International Journal of Multidisciplinary and Current Educational Research (IJMCER)
ISSN: 2581-7027 ||Volume|| 5 ||Issue|| 6 ||Pages 111-118 ||2023||

Synergizing Big Data Analytics and Natural Language Processing: A Comprehensive Review of Techniques and Emerging Trends

Leonidas Theodorakopoulos
Department of Management Science and Technology, University of Patras.

ABSTRACT : The integration of big data analytics and natural language processing (NLP) has led to exciting developments in the field of text analysis. This paper explores the techniques and applications of big data analytics and NLP, including sentiment analysis, named entity recognition, topic modeling, and summarization. We discuss real-world applications such as social media monitoring, customer feedback analysis, and healthcare data analysis. Furthermore, we highlight emerging trends and future directions in the field, such as deep learning-based NLP models, knowledge graphs, and explainable AI.
.

KEYWORDS: named entity recognition, topic modeling, summarization, social media monitoring, customer feedback analysis

I. INTRODUCTION

Big data analytics and natural language processing (NLP) are two fields that are becoming increasingly important in today's data-driven world. Big data analytics involves the processing and analysis of large volumes of data, often from diverse sources, to extract insights and inform decision-making. NLP, on the other hand, is concerned with the processing and analysis of human language data, such as text and speech.Combining these two fields can have several benefits. Firstly, NLP can help in extracting insights from large textual datasets that are difficult or impossible to analyze using traditional methods. For example, sentiment analysis can help businesses to understand customer opinions and feedback at scale, while named entity recognition can help to identify important entities such as people, organizations, and locations in unstructured text. Secondly, big data analytics can enable more accurate and efficient NLP models. For example, by using large-scale text corpora to train NLP models, researchers can improve the accuracy of tasks such as language translation, text classification, and information extraction. Additionally, big data analytics can help to overcome some of the challenges associated with NLP, such as the need for labeled training data, by enabling unsupervised and semi-supervised learning approaches [1]. The combination of big data analytics and NLP can unlock new opportunities for data-driven insights and decision-making, particularly in industries such as healthcare, finance, and social media. However, there are also several challenges associated with this approach, such as the need for specialized skills and infrastructure, the need to ensure data privacy and security, and the need to address ethical considerations such as bias and fairness [2].

II. TECHNIQUES AND APPROACHES

Big data analytics with natural language processing (NLP) involves the use of various techniques and approaches to extract insights and information from large volumes of textual data. Some of the most common techniques used in this field include:

A. Sentiment Analysis : Sentiment analysis is a popular technique in natural language processing (NLP) that involves the automated analysis of text to determine the sentiment or opinion expressed by the author [3]. It is a valuable tool for businesses, governments, and researchers who want to understand the public sentiment towards a product, service, policy, or idea. Sentiment analysis can be used for a variety of tasks, including brand monitoring, customer feedback analysis, and social media analysis [4]. For example, businesses can use sentiment analysis to monitor customer feedback on social media platforms and identify potential problems with their products or services. By analyzing the sentiment of customer reviews, businesses can gain valuable insights into what aspects of their products or services are most appealing to customers, and what areas may need improvement [5]. Sentiment analysis is typically performed using machine learning techniques, such as supervised learning or deep learning. In supervised learning, a model is trained on a labeled dataset of text, where each example is labeled with the sentiment expressed in the text [6]. The model learns to recognize patterns in the text that are associated with positive, negative, or neutral sentiment. Once the model is trained, it can be used to predict the sentiment of new, unlabeled text. Notably, in educational settings, administrators and staff at educational institutions can use NLP semantic and sentiment analysis to study students' responses to current instruction and changes in their academic and social environments [7]. This can be useful for determining whether a particular curriculum or teaching strategy is being well received and for identifying students who may be experiencing difficulties of some sort [8]. Educators can also utilize Natural Language Processing techniques to examine the level of collaboration between students in the classroom [9]. Researchers have begun to employ social network analysis techniques to linguistic data in order to identify patterns of student collaboration in online discussion forums and Massive Open Online Courses (MOOCs) [10]. Moreover, Natural Language Processing Techniques are indispensable assets for future leaders [11]. Executives, supervisors, and leaders are always on the lookout for development opportunities that will help them become better communicators, talent evaluators, and ultimately transform the organization [12]. Furthermore, the process has its origins in the introduction of natural language processing technology, also known as "discourse analysis," which is the study of the relationships between naturally occurring connected sentences, whether they are spoken or written [13]. Companies that use digital tools to activate strategies or facilitate structured dialogue as part of their leadership development initiatives can, for instance, leverage the data generated by natural language processing to gain a deeper understanding of leadership dynamics [14], influence, and culture. This does not entail eavesdropping on private conversations or identifying individuals by name, but over time, patterns emerge that assist executives in identifying unintended consequences and making more informed decisions [15]. It can assist executives in discovering how their communications are perceived by cohorts (e.g., anonymously segmented by role or geography) within their organization [16]. And with more than 84 percent of businesses recognizing the significance of "people analytics," it is more essential than ever to comprehend natural language processing and how it operates.

