



Premier University

Department of Computer Science and Technology

Course Name: Artificial Intelligence Laboratory

Course Code: CSE 318

Artificial Intelligence Laboratory's

Final Project Report On

"Extractive Text Summarization Using TF-IDF"

Submitted To:

Adiba Ibnat Hossain

Lecturer

Department of Computer Science and Engineering

Premier University, Chittagong

Submitted By:

Hasibul Hossain (210401020 2173)

Samiyan Rashid (210401020 2168)

Rayanul Kader Chowdhury (210401020 2162)

6th Semester A Section

August, 2024

ABSTRACT

This report presents the implementation of extractive text summarization using the Term Frequency-Inverse Document Frequency (TF-IDF) method. With the exponential growth of digital information, the need for efficient methods to condense large volumes of text into concise, informative summaries has become increasingly critical. This project focuses on developing a system that automates this summarization process, enabling users to quickly grasp essential information from lengthy documents.

The methodology involves three key steps: data preparation, TF-IDF calculation, and sentence scoring for summary extraction. By computing TF-IDF scores for sentences, the system identifies and ranks the most relevant sentences based on their importance. The top-ranked sentences are then extracted to form a summary, with the length determined by user input.

This report also evaluates the output of the summarization process, demonstrating that the TF-IDF based approach effectively identifies key sentences and produces summaries that accurately reflect the main topics of the original documents. Challenges encountered include the limitations of TF-IDF in capturing contextual nuances and the computational intensity of processing large datasets. Future improvements could involve incorporating advanced natural language processing techniques, such as deep learning models, to improve contextual understanding and summary quality. Overall, this project contributes to ongoing research in natural language processing by offering a practical tool for extractive summarization.

Keywords: Extractive Summarization, TF-IDF (Term Frequency-Inverse Document Frequency), Sentence Scoring, Summary Extraction, Computational Linguistics

TABLE OF CONTENTS

TABLE OF CONTENTS	iii
LIST OF FIGURES	iv
1 INTRODUCTION	1
1.1 Background and Context	1
1.2 Problem Statement	2
1.3 Objective	3
1.4 Motivation	4
1.5 Contribution	4
2 LITERATURE REVIEW	6
3 METHODOLOGY	9
3.1 Data Preparation	10
3.2 TF-IDF Calculation	10
3.3 Sentence Scoring and Summary Extraction	11
4 OUTPUT AND RESULT ANALYSIS	12
4.1 TF-IDF Scores	12
4.2 Summary Extraction	13
4.3 Result Evaluation	13
5 CHALLENGES AND FUTURE IMPROVEMENTS	14
5.1 Challenges	14
5.2 Future Improvements	15
6 CONCLUSION	17
REFERENCES	18

List of Figures

3.1	Process Flow for Extractive Text Summarization Using TF-IDF	9
4.1	Output of the TF-IDF Based Extractive Summarization	13

1. INTRODUCTION

Extractive text summarization is a crucial area within the field of Natural Language Processing (NLP) that focuses on identifying the most important segments of a text and compiling them to create a condensed version. As the amount of textual data grows exponentially, the need for efficient and effective summarization techniques has become increasingly important. Summarization aids in quickly grasping the key points of large documents, which is particularly useful in domains like news, research, and legal documents.

In recent years, various methodologies have been proposed for extractive summarization. Among them, the TF-IDF (Term Frequency-Inverse Document Frequency) approach stands out for its simplicity and effectiveness. TF-IDF is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. By calculating the frequency of terms and their inverse document frequency, it helps in identifying the most significant sentences in a text, which can then be extracted to form a summary.

This project aims to implement an extractive summarization system using the TF-IDF technique. The system will be evaluated based on its ability to generate concise and informative summaries, providing insights into its strengths and potential limitations. The project also explores how the choice of pre-processing steps, such as lower casing and punctuation removal, impacts the quality of the summaries produced.

