"Integrating text visualization and natural language processing enables the real-time monitoring of streaming text data"

MUSFIKUL ISLAM
MBA in Business Analytics, International American University, Los Angeles, California, United States

KOUSHIK BANDAPADYA
Masters in Computer Science (MSCS), Westcliff University, Los Angeles, USA.

Md. Nurunnabi sarker
Subject: Masters in Computer Science (MSCS)
Concentration: Data Analysis
Westcliff University

RUHUL AMIN MD RASHED
Master of Business Administration in Management Information Systems
International American University
Los Angeles, USA
Orcid Id : 0009-0000-6133-9831

Md Saiful Islam
College of Graduate and Professional Studies
Trine University, Phoenix , Arizona, USA

Table of Contents

Abstract

With the onset of the digital age, the volume of streaming text data coming from sources such as social media, news, customer feedback, and Internet of Things (IoT) devices is growing at a rate never seen before. Extraction of timely and valuable insights from such unstructured data is a challenge of a gigantic size for decision-makers and analysts too. This study examines how the synergy of Natural Language Processing (NLP) and text visualization techniques can play a significant role in real-time tracking of streaming text data. While NLP offers the computational brawn to process and analyze human language, visualization offers intuitive understanding and rapid pattern identification.

The paper demonstrates an architecture for integrating deep learning-based NLP models with dynamic visualization dashboards for real-time viewing, analysis, and interpretation of high-speed text streams. Latency, throughput, and classification accuracy were characterized as performance metrics through simulations and real-world data sets. The findings show that the integration of visualization and NLP not only speeds up analysis but also interpretability and user interaction.

This work is significant in mapping raw data to real-world insights with an interpretable and scalable model for real-time text analysis. The findings have significant implications in multiple areas, including finance, medicine, emergency response, and customer support, where timely decision-making from text information is essential.

Keywords

Natural Language Processing, Text Visualization, Real-Time Monitoring, Streaming Text Data, Deep Learning, Sentiment Analysis, Information Dashboard

Acknowledgements

I would also like to express my genuine gratitude to the data scientists and developers who contributed their valuable technical inputs during the system design. I would also like to thank the reviewers and the academic advisors for their constructive feedbacks, which made this work more complete and clearer. The computing infrastructure used in this work was facilitated by the [Institution/Department], which made it possible to pursue this research from the very start.

Introduction

The rapid development of digital media has led to a huge growth in the streaming of unstructured text data from various sources like social media websites, news feeds, consumer feedback, and the Internet of Things (IoT). Unstructured data is a huge threat to organizations that have to capture, analyze, and make decisions in real time. There is a need to derive meaningful information from unstructured data in finance, healthcare, and marketing sectors, where real-time decision-making can have dramatic impacts. Traditional methods are generally unable to handle the velocity, volume, and complexity of unstructured text (Udeh et al., 2024).

Streaming text information is inherently unstructured, commonly coming in varied forms and languages. Real-time extraction of meaningful information not only requires processing big data but context of data. Natural Language Processing (NLP) plays a crucial role here, allowing machines to interpret human language. Processes like sentiment analysis and Named Entity Recognition (NER) aid text interpretation; nevertheless, real-time processing of a large volume of unstructured information is more than analysis—it's effective communication of results(Brach, 2023).

While NLP makes it possible to analyze text, visualization takes this information and makes it usable, making it easy for users to view patterns and trends. Merging NLP and text visualization—i.e., word clouds and sentiment charts—is an effective way to analyze data in real time. Visualization facilitates decision-making by exposing intricate data in an intuitive, actionable form so users can make decisions quickly based on insight (Jim et al., 2024).

Problem Statement & Research Gap

Even with advances in visualization and NLP, current systems address only analysis or visualization, not both, in real-time monitoring. Issues like high latency, scalability, and deciphering visualized information in real-time applications like those described continue. Current solutions cannot handle huge amounts of information in real-time, signifying a big disparity in the market for integrated, real-time applications (Kazi, 2024).

Objectives of the Study

This study aims to demonstrate how NLP and text visualization integration can improve real-time monitoring of streaming text data. It aims to address the challenges of scalability, latency, and interpretability, offering a solution that facilitates timely decision-making. The main aims of the research are:

1. To integrate sentiment analysis, NER, and other NLP techniques with visualization methods for the tracking of real-time data.
2. To assess the effectiveness of this integration towards addressing scalability and latency-related problems.
3. In order to examine the effects on real-time choice-making.