B. Deep Learning Techniques ***:*** Deep learning techniques such as neural networks have also been used for sentiment analysis with great success. These models are particularly useful for dealing with complex and nuanced language, as they can capture more subtle patterns in the text. However, there are several challenges associated with sentiment analysis, particularly when dealing with informal language or sarcasm. Additionally, sentiment analysis models can be biased, particularly when trained on datasets that are not representative of the population. Therefore, it is important to carefully select training data and evaluate models on a diverse range of examples to ensure they are accurate and unbiased [17].

C. Named Entity Recognition : Named entity recognition (NER) is an important task in natural language processing (NLP) that involves identifying and categorizing named entities in unstructured text. Named entities are typically proper nouns that refer to specific people, organizations, locations, or other entities. NER can be used for a variety of tasks, including information extraction, entity disambiguation, and event extraction. Information extraction involves automatically extracting structured data from unstructured text. Named entities are often important pieces of information in text, and NER can be used to extract this information automatically. For example, NER can be used to extract the names of people, organizations, and locations mentioned in news articles or other textual sources. Entity disambiguation involves resolving ambiguous references to named entities. For example, the name "Washington" could refer to the city or the state, and NER can be used to determine which entity is being referred to base on the context in which the name appears. Event extraction involves identifying and extracting information about events that occur in text. Named entities are often important components of events, and NER can be used to extract this information automatically. For example, NER could be used to extract the names of people, organizations, and locations involved in a news event, such as a political scandal or natural disaster. NER is typically performed using machine learning techniques, such as supervised learning or deep learning. In supervised learning, a model is trained on a labeled dataset of text, where each example is labeled with the named entities present in the text. The model learns to recognize patterns in the text that are associated with named entities and can then be used to predict named entities in new, unlabeled text. There are several challenges associated with NER, including ambiguity and variability in naming conventions, as well as the need for domain-specific knowledge to accurately recognize named entities. Therefore, it is important to carefully select training data and evaluate models on a diverse range of examples to ensure they are accurate and robust [18].

D. Topic Modeling : Topic modeling is a powerful technique in natural language processing (NLP) that involves identifying the underlying topics or themes in a collection of documents. It is a useful tool for a variety of tasks, including document clustering, summarization, and recommendation systems. Document clustering involves grouping similar documents together based on their topic or theme. Topic modeling can be used to automatically identify the topics present in a collection of documents, which can then be used to cluster the documents based on their topic or theme. Topic modeling can be used to identify the most important topics present in the documents, which can then be used to generate a summary that captures the key themes of the documents. Recommendation systems involve suggesting relevant documents or other items to users based on their interests or preferences. Topic modeling can be used to identify the topics present in a user's past interactions with the system, which can then be used to suggest other documents or items that are relevant to those topics. Topic modeling is typically performed using unsupervised machine learning techniques, such as Latent Dirichlet Allocation (LDA) or Non-negative Matrix Factorization (NMF). In LDA, each document is represented as a mixture of topics, and each topic is represented as a probability distribution over the words in the corpus. The goal of the model is to identify the topic distribution for each document and the word distribution for each topic that best explains the observed data. NMF is a similar technique that involves factorizing a matrix of word frequencies into two lower-dimensional matrices representing topics and word weights. NMF has the advantage of being more interpretable, as it directly maps words to topics and can be easily visualized. However, there are several challenges associated with topic modeling, including selecting the appropriate number of topics, handling noisy or ambiguous data, and interpreting the resulting topics. Therefore, it is important to carefully select appropriate models and evaluate them on a diverse range of examples to ensure they are accurate and useful for the intended application.

