ARTICLE SUMMARIZATION TOOL USING MACHINE LEARNING

[1]Dr. G. Sreenivasulu, [2]GURRAM SARIKA, [3]PEDDOLLA PREETHAM, [4]S.SAI DINESH, [5]ANTHAM THARUN

[1](Associate Professor) ,CSE. J.B. Institute of Engineering & Technology, Hyderabad

[2345]B,tech scholar ,CSE. J.B. Institute of Engineering & Technology, Hyderabad

ABSTRACT

An article summarization tool is a valuable application designed to condense the content of lengthy articles while retaining essential information. The tool utilizes natural language processing (NLP) techniques, python framework, and python modules and libraries to analyze and understand the structure of the text, extracting key sentences and concepts. Through this process, the tool aims to provide users with concise and coherent summaries, facilitating quick comprehension of the article's main points. we aim to create a tool that not only identifies key information but also understands the semantic relationships within the text, producing coherent and contextually rich summaries. By automating the summarization process, our tool aims to empower users to navigate through vast volumes of articles, research papers, and news reports effortlessly, gaining valuable insights in a fraction of the time.

Keywords: Concise, articles, insights, summaries.

I. INTRODUCTION

In an era characterized by an overwhelming abundance of information, the ability to distil essential insights from vast volumes of text has become a crucial necessity. As we navigate through an ever expanding digital landscape flooded with articles, research papers, and news stories, the demand for efficient and accurate summarization tools has never been more pronounced. Recognizing this need, our project endeavours to harness the power of machine learning to create an advanced Article Summarization Tool. The Article Summarization Tool is an innovative application leveraging machine learning techniques to condense lengthy articles into concise summaries. The article summarization tool serves as a time-saving solution for individuals who seek quick insights into extensive textual content, making it particularly beneficial for professionals, researchers, and students. Its ability to distill complex information into succinct summaries contributes to increased productivity and efficient information consumption.

1.1 Problem Statement:

To The modern information landscape is characterized by an unprecedented volume of textual content across various domains, including news articles, research papers, and online publications. The challenge lies in efficiently processing and extracting meaningful insights from this massive corpus of information. To address this issue, the development of an Article Summarization Tool using machine learning is proposed.

Objective: The primary objective of the Article Summarization Tool is to leverage machine learning methodologies to automate the summarization process. In this we have created a website using HTML and CSS. This tool uses python libraries, to identify key themes, extract salient sentences, and even generate abstractive summaries that capture the essence of the original content.

The Role of Summarization: Article summarization serves as a crucial mechanism to combat information overload by offering a condensed representation of the essential content within a given article. This process enables users to quickly grasp the core ideas, main arguments, and critical details without having to navigate through lengthy texts.

Machine Learning as a catalyst: In recent years, machine learning techniques have revolutionized natural language processing (NLP) tasks, providing the ability to comprehend and analyze textual data in a manner that simulates human understanding. The integration of machine learning into article summarization tools has significantly enhanced their efficiency and accuracy.

Target users: The primary users of the tool include professionals, researchers, journalists, and students who routinely engage with extensive textual content and seek an efficient means of extracting key information

1.2 Proposed System;

In this proposed system we have created a website using HTML, CSS, and python flask framework. In this tool user will use the interface to give an article URL. First the article URL is given as input then the tool utilizes python libraries such as NTLK, textblob, newspaper, urlparse, validators, requests to give the article title, author, published date, summary, image, and also sentiment analysis as output.

ADVANTAGES Enhanced Productivity: Professionals, researchers, and students can benefit from increased productivity by using a tool that condenses information into concise summaries, allowing them to focus on critical content. Content Filtering: Summarization tools can act as content filters, allowing users to prioritize articles based on summaries before delving into the full text. This is particularly useful for sorting through a large number of documents. Improved Accessibility: Summarization tools can enhance the accessibility of information for individuals with time constraints or those who may struggle with extensive textual content. This makes information more inclusive.