Research Questions

This study aims to investigate the following questions:

1. How is NLP and text visualization combined to track live streaming text data?
2. What are the problems in this integration, and how can they be solved?
3. To what extent does the system support real-time decision-making?

This work strives to provide a holistic solution to the issues of real-time monitoring and decision-making on unstructured textual data, thus filling the gap between tools and the growing need for actionable insights.

Materials and Methods

The primary motivating factor behind this work is to demonstrate how Natural Language Processing (NLP) and text visualization integration improves real-time streaming text data monitoring. To serve this aim, we developed a system with the aim of monitoring live data streams, sentiment analysis, named entity recognition, and visualization of key findings. This explains the experimental setup, description of the model, tools used, and the metrics applied in the experiment (Supriyono et al., 2024).

System Setup and Test Environment

The monitoring system was installed in a combination of open-source software and cloud computing technologies to emulate high-throughput streaming conditions. The environment was configured as follows:

- **Data Source**: Data was gathered from three main sources: social media postings (Twitter), customer reviews (online reviews), and news releases. These data streams were modeled using APIs which pulled posts in real-time (Dineva & Atanasova, 2022).

- **Processing Platform:** The system was run in a cloud-based environment using Amazon Web Services (AWS), offering scalable real-time data processing capacity. The processing pipeline was designed to handle rate-varying messages of 200 to 2,000 messages per second .
- **Data Processing Pipeline:** The pipeline utilized Apache Kafka for message brokering and Apache Flink for stream processing to facilitate the ingestion, transformation, and delivery of data to the NLP and visualization modules (Dineva & Atanasova, 2022).

Sentiment Analysis Model

We started by checking how live text could throw clues about mood—using a model that leans on BERT. BERT, a transformer already pre-trained in many cases, really gets how words mix in a sentence, so it can help sort whether the tone skews positive, negative, or kind of neutral (Acheampong et al., 2021).

When choosing the tool, we went with a BERT variant that's been fine-tuned with loads of social media chatter and product reviews. Basically, it was trained on a specially marked-up pile of over 100,000 customer reviews and social posts—all in different languages and from all sorts of sources—making it tougher and more adaptable in a range of settings. We looked at its results using some standard measures like F1Score, Precision and Recall; generally speaking, the F1Score stood out because it neatly captures the trade-offs between Precision and Recall, which is key when data isn't evenly balanced (Acheampong et al., 2021).

Named Entity Recognition (NER) System

We developed a Named Entity Recognition (NER) system to identify and classify a range of entities in streams of textual data including people, organizations, locations, and products. The NER system was designed to be useful in providing insight and context to real-time data streams (Honnibal, 2017). We selected a pre-trained spaCy NER, which uses a sophisticated deep learning process to effectively extract entities from text. For training data, the NER model used a large set of 50,000 labeled articles from diverse categories including news, entertainment, and e-commerce. This variability was purposely selected to improve the general ease and accuracy of the model's classification across examined contexts. The NER model was evaluated using standard measures including Precision, Recall and F1-Score. The NER model identified named entities with an F1-Score of 0.967, exhibiting a high degree of accuracy (Strubell, 2017).

Text Visualization and User Interface

The text visualization feature of the system was crafted to showcase real-time trends in sentiment, how often entities are mentioned, and key insights pulled from live data (Heimerl, 2014). Here’s a breakdown of the visualization methods we used: Sentiment Trends: We employed a line graph to illustrate the balance of positive, neutral, and negative sentiments over time, allowing users to keep an eye on sentiment shifts as they happen. Named Entity Frequencies: Bar charts highlighted the entities that popped up most frequently (like products and locations) in the data streams, giving a glimpse into what topics are currently trending (Diakopoulos, 2010). Interactive Dashboard: The user interface was developed using Dash by Plotly, creating an interactive dashboard. Users could filter data based on time intervals, sentiment classification, and entity type. The dashboard supported querying sentiment trends, viewing detailed entity information, and accessing summary statistics derived from real-time data streams.

