT-2 shows better results than T5. Khaliq et al (2022) compared T5 and GPT using BLEU score, where it came out that T5 has better results than GPT. However, this metric is intended for machine translation evaluation (Papineni et al., 2002). A comparison of XLNet was found only with BERT and RoBERTa models, where XLNet showed better results than BERT in most cases. (Liu, 2019) However, a comparison that focused strictly on evaluating the generated text of each model was not found. The research conducted above found that the evaluation results of all the models tested were higher than 62%. It showed that although the T5 model has

Models	precision	recall	f1	Numbers of parameters
XLNet base	0.626486	0.705383	0.665162	110 million
GPT2	0.649675	0.717464	0.684829	1 500 million
T5-base	0.6751	0.682623	0.680336	220 million
GPT	0.633322	0.711081	0.671812	117 million

Tab. 2 Model evaluation results
Source: numbers of parameters: Radford, Alec, et al., 2019; Jangir, S. 2021; Raffel, C. et al., 2019, Nguyen-Mau, T., 2024

significantly fewer parameters than the GPT-2 model, when precision is used, the T5 model performs better than the GPT-2.

Acknowledgements

The publication was supported by the IGA doctoral project (Internal Grant Agency of the PEF) and specifically by the project IGA-PEF-DP-23-023.

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

REFERENCES

BROWN, T. B., MANN, B., RYDER, N., SUBBIAH, M., KAPLAN, J., DHARIWAL, P., NEELAKANTAN, A., SHYAM, P., SASTRY, G., ASKELL, A., AGARWAL, S., HERBERT-VOSS, A., KRUEGER, G., HENIGHAN, T., CHILD, R., RAMESH, A., ZIEGLER, D. M., WU, J., WINTER, C. ... and AMODEI, D. 2020. Language Models are Few-Shot Learners (Version 4). *arXiv*. https://doi.org/10.48550/ARXIV.2005.14165

CLARK, K., LUONG, M.-T., LE, Q. V. and MANNING, C. D. 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators (Version 1). *arXiv*. https://doi.org/10.48550/ARXIV.2003.10555

DEVLIN, J., CHANG, M.-W., LEE, K. and TOUTANOVA, K. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Version 2). *arXiv*. https://doi.org/10.48550/ARXIV.1810.04805

JANGIR, S. 2021. *Finetuning BERT and XLNet for Sentiment Analysis of Stock Market Tweets using Mixout and Dropout Regularization*. Technological University Dublin. https://doi.org/10.21427/K0YS-5B82

KHALIQ, Z., FAROOQ, S. U. and KHAN, D. A. 2022. A deep learning-based automated framework for functional User Interface testing. *Information and Software Technology*, 150, 106969. Elsevier BV. https://doi.org/10.1016/j.infsof.2022.106969

KIM, H., HESSEL, J., JIANG, L., WEST, P., LU, X., YU, Y., ZHOU, P., BRAS, R. L., ALIKHANI, M., KIM, G., SAP, M. and CHOI, Y. 2022. SODA: Million-scale Dialogue Distillation with Social Commonsense Contextualization (Version 3). *arXiv*. https://doi.org/10.48550/ARXIV.2212.10465

KIM, H., YU, Y., JIANG, L., LU, X., KHASHABI, D., KIM, G., CHOI, Y. and SAP, M. 2022. ProsocialDialog: A Prosocial Backbone for Conversational Agents (Version 2). *arXiv*. https://doi.org/10.48550/ARXIV.2205.12688

LAN, Z., CHEN, M., GOODMAN, S., GIMPEL, K., SHARMA, P. and SORICUT, R. 2019. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations (Version 6). *arXiv*. https://doi.org/10.48550/ARXIV.1909.11942

LEE, Y.-J., KO, B., KIM, H.-G. and CHOI, H.-J. 2022. DialogCC: Large-Scale Multi-Modal Dialogue Dataset (Version 1). *arXiv*. https://doi.org/10.48550/ARXIV.2212.04119

LI, Q., LI, P., REN, Z., REN, P. and CHEN, Z. 2020. Knowledge Bridging for Empathetic Dialogue Generation (Version 3). *arXiv*. https://doi.org/10.48550/ARXIV.2009.09708

LI, Y., SU, H., SHEN, X., LI, W., CAO, Z. and NIU, S. 2017. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset (Version 1). *arXiv*. https://doi.org/10.48550/ARXIV.1710.03957

LIU, Y., OTT, M., GOYAL, N., DU, J., JOSHI, M., CHEN, D., LEVY, O., LEWIS, M., ZETTLEMOYER, L. and STOYANOV, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach (Version 1). *arXiv*. https://doi.org/10.48550/ARXIV.1907.11692

