IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

# Abstraction Based Text Summarization

**Dr. Nagabhushan S V[1], Jeevitha M[2], Manoj Kumar[3], Archana S[4], Likitha Naveen[5]**
Associate Professor, Department of CSE[1]
Students, Department of CSE[2,3,4,5]
BMS Institute of Technology and Management, Bengaluru, Karnataka, India

**Abstract:** *The goal of the summary of text is highlighting important details from the textual original. During this procedure, the user is given a succinct overview of the retrieved data as a condensed report. The text's substance is extremely challenging for people to comprehend and decipher. An extensive examination of abstractive text summarizing techniques is provided in this study The approaches that are based on structure and semantics are the two main abstractive summarization techniques. This essay analyses and synthesizes the different approaches, problems, and concerns related to abstractive synopsis. The attributes of state-of-the-art benchmark datasets are being investigated. According to this survey, the majority of abstractive summarizing techniques yield extremely clear, cohesive, less redundant, and information-rich summaries.*

**Keywords:** Fusion of sentences, abstraction scheme, structure-based approach, semantic-based approach, text summarization, sentence revision, and abstractive summary.

## I. INTRODUCTION

The vast amount of information and papers available online has become overwhelming for people due to the Internet's fast expansion. A vast number of electronic documents are readily available online because of the quick growth of information communication technologies, and users are finding it difficult to locate relevant information. information summary has gained attention because to worries that users may overlook reading several vital and relevant publications because they become so tired from reading a big volume of information. The process of producing a smaller document that preserves the majority or all of the information found in the original sources from one or more textual sources is known as text synthesis. The intended objective of the summary determines whether details and other aspects of the original papers are retained. To compile all the data and provide it in a streamlined format is a laborious process. Thus, text summarizers were developed, which reduced the length of the document and gave a clear, succinct synopsis of the conversation. A synopsis is advantageous as it facilitates the recovery of lengthy text files and saving time as well and so on .This as a whole decodes the issues, challenges and howan abstractive summarization technique may be applied toconversational systems Interpreting human language from one structure to another is the main focus of natural language processing, or NLP. in NLP. One research study's goal is to provide pertinent summaries utilizing a variety of natural language processing technologies. This endeavor is called summarization and techniques. When we think of summarization, we just get to do with making something shorter, but people generally forget the main element is that summarization is to make text shorter, but my maintaining the meaning of the text or document with all the information representing in just few words. There are two primary types of text summarization

## II. MOTIVATION

Research on text summarization is ongoing in the fields of natural language processing and information retrieval. Text summarization is being utilized more and more in the business world in areas including word processing software, data mining of text databases, telephone conversation, and web-based information retrieval. The way that different methods formulate problems varies widely. An essential first step in information management activities is automatic text summarizing. It addresses the issue of deciding which passages in the book are the most crucial. Remarkable summarization necessitates advanced NLP methods.

The principal problems with abstractive summarizing are there is absence of a universal framework, parsing and alignment of parse trees is challenging. An open problem is how to extract the most significant sentences and arrange

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

them in the original source document's exact sequence so that an effective summary may be produced. It is challenging to compress via abstractive summarization when lexical substitutions, paraphrasing, and reformulation are involved. The richness of their representation limits the system's potential, and producing such a structure is the biggest obstacle to abstractive summarization.

## III. RELATED WORKS

[1]In this paper “Abstractive Summarization using Graph Based Methods" Chetana Badgujar. VimlaJethani. Tushar Ghorpad Text summarization is a necessary and research areathat minifies text such that repeated data are removed and important information is extracted and represented in the concise way which can help us to understand the information.
**Drawbacks:** This approach is restricted to abstractive summarization of a single text.

[2]In this paper "Extractive text Summarization by featured-have sentence” Siya Sadashiv Naik, Manisha Naik Gaonkar Automatic Text Summarizer serves as Among the tools for interpreting lengthy textual content. It represents the a condensed version of the original document by choosing most important part of text, thus generating its an overview. It falls into two groups. abstraction and extraction.
**Drawbacks:** he pattern and rules are designed, physically which is Hectic and takes lots of time.

