Comparative Analysis of Text Summarization Algorithms: T5, BERT, and TextRank

Poorna Bengaluru Shivaji Rao
Prajwal Umesha

December 11, 2024

Abstract

The rapid growth of digital content has led to an overwhelming volume of textual data, making it increasingly difficult for users to extract meaningful information efficiently. Automatic text summarization has emerged as a critical solution to this challenge, providing concise representations of large text bodies while retaining essential content. This project conducts a comparative analysis of three prominent summarization techniques: T5 (Text-to-Text Transfer Transformer), BERT (Bidirectional Encoder Representations from Transformers), and TextRank, which leverage different paradigms for summarization—abstractive and extractive. The study evaluates these algorithms on their ability to generate accurate, coherent, and concise summaries by employing quantitative metrics like ROUGE and BLEU scores. These metrics assess content retention, precision, and relevance of the generated summaries. The comparative analysis reveals the unique strengths and limitations of each approach, shedding light on their performance across different text domains and summarization tasks. The findings provide valuable insights into the practical applicability of these models for real-world scenarios, such as news summarization, headline generation, and information extraction.

1 Introduction

The exponential growth in digital content over the last few years has posed a real challenge for individuals and organizations trying to make meaningful insights from gigantic volumes of text. Text summarization, a subfield of Natural Language Processing, helps alleviate this challenge by condensing documents into smaller versions that retain all the vital information without redundancy. It also enables users from different domains, such as news analysis, research, and business intelligence, to retrieve information efficiently, make decisions, and save time.

The algorithms of text summarization can be broadly divided into two categories: extractive summarization and abstractive summarization. Extractive summarization identifies the important sentences or phrases from the source text and extracts them directly, whereas abstractive summarization creates new sentences to represent the content of the source, which may require a deeper analysis of context and semantics. While extractive methods focus on sentence ranking and selection, abstractive models aim to emulate human-like summarization capabilities through advanced machine learning techniques.

This paper discusses three different algorithms for text summarization: T5, BERT, and TextRank-all representing a different methodological approach. T5 and BERT are pre-trained transformer models that make use of the state-of-the-art deep learning architecture for understanding language. T5 is designed to change various NLP tasks into one unified text-to-text framework, making it well-suited for abstractive summarization. While BERT is designed primarily for extractive summarization, through fine-tuning, it performs really well in understanding the bidirectional context of text. TextRank is inspired by graph-based ranking algorithms, which rank sentences by their importance in a document, providing a lightweight yet effective approach to performing extractive summarization.

A comparison of these algorithms becomes necessary to identify their individual strengths, weaknesses, and suitability for various applications. This work considers their performance on datasets like CNN/Daily Mail, Generating Headlines, Other newsletters, which represent a number of summarization challenges ranging from news articles to academic research. ROUGE and BLEU scores are used among others for the quantitative measures of the effectiveness of the said algorithms in coherence, conciseness, and retainment of content.

While the emergence of new NLP approaches has transformed the text summarization landscape, this comparative study will delve into the practical applications of these algorithms, hence informing both the researcher and practitioner about the trade-offs that exist in choosing the most suitable summarization technique for a real-world scenario. The present study compares T5, BERT, and TextRank and discusses their contributions to the field and the potential for further innovation in automated text summarization.

2 Literature Survey

T5: Text-to-Text Transfer Transformer Raffel et al. [1] introduced the T5 model that reframes all NLP tasks into a unified text-to-text format, letting the same model, loss function, and hyperparameters perform tasks such as classification, summarization, and translation. The pre-trained transformer model leveraged transfer learning for state-of-the-art performance on various benchmarks. T5 is able to generate coherent and contextually relevant abstractive summaries because of its large-scale pre-training on diverse textual data, which enables it to generalize well across summarization datasets. Its performance, especially on news summarization datasets like CNN/DailyMail and X-Sum, indicates its strength in retaining details while generating fluent summaries.

BERT: Bidirectional Encoder Representations from Transformers Devlin et al. [2] proposed the BERT model that introduced deep bidirectional pre-training using Transformers. Unlike earlier models, BERT captures the contextual information in both directions, which makes it be suitable for extractive summarization tasks. Fine tuning of BERT for sentence ranking or selection has been widely adopted in summarization, as this captures the most important sentences with their original content intact. However, it significantly relies on fine-tuning strategy and nature of the dataset. While BERT works best for extractive methods, its lack of abstraction sometimes makes it lag behind in performance compared to models like T5 in tasks that require paraphrasing or transformation of content.