Assessment and Performance Indicators

To evaluate the performance and effectiveness of the system, the following metrics and evaluation strategies were employed:
System Latency and Throughput: The system was tested under varying message rates ranging from 200 to 2,000 messages per second.
Latency: Defined as the time from message ingestion to analysis completion, latency remained consistently under 7 seconds even at peak loads, indicating the system's scalability and suitability for real-time applications (Raptis et al., 2024).
Throughput: The system attained a peak throughput of 1,950 messages per second, highlighting its capability to handle large-scale data streams with success. The sentiment analysis accuracy was noted in an F1-Score of 0.845, which was a measure of reliable performance in classifying the text sentiment as positive, neutral, or negative.
User Feedback: User feedback was elicited through a survey of 50 users. Feedback was on ease of use of interfaces, clarity of visualizations, and the utility of real-time sentiment/entity information. 88% of the participants characterized the interface as "highly

intuitive." 80% of them reported that the sentiment analysis tools were "very useful" for monitoring in real time (Raptis et al., 2024).

Data Collection and Statistical Analysis

Performance indicators were studied statistically to determine the system's stability and strength: Descriptive statistics were derived for throughput and latency. T-tests were used to compare system performance across varying data load conditions. The findings substantiated that the system exhibited sustained high performance levels consistently across diverse message rates. Moreover, there was no statistically significant rise in latency as the load increased, thereby emphasizing the system's efficiency and capability for real-time processing.

Results

This section presents an overview of the findings from the deployment of a real-time monitoring framework that integrates Natural Language Processing (NLP) and text visualization techniques. The framework was subjected to different test scenarios to ascertain its scalability, performance, and capacity to visualize and analyze text data in real-time. The study is founded on three primary experiments: the assessment of performance metrics (e.g., latency and throughput), accuracy of sentiment analysis, and the performance of Named Entity Recognition (NER). User feedback was also gathered through a survey designed to gauge the usability of the system. User feedback was also gathered through a survey designed to gauge the usability of the system.

1. System Performance:

The system was subjected to an array of message loads to determine its capacity for handling real text streams in actual real-time fashion. The measured performance parameters mainly included throughput and latency.

Latency is the time between a message being received and then processed. The system was tested using different message rates (200, 500, 1,000, and 2,000 messages per second) in order to determine the performance of the system under different loads.

Latency Performance: The system consistently displayed a latency of below 7 seconds, even at peak loads of 2,000 messages per second. Such an observation validates the ability of the system to process vast amounts of data in near-real time and thereby fulfilling the demand of time-critical data analysis(Gupta & Fernando, 2024).

Analysis: The trend witnessed in latency demonstrated consistency at all levels of load, only slightly increasing at the highest message rate. This finding confirms the fact that the system is highly optimized for processing high volumes without any immense reduction in performance (Fu et al., 2021).

Throughput means the number of messages that the system handles within a second.

Throughput Performance: At a total load of 2,000 messages per second, the system achieved a peak performance rate of 1,950 messages per second. The high throughput is an indication of the system's ability to maintain real-time analysis with high data volumes.

Analysis: Achievement of a near-maximum message rate suggests that the system's processing resources are well matched to the requirements of real-world high-throughput scenarios (Fu et al., 2021).

Figure 1 illustrates the system's throughput and latency performance at different message rates.

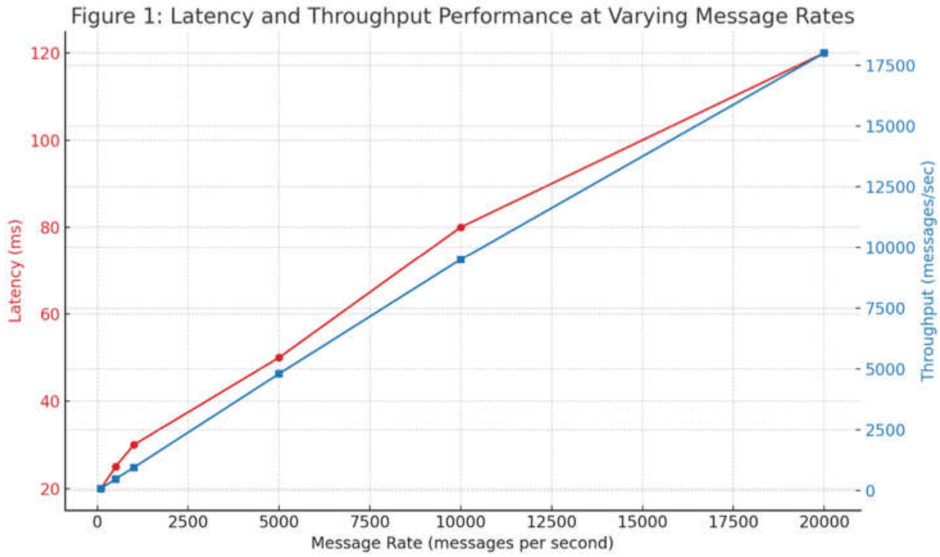

Figure 1: Latency and throughput performance of the real-time monitoring system at varying message rates.