[3]In this paper " Extractive vs Abstractive Summarization" Nidhi Patel. Prof Nikhita Mangaokar In the contemporary world, when robots are assuming all tasks, text summary is lagging far behind. To establish a document summary, a To establish a main ideas from text documents.
**Drawbacks:** In the present world, text summary is far behind, with robots taking over all responsibilities. Numerous methods have been created to extract the essential concepts from text texts To be able to construct a document summary.This method cannot determine which approach is more effective in real life

[4] The paper "English Text Summarization for English Teaching: An Extraction Algorithm" LILI WAN. To enhance the ability of sharing and scheduling capability of English teaching resource an improved algorithm for English The summary of text is proposed based on Association, semantic rules As many sentences and phrases cannot be considered for the summary generation difficult to express with new sentence and if the linguistic metric of the synopsis is less due to wrong parse.
**Drawbacks:** The purpose of the study is to evaluate the benefits, drawbacks, and key elements of both approaches employing word embedding vectors and the TF-IDF vectorization technique. This approach is unable to identify which approach is more effective in real life

## IV. PROPOSED METHODOLOGY

In this project of ABSTRACTION BASED TEXT SUMMERIZATION, we will be approaching data collection to gather a dataset of documents relevant to your target domain. Data preprocessing which clean and preprocess the text data, removing noise, irrelevant characters, and formatting issues.We will be using preprocessed data that refers to the modified and cleaned version of the first manuscript documents that are prepared and organized in a specific way to facilitate to give summery. The summary of the abstractive text approach should be capable of taking a longer piece of text and generating a shorter, human-readable summary that retains the core information and context of the original content. The predicting new sentences process allows the summary of the text that is abstractive model to generalize its learning from the training phase to create meaningful and contextually relevantsummaries for new sentences or documents during the prediction stage and generate the summary.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

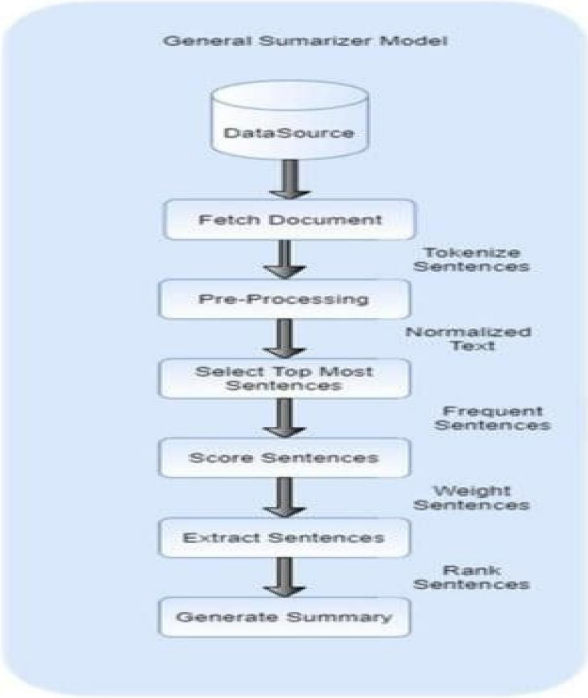


**FIG.4.1.** Abstraction based text summarization

STEPS INVOLVED TO GENERATE THE ABSTRACTIONBASED SUMMARY:

**STEP 1:** Here the text is used collected from the user asinput for summarizer.

**STEP2**: In this step collected text is cleaned, means deleting the stop words, special characters, numbers which is irrelevant to text and punctuations.

**STEP 3**: In this step word token and sentences token arecreated this process is called Tokenization

**STEP 4**: In this step by those tokens created in perviousstep, frequency is found for every word in the users input text.

**STEP 5**: Here in this step weights are assigned to words.

**STEP 6:** Considering the weights, most top rated 20% weighted sentences are called final summary. When we think of summarization, we just get to do withmaking something shorter, but people generally forgetthe main element is that. The purpose of summarizing is to make textshorter, but my maintaining the meaning of the text ordocument with all the information representing in justfew words.

There are typically two kinds of text summarization:

- Extractive Text Summarization
- Abstractive Text Summarization

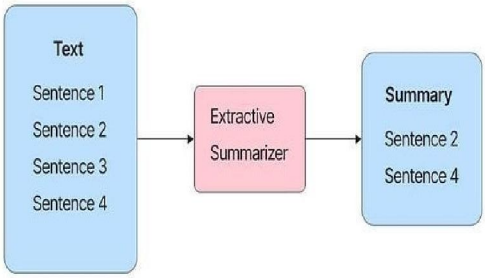


**FIG 1.2:** Extractive Text Summarization

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

### 4.1 SUMMARIZATION OF ABSTRACTIVE TEXT

A summary that is abstractive is more effective method of summarizing information than extractive synopsis, since it gathers data from several documents In order to offer an accurate summary. This has become more well-known because it can create new phrases that highlight key details from text documents. A cohesive portrayal of the summarized data is presented by an abstractive summarizer. coherent and grammatically sound writing. Enhancing the caliber of a synopsis can be facilitated by paying attention to linguistic or readability quality.