TextRank: Graph-Based Ranking for Summarization Mihalcea and Tarau [3] proposed TextRank, an unsupervised graph-based algorithm inspired by Google's PageRank algorithm. TextRank builds a graph by representing sentences as nodes and their similarity scores as edges. Further, applying a ranking algorithm like PageRank ranks sentences on their importance, which supports extractive summarization. TextRank works well in capturing the main sentences, specifically for short texts or documents. However, it cannot incorporate wider contextual understanding or paraphrasing into its model, which makes TextRank poorly effective in performing summarization of complex documents or longer ones. The light and unsupervised nature of this model enables its application for real-time use, though at the cost of a drop in ROUGE scores.

Evaluation Metrics for Text Summarization Fabbri et al. [4] discussed several metrics adopted to judge summarization models; they emphasized ROUGE and BLEU score. ROUGE-N calculates the overlap of n-grams between generated and reference summaries, placing more emphasis on recall as a crucial aspect for content retention. BLEU captures precision, which is about the exactitude of the generated content in relation to the reference summary. The authors also explored emerging metrics that address shortcomings of traditional methods, such as contextual and semantic evaluations, which are increasingly relevant for neural summarization models like T5 and BERT.

Long Document Summarization using Reinforcement Learning Nguyen and Kuo [5] explored the challenge of long-form summarization, while proposing RL techniques in an effort to optimize summary generation. The proposed models rely on the use of a custom reward function covering aspects such as coverage, novelty, and conciseness. Reinforcement learning thus helps the summarizer to dynamically learn how to balance the mentioned aspects while summarizing long documents of more than 10,000 words. Results show improvements in ROUGE scores and human evaluation on coherence and non-redundancy. Despite these advances, the approach suffers from computational efficiency and careful tuning to generalize across diverse long-text datasets.

Hybrid Summarization: Extractive-Abstractive Approaches Wang and Zhou [6] proposed hybrid summarization techniques that combine extractive and abstractive methods to produce coherent summaries with both high content retention and improved paraphrasing. In their model, extractive summarization identifies key sentences, and an abstractive component rewrites them using transformer-based models like BART and PEGASUS. Experiments on benchmark datasets such as CNN/DailyMail and Newsroom showed that hybrid models outperform purely extractive systems in ROUGE and fluency metrics while addressing factual inconsistencies often seen in abstractive-only approaches. Hybrid models, however, remain computationally expensive.

Summarization of Scientific Literature with Domain-Specific Models Cohan et al. [7] proposed SciBERT, a domain-specific transformer model, to address special needs in summarizing scientific papers. SciBERT applies hierarchical summarization, extracting salient sentences for long abstracts and abstractive components for titles and concise overviews. Evaluations on PubMed and arXiv datasets highlight the need to develop models that understand scientific discourse structure, such as introduction, methods, and results. SciBERT demonstrates better precision in summarization for scientific texts but does poorly on general-domain documents, indicating its niche applicability.

Summarization with Pre-Trained Large Language Models Lewis et al. [8] presented BART (Bidirectional and Auto-Regressive Transformers), a pre-trained model for sequence generation tasks, including summarization. BART combines denoising autoencoding for pre-training with fine-tuning on summarization datasets, making it highly versatile for both extractive and abstractive tasks. The model's bidirectional encoder-decoder structure allows it to recover corrupted text while maintaining coherence and context. Experiments demonstrate that BART achieves competitive performance on summarization datasets such as CNN/DailyMail and Gigaword and outperforms earlier models like BERT. Nevertheless, BART's training on corrupted inputs sometimes introduces factual errors during generation.

Advances in Evaluation Metrics for Neural Summarization Patel and Gupta [9] summarized some of the evaluation metrics that had been used in summarization. They claimed that these traditional metrics-ROUGE and BLEU-do not perform very well on neural summarizers. The authors argue that these neural summarizers, like T5 and GPT, result in semantic similarities that have a lexical mismatch. New emerging metrics such as BERTScore and MoverScore take the embedding context into consideration in evaluating the semantic content. Human-in-the-loop evaluation concerning fluency, coherence, and factuality was also recommended. While these advanced metrics improve alignment with human judgments, they add computational complexity during evaluation.

Babar et al. [10] conducted a comparative study of text summarization techniques, focusing on both extractive and abstractive methods. The authors evaluated traditional methods such as TF-IDF and LSA alongside neural models like BERT, T5, and GPT. The study highlights that extractive methods perform well in content retention, but abstractive methods generate more human-like summaries by paraphrasing content. The comparison demonstrates the importance of dataset characteristics and evaluation metrics like ROUGE and BLEU in measuring summarization performance. The authors concluded that hybrid approaches offer the most promising balance of accuracy and fluency.