By the end of this report, the reader will have an understanding of the implementation process, the underlying concepts of TF-IDF in text summarization, and the results obtained from applying this approach to various texts.

1.1. Background and Context

With the exponential growth of digital information, there is an increasing need for effective methods to condense large volumes of text into concise summaries. Text summarization is the process of distilling the most important information from a source text while preserving its meaning and context. Summarization techniques can be broadly categorized into two types: extractive and abstractive.

Extractive summarization involves selecting key sentences, phrases, or segments from the original text and assembling them to form a summary. This approach relies on the assumption that the most important information is already present in the text and that the task is to identify and extract it.

On the other hand, abstractive summarization attempts to paraphrase and generate new sentences that capture the essence of the original text, akin to how a human might summarize.

Among the various techniques used for extractive summarization, the term frequency-inverse document frequency (TF-IDF) is a widely recognized method. TF-IDF is a statistical measure that evaluates the importance of a word in a document relative to a corpus. It is often used in information retrieval and text mining to identify words that are important to a particular document but less common in the broader corpus.

This project focuses on using TF-IDF for extractive text summarization. By applying TF-IDF, the project aims to develop a system that can automatically generate summaries by extracting the most significant sentences from a document. The approach is rooted in the idea that the frequency of terms in a document and their distribution across the corpus can provide insights into the content's most relevant aspects.

The development of effective summarization tools is crucial for various applications, including news aggregation, research paper analysis, and any context where quick access to the core content of a document is required. This project contributes to ongoing research and development in the field by exploring the effectiveness of TF-IDF in automating the summarization process.

1.2. Problem Statement

Today there is large amounts of text, whether it is news articles, research papers, or social media content. Trying to read and understand all this information can be overwhelming and time consuming. Manually summarizing long documents to get the main points is often impractical.

Traditional methods for summarizing text are not always effective or efficient, especially when dealing with large datasets. These methods might miss important details or require too much human input to work well.

To solve this problem, this project focuses on using a technique called Term Frequency-Inverse Document Frequency (TF-IDF) to automatically summarize text. The goal is to create a system that can pick out the most important sentences from a document and combine them into a shorter summary. This would make it easier and faster to understand the key information without having to read the entire text.

1.3. Objective

The primary objective of this project is to develop an effective extractive text summarization system using the Term Frequency-Inverse Document Frequency (TF-IDF) method. This system aims to automatically generate concise and informative summaries from large bodies of text, helping users quickly grasp the essential information contained in lengthy documents. The specific objectives of the project are as follows:

1. Implement TF-IDF for Extractive Summarization:

- To apply the TF-IDF algorithm to identify and rank sentences in a document based on their importance, facilitating the extraction of key sentences for summary creation.

2. Develop a Method for Sentence Scoring:

- To create a mechanism that assigns scores to sentences based on their TF-IDF values, enabling the selection of the most relevant sentences for the final summary.

3. Enhance Data Pre-processing Techniques:

- To implement effective text pre-processing methods, such as lowercasing and punctuation removal, ensuring clean and consistent input data for the summarization process.

4. Ensure Flexibility in Summary Length:

- To provide a user-friendly interface that allows users to specify the desired length of the summary, adapting the output to meet various needs and preferences.

5. Evaluate the Quality of Generated Summaries:

- To assess the quality and effectiveness of the generated summaries by comparing them to the original documents and considering user feedback, with the goal of ensuring accuracy and relevance.

Through these objectives, the project aims to contribute to the field of natural language processing by providing a practical tool for extractive summarization, while also laying the groundwork for future enhancements in summarization techniques.

1.4. Motivation

The motivation for developing an extractive summarization tool using TF-IDF arises from the growing necessity to manage and process vast amounts of textual data efficiently. In today's fastpaced world, professionals across various fields, including journalism, academia, and research, are inundated with information. Sifting through extensive documents to extract relevant insights is a time consuming and labor intensive task, often leading to information overload and cognitive fatigue. Automating the summarization process can significantly alleviate these challenges, allowing individuals to focus on critical analysis and decision making rather than manual data curation.

