

The 8th International Conference on Economics and Social Sciences
Exploring Global Perspectives:
The Future of Economics and Social Sciences
June 5-6, 2025
Bucharest University of Economic Studies, Romania

Understanding Regional Development Strategies in Romania: A Text Mining Approach

Alina Georgiana PROFIROIU¹, Sebastien BOURDIN²,
Daniela-Luminița CONSTANTIN¹, Ionuț NICA^{1*}, Ștefan IONESCU¹

DOI: 10.24818/ICESS/2025/033

Abstract

This study aims to tackle Romania's regional development strategies for the 2021-2027 timespan by using a text mining analysis to identify the main themes, but also the interconnections within policy documents. The conducted analysis is considering the regional development plan of each of the seven development regions, while for the South-West Oltenia region, a different approach that relies on the Smart Specialisation Strategy (RIS3) was used, since this region did not have any conventional development plan. To properly extract those terms that are the most frequent, but also the topic distributions, Natural Language Processing (NLP) techniques were ran against a Python context. The results of using these techniques stated the fact that themes such as "regional development", "infrastructure", "modernisation", "public services" and "economic growth" are the main themes detected in all of the considered documents, underscoring that there is a strong focus on sustainability, socioeconomic transformation and strategic planning. More than this, by considering the analysis that puts into perspective those regions that are developed, such as, for example, Bucharest-Ilfov and the North-East region, either distinct regional priorities or policy approaches can be proved. By doing this, we can enhance our understanding of the territorial disparities that may apply within a unified national strategy. With the help of the text mining techniques, dominant policies areas, like investment priorities in transportation, digitalisation and labour market development were identified. These findings underscore both the similarities, but also the divergences with regards to Romania's regional development strategies, basically pointing out the important insights that may be considered by either policymakers or researchers that are interested in the evaluation of data-driven public strategies. This research is able to contribute to the area of computational policy analysis by proving the usefulness of text mining techniques when it comes down to our understanding of the strategic documents that are rather large-scale, thus

¹ Bucharest University of Economic Studies, Bucharest, Romania, ionut.nica@csie.ase.ro.

² EM Normandie, Paris, France.

* Corresponding author.

assisting the decisions with regards to regional governance and economic planning to be made in a better-informed manner.

Keywords: place-based policies, local development and innovation, collaborative networks for local development, text mining, regional development strategies, natural language processing.

1. Introduction

When it comes to the regional development, it is important to mention the fact that this matter is a key factor in the context of shaping the social landscape and the economics of any country, Romania being no exception. Since the country consists in several diverse regions, considering their own unique strengths and weaknesses, adapted strategies approaches for both local and national levels were imperious (Bacter et al., 2024; Zamfir & Stoica, 2023; Rotaru et al., 2024). Recently, decision-making became more and more reliant on the data-driven outputs (Abir et al., 2024; Attah et al., 2024), since this approach is fundamental with regards to better understanding the factors that are influencing regional development (Vaz, 2024), whilst also being able to offer insights in the context of potential growth areas. Since the current traditional methods of regional analysis count mainly on quantitative data, the steady increase in the availability of text-related data from various sources, such as reports, policies or other official documents suggested new areas that can be taken into account in the context of understanding regional development strategies (Nielbo et al., 2024; Di Marino et al., 2025).

The current research applies text mining techniques in order to examine the development strategies of various regions in Romania. The main target of this study is to find those primary themes, priorities, and challenges in each region's development strategies with the help of an analysis over the textual data of the key regional documents. It is noteworthy that text mining represents an efficient data extraction tool, putting into perspective the key factors that are being considered in the decision-making process in regional development, from infrastructure and economic growth to social inclusion and environmental sustainability. By choosing this approach, our aim is to come up with a clearer image of how Romania's regions are evolving, but also to underscore what are those employed strategies that encourage an improved quality of life for the country's population and a more sustainable growth.

This study is centred around all of Romania's eight development regions in Romania, applying text mining techniques, such as word clouds and frequency analysis, to identify the most impactful themes and strategic priorities that were highlighted in the regional planning documents. With the help of these tools, the paper is able to provide a new point of view with regards to the regional development, emphasising the key policy areas that are shaping Romania's growth trajectory whilst offering important information about future development planning.