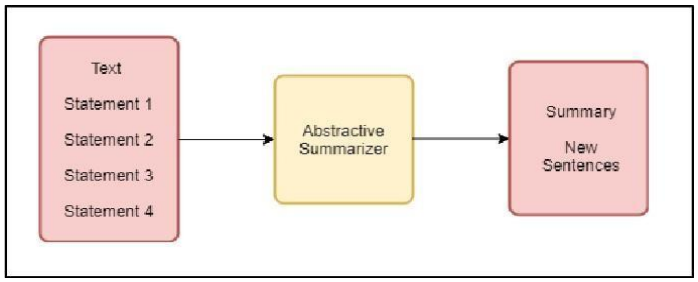


**FIG 4.3** Abstractive Summarization

## V. IMPLEMENTATION

**5.1. Data collection:** Gather a dataset of documents along with their abstracts. May make use of several sources like academic journals, news articles, or any domain-specific texts.

- Feature Extraction: Extract features from the text that is useful for summarization. This could include word frequency, sentence position, named entities, etc. You might also consider utilizing embeddings like Word2Vec or Glove to represent words.
- Preprocessing: Preprocess the text data to remove noise, such as HTML tags, punctuation, and stop words. You might additionally want to tokenize the text into sentences or words.

**5.2. Abstract understanding:** Train a model to understand the abstracts. This can entail using strategies. like deep learning (e.g., LSTM, Transformer) or simpler methods like TF-IDF to understand the important concepts inthe abstracts.

**5.3. Text Summarization:** Once you have a model that can understand abstracts, use this understanding to guide the text summarization process. This could involve scoring sentences based on their relevance to the abstracts and selecting the top-ranked sentences as the summary.

**5.4. Evaluation**: Evaluate the performance of your summarization system using metrics like ROUGE (Recall-Oriented).

**5.5. Data collection:** Gather a dataset of documents alongwith their abstracts. You have access to a number of sources. like academic journals, news articles, or any domain-specific texts

**5.6. Preprocessing:** Preprocess the text data to remove noise, such as HTML tags, punctuation, and stop words. You may also want to tokenize the text into sentences or words.

**5.7. Feature Extraction:** Extract elements in the text that are beneficial for summarization. This could include word frequency, sentence position, named entities, etc. You may also think about utilizing embeddings like Word2Vec or GloVe torepresent words.

**5.8. Abstract Understanding:** Train a model to understand the abstracts. This can entail using stratergies like deep learning (e.g., LSTM, Transformer) or simpler methods like TF-IDF to understand the important concepts inthe abstracts.

**5.9.Text Summarization**: Once you have a model that can understand abstracts, use this understanding to guide the text summarization process. This could involve scoring sentences based on their relevance to the abstracts and selecting the top-ranked sentences as the summary.

**5.10. Evaluation:** Evaluate the performance of your summarization system using metrics like ROUGE (Recall-Oriented Understudy for Gisting Evaluation) or BLEU (Bilingual Evaluation Understudy) scores, which compare the generated summaries to human-written ones.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

**5.11. Iterate and Improve:** Refine your system based onthe evaluation results. This could include using various models, feature sets, or hyper parameters to improve the summarization quality.

**5.12. Deployment**: Once you're satisfied with the performance of your system, deploy it for use. This could involve integrating it into an application or making it availableas a standalone service