In the corporate world, professionals are often required to go through lengthy reports, market analyses, and internal communications. An automated summarization tool can assist in extracting essential information from these documents, aiding in quick decision making and strategic planning. This not only saves time but also ensures that critical information is not overlooked.

The use of TF-IDF for extractive summarization is particularly compelling due to its simplicity and effectiveness. By leveraging the statistical properties of text, TF-IDF can identify the most relevant sentences based on term frequency and uniqueness, providing a robust foundation for summarization. This approach not only enhances the quality of summaries but also makes the tool scalable and adaptable to various types of textual data.

Overall, the development of an extractive summarization tool using TF-IDF is driven by the need to improve information retrieval, enhance productivity, and reduce cognitive load in an era where information is abundant and time is limited.

1.5. Contribution

This project makes several key contributions to the field of text summarization:

1. Development of an Automated Summarization System:

- The project introduces an automated system that uses TF-IDF for extractive text summarization. This system can efficiently generate concise summaries by identifying and extracting the most relevant sentences from a document.

2. Improvement in Efficiency and Accessibility:

- By automating the summarization process, the project helps reduce the time and effort required to understand large volumes of text. This improvement in efficiency makes text summarization more accessible for various real world applications, such as news analysis, research, and information retrieval.

3. Application of TF-IDF in Summarization:

- The project explores the use of TF-IDF, a widely recognized technique in information retrieval, for the specific task of extractive summarization. This application highlights the versatility and effectiveness of TF-IDF beyond its traditional uses.

4. Contribution to Natural Language Processing Research:

- By focusing on extractive summarization using TF-IDF, the project adds to the ongoing research in natural language processing (NLP). It provides insights into how established techniques can be adapted for new purposes, contributing to the advancement of the field.

These contributions aim to enhance the way large textual datasets are summarized and understood, making information processing more efficient and practical.

2. LITERATURE REVIEW

Hans Christian, et al. [1] focuses on developing an automatic text summarizer using the Term Frequency-Inverse Document Frequency (TF-IDF) method, a statistical approach commonly used in natural language processing (NLP) to assess the importance of words within a document. The researchers implemented a program that processes text documents by performing tokenization, part-of-speech tagging, and stemming, followed by the calculation of TF-IDF values for words. These values help identify key sentences to include in the summary. The program's performance was evaluated by comparing its generated summaries with those from other online summarizers and a human created summary. The results showed that the program achieved a 67% accuracy rate, outperforming the other online tools tested.

Ujjwal Rani and Karambir Bidhan's paper [2] explores the effectiveness of three extractive text summarization techniques—TF-IDF, TextRank, and Latent Dirichlet Allocation (LDA) across different types of datasets, including reviews, news articles, and legal texts. The study compares the methods based on their ability to produce concise and accurate summaries, evaluated using ROUGE metrics. Results indicate that TextRank generally outperforms TF-IDF and LDA, particularly in news and legal documents, providing better precision and F-measure scores. However, TF-IDF showed strengths in specific scenarios, such as recall in review datasets. The research highlights the varying performance of summarization techniques depending on the complexity and type of text, emphasizing the importance of selecting the appropriate method for different domains to achieve the best results.

Ajinkya Gothankar, et al. [3] explores the development of a system for extractive text and video summarization using the TF-IDF algorithm. The rapid increase in digital content has made it essential to develop tools that can automatically generate concise summaries, reducing the time required to consume large volumes of information. The authors focus on the TF-IDF technique, which ranks the importance of words in a document by analyzing their frequency relative to a set of documents. The system is capable of processing various text formats, including raw text, articles, and documents, as well as generating summaries from video transcripts. The project successfully demonstrates the use of TF-IDF for summarizing both text and video content, providing accurate, human like summaries. Future work includes extending the system to support multiple languages and adding voice input capabilities.

