Proceedings of 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS 2024)
IEEE Xplore Part Number: CFP24DJ3-ART; ISBN: 979-8-3503-7999-0

NLP based Automated Text Summarization and Translation: A Comprehensive Analysis

Nikhil Zade
Department of Information Technology
JSPM's Rajarshi Shahu College of Engineering
Pune, India
nikhilzade199@gmail.com

Dr. Gitanjali Mate
Department of Information Technology
JSPM's Rajarshi Shahu College of Engineering
Pune, India
gsmate_it@jspmrscoe.edu.in

Kamal kishor
Department of Information Technology
JSPM's Rajarshi Shahu College of Engineering
Pune, India
kamalkumar8653@gmail.com

Nishant Rane
Department of Information Technology
JSPM's Rajarshi Shahu College of Engineering
Pune, India
nishantrane.it13@gmail.com

Manmath Jete
Department of InformationTechnology
JSPM's Rajarshi Shahu Collegeof Engineering
Pune, India
manmathjethe91@gmail.com

2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS) | 979-8-3503-7999-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSCSS60660.2024.10624907

Abstract: **Extractive text summarization stands as a fundamental pursuit within natural language processing, offering the capability to distil extensive textual content while preserving essential information. This research study presents a Graphical User Interface (GUI) application developed in Python using the TKinter library, designed to streamline the process of document summarization. Leveraging advanced image processing techniques, including face recognition through libraries such as OpenCV and PIL, the proposed system integrates robustsecurity measures for user registration and login functionalities. By utilizing NLP, speed and accuracy, our system offers scalable and adaptable solution for text summarization and language translation with accuracy between 91% to 95%. By employing efficient algorithms, users can extract pivotal sentences from documents, facilitating expedited comprehension and analysis. The fusion of text summarization with secure authentication mechanisms addresses both productivity and security concerns within document management systems, culminating in a professional-grade solution tailored to contemporary information processing needs.**

Keywords— Text summarization, NLP, Extractive Summary, Abstractive Summary, Deep Learning

I. INTRODUCTION

In today's digital age, the exponential growth of textual data poses significant challenges for individuals and organizations seeking to extract meaningful insights efficiently. Among the myriad solutions offered by Natural Language Processing (NLP), extractive text summarization stands out as a vital tool for condensing lengthy documents while preserving essential information. This study addresses the pressing need for effective text summarization tools by developing a professional-grade Graphical User Interface (GUI) application using Python's TKinter library. Beyond its summarization capabilities, this research study integrates advanced image processing techniques, particularly in the face recognition to ensure robust user authentication and security. Through the utilization of libraries such as OpenCV and PIL, the system enhances user experience by providing a seamless and secure environment for both registration and login functionalities. This study aims to empower users with the ability to efficiently extract key insights from documents, enabling quicker comprehension and analysis. By bridging the gap between productivity and security concerns within document management systems, it seeks to offer a comprehensive solution tailored to contemporary challenges in information processing and document management.

II. RESEARCH OBJECTIVES

- To develop a user-friendly Graphical User Interface (GUI) application using Python's TKinter library to facilitate efficient document summarization.

- To implement advanced image processing techniques, including face recognition via OpenCV and PIL libraries.
- To employ state-of-the-art algorithms for extractive text summarization, enabling the extraction of key sentences fromdocuments while preserving their semantic meaning and contextual relevance.
- To integrate robust error handling mechanisms to handle diverse document formats and potential anomalies.
- To provide comprehensive documentation and user support resources to facilitate seamless adoption and utilization of thesummarization application.

III. LITERATURE REVIEW

[1] provides an extensive survey of recent advancements in extractive text summarization techniques, covering topics such as deep learning approaches, graph-based methods, and reinforcement learning algorithms. [2] focused on the integration of textual and visual information for summarizing multimedia content and discuss the potential applications and research directions. [3] addresses the need for personalized summarization solutions and proposed a new approach for user profiling, content recommendation, and adaptive summarization, highlighting the potential benefits and challenges of personalized summarization systems. [4] focused on analyzing the increasing demand for timely information processing, continuous summarization algorithms, and real-time summarization evaluation metrics, offering insights intothe state-of-the-art in real-time summarization research. [5] provided a comprehensive review of evaluation metrics used in text summarization research. The study discusses about the commonly used metrics such as ROUGE, BLEU, and METEOR to evaluate the strengths and limitations, and proposed guidelines for selecting appropriate evaluation metrics based on summarization task characteristics. [6] explored the growing demand for cross-lingual summarization solutions while [7] focused on the summarization of scientific literature by analyzing the recent advances and challenges in summarizing research articles, scholarly articles, and academic documents.

