

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Advanced News Analytics : Sentimental and Summarization Analysis of news

Aayushi Singh^{*1}, Prakriti Patkar², Yash Kushwaha³, Dr. Nachiket Tapas⁴

Department of Computer Science and Engineering, Chhattisgarh Swami Vivekanand Technical University, Bhilai, India *Corresponding author: <u>aayushi.singh.0317@gmail.com</u>

ABSTRACT :

In an age where digital news is created at a never-before-seen pace, analyzing large amounts of information for relevance, sentiment, and credibility has emerged as a major challenge. This work introduces an AI-based news analytics system integrating deep learning models for news summarization and sentiment analysis. Through the use of **DistilBERT** and **RoBERTa** for sentiment classification and **BART** for abstractive summarization, the system is able to accurately analyze emotional tone and extract key information from news pieces. Comparative tests across lexicon-based (VADER, TextBlob), machine learning (Naive Bayes, SVM), and deep learning models prove that transformer-based models perform higher in terms of sentiment accuracy and coherence in summaries. The software is hosted as a full-stack web application with APIs (NewsAPI, Tweepy, Reddit PRAW) and has a strong backend pipeline for real-time sentiment and summary generation. The research advances digestible content, personalized news delivery, and decision-making, enabling scalable, smart news consumption systems.

Keywords: Sentiment Analysis, News Summarization, Machine Learning, Deep Learning, BERT, RoBERTa, TextRank, AI-based News Analytics, Abstractive Summarization, Extractive Summarization, NLP-based News Analysis.

INTRODUCTION

The sheer amount of news printed every day creates enormous challenges for users trying to consume and understand information efficiently. Conventional approaches to news analysis are unable to properly evaluate sentiment and decide on article relevance. Moreover, as the dissemination of misinformation is spreading, there arises an urgent requirement for automated systems that can supply not only sentiment classification but also proper summarization of news content.

This research suggests an AI-powered web portal aimed at addressing these problems through a holistic approach to sentiment analysis and news summarization. The system assesses the emotional tone of news stories based on conventional, machine learning, and deep learning algorithms. It also creates brief but informative summarizes for easier reading and offers personalized news suggestions based on user sentiment options. The study seeks to close the gap between summarization and sentiment classification so that users are provided with relevant, digestible, and unprejudiced news content. By merging current deep learning models with conventional means, the platform improves the quality of news consumption and aids in more informed decision-making.

OBJECTIVES

The primary objective of this research is to design and implement a robust AI-driven system that:

- Automates sentiment analysis using both classical and deep learning techniques.
- Summarizes news articles to ensure information clarity and conciseness.
- Performs comparative evaluation between traditional, ML, and DL models.
- Offers real-time news analysis through a web-based platform.
- Enhances user engagement through tailored sentiment-based recommendations.
- Improves the quality and efficiency of news consumption and decision-making.

LITERATURE REVIEW

Existing work in news sentiment and summarization has mostly been based on lexicon-based models or traditional machine learning algorithms. TextBlob and VADER are widely applied for rule-based sentiment analysis but are not good at handling sarcasm and contextual meaning. Naive Bayes and SVM models provide average performance but need heavy feature engineering. With the introduction of deep learning, models such as LSTM enhanced sequential text processing but not the global dependency modeling ability of the transformer architecture. Current research emphasizes that transformerbased models (BERT, RoBERTa, DistilBERT) produce better performance in sentiment classification tasks. Likewise, summarization work has shifted from TextRank extractive approaches to abstractive methods with transformer architectures such as T5, BART, and Pegasus, providing more fluent and informative summaries. Our research extends these advances by combining both tasks within a single scalable platform, confirmed through extensive experimentation.

SYSTEM DESIGN

Following is a basic system design description along with a block diagram description for adavance news analytics. **System Design Overview:**

- Data Layer: Gathers multi-source data through NewsAPI, Tweepy, and Reddit PRAW.
- Preprocessing Module: Tokenization, lemmatization, noise removal, and deduplication.
- Sentiment Engine: Combines VADER/TextBlob, Naive Bayes/SVM, and DistilBERT/RoBERTa for comparative sentiment analysis.
- Summarization Module: Utilizes BART for abstractive summaries; TextRank and BERTSUM for extractive summaries.
- Backend Server: Developed with Flask/FastAPI for API orchestration.
- Frontend Interface: React.js for dynamic news browsing and interaction.
- Storage Layer: MongoDB/PostgreSQL for storing news, sentiment, and summaries.

IMPLEMETATION

Implementation entailed a series of careful steps toward developing and hosting the AI-powered news analytics system. First, real-time news information was fetched through external APIs to have a constant flow of varied material from sources such as NewsAPI, Twitter, and Reddit. The obtained information passed through intense preprocessing using tools such as NLTK and SpaCy for tokenization, lemmatization, and removal of noise to improve text clarity and uniformity.

Subsequent to this, sentiment classification models varied from traditional algorithms such as Naive Bayes and SVM to deep learning transformers such as DistilBERT and RoBERTa were trained and tested on tagged datasets to gauge their performance. For summarization, the system incorporated Hugging Face's pipelines such as BART for abstractive summarization. Tools such as matplotlib were utilized to represent sentiment trends and summary performance for comparative evaluation.