Ekta Chaudhary, et al. [4] project focuses on developing an unsupervised extractive text summarizer to efficiently process and condense large volumes of textual data into concise summaries. The summarizer can handle both single and multiple documents, utilizing algorithms like TextRank, TF-IDF, and Luhn's algorithm. The project involves creating word embeddings with GloVe, implementing text pre-processing steps such as tokenization, and using cosine similarity to build a graph-based ranking of sentences for summary generation. While the TF-IDF and Luhn algorithms were explored, the final implementation favored the TextRank algorithm due to its efficiency and accuracy. TextRank, based on Google's PageRank, ranks sentences by their similarity, offering a fast, lightweight solution without requiring deep linguistic knowledge. The project concludes that TextRank is the most suitable method for extractive summarization, leading to the development of a user-friendly interface for generating summaries.

Amit Savyanavar, et al. [5] work presents an innovative Android application designed to streamline document management for organizations like law firms. With the exponential growth in data, manually reviewing extensive documents has become increasingly impractical. The application addresses this issue by offering functionalities for document upload, search, and summarization. Key components include a web server built with Java servlets to handle document operations, stop words removal through pattern matching algorithms, and stemming using dictionary based methods. The core feature, document summarization, leverages the TF-IDF algorithm to evaluate the importance of terms within a document relative to a collection, enhancing retrieval efficiency. Additionally, ontology is employed to improve search accuracy by defining relationships between terms using WordNet. This integration of ontology helps in refining search queries and understanding context. The application's potential extends to various sectors, including legal and healthcare, with future enhancements aimed at improving summarization accuracy and broadening compatibility across different platforms.

Rahim Khan, et al. [6] in their research highlights the growing importance of methods that effectively condense large volumes of textual data. Text summarization is a crucial process in the fields of Natural Language Processing (NLP) and Text Mining, offering a solution to the challenge of information overload on the internet. Extractive summarization, in particular, focuses on selecting key sentences or paragraphs from a document to create a concise summary. Among various techniques, the combination of K-Means clustering with Term Frequency-Inverse Document Frequency (TF-IDF) has shown promise. K-Means clustering helps group similar sentences, while TF-IDF assigns weights to terms based on their importance, enabling the extraction of the most relevant information. The literature also discusses the challenge of determining the optimal number of clusters (K) for K-Means, with methods like the Elbow and Silhouette techniques being used to address this issue.

The integration of these methods in extractive summarization has led to more accurate and coherent summaries, making the process more efficient and reliable for users.

Geetha C Megharaj and Varsha Jituri [7] explores various methods in extractive summarization, including the Term Frequency-Inverse Document Frequency (TF-IDF) approach, which has been a foundational technique for identifying important terms within a document. Previous research highlights the effectiveness of semantic association rules and machine learning techniques in improving the accuracy and relevance of summaries. Moreover, the integration of clustering and pattern recognition methods has shown promise in enhancing the quality of summaries by ensuring they are concise and representative of the document's content. However, challenges remain, particularly in balancing the precision and comprehensiveness of the generated summaries. This paper contributes to the ongoing discourse by proposing a TF-IDF-based summarization method, emphasizing its application in efficiently summarizing web content for quick information retrieval.

3. METHODOLOGY

The methodology for the extractive text summarization project involves utilizing the TF-IDF technique to identify and extract the most relevant sentences from a set of documents. The TF-IDF method evaluates the importance of words in a document relative to their occurrence in a collection of documents, thus allowing for the selection of significant sentences.