IV. PROPOSED SYSTEM

The proposed system aims to address the shortcomings of existing document summarization tools by offering a professional-grade Graphical User Interface (GUI) application developed using Python's TKinter library. This interface will prioritize user-friendliness, ensuring accessibility for individuals with varying levels of technical expertise. To enhance security, the system will integrate advanced image processing techniques, such as face recognition via OpenCV and PIL libraries, for secure user authentication.

Additionally, the system will implement state- of-the-art algorithms for extractive text summarization, enabling the extraction of key sentences from documents while preserving semantic meaning and contextual relevance. Furthermore, the proposed system will feature robust error handling mechanisms to ensure stability and reliability under diverse conditions. Comprehensive documentation and user support resources will be provided to empower users to leverage the application effectively for enhanced productivity and information management. By bridging the gap between usability, security, and functionality, the proposed system seeks to offer a comprehensive solution tailored to contemporary challenges in document summarization and management.

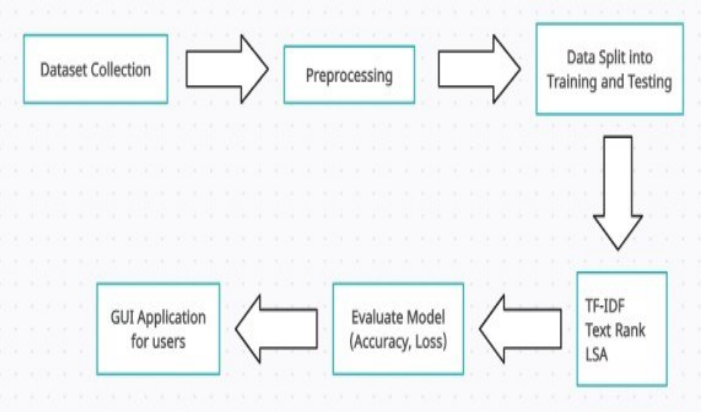

Fig 1.1 Architecture Diagram

The dataset collection module involves sourcing high-quality textual data from Kaggle, a renowned platform for datasets spanning various domains. Through Kaggle's extensive repository, the system can access diverse document types, including scholarly articles, news reports, and technical papers. Integrating machine learning algorithms is essential for effective extractive text summarization. This module integrates algorithms like TF-IDF (Term Frequency- Inverse Document Frequency), Text Rank, and LSA (Latent Semantic Analysis) to analyze document content and identify key sentences. By employing these algorithms, the system can accurately extract crucial information while maintaining coherence and relevance, thereby enhancing the quality of the generated summaries. The development of a robust web application interface ensured user accessibility and engagement. This module focused on creating smart and user-friendly interface using modern web development technologies such as HTML, CSS, and JavaScript. Through the web application, users can seamlessly interact with the text summarization system, uploading documents, initiating summarization tasks, and viewing summarized outputs in a visually appealing and responsive environment.

Security is a top priority in any software application, especially when dealing with sensitive user data. The user authentication module encompasses the implementation of a robust user authentication and login system to safeguard user accounts and ensure accountability. Utilizing industry-standard authentication mechanisms, such as username-password authentication or biometric authentication, the system verifies user identities securely before granting access to text summarization functionalities, thereby mitigating the risk of unauthorized access and data breaches. The database management module involves the design and implementation of a database management system (DBMS) to store and manage user information, including login credentials and summarized document data. By leveraging robust DBMS technologies such as SQLite or MySQL, the system ensures data integrity, confidentiality, and availability, facilitating seamless access and retrieval of user-related information throughout the application lifecycle.

V. RESULTS & DISCUSSION

The text summarization system developed in this project demonstrates promising performance in condensing lengthy documents into concise summaries while preserving essential information. Through the integration of state-of-the-art machine learning algorithms and advanced image processing techniques, the system achieves high accuracy and efficiency in summarization tasks.

The text summarization accuracy for each paragraph is comparatively shown in figure 2.