LOWE, R., POW, N., SERBAN, I. and PINEAU, J. 2015. The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems (Version 3). *arXiv*. https://doi.org/10.48550/ARXIV.1506.08909

NGUYEN-MAU, T., LE, A.-C., PHAM, D.-H. and HUYNH, V.-N. 2024. An information fusion based approach to context-based fine-tuning of GPT models. *Information Fusion*, 104, 102202. Elsevier BV. https://doi.org/10.1016/j.inffus.2023.102202

OPENAI, ACHIAM, J., ADLER, S., AGARWAL, S., AHMAD, L., AKKAYA, I., ALEMAN, F. L., ALMEIDA, D., ALTENSCHMIDT, J., ALTMAN, S., ANADKAT, S., AVILA, R., BABUSCHKIN, I., BALAJI, S., BALCOM, V., BALTESCU, P., BAO, H., BAVARIAN, M. ... ZOPH, B. 2023. GPT-4 Technical Report (Version 4). *arXiv*. https://doi.org/10.48550/ARXIV.2303.08774

PAPINENI, K. et al. 2002. Bleu: a method for automatic evaluation of machine translation. In: *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, p. 311–318.

RADFORD, A., NARASIMHAN, K., SALIMANS, T. and SUTSKEVER, I. 2018. *Improving language understanding by generative pre-training*. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

RADFORD, A., WU, J., CHILD, R., LUAN, D., AMODEI, D. and SUTSKEVER, I. 2019. *Language Models are Unsupervised Multitask Learners*. https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

RAFFEL, C., SHAZEER, N., ROBERTS, A., LEE, K., NARANG, S., MATENA, M., ZHOU, Y., LI, W. and LIU, P. J. 2019. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer (Version 4). *arXiv*. https://doi.org/10.48550/ARXIV.1910.10683

RAHALI, A. and AKHLOUFI, M. A. 2023. End-to-End Transformer-Based Models in Textual-Based NLP. *AI*, 4(1), 54–110. MDPI AG. https://doi.org/10.3390/ai4010004

SAI, A. B., MOHANKUMAR, A. K. and KHAPRA, M. M. 2020. A Survey of Evaluation Metrics Used for NLG Systems (Version 2). *arXiv*. https://doi.org/10.48550/ARXIV.2008.12009

SIVARAJKUMAR, S. and WANG, Y. 2022. HealthPrompt: A Zero-shot Learning Paradigm for Clinical Natural Language Processing (Version 1). *arXiv*. https://doi.org/10.48550/ARXIV.2203.05061

STASTNY, J. and SKORPIL, V. 2007. Analysis of Algorithms for Radial Basis Function Neural Network. In: *Personal Wireless Communications*. Springer New York, vol. 245, pp. 54-62, ISSN 1571- 5736, ISBN 978-0-387-74158-1, WOS:000250717300005.

STASTNY, J., SKORPIL. V., BALOGH, Z. and KLEIN, R. 2021. Job shop scheduling problem optimization by means of graph-based algorithm. *Applied Sciences*, 11(4), 1921. ISSN 2076-3417. https://doi.org/10.3390/app11041921

THOPPILAN, R., DE FREITAS, D., HALL, J., SHAZEER, N., KULSHRESHTHA, A., CHENG, H.-T., JIN, A., BOS, T., BAKER, L., DU, Y., LI, Y., LEE, H., ZHENG, H. S., GHAFOURI, A., MENEGALI, M., HUANG, Y., KRIKUN, M., LEPIKHIN, D., QIN, J. ... and LE, Q. 2022. LaMDA: Language Models for Dialog Applications (Version 3). *arXiv*. https://doi.org/10.48550/ARXIV.2201.08239

TUNSTALL, Lewis, WERRA, Lenadro von, WOLF, Thomas and GÉRON, Aurélien. 2022. *Natural language processing with transformers: building language applications with Hugging Face*. Revised edition. Beijing: O'Reilly. ISBN 978-1-098-13679-6

YANG, Z., DAI, Z., YANG, Y., CARBONELL, J., SALAKHUTDINOV, R. and LE, Q. V. 2019. XLNet: Generalized Autoregressive Pretraining for Language Understanding (Version 2). *arXiv*. https://doi.org/10.48550/ARXIV.1906.08237

ZHANG, T., KISHORE, V., WU, F., WEINBERGER, K. Q. and ARTZI, Y. 2019. BERTScore: Evaluating Text Generation with BERT (Version 3). *arXiv*. https://doi.org/10.48550/ARXIV.1904.09675

ZIEGLER, D. M., WU, J., WINTER, C., ... and AMODEI, D. 2020. Language Models are Few-Shot Learners (Version 4). *arXiv*. https://doi.org/10.48550/ARXIV.2005.14165