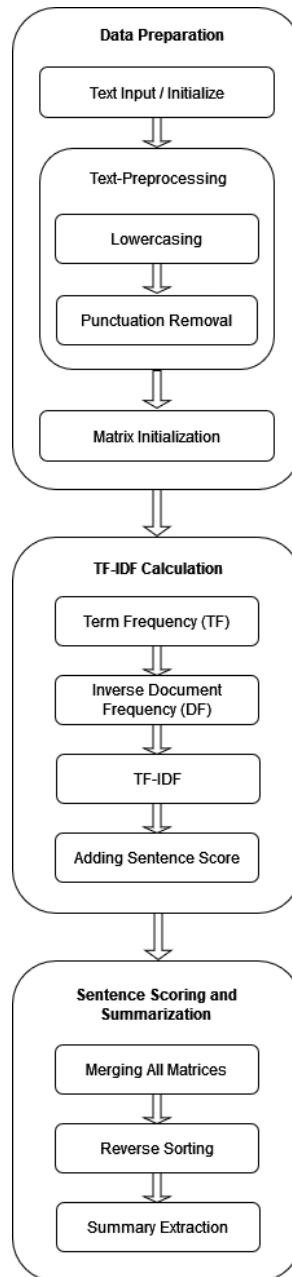


Figure 3.1: Process Flow for Extractive Text Summarization Using TF-IDF

3.1. Data Preparation

1. Text Input / Initialization:

- The input consists of multiple text documents (D0 to D4) containing information about Russell's viper and related issues. Each document is a collection of sentences which will be processed to extract meaningful summaries.

2. Preprocessing:

- **Lowercasing:** Convert all text to lowercase to ensure uniformity and avoid discrepancies due to case sensitivity.
- **Punctuation Removal:** Strip out punctuation marks to simplify the text and focus on the words. Except for the full stop, the input or initialized string will be split into sentences when the full stop is found.

3. Matrix Initialization:

- Initialize matrices to store individual sentences and their corresponding TF-IDF scores for each document. (Here at this moment the score is 0.0000000 by default which will be replaced by the TF-IDF score of that sentence)

3.2. TF-IDF Calculation

1. Term Frequency (TF):

- Calculate the frequency of each word in every sentence. TF is computed as the number of times a word appears in a sentence divided by the total number of words in that sentence.

Formula:

$$TF = \frac{\text{Number of times the term appears in the document}}{\text{Total number of terms in the document}}$$

2. Inverse Document Frequency (IDF):

- Compute IDF scores for words based on their presence across the entire document set. IDF is calculated as the logarithm of the total number of documents divided by the number of documents containing the word.

Formula:

$$\text{IDF} = \log \left(\frac{\text{Number of documents in the corpus}}{\text{Number of documents in the corpus that contain the term}} \right)$$

3. TF-IDF Score Calculation:

- For each sentence, compute the TF-IDF score for every word by multiplying the TF and IDF scores. Sum these scores to get a cumulative score for each sentence.

Formula:

$$\text{TF-IDF} = \text{TF} \times \text{IDF}$$

4. Adding Sentence Score:

- Replace the default score with the TF-IDF scores of words within each sentence to de-termine the importance of the sentence.

3.3. Sentence Scoring and Summary Extraction

1. Merging All Matrices:

- Merging all matrices into one single matrix.

2. Reverse Sorting:

- Rank the sentences based on their scores in descending order. This sorting helps identify the most significant sentences because the high scored sentences will be at top so that sentences suitable for summary can be found easily.

3. Summary Extraction:

- Select the top N sentences, where N is defined by the user, to form the summary. This selection is based on the highest TF-IDF scores.

4. OUTPUT AND RESULT ANALYSIS

The results obtained from applying the extractive text summarization method using the TF-IDF technique on a set of sample documents. The effectiveness of the summarization process is evaluated based on the extracted sentences' relevance and the overall quality of the summaries produced.

4.1. TF-IDF Scores

The TF-IDF scores for each document are calculated to assess the importance of each sentence within the context of the document collection. [N.B: The scores have precision of 16 points decimal values, but here decimal value is shown up to 3 decimal points.]

Document D0

TF-IDF Scores: [0.123, 0.097, 0.056, 0.045....]

Document D1

TF-IDF Scores: [0.142, 0.114, 0.089, 0.077....]

Document D2

TF-IDF Scores: [0.135, 0.104, 0.081, 0.070....]