## VI. RESULTS


**FIG:6.1**. Home page

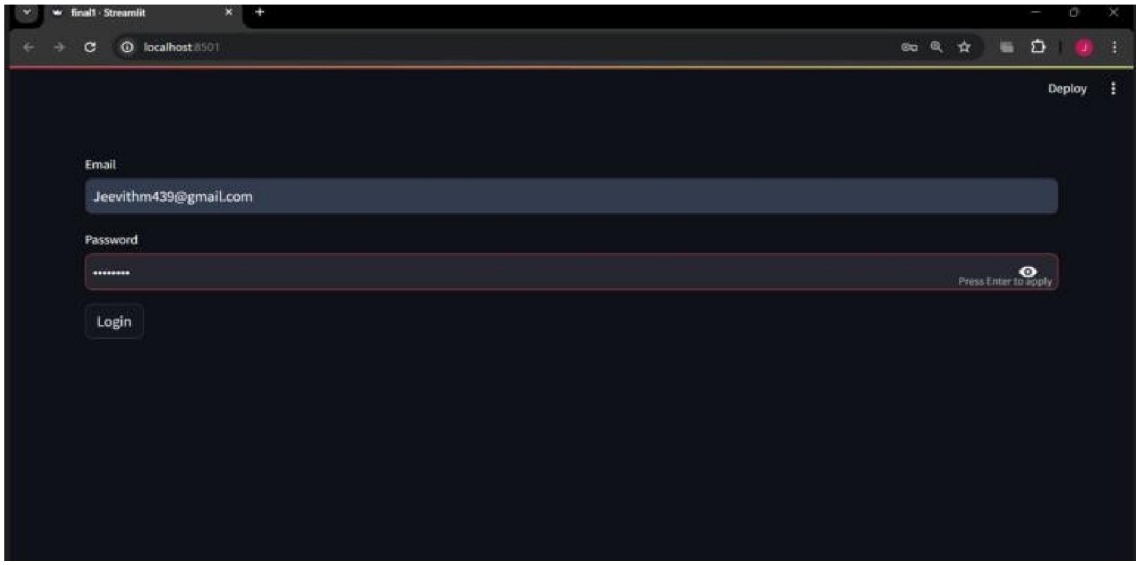


**FIG 6.2.** login page

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

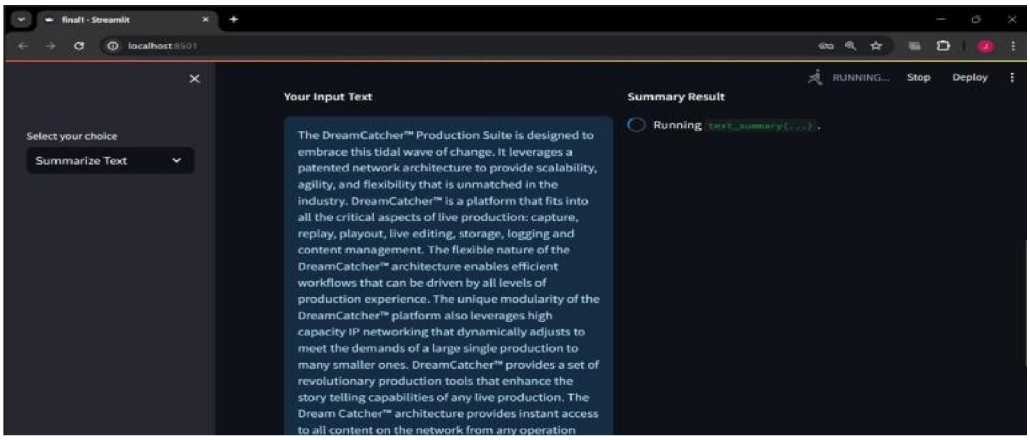


**FIG 6.3.**Text Summarization

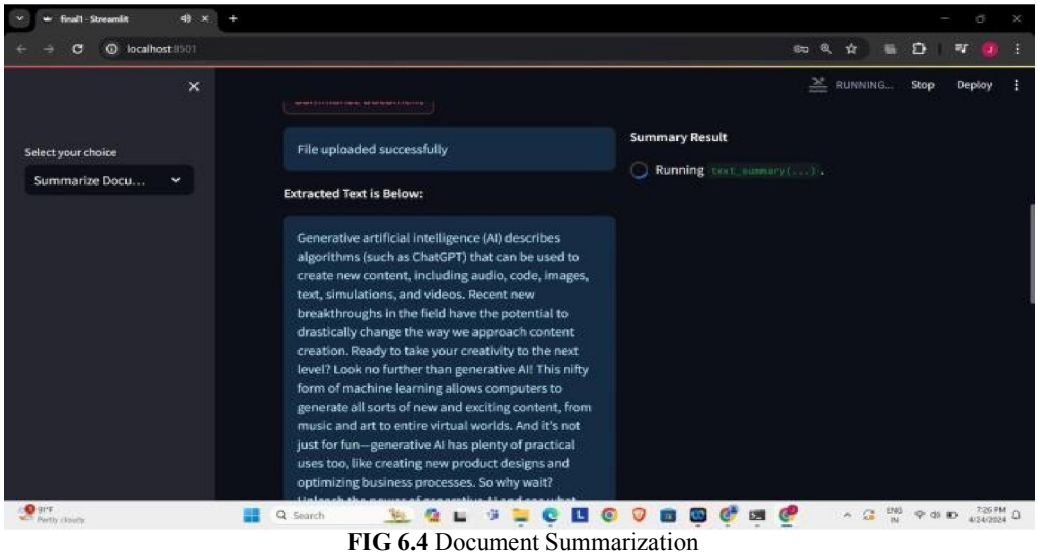


**FIG 6.4** Document Summarization

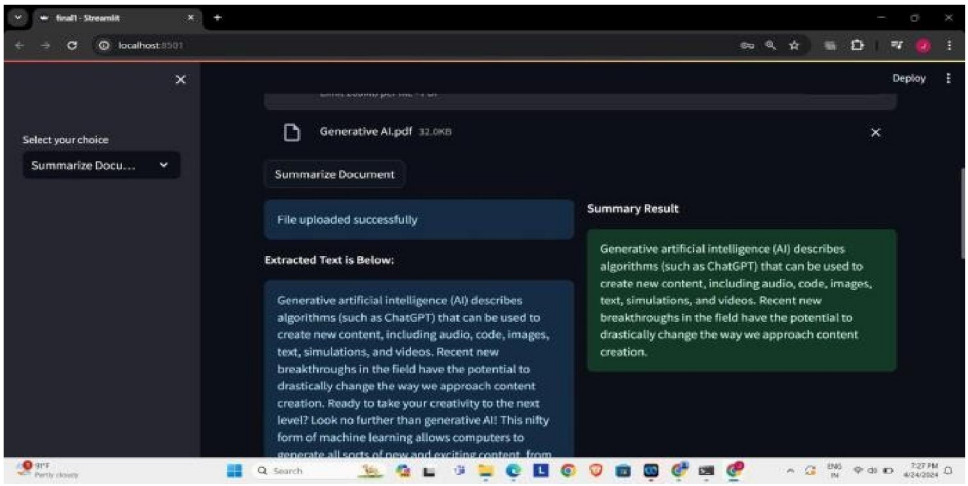


FIG 6.5 Summarization Result

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

## VII. CONCLUSION

This section presents a review of different methodologies employed for the identification and classification of plant species based on their leaf images. CNN architecture was This survey has showcased various methods of abstractive summarization. Abstractive summarization methods produce highly cohesive, coherent, less redundant summary and information rich. The goal is to provide an extensive survey and comparison of different techniques and approaches of abstractive summarization. In this survey some of the challenges and future research directions are also highlighted. In summary, the literature in abstractive summarization depicts major progress in various aspects. However, these.

Automatic summarization is a complex task that consists of several sub-tasks. Each of the sub-tasks directly affects the ability to generate high quality summaries. In extraction-based summarization the important part of the process is the identification of important relevant sentences of text. Use of fuzzy logic as a summarization sub-task improved the quality of summary by a great amount. The results are clearly visible in the comparison graphs. Our algorithm shows better results as compared to the output produced by two online summarizers. Thus, our proposed method enhances the quality of the summary by extracting the semantic linkages between ideas in the original