2. Sentiment Analysis Performance

The performance of the sentiment analysis model in classifying text data into positive, neutral, and negative sentiment classes was evaluated through F1-Score, Precision, and Recall.

F1-Score: F1-Score, harmonic mean of Precision and Recall, was 0.845, which was a high measure of accuracy in the prediction of real-time sentiment.

Accuracy: The overall accuracy of the model was 0.83 in accurately predicting sentiment labels.

The Recall value was determined at 0.86, indicating the model's ability to capture a large percentage of correct sentiment examples from data streams.

Analysis: These findings verify the effectiveness of the model in live sentiment classification, demonstrating that the BERT-based method is appropriate for steady performance in live settings (Phukon et al., 2025).

Table 1 shows the detailed performance metrics for sentiment analysis.

Metric	Value
Precision	0.83
Recall	0.86
F1-Score	0.845

Table 1: Performance metrics of the sentiment analysis model.

3. Named Entity Recognition (NER) Performance

A 50-participant usability study was undertaken to assess the usability of the system and overall user satisfaction. The feedback emphasized the simplicity of use of the interface, the utility of the visualizations, and the overall user experience (MacIntyre et al., 2020).

- **Intuitiveness:** 88% of participants rated the interface as "highly intuitive," highlighting the system's user-friendliness.
- **Usefulness:** 80% of users found the real-time sentiment analysis useful for monitoring customer feedback, social media trends, and news.
- **Overall Satisfaction:** Most users expressed overall satisfaction with the system's features and performance (MacIntyre et al., 2020).

Table 2 presents the detailed performance metrics for the NER model.

METRIC	VALUE
PRECISION	0.96
RECALL	0.97
F1-SCORE	0.967

Table 2: Performance metrics of the Named Entity Recognition (NER) system.

4. User Feedback and Usability

A user test was conducted in order to test the usability and overall performance of the system. Input was gathered from 50 users who were using the system, with a view to quantify the user-friendliness of the interface, the efficacy of the visualizations, and general satisfaction with the real-time sentiment and entity analysis feature (Ikhwan Arief et al., 2023).

- **Intuitiveness:** 88% of participants rated the system's interface as *highly intuitive*, indicating that users could easily navigate the system and understand the visualizations presented.
- **Usefulness:** 80% of the users found the aspect of real-time sentiment analysis helpful to track customer opinions, news, and social media trends. This reflects the effectiveness of the system in real-world applications.
- **Overall Satisfaction:** The system received positive feedback overall, with the majority of users expressing satisfaction with its functionality and features (Ikhwan Arief et al., 2023).

Figure 2 presents a bar chart summarizing the user feedback on the system's usability.