E. Summarization : Summarization is a crucial task in natural language processing that involves generating a shorter version of a longer document while retaining the most important information. Summarization can be useful for a variety of applications, including news article summarization, document summarization for legal or medical documents, and automatic summarization of customer feedback. News article summarization involves automatically generating a summary of news articles. This can be useful for readers who want to quickly scan through multiple articles to get a sense of the news without having to read the entire article. Summarization algorithms can be trained on large datasets of news articles and can use techniques such as named entity recognition and keyword extraction to identify the most important information in the article. Document summarization for legal or medical documents can be useful for professionals who need to quickly understand the content of lengthy legal or medical documents. Summarization algorithms can identify key terms and concepts in the document and generate a summary that captures the most important information. Automatic summarization of customer feedback can be useful for businesses that receive a large volume of feedback from customers [19]. Summarization algorithms can identify common themes in the feedback and generate a summary that highlights the most frequently mentioned issues or concerns. But there are several challenges associated with summarization, including dealing with ambiguity, capturing the intended meaning of the text, and maintaining coherence and readability in the summary. Various approaches have been proposed to address these challenges, including extractive summarization, which involves selecting the most important sentences or phrases from the original document, and abstractive summarization, which involves generating new sentences that capture the essence of the original document.

Additionally, the evaluation of summarization algorithms is also challenging, as there may be multiple correct ways to summarize a document. Therefore, it is important to carefully evaluate summarization algorithms using appropriate metrics and benchmarks to ensure their effectiveness for the intended application. While these techniques can be powerful tools for analyzing large volumes of textual data, they also face several challenges. For example, accuracy can be a challenge for techniques such as sentiment analysis and named entity recognition, particularly when dealing with complex or ambiguous text. Scalability can also be an issue, as these techniques can be computationally expensive and require specialized infrastructure. Additionally, interpretability can be a challenge, as it can be difficult to understand how these techniques arrive at their results, which can make it difficult to trust and act on the insights they provide [20]. Overall, while big data analytics with NLP presents many exciting opportunities, it is important to carefully consider these challenges and develop approaches that can address them. This may involve developing more accurate and scalable models, improving interpretability and transparency, and addressing ethical considerations such as bias and fairness.

III. REAL WORLD APPLICATIONS OF BDA AND NLP

Real-world applications of big data analytics and NLP are diverse and rapidly growing. Here are some examples: *Social media monitoring* has become an increasingly important tool for companies to understand their customers and the market. With the vast amount of data available on social media platforms, companies can gain valuable insights into customer sentiment, brand reputation, and market trends. Sentiment analysis, a technique commonly used in social media monitoring, involves analyzing text to determine the sentiment or opinion of the author. By monitoring social media conversations about their brand and products, companies can gain a better understanding of how their customers feel about their products, services, and brand overall. This information can help companies identify potential issues or opportunities for improvement, and make data-driven decisions about marketing, customer service, and product development [21]. For example, a company may use social media monitoring to track customer sentiment towards a new product launch. By analyzing social media conversations about the product, the company can identify any issues or concerns that customers may have and adjust their marketing strategy or product features accordingly. They can also identify positive sentiment and use it to their advantage in their marketing and advertising efforts. In addition to sentiment analysis, social media monitoring can also be used to track mentions of competitors, industry trends, and customer needs and preferences [22]. This information can help companies stay competitive and adapt to changes in the market. Overall, social media monitoring and sentiment analysis can provide companies with valuable insights that can help them make data-driven decisions and improve their products and services. It is becoming an increasingly important tool in today's digital age, where social media plays a significant role in shaping public opinion and influencing purchasing decisions [23].

Customer feedback analysis is also a critical component of any company's customer experience strategy. It involves collecting and analyzing customer feedback data, such as survey responses, product reviews, and customer support tickets, to identify common themes and sentiment. This data can be used to understand customer needs and preferences, identify areas for improvement, and measure customer satisfaction. One of the primary benefits of customer feedback analysis is that it can help companies identify customer pain points and areas for improvement. By analyzing customer feedback data, companies can gain insights into the issues and challenges that customers face with their products and services [24]. This information can help companies make targeted improvements to their products and services, thereby improving the customer experience and increasing customer satisfaction. For example, a company may analyze customer feedback data to identify common complaints about a specific product [25]. They may discover that customers are having difficulty using a certain feature, or that the product is not meeting their expectations in some way. Armed with this information, the company can make improvements to the product or its documentation and provide additional customer support to address the issue. Customer feedback analysis can also help companies measure customer satisfaction and identify areas where they excel. By analyzing customer feedback data, companies can identify what is working well and what customers appreciate about their products and services. This information can be used to build on strengths and enhance the customer experience even further. Customer feedback analysis is an essential tool for companies looking to improve their products and services and enhance the customer experience. By collecting and analyzing customer feedback data, companies can gain valuable insights into customer needs and preferences, identify areas for improvement, and measure customer satisfaction [26].