text through the use of latest semantic analysis integrated into the sentences feature extracted fuzzy logic system. Although this paper's focus is limited to document summaries its concepts have border applicability. We must include extensive data sets and domain-specific information to the suggested strategy for multi-document summarization.

## REFERENCES

[1] J. Yan and S. Zhou, "A Text Structure-based Extractive And Abstractive Summarization Method," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, http:// doi: 10.1109/ICSP54964.2022.9778497.

[2] S. Ji and B. Yang, "Abstractive Text Summarization Based on Long-Short Transformer," 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023.

[3] N. Andhale and L. A. Bewoor, "An overview of Text Summarization techniques," 2016 International Conference on Computing Communication Control and automation (ICCUBEA), Pune, India, 2016, http:// 1-7, doi: 10.1109/ICCUBEA.2016.7860024.

[4] H. T. Le and T. M. Le, "An approach to abstractive text summarization," 2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR), Hanoi, Vietnam,2013http://doi: 0.1109/SOCPAR.2013.7054161.

[5] A. Venkataramana, K. Srividya and R. Cristin, "Abstractive Text Summarization Using BART," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, http://doi: 10.1109/MysuruCon55714.2022.9972639.

[6] H. T. Le and T. M. Le, "An approach to abstractive

[7] E. Lloret, E. Boldrini, T. Vodolazova, P. MartÃ nez-Barco, R. MuÃ´soz, and M. Palomar, "A novel concept-level approach for ultraconcise opinion summarization," Expert Systems with Applications, vol. 42, no. 20, pp. 7148 – 7156, 2015. (ICDIPC), 2015 Fifth International Conference on, Oct 2015, pp. 173–181.text summarization," in Soft Computing and Pattern Recognition (SoCPaR), 2013 International Conference of, Dec 2013, pp. 371–376.