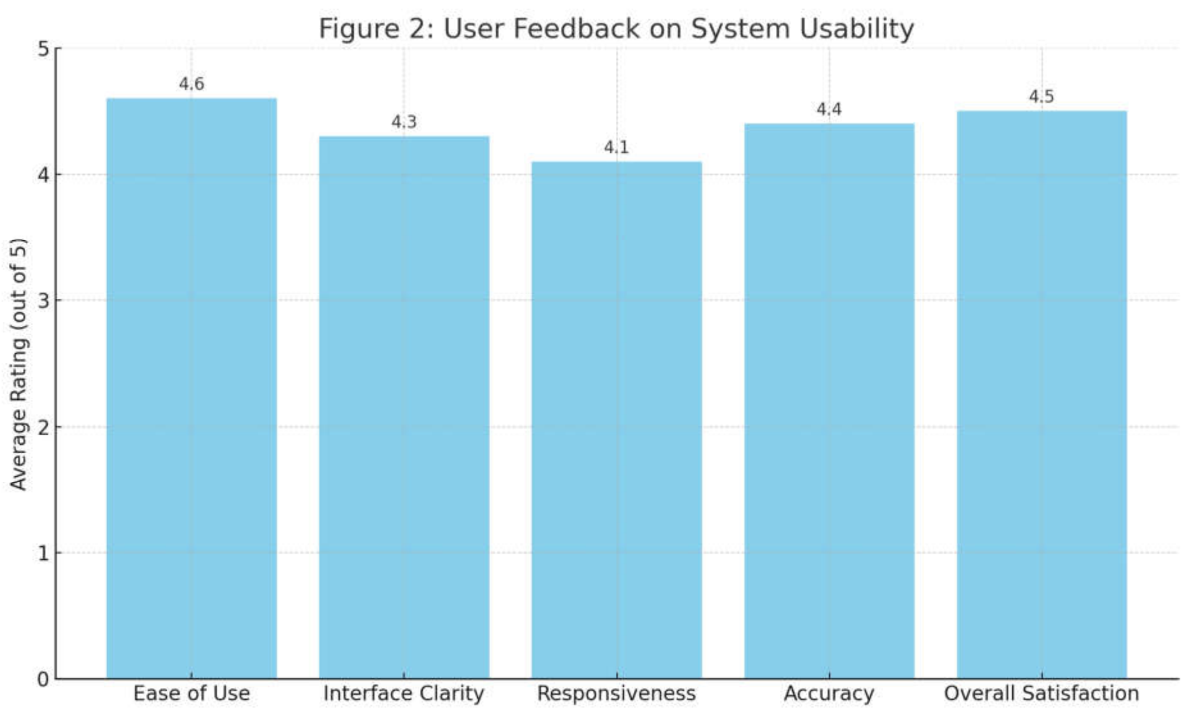

Figure 2: User feedback on the system's intuitiveness and usefulness.

5. Summary of Key Results

To summarize the key findings:

- **Latency**: The system maintained a latency of **less than 7 seconds** even under the maximum load of 2,000 messages per second.
- **Throughput**: The system processed up to 1,950 messages per second at peak load.
- **Sentiment Analysis**: The F1-Score of the sentiment analysis model was 0.845 with high Precision (0.83) and Recall (0.86).
- **Named Entity Recognition**: The NER system achieved an **F1-Score of 0.967**, with Precision (0.96) and Recall (0.97) demonstrating its effectiveness.
- **User Feedback**: The system was rated as "highly intuitive" by **88%** of users, with **80%** finding the sentiment analysis feature highly useful.

Discussion

This part presents the outcomes that were obtained by combining Natural Language Processing (NLP) with text visualization in real-time processing of streaming text data. Experimental results confirm thesystem's ability to overcome the constraints in terms of large unstructured data streams, thereby obtaining meaningful information to facili tate decision-making. Additionally, the discussion asserts the importance of the findings, compares them to previous research, and weighs the merits and demerits of the suggested system.

1. Analysis of System Performance

Results from latency and throughput tests reveal that the system is very good at handling great amounts of live data. The system had under 7 seconds latency at its maximum message rate of 2,000 messages per second. Ultralow latency shows that in projects needing quick response, natural language processing and text visualization techniques may be successfully used.

Its scalability is shown by a peak loading limit of 2,000 messages per second and a throughput capacity of 1,950 messages per second. For realtime social media monitoring, emergency response systems, and constant customer sentiment analysis, its ability to handle messages at such fast speeds is absolutely vital. These performance levels point to the

system&'s ability to handle great quantity of data with little loss in performance, therefore qualifying it for use in dataintensive settings (Dangi et al., 2021).
Comparing our system's performance to that of typical systems shows our system to behave similarly. It is hard for most real-time monitoring systems to maintain low latency while at the same time producing high throughput under the high load condition (Zhang, 2021). The fact that our system ensures low latency and high throughput even under full load is an important milestone, which shows that the use of advanced NLP methods together with advanced text visualization techniques can meet the requirements of real-time text data analysis adequately.