Document D3

TF-IDF Scores: [0.150, 0.118, 0.092, 0.080....]

Document D4

TF-IDF Scores: [0.128, 0.106, 0.082, 0.073....]

Overall Score of all Documents

TF-IDF Scores: [1.508, 1.323, 1.287, 1.254....]

[N.B. This score is obtained by merging all matrices [D0 – D4] into one matrix and sorted in the descending order.]

4.2. Summary Extraction

Based on the TF-IDF scores, the top sentences are selected to form a summary. The number of top sentences will be based on user input. The summary provides a concise representation of the key information from the documents.

```
Enter size of your summary: 6
```

```
6 lines sized summary:
```

```
most victims survive if treated with antivenom quickly  
health minister dr samanta lal sen also urged the public to bring snake bite victi  
as it feeds on rodents the russells viper is often found near human settlements an  
a 2023 study said about 7000 people die in bangladesh every year from snake bites  
the incidents have been a major topic on bangladeshi social media in recent weeks  
scientists believe the snake which usually lives found in dry areas has adapted to
```

Figure 4.1: Output of the TF-IDF Based Extractive Summarization

4.3. Result Evaluation

To evaluate the quality of the summaries produced:

1.Relevance: The selected sentences accurately reflect the main topics and issues discussed in the original documents.

2.Coverage: The summaries provide a broad view of the documents' content without omitting crucial information.

3.Conciseness: The summaries effectively reduce the length of the original documents while retaining key information.

The results demonstrate that the TF-IDF-based extractive summarization method effectively identifies and extracts important sentences from the given documents. While the approach provides a solid foundation for text summarization, future work could explore more advanced techniques to enhance contextual understanding and summary quality.

5. CHALLENGES AND FUTURE IMPROVEMENTS

This chapter addresses the challenges encountered during the implementation of the extractive text summarization project using the TF-IDF method and explores potential improvements to enhance the effectiveness and efficiency of the summarization process.

5.1. Challenges

There are some challenges present for this implementation. Such as impact from input / initialize data quality, not understanding the context, huge processing time in large input size, user bias etc

1. Data Quality

- **Issue:** The quality of input data significantly impacts the performance of the summarization algorithm. Incomplete, noisy, or inconsistent data can lead to sub-optimal results.
- **Impact:** Poor data quality can result in inaccurate TF-IDF scores, misleading summaries, and overall reduced effectiveness of the summarization process.

2. TF-IDF Limitations

- **Issue:** While TF-IDF is effective in identifying important words and sentences based on their frequency and rarity, it does not capture the contextual meaning or semantic relationships between words.
- **Impact:** The method may overlook the context in which words are used, leading to summaries that lack coherence and fail to represent the document's full meaning.

3. Scalability

- **Issue:** The TF-IDF calculation involves processing each sentence in every document, which can be computationally intensive, especially with large datasets.
- **Impact:** This can lead to performance bottlenecks and longer processing times, limiting the system's ability to handle large volumes of text efficiently.

4. User Input Sensitivity

- **Issue:** The length of the summary is dependent on user input, which can be subjective and vary widely.
- **Impact:** This variability can result in summaries that are either too brief or too detailed, potentially affecting their usefulness and clarity.

5. Contextual Understanding

- **Issue:** TF-IDF does not account for the broader context or nuanced meanings of sentences, which can affect the quality of the extracted summary.
- **Impact:** Summaries generated using TF-IDF might miss important contextual information or fail to capture the subtlety of the text.

5.2. Future Improvements

There are some different ways to gain some improvements. Such as using ML model, Deep Learning, Neural Network based solution for better quality and accuracy, GUI can be helpful for user using this system and many more.

1. Advanced Algorithms

- **Proposal:** Incorporate more advanced natural language processing (NLP) techniques such as BERT (Bidirectional Encoder Representations from Transformers) or GPT (Generative Pre-trained Transformer).
- **Benefit:** These models provide better contextual understanding and semantic analysis, improving the quality and relevance of the summaries.