2. Literature Review

Recently, the convergence of text mining analysis towards various areas gained significant attention, as there is an increasing trend in studies related to areas like climate change, economic resilience, sentiment analysis, regional development, and social media research. This section reviews relevant literature that is focused over highlighting the usage of text mining and related techniques in across various areas, providing a more clearer picture about the key themes and emerging trends.

As the usage of machine learning (ML) approaches became more and more increasing, it is obvious that these approaches are starting to play a key role in addressing global challenges. Niculae et al. (2024) conducted a bibliometric analysis that showcased a 37.79% annual growth in climate-related researches that were making use of ML. With the assistance of techniques such as Latent Dirichlet Allocation (LDA), the study was able to pinpoint key thematic areas, such as agriculture, hydrology, biodiversity, and forecasting models, thus demonstrating the interdisciplinary nature of climate research. With regards to the Romanian consumer reviews, Briciu et al. (2024) performed a sentiment analysis that had its methodology revolving around several ML and deep learning techniques in order to analyse reviews on different levels. Their study was able to prove the fact that binary classification tasks were performed in a more effective way when convolutional neural networks (CNNs) were used, considering that an 0.93 AUC score was registered, basically highlighting the power that ML may have not only in enabling us to better understand consumer behaviour, but also in enhancing decision-making. Especially after 2020, the idea of economic resilience has advanced due to global challenges, such as the COVID-19 pandemic. In their paper, Pernici and Stancu (2024) had a multidisciplinary approach, focusing over the impact of external factors, macroeconomic modelling and the development of those strategies that are long-term oriented, for which the researchers used sentiment analysis and text mining in order to explore the shift in the definition of economic resilience. Concerning the regional development, agribusiness and rural tourism are playing a vital role in economic growth.

Lequeux-Dincă and Teodorescu (2024), in their research, used semantic analysis in order to properly explore rural tourism governance for the South Bukovina area, by this being able to highlight how important public-private partnerships, EU funding and tourism events are when it comes down to promoting a sustainable development. To better comprehend the labour market trends, text mining has also been applied. The paper proposed by Davidescu et al. (2024) analysed LinkedIn job descriptions for entry-level positions in Romania, identifying key competencies such as experience, teamwork, and business development, highlighting the growing importance of data-driven skills and software knowledge in the labour market.

Lastly, Delcea et al. (2024) studied the growing applicability of AI in researches that are either local or regional, especially for topics such as resource management or urban planning. The researchers' bibliometric analysis underscored the fact that there is an increasing trend in the usage of AI methods, such as artificial neural networks or ML, when complex regional challenges are tackled.

The literature reviewed highlights the broad applicability of text mining and ML techniques across various fields, offering valuable tools for uncovering patterns, identifying emerging trends, and informing decision-making processes.

3. Methodology

As there is a multiphase planning across various stakeholders with regards to regional development strategies, large amounts of textual information is generated, much of which is unstructured (Wu et al., 2022). The available information is being stocked in several formats, like reports, policy documents, and official communications, the information available for each of the formats representing an essential criterion in detecting key themes and strategies in the planning process. More than this, the comprehension of the development priorities and improvement of decision-making relies heavily on two factors: an efficient extraction of key themes and data analysis. As the late 1940s brought up machine translation and information retrieval, a notable progress over time in regards to textual analytics was registered. By 1960s, computational linguistics was already starting to include different aspects of ML, such as semantics, syntax and morphology (Craig & Sommerville, 2006).

Text analysis has evolved significantly in recent years, having been developed in the late 1940s together with machine translation. Until the 1960s, it progressively evolved from the perspective of computational linguistics which included several aspects including morphology, semantics, and syntax (Chowdhary, 2020). The 1990s generated an impact on the attention of text mining and natural language processing due to significant technological advances, but also their applicability in different fields, such as health (Van Leeuwen et al., 2024), urban resilience (Hadj Kilani & Dhaher, 2024), and regional development (Shao et al., 2024; Bucur et al., 2024). Unlike text mining, NLP brings an additional component that provides the necessary tools for text processing (Dai et al., 2024), like tokenisation (Vayadande et al., 2024), stemming (Abidin et al., 2024), but also for more advanced analysis. All of these are fundamental to be able to analyse both key documents on a large scale and to extract relevant information for regional development strategies (Shamshiri et al., 2024).