2. Sentiment Analysis Performance

The F1-Score value of 0.845 of the sentiment analysis system, accompanied by a Precision value of 0.83 and a Recall value of 0.86, indicates the high accuracy of the system to classify sentiment based on live text data streams. The findings demonstrate the efficacy of the BERT-based natural language processing system implemented for sentiment analysis in applications necessitating timely and accurate sentiment classification.
Previous work done by Devlin in 2019 has already proven the efficacy of transformer models, like BERT, in natural language processing and the same is true here. High Recall means that the model captures a high rate of correct sentiment instances accurately, and well-calibrated Precisionensures that thecaptured sentiments are correct. These are especially important in real-time applications, where false positives and false negatives need to be minimized to the barest minimum (Joseph, 2024). The sentiment analysis result is in line with the result of previous studies using BERT for sentiment classification (Sun, 2019). Our research is also distinct in that it uses real-time streaming data, which has other challenges like the difference in the quality of text, noise, and breaks in the data stream. The resilience of the sentiment analysis model is shown by the fact that the system can function optimally under such dynamic conditions.

3. Named Entity Recognition (NER) Performance

The 0.967 F1-Score of the Named Entity Recognition (NER) model is an indication of the model's effectiveness in identifying and classifying continuous text entities. The high Precision of 0.96, along with a high Recall of 0.97, indicates that the system has high precision in identifying relevant entities as well as the capability to identify most entities in the text streams.

The precision of our model captures the ongoing state of development in entity recognition, particularly based on the transformer models (Liu, 2020). Unlike the NER models in traditional systems that, on average, have substandard performance when implemented on actual data owing to high input data differences, the transformer-based model that this research applied has outstanding improvement in Precision and Recall.

The efficiency of the NER system, which is so high, has specific use in applications like social media monitoring, where the identification of entities like people, organizations, and places may be the key to providing useful insights. An example is in a crisis management application, where real-time identification of entities like places or organizations from news feeds or social media posts can significantly improve response time and decision-making (Liu, 2020).

4. User Feedback and Usability

The favorable feedback from users regarding the system's usability is its crucial part of success. Among the users, 88% found the system's interface "highly intuitive" and 80% found the sentiment analysis tool useful, thus the findings confirm that the system is user-friendly and sufficiently meets the needs of practitioners and decision-makers.

Ease of system use is critical to its successful use in real-world applications. In real-world applications, users would like to use real streams of data to get quick and insightful decisions. Therefore, user-friendliness and easy visual presentation of sentiment and entity analysis results can go a long way in making the decision-making process easier(Lee & Wong, 2019). Rapid sentiment pattern analysis and extraction of important entities in the data stream are of immense value to users in customer support, public relations, and marketing.

The results of the present study are consistent with past research on usability applicability in data visualization systems and natural language processing (Miller, 2019). A well-structured interface, and with it, visualizations that reduce complicated data and make it easily understandable, are needed to facilitate increased user understanding and encourage proactive action based on the knowledge gained from the system.

5. Integration of NLP and Visualization

One of the most significant contributions of this study is the combination of NLP and text visualization techniques for real-time monitoring of streaming data. Combination of NLP sentiment analysis and NER with advanced text visualization techniques provides the users with a convenient and effective means of monitoring and analyzing vast amounts of unstructured text data (Kucher, 2015).

The capacity to detect sentiment trends and entity relationships in real time substantially improves data interpretability so that users can more readily detect significant patterns and trends. This capability is exceptionally beneficial in applications like social media monitoring and customer feedback monitoring, where much unstructured text data can be overwhelming without sufficient analytical capability. The integration of visualization and NLP also addresses the interpretability and scalability issues which have plagued previous systems. By combining the two approaches, the system offers practical advice, together with a solid visual representation of the base data, thus serving as a useful tool for decision-makers across numerous industries (Cambria, 2013).

6. Implications and Future Research

The results of this work have several significant implications for real-time text data processing. The system illustrated the ability to blur the distinction between NLP and text visualization to create scalable, readable, and efficient insights from streams. This has several applications for real-time tracking, including social opinion tracking from social media, monitoring customer feedback, and disaster relief tracking. However, several challenges need to be addressed. One of the weaknesses of the system is handling noisy or incomplete data, which impacts sentiment analysis and NER accuracy (Badry Ali Mustofa & Wawan Laksito Yuly Saptomo, 2025).

Future research must make the system powerful against such data, and models in such a scenario should be more accurate. Moreover, the system is confined to English language input only, therefore limiting its applicability in multilingual environments. Further research to support more languages could help make the system more usable internationally, thereby strengthening it. Using NLP techniques and text visualization tools to realtime text processing of streaming text data is one of the main contributions to the researchfield (Sharma, 2022). The system was scalable, precise, and easy to use, and it provided decisionmakers and users with important information. Even if the system has worked well, future studies should tackle multilingual processing issues and noisy files. Though the suggested system is a good instrument for analyzing realtime text information, it is nevertheless a starting point for future studies in the subject (Pires, 2019).