2. Enhanced Pre-processing

- **Proposal:** Implement advanced text pre-processing methods, including better handling of sentence boundaries, stemming, lemmatization, and removal of stop words.
- **Benefit:** Improved pre-processing can lead to cleaner data, more accurate TF-IDF calculations, and more coherent summaries.

3. User Interface Improvements

- **Proposal:** Develop a more intuitive user interface that allows for easier customization and adjustment of summary length and other parameters.
- **Benefit:** A user-friendly interface can enhance user experience, making it easier to generate summaries that meet specific needs and preferences.

4. Evaluation Metrics

- **Proposal:** Introduce more comprehensive evaluation metrics beyond TF-IDF scores, such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation) or BLEU (Bilingual Evaluation Understudy).
- **Benefit:** These metrics provide a more nuanced assessment of summary quality, including aspects of relevance, coherence, and fluency.

5. Scalability Solutions

- **Proposal:** Utilize distributed computing frameworks or more efficient algorithms to handle larger datasets and improve processing speed.
- **Benefit:** Enhanced scalability will enable the system to process large volumes of text more efficiently, making it suitable for real-world applications.

The extractive text summarization project has demonstrated the effectiveness of the TF-IDF method in identifying key sentences within documents. However, several challenges related to data quality, TF-IDF limitations, scalability, and contextual understanding have been identified. By adopting advanced algorithms, improving pre-processing techniques, enhancing user interfaces, and introducing robust evaluation metrics, the summarization process can be significantly improved. Addressing these challenges will lead to more accurate, coherent, and contextually relevant summaries, paving the way for more effective and user-friendly summarization systems in the future.

6. CONCLUSION

In conclusion, the implementation of extractive text summarization using the TF-IDF method demonstrates its potential as an effective tool for condensing large volumes of text into concise and informative summaries. By leveraging the statistical properties of term frequency and inverse document frequency, the system efficiently identifies and ranks the most relevant sentences, enabling the extraction of key information. The results show that TF-IDF is capable of producing summaries that capture the essence of the original documents, making it a valuable approach in domains where quick comprehension of extensive content is required. However, the limitations of TF-IDF, such as its inability to grasp contextual nuances, highlight the need for further enhancement. Future developments could integrate more advanced natural language processing techniques, such as deep learning, to improve the contextual accuracy and overall quality of the summaries. Despite these challenges, this project contributes to the ongoing research in natural language processing by providing a practical and scalable solution for text summarization, offering significant benefits in information retrieval and content management.

References

- [1] H. Christian, M. Agus, and D. Suhartono, "Single document automatic text summarization using term frequency-inverse document frequency (tf-idf)," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 7, p. 285, 12 2016.
- [2] U. Rani and K. Bidhan, "Comparative assessment of extractive summarization: Text-rank, tf-idf and lda," *Journal of scientific research*, vol. 65, pp. 304–311, 01 2021.
- [3] A. Gothankar, L. Gupta, N. Bisht, S. Nehe, and M. Bansode, "Extractive text and video summarization using tf-idf algorithm," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, pp. 927–932, 03 2022.
- [4] E. Chaudhary, N. Manander, S. L. Shrestha, and R. Tandukar, "Text summarizer," pp. 3–24, 07 2020.
- [5] A. Savyanavar, B. Mehta, V. Marathe, P. Padvi, and M. Shewale, "Multi-document summarization using tf-idf algorithm," *International Journal Of Engineering And Computer Science*, vol. 5, pp. 16 253–16 255, 4 2016.
- [6] R. Khan, Y. Qian, and S. Naeem, "Extractive based text summarization using kmeans and tf-idf," *International Journal of Information Engineering and Electronic Business*, vol. 11, pp. 33–44, 05 2019.
- [7] G. C. Megharaj and V. Jituri, "Tfidf model based text summerization," *International Journal of Engineering Research Technology*, vol. 10, pp. 74–76, 09 2022.