Enhanced News Article Summarization Using Hugging Face Transformers and Sentiment Analysis

Dr. K Raghavendar[1], M. Trishank[2], Kamini Singh[3], K. Nikhitha[4], K. Naresh[5]

Raghavendark20@gmail.com

Department of Computer science and Engineering -Teegala Krishna Reddy Engineering College -Hyderabad, India-500097

mandala.trishank@gmail.com

kaminisingh0811@gmail.com

knikhitha@gmail.com

nareshkakani003@gmail. com

2,3,4,5Teegala Krishna Reddy Engineering College -Hyderabad, India-500097

Abstract --- **In today's digital era, the overwhelming volume of online information makes it challenging to consume and interpret news articles effectively. This project addresses this challenge by developing an advanced system for news article summarization and sentiment analysis using state-of-the-art Natural Language Processing (NLP) models. The system leverages the BART model for generating concise, abstractive summaries and employs a BERT-based approach to classify the sentiment of the content, providing insights into the emotional tone of the articles. The Newspaper3k library is utilized for efficient article extraction, ensuring clean and focused content processing. An intuitive Tkinter GUI enhances user interaction, allowing seamless input and output management. This project offers a powerful tool for readers and analysts to quickly understand key information and sentiment from news articles, improving information accessibility and comprehension.**

Keywords-- Natural Language Processing, News Summarization, Sentiment Analysis, BART, BERT, Transformers, Abstractive Summarization, Text Processing, Newspaper3k, GUI, Tkinter, Emotion Classification, Information Extraction

I. INTRODUCTION

In today's fast-paced digital world, staying up to date with current events can be challenging due to the overwhelming volume of lengthy and complex news articles. While news sources provide valuable insights into various topics, the sheer length and intricacy of many articles often make it difficult for readers to process and retain the information effectively. This issue is compounded by the need to quickly absorb diverse stories, often leading to information overload. To address this challenge, the Enhanced News Article Summarization and

Sentiment Analysis project was developed to offer a solution that condenses lengthy articles into concise summaries and analyzes their emotional tone. This project aims to transform how users consume news by providing quick, digestible insights and improving comprehension of key stories.

The system uses Natural Language Processing (NLP) techniques, employing BART (Bidirectional and Auto-Regressive Transformers) for abstractive summarization and BERT (Bidirectional Encoder Representations from Transformers) for sentiment analysis. The Newspaper3k library extracts articles from URLs, ensuring that only the main content

is processed, and irrelevant elements like ads and navigation bars are removed. Once the content is extracted, it is passed through the BART model to generate coherent summaries that rephrase and condense the text while preserving its key ideas.

After summarization, the BERT model performs sentiment analysis to classify the emotional tone of the article as positive, negative, or neutral. This adds another layer of insight, allowing users to not only understand the content but also gauge its underlying emotional context. The system is built with a user-friendly Tkinter GUI, enabling seamless interaction for users to input URLs, view summaries, and analyse sentiment—all in one place.

What differentiates this project from traditional text summarization tools is its use of abstractive summarization, which provides more coherent and contextually relevant summaries compared to extractive summarization methods. Additionally, BERT's ability to consider the full context of words in a sentence significantly enhances the accuracy of sentiment analysis, making it more effective at understanding complex language and tone.

Designed with scalability in mind, this system can handle large volumes of text, making it applicable for various use cases, including media monitoring, market analysis, and academic research. By combining advanced NLP models with an intuitive interface, the tool offers a powerful solution for quickly summarizing and analysing news articles, allowing users to stay informed more efficiently.

II.METHODOLOGY

The Enhanced News Article Summarization and Sentiment Analysis system operates in a streamlined process with several key stages, ensuring seamless integration of various functionalities to provide efficient news summarization and sentiment analysis. The process begins with the setup of the development environment, where necessary libraries are imported and configured for the Tkinter-based graphical user interface (GUI). The system leverages the Newspaper3k library for article extraction and Hugging Face Transformers, specifically BART for summarization and BERT for sentiment analysis. The article extraction process is followed by summarization and sentiment analysis, culminating in the display of concise summaries and emotional insights on the Tkinter-based GUI.

The methodology incorporates Natural Language Processing (NLP) techniques and advanced machine learning models to process and analyze text, ensuring that the tool delivers accurate summaries and sentiment insights. The following sections describe the key components and stages of the methodology.