Contact information

Michaela Veselá: email: xvesela@mendelu.cz
Oldřich Trenz: email: oldrich.trenz@mendelu.cz

ENHANCING MICRO-CREDENTIALS WITH BLOCKCHAIN

Martin Základsník[1], Veronika Konečná[1], Oldřich Faldík[1], Oldřich Trenz[1], Andrej Gono[1]

[1]Department of Information Technology, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

ABSTRACT

The article addresses the problem of using micro-credentials in the educational process and explores the possibilities of their deployment through blockchain technology. The topic of micro-credentials as the first step in the process of digitalization of the educational process is presented along with the setting of its trustworthiness. A variety of advantages can be associated with micro-credentials, including the confirmation of individual transactions at the level of the educational process.

One of the main problems in the centralized storage of micro-credentials is the risk of unauthorized access and the possibility of leakage of sensitive information. This paper proposes the implementation of blockchain technology as a way to decentralize data storage. This would eliminate the threat of unauthorized access and provide a higher level of data security and integrity.

Challenges associated with a centralized certificate authority such as scalability issues and outages are also discussed. It can be evaluated that blockchain can provide a robust and reliable framework for digitizing certificates in the education sector.

The conclusions of the paper highlight the benefits of decentralization through blockchain and the need to open up the certification network for corporate certificates. Overall, the paper discusses the importance and benefits of using blockchain technology to enhance the security and efficiency of digital certificates in the education sector.

Keywords: blockchain, digitalization, micro-credentials

JEL Code: L8, O3

1 INTRODUCTION

Digitalization is one of the most discussed topics in recent years. Every industry is affected by digitalization in some way. Within the education sector, many certificates or diplomas are being issued, which currently exist mainly in paper form. The disadvantage of paper certificates is that verifying their authenticity can be complicated and they are therefore relatively easy to falsify. Proving a paper certificate is also complicated by the difficulties of certified copies and so on. The solution may be to digitalize these certificates.

https://doi.org/10.11118/978-80-7509-990-7-0201

The first step of digitalization in education is the digitalization of individual successfully completed courses, the micro-credentials. Micro-credentials have several advantages. They provide students with a visual and tangible record of their skills or achievements. By taking inspiration from gaming badges, they can create an element of competition and reward that is typical in a gaming environment. Earning a badge or micro-certificate can be seen as a reward for achieving a specific goal or skill (Gish-Lieberman, 2021). These certificates can also then be used to prove completion of a course without the student completing the entire study programme.

Another key step in the process of digitalization of education is the storage of digital certificates in a central database. While this offers advantages in accessibility and management, it brings with it security challenges. A central database may face the risk of unauthorized access or leakage of sensitive information contained in digital certificates. In addition, if this database is down or unavailable, authentication of certificates can be complicated (Hada, 2023).

One possible solution to these challenges may be the implementation of blockchain technology. Blockchain offers a decentralized way of storing data, which can eliminate the risk of unauthorized access or manipulation of certificates. This would ensure a higher level of security and data integrity. Moreover, due to decentralization, it would be possible to authenticate certificates even in the event of central system failures. Thus, the implementation of blockchain could provide a robust and reliable framework for digitizing certificates in education (Zaklasnik, 2023).

The aim of the article is to outline the possibilities of extending the existing platform (Masaryk University, 2023) for issuing micro-credentials created by Masaryk University and Charles University with the use of blockchain technology. The Masaryk University provides a central verification point and storage place for micro-credentials and Charles University provides a central catalogue of courses for which can be issued micro-credentials issued. This approach has several limitations related to scalability, openness and resiliency. (Charles University, 2024)

Proposed solution will ensure decentralization and higher security of the whole system and the possibility of involving other educational institutions and companies, which will create a unique unified platform in which the user can have all the certificates from all different institutions. The A sub-goal is then to introduce the JSON-LD data format for micro-credentials and the process of issuing and storing micro-credentials data on the blockchain.

2 EXISTING SOLUTIONS

Typically, the existing solutions are possible to integrate with own Learning Management Systems (LMS) and main publishing platforms, meaning they can be easily integrated into university systems. They offer the possibility to share credentials on social platforms such as LinkedIn, Facebook, X and others, as well as manage the public profiles for both issuer and recipient and download the credential in various formats. The solutions often provide data analytic tools for monitoring the issuance of certificates, as well as their performance and impact. They are usually well scalable and being offered to small, as well as enterprise-sized institutions. The blockchain layer is often included in a broader subscription model.