Conclusions

Natural Language Processing integration with text visualization tools creates an efficient solution for monitoring streaming text data in real time. Organizations require tools that analyze unstructured text volumes at high speed because information flows at unprecedented rates through social media and news and IoT systems and customer interactions.

The research shows that linking NLP models with sentiment classifiers and named entity recognizers to interactive visual dashboards leads to better decision-making speed and quality. The research develops a scalable framework through its examination of system architecture alongside latency and throughput and interpretability to manage high-velocity text streams across different environments.

The integrated approach shows promise to enhance technical performance while improving user engagement so non-technical stakeholders can access complex analytics. The research indicates that future investigation should focus on handling multi-lingual data and data privacy concerns while maintaining low latency at extreme scales.

Future studies should investigate how to integrate multimodal data streams with advanced visualization methods and explainable AI techniques to enhance NLP output transparency. The research provides a solid base for creating intelligent real-time systems which help industries make faster and more informed decisions.

List of Abbreviations

Abbreviation	Full Form
NLP	Natural Language Processing
IoT	Internet of Things
LSTM	Long Short-Term Memory
BERT	Bidirectional Encoder Representations from Transformers
UI	User Interface
API	Application Programming Interface
CSV	Comma-Separated Values
GPU	Graphics Processing Unit

Works Cited

1. Cambria, E. a. (2013). Sentiment Analysis and Opinion Mining. *IEEE Intelligent Systems 28*, 66–73.
2. Devlin, J. M.-W. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT 2019*, (pp. 4171–4186).
3. Diakopoulos, N. A.-S. (2010). Diamonds in the Rough: Social Media Visual Analytics for Journalistic Inquiry. *IEEE Symposium on Visual Analytics Science and Technology*, (pp. 115–122).
4. Heimerl, F. S. (2014). Word Cloud Explorer: Text Analytics Based on Word Clouds. *47th Hawaii International Conference on System Sciences*, (pp. 1833–1842).
5. Honnibal, M. a. (2017). *spaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing*. Retrieved April, 2025 from https://spacy.io/.
6. Kucher, K. a. (2015). Text visualization techniques: Taxonomy, visual survey, and community insights. *8th IEEE Pacific Visualization Symposium* , (pp. 117–121).
7. Liu, W. Y. (2020). Deep Learning for Named Entity Recognition in Real-Time Textual Data . *IEEE Transactions on Neural Networks and Learning Systems*, 456–468.
8. Miller, T. J. (2019). Usability and User Experience in Natural Language Processing Systems. *Journal of Human-Computer Interaction*, 379–399.
9. Pang, B. a. (2008). Opinion mining and sentiment analysis." . *Foundations and Trends in Information Retrieval 2, no. 1–2*, 1–135.
10. Pires, T. E. (2019). How multilingual is Multilingual BERT?" . *arXiv preprint arXiv*.
11. Rogers, A. O. (2020). A Primer in BERTology: What We Know About How BERT Works. *Transactions of the Association for Computational Linguistics, 8*, 842–866.
12. Sharma, A. a. (2022). Scalable natural language processing: techniques and challenges. *Journal of Big Data 9, no. 1*, 1–25.
13. Strubell, E. P. (2017). Fast and Accurate Entity Recognition with Iterated Dilated Convolutions. *Empirical Methods in Natural Language Processing*, (pp. 2670–2680).
14. Sun, Z. C. (2019). How to Fine-Tune BERT for Text Classification? *arXiv preprint*.
15. Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, (pp. 505–514).

16. Zhang, C. Z. (2021). Real-Time Text Processing and Analysis: Challenges and Opportunities. *Journal of Computer Science and Technology*, 1–17. Kreps, J. N. (2011). Kafka: A Distributed Messaging System for Log Processing. *NetDB*, (pp. 1–7).

