MANGALORE UNIVERSITY

A Mini Project Report On

"KANNADA TEXT SUMMARIZATION"

Carried out and Submitted

To

DEPARTMENT OF POST-GRADUATE STUDIES AND RESEARCH IN COMPUTER SCIENCE

2024-25

Submitted By

ANANYA S

Reg. No: P05AZ23S038003

II Year M.Sc. Computer Science

Under the Guidance of

Dr. H. L. Shashirekha

Professor, Department of Computer Science

Mangalore University

Mangalagangothri-574199

DEPARTMENT OF POSTGRADUATE STUDIES AND RESEARCH IN COMPUTER SCIENCE
MANGALORE UNIVERSITY, MANGALAGANGOTHRI-574199

MANGALORE UNIVERSITY

DEPARTMENT OF POSTGRADUATE STUDIES AND RESEARCH IN COMPUTER SCIENCE

Mangalore University

Mangalagangothri, Mangalore-574199

CERTIFICATE

This is to certify that the project work entitled **"KANNADA TEXT SUMMARIZATION"** has been successfully carried out in the Department of Post- Graduate Studies and Research in Computer Science by **Ms. ANANYA.S(Reg. No: P05AZ23S038003),** student of third semester Master of Computer Science, under the supervision and guidance of **Dr. H. L. Shashirekha,** Professor, Department of Post-Graduate Studies and Research in Computer Science, Mangalore University. The project is partial fulfilment of the requirements for the award of **Master of Computer Science** by **Mangalore University** during the academic year **2024-25.**

Internal Guide **Chairperson**

Internal Examiner **External Examiner**

Submitted for the viva-voce examination held on____________

Date: ____________

DECLARATION

This project work entitled **“KANNADA TEXT SUMMARIZATION”** has been successfully carried out by me under the supervision and guidance of **Dr. H. L. Shashirekha,** Professor, Department of Post-Graduate Studies and Research in Computer Science, Mangalore University, Managalagangotri. This project is submitted in partial fulfilment for the award of **Master of Science in Computer Science** degree by **Mangalore University** during the academic year 2024-25. This work or any part of this work has not been submitted to any other University or Institute/School for the award of any other Degree or Diploma.

Date: **Name: ANANYA S**

Place: Mangalagangothri **Reg No: P05AZ23S038003**

ABSTRACT

Natural Language Processing is a vast area which has great importance when people started to interpret human language from one from to another. There is a lot of information available on the internet today, but it is difficult to read the long text quickly and efficiently. Summarization is one of the research works in NLP which concentrates on providing meaningful summary of the long text using various NLP tools and techniques. Text summarization methods generate summaries of the relevant information from original content. Text summarization methods are two types: abstractive and extractive. For English, numerous text summarization techniques exist in the literature. But for Indian languages, there are only a few techniques developed.

Therefore, brevity should be able to reduce long words into short sentences with shorter words describing the same subject.

For Kannada text summarization task, TF-IDF feature extraction and LDA model were employed to extract keywords and rank text. The algorithm automatically extracts keywords for summarizing texts in Kannada datasets.

Keywords: Text summarization, Natural language processing, Indian languages, Language dependency, Latent Dirichlet allocation.

1.INTRODUCTION

Natural Language Processing (NLP) revolutionizes how we interact with and make sense of textual data. Text summarization, a prominent application of NLP, addresses the challenge of condensing lengthy documents, articles, or passages into shorter versions while retaining the core ideas and critical information. Text summarization task can be classified into two categories, extractive summarization and abstractive summarization. Text Summarization process can be highly beneficial in numerous applications, including news aggregation, research synthesis, and content curation. By leveraging machine learning (ML) methods, summarization models can learn to identify and extract key sentences and concepts from large volumes of text, significantly reducing the time and effort required to comprehend lengthy documents. The effectiveness of these models is typically evaluated using metrics such as Rouge, F1 score, precision, and recall to ensure they produce relevant and coherent summaries.

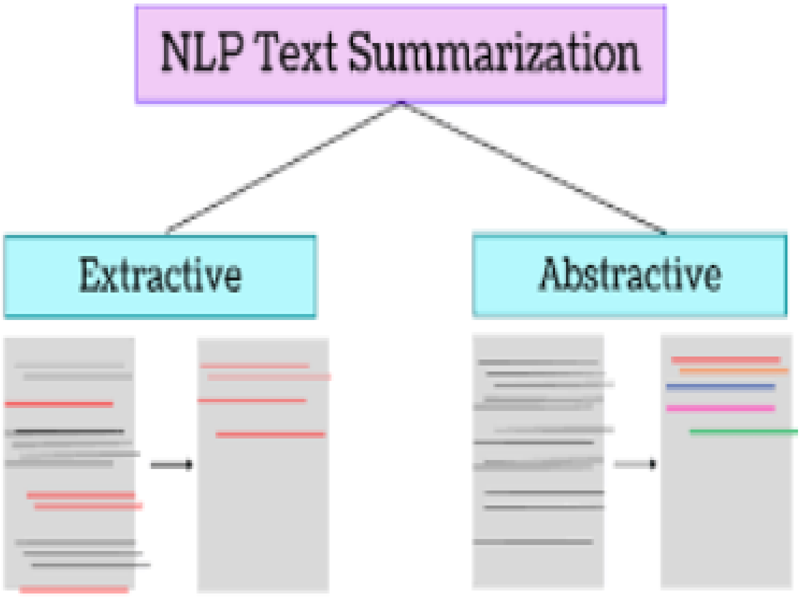


Fig-1: Extractive and Abstractive

1.1 Extractive Summarization

It involves selecting and rearranging sentences directly from the source text and extracting the most informative sentences to form a coherent summary. Extractive summarization methods use text mining approaches. The process of Extractive text summarization includes the following:

I. Text Pre-Processing
II. Sentence Scoring
III. Sentence Scoring
IV. Post-processing

1.2 Abstractive Summarization

Abstractive summarization goes a step further by generating new sentences that has the same meaning of the original content. This approach requires a deeper understanding of language and context, often involving techniques such as language generation models.