2.1 Credly

Credly enables users to issue, manage and claim digital credentials. To claim a credential on Credly, the recipient must create an account. Issuing digital credentials on Credly is facilitated through the paid dedicated Acclaim platform. The issuance can be one-time manual or automated, using Credly's API integrated into existing systems. In the case of Credly, blockchain is not the underlying platform but rather an extension available as part of the Acclaim subscription. To publish the credential on the blockchain, first, the issuer much enable this possibility and then it must be approved by the receiver as well or the receiver can claim the credential without the blockchain verification (Credly, 2024).

2.2 Sertifier

Sertifier provides digital certificates and badge solutions for startups, small to medium-sized businesses, and enterprises. Their certificate offerings range from product certificates and online courses to employee training. Among their clients are Paypal, Cisco, Coachhub, Johnson & Johnson, and others, with over 8 million credentials issued. Sertifier operates in a hybrid mode, combining traditional storage and blockchain technologies. Two out of three subscription plans include blockchain features, providing enhanced security, authentication, and streamlined verification processes to prevent data tampering. They use the robust blockchain platform Velocity Network™ to ensure the integrity of every credential and enable real-time verification anywhere and anytime. Moreover, the Verified Wallet is necessary for storing credentials.

By digitally signing the credentials, the issuer guarantees their authenticity. The credentials are then issued and stored on the blockchain. The recipient must confirm the credential accuracy, claim it, and add it to their own blockchain-verified portfolio (Sertifier, 2024).

2.3 CredSure

CredSure is credentialing platform utilized by over a million users worldwide. Is uses blockchain for the credential's verification and stores all the credentials on the blockchain. They primarily focus on eliminating the issues connected to paper certificates and offer a platform for the translation of paper docs to digitalized files.

They enable credential collection, data evaluation, or ensuring trust. In the case of CredSure, both issuer and receiver of the credentials need their blockchain wallet where the credentials will be stored.

Users and verifiers can access the credential using unique URL or QR code on the verification page, that displays the issued course or product. CredSure also enables automation of the issuance and sending of credentials and badges, while providing technical support (Credsure, 2024).

2.4 EvidenZ

EvidenZ is a framework for blockchain verification of credentials, primarily focusing on higher education. Currently, 232 institutions across 25 countries are involved. It prioritizes user-friendliness and supports multiple devices and can operate on any blockchain.

For the issuance of every certificate, the essential part of the process is burning a part of the BCDT token. Among others, EvidenZ collaborates with Avalanche, Binance, and uses aleph.im for data storage on a secure and decentralized network. EvidenZ stores only encrypted data on the blockchain, rather than the actual documents or their hash fingerprints. Before issuing the data, it first verifies the identity of data issuers to establish trust. It is also convenient that no registration is needed to claim the credential, and owning a blockchain wallet is not required (although an email confirmation is needed to claim the credential). This is because EvidenZ uses shareable URLs for permanent access to the credential (EvidenZ, 2024).

3 METHODOLOGY AND DATA

The authors Selvaratnam and Sankey (2021) identify a number of disadvantages of paper certificates in the context of micro-credentials. Paper certificates are static, easy to falsify, difficult to share, and lacking in detailed information about the qualifications achieved. In addition, the production of paper certificates has a negative impact on the environment. In contrast, digital micro-credentials offer flexibility, robust security, ease of sharing, detailed information and environmental friendliness. The authors highlight that digital micro-credentials are

better suited to modern trends in education and the labour market and represent a more efficient and modern alternative to paper certificates.

3.1 Current micro-service infrastructure

The current micro-credentials system at Masaryk University integrates several technological components to ensure efficient course management, issuance, and verification of micro-credentials.

The system issues micro-certificates in electronic format (JSON-LD) following European Digital Credentials (EDC) specifications. Each micro-certificate contains essential details about the learner‘s achievements, issuer information, and validity period. If incorrect data is found post-issuance, the certificate can be invalidated, marking it as invalid and disabling its verification links without deleting it from the database.

Issued micro-certificates are accessible to recipients through digital wallets or direct email. They are not stored in a public database, preserving privacy and security. Verification is facilitated by a web-based tool, EDCI Viewer, allowing third parties to upload the micro-certificate file for validation, thus ensuring integrity and authenticity.

The system generates a PDF version of the micro-certificate, including a verification link. This representation meets standard requirements and can be digitally signed by the issuing institution for additional authenticity. Utilizing RESTful API for secure data transfer over HTTPS, the system ensures all API requests are authenticated, allowing only authorized clients to access or modify data.

A central database stores all issued micro-certificates and their metadata, enabling easy retrieval and management. The system supports automated updates and maintenance of certificate validity, including the renewal of electronic seals. There is also web interface, which caters to different user groups, including public verifiers, administrative personnel, and recipients of micro-certificates. This simplifies the processes of certificate issuance, management, and verification.