For the analysis of regional development strategies in Romania for the 2021-2027 period, natural language processing (NLP) techniques and network analysis were applied to identify the dominant themes and interconnections between key concepts in policy documents. The methodological approach was executed in several steps:

- **Data collection and preprocessing:** The PDF format regional policy documents were acquired, Python and the `<<PyPDF2>>` library being responsible for extracting each file's text. This step was about going through all of the documents' pages, extracting the text in a raw format.
- **Text cleaning:** In order to assure the accuracy if the analysis, irrelevant words, known as stop words, had to be removed. The stop words are related to Romanian common words, like "este", "în", "pe" and "sau", words that are not able to add any value to the analysis. More than this, the stop words list was customised in

such a way that words specific only to the Romanian language are detected, thus improving the quality of the results.

- **Creating a text representation model:** The `<<CountVectorizer>>` from the `<<sklearn>>` helped in transforming the text into term vectors, exemplifying the word frequency that is present for each document. Additionally, based on the provided text, the `<<CountVectorizer>>` is creating a matrix that consists in token counts, where each of the matrix's columns representing a unique term, and the corresponding values describe the recurrence of those terms across the considered documents. To ensure the fact that stop words are completely excluded from the analysis, but also that the identification of key themes is as accurate as possible, the custom stop words list was applied against the matrix.
- **Word frequencies:** To compute the frequency of each term across all of the documents, processing the matrix generated by the `<<CountVectorizer>>` was required. The `<<sum(axis=0)>>` function helped the analysis by aggregating each term's occurrence, the result being used to create a key-value dictionary, the key is the word, whereas the value being the frequency of that word.
- **Creating the Word Cloud:** After cleaning up the data and turning it into numbers, the word cloud approach was applied to show the most frequent words in the documents. The Word Cloud from the `<<wordcloud>>` library was used to create a visual representation of the most frequent words, removing stop words and customising the dimensions and background colour for maximum clarity.
- **Display:** The word cloud is displayed using `<<matplotlib>>` to visualise the results. The word cloud's size and style are changed to make it easier to read, and stop words are left out of the visual output so that only useful words are shown.

This approach uses text mining and natural language processing on Romanian regional policy papers to find important topics and look at how often they appear. Lastly, the word cloud visualisation makes it easy to see and understand the main ideas in the policy documents.

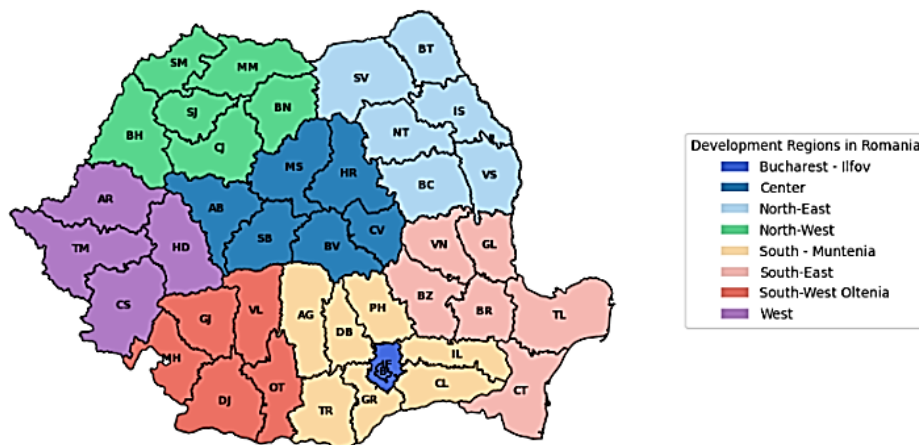
4. Results

To explore how regional development priorities are connected across Romania, this study conducts a text mining analysis on a corpus composed of strategic planning documents from all eight Romanian development regions. The chosen papers comprise formal plans for regional growth and smart specialisation initiatives for the years 2021 to 2027. These writings show the unique goals, subject areas, and territorial visions that were created at the regional level. They do this by taking into account both national policy directives and the demands of the local economy and society.

The European Union uses the NUTS2 classification (Nomenclature of Territorial Units for Statistics – Level 2) to show the development areas of Romania in Figure 1. These areas are used to study and put into action regional strategies, especially when it comes to getting money from the European Union. Even though they are not legal entities (i.e., they are not administrative units), they are very

important for: (i) regional statistical analysis; (ii) planning for territorial development; and (iii) using and absorbing European Union (EU) funding.

Figure 1. Map of development regions in Romania



Source: created by the authors using Python.

