



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Automatic Text Summarization and Text Generation for News Articles

Sahana M J, Vaishnavi H S, Vinutha G, Virshitha G, Ms. Madhu D

Department of Computer Science and Engineering, PESITM, Shivamogga, Karnataka, India

Department of Computer Science and Engineering, PESITM, Shivamogga, Karnataka, India

Department of Computer Science and Engineering, PESITM, Shivamogga, Karnataka, India

Department of Computer Science and Engineering, PESITM, Shivamogga, Karnataka, India

Assistant Professor, Department of Computer Science and Engineering, PESITM, Shivamogga, Karnataka, India

ABSTRACT: In the era of information overload, newspapers face the challenge of efficiently presenting news content to their readers. This project addresses this issue by developing a system for automatic text summarization and text generation specifically for newspapers. The proposed system leverages advanced natural language processing (NLP) techniques to analyze and condense lengthy articles into concise summaries while preserving the key information. The first component of the system focuses on automatic text summarization, employing state-of-the-art extractive and abstractive summarization methods. Extractive summarization identifies and selects essential sentences from the original text, while abstractive summarization generates coherent summaries by paraphrasing and restructuring the content. The second component involves text generation, where the system is designed to create engaging and relevant content for newspaper articles. Utilizing advanced language models, the system generates coherent and contextually appropriate sentences to expand on summarized information or produce additional content for publication. This ensures that the generated text aligns seamlessly with the newspaper's editorial style and maintains a high level of readability. The impact of this project extends beyond traditional newsrooms, providing a valuable tool for enhancing efficiency in content production, especially in the fast-paced and dynamic landscape of news reporting. By automating text summarization and generation, newspapers can streamline their operations, increase content output, and deliver more digestible and engaging news to their audience. Keywords—Machine Learning, Natural language processing, Summarization, Extractive, Generation, Automatic text.

I. INTRODUCTION

In today's world we need to wish to have updated news so we need time to read things. For readers seeking timely and comprehensive updates are important. To address this, our project, titled "Automated Text Summarization and Generation for News Articles," with the help of machine learning (ML) techniques and algorithm, module like newspaper3k we can use to get related news from internet, reader are more interested to get to know what's ongoing, so article plays an important role in it. The summarization and generation of news content will be time reducing. Through advanced Natural Language Processing (NLP) and ML models, we aim is to lengthy articles into concise and informative summaries, providing readers with a quick yet comprehensive overview of the about 2 Automatic text summarization and text generation for news articles article. Additionally, our project not only to generate the article summary and also make a way to article and we have option like favourite and where they can select where personal favourite news by that not it helps readers to deeper into the content .By incorporating state-of-theart NLP libraries and ML models such as BERT, GPT, or T5(A text generation and prediction system Pretraining on new corpora using BERT and GPT-2 in this paper they have used this library and models), our system will exhibit an enhanced understanding of contextual data, ensuring the accuracy and coherence of the generated summaries. The user interface of the system will empower readers to effortlessly access both the full articles and their condensed versions, and also to choose their favourite category in it .Scalability and performance optimization will be prioritized to accommodate the vast volume of newspaper articles, providing a responsive and efficient experience for users. A unique aspect of our project lies in its commitment to user agency. By incorporating a feature that enables users to mark articles as favourites, our platform transforms into a personalized news. At our text summarization mechanism lies the Newspaper3k module, a robust tool that not only simplifies the summarization process but also enhances the accuracy and relevance of the generated content. And we have used streamlit an open-source Python library that facilitates the creation of web applications for data science and machine learning projects with minimal effort.

II. RELATED WORK

[1] Automatic text summarization of news articles in this paper they have proposed about to give the increase in size and number of documents available online, an efficient automatic news summarizer is the need of the hour. In this paper, they propose a technique of text summarization which focuses on the problem of identifying the most important portions of the text and producing coherent summaries. In their methodology, they do not require full semantic interpretation of the text, instead we create a summary using a model of topic progression in the text derived from lexical chains. [2] Automatic text generation and summarization for academic writing in this paper, they discuss the implications of automatic text generation for academic writing. They first review the current state of the technology and how it is being used. And then discuss the implications of using automatic text generators for academic writing, including the need for users to be aware of the limitations of the technology and how to use it effectively. They also discuss how the use of automatic text generation can change the traditional stages of writing, and how the content generated by these systems is not justified by semantic or extralinguistic criteria. [3] A text generation and prediction system: Pre-training on new corpora using BERT and GPT-2 where in this starting word to make a sentence or filling in sentences is an important direction of natural language processing. From one aspect, it reflects whether the machine can have human thinking and creativity. They train the machine for specific tasks and then use it in natural language processing, which will help solve some sentence generation problems, especially for application scenarios such as summary generation, machine translation, and automatic question answering. [4] Extractive automatic text summarization using spacy in python & NLP the amount of data produced at every instance is very huge; there is an ultimate need to develop a machine that can reduce the length of the texts automatically. Moreover, applying text summarization gears up the procedure of researching, reduces reading time, and increases the amount of important information being generated in the specific field. [5] Information retrieval and processing system for news articles in English in this focus on designing a tool to extract details from English news articles and present it to the user, in an organized manner. A predefined set of websites are crawled and the details are stored. The details extracted by the tool are named entities such as location, person, and organization mentioned in the news, news summary and important keywords pertaining to each news article. [6] Data to text generation with iterative text editing here They approach maximizes the completeness and semantic accuracy of the output text while leveraging the abilities of recent pre-trained models for text editing (Laser Tagger) and language modeling (GPT-2) to improve the text fluency. To this end, they first transform data items to text using trivial templates, and then we iteratively improve the resulting text by a neural model trained for the sentence fusion task.