We aim to achieve three main goals in the extension of Masaryk University‘s existing platform for issuing micro-credentials with blockchain technology:

1. Decentralization and increased security: by using blockchain technology, the platform will move from a centralized model to a decentralized one. This transition increases the overall security of the system because the data is distributed across a network of nodes, making it resistant to tampering and unauthorized changes.
2. Openness and scalability: implementation of the blockchain simplifies the involvement of new entities outside Masaryk University. This allows other educational institutions and companies to join the platform, which supports collaboration and creates a uniform ecosystem for micro-credentials.
3. Standardization: JSON-LD (JavaScript Object Notation for Linked Data) is the W3C standard used for micro-credentials. JSON-LD offers a standardized and machine-readable format that is not dependent on any one platform or system.

This article uses both qualitative and quantitative research approaches. Within the qualitative analysis, it investigates existing literature related to blockchain technology in the context of educational certificates, specifically micro-credentials. The quantitative analysis then examines the technical aspects of blockchain implementation. This includes an analysis of existing blockchain platforms and protocols suitable for issuing and verifying micro-credentials.

Blockchain technology was selected for a decentralized approach of issuing certificates, with emphasis on speed, energy efficiency and security. Proof of Authority (PoA) algorithm was preferred for its speed, proven reliability and wide use in private blockchains.

The feasibility of PoA as a consensus algorithm in blockchain networks have been explored by Joshi (2021), who emphasizes the need for security, reliability, and speed in such

algorithms. Ali, Sahib and Waleed (2019) further discuss the use of PoA in preserving authentication and authorization on the blockchain, highlighting its potential in addressing the challenges of big data size and verification time. Alrubei, Ball and Rigelsford (2021) proposes a novel consensus mechanism, Honesty-based Distributed Proof-of-Authority (HDPoA), which combines PoA and Proof-of-Work (PoW) to enhance security and reduce confirmation time in IoT-blockchain applications. Lastly, Khalil, Aziz and Asif (2021) suggest the use of PoA in the Ethereum blockchain platform for charitable organizations, emphasizing its potential to enhance trust and transparency.

Trusted authority-formed voters were selected as participants in the network, contributing by issuing and verifying certificates. The evaluation of different blockchains led to the preference of Proof of Authority over other algorithms to achieve optimal security and efficiency results. Considering privacy and legal issues, an analysis of challenges related to the right to be forgotten and certificate updates was also considered.

Limitations

This paper only covers the design concept for the platform extension. Full implementation of the proposed solution would require further research and development, including pilot testing.

4 RESULTS

Masaryk University uses JSON-LD format for their micro-credentials platform.

JSON-LD (JSON for Linked Data) provides a standardized format for structuring and representing data (W3C, 2022). Among the advantages is that JSON-LD is human readable. Each micro credential can be used to record all certificate data such as student identity, course completion data, number of credits, etc. This structured data allows for efficient processing across multiple systems and platforms, including integration with a blockchain platform. Below is an example of what a micro-credential looks like in JSON-LD format.

```
{
 “@context”: [
   “https://www.w3.org/2018/credentials/v1”,
   “https://www.w3.org/2018/credentials/examples/v1”
 ],
 “id”: “http://example.edu/credentials/58473”,
 “type”: [“VerifiableCredential”, “AlumniCredential”],
 “issuer”: “https://example.edu/issuers/565049”,
 “issuanceDate”: “2010-01-01T00:00:00Z”,
 “credentialSubject”: {
   “id”: “did:example:ebfeb1f712ebc6f1c276e12ec21”,
   “alumniOf”: {
     “id”: “did:example:c276e12ec21ebfeb1f712ebc6f1”,
     “name”: [{
       “value”: “Example University”,
       “lang”: “en”
     }, {
       “value”: “Exemple d’Université”,
       “lang”: “fr”
     }]
   }
 },
 “proof”: { ... }
}
```

Fig. 1: Verifiable Credentials Data Model v1.1.
Source: W3C, 2022

A decentralized approach provided by blockchain could make the process of issuing certificates faster and not dependent on one central authority, when it comes to certificates that require human input and aren't issued automatically by digital authority.

When choosing a blockchain suitable for this use-case, we identified multiple key requirements. To be proven beneficial over typical centralized databases, it should be fast enough, energy efficient and secure, while allowing decision-making for a selected group of members. This approach prefers private blockchains and specific consensual algorithms. It eliminates, for example, widely spread Proof of Work, that is secure, but slow and energy consuming. It also makes Proof of Stake to be questionable, potentially allowing one party to own more voting power, unless this concern is addressed at the beginning. (Xiao et al., 2020) We also intend to avoid experimental and untested solutions. As a suitable solution now, it seems to be a Proof of Authority consensual algorithm, as it is fast, tested, not too complex for implementation and used widely in private blockchains. The voters are selected, trustworthy authorities, and as motivation could serve their reputation and right to stay in the ecosystem, instead of transaction fees (Azbeg, 2021. Khalil, 2021). The voters would be companies and academic institutions involved in the network, issuing certificates and validating those of theirs, as well as accepting new members into the network. This way, every new incoming authority will gain a voting right and the ecosystem will be naturally scaled-up.