Healthcare data analysis is an essential component of modern healthcare delivery, as it can help to improve patient outcomes, reduce costs, and enhance the quality of care. With the increasing digitization of healthcare data, there is an ever-growing volume of patient data available, including electronic health records (EHRs), medical imaging, and genomics data. However, this data is often unstructured and difficult to analyze, which is where natural language processing (NLP) techniques can be particularly valuable. NLP techniques can be used to extract relevant information from unstructured medical text data, such as clinical notes, radiology reports, and pathology reports. Named entity recognition, for example, can be used to identify and categorize entities such as medical conditions, medications, and procedures mentioned in clinical notes, while topic modeling can be used to identify the underlying themes or topics in a corpus of clinical notes. By analyzing this information, healthcare providers can gain insights into patient care, such as identifying patients at risk of developing certain medical conditions or predicting treatment outcomes. For example, by analyzing EHR data, providers may be able to identify patients at risk of developing chronic conditions, such as diabetes or heart disease, and intervene early to prevent or manage these conditions. Medical imaging data is another rich source of patient data that can benefit from NLP techniques. By analyzing radiology reports and imaging data, healthcare providers can identify patterns and insights that may not be immediately visible to the human eye. For example, by analyzing chest X-rays, providers may be able to detect early signs of lung cancer or identify other respiratory conditions [27].

Genomics data is also an area where NLP techniques can be particularly valuable. By analyzing genetic data, providers can identify patients at risk of certain genetic conditions or develop personalized treatment plans based on an individual's genetic makeup. Overall, healthcare data analysis using NLP techniques has the potential to revolutionize healthcare delivery by providing new insights into patient care and enabling more personalized and effective treatments [28]. As the volume of healthcare data continues to grow, the role of NLP techniques in healthcare data analysis is likely to become increasingly important.

IV. EMERGING TRENDS AND FUTURE DIRECTIONS IN THE FIELD OF BIG DATA ANALYTICS AND NLP

Deep learning-based NLP models are a recent development in the field of NLP that use neural networks to learn and extract complex patterns from large amounts of textual data. These models are capable of handling various NLP tasks such as language translation, language generation, sentiment analysis, and named entity recognition, among others. One of the breakthroughs in deep learning-based NLP models has been the development of transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer). These models use self-supervised learning techniques to pre-train on large amounts of text data and then fine-tune on specific NLP tasks (*Mazzei, 2020*). The pre-training process allows the models to learn the context and meaning of words and phrases, leading to more accurate and robust performance on downstream NLP tasks. BERT, for example, has been shown to outperform previous state-of-the-art models on various NLP benchmarks, including sentence classification, question-answering, and named entity recognition. It has also been used in real-world applications such as chatbots and search engines [29]. GPT, on the other hand, is a generative model that can be used for language generation tasks, such as text completion and summarization. It has been shown to produce human-like responses in chatbots and can also be used for content generation in marketing and advertising [30]. In addition to transformer-based models, other deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have also been applied to NLP tasks, with promising results. These models can handle tasks such as sentiment analysis and machine translation. Deep learning-based NLP models have shown great potential for improving the accuracy and efficiency of NLP tasks, leading to a range of practical applications in industries such as healthcare, finance, and customer service [31].