A word cloud based on the Central Region Development Strategy 2021-2027 was created in Figure 2, illustrating both the frequency of keywords and their significance. It can be seen that “regional development”, “projects”, “infrastructure” and “services” are the identified keywords that highlight significant priorities in economic growth, transportation, healthcare and public services.

The analysis highlights several critical areas that may be of interest to specialists in the field. First, we can see that, based on the keywords in the word cloud, the Centre Region focuses on (i) **Regional economic development**, which highlights the need for investments in infrastructure, but also for the creation of new jobs to stimulate the local economy. The second area observed is (ii) **Infrastructure and public services**; this focuses more on the most important improvements in the transport, education, and healthcare infrastructure, while addressing these services in both rural and urban areas; Another important aspect identified is (iii) **Sustainable growth and labour market integration**. A functional labour market, supported by policies that encourage both public and private investments, can contribute to generating inclusive and long-term economic progress. (iv) **Social policies and inclusion**, another area observed, focuses on support policies for socially vulnerable people, social inclusion security, and balanced spatial development. (v) **Tourism and Rural Environment**, which addresses the economic importance of tourism and rural development, particularly through tourism infrastructure projects, is another critical area identified. Finally, (vi) **Modernisation and Sustainability**, which aims to transform the region while ensuring that the economic development framework is environmentally and economically sustainable.

Figure 3. Word cloud based on the North-East Region Development Strategy



Source: created by the authors using Python.

In Figure 4, the Bucharest-Ilfov region's word cloud is able to put into perspective some major themes, such as “development”, “București”, and “Ilfov”, therefore indicating a focus over encouraging growth for both the capital city and its surrounding areas. Terms like “infrastructure” and “transport” underline the region's investment in physical and digital connectivity. Economic growth is driven by words like “investiții” (investments) and “afaceri” (businesses), while social inclusion is tackled with the help of terms like “social” and “sprijinire” (support). The word cloud also capable of emphasising how important it is to have a well-balanced urban and rural development, with references to “rural” (rural environment) and “orașe” (cities).

Based on the word cloud analysis’, a few distinct key themes arise for the București-Ilfov region: (i) **Balanced Urban and Suburban Development** - its aim being to contribute to a synchronised growth between Bucharest and its surrounding regions, by addressing concerns that are related to urban integration; (ii) **Economic Diversification** – encourages the expansion beyond traditional sectors, only by assisting SMEs, digital services, and the industries that are eco-friendly; (iii) **Sustainable Development and Environmental Concerns** – this topic is putting an emphasis over several factors, like smart urban planning, sustainability, and a more efficient waste management; (iv) **Tourism Development** – capitalises on cultural and historical resources to enhance regional tourism and economic activity; (v) **Public Administration and Governance** – this topic is described by the necessity of having the regional strategies aligned to the national policies, through effective institutional collaboration; (vi) **Population Dynamics** – takes into account demographic trends, urbanisation, and migration when planning infrastructure and services; (vii) **Social Services and Quality of Life** – this topic aims at giving all citizens a better access to healthcare, education, and other necessities.

Integration and Alignment – underlines the use of EU funding and cooperation to support development, especially in infrastructure and innovation; **(vii) Urban and Regional Dynamics** – reflects the interactions happening amongst regions, but also the urban-rural integration, possibly influenced by proximity to other economic hubs.

Since the region's word cloud also contains keywords like “Facebook”, “www”, “inovare” (innovation), and “afaceri” (business), an increasing focus over digital technologies and digital economy in the North-West Region is suggested. Therefore, three interconnected areas of interest are further highlighted. First, the **digitalisation of business** is gaining momentum, as the local enterprises are relying more and more on social media and digital tools for visibility, marketing, and communication. Second, **social media and community engagement** appear to play a strategic role in regional development, thus establishing public participation, innovation diffusion, and collaboration among stakeholders. Third, **the process of assisting digital startups and innovation ecosystems** appears to be a key regional priority, pointing to a focus on entrepreneurship, IT growth, and digital transformation.