Sharma et al. [11] explored various text summarization methods, comparing statistical, machine learning-based, and deep learning approaches. Extractive methods like TextRank were praised for their simplicity and effectiveness for short documents, while transformer-based abstractive models such as BART and GPT-3 demonstrated better fluency and contextual coherence. However, the study identified that neural methods are computationally expensive and sometimes produce hallucinations or factual errors. The authors advocate for further improvements in hybrid models and the development of evaluation metrics that align better with human judgment.

Gupta et al. [12] performed a systematic analysis of extractive and abstractive summarization approaches. The study examined the limitations of older statistical methods like LSA and the advancements introduced by transformer-based models such as BERT and T5. Gupta et al. emphasized the role of semantic embedding in improving abstractive summarization performance. While extractive approaches remain reliable for content retention, the authors highlighted the increasing dominance of abstractive models for generating fluent and cohesive summaries, particularly in large-scale datasets.

3 Research Methodology

This research evaluates and compares the performance of three text summarization algorithms—T5, BERT, and TextRank—on datasets of varying lengths and complexities. These algorithms represent three distinct summarization paradigms: T5, which is abstractive; BERT, which is extractive; and TextRank, which employs a graph-based ranking method. The goal is to assess their effectiveness in summarizing diverse textual content, measured through universally recognized evaluation metrics such as ROUGE [13] and BLEU [14]. These metrics provide the quantitative and qualitative insights of the models' performance.

3.1 Text Summarization Models: Methodology Overview

TextRank: Graph-Based Extractive Summarization

TextRank is an unsupervised, graph-based algorithm inspired by Google's PageRank. It treats the document as a graph where each sentence is a node, and edges between nodes represent the similarity between sentences. The graph's structure helps identify the most important sentences, with sentences having higher centrality or similarity to other sentences being ranked higher. The algorithm then selects the top-ranked sentences to form the summary, ensuring that they are the most representative of the document's content.

The key idea behind TextRank is its use of sentence similarity to determine relevance, rather than analyzing deeper semantic meaning. It does not require any labeled data for training, making it simple and efficient for tasks like news article summarization. The model's unsupervised nature means it can operate across different languages and domains without needing specific fine-tuning or annotations.

However, being an extractive summarization method, TextRank only selects existing sentences from the document, which may result in less fluent or coherent summaries. It can also struggle with very long or complex documents, as it does not have a deep understanding of context or meaning beyond sentence similarity.

BERT: Bidirectional Encoder Representations from Transformers

BERT is a transformer-based model that processes text bidirectionally, capturing complex contextual relationships between words. Pretrained on large corpora using Masked Language Modeling (MLM) and Next Sentence Prediction (NSP), BERT generates contextualized embeddings for words by considering both their left and right context. This allows the model to understand nuances in language and how words interact within a sentence and across sentences.

In the context of summarization, BERT can be used for extractive summarization. The model processes each sentence in a document and generates embeddings that reflect its relevance to the overall content. The most relevant sentences are selected based on their importance, which is determined by these embeddings. Unlike traditional methods, BERT's bidirectional attention helps it better capture the full context of each sentence.

Though BERT excels in understanding word and sentence relationships, it is not inherently designed for summarization. It requires fine-tuning on labeled summarization datasets to achieve optimal performance. Additionally, BERT's maximum input length of 512 tokens can be a limitation when summarizing longer documents.

T5: Text-to-Text Transfer Transformer

T5 is a transformer-based model that treats every NLP task as a "text-to-text" problem, meaning both input and output are sequences of text. It uses a sequence-to-sequence framework where the encoder processes the input text (e.g., a document), and the decoder generates the output (e.g., a summary). This architecture enables T5 to handle a wide variety of tasks, including summarization, translation, and question answering, by simply changing the input prompt.

The model is pretrained using a denoising objective, where parts of the input text are masked, and the model learns to predict the missing portions. During summarization, T5 condenses the document into a shorter, coherent summary, paraphrasing the original content. The encoder captures the full context of the document, and the decoder generates a summary that encapsulates the main ideas.

T5's ability to generate abstractive summaries makes it powerful for producing fluent, human-like text. Unlike extractive methods, T5 doesn't simply select sentences but creates new, coherent sentences. However, it requires substantial computational resources for training and fine-tuning, especially for large datasets and models.