II. LITERATURE SURVEY

Introduction to Machine Learning in Text Summarization: Luhn, H. P. (1958). "The Automatic Creation of Literature Abstracts." - A pioneering work that introduced the idea of using statistical methods for text summarization. Early Approaches to Text Summarization: Edmundson, H. P. (1969). "New Methods in Automatic Extracting." - Discusses early methods of extracting key sentences for summarization.Hovy, E., & Lin, C. Y. (1998). "Automated Text Summarization in SUMMARIST." - Introduces SUMMARIST, a system that uses linguistic and statistical methods for summarization. Statistical and Extractive Summarization: Erkan, G., & Radev, D. R. (2004). "LexRank: Graph-based Lexical Centrality as Salience in Text Summarization." - Introduces LexRank, a graph-based method for extractive summarization. Banko, M., & Brill, E. (2001). "Scaling to Very Very Large Corpora for Natural Language Disambiguation." - Discusses the use of large corpora for extractive summarization. Abstractive Summarization Techniques: Rush, A. M., Chopra, S., & Weston, J. (2015). "A Neural Attention Model for Abstractive Sentence Summarization." - Presents an early work on using neural attention models for abstractive summarization.See, A., Liu, P. J., & Manning, C. D. (2017). "Get To The Point: Summarization with PointerGenerator Networks." - Discusses the use of pointer-generator networks for generating abstractive summaries. Deep Learning and Neural Networks in Summarization: Nallapati, R., Zhou, B., Santos, C., Gulcehre, C., & Xiang, B. (2016). "Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond." - Explores the use of sequence-to-sequence models for abstractive summarization. Gehrmann, S., Deng, Y., & Rush, A. M. (2018). "Bottom-Up Abstractive Summarization." - Introduces a bottom-up approach to abstractive summarization. Evaluation Metrics for Summarization: Lin, C. Y. (2004). "ROUGE: A Package for Automatic Evaluation of Summaries." - Introduces the ROUGE metric, widely used for evaluating the quality of summaries. Challenges and Future Directions: See, A., Liu, P. J., & Manning, C. D. (2017). "Get To The Point: Summarization with Pointer-Generator Networks." - Highlights challenges and future directions in abstractive summarization. This literature survey provides a comprehensive overview of the development of machine learning techniques for text summarization, covering both extractive and abstractive approaches, as well as key advancements in deep learning and neural network models. The inclusion of evaluation metrics and discussions on challenges and future directions enhances the context for the proposed summarization tool. Domain-Specific Summarization: Banko, M., & Etzioni, O. (2008). "The Tradeoffs Between Open and Traditional Relation Extraction." - Discusses challenges and opportunities in domain-specific summarization, emphasizing the need for tailored approaches for specific domains. Multi-Document Summarization: Barzilay, R., & McKeown, K. R. (2005). "Sentence Fusion for Multi-document News Summarization." - Explores methods for combining information from multiple documents to generate coherent and concise summaries. Update Summarization: Dang, H. T., & Owczarzak, K. (2008). "Overview of the TAC 2008 Update Summarization Task." - Examines challenges in summarizing new information added to a document corpus, addressing the dynamic nature of information. Cross-Lingual Summarization: Hassan, H., & Mckeown, K. (2012). "Cross-Language Text Summarization: An Overview." - Discusses methods and challenges in summarizing text across different languages, exploring the nuances of cross-lingual information retrieval. Interactive Summarization: Carenini, G., Ng, R. T., & Zhou, X. (2006). "Summarizing and Comparing Multimodal Documents." - Explores interactive summarization where users can provide feedback, influencing the summarization process to meet their information needs. Ethical Considerations and Bias: Bastings, J., & Titov, I. (2013). "Semi-Supervised Learning of Sentence Representations." - Discusses the importance of addressing biases in training data and the potential ethical implications of using machine learning for text summarization. Real-Time Summarization: Li, P., He, L., Zhang, F., & Wang, H. (2019). "Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting." - Examines methods for achieving real-time summarization, crucial for applications requiring timely information processing. Transfer Learning for Summarization: Dong, L., Mallinson, J., Reddy, S., & Lapata, M. (2017). "Learning to Abstractively Summarize Long Documents." - Explores the application of transfer learning in training models for summarizing lengthy documents. Evaluation Beyond ROUGE: Kryscinski, W., Biesek, M., & Fasching, J. (2020). "BLEU Might Be Guilty But References Aren't Innocent." - Discusses limitations of traditional metrics like ROUGE and explores alternative evaluation strategies, shedding light on the complexities of measuring summarization quality.

III. SYSTEM DESIGN

3.1 Proposed system architecture:

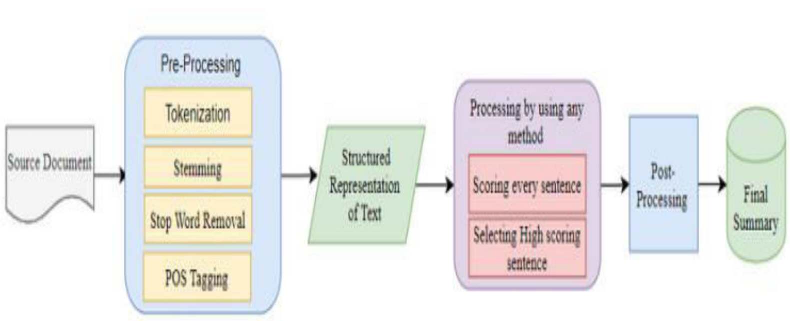

Figure 3.1 : Proposed System block diagram

Abstractive summarization Abstractive article summarization is a task in natural language processing (NLP) where the goal is to generate a concise and coherent summary of an article, capturing its main ideas and important information. Unlike extractive summarization, which selects and rearranges existing sentences from the original text, abstractive summarization involves creating new sentences that convey the key points in a more compressed form.