Text summarization for Indian languages is challenging due to the lack of adequate tools like appropriate taggers, parsers, synsets, etc. It can also categories into several other domains such as single and multi document summarization, query and topic focused summarization, monolingual and multilingual summarization etc. Single document summarization is simply defined as summary of text from a single document and summary from more than one document is called multi document. summary process which involves summarization on the basis of Interrogating phrase is called query based summarization. Here, the system itself generate the topic according to given query and create summary of topic related documents. The topic based summarization involves summarization of topic related documents. It will help in finding a quick view of several documents in less time.

In the context of the Kannada language, text summarization poses unique challenges due to its rich morphology and syntactic structure. This project aims to address these challenges by developing a sophisticated system capable of processing and summarizing Kannada text. This project involves collecting a diverse set of Kannada documents, preparing them for analysis, and identifying the critical elements that contribute to effective summarization. By incorporating Kannada-specific linguistic features and using advanced NLP techniques, the system strives to generate summaries that accurately reflect the original content, thereby enhancing the accessibility and comprehensibility of Kannada literature and information. The research project developed a comprehensive tool for summarizing Kannada text using NLP techniques. The project involved several key steps: scraping Kannada text from web sources, processing and cleaning the text data, and implementing a summarization algorithm. A frequency-based approach was utilized to score sentences, filtering out stop words and assigning scores based on word frequencies. The summarization model was designed to identify and retain the most significant sentences, producing a concise summary of the input text. Latent Dirichlet Allocation (LDA) is a generative probabilistic model used for topic modelling, which is the process of identifying topics present in a collection of documents. In this project LDA model was used, LDA used to generate the topic vector for the supplied document. The performance of the summarization model was evaluated using metrics such as Rouge, recall, F1 score, and precision, providing insights into its effectiveness and areas for further improvement. This project not only advances Kannada text processing but also contributes to the broader field of language technology.

2.RELATED WORKS

[1] The paper **"Saaramsha: Leveraging NLP for Efficient Kannada Text Summarization"** was published by **P. Maharshi Reddy et.al.** This paper presents an extractive summarization approach for Kannada text using Natural Language Processing (NLP). It employs Term Frequency (TF) and Google Suggestion Score (GSS) coefficients to identify and rank key sentences. The methodology includes web scraping, text processing, keyword extraction, and sentence ranking to generate concise summaries. The model was tested on a Kannada news dataset, ensuring efficient summarization while retaining key details. A user-friendly interface allows document uploads and summarization. Future work aims to incorporate abstractive summarization using transformers and enhance web scraping techniques for better content extraction. This research contributes to improving Kannada NLP and facilitates efficient news summarization and information retrieval.

[2] The paper "**Text summarization using Natural Language Processing (NLP)"** was published by **B Rajesh et.al.** this paper explains about to extract key information from large volumes of text efficiently. The study discusses two primary approaches: **extractive summarization**, which selects important sentences directly from the text, and **abstractive summarization**, which generates new sentences while preserving the original meaning. In this paper they implemented an extractive summarization model using SpaCy, NLTK, and Heapq in Python. Their approach involves pre-processing text, sentence scoring, and selection based on word frequency and importance. The model was evaluated using ROUGE metrics, showing promising recall and precision scores. A Flask-based web application was developed to provide a user-friendly summarization tool. The study highlights the practical applications of text summarization in news reporting, research, online product reviews, and social media marketing. Future enhancements include integrating deep learning models for improved abstractive summarization and expanding functionality for diverse text formats. The findings reinforce the significance of NLP-based summarization in managing

information overload, improving accessibility, and enhancing user experience in various industries.

[3] The paper "**Automated Kannada Text Summarization using Sentence Features"** was published by **Arpitha Swamy et.al.** This paper addresses the challenge of summarizing Kannada text documents using an extractive approach. The proposed system ranks sentences based on five key features: Term Frequency, Term Frequency-Inverse Sentence Frequency, Keywords, Sentence Length, and Sentence Position. These features are used to score and select the most important sentences for inclusion in the summary. The system was evaluated on a dataset of 50 Kannada documents from five different categories, with performance assessed using the ROUGE toolkit. The results showed that the proposed method achieved good precision and recall scores, demonstrating effectiveness in producing summaries comparable to human-generated ones. The study highlights the need for improved text summarization in Indian languages and suggests future enhancements by incorporating additional statistical and linguistic features.

[4] The paper **Extractive Summarization of Kannada Multi Documents using LDA** was published by **Veena R et.al**. This paper explores the use of **Latent Dirichlet Allocation (LDA)** for summarizing multiple Kannada-language documents. The study focuses on extractive summarization, where key sentences are selected from the original text to generate a concise summary while preserving the original meaning. The authors discuss the challenges of Kannada text processing, including its rich morphology and lack of extensive NLP resources. They apply **LDA for topic modelling**, identifying the most relevant topics within a document set. The methodology involves **preprocessing**, **topic extraction using LDA**, and **sentence ranking** to create summaries. The paper evaluates the effectiveness of the approach through experiments and compares the results with other summarization techniques. The study contributes to Kannada natural language processing (NLP) by enhancing summarization techniques, which could be useful for various applications like news aggregation and document summarization.