The output of the model is filtered by a simple heuristic and reranked with an off-the-shelf pre-trained language model. They evaluate their approach on two major datato-text datasets (Web NLG, Cleaned E2E) and analyze its caveats and benefits. Furthermore, they show that their formulation of data-to-text generation opens up the possibility for zero-shot domain adaptation using a general- domain dataset for sentence fusion.

III. PROPOSED SYSTEM

This project proposes a novel machine learning-based system that leverages the accurate and proper information about article. Our innovative system design will have to redefine the user experience by seamlessly integrating two pivotal features: text summarization and text generation. At its core, our project focuses on distilling the essence of lengthy articles into concise and informative summaries, enabling users to grasp the key insights expeditiously. Moreover, we empower users to curate their news feeds by implementing a feature that allows them to designate favourite articles, tailoring their content consumption to align with individual preferences.

In tandem with text summarization, our system's text generation capability enriches the user experience by offering insightful and contextually relevant content. Together, these features harmonize to create a dynamic and personalized news platform, fostering a symbiotic relationship between cutting edge technology and the evolving landscape of news. Our innovation extends to empowering users to curate their own news haven. With a simple selection of favourite articles, individuals can tailor their news feeds to align perfectly with their unique preferences. Powered by Newspaper3k, our text generation feature adds an extra layer of insight, providing users with contextually rich content that goes beyond the ordinary. This project is not just about technology; it's about revolutionizing the way we interact with news, seamlessly blending cutting- edge capabilities with the ever-evolving landscape of journalism for a more personalized and enriching experience. At our text summarization mechanism lies the Newspaper3k module, a robust tool that not only simplifies the summarization process but also enhances the accuracy and relevance of the generated content. And we have used streamlite an open-source Python library that facilitates the creation of web mental health assessments applications for data science and machine learning projects with minimal effort.

Develop an intuitive and user-friendly interface that encourages engagement. Design a simple, clear, and empathetic UX that takes into account the potential sensitivity of the subject matter. Provide users with clear instructions and guidance throughout the assessment process to ensure a seamless user experience. Utilize data analysis techniques to derive meaningful insights and identify patterns in the collected data. Consider integrating machine learning algorithms to improve accuracy and personalize assessments. Offer users personalized suggestion to improve and promote understanding. The system should be easy to use and also it should not have problem with news. News should be accurate and should not have false news. News should be latest and also real and project will project a best version where we can find all news in a platform.

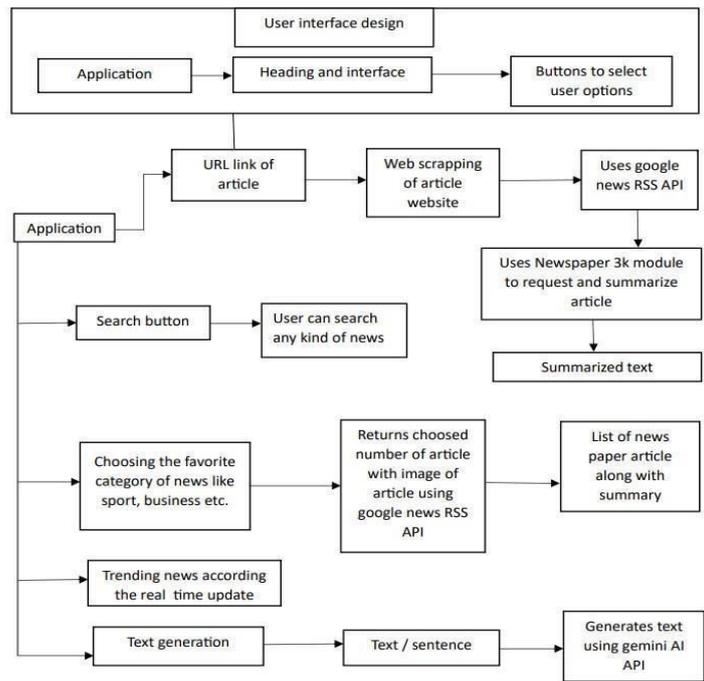


Fig 1: System Architecture.