As we try to think about a long-lasting solution, it is worth considering the infrastructure to be quantum resilient, as the academic sector could be prone to possible future quantum attacks. A lot of blockchains aren't easily adjustable and don't implement quantum resistant algorithms at the moment (Yang, 2022).

Although the certificate owners would be able to own the certificate and decide whether to make it private or not, there are still privacy and legal issues in this aspect. A potential challenge is the whole legal framework itself, which isn't well established for blockchain yet. Blockchain nature could challenge some legal requirements, such as the right to be forgotten defined by GDPR, that contradicts with the immutable nature of blockchain (Wolford, 2023). A subsequent issue is the need to update the certificate, for example in case it becomes invalid or was issued wrong. This particular aspect of changing policies should be discussed prior, when setting up the blockchain rules. In this case it should be noted that changes into blockchain transactions are very often hard to implement and could introduce more vulnerabilities.

Another potential issue could be initial investment into the new infrastructure, which should be calculated if it is worth the potential advantages. On the other hand, blockchain eliminates the single point of failure risk. The centralized databases, of course, use replicas and back-ups, that usually stays more storage efficient than blockchain (as every involved party needs to have their own blockchain replica), although the centralized databases also need to be managed, giving space to more failures and security issues. Therefore, the proposed solution of blockchain infrastructure should also address ways to optimize storage efficiency.

The process of creating and distributing micro-credentials in an educational institution involves several steps. The core components of the system include the educational institution, a JSON-LD generator, a cryptographic hashing function, a blockchain network, a QR code generator, and a student application.

The blockchain network uses the Proof of Authority consensus algorithm, chosen for its transaction processing capabilities and minimal energy consumption. Validators in this network are trusted entities, such as academic institutions ensuring data integrity and security. Smart contracts are utilized to automate the issuance, verification, and revocation of credentials on the blockchain.

Each micro-credential is structured in the JSON-LD format, enabling seamless integration with the blockchain. This standardization supports the representation of detailed credential data, including student identity, course completion data, and credit count. Credentials are issued in both digital format for blockchain records and as downloadable PDFs for offline use. Each PDF certificate includes a QR code that links back to the blockchain record for verification purposes.

All credential data are hashed using cryptographic functions (SHA-256) and stored on the blockchain, safeguarding against tampering and unauthorized alterations. Verification of credentials can be performed through decentralized nodes, ensuring availability and reliability even if a central system fails.

Important part of the system are RESTful APIs to connect the blockchain with existing Learning Management Systems (LMS) and the Student Information System (SIS). These APIs facilitate secure data exchanges and real-time updates of credential statuses.

The practical process then starts, when the student completes a course or training for which the student can earn a micro-credential. After successful course completion, students can opt for a digital version of their diploma. If they choose digitization, their consent must be obtained and stored in line with GDPR regulations. With consent, the institution can generate a "Verifiable Credential" (digital diploma adhering to W3C standards) containing relevant information such as course title, student name, completion date and other metadata. The student can then download a micro-credential from the institution via a secure online portal and store it in their digital wallet, which can be a software application on a mobile device (Základsník, 2023).

At the same time, a hash is generated using a cryptographic hash function, such as SHA-256, from the content of the certificate. A hash is a unique and fixed string of characters that represents the entire contents of a micro-credential. This hash is stored on the blockchain, which secures it from manipulation and allows independent verification of the micro-credential.

It is also possible to generate a PDF certificate. This certificate is used for offline presentation and contains all relevant information about the micro-credential, including the QR code. The QR code makes it easy to verify the validity of the certificate by scanning it with a smartphone or other device.

The image below shows the entire flow of generating and storing a micro-credential.