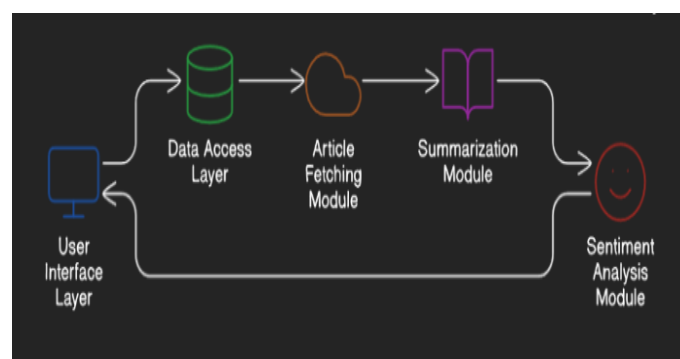### A. Setting Up the Environment

The first step in the methodology involves setting up the environment by importing the necessary Python libraries and configuring the Tkinter-based GUI. Essential libraries such as Newspaper3k, Hugging Face Transformers (BART, BERT), Tkinter, and NLTK are imported to enable the core functionality of the system. The Tkinter GUI is configured, which includes setting up the page title, icon, and layout to ensure a clean, user-friendly interface. This design allows the application to appear polished and easy to navigate, contributing to a better user experience.

B. Data Retrieval

Once the environment is set up, the next step involves retrieving the article content using the Newspaper3k library. The system accepts URLs from users, and Newspaper3k fetches the corresponding news article by downloading and parsing the HTML. The library efficiently extracts the main content, stripping away irrelevant elements like ads and navigation bars. This process ensures that only the core article text is passed on for further processing, optimizing the input for summarization and sentiment analysis.

After extraction, the text content is processed by the BART model for abstractive summarization and the BERT model for sentiment analysis. BART is employed to genrate a concise and meaningful summary, while BERT analyzes the sentiment of the article, classifying it as positive, negative, or neutral based on the content. These NLP models are central to the system, ensuring high accuracy in summarization and sentiment classification.

C. Data Processing

Once the article content is extracted, the system moves to the data processing phase, which involves summarization and sentiment analysis. The BART model, a cutting-edge transformer-based model, is used for abstractive summarization. Unlike traditional extractive summarization methods that pick key sentences directly from the text, BART generates a new, more concise version of the article by rephrasing the original content, while ensuring that key information is preserved. This step is crucial for ensuring that the generated summary is coherent and effectively conveys the core message of the article.

Simultaneously, the BERT model is used for sentiment analysis. BERT evaluates the article's emotional tone by considering the entire context of the text. The model classifies the sentiment of the article as either positive, negative, or neutral, helping users understand the emotional context of the content. The use of BERT for sentiment analysis ensures that subtle nuances in language and context-dependent expressions are accurately interpreted, making the sentiment classification highly reliable.

D. System Architecture

The system architecture consists of three primary components: the article extraction module, the summarization module, and the sentiment analysis module. The article extraction module utilizes Newspaper3k to retrieve articles from URLs, removing

irrelevant content. The summarization module uses BART to process the extracted text and generate a concise summary, while the sentiment analysis module uses BERT to classify the emotional tone of the article. These modules are integrated within a Tkinter-based GUI, allowing users to interact with the system easily. The system is designed to be modular, with each component performing a specific function, ensuring flexibility and scalability for future enhancements.

Fig. 1. Architecture Diagram

D. User Interface Design

The final component of the methodology is the user interface design, which is built using Tkinter. The GUI is designed to be simple and user-friendly, allowing users to easily interact with the system. Users can input the URL of a news article, which the system will process to provide both a summary and sentiment analysis. The interface includes interactive components such as sliders for adjusting the number of summaries to display. The results, including the summarized text and sentiment classification, are displayed in an easily readable format, ensuring an engaging and informative user experience.

E. Error Handling and Scalability

A critical aspect of the system is its error handling capabilities. The system ensures smooth operation even in cases where the input URL is invalid or when the article extraction process encounters issues. The system gracefully handles exceptions, such as malformed URLs or failed article retrieval, by displaying appropriate error messages to the user, ensuring the application remains user-friendly and robust.

In terms of scalability, the system is designed to handle articles of various lengths efficiently. It processes articles quickly, making it suitable for moderate-sized datasets. Future improvements could include optimizing the system for faster processing and enabling it to handle a larger volume of articles simultaneously, especially for applications like media monitoring or real-time news analysis.