a. Knowledge graphs : Knowledge graphs are becoming increasingly important in the field of big data analytics and NLP. They provide a way to organize and integrate structured and unstructured data from various sources, enabling more efficient and effective data analysis. In a knowledge graph, entities are represented as nodes, and the relationships between them are represented as edges. For example, in a medical knowledge graph, a patient could be a node, and the relationships between the patient and their medical conditions, medications, and medical history could be represented as edges. Knowledge graphs can be constructed automatically from large amounts of unstructured data using NLP techniques such as named entity recognition and relationship extraction [32]. They can also be manually constructed by experts in a specific domain. Once constructed, knowledge graphs can be queried using graph-based query languages such as SPARQL, allowing users to extract meaningful insights and answer complex questions [33]. Knowledge graphs have a wide range of applications, including in healthcare, finance, e-commerce, and more. In healthcare, knowledge graphs can be used to analyze patient data to identify patterns and predict outcomes. In finance, knowledge graphs can be used to analyze financial data to detect fraud and make investment decisions [34]. In e-commerce, knowledge graphs can be used to make personalized product recommendations based on a customer's interests and past behavior. As the volume and complexity of data continue to increase, knowledge graphs are likely to become even more important in the field of big data analytics and NLP, enabling more efficient and effective data analysis and decision-making [35].

b. Machine learning models : Machine learning models, can be complex and difficult to understand, making it challenging to explain the rationale behind their decisions. Explainable AI (XAI) aims to address this challenge by providing methods and techniques for understanding how a model arrives at its predictions or decisions. In the context of healthcare, for example, explainability is important because the decisions made by a model can have significant consequences for patients. If a model recommends a certain treatment or diagnosis, it is important for doctors and patients to understand how the model arrived at that recommendation. Similarly, in finance, explainability is important for regulatory compliance and risk management. There are various approaches to achieving explainability in machine learning models. One approach is to use local explanation methods, which provide an explanation for a specific prediction made by the model. Examples of local explanation methods include LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations). Another approach is to use global explanation methods, which provide an overall understanding of how the model works. Examples of global explanation methods include feature importance analysis, decision trees, and rule extraction. XAI is an active area of research, and there are ongoing efforts to develop new methods and techniques for improving the interpretability of machine learning models [36]. This is particularly important as machine learning models become more prevalent in areas such as healthcare, finance, and autonomous vehicles, where the consequences of model errors can be significant. Overall, the field of big data analytics and NLP is rapidly evolving and offers numerous opportunities for research and application in a wide range of industries.

Finally, Natural language processing (NLP) could be contributed to gamification contexts. NLP enables AI to analyze and comprehend human language, enabling the creation of more immersive and interactive experiences. A gamified learning platform, for instance, could use NLP to analyze student responses and provide individualized feedback based on each individual's strengths and deficiencies. This could result in a more engaging and effective learning experience that is better tailored to the requirements of each student [37]. Also, NLP can be used to make gamification experiences more dynamic and adaptive. By analyzing user data with machine learning, gamification designers can create experiences that adapt and evolve over time. This may include personalized challenges that become more difficult as the user advances or dynamic rewards that vary based on the user's actions.

V. CONCLUSION

In conclusion, the integration of big data analytics and natural language processing has the potential to revolutionize the way we extract insights from textual data. With the help of NLP techniques such as sentiment analysis, named entity recognition, topic modeling, and summarization, we can derive meaningful information from large volumes of unstructured text data. Real-world applications such as social media monitoring, customer feedback analysis, and healthcare data analysis demonstrate the practical value of these techniques. Furthermore, emerging trends such as deep learning-based NLP models, knowledge graphs, and explainable AI are likely to play a significant role in shaping the future of this field. Despite the significant progress made in the field of big data analytics and NLP, there are still many challenges and opportunities for future research. One important area of research is to develop more accurate and efficient NLP models, particularly in areas such as sentiment analysis and summarization. Another area of research is to improve the scalability of these models, enabling them to handle even larger volumes of data. In addition, there is a need for research on the interpretability of these models, particularly in applications where the decisions made by the model have significant impact. Finally, there is a need for research on the ethical and social implications of these technologies, particularly in areas such as privacy, bias, and fairness. By addressing these challenges and opportunities, we can continue to push the boundaries of what is possible with big data analytics and NLP and unlock even greater value for society

REFERENCES

[1] Agerri, R., Artola, X., Beloki, Z., Rigau, G., & Soroa, A. (2015). Big data for Natural Language Processing: A streaming approach. Knowledge-Based Systems, 79, 36-42.