To ensure that the system is accessible and scalable, a RESTful web service was created to deliver the predictions to users via a React-based frontend. The entire system was tested locally and then extended to cloud platforms for real-time, scalable performance in production environments.

WORKING PRINCIPLE

Here is the flow chart representation of model:



The system runs in a real-time pipeline that starts with multi-source news gathering via APIs (NewsAPI, Tweepy, Reddit PRAW). Raw data are cleaned and normalized via a preprocessing pipeline involving tokenization, lemmatization, and noise filtering via SpaCy and NLTK. The processed data is put through a multi-model sentiment engine that includes lexicon-based algorithms (VADER, TextBlob), traditional ML models (Naive Bayes, SVM), and **deep learning-based classifiers (DistilBERT, RoBERTa)** to identify sentiment polarity. Simultaneously, the text is input into summarization models: TextRank for extractive summary and **BART** for abstractive summary. The output is saved in a database and served using a RESTful API. The frontend retrieves and displays real-time sentiment distribution and summary highlights via dynamic React components and interactive charts. The architecture allows for uninterrupted, context-based analysis and display of news content, enhancing user comprehension and involvement.

EXPERIMENTAL RESULT

Experiments were conducted on datasets collected from NewsAPI, Twitter, and Reddit. DistilBERT achieved ~88% accuracy, outperforming VADER (65%) and Naive Bayes (72%). RoBERTa scored highest in F1-measure (~0.89). For summarization, BART outperformed TextRank in ROUGE-1, and ROUGE-L. Average summary length was reduced by 60% while maintaining context. A user study showed improved content understanding and decision-making in 70% of test users.

Below is the pie chart of sentimental analysis: DistilBERT and RoBERTa



Below is the text summarization model accuraccy: textRank and Bart-large-CNN



APPLICATIONS

The AI News Analytics tool has several practical uses to enhance how people read and understand digital news. Through the integration of sentiment analysis and summarization in one platform, the tool increases the relevance and readability of online news. It can be used by journalists and media outlets to track how the public feels about emerging stories, policies, or events in real-time, allowing them to do responsive and directed reporting. Policy makers and analysts find it useful to have rapid sentiment information and concise summaries across varied news sources for data-driven decision-making. End-users are facilitated with personalized news streams through the analysis of emotional tone and thematic content, and the reduction of information overload and cognitive fatigue. The system also has utility in education and research, where sentiment-aware news summarization can aid in social media analysis, studies of political discourse, and behavioral research. The implementation of the solution as a web application also provides for integration with browser extensions, mobile applications, and third-party platforms, making it more accessible and far-reaching.

FUTURE SCOPE

Further improvements to this system can go a long way toward expanding its potential and user base. One of the major areas includes adding multilingual support for sentiment analysis as well as summarization, making it possible for non-English content to be properly understood. Adding real-time fact-checking APIs can also further increase the credibility of the system by actively fighting disinformation. Making the summarization and sentiment response voice-based can also make the platform more accessible, particularly to visually impaired individuals or those who find auditory interaction preferable.

Personalization components such as user profiling can tune content suggestions in accordance with user preferences, habits, or sentiment patterns. In order to continue performing at a high level and remaining relevant, continuous learning systems will be embedded, enabling models to learn from realtime feedback from users and fresh data. Lastly, rolling out a mobile app version of the platform will improve accessibility and convenience, and users will be able to keep themselves updated anytime and anywhere.

CONCLUSION

This work introduces a real-time, scalable AI solution for summarization and sentiment of news articles. Building upon the latest state-of-the-art models such as **DistilBERT**, **RoBERTa**, **and BART**, our platform benefits over conventional and ML methods in terms of sentiment accuracy and summary informativeness. Implemented as a full-stack application, the platform is realized with real-world usability in filtering, analysis, and presentation of digital news content. Through comparative assessment and real-time testing, the system demonstrates its potential to reshape the way news is read, enabling informed decision-making and targeted engagement.

REFERENCES

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2018.

Available: https://arxiv.org/abs/1810.04805

[2] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper, and lighter," arXiv preprint arXiv:1910.01108, 2019.

Available: https://arxiv.org/abs/1910.01108

[3] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to Fine-Tune BERT for Text Classification?" arXiv preprint arXiv:1905.05583, 2019.

Available: https://arxiv.org/abs/1905.05583

[4] L. Tunstall, L. von Werra, and T. Wolf, Natural Language Processing with Transformers, O'Reilly Media, 2022.

Available: https://www.oreilly.com/library/view/natural-language-processing/9781098103231/

[5] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," arXiv preprint arXiv:1910.13461, 2019.

Available: https://arxiv.org/abs/1910.13461

[6] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," arXiv preprint arXiv:1910.10683, 2019.

Available: https://arxiv.org/abs/1910.10683

[7] Y. Liu and M. Lapata, "Text Summarization with Pretrained Encoders," arXiv preprint arXiv:1908.08345, 2019.

Available: https://arxiv.org/abs/1908.08345