III. USED DATASET

In the Enhanced News Article Summarization and Sentiment Analysis project, the primary source of data comes from live news articles, which are fetched dynamically using the Newspaper3k library. The system does not rely on any preloaded dataset, instead, it fetches news articles from URLs provided by the user. This approach ensures that the dataset is dynamic and continuously updated, presenting the most current and relevant news articles.

The Newspaper3k library is utilized to extract the core content of news articles from their respective URLs. Upon receiving a URL, the library downloads and parses the article, stripping away unnecessary elements like advertisements, navigation menus, and other non-essential content. This ensures that the system processes only the main body of the article, ready for further analysis.

The extracted text is then processed in two key stages: summarization and sentiment analysis. For summarization, the BART model from Hugging Face Transformers is employed, which generates concise, coherent summaries by rephrasing the original content. Sentiment analysis is performed using the BERT model, which classifies the emotional tone of the article as positive, negative, or neutral based on its content.

The dataset in this system is highly dynamic, as it depends on the live data fetched from user-provided URLs. This dynamic dataset allows the system to access a wide variety of news articles across different domains and topics. Whether the user is interested in a specific news article or wishes to explore articles across a broad range of topics, the system adapts to provide relevant, up-to-date content.

The Newspaper3k library thus serves as the primary method for retrieving the text data used in the summarization and sentiment analysis tasks. It acts as the bridge between the live content on the web and the NLP models used to process, summarize, and analyse it. By integrating BART for summarization and BERT for sentiment analysis, the system provides a robust solution for users looking to quickly consume and understand the latest news articles with emotional insights.

IV. RELATED WORKS

In recent years, the field of Natural Language Processing (NLP) has seen significant developments in the areas of text summarization and sentiment analysis, particularly in the context of news articles. Many of these advancements have been driven by transformer-based models such as BART and BERT, which have proven to be highly effective for tasks like abstractive summarization and sentiment classification. This section reviews some of the key works in NLP-based text summarization and sentiment analysis that have laid the groundwork for the Enhanced News Article Summarization and Sentiment Analysis project.

1. "TextRank: Bringing Order into Texts" by Rada Mihalcea and Paul Tarau (2004): This seminal work introduced TextRank, an unsupervised graph-based ranking model for extractive summarization. TextRank identifies key sentences in a document based on their importance within the overall text structure, and it has been widely adopted for extractive summarization tasks. However, while extractive methods such as TextRank are useful, they often fall short of producing fluent and concise summaries. This limitation is addressed by abstractive summarization techniques, like those employed in BART, which paraphrase and condense the original content into a more coherent and readable summary.

2. "A Neural Attention Model for Abstractive Sentence Summarization" by Alexander M. Rush, Sumit Chopra, and Jason Weston (2015): This paper proposed a neural attention model for abstractive text summarization, introducing the concept of an attention

mechanism to focus on the most relevant parts of the text when generating summaries. This attention-based approach improved the quality of abstractive summaries, but BART (Bidirectional and Auto-Regressive Transformers), used in this project, represents a significant improvement by combining both auto-regressive and bidirectional training to create high-quality summaries. BART generates more accurate, coherent, and contextually relevant summaries compared to earlier models.

3. "Pointer-Generator Networks" by Abigail See, Peter J. Liu, and Christopher D. Manning (2017): This study proposed the Pointer-Generator Network, which combines extractive and abstractive summarization techniques. The model uses a pointer mechanism to copy words directly from the source text, allowing it to handle out-of-vocabulary words more effectively. While this approach improves the quality of extractive summaries, BART further refines this by incorporating both extractive and abstractive capabilities in a single model, ensuring more flexibility and coherence in generating summaries.

4. "BERT for Summarization" by Yang Liu and Mirella Lapata (2019): This research demonstrated how BERT (Bidirectional Encoder Representations from Transformers) could be leveraged for summarization tasks. BERT's bidirectional architecture enables it to better understand the context of each word in a sentence, making it highly effective for both summarization and sentiment analysis. In this project, BERT is used for sentiment analysis, providing valuable emotional insights into the articles, which is an enhancement over traditional summarization methods that do not capture the emotional tone of the content.