IV. METHODOLOGY

State the main ideas. Identify the most important details that support the main ideas. Summarize in your own words. Do not copy phrases or sentences unless they are being used as direct quotations. Express the underlying meaning of the article, but do not critique or analyse. The process of selecting and presenting news content is defined as identifying news value. This involves evaluating factors like extraordinariness, threshold, negativity. this system as your virtual news concierge, intuitively presenting you with concise summaries that capture the heart of a story, alongside headlines crafted to captivate your curiosity. The technology driving this project is not just sophisticated; it's designed to be user-friendly, with a web interface that beckons exploration and invites readers to effortlessly toggle between full articles and their condensed versions. News articles such as sports game reports are often thought to closely follow the underlying game statistics, but in practice they contain a notable amount of background knowledge, interpretation, insight into the game, and quotes that are not present in the official statistics. This poses a challenge for automated data-to-text news generation with real world news.

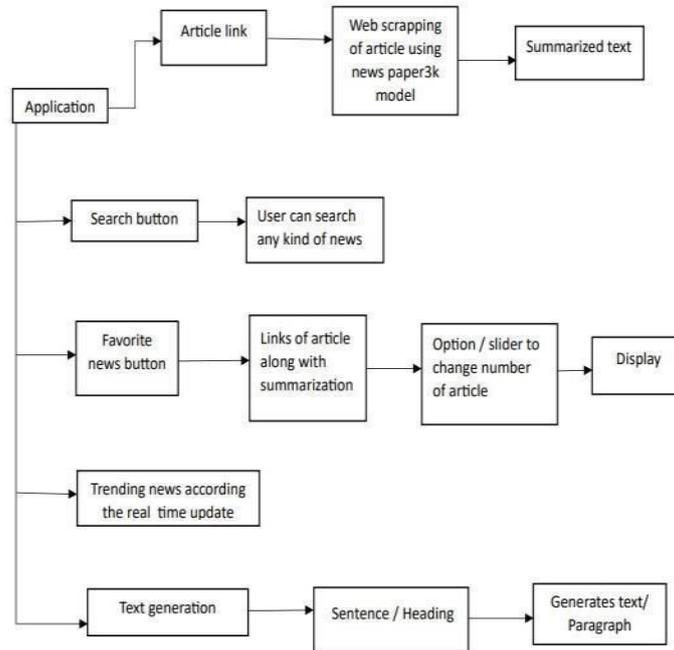


Fig 2 : Workflow of methods.

V. RESULTS

Text summarization and generation for a news articles application involves condensing articles into concise summaries and creating engaging headlines for different categories like favourite news and trendy news. The output could feature personalized favourite news summaries tailored to user preferences, providing a quick overview of their preferred topics. And also the “Favourite news” section could provide comprehensive, personalized summaries of articles related to topics the user has marked as favourites. The “Trendy news” section would present the latest and most popular stories in a concise format and also search option where user can search any topics. And for text generation gemini pro is used.



Fig:3 User interface of project.

This figure represents the starting page, where user can see option in our project. A summarised news and text generation.



Fig:4 Choose your category.

This figure represent the option of choosing there thing, like user can select their category like here favourite news and trending news.

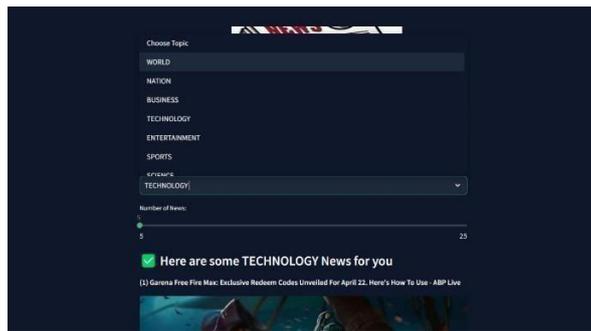


Fig:5 Option available inside favourite category.

This figure represent option like world, nation business, technology, entertainment, sports and also science etc. inside favourite category.



Fig:6 Trending news category.

Here in this figure it will show the trending news and also user can select how many news user need.



|| Volume 13, Issue 5, May 2024 ||
| DOI:10.15680/IJIRSET.2024.1305191 |

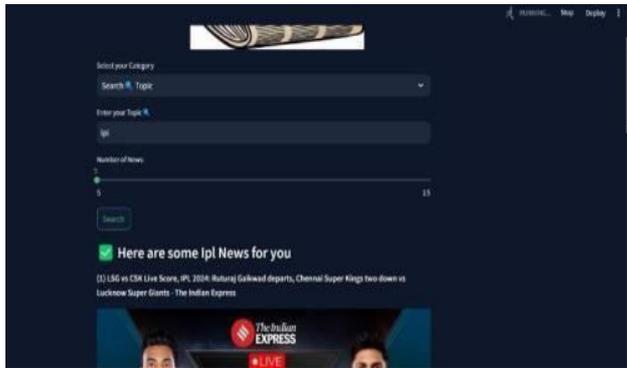


Fig: 7 Search your topic option.