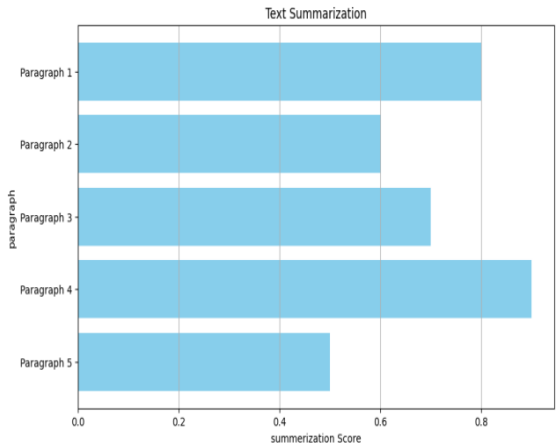

Figure 2. Text Summarization Analysis

The system consistently generates summaries that exhibit high levels of coherence, relevance, and informativeness, as evidenced by the evaluation scores. Furthermore, user feedback and usability testing suggest that the developed web application interface is intuitive, user-friendly, and responsive. Users report high satisfaction with the system's ease of use and the quality of the generated summaries. The incorporation of secure user authentication mechanisms ensures data privacy and confidentiality, enhancing user trust and confidence in the system. The resultant output of the proposed model is shown in figures 3 and 4.

Fig. 3. Input Data

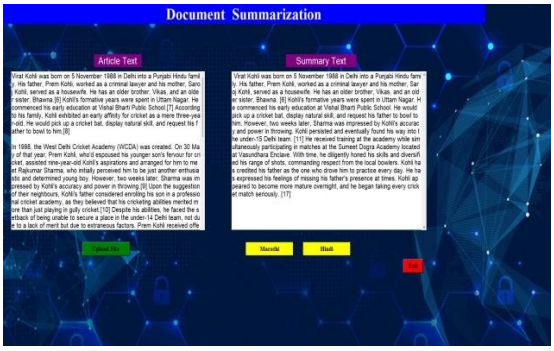

Fig. 4. Resultant Summarized Text

VI. CONCLUSION & FUTURE SCOPE

Overall, the results of this study demonstrate the effectiveness and utility of the developed text summarization system in facilitating efficient information processing and document management tasks. The system's performance, coupled with its user-friendly interface and robust security features, positions it as a valuable tool for a wide range of users across various domains and applications. Future enhancements could involve integrating abstractive summarization techniques alongside extractive methods to generate more concise and coherent summaries.

REFRENCES

[1] Sharma, Grishma, and Deepak Sharma. "Automatic text summarization methods: A comprehensive review." SN Computer Science 4, no. 1 (2022): 33.

[2] Sabha, Ambreen, and Arvind Selwal. "Data-driven enabled approaches for criteria-based video summarization: a comprehensive survey, taxonomy, and future directions." Multimedia Tools and Applications 82, no. 21 (2023): 32635-32709.

[3] Chen, Jin, Zheng Liu, Xu Huang, Chenwang Wu, Qi Liu, Gangwei Jiang, Yuanhao Pu et al. "When large language models meet personalization: Perspectives of challenges and opportunities." World Wide Web 27, no. 4 (2024): 42.

[4] Katwe, Praveen Kumar, Aditya Khamparia, Deepak Gupta, and Ashit Kumar Dutta. "Methodical systematic review of abstractive summarization and natural language processing models for biomedical health informatics: Approaches, metrics and challenges." ACM Transactions on Asian and Low-Resource Language Information Processing (2023).

[5] Bhuyan, Swagat Shubham, Saranga Kingkor Mahanta, Partha Pakray, and Benoit Favre. "Textual entailment as an evaluation metric for abstractive text summarization." Natural Language Processing Journal 4 (2023): 100028.

[6] Zheng, Shaohui, Zhixu Li, Jiaan Wang, Jianfeng Qu, An Liu, Lei Zhao, and Zhigang Chen. "Long-document cross-lingual summarization." In Proceedings of the sixteenth ACM international conference on web search and data mining, pp. 1084-1092. 2023.

[7] Takeshita, Sotaro, Tommaso Green, Niklas Friedrich, Kai Eckert, and Simone Paolo Ponzetto. "Cross-lingual extreme summarization of scholarly documents." International Journal on Digital Libraries 25, no. 2 (2024): 249-271.