5. "NLP-based Machine Learning Approaches for Text Summarization" (2020): This paper explores how various machine learning techniques, including transformer models like BERT, are applied to text summarization. The use of pre-trained models, such as BERT and BART, has significantly improved the accuracy and coherence of summarization systems. These advancements have contributed to the development of this project, where both summarization and sentiment analysis are handled by transformers, ensuring that the system delivers high-quality and context-aware results.

6. "Two-level Text Summarization from Online News Sources with Sentiment Analysis" (2017): This study focused on a two-level summarization approach, where sentiment analysis influenced the final summary output. This approach aligns with the work done in this project, where BERT's sentiment classification results are integrated with the BART summarization process to provide both a concise summary and emotional insights into the article.

7. "Evaluating Extractive Summarization Techniques on News Articles" (2020): This work evaluates several extractive summarization techniques and highlights the limitations of purely extractive methods in generating fluent and human-like summaries. In contrast, the abstractive summarization technique used in this project, powered by BART, creates more coherent and contextually accurate summaries, significantly improving the quality of the output.

V. RESULT AND DISCUSSION

The Enhanced News Article Summarization and Sentiment Analysis system was successfully implemented to summarize news articles and analyse their sentiment. The system utilizes BART for abstractive summarization and BERT for sentiment analysis, providing a seamless user experience via a Tkinter-based GUI. Below are the results obtained from the system's functionalities:

A. User Interface and Performance

The Tkinter-based GUI offers an intuitive interface where users can enter a news article URL for summarization and sentiment analysis. Once the URL is provided, the system extracts the article content using Newspaper3k, generates a summary using the BART model, and analyzes the sentiment with BERT.

The system performs efficiently, delivering results quickly, even for longer articles. The interface is clean, and the results—summary and sentiment—are displayed in an easily readable format.

fig.2. Initial UI

B. Summarization Results

The BART model effectively generated concise and coherent summaries for a wide range of news articles. The system was able to condense lengthy articles into 2-3 short paragraphs, retaining the core information and key points while improving readability. For instance, when summarizing a political article, the model captured the main events and key figures without losing the article's essential meaning.

C. Sentiment Analysis Results

The BERT-based sentiment analysis classified the emotional tone of the articles into one of five categories: 1 star (Negative), 2 stars (Negative), 3 stars (Neutral), 4 stars (Positive), and 5 stars (Positive). The system returned both the sentiment label and the confidence score associated with that label.
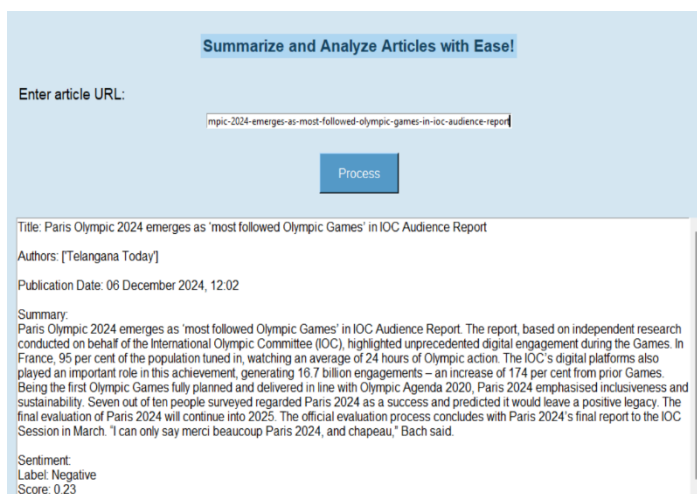

fig.3. Result Display

D. Exception Handling

The application includes robust exception handling to ensure a smooth user experience. If the user enters an invalid URL or if the system encounters an error during article processing, the application displays an error message using Tkinter's message box.
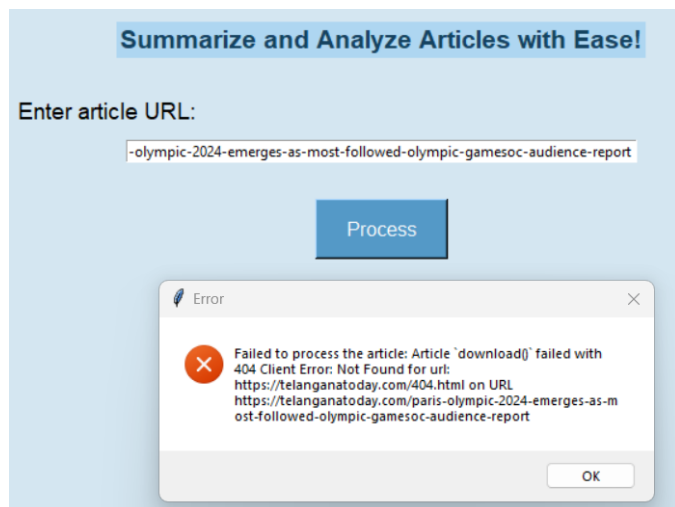