In this figure user can search any news we want, and also how many news we need. Search option we collect related thing that user have searched for.

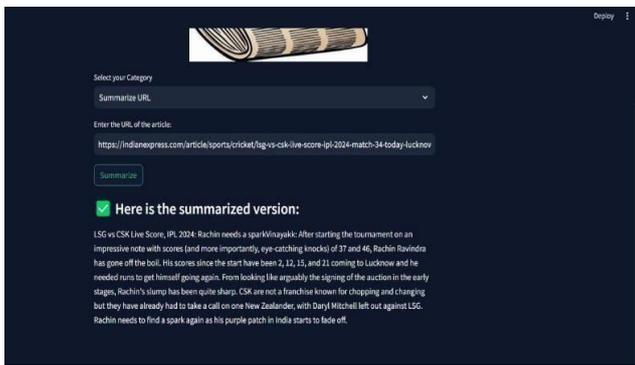


Fig: 8 Summarized news article.

Here in this figure user can paste an url link which from an article which user want to summarized and that need to be paste and after clicking on summarize button they will get summarized version of that article.

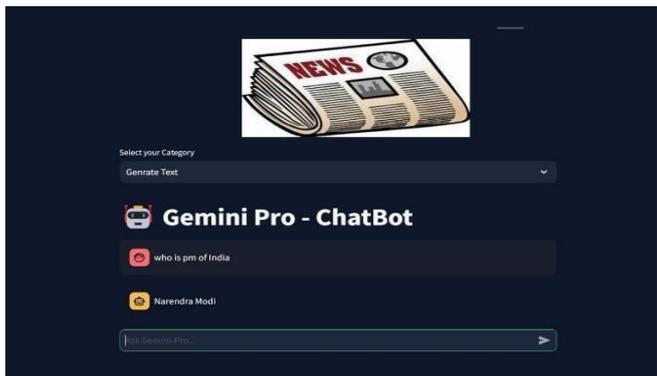


Fig: 9 Gemini Pro chatbot.

In this figure user can generate text using gemini pro. Where text will be generate for user input.

VI. CONCLUSION

The integration of text summarization and text generation technologies holds immense promise for the newspaper industry, in a new era of efficient content creation and consumption. The application of these technologies streamlines editorial processes, enabling newsrooms to produce proper and engaging summaries of complex articles. This not only enhances the overall productivity of journalists but also caters to the evolving preferences of readers who seek information in a time-efficient manner. The advent of abstractive summarization techniques, coupled with advanced natural language processing models, ensures that the generated summaries maintain context and coherence. Realtime updates and breaking news capabilities is essential to keep pace with the dynamic nature of the news industry.

VII. FUTURE WORK

The work could be expanded to include a user profile system, allowing individuals to create and maintain their profiles. This feature would enable users to establish personalized accounts within the platform, where they can securely store their score histories and performance records. Using this functionality, users can easily access and review their past assessments, track their progress over time, and identify areas for improvement.

REFERENCES

- [1]. Prakhar sethi and Samer sonawane, 2017, Automatic text summarization of news articles.
- [2]. Alice Delorme Benites and chris M Anson,2023, Automatic text generation and summarization for academic writing.
- [3]. Yuanbin QU, Peihan Liu, Wei song and Lizhen Liu,2020, A text generation and prediction system:Pre-training on new corpora using BERT and GPT-2.
- [4]. Swaranjali Jugran, Ashish Kumar, Bhupendra Sindh Tyagi and Vivek Anand,2021, Extractive automatic text summarization using spacy in python & NLP.
- [5]. Rohith Sathyajit and Sandhya Harikumar, 2019, Information retrieval and processing system for news articles in English.
- [6]. Zdenek Kasner,2020, Data-to-text generation with iterative text editing.
- [7]. Santoshi Kumari and T P Pushphavati,2022, Qustion answering and text generation using BERT and GPT-2 model.
- [8]. Noureen Fatima, Ali Shariq Imran and Zenun Kastrati, 2019, A systematic literature review on text generation using deep neural network models.
- [9]. Prabhudas Janjanam and CH Pradeep, 2019, Text summeraization: An essential study.
- [10]. Wenhao Yu, Chenguangn Zhu and Zaitang Li, 2022, A survey of knowledge-enhanced text generation.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details