[5] The paper **Multi-Document Summarization Using K-Means and Latent Dirichlet Allocation (LDA) – Significance Sentences** was published by **Shiva Twinandilla** et.al. The paper explores multi-document summarization using a novel approach that combines K-Means clustering and Latent Dirichlet Allocation (LDA) with Significance Sentences. The study addresses the issue of redundancy in online news, often leading to "yellow journalism," making it hard to distinguish factual information. While LDA-Significance Sentences alone provide effective summarization, they lack topic organization. By integrating K-Means clustering, the method groups documents by topic before summarization, improving coherence. Experiments show that the best alpha value is 0.001, achieving a ROUGE-1 score of 0.5545, and the optimal summarization level is 30%, with a score of 0.6118. The effectiveness of the method depends on correctly clustering documents, as improper clustering lowers summarization quality.

[6] The paper **"Comparative Assessment of Extractive Summarization: TextRank, TF-IDF and LDA"** was published by **Ujjwal Rani** et.al. The paper explores automatic text summarization, a key area in Natural Language Processing (NLP) that helps extract important information from large text documents. It compares three extractive summarization techniques: TF-IDF (Term Frequency-Inverse Document Frequency), TextRank, and Latent Dirichlet Allocation (LDA). The study evaluates these methods using three datasets—reviews, news articles, and legal texts—and assesses their effectiveness using ROUGE metrics. The results indicate that TextRank performs best overall, especially for news and legal texts, due to its ability to rank sentences based on graph-based relationships. TF-IDF performs well for recall but lags in precision and F-measure. LDA, while effective for topic modelling, does not outperform TextRank. The study suggests that TextRank is the most reliable method for automatic summarization. Future research could explore hybrid approaches, improved similarity measures, and supervised LDA to enhance summarization accuracy.

[7] The paper **"Text Summarization for Indian Languages: Finetuned Transformer Model Application"** was published by **V Ilanchezhiyan** et.al. This paper explores automated text summarization for Indian languages, particularly Hindi,

Gujarati, and Bengali, using transformer-based models. The study is part of the ILSUM task at FIRE 2023, where the authors' model, m2023-t5-base, ranked second for English summarization. The methodology involves fine-tuning pre-trained models (mT5-small, mT5-base) to generate concise and meaningful summaries. The workflow includes translating Indian language texts into English, summarizing them using T5-base, and back-translating the output. Evaluation metrics such as ROUGE and BERT scores indicate improved performance after fine-tuning. Results show significant improvement in summarization quality, demonstrating the effectiveness of transformer-based models for Indian languages. The study highlights the challenges of code-mixing and script-mixing and suggests future improvements for multilingual NLP applications. The research contributes to advancing automatic summarization techniques for low-resource languages.

[8] The paper **"Trends in Extractive and Abstractive Techniques in Text Summarization"** was published by Neelima Bhatia and Arunima Jaiswal explores two primary approaches to text summarization: extractive and abstractive. Extractive summarization selects key sentences based on statistical and linguistic features, while abstractive summarization generates new sentences by understanding the content. The paper discusses various extractive techniques, such as TF-IDF, clustering, graph theory, machine learning, neural networks, and fuzzy logic. It also covers abstractive techniques, including structured-based (tree-based, ontology, and rule-based) and semantic-based approaches (semantic graphs and multimodal models). Challenges include redundancy, lack of coherence in extractive methods, and representation issues in abstractive summarization. The paper concludes that while extractive summarization is more widely used, abstractive techniques hold promise for more meaningful and concise summaries.

[9] The paper "**Accountability of NLP Tools in Text Summarization for Indian Languages"** was published by Pradeepika Verma and Anshul Verma. This research paper discusses the challenges and advancements in automatic text summarization for Indian languages. It highlights the exponential growth of digital information and the necessity of summarization tools for quick access to relevant data. While such tools are well-developed for English, Indian languages face difficulties due to limited NLP resources. The paper surveys existing summarization techniques for various Indian languages, including Hindi, Punjabi, Bengali, Marathi, Tamil, and Kannada, outlining

methods like graph-based, statistical, and machine learning approaches. It identifies key issues such as the lack of stop-word lists, stemming tools, and sentence boundary detection rules. The study also compares summarization performance between Indian and English languages, demonstrating the impact of NLP tool availability on accuracy. The authors conclude that enhancing NLP tools and exploring advanced approaches, such as neural networks, can improve text summarization for Indian languages.

[10] The paper **"An extractive text summarization approach using tagged-LDA based topic modelling"** was published by Ruby Rani & D. K. Lobiyal. The paper presents an extractive text summarization approach using a tagged-LDA-based topic modelling technique, focusing on Hindi novels and stories. Due to the lack of available corpora and processing tools for Hindi, the authors compiled a dataset and developed a linguistic feature-rich topic modeling method for automatic summarization. The proposed approach includes four sentence-weighting variants: Lexical LDA, Sliding Window LDA, Relative Sentence Weighting LDA, and Integrated Sentence Weighting LDA. The method enhances summary coherence and diversity while minimizing redundancy through a smoothing technique. The evaluation, based on gist diversity, retention ratio, and ROUGE scores, demonstrates that the proposed models outperform traditional topic modeling and baseline approaches. The method is also tested on English datasets for comparative performance analysis, confirming its effectiveness in generating concise, informative, and coherent summaries.