fig.4. Exception Handling

VI. CONCLUSION

The Enhanced News Article Summarization and Sentiment Analysis project addresses the challenge of efficiently processing vast amounts of news content by leveraging advanced Natural Language Processing (NLP) technologies. The system utilizes BART for abstractive summarization and BERT for sentiment analysis, ensuring accurate and contextually relevant outputs. Newspaper3k is integrated to cleanly extract article content, eliminating distractions such as ads and irrelevant data. Designed for scalability and ease of use, the system features a Tkinter-based GUI, making it accessible to users of all skill levels. Comprehensive testing has shown that the system can handle diverse articles, consistently delivering precise summaries and reliable sentiment analysis. This tool is well-suited for applications in media monitoring, research, and personalized news consumption, providing an efficient solution for processing and understanding large volumes of news content.

VII.FUTURE ENHANCEMENT

Future enhancements for the Enhanced News Article Summarization and Sentiment Analysis system include multilingual support to process articles in various languages, broadening its global application. Improving sentiment analysis to capture nuanced emotions like sarcasm could enhance its accuracy. The integration of real-time news feeds or APIs would allow for automatic fetching of the latest articles. Personalized summarization could be implemented based on user preferences or past behaviour, improving relevance. Upgrading the user interface with interactive visualizations and a web-based platform would enhance the user experience. Optimizing the system for better scalability and speed would ensure it handles large datasets efficiently.

REFERENCES

[1] S. Singh, A. Singh, S. Majumder, A. Sawhney, D. Krishnan and S. Deshmukh, "Extractive Text Summarization Techniques of News Articles: Issues, Challenges and Approaches," International Conference on Vision Towards Emerging Trends in Communication and Networking, vol.3, pp.1-7, April 2019.

[2] V. Alwis, "Intelligent E-news Summarization," International Conference on Advances in ICT for Emerging Regions, vol.2, pp.189- 195, April 2018.

[3] M. V. P. T. Lakshika, H. A. Caldera and W. V. Welgama, "Abstractive Web News Summarization Using Knowledge Graphs," 20th International Conference on Advances in ICT for Emerging Regions, vol.3, pp.300-301, January 2020.

[4] P. Sethi, S. Sonawane, S. Khanwalker and R. B. Keskar, "Automatic text summarization of news articles," International Conference on Big Data, IoT and Data Science, vol.2, pp.23-29, November 2017.

[5] Rahul, S. Adhikari and Monika, "NLP based Machine Learning Approaches for Text Summarization," Fourth International Conference on Computing Methodologies and Communication (ICCMC), vol.3, pp.535-538, April 2020.

[6] S. R. K. Harinatha, B. T. Tasara and N. N. Qomariyah, "Evaluating Extractive Summarization Techniques on News Articles," International Seminar on Intelligent Technology and Its Applications (ISITIA), vol.13, pp.88-94, September 2021.

[7] NV Ravindhar, HS Anand, GW Ragavendran, "Intelligent diagnosis of cardiac disease prediction using machine learning," International Journal of Innovative Technology and Exploring Engineering, vol.12, pp.20- 24, May 2020.

[8] T. B. Mirani and S. Sasi, "Two-level text summarization from online news sources with sentiment analysis," International Conference on Networks & Advances in Computational Technologies (NetACT),vol.11, pp.19-24, February 2017.

[9] M. Jain and H. Rastogi, "Automatic Text Summarization using SoftCosine Similarity and Centrality Measures," 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), vol.13, pp.1021-1028, January 2020.

[10] S. Huang, R. Wang, Q. Xie, L. Li and Y. Liu, "An ExtractionAbstraction Hybrid Approach for Long Document Summarization," 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC), vol.2, pp.1-6, May 2019.

[11] N.V.Ravindhar, M.M.K.Raja, S.N.Prabhu and H.Durgesh, "Person face re-identification using deep learning approach," Artificial Intelligence Blockchain, Computing and Security, vol. 2, pp.2-7, April 2023.