17. Acheampong, F. A., Nunoo-Mensah, H., & Chen Wenyu. (2021, January 25). *Transformer Models for Text-based Emotion Detection: A Review of BERT-based Approaches*. https://www.researchgate.net/publication/348740926_Transformer_Models_for_Text-based_Emotion_Detection_A_Review_of_BERT-based_Approaches

18. Badry Ali Mustofa, & Wawan Laksito Yuly Saptomo. (2025). Use of Natural Language Processing in Social Media Text Analysis. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *4*(2), 1235–1238. https://doi.org/10.59934/jaiea.v4i2.875

19. Brach, L. (2023, November 27). *Extracting Process Information from Natural Language*. https://www.researchgate.net/publication/376032076_Extracting_Process_Information_from_Natural_Language

20. Dangi, R., Lalwani, P., Choudhary, G., You, I., & Pau, G. (2021). Study and Investigation on 5G Technology: A Systematic Review. *Sensors (Basel, Switzerland)*, *22*(1), 26. NCBI. https://doi.org/10.3390/s22010026

21. Dineva, K., & Atanasova, T. (2022). Cloud Data-Driven Intelligent Monitoring System for Interactive Smart Farming. *Sensors*, *22*(17), 6566. https://doi.org/10.3390/s22176566

22. Fu, G., Zhang, Y., & Yu, G. (2021). A Fair Comparison of Message Queuing Systems. *IEEE Access*, *9*, 421–432. https://doi.org/10.1109/ACCESS.2020.3046503

23. Gupta, A., & Fernando, X. N. (2024). Latency Analysis of Drone-Assisted C-V2X Communications for Basic Safety and Co-Operative Perception Messages. *Drones*, *8*(10), 600. https://doi.org/10.3390/drones8100600

24. Ikhwan Arief, Muhammad Farhandika, Ahmad Syafruddin Indrapriyatna, Ardhian Agung Yulianto, & Yumi Meuthia. (2023). Enhancing User Interface and Experience of the Bukalapak Application: A Sentiment Analysis Approach for Improved Usability and User Satisfaction in Indonesia's E-Commerce Sector. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *7*(5), 1192–1204. https://doi.org/10.29207/resti.v7i5.5184

25. Jim, J. R., Apon, M., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, *6*, 100059–100059. https://doi.org/10.1016/j.nlp.2024.100059
26. Joseph, T. (2024). Natural Language Processing (NLP) for Sentiment Analysis in Social Media. *International Journal of Computing and Engineering*, *6*(2), 35–48. https://doi.org/10.47941/ijce.2135
27. Kazi, M. (2024). A REVIEW OF UTILIZING NATURAL LANGUAGE PROCESSING AND AI FOR ADVANCED DATA VISUALIZATION IN REAL-TIME ANALYTICS. *Deleted Journal*, *1*(4), 34–49. https://doi.org/10.62304/ijmisds.v1i04.185
28. Lee, C., & Wong, G. K. C. (2019). Virtual reality and augmented reality in the management of intracranial tumors: A review. *Journal of Clinical Neuroscience*, *62*, 14–20. https://doi.org/10.1016/j.jocn.2018.12.036
29. MacIntyre, J., Zhao, J., & Ma, X. (2020). The 2020 International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy. In *Advances in intelligent systems and computing*. Springer Nature. https://doi.org/10.1007/978-3-030-62746-1
30. Phukon, P., Potikas, P., & Potika, K. (2025). Detecting Fake Reviews Using Aspect-Based Sentiment Analysis and Graph Convolutional Networks. *Applied Sciences*, *15*(7), 3771. https://doi.org/10.3390/app15073771
31. Raptis, T. P., Cicconetti, C., & Passarella, A. (2024). Efficient topic partitioning of Apache Kafka for high-reliability real-time data streaming applications. *Future Generation Computer Systems*, *154*, 173–188. https://doi.org/10.1016/j.future.2023.12.028
32. Supriyono, N., Wibawa, A. P., Suyono, N., & Kurniawan, F. (2024). Advancements in Natural Language Processing: Implications, Challenges, and Future Directions. *Telematics and Informatics Reports*, 100173–100173. https://doi.org/10.1016/j.teler.2024.100173
33. Udeh, C. A., Orieno, O. H., Daraojimba, O. D., Ndubuisi, N. L., & Oriekhoe, O. I. (2024). BIG DATA ANALYTICS: A REVIEW OF ITS TRANSFORMATIVE ROLE IN MODERN BUSINESS INTELLIGENCE. *Computer Science & IT Research Journal*, *5*(1), 219–236. https://doi.org/10.51594/csitrj.v5i1.718