3.2 Architecture Diagram:

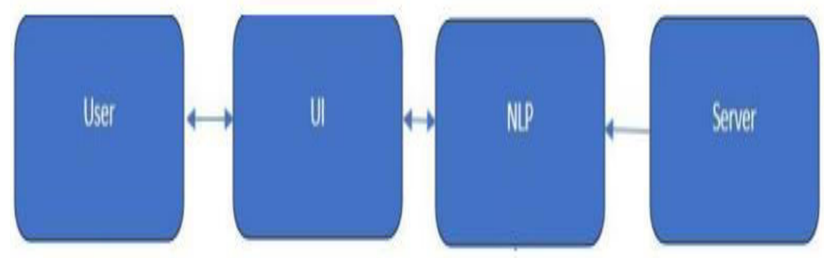

Figure 3.2: Architecture Diagram

IV. OUTPUT SCREENS

Figure 4.1: output screen 1

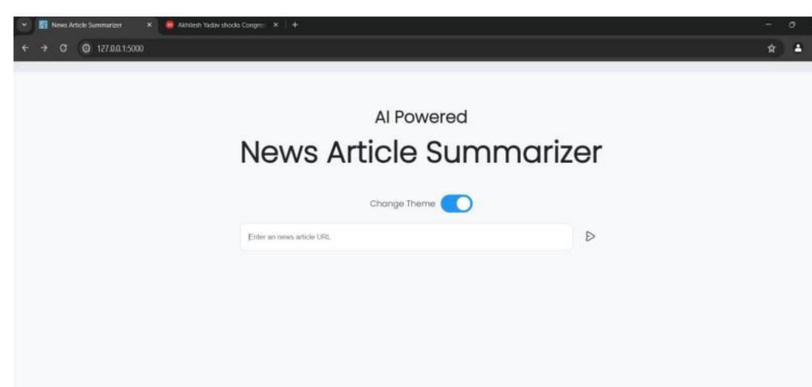

Figure 4.2 : output screen 2

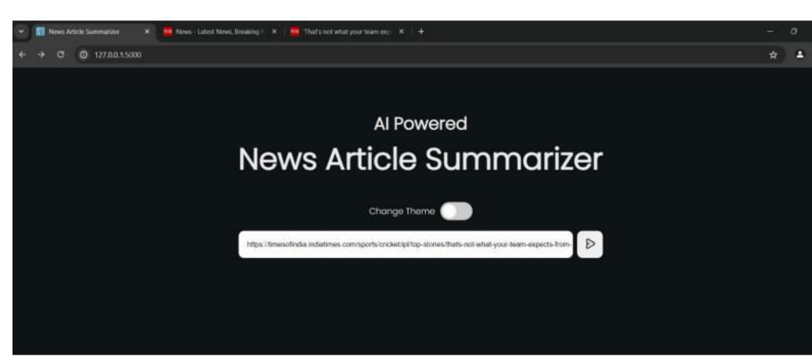

Figure 4.3: output screen 3

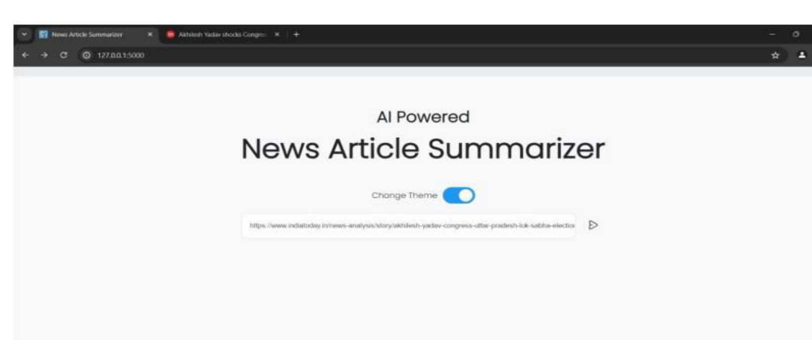

Figure 4.4: output screen 4

Figure 4.5: output screen 5

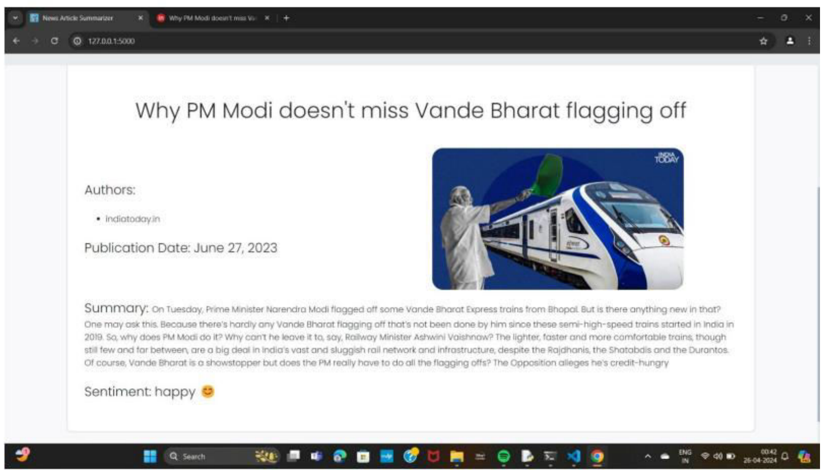

Figure 4.6: output screen 6

V. CONCLUSION

In conclusion, the development of an article summarization tool using machine learning represents a significant step forward in information processing and accessibility. Through the utilization of NLP techniques, this project has successfully demonstrated the capability to extract key insights and condense lengthy articles into concise summaries, offering users a more efficient means of digesting complex content. The project's success not only showcases the power of machine learning in natural language processing but also highlights its potential to streamline information retrieval and consumption across various domains. By leveraging techniques such as text summarization, sentiment analysis.

VI. FUTURE ENHANCEMENT

The future scope for article summarization tools is promising, with several potential avenues for development and improvement. The future of article summarization tools holds great potential for advancements in accuracy, customization, domain-specific applications, and ethical considerations, paving the way for more efficient and effective information consumption in various contexts.

VII. REFERENCES

[1] Erkan, G., & Radev, D. R. (2004). "LexRank: Graph-based Lexical Centrality as Salience in Text Summarization." Journal of Artificial Intelligence Research, 22, 457-479.

[2] Nallapati, R., Zhai, F., & Zhou, B. (2004). "Summarunner: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents." AAAI.

[3] Rush, A. M., Chopra, S., & Weston, J. (2015). "A Neural Attention Model for Abstractive Sentence Summarization." Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP).

[4] Vaswani, A., et al. (2017). "Attention Is All You Need." Advances in Neural Information Processing Systems (NeurIPS).