Figure 6. Word cloud based on the North-West Region Development Strategy



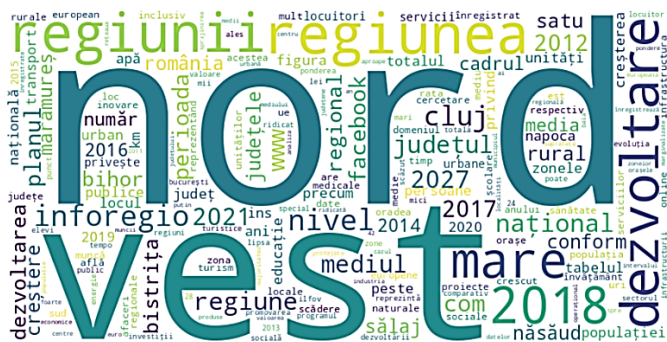
Source: created by the authors using Python.

About West Region's word cloud, Figure 7 is able to describe several relevant themes that are basically explaining what are the region's development priorities and ongoing strategies. Amongst world cloud's most frequent words, terms such as “regiunea Vest” (West Region), “dezvoltare” (development), and “creștere” (growth) are present, suggesting how important the economic development and growth strategies are for this region. More than this, these keywords also indicate the fact that the regions is focused over improving its economic landscape, especially when it comes to stimulating local economies, attracting new investments or enhancing various sectors, like industry, services and infrastructure. Key themes identified in the West Region's development strategy include: **(i) Economic Growth and Investment** – topic focused over attracting national and international investments and diversifying the regional economy; **(ii) Infrastructure and Transport Development** – priority accorded for having an improved connectivity in

order to support economic activity and regional mobility; *(iii) Social Services and Education* – aims at not only strengthening the healthcare and education services, but also at ensuring an equal access to essential services; *(iv) Environmental Sustainability* – promotes responsible development while protecting natural resources; *(v) European Integration and Alignment* – ensures that the regional strategies are aligning to EU’s policies and funding instruments; *(vi) Digitalisation and Innovation* – emphasises the support for digital infrastructure, tech entrepreneurship, and modernisation of public services; *(vii) Sustainability and Green Growth* – this topic is revolving around integrating renewable energy and sustainable practices into long-term development goals.

As shown in the word cloud, the West Region’s development priorities are concentrated towards economic growth, investment, infrastructure development, sustainability and education.

Figure 7. Word cloud based on the West Region Development Strategy



Source: created by the authors using Python.

In the context of Figure 8, it describes the word cloud that has been generated from the South-Muntenia Region Development Strategy for the 2021-2027 timeframe. About the most prominent themes, the presence of terms such as “Sud Muntenia” (South-Muntenia), “regiunea” (region), “dezvoltare” (development), and “reziduu” (residue) is highlighting the fact that the region's focus is over development, environmental concerns, but also over the socioeconomic transformation. The high occurrence of words like “muncă” (work), “transport”, and “mediu rural” (rural environment) indicates that there are some major directions of development, such as labour market interventions, infrastructure improvement and the need of decreasing the urban-rural disparities. Some fundamental priority themes are emerging, based on this analysis: *(i) Economic Development and Labour Market* – strategic emphasis that is being put over several factors, like job creation, improving employment conditions, and boosting regional productivity; *(ii) Environmental Protection and Waste Management* – this topic is pointing out the strategic efforts that are being made towards achieving the ecological sustainability, this in order to reduce waste (e.g., “reziduu fix”) .; *(iii) Transport and Infrastructure* – strategic development plans that aim at increasing the investments

funds and implement regional projects, as these factors contribute to Romania's political and economic integration within the European framework. The EU funds are used by any of the regions in order to support innovation projects and infrastructure. **Urban-rural division** proved to be a common topic for all areas, where if the developed regions focused over the growth of smart cities and urban development, the less developed regions had their focus shifted towards rural development, agriculture and tourism. As such, the strategic plans of either more developed or less developed areas aim to reduce disparities between urban and rural areas. In particular, the comparison between more developed and less developed regions, such as Bucharest-Ilfov with North-East, underscores how the geographical context, development level, or the institutional capacity are shaping divergent strategic priorities. As such, this comparative point of view enhances the study's analytical value by revealing both convergence and asymmetries across Romania's regional policy landscape.

This study took into account all eight of Romania's development regions, seven of which were based on regional development plans, whereas one region (South-West Oltenia) was analysed using the Regional Smart Specialisation Strategy (RIS3). The findings were gathered through text mining techniques that proved to be effective tools in revealing not only the common policy themes, but also the regional differences, directly from strategic planning documents.

