



TEXT SUMMARIZATION TOOL: APPLICATIONS AND RESULTS

¹Aniketh Jadhav, ²Ninad Ghandat, ³Shubham Pawar, ⁴Mayur Chaudhari

^{1, 2, 3, 4} Student,

Department Of Information Technology,
MET Bhujbal Knowledge City Institute of Engineering,
Nashik, Maharashtra, India.

Abstract: Text summarization is a process of extracting the context of a large document and summarize it into a smaller paragraph or a few sentences. Text summarization plays a vital role in saving time in our day-to-day life. It is also used in many bigger project implementations of classification of documents or in search engines. This paper presents a method of achieving text summaries accurately using deep learning methods. Polarity analyzes each and every word from the article, and generates a number which denotes the sentiment of the article. Reading long texts and articles is not possible for everyone in this smart and fast day-to-day life. The tool will be made with the help of Natural Language Toolkit. This tool aims to deliver short yet meaningful summaries of text documents to the end users.

Index Terms - Deep Learning, Natural Language Processing, Abstractive summary, Extractive summary, Polarity, Natural Language Toolkit.

I. INTRODUCTION

Text summarization is a process of producing brief and concise summary by capturing the vital information and the comprehensive meaning. Text summarization is achieved by natural language processing techniques by using algorithms like page rank algorithms etc. While these algorithms fulfil the objective of text summarization, they cannot generate new sentences which are not in the document like humans. They can also have grammatical errors. This is where Deep Learning comes to our rescue. The use of deep learning builds an efficient and fast model for text summarization. The use of deep learning methods helps us generate summaries which can be formed with new phrases and sentences and also which are grammatically correct. This results-based paper aims to contribute to the existing research on text summarization techniques, and its applications.

Users can conveniently access the tool through a user-friendly web interface or integrate it into existing applications via an API, ensuring seamless accessibility and integration into their workflow. As a result, our text summarization tool contributes to enhancing productivity, knowledge acquisition, and decision-making processes. By automating the laborious task of manual summarization, it empowers users to rapidly extract critical insights from large volumes of text, thereby reducing information overload and enabling efficient knowledge utilization. The tool's reliability, flexibility, and high-quality summaries position it as an indispensable asset in the digital age, unlocking new possibilities for information extraction and interpretation.

II. IMPLEMENTATION

This project's motive is to create a free and easy to use summarization tool that can be used by anyone for any purpose. The accessibility of tool and the user interface should be as simple as possible so that even a person from non-technical background can figure out how to use the tool.

In this project, we have implemented the tool using Python programming language.

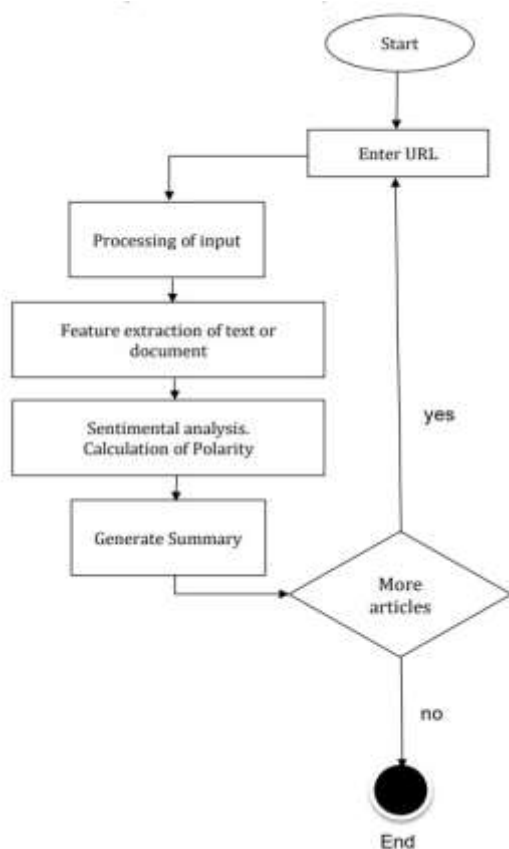


Figure: Activity diagram

1] Pre-processing:

Here, tasks like removal of stop words, punctuation marks, special characters, etc. is performed.

2] Feature Extraction:

The important keywords appearing in the text or article are given higher preference than regular appearing words (stop words.)

3] Sentiment Analysis:

An overview of the emotion expressed in the article (happiness, sadness, humor, etc.) is analyzed.

4] Polarity:

It represents the sentiment or polarity of the summarized text or article (positive, negative, or neutral.)

5] Generation of summary:

The final precise summary is generated with the help of Python's libraries.

Text summarization tool can be used in any field, like for example content creation, books and literature, financial research, chatbots, video conferencing along with others, writing better reports, finding more relevant information in less time, etc. Text summarization tool can be also used for Information Retrieval purposes.