3.METHODOLOGY

Dataset for Kannada text summarization is collected from articles of different domains including latest news, national and international news, sports, entertainment, astrology, jobs, business from https://Kannada.news18.com website using a handwritten scraper. 23,166 URLs are collected at the early phase of dataset collection. Unique articles are filtered out during pre-processing so that the summary (description) contains between 15 and 40 tokens. The process has resulted into the dataset with 16,975 unique articles. Average number of tokens in the summary (description) is 23.

3.1.Preprocessing

Natural language processing always involves text preprocessing. Once we have the text, we subject it to standard text processing procedures to prepare it for summarization. This involves Tokenization, Remove Newlines, Tabs, and Zero-width Characters, Replace Certain Special Characters, Remove Digits, URLs, Mentions, HTML Tags, Remove Special Characters and Unwanted Punctuations, Remove Remaining Non-Alphanumeric Characters, Stop Word Removal.

- **Tokenization**
  Tokenization simplifies the text into manageable units for further processing.
- **Remove Newlines, Tabs, Zero-width Characters**
  `\n`, `\r`, `\t`, `\u200d` these characters are replaced because, these characters are whitespace-like but may interfere with tokenization.
- **Remove Digits, URLs, Mentions, HTML Tags**
  URLs are unnecessary in text analysis. Mentions are often present in social media text but are not relevant in linguistic analysis. HTML tags are not useful in text analysis.
- **Remove Special Characters and Unwanted Punctuations**
  "@!?.,;:'$\"₹()[]{}<>।॥-|/*0123456789\u200C\u00A0\u200d-"these characters were removed using a loop to replace each unwanted character

with a space. Special characters and punctuation do not contribute to meaningful tokenization.

- **Stop Word Removal**
  The commonly occurring words are called stop words. In our project we use 100 stop word list. Stop words are common words (e.g.,” ಎಂದು” “ಮತ್ತು”) that do not contribute meaningfully to analysis. Which have less importance in the conclusion of document, are discarded for summarization process. In Kannada words like (and), (it), ಇದೆ (is), etc. are frequently used stop words in sentences.
- **Remove Remaining Non-Alphanumeric Characters**
  In this stage, only words and numbers are retained, filtering out Punctuation marks, other symbols and other unwanted characters. Ensures only letters and numbers remain.

3.2. Feature Extraction

TF-IDF

This approach is termed as Term frequency-inverse document frequency is a statistical extraction approach that works by the comparison of the frequency of words in a particular document with the inverse proportion frequency of that word in other documents. It means if a word appears frequently in a document then it might be assumed by the user that it is important for the document. But if the same word appears frequently in other documents also then that word is not significant at all. Term Frequency is a measure to determine how frequently a term appears in the document.

$$tf(w) = \frac{\text{Total count of appearance of a term } w \text{ in D}}{\text{Total count of terms in D}} \quad (2)$$

idf(w) = log (Total number of documents / Number of documents containing the term w) (3)

D here represents a particular document

Dn here represents collection of documents

Hence TF-IDF is calculated in the following way for w which is a word in the document:

tf-idf (w) = (2) * (3).

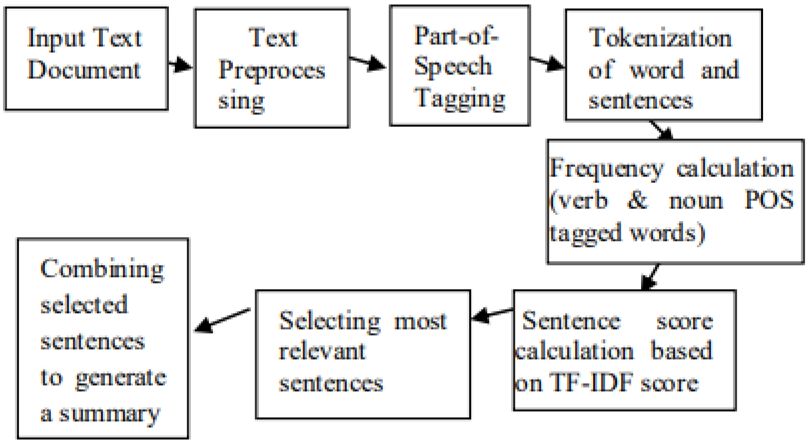


Fig: Flow Chart for TF-IDF Approach

3.3. Latent Dirichlet Allocation

Latent Dirichlet Allocation is an unsupervised approach based on probabilistic algorithm extensively used for topic modelling. Topic modelling comes up with the ways to understand, organize, and generate summaries of large documents. The first step towards this approach is the extraction of text data from the documents and then preprocessing the text. Preprocessing includes cleaning of text data, stop word removal and next step includes application of LDA for topic modelling. LDA represents the documents as the combination of topics as it breaks down the text document into topic clusters, which are based on probability distribution to mark the importance of the topic with regards to document.