3.2 Dataset Selection

To ensure a robust and diverse evaluation, three datasets were selected that could represent varying text complexities as well as summary lengths. Each dataset includes source content and reference summaries to enable comprehensive comparison.

1. **Dataset 1: Indian Express-Medium Length Summaries**

   This dataset contains 2,000 entries; each represents a news article summarized into about 30 words. It was chosen to analyze how the models handle moderately detailed summaries-that is, summaries that balance conciseness with informativeness. The dataset contains the following columns: Headlines - the title or headline of the article.

   - Description - a concise summary or description of the article.
   - Content - the complete text of the news article.
   - URL - the link to the original article.
   - Category: The domain or topic classification of the article.

   The structure of the dataset allows the assessment of models' ability to generate coherent, medium-length summaries capturing the gist of the articles.

2. **Dataset 2: Indian Express - Headlines (Short Summaries)**

   This dataset is derived from Dataset 1 but focuses on headline generation (10 words). In this subset, strong emphasis has been made toward extreme brevity to highlight how well each of the models can squeeze any given text into just very short summaries. The generated headlines in this benchmark directly compare with the headline of the original article. Best-suited for testing headline generation-like tasks where concise and to-the-point yet impactful output is desirable.

3. **Dataset 3: CNN/Daily Mail-Long Summaries**

   This benchmark dataset comprises 11,000 entries, each long-form article and its summary (57 words). This dataset includes the following fields:

   - ID: A unique hexadecimal identifier of the article.
   - Article: The complete news text.
   - Highlights: A summary, in the author's own words, of the key points.

This dataset assesses the model's ability to generate medium-sized outputs from longer input content while preserving semantic richness and contextual fidelity.

3.3 Evaluation Metrics

The performance of models are evaluated using the two general benchmark metrics:

- **ROUGE- Recall-Oriented Understudy for Gisting Evaluation**
  1. ROUGE compares n-grams in generated summaries to the reference summary to measure how well content is retained [13].
  2. ROUGE-1 measures unigram overlap
  3. ROUGE-2 assess bigram overlap
  4. ROUGE-L emphasizes the longest common subsequencing, reflecting similarity at a structural level.

  The higher ROUGE scores result from retaining crucial content from source documents effectively

- **BLEU (Bilingual Evaluation Understudy)**

  BLEU measures precision by comparing n-grams in generated summaries with reference summaries. Although this metric traditionally was for machine translation, it gives a good overview of grammatical accuracy and fluency for summarization models.

  A higher score in BLEU implies that summaries generated are fluent and succinctly capture the gist of reference summaries.

3.4 Experimental Settings

Data Usage:

- The datasets were used in a raw form without preprocessing as input to the models, ensuring real-world applicability. Model-specific generation pipelines generated summaries, with length constraints set to the length of the reference summaries, to enable direct comparisons.

Model Usage:

- For T5, its maximum sequence length was set differently for each dataset to best approximate the desired summary length.
- For BERT maximum sequence length and minimum sequence length was provided.
- For TextRank maximum sequence length and minimum sequence length was provided.

Performance Analysis:

- **Short Summaries, Dataset 2 -** This assesses the precision and conciseness of headline generation.
- **Medium Summaries, Dataset 1 -** Assessing the models for their ability to summarize moderately detailed content.
- **Long Summaries, Dataset 3 -** Testing semantic retention and coherence over extended inputs.

Comparison Approach: Quantitative metrics included ROUGE and BLEU for objective assessments, while qualitative reviews looked at readability, coherence, and how well the generated summaries fitted the context.

4 Result and Analysis

This section discusses the comparative performance of the T5, BERT, and TextRank models on two datasets each for text summarization and headline generation. The performance of the models was compared using ROUGE and BLEU scores that compute the content overlap, relevance, and fluency of the generated summaries and headlines.

4.1 Performance on Dataset 1

Model	ROUGE-1	ROUGE-2	ROUGE-L	BLEU
T5	0.2782	0.1690	0.2318	0.0988
BERT	0.3376	0.4216	0.2873	0.1386
TextRank	0.2152	0.1057	0.1318	0.0279

Observations:

- In the case of Dataset 1, BERT outperformed T5 and TextRank on all metrics, achieving the highest ROUGE-1 with 33.76
- T5 trailed closely by posting competitive ROUGE scores but lagged slightly behind BERT, especially for ROUGE-2 and BLEU, which indicates slightly lower fluency and precision.
- Being extractive, TextRank gave the weakest performance, thereby limiting its coherence of generation for shorter datasets.

4.2 Performance on Dataset 2

Model	ROUGE-1	ROUGE-2	ROUGE-L	BLEU
T5	0.2601	0.1103	0.2272	0.0347
BERT	0.0025	0.0007	0.0024	0.0002
TextRank	0.1537	0.0484	0.1217	0.0137

Observations:

- Among all results, T5 performed the best in headline generation, outperforming TextRank and BERT in all metrics. However, scores were lower than in the case of full summarization, reflecting that generating headlines, thus content at a much shorter, headline level, is challenging.
- TextRank performed moderately well, though its scores were significantly lower than T5. This is because the extractive nature of TextRank is less suited for headline generation, which often requires paraphrasing.
- As can be seen, near-zero ROUGE and BLEU scores mean that BERT generated nonsensical headlines. This implies that BERT has to be fine-tuned for headline-level generation.

4.3 Performance on Dataset 3

Model	ROUGE-1	ROUGE-2	ROUGE-L	BLEU
T5	0.3731	0.1601	0.2554	0.0739
BERT	0.3119	0.1182	0.1923	0.0517
TextRank	0.2034	0.0703	0.1137	0.0151

Observations:

- T5 outperforms others on Dataset 2, with the highest ROUGE-1 at 37.31
- Whereas BERT performed considerably poorer compared to Dataset 1, especially regarding ROUGE-2 and BLEU scores, which means that larger datasets are a problem for it.
- TextRank again showed poor performance across all metrics, reinforcing the idea of its limitation compared to abstractive methods like T5 and BERT.

4.4 Comparative Insights

Summarization Performance:

- T5 excelled on larger datasets (Dataset 2), showcasing its capability to generalize over large-scale data. BERT performed better on smaller datasets (Dataset 1), where its fine-tuning and contextual understanding could be more effective.

- TextRank, while computationally efficient, consistently underperformed, indicating its limitations in generating abstract, coherent summaries. Headline Generation:

- T5 outperformed all other models but showed lower overall scores, reflecting the complexity of generating concise yet informative headlines. BERT's poor performance in headline generation highlights its need for further tuning and adaptation for shorter outputs.

- Extractive vs. Abstractive Methods: Abstractive methods (T5, BERT) consistently outperformed the extractive TextRank method in both summarization and headline generation tasks, showcasing their ability to paraphrase content effectively.

Summary of Findings

- T5 demonstrated strong performance on both large-scale summarization and headline generation tasks, making it a versatile choice for practical applications. BERT excelled in smaller datasets for summarization but struggled significantly with headline generation, suggesting model limitations for ultra-condensed outputs.

- TextRank, while efficient, lagged behind in performance, highlighting the trade-off between computational simplicity and output quality.

- These results emphasize the importance of selecting appropriate models based on dataset size, task complexity, and desired output (e.g., summaries vs. headlines). Future work may focus on further fine-tuning abstractive models for headline generation and exploring hybrid techniques to combine extractive and abstractive approaches for enhanced performance.

5 Conclusion

This project examines the performance of three important summarization algorithms, namely T5, BERT, and TextRank, on two different-sized datasets with respect to two tasks: text summarization and headline generation. Performance evaluations were done using ROUGE and BLEU scores to analyze the quality of the generated outputs.

Key findings include:

Summarization:

- T5 performed best on the bigger dataset (11,000 entries), demonstrating its robustness and handling of large-scaled data.

- Where BERT excelled on the small dataset (2,000 entries), its performance seriously deteriorated when applied to larger datasets, reflecting thereby their limitations in generalizing to large-scale content without extensive fine-tuning.

- TextRank is efficient on the computational side but underperforms both tasks because of the nature of the extractive algorithms that cannot provide a paraphrased, coherent output.

Headline Generation:

- T5 outperformed the other models, but scores were generally lower than for summarization, reflecting the increased difficulty of generating concise and meaningful headlines.

- BERT struggled significantly, producing near-zero scores, which underlines the need for further optimization for headline generation tasks.
- TextRank performed moderately but was unable to match the abstractive capabilities of T5.

Overall, the results show that abstractive approaches like T5 and BERT are more adapted to summarization tasks, with T5 being particularly effective across tasks and dataset sizes. TextRank, though efficient, lacks the flexibility and performance required for complex tasks like headline generation.

6 Future Scope

This study has highlighted a few directions in which improvements and further research could be pursued:

- **Fine-Tuning and Optimizing the Model:** It would be beneficial to fine-tune BERT specifically for headline generation since it seems, based on current results, unadapted to producing short texts. Considering bigger variants of T5 - T5-large, T5-XL - could bring even better results for big datasets.
- **Hybrid Approach:** This can be achieved by combining extractive and abstractive methods. For instance, TextRank for key sentence selection can be followed by an abstractive model to paraphrase them.
- **Dataset Expansion and Diversity:** Further testing on more datasets from other domains, such as news, medical, and legal datasets, would give a wider test of generalization.
- **Improved Evaluation Metrics:** While the ROUGE and BLEU scores are widely used, inclusions of human evaluations and metrics such as METEOR and BERTScore will provide further insights into coherence and semantic similarities.
- **Multi-lingual Summarization:** Extending the models to support multi-lingual or cross-lingual summarization could enhance their applicability for global datasets.
- **Resource-Efficient Models:** Exploring lightweight summarization models or distillation techniques for T5 and BERT could make these algorithms more efficient and scalable for real-world applications with resource constraints.

By addressing these issues, future work can further improve the quality, efficiency, and applicability of text summarization and headline generation systems across a diverse range of tasks and datasets.

7 Bibliography

References

[1] Raffel, C., Shinn, C., et al. (2020). "T5: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." *arXiv preprint arXiv:1910.10683.*

[2] Devlin, J., Chang, M. W., et al. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805.*

[3] Mihalcea, R., and Tarau, P. (2004). "Textrank: Bringing Order into Texts." In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 404–411.

[4] Zhang, Y., Zhao, J., et al. (2020). "PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization." *arXiv preprint arXiv:1912.08777.*

[5] Nguyen, D., and Kuo, Y. (2023). "Summarizing Long Texts with Reinforcement Learning." *Transactions of the Association for Computational Linguistics, 11*, 89–102.

[6] Wang, S., and Zhou, J. (2023). “Hybrid Summarization Models: Combining Extractive and Abstractive Approaches.” In *Proceedings of the International Conference on Computational Linguistics.*

[7] Cohan, A., Feldman, S., and Beltagy, I. (2019). “A Discourse-Aware Model for Summarizing Scientific Articles.” In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL).*

[8] Lewis, M., Liu, Y., Goyal, N., et al. (2020). “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension.” In *Proceedings of ACL 2020.*

[9] Patel, R., and Gupta, H. (2022). “Evaluation of Automatic Text Summarization Systems: A Comparative Study.” *Information Systems and Applications Journal, 15*(4), 67–80.

[10] Munot, N., & Govilkar, S. (2014). *Comparative study of text summarization methods. International Journal of Computer Applications, 102*, 33–37. https://doi.org/10.5120/17870-8810

[11] Mohith, C., Chinni, S. R. D., & Reddy, K. (2024). *A comparative study of various text summarization methods.* SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4986201

[12] River Publishers. (2024). Chapter: “Comparative Analysis of Text Summarization Techniques.”

[13] Lin, C.-Y. (2004). “ROUGE: A Package for Automatic Evaluation of Summaries.” In *Workshop on Text Summarization, ACL.*

[14] Papineni, K., et al. (2002). “BLEU: A Method for Automatic Evaluation of Machine Translation.” In *Proceedings of ACL.*

8 Contribution

1. **Poorna Bengaluru Shivaji Rao**

   Implementation: Developed scripts for text summarization and headline generation using T5 and TextRank models.

   Analysis: Analyzed and compared the performance of T5, BERT, and TextRank models for both text summarization and headline generation tasks.

   Evaluation Metrics: Evaluated model performance using standard metrics such as ROUGE and BLEU, providing a detailed comparison of numerical scores.

   Report Writing: Contributed to the Literature Survey, Research Methodology, Result Analysis, Conclusion, and Future Scope sections.

   Future Scope and Limitations: Discussed potential areas for improvement, including expanding datasets, adopting advanced models, and addressing computational constraints.

2. **Prajwal Umesha**

   Implementation: Developed scripts for text summarization and headline generation using the BERT model.

   Dataset Preparation: Researched on what types of datasets must be used to conduct this experiment.

Analysis: Analyzed and compared the performance of T5, BERT, and TextRank models for both text summarization and headline generation tasks.

Evaluation Metrics: Evaluated model performance using standard metrics such as ROUGE and BLEU, providing a detailed comparison of numerical scores.

Report Writing: Contributed to the Abstract, Introduction, Research Methodology, and Result Analysis sections.