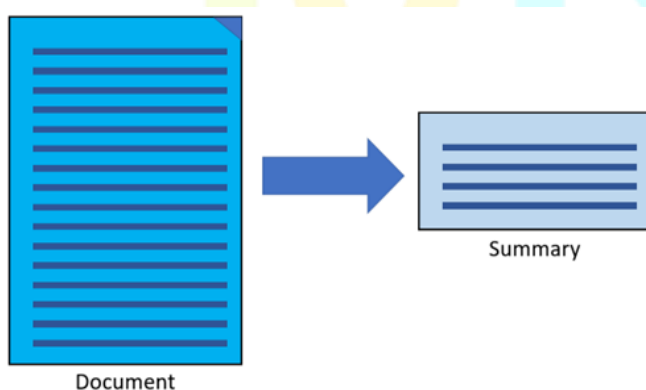


Figure: Extractive Summarization

III. REQUIREMENTS

Software Requirements:

- Programming Language: Python 3.11
(Any latest version)
 - Tools: PyCharm, Python IDLE, Visual Studio Code
 - Python libraries: Natural Language Toolkit (NLTK)
- Natural Language Processing (NLP)
Tkinter (for graphical user interface)

Hardware Requirements:

- Processor: Intel Core i5 or above.
- Ram: 4GB
- Hard Disk: 128GB or above

IV. RESULTS

The input will be the URL of that an article or text document. The tool will start with the processing of the contents in the article. NLP will parse the contents of the web page. The feature extraction will be done by the newspaper library. The newspaper library will extract the useful information from the web page and discard the irrelevant information. The TextBlob library will perform sentimental analysis of the web page contents, along with the calculation of polarity. With the help of these processes, a long text or article gets converted into a concise and precise summary.

The output obtained will be in the following format:

1. Title of article.
2. Author of the article.
3. Date of publication.
4. Summary.
5. Polarity.

The following screenshots explain the GUI of the tool.



Figure: User Interface



Figure: example URL with generated summary

V. APPLICATIONS

[1] News Collection: Text summarization can be used to automatically generate precise summaries of news articles, allowing users to quickly scan and gather information from multiple sources. [2] Document Summarization: Summarizing lengthy documents such as research papers, legal documents, or business reports can help readers extract key insights and save time. [3] Social Media Analysis: Summarizing social media posts or comments can provide a concise overview of discussions, current trends, or sentiment analysis on specific topics. [4] Chatbots and Virtual Assistants: Text summarization helps chatbots and virtual assistants understand and respond to user queries by extracting relevant information from large text sources (examples: Google's Alexa, Apple's Siri, etc.) [5] Email Filtering: Summarizing emails or categorizing them based on their content can help users in prioritizing and managing large volumes of inbox, unread messages. [6] Legal and Contract Analysis: In the legal domain, summarizing legal documents, contracts, or case briefs can be helpful for lawyers and legal professionals in reviewing and analyzing important information quickly. [7] E-Learning and Education: Summarization can be used in e-learning platforms to shrink educational content, textbooks, or research papers into concise summaries for students and researchers. [8] Data Extraction for Data Science: Text summarization techniques can assist in extracting essential information from unstructured textual data for further analysis as well as data modeling.

These applications demonstrate the adaptability and value of text summarization in enhancing productivity, information retrieval, and decision-making processes across many industries and domains.

VI. CONCLUSION

The increasing growth of the Internet has made a huge amount of information available. It is difficult for humans to summarize large amounts of text. Thus, there is an immense need for automatic summarization tools in this age of information overload. The International Data Corporation (IDC) projects that the total amount of digital data circulating around the world, annually would hit 180 zettabytes by 2025. That is an enormous amount of data circulating in the digital world. Classification, and other computer science fields like Information Retrieval. Access time for information searching can be improved. Applying text summarization reduces reading time and accelerates the process of researching for information playing a major role in current era of rapid development and digitalization. Humans are generally quite good at this task as we have the capacity to understand the meaning of a text document and extract salient features to summarize the documents using our own words. However, the automatic methods for text summarization of articles are crucial in today's world where there is an over-abundance of data and lack of manpower as well as time to interpret the data.

REFERENCES

- [1] Ravali Boorugu, G. Ramesh – “A Survey on NLP Based Text Summarization for Summarizing Product Reviews” – 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)
- [2] N. Andhale – “An overview of Text Summarization Techniques” – 2016 International Conference on Computing Communication Control and automation (ICCUBEA)
- [3] P. Janjanam, CH Pradeep Reddy – “Text Summarization: An Essential Study” – 2019 International Conference on Computational Intelligence in Data Science (ICCIDS)
- [4] J.N. Madhuri, R. Ganesh Kumar – “Extractive Text Summarization Using Sentence Ranking” 2019 International Conference on Data Science and Communication (IconDSC)
- [5] M. Patel, Satyadev Vyas, K. Maurya – “Machine Learning Approach for Automatic Text Summarization Using Neural Networks” – 2018 International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE) [Vol. 7, Issue 1]
- [6] Oguzhan Tas, Farzad Kiyani – “A Survey Automatic Text Summarization” – 2017 PressAcademia Procedia 2nd World Conference on Technology, Innovation and Entrepreneurship DOI: 10.17261/Pressacademia.2017.591 PAP-WCTIE-V.5-2017(29)-p.204- 213
- [7] Surabhi Adhikari, Rahul, Monika – “NLP based Machine Learning Approaches for Text Summarization” –2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)
- [8] Pradeepika Verma, Anshul Verma – “Accountability of NLP Tools in Text Summarization for Indian Languages” 2020 Journal of Scientific Research (Volume 64, Issue 1)
- [9] M. F. Mridha, Aklima Akter Lima, Kamruddin Nur, Sujoy Chandra Das, Mahmud Hasan – “A Survey of Automatic Text Summarization: Progress, Process and Challenges” 2021 IEEE Access (Volume: 9) (Page number: 156043 – 156070)