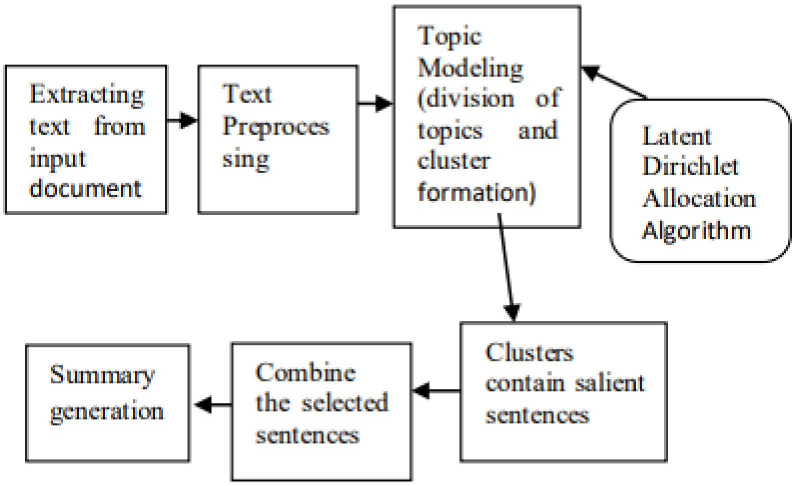


Fig: Flow Chart for the LDA approach

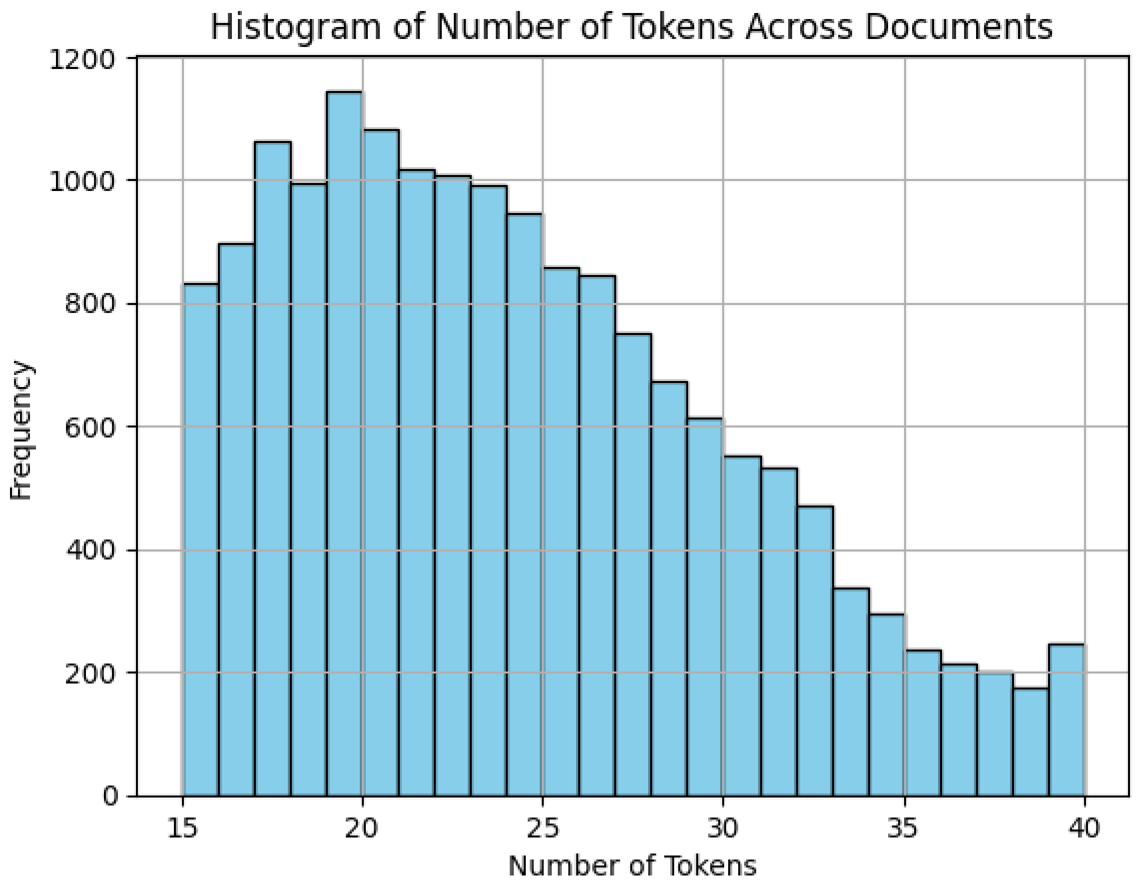
The clusters generated, contains the salient sentences from the original text document. The relevant topics which are identified would be linked to each cluster. The sentences of the source document are assigned to different recognized topics, in order to increase coverage. The property of topic selection from the document enhances the summarization process. LDA approach used in this paper isolates the five topics which are distinct from each other to generate a summary.

4.EXPERIMENT AND RESULTS

4.1.DATASET

The dataset is created for Kannada dataset for text summarization by collecting articles from different domains including latest news, national and international news, sports, entertainment, astrology, jobs, business from https://Kannada.news18.com website using a handwritten scraper. 23,166 urls are collected. Unique articles are filtered out during pre-processing so that the summary (description) contains between 15 and 40 tokens. This process has resulted into the dataset with 16,975 unique articles. Average number of tokens in the summary (description) is 23.

```
Average number of tokens per document: 23.83522827687776
Maximum number of tokens in a document: 40
Minimum number of tokens in a document: 15
```



4.2.Evaluation

The generated system summary is evaluated by comparing it to the human summary using the ROUGE toolkit. There are different ROUGE measures - ROUGE1, ROUGE2, ROUGEL, and ROUGES etc. ROUGE-L: Measures the Longest Common Subsequence (LCS) between the generated and reference summaries, capturing sentence-level structure similarity. We have used ROUGE1 measure to evaluate the system generated summaries. ROUGE is termed as Recall Oriented Understudy for Gisting Evaluation. This is a way of doing evaluation that computes the similarity between system produced summaries and the human-generated summary. The generated system summaries are evaluated using ROUGE toolkit in terms of evaluation measures - Recall, Precision and F-score. Precision can be defined as the ratio of count of common sentences present in both system and model summaries over the total count of sentences present in the system summary. Recall is defined as the ratio of number of common sentences present in both system and model summaries and the total count of sentences present in the model summary. F-score is defined as a composite measure that combines recall and precision.