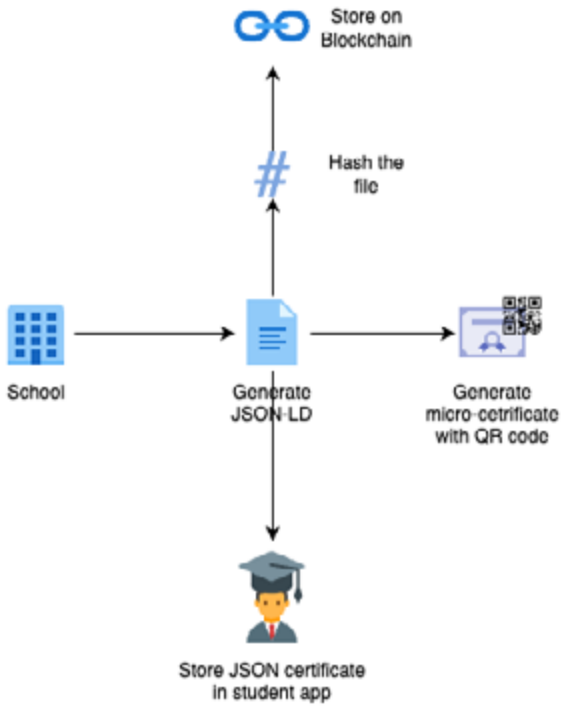


Fig. 2: The process of generating and storing micro--credentials on the blockchain, 2024

5 DISCUSSION AND CONCLUSIONS

Traditional paper certificates in education are prone to falsification and fraud as mentioned by the authors Selvaratnam and Sankey (2021). As well as that, students may find it difficult to prove the legitimacy of the certificates issued in different institutions.

In this paper, we research and propose the implementation of micro-credentials and blockchain for digitizing education. According to experts, the use of Blockchain technology in education is a potential use case (Casey, 2019. Clark, 2016).

Our proposed extension of Masaryk University's existing platform for issuing micro-credentials using blockchain technology offers several advantages over the current centralized system. By decentralizing the platform, we ensure the overall security and integrity of the micro-credentials issuance process, where data is protected from tampering and modification. Compared to similar existing solutions, our proposal does not issue credentials solely in digital form, but rather serves as an extension of paper certificates and both formats are supposed to coexist alongside each other. Additionally, having a private blockchain with a proof of authority algorithm grants involved parties a voting right a decision-making option, which could help our solution to evolve in alignment with academic sector needs

In addition, by implementing blockchain, we enable the involvement of new educational institutions beyond Masaryk University. In this way, we are creating a universal platform and ecosystem where students can have all the certificates they have achieved.

The future work will involve a comprehensive requirement analysis and system design identifying and engaging key stakeholders, including academic institutions, certification authorities, and potential employers.

Next step is setting up the blockchain network that supports the PoA consensus algorithm. The configuration of blockchain nodes for participating institutions will follow, ensuring each node is correctly integrated into the network and capable of issuing and verifying micro-credentials.

Next development will involve finalizing the JSON-LD format for micro-credentials. This format must include all necessary metadata and adhere to W3C standards, ensuring interoperability and ease of use. There will be an intuitive user interface for facilitating the issuance, management, and verification of micro-credentials by both issuers and recipients. Integrating the blockchain system with existing Learning Management Systems (LMS) at participating institutions will be another critical task.

Detailed testing will be conducted to ensure the system works as intended and is secure. Functional testing will verify that all system components operate correctly, while security testing will identify and mitigate potential vulnerabilities. A pilot deployment involving a limited number of courses and students will follow. The system's performance will be evaluated in terms of speed, scalability, and user satisfaction.

Acknowledgements

This paper was supported by the project CZ.02.1.01/0.0/0.0/16_017/0002334 Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education.

REFERENCES

ALI, W. A., SAHIB, N. M. and WALEED, J. 2019. Preservation Authentication and Authorization on Blockchain. In: *2019 2nd International Conference on Engineering Technology and its Applications (IICETA)*, pp. 83-88. https://doi.org/10.1109/IICETA.2019.8835346

ALRUBEI, S. M., BALL, E. and RIGELSFORD, J. M. 2021. Securing IoT-Blockchain Applications Through Honesty-Based Distributed Proof of Authority Consensus Algorithm. In: *2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, pp. 1-7. https://doi.org/10.1109/CyberSA52016.2021.9435428

AZBEG, K., OUCHETTO, O., JAI ANDALOUSSI, S. and LAILA, F. 2020. An Overview of Blockchain Consensus Algorithms: Comparison, Challenges and Future Directions. In: *Advances on Smart and Soft Computing*, pp. 357-369. https://doi.org/10.1007/978-981-15-6048-4_31

BLOCKCERTS. 2024. *Blockchain Credentials*. https://www.blockcerts.org/

CASEY, J. 2019. *My Skills Project* [White Paper]. IFI Charitable Trust. https://myskills.org.uk/wp-content/uploads/2019/03/My-Skills-White-Paper-25-3-19.pdf

CLARK, D. 2016. Ten ways Blockchain could be used in education. *oeb-insights* [online]. https://oeb.global/oeb-insights/10-ways-blockchain-could-be-used-ineducation/

CREDLY. 2024. *Digital Credentials*. https://info.credly.com/

CREDSURE. 2024. Certif-ID International GmbH. *CredSure* [online]. https://credsure.io/