[2] Alblawi, A. S., & Alhamed, A. A. (2017). Big data and learning analytics in higher education: Demystifying variety, acquisition, storage, NLP and analytics. 2017 IEEE Conference on Big Data and Analytics (ICBDA). https://doi.org/10.1109/icbdaa.2017.8284118

[3] Antonopoulou, H., Theodorakopoulos, L., Halkiopoulos, C., & Mamalougkou, V. (2023). Utilizing Machine Learning to Reassess the Predictability of Bank Stocks. Emerging Science Journal, 7(3), 724–732. https://doi.org/10.28991/esj-2023-07-03-04

[4] Chen, X., & Sun, Y. (2021). An Intelligent System to Improve Athlete Depression and Eating Disorder using Artificial Intelligence and Big Data Analysis. NLP Techniques and Applications. https://doi.org/10.5121/csit.2021.111910

[5] Gudivada, V. N., Rao, D., & Raghavan, V. V. (2015). Big data driven natural language processing research and applications. In Handbook of statistics (Vol. 33, pp. 203-238). Elsevier.

[6] Karras, C., Karras, A., Theodorakopoulos, L., Giannoukou, I., & Sioutas, S. (2022, August). Expanding queries with maximum likelihood estimators and language models. In The International Conference on Innovations in Computing Research (pp. 201-213). Cham: Springer International Publishing.

[7] Kanavos, A., Theodorakopoulos, L., & Vonitsanos, G. (2019, July). NoSQL Approach for Recommendation of Highly Rated Products. In 7th International Conference on Contemporary Marketing Issues (ICCMI).

[8] Khader, M., Awajan, A., & Al-Naymat, G. (2018, November). The effects of natural language processing on big data analysis: Sentiment analysis case study. In 2018 International Arab Conference on Information Technology (ACIT) (pp. 1-7). IEEE.

[9] Li, L., Geissinger, J., Ingram, W. A., & Fox, E. A. (2020). Teaching natural language processing through big data text summarization with problem-based learning. Data and Information Management, 4(1), 18-43.

[10] Liu, X., Shin, H., & Burns, A. C. (2021). Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. Journal of Business research, 125, 815-826.

[11] Mackey, T. K., Li, J., Purushothaman, V., Nali, M., Shah, N., Bardier, C., ... & Liang, B. (2020). Big data, natural language processing, and deep learning to detect and characterize illicit COVID-19 product sales: Infoveillance study on Twitter and Instagram. JMIR public health and surveillance, 6(3), e20794.

[12] Maddox, T. M., & Matheny, M. A. (2015). Natural language processing and the promise of big data: small step forward, but many miles to go. Circulation: Cardiovascular Quality and Outcomes, 8(5), 463-465.

[13] Marquez, J. L. J., Carrasco, I. G., & Cuadrado, J. L. L. (2018). Challenges and opportunities in analytic-predictive environments of big data and natural language processing for social network rating systems. IEEE Latin America Transactions, 16(2), 592-597.

[14] Mazzei, M. (2020). A Machine Learning Platform for NLP in Big Data. Intelligent Systems and Applications, 245–259. https://doi.org/10.1007/978-3-030-55187-2_21

[15] Pais, S., Cordeiro, J., & Jamil, M. L. (2022). NLP-based platform as a service: a brief review. Journal of Big Data, 9(1). https://doi.org/10.1186/s40537-022-00603-5

[16] Panteli, A., Kompothrekas, A., Halkiopoulos, C., Boutsinas, B. (2021). An Innovative Recommender System for Health Tourism. In: Katsoni V., van Zyl C. (eds) Culture and Tourism in a Smart, Globalized, and Sustainable World. Springer Proceedings in Business and Economics. Springer, Cham. DOI:10.1007/978-3-030-72469-6_42

[17] Pasch, S., & Ehnes, D. (2022). NLP for Responsible Finance: Fine-Tuning Transformer-Based Models for ESG. 2022 IEEE International Conference on Big Data (Big Data). https://doi.org/10.1109/bigdata55660.2022.10020755

[18] Patel, J. M. (2020). Natural Language Processing (NLP) and Text Analytics. Getting Structured Data from the Internet, 135–223. https://doi.org/10.1007/978-1-4842-6576-5_4

[19] Pendyala, V. S., Atrey, N., Aggarwal, T., & Goyal, S. (2022). Enhanced Algorithmic Job Matching based on a Comprehensive Candidate Profile using NLP and Machine Learning. 2022 IEEE Eighth International Conference on Big Data Computing Service and Applications (BigDataService). https://doi.org/10.1109/bigdataservice55688.2022.00040

[20] Ray, J., Johnny, O., Trovati, M., Sotiriadis, S., & Bessis, N. (2018). The rise of big data science: A survey of techniques, methods and approaches in the field of natural language processing and network theory. Big Data and Cognitive Computing, 2(3), 22.