The lack of comprehensive lexicon that is specifically designed for the Romanian language proved to be one of the analysis' main limitation. Even if lexicons such as Bing or NRC exist for the English language, fewer resources are available for Romanian text, as such leading to an increased complexity with regards to the sentiment analysis and the text mining tasks. A dedicated Romanian lexicon could significantly improve the quality and accuracy of future text mining analyses in this language.

In conclusion, each of Romania's areas have a set of strategies that are adapted to their specific needs and characteristics, yet all of them raised a common list of concerns in regards to human capital growth, sustainable economic development, infrastructure and public service improvement and environmental protection. By addressing these concerning themes, a contribution towards ensuring a balanced and sustainable development across all regions of Romania will be brought.

Acknowledgement

This work was funded by the EU's NextGenerationEU instrument through the National Recovery and Resilience Plan of Romania - Pillar III-C9-I8, managed by the Ministry of Research, Innovation and Digitisation, within the project entitled „Place-based Economic Policy in EU's Periphery – fundamental research in collaborative development and local resilience. Projections for Romania and Moldova (PEPER)”, contract no. 760045/23.05.2023, code CF 275/30.11.2022.

References

- [1] Abidin, Z., Junaidi, A., Wamiliana. (2024). Text stemming and lemmatization of regional languages in Indonesia: A systematic literature review. *Journal of Information Systems Engineering and Business Intelligence*, 10, 217-231. <https://doi.org/10.20473/jisebi.10.2.217-231>.
- [2] Abir, S. I., Sarwer, M. H., Hasan, M., Sultana, N., Dolon, M. S. A., Arefeen, S. M. S., Shimanto, A. H., Zakaria, R. M., Al Shiam, S. A., Saha, T. R. (2024). Accelerating BRICS economic growth: AI-driven data analytics for informed policy and decision making. *Journal of Economics, Finance and Accounting Studies*, 6, 102-115. <https://doi.org/10.32996/jefas.2024.6.6.8>.
- [3] Abir, S. I., Sarwer, M. H., Hasan, M., Sultana, N., Dolon, M. S. A., Arefeen, S. M. S., Shimanto, A. H., Zakaria, R. M., Al Shiam, S. A., Saha, T. R. (2024). Accelerating BRICS economic growth: AI-driven data analytics for informed policy and decision making. *Journal of Economics, Finance and Accounting Studies*, 6, 102-115. <https://doi.org/10.32996/jefas.2024.6.6.8>.
- [4] Attah, R. U., Garba, B. M. P., Gil-Ozoudeh, I., Iwuanyanwu, O. (2024). Leveraging geographic information systems and data analytics for enhanced public sector decision-making and urban planning. *Magna Scientia Advanced Research Review*, 12, 152-163. <https://doi.org/10.30574/msarr.2024.12.2.0191>.
- [5] Briciu, A., Călin, A.-D., Miholca, D.-L., Moroz-Dubenco, C., Petrașcu, V., Dascălu, G. (2024). Machine-learning-based approaches for multi-level sentiment analysis of Romanian reviews. *Mathematics*, 12, 456. <https://doi.org/10.3390/math1203045611>.
- [6] Bucur, C., Tudorica, B., Andrei, J. V., Dusmanescu, D., Paraschiv, D., Teodor, C. (2024). Sentiment analysis of global news on environmental issues: Insights into public perception and its impact on low-carbon economy transition. *Frontiers in Environmental Science*, 12, 1360304. <https://doi.org/10.3389/fenvs.2024.1360304>.
- [7] Chowdhary, K. R. (2020). Natural language processing. În *Fundamentals of artificial intelligence* (603-649). Springer India.
- [8] Craig, N., Sommerville, J. (2006). Information management systems on construction projects: Case reviews. *Records Management Journal*, 16, 131-148. <https://doi.org/10.1108/09565690610713192>.
- [9] Dai, S., Li, K., Luo, Z., Zhao, P., Hong, B., Zhu, A., Liu, J. (2024). AI-based NLP section discusses the application and effect of bag-of-words models and TF-IDF in NLP tasks. *Journal of Artificial Intelligence and Geospatial Systems*, 5, 13-21. <https://doi.org/10.60087/jaigs.v5i1.149>.
- [10] Davidescu, A. A., Manta, E. M., Cojocaru, M. R. (2024). The transition from higher education to employment in Romania: A deep understanding of the job market using text analysis. În A. M. Thake, K. Sood, E. Özen, & S. Grima (Eds.), *Contemporary studies in economic and financial analysis* (69-86). Emerald Publishing Limited.
- [11] Delcea, C., Nica, I., Ionescu, Ș., Cibu, B., Țibrea, H. (2024). Mapping the frontier: A bibliometric analysis of artificial intelligence applications in local and regional studies. *Algorithms*, 17, 418. <https://doi.org/10.3390/a17090418>.
- [12] Di Marino, M., Tiitu, M., Rehunen, A., Chavoshi, S. H., Lapintie, K. (2025). Multilocality in the regions of Oslo and Helsinki: A regional planning perspective after the COVID-19 pandemic. *Regional Studies*, 59, 2355290. <https://doi.org/10.1080/00343404.2024.2355290>.