Precision = (Number of overlapping words) / (Total words in generated summary)

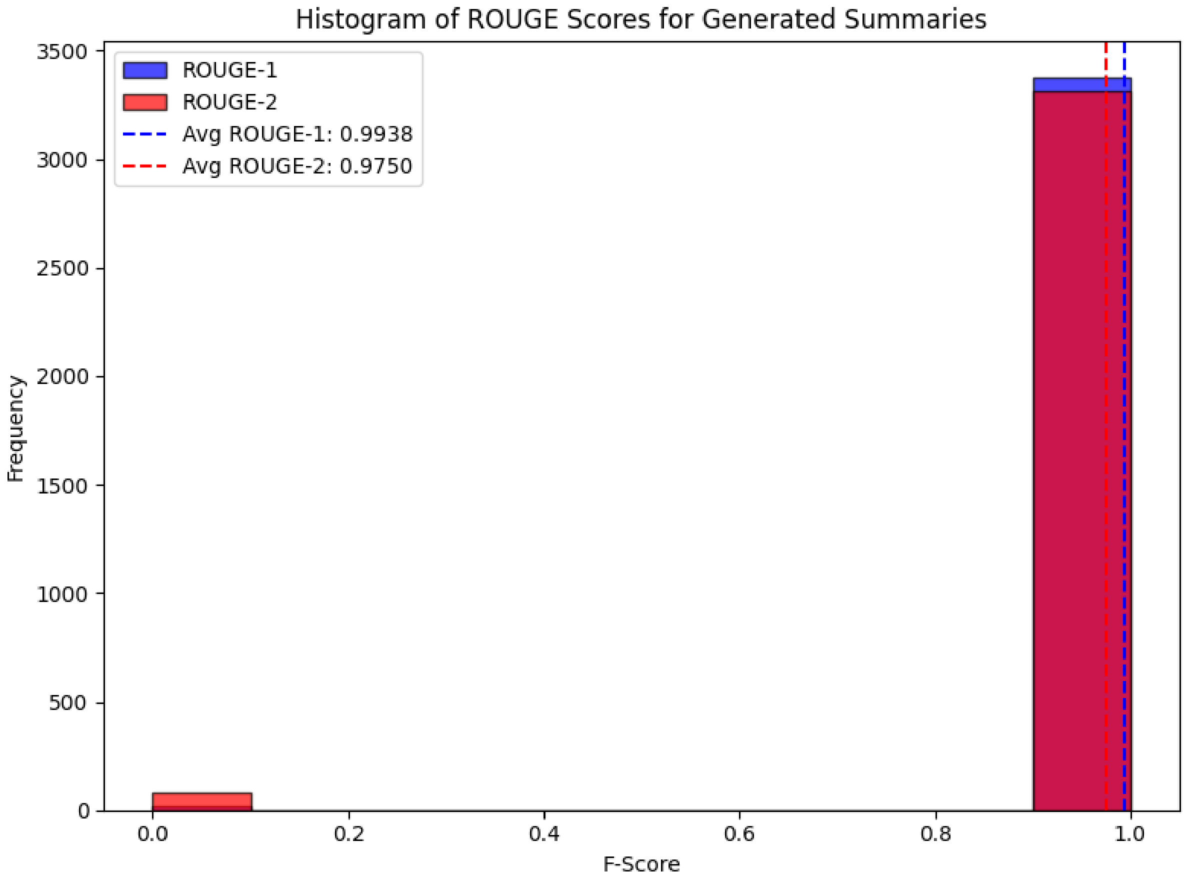
Recall = (Number of overlapping words) / (Total words in reference summary)

F1-score = Harmonic mean of Precision and Recall

4.3.Result

The performance of the system is assessed based on the average ROUGE scores obtained for the datasets. The approaches used for text summarization are TF-IDF, and LDA. The analysis of ROUGE scores reveals that the performance of the summarization approaches varies depending on the dataset, with TextRank often providing the best overall performance. Graphs are provided to visualize the performance of the three summarization approaches in relation to the kannada datasets.