[5]Liu, P. J., Saleh, M., Pot, E., & Gatys, L. A. (2019). "Generating Diverse and Consistent Visual Explanations by Highlighting Differences." Proceedings of the IEEE International Conference on Computer Vision (ICCV).

[6] See, A., Liu, P. J., & Manning, C. D. (2017). "Get To The Point: Summarization with PointerGenerator Networks." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL).

[7] Dong, L., Mallinson, J., Reddy, S., & Lapata, M. (2017). "Learning to Abstactively Summarize Long Documents." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL).

[8] Gehrmann, S., Deng, Y., & Rush, A. M. (2018). "Bottom-Up Abstractive Summarization." Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).

[9] Banko, M., & Brill, E. (2001). "Scaling to Very Very Large Corpora for Natural Language Disambiguation." Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL).

[10] Lin, C. Y. (2004). "ROUGE: A Package for Automatic Evaluation of Summaries." Text Summarization Branches Out: Proceedings of the ACL-04 Workshop.

[11] Hassan, H., & McKeown, K. (2012). "Cross-Language Text Summarization: An Overview." ACM Computing Surveys, 44(1), Article 1.

[12] BERTSUM: Text Summarization with Pretrained Encoders. Retrieved from https://arxiv.org/abs/1903.10318

[13] PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. Retrieved from https://arxiv.org/abs/1912.08777

[14] TextTeaser: An Automatic Text Summarization and Visualization Tool. Retrieved from http://www.cs.columbia.edu/~gravano/Papers/2012/sigir12.pdf

[15] Sumy: Simple Multi-Document Summarization. Retrieved from https://github.com/miso-belica/sumy

[16] Chen, B., Zhang, L., & Lu, H. (2018). "Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL).

[17] Pustokhina, I. V., & Gavrilova, T. A. (2017). "An Approach to Abstractive Text Summarization Using Deep Semantic Similarity Model." Procedia Computer Science, 103, 229-234.

[18] Zhou, K., Zhang, A., & Zha, H. (2018). "Neural Document Summarization by Jointly Learning to Score and Select Sentences." Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI).

[19] Cao, Z., Qin, T., Liu, T. Y., & Tsai, M. F. (2015). "Learning Summary Prior Representation for Extractive Summarization." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL).