Real-Time News & Social Media Sentiment Analyzer + Summarizer

1st Mr. Adnan Khan
B.Tech Final Year Student
Dept. of Computer Science & Engineering
MGM's College of Engineering, Nanded, Maharashtra, India
s22_khan_adnan@mgmcen.ac.in

2nd Mr. Ganesh Kalapad
B.Tech Final Year Student
Dept. of Computer Science & Engineering
MGM's College of Engineering, Nanded, Maharashtra, India
s22_kalapad_ganesh@mgmcen.ac.in

3rd Mr. Aniket Nagare
B.Tech Final Year Student
Dept. of Computer Science & Engineering
MGM's College of Engineering, Nanded, Maharashtra, India
s22_nagare_aniket@mgmcen.ac.in

4th Mr. Suhas Dube
B.Tech Final Year Student
Dept. of Computer Science & Engineering
MGM's College of Engineering, Nanded, Maharashtra, India
s22_dube_suhas@mgmcen.ac.in

5th Dr. Shital Y. Gaikwad
Guide - Senior Assistant Professor
Dept. of Computer Science & Engineering
MGM's College of Engineering, Nanded, Maharashtra, India
gaikwad_shital@mgmcen.ac.in

***Abstract* - The rapid increase in online news content and the spread of misinformation have made it difficult for users to identify reliable information, understand public sentiment, and consume lengthy articles efficiently. To address these challenges, this project introduces an intelligent real-time news analysis system that integrates news fetching, automatic summarization, sentiment analysis, and fake news detection into a unified platform. The system collects news articles from the web using automated scraping methods and processes them through advanced NLP models to generate concise and meaningful summaries, enabling users to understand key information at a glance.**

Sentiment analysis is performed using the VADER sentiment model, which evaluates the emotional polarity of the news content and classifies it into positive, negative, or neutral categories, providing insights into public mood and media tone. To ensure information reliability, a machine-learning–based fake news detection module is incorporated, which analyzes linguistic patterns and article characteristics to determine whether the fetched news is credible or suspicious.

The platform is designed with a clear and user-friendly interface that presents summaries, sentiment scores, and fake news predictions in real time, significantly reducing cognitive load and helping users make quick, informed decisions. By combining multiple analytical components into a single system, this project enhances the accuracy, trustworthiness, and accessibility of digital news consumption.

Index Terms* - *Abstractive Summarization, BART, T5, Natural Language Processing, Sentiment Analysis, Fake News Detection, Real-Time News Analysis.

I. INTRODUCTION

In today's digital age, the rapid increase in online news content and social media activity has led to an overwhelming amount of information for users to process. This information overload, compounded by the spread of misinformation, has made it difficult for individuals to identify reliable content, understand public sentiment, and efficiently consume lengthy articles. To address these challenges, this research focuses on the development of an intelligent real-time news analysis system that integrates news fetching, automatic summarization, sentiment analysis, and fake news detection.

The system collects news articles and social media content through web scraping, processing the data with advanced NLP models like T5 and BART for summarization. These models condense lengthy content into concise, meaningful summaries, enabling users to quickly grasp key information. Sentiment analysis is conducted using the VADER model, which classifies content into positive, negative, or neutral categories, providing insights into the emotional tone of the news. Additionally, a fake news detection module evaluates the credibility of the articles by analyzing linguistic patterns and article characteristics.

By combining real-time data processing with sentiment analysis and summarization, the platform enhances the accuracy, trustworthiness, and accessibility of digital news and social media consumption. This research proposes a unified solution that simplifies content consumption, reduces cognitive load, and helps users make informed decisions quickly. Future enhancements may include multilingual support, integration with diverse social media platforms, and optimization of transformer models for improved performance.

II. RELATED WORK

The rapid growth of digital news content has motivated extensive research in the fields of text summarization, sentiment analysis, and misinformation detection. Early approaches to news summarization primarily relied on extractive methods, where important sentences were selected based on statistical features such as term frequency, sentence position, and keyword relevance. Although effective to some extent, these methods often failed to preserve coherence and contextual meaning.

With advancements in deep learning, abstractive summarization techniques have gained prominence. Transformer-based models such as BART and T5 have demonstrated state-of-the-art performance in generating human-like summaries. BART combines bidirectional encoding with autoregressive decoding, enabling it to understand global context while generating fluent summaries. It has been widely used for news summarization tasks on benchmark datasets such as CNN/Daily Mail and XSum. Similarly, the T5 model reformulates all natural language processing tasks into a text-to-text framework, providing flexibility and improved transfer learning across multiple domains.

Sentiment analysis has also evolved significantly over time. Traditional machine-learning models required extensive feature engineering, whereas lexicon-based approaches such as VADER (Valence Aware Dictionary and Sentiment Reasoner) have proven effective for analyzing sentiment in news articles and social media content. VADER is particularly suitable for real-time applications due to its computational efficiency and ability to capture polarity intensity.

Fake news detection has become an important research area due to the increasing spread of misinformation. Existing approaches utilize machine-learning and natural language processing techniques to analyze writing style, semantic inconsistencies, and source credibility. Techniques such as bag-of-words, TF-IDF, and neural network-based classifiers have been employed to distinguish between reliable and deceptive content.

Despite significant progress, many existing systems focus on only one aspect of news analysis, such as summarization or sentiment detection, rather than providing an integrated solution. This project addresses this limitation by combining real-time news extraction, abstractive summarization using BART and T5, sentiment analysis using VADER, and fake news detection into a unified and user-friendly platform.

III. PROPOSED SYSTEM

The proposed system presents an intelligent real-time news analysis framework that integrates automated news extraction, abstractive text summarization, sentiment analysis, and reliability assessment into a unified platform. The primary objective of the system is to help users quickly understand large volumes of news content while gaining insight into the emotional tone and credibility of the information.

In the proposed approach, news data is collected in real time from online sources using web scraping techniques and URL-based content extraction. The retrieved articles undergo preprocessing steps such as noise removal, tokenization, stop-word elimination, and language detection to ensure clean and structured input for further analysis.

For sentiment analysis, the system employs the VADER (Valence Aware Dictionary and Sentiment Reasoner) model, which is specifically designed for analyzing sentiment in real-time textual data such as news articles and social media content. VADER computes sentiment polarity scores by evaluating lexical features and contextual sentiment intensity, classifying the content into positive, negative, or neutral categories. This enables users to understand the overall emotional tone of news events and media narratives instantly.

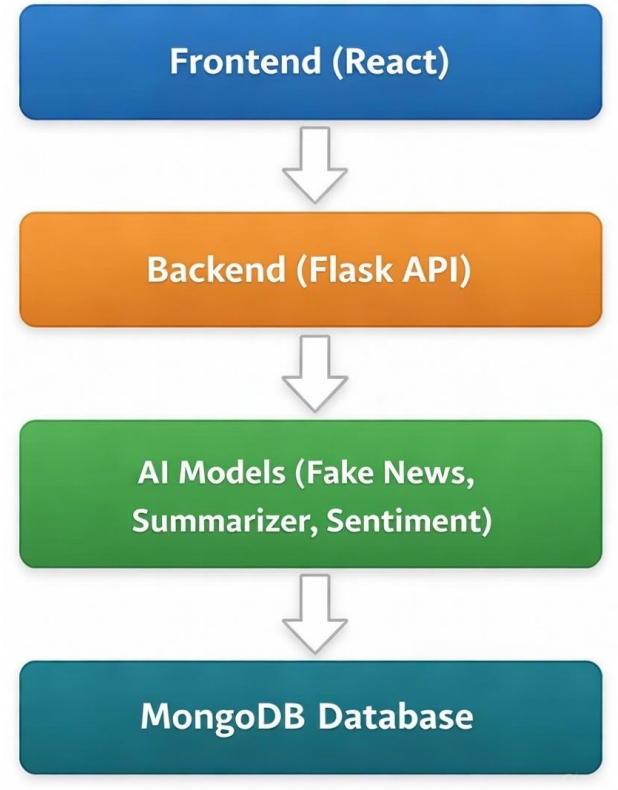
To enhance content comprehension, the processed news articles are passed through transformer-based abstractive summarization models such as BART and T5. These models generate concise and meaningful summaries that preserve the core context of the original article, significantly reducing reading time while maintaining semantic accuracy.

Additionally, the system incorporates a machine-learning–based fake news detection module that analyzes linguistic patterns, keyword distributions, and article characteristics to determine the credibility of the content. This provides an additional layer of trustworthiness, helping users identify potentially misleading or suspicious information.

The results are presented through an interactive and user-friendly interface that displays summarized text, sentiment scores, keyword highlights, and credibility indicators in

real time. Visualization tools are used to represent sentiment polarity trends, allowing users to quickly interpret public opinion and media tone.

III. SYSTEM ARCHITECTUR



The system architecture of the proposed real-time news sentiment analysis and summarization platform is designed in a layered manner to ensure modularity, scalability, and efficient processing. The architecture consists of four primary layers: User Input Layer, Processing Layer, Analysis Layer, and Output Layer.

The User Input Layer serves as the entry point of the system, allowing users to provide input in multiple formats such as raw text, PDF documents, images, or URLs of online news articles. This flexibility enables the system to handle diverse sources of news data.