Average ROUGE-1 Score: 0.9938
Average ROUGE-2 Score: 0.9750

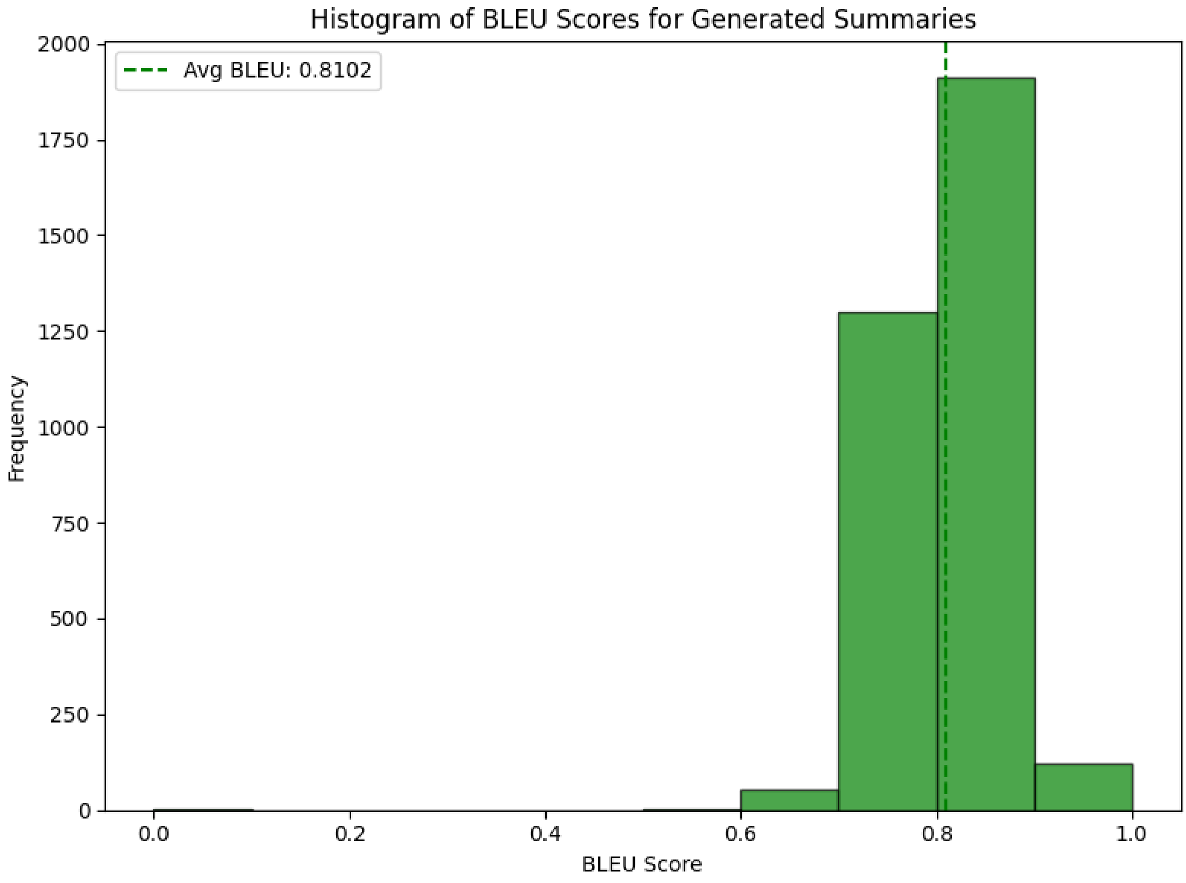


ROUGE Evaluation Report:

Metric	Precision	Recall	F1-Score
ROUGE1	0.00	0.03	0.01
ROUGE2	0.00	0.00	0.00
ROUGEL	0.00	0.03	0.01

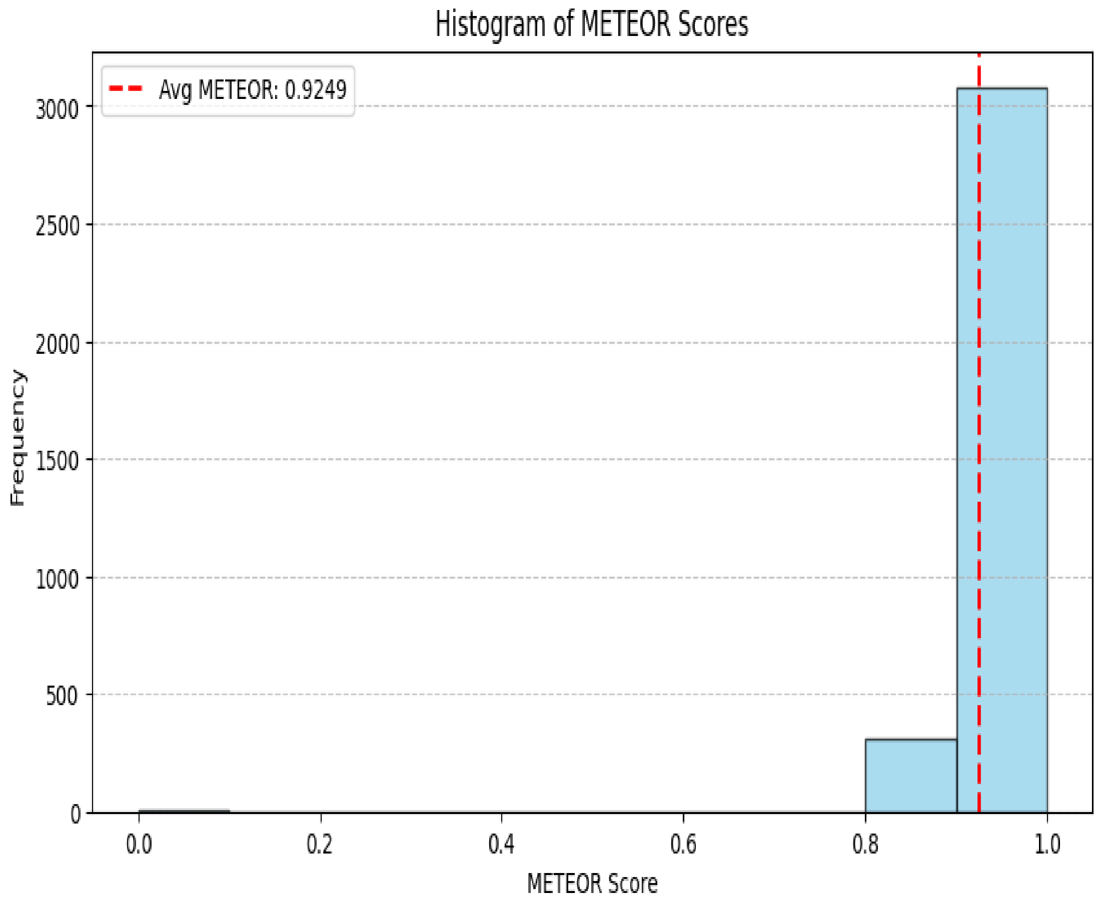
The BLEU score is commonly used in machine translation and summarization to measure the similarity between generated text and reference text. Calculate BLEU scores for a set of summaries by comparing them with their corresponding original documents. The results indicate an average BLEU score of **0.8102 .**