- [13] Hadj Kilani, B., Dhaher, N. (2024). Building urban resilience through megaevents: A systematic review using text mining and natural language processing (NLP). *Qeios*. <https://doi.org/10.32388/M3BIFL.2>.
- [14] Lequeux-Dincă, A.-I., Teodorescu, C. (2024). Governance and development of tourism in rural areas through the lens of media in South Bukovina (Romania). *Agri-culture*, 14, 1462. <https://doi.org/10.3390/agriculture14091462>.
- [15] Niculae, A.-M., Oprea, S.-V., Văduva, A.-G., Bâra, A., Andreescu, A.-I. (2024). Assessing the role of machine learning in climate research publications. *Sustainability*, 16, 11086. <https://doi.org/10.3390/su162411086>.
- [16] Nielbo, K. L., Karsdorp, F., Wevers, M., Lassche, A., Baglini, R. B., Kestemont, M., Tahmasebi, N. (2024). Quantitative text analysis. *Nature Reviews Methods Primers*, 4, 25. <https://doi.org/10.1038/s43586-024-00302-w>.
- [17] Pernici, A., Stancu, S. (2024). Exploring the concept of economic resilience pre- and post-2020: A literature review using text mining and sentiment analysis. În K. Arai (Ed.), *Intelligent computing* (105-127). Springer Nature Switzerland.
- [18] Rotaru, A., Hegedus, C., Pop, I., Vatca, S., Vatca, A., Andronie, L. (2024). Assessment of regional development level in Romania through principal component analysis. *Journal of Central European Agriculture*, 25, 588-605. <https://doi.org/10.5513/JCEA01/25.2.4176>.
- [19] Rotaru, A., Hegedus, C., Pop, I., Vatca, S., Vatca, A., Andronie, L. (2024). Assessment of regional development level in Romania through principal component analysis. *Journal of Central European Agriculture*, 25, 588-605. <https://doi.org/10.5513/JCEA01/25.2.4176>.
- [20] Shamshiri, A., Ryu, K. R., Park, J. Y. (2024). Text mining and natural language processing in construction. *Automation in Construction*, 158, 105200. <https://doi.org/10.1016/j.autcon.2023.105200>.
- [21] Shao, R., Lin, P., Xu, Z. (2024). Integrated natural language processing method for text mining and visualization of underground engineering text reports. *Automation in Construction*, 166, 105636. <https://doi.org/10.1016/j.autcon.2024.10563621>.
- [22] Van Leeuwen, J. R., Penne, E. L., Rabelink, T., Knevel, R., Teng, Y. K. O. (2024). Using an artificial intelligence tool incorporating natural language processing to identify patients with a diagnosis of ANCA-associated vasculitis in electronic health records. *Computers in Biology and Medicine*, 168, 107757. <https://doi.org/10.1016/j.combiomed.2023.107757>.
- [23] Vayadande, K., Kale, D. R., Nalavade, J., Kumar, R., Magar, H. D. (2024). Text generation & classification in NLP: A review. În A. Dey, S. Nayak, R. Kumar, & S. N. Mohanty (Eds.), *How machine learning is innovating today's world* (25-36). Wiley.
- [24] Vaz, E. (2024). Regional knowledge economies: Exploring the intersection of technology, geography, and innovation in the digital era. Springer Nature Switzerland.
- [25] Wu, C., Li, X., Guo, Y., Wang, J., Ren, Z., Wang, M., Yang, Z. (2022). Natural language processing for smart construction: Current status and future directions. *Automation in Construction*, 134, 104059. <https://doi.org/10.1016/j.autcon.2021.104059>.