The Processing Layer is responsible for extracting textual content from the provided inputs. For PDFs, PyPDF2 is used to extract text, while optical character recognition (OCR) using pytesseract is applied for image-based documents. Web-based news articles are retrieved using web scraping techniques implemented with BeautifulSoup. The extracted text undergoes preprocessing steps such as tokenization, stop-word removal, and language detection to ensure clean and structured input.

The Analysis Layer performs the core functionalities of the system. Abstractive summarization is carried out using transformer-based models such as BART and T5 to generate concise summaries. Sentiment analysis is performed using the VADER model, which computes polarity scores and categorizes sentiment as positive, negative, or neutral. In parallel, a fake news detection module evaluates the credibility of the content using machine-learning techniques based on linguistic and statistical features.

V. IMPLEMENTATION

The proposed system is implemented using Python due to its extensive support for natural language processing and machine learning libraries. Web scraping is performed using the BeautifulSoup library to collect real-time news articles from online sources. Text extraction from PDFs and images is handled using PyPDF2 and pytesseract, respectively.

For text summarization, pre-trained transformer models BART and T5 are utilized and fine-tuned for news summarization tasks. These models encode the input text and generate abstractive summaries that preserve contextual meaning while significantly reducing content length.

Keyword extraction is implemented using the CountVectorizer method from the Scikit-learn library, which identifies important terms by removing stopwords and emphasizing frequently occurring words. Sentiment analysis is implemented using the VADER sentiment analyzer, which produces compound sentiment scores and classifies text polarity efficiently.

The system interface is developed using Streamlit, enabling real-time interaction and dynamic visualization of sentiment results using graphical tools. The modular implementation allows easy integration of additional models and future enhancements.

VI. RESULTS AND ANALYSIS

The experimental results demonstrate that the proposed system effectively processes various input formats and generates accurate summaries, sentiment classifications, and keyword extractions. Both BART and T5 models produce high-quality abstractive summaries, with BART preserving more contextual detail and T5 generating more concise outputs.

Summarization performance is evaluated using ROUGE metrics, which indicate strong overlap between system-generated summaries and reference summaries. The keyword extraction results successfully highlight core themes of the input text, aiding rapid content understanding.

Sentiment analysis using VADER achieves reliable classification of news articles into positive, neutral, and negative categories. The sentiment scores align well with manually annotated sentiment labels, particularly for news and short-form textual content. Visual sentiment representations further enhance user comprehension.

The system maintains stable performance in real-time scenarios, demonstrating its suitability for continuous news monitoring and analysis applications.

VII. CONCLUSION AND FUTURE WORK

This research presents an integrated real-time news analysis system that combines abstractive text summarization, sentiment analysis, keyword extraction, and fake news detection into a unified framework. By leveraging transformer-based models such as BART and T5 along with the VADER sentiment analyzer, the system enables users to quickly comprehend large volumes of news content while gaining insights into emotional tone and information reliability.

The proposed system significantly reduces information overload and supports informed decision-making for students, researchers, journalists, and general users. Its modular architecture and real-time processing capabilities make it a practical solution for modern digital news consumption.

Future work will focus on incorporating multilingual support to analyze news in multiple languages, integrating social media data for broader sentiment analysis, and employing more advanced deep learning models for improved fake news detection accuracy. Additional enhancements may include real-time trend analysis and personalized news recommendations.

REFERENCES

[1] Alguliyev, R.M.; Aliguliyev, R.M.; Isazade, N.R.; Abdi, A.; Idris, N. (2019). COSUM: Text summarization based on clustering and optimization. Expert Systems, 36, e12340.

[2] Liu, F., Flanigan, J., Thomson, S., Sadeh, N., & Smith, N.A. (2015). Toward abstractive summarization using semantic representations.

[3] Lloret, E., Roma-Ferri, M. T., & Palomar, M. (2011). Compendium: A text summarization system for generating abstracts of research papers. In International Conference on Application of Natural Language to Information Systems (pp. 3–14). Springer.

[4] Sankarasubramaniam, Y., Ramanathan, K., & Ghosh, S. (2014). Text summarization using Wikipedia. Information Processing & Management, 50(3), 443–461.

[5] Song, S., Huang, H., & Ruan, T. (2019). Abstractive text summarization using LSTM-CNN based deep learning. Multimedia Tools and Applications, 78, 857–875.

[6] Yousefi-Azar, M., & Hamey, L. (2017). Text summarization using unsupervised deep learning. Expert Systems with Applications, 68, 93–105.

[7] Text Summarization with Sentimental Analysis. (2021). International Journal of Innovative Research in Computer Science & Technology, 9(4), 18–21.