DNV. 2024. Traceable trusted certificates. *www.dnv.com* [online]. https://www.dnv.com/assurance/certificates-in-the-blockchain/

EVIDENZ. 2024. *The ultimate blockchain digital credentialing technology*. https://www.evidenz.io/

GISH-LIEBERMAN, J. J., TAWFIK, A. and GATEWOOD, J. 2021. Micro-Credentials and Badges in Education: a Historical Overview. *TechTrends*, 65, 5-7. https://doi.org/10.1007/s11528-020-00567-4

ALSOBHI, Hada A., ALAKHTAR, Rayed A., UBAID, Ayesha, HUSSAIN, Omar K., HUSSAIN, Farookh Khadeer 2023. Blockchain-based micro-credentialing system in higher education institutions: Systematic literature review. *Knowledge-Based Systems*, 265, ISSN 0950-7051. https://doi.org/10.1016/j.knosys.2022.110238

CHARLES UNIVERSITY. 2024 *Zlepšení prostupnosti vzdělání na úrovni vysokých škol pomocí mikrocertifikátů*. https://cczv.cuni.cz/CCZV-572.html

CHARLES UNIVERSITY. 2023. *Micro-Credentials at Charles University*. https://cczv.cuni.cz/CCZV-572-version1-prirucka___micro_credentials_na_uk_eng.pdf

JOSHI, S. 2021. Feasibility of Proof of Authority as a Consensus Protocol Model. *ArXiv,* abs/2109.02480.

KHALIL, I., AZIZ, O. and ASIF, N. 2021. Blockchain and Its Implementation for Charitable Organizations. In: *2021 International Conference on Innovative Computing (ICIC)*, pp. 1-10.

MASARYK UNIVERSITY. 2023. Mikrocertifikáty ve vzdělávání. MU je zavádí jako jedna z prvních v Evropě. *MUNI* [online]. https://www.em.muni.cz/udalosti/16554-mikrocertifikaty-ve-vzdelavani-mu-je-zavadi-jako-jedna-z-prvnich-v-evrope

SELVARATNAM, Ratna Malar and SANKEY, Michael. 2021. An integrative literature review of the implementation of micro-credentials in higher education: Implications for practice in Australasia. *Journal of Teaching and Learning for Graduate Employability*, 12, 1-17. https://doi.org/10.21153/jtlge2021vol12no1art942

SERTIFIER. 2024. *Sertifier*. https://sertifier.com/

VARSHINEE, C., GEEANESWARI, N. and ROOPESH, S. 2021. The future of continuous learning–Digital badge and microcredential system using blockchain. *Global Transitions Proceedings*, 2, 355-361. ISSN 2666-285X. https://doi.org/10.1016/j.gltp.2021.08.026

WOLFORD, B. 2023. Everything you need to know about the "Right to be forgotten". *GDPR.eu* [online]. https://gdpr.eu/right-to-be-forgotten/

WORLD WIDE WEB CONSORTIUM–W3C. 2022. *Verifiable Credentials Data Model v1.1*. https://www.w3.org/TR/vc-data-model/

XIAO, Y., ZHANG, N., LOU, W. and HOU, Y. T. 2020. A Survey of Distributed Consensus Protocols for Blockchain Networks. *IEEE Communications Surveys & Tutorials*, 22(2), 1432-1465. https://doi.org/10.1109/COMST.2020.2969706

YANG, Z., SALMAN, T., JAIN, R. and PIETRO, R. 2022. Decentralization Using Quantum Blockchain: A Theoretical Analysis. *IEEE Transactions on Quantum Engineering*, 3, 4100716. https://doi.org/10.1109/TQE.2022.3207111

ZÁKLASNÍK, M., FARANA, R. and SURMA, S. 2023. Digitization of University Diplomas in the European Union. In: SILHAVY, R. and SILHAVY, P. (Eds.). S*oftware Engineering Research in System Science. CSOC 2023*. Lecture Notes in Networks and Systems, vol 722. Springer, Cham. https://doi.org/10.1007/978-3-031-35311-6_39

Contact information

Martin Zaklasnik: email: xzaklas1@mendelu.cz
Veronika Konecna: email: 463407@mail.muni.cz
Oldrich Faldik: email: oldrich.faldik@mendelu.cz
Oldrich Trenz: email: oldrich.trenz@mendelu.cz
Andrej Gono: email: xgono@mendelu.cz

26TH ANNUAL INTERNATIONAL CONFERENCE

ECONOMIC COMPETITIVENESS AND SUSTAINABILITY 2024

PROCEEDINGS

Editors: Petr David, Hana Vránová

Publisher: Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
Edition: First from 2024

ISBN 978-80-7509-990-7 (online ; pdf)
https://doi.org/10.11118/978-80-7509-990-7