[21] Sharma, R., Agarwal, P., & Arya, A. (2022). Natural Language Processing and Big Data: A Strapping Combination. In New Trends and Applications in Internet of Things (IoT) and Big Data Analytics (pp. 255-271). Cham: Springer International Publishing.

[22] Tarnowska, K. A., & Ras, Z. (2021). NLP-Based Customer Loyalty Improvement Recommender System (CLIRS2). Big Data and Cognitive Computing, 5(1), 4. https://doi.org/10.3390/bdcc5010004

[23] Theodorakopoulos, L., Antonopoulou, H., Mamalougou, V., & C. Giotopoulos, K. (2022a). The drivers of volume volatility: A big data analysis based on economic uncertainty measures for the Greek banking system. Banks and Bank Systems, 17(3), 49–57. https://doi.org/10.21511/bbs.17(3).2022.05

[24] Theodorakopoulos, L., Halkiopoulos, C., & Papadopoulos, D. (2022b, September). Applying Big Data Technologies in Tourism Industry: A Conceptual Analysis. In International Conference of the International Association of Cultural and Digital Tourism (pp. 337-352). Cham: Springer International Publishing.

[25] Yayah, F. C., Ghauth, K. I., & Ting, C.-Y. (2018). Application of NLP on Big Data Using Hadoop: Case Study Using Trouble Tickets. Advanced Science Letters, 24(10), 7696–7702. https://doi.org/10.1166/asl.2018.13002

[26] Gkintoni, E., Koutsopoulou, I., Antonopoulou, H., & Christopoulos, P. (2021). CONSEQUENCES OF THE COVID-19 PANDEMIC ON GREEK STUDENTS'MENTAL HEALTH: QUALITY OF LIFE AND TRAUMA STRESSFUL EVENTS CORRELATION. In ICERI2021 Proceedings (pp. 2615-2621). IATED.

[27] Thanasas, G. L., Theodorakopoulos, L., & Lampropoulos, S. (2022). A Big Data Analysis with Machine Learning techniques in Accounting dataset from the Greek banking system. *Available at SSRN 4175698*.

[28] Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational intelligence magazine*, *9*(2), 48-57.

[29] Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243*.

[30] Li, J., Chen, X., Hovy, E., & Jurafsky, D. (2015). Visualizing and understanding neural models in NLP. *arXiv preprint arXiv:1506.01066*.

[31] Mihalcea, R., Liu, H., & Lieberman, H. (2006). NLP (natural language processing) for NLP (natural language programming). In *Computational Linguistics and Intelligent Text Processing: 7th International Conference, CICLing 2006, Mexico City, Mexico, February 19-25, 2006. Proceedings 7* (pp. 319-330). Springer Berlin Heidelberg.

[32] Zhou, M., Duan, N., Liu, S., & Shum, H. Y. (2020). Progress in neural NLP: modeling, learning, and reasoning. *Engineering*, *6*(3), 275-290.

[33] DeYoung, J., Jain, S., Rajani, N. F., Lehman, E., Xiong, C., Socher, R., & Wallace, B. C. (2019). ERASER: A benchmark to evaluate rationalized NLP models. *arXiv preprint arXiv:1911.03429*.

[34] Halkiopoulos, C., Gkintoni, E., Antonopoulou, H., & Skouroliakos, L. (2021). Behavioral Analysis of Personality, Branding and Emotional State in e-Sports. *Technium Soc. Sci. J.*, *24*, 434.

[35] Michalopoulos, D., Karras, A., Karras, C., Sioutas, S., & Giotopoulos, K. (2022). Neuro-Fuzzy Employee Ranking System in the Public Sector. *Volume 358: Fuzzy Systems and Data Mining VIII*, 325-333.

[36] Spirakis, P., Tampakas, B., & Antonopoulou, H. (1995). Distributed protocols against mobile eavesdroppers. In *Distributed Algorithms: 9th International Workshop, WDAG'95 Le Mont-Saint-Michel, France, September 13–15, 1995 Proceedings 9* (pp. 160-167). Springer Berlin Heidelberg.

[37] Voyiatzis, I., & Antonopoulou, H. (2007, September). Decoder-based Decompression for test sets containing don't cares. In *2007 International Conference on Design & Technology of Integrated Systems in Nanoscale Era* (pp. 229-232). IEEE.