Average BLEU Score: 0.8102



The METEOR (Metric for Evaluation of Translation with Explicit Ordering) is a metric that was originally developed to assess the quality of machine translations, but it can also be used to assess the quality of machine summaries. It evaluates the similarity between the generated summary and the reference summary, taking into account grammar and semantics. METEOR scores are given on a scale of 0 to 1, with higher values indicating greater similarity between the generated summary and the reference summary.

Average METEOR Score: 0.9249



5. CONCLUSION

Text Summarization is the process of compressing of the original documents into the summaries in such a way that the generated summaries can be substituted with the original document without making any compromise with the information delivered from the document. Abstractive and Extractive are two widely used approaches for summarization. In this paper, the extraction based methodologies are implemented.TF-IDF, and LDA are widespread techniques that are used for document summarization. In this project dataset is collected from different domains including latest news, national and international news, sports, entertainment, astrology, jobs, business. The evaluation is done using the ROUGE metrics using Recall, Precision and F-measure criteria. Rouge metrics gives a average of 0.9 value. Precision, recall,F1-score values range between 0 to 1.Here we used another metrics also that is BLEU. Average BLEU score for the model is 0.81 ,average Rouge score is 0.9 ,compare to BLEU average Rouge score is better.

FUTURE WORKS

The generated summaries are evaluated using ROUGE toolkit with recall, precision and f-score evaluation measures. The performance of this proposed system is good in terms of average recall, average precision and average f-score values. In the future, the performance of the proposed system may further be enhanced by considering more statistical and linguistic features in the process of sentence scoring and sentence ranking. The evaluation value we got is not good enough, in future try to improve the value. Try to test this dataset with other models also and also evaluate the performance.

6. REFERENCES

- Ujjwal Rani and Karambir Bidhan .“Comparative Assessment of Extractive Summarization: TextRank, TF-IDF and LDA” Journal of Scientific Research Institute of Science, Banaras Hindu University, Varanasi, India. (Vol .65, Issue 1, 2021, pp.304-311).

- P. Maharshi Reddy, Lalam Aakash, Likhithraj A, Rashmi K.B. “Saaramsha: Leveraging NLP for Efficient Kannada Text Summarization” International Journal of Computer Applications (0975 – 8887).(vol. 186 – No.31, July 2024, pp-1-4).

- Veena R, Dr. D. Ramesh, Dr. Hanumanthappa M. “Extractive Summarization of Kannada Multi Documents using LDA”.(pp-561-571).

- B Rajesh, K Nimai Chaitanya, P Tejesh Govardhan, K Krishna Mahesh, B Sudarshan “Text Summarization using NLP” International Research Journal of Engineering and Technology (IRJET). (Vol: 11 Issue: 03 | Mar 2024, PP.702-708).

- Kishore Kumar Mamidala, Suresh Kumar Sanampudi, “TEXT SUMMARIZATION FOR INDIAN LANGUAGES: A SURVEY” International Journal of Advanced Research in Engineering and Technology (IJARET). (Vol. 12, Issue 1, January 2021, pp. 530-538).

- Arpitha Swamy, Srinath S,” Automated Kannada Text Summarization using Sentence Features”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019. (Vol-8 Issue-2, July 2019,pp.470-474)

- Neelima Bhatia, Arunima Jaiswal. “Trends in Extractive and Abstractive Techniques in Text Summarization”.(Vol.117 – No. 6, May 2015,pp-21-24)

- Sunitha.C, Dr.A.Jaya, Amal Ganesh, “A Study on Abstractive Summarization Techniques in Indian Languages”, fourth international conference on Recent trends in Computer science and Engineering. (pp.25 – 31)

- Pradeepika Verma and Anshul Verma,” Accountability of NLP Tools in Text Summarization for Indian Languages”. Journal of Scientific Research Institute of Science, Banaras Hindu University, Varanasi, India.(Vol. 64, Issue 1, 2020,pp.358-363).

- Ruby Rani1 & D. K. Lobiyal ,“An extractive text summarization approach using tagged-LDA based topic modelling”. Received: 18 February 2020 /Revised: 26 June 2020 /Accepted: 6 August 2020 Published online :19 september 2020. Springer Science Business Media, LLC, part of Springer Nature 2020.(pp.3275–3305).