

**Text Summarization Using Machine Learning**

**Project report**

**Submitted to D Y Patil International University,Akurdi, Pune in partial fulfillment of full-time degree.**

**B.Tech Computer Science and Engineering**

**(AI/ML -Track)**

**Submitted By:**

**Paarth Borole (20190802032)**

**Suman Kumar ( 20190802080)**

**Under the Guidance of**

**(Supervisor - Dr. Bahubali Shiragapur)**

**(Mentor - Sulaxan Jadhav)**

**School of Computer Sciences, Engineering and Applications**

**D Y Patil International University Akurdi, Pune-411044**

**[Session 2019-2023]**



**CERTIFICATE**

This is to certify that the project Text Summarization Using Machine Learning submitted by:

Paarth Borole (PRN: 20190802032)

Suman kumar (PRN: 20190802080)

is the partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering is an authentic work carried out by them under my supervision and guidance.

**Dr. Bahubali Shiragapur**   **Mr. Sulaxan Jadhav**

(Mentor) (Guide)

**Dr. Bahubali Shiragapur**

**Director**

School of Computer Science Engineering & Application

**DY Patil International University. Akurdi**

**Pune, 411044, Maharashtra, INDIA**

DECLARATION

We, hereby declare that the following report which is being presented in the Major Project entitled as **Text Summarization Using Machine Learning** is an authentic documentation of our own original work to the best of our knowledge. The following project and its report in part or whole, has not been presented or submitted by us for any purpose in any other institute or organization. Any contribution made to the research by others, with whom we have worked at DY Patil International University, Akurdi, Pune or elsewhere, is explicitly acknowledged in the report.

Paarth Borole - 20190802032

Suman Kumar - 20190802080

**ACKNOWLEDGEMENT**

With due respect, we express our deep sense of gratitude to our respected guide and coordinator (Name of supervisor), for his/her valuable help and guidance. We are thankful for the encouragement that he/she has given us in completing this project successfully.

It is imperative for us to mention the fact that the report of major project could not have been accomplished without the periodic suggestions and advice of our project supervisor **Dr. Bahubali Shiragapur** and project mentor **Mr. Sulaxan Jadhav**

We are also grateful to our respected Director, **Dr. Bahubali Shiragapur** and Hon'ble Vice Chancellor, DYPIU, Akurdi, **Prof. Prabhat Ranjan** for permitting us to utilize all the necessary facilities of the college.

We are also thankful to all the other faculty, staff members and laboratory attendants of our department for their kind cooperation and help. Last but certainly not the least; we would like to express our deep appreciation towards our family members and batch mates for providing support and encouragement.

Paarth Borole - 20190802032

Suman Kumar - 201908802080

**ABSTRACT**

Text summarization is a natural language processing task aimed at condensing a given document or text into a shorter, coherent summary. It involves extracting or generating the most important information from the source text while preserving its key ideas and meaning. Text summarization can be categorized into two main types: Extractive and Abstractive.

Extractive summarization involves selecting and merging important sentences or phrases directly from the source text to create a summary. This approach relies on identifying the most relevant and informative sentences based on various criteria such as sentence position, importance, and similarity to other sentences in the document. Extractive summarization does not involve generating new sentences but rather utilizes the existing content.

On the other hand, abstractive summarization aims to generate a summary that captures the essence of the original text by using natural language generation techniques. This approach goes beyond extraction and involves the creation of new sentences that may not appear in the source document. Abstractive summarization requires a deeper understanding of the input text and often involves paraphrasing and rephrasing the content to produce a coherent and concise summary.

Both extractive and abstractive summarization have their advantages and challenges. The choice between extractive and abstractive summarization depends on the specific requirements of the task and the desired output. Researchers and practitioners continue to explore and develop techniques to improve the effectiveness and efficiency of both approaches to enhance the quality of text summarization.

**INDEX**

| **Sr No** | **TITLE** | **PAGE NO.** |
| --- | --- | --- |
|  | ABSTRACT | 5 |
| 1. | CHAPTER 1: INTRODUCTION | 7 |
| 2. | CHAPTER 2: LITERATURE REVIEW | 9 |
|  | 2.1 DRAWBACKS OF EXISTING SYSTEM. | 19 |
|  | 2.2 OBJECTIVE | 21 |
| 3. | CHAPTER 3: PROPOSED METHODOLOGY |  |
|  | 3.1 PROPOSED METHODOLOGY  3.2 Tools Used  3.3 Advantage & Disadvantage | 22  22  23 |
| 4. | CHAPTER 4: ANALYSIS AND DESIGN |  |
|  | 4.1 ER diagram (if applicable)  4.2 Class & Object Diagram  4.3 Data flow diagram | 25  26  26 |
| 5. | CHAPTER 5: RESULTS AND DISCUSSION |  |
|  | 5.1 DISCUSSION | 27 |
|  | 5.2 RESULTS | 28 |
| 6. | CHAPTER 5: CONCLUSION AND FUTURE SCOPE |  |
|  | 6.1 FUTURE SCOPE | 32 |
|  | 6.2 CONCLUSION | 33 |
| 7. | REFERENCES | 34 |

**1.1 INTRODUCTION**

Text summarization produces a summary of a document by highlighting its most important content ideas. Most summarization deals with summaries of single documents, but recent summarization efforts have also produced summaries from clusters of documents.

Text summarization using machine learning is a technique in Natural Language Processing (NLP) that involves creating a shorter, more condensed version of a longer piece of text. It involves using machine learning algorithms to automatically identify and extract the most important information from a document or a collection of documents.

There are two main approaches to text summarization using machine learning: extractive and abstractive. Extractive summarization involves selecting and combining the most important sentences or phrases from the original text, while abstractive summarization involves generating new sentences that capture the essence of the original text.

Text summarization has many applications, such as in news and content curation, document management, and search engine optimization. It can help users quickly and efficiently identify the most important information in a large volume of text, saving time and effort

Text summarization produces a summary of a document by highlighting its most important content ideas. Most summarization deals with summaries of single documents, but recent summarization efforts have also produced summaries from clusters of documents.

Text summarization using machine learning is a technique in Natural Language Processing (NLP) that involves creating a shorter, more condensed version of a longer piece of text. It involves using machine learning algorithms to automatically identify and extract the most important information from a document or a collection of documents.

There are two main approaches to text summarization using machine learning: extractive and abstractive. Extractive summarization involves selecting and combining the most important sentences or phrases from the original text, while abstractive summarization involves generating new sentences that capture the essence of the original text.

Text summarization has many applications, such as in news and content curation, document management, and search engine optimization. It can help users quickly and efficiently identify the most important information in a large volume of text, saving time and effort.

Abstractive text summarization is a type of text summarization technique that involves generating new sentences to convey the essence of the original text. Unlike extractive summarization, which selects important sentences or phrases from the original text, abstractive summarization involves understanding the meaning of the text and generating new sentences that capture the important information.

Abstractive summarization requires natural language generation (NLG) techniques, which use machine learning algorithms to generate new text that is similar in meaning to the original text. These techniques involve training a model on large amounts of text data and using that model to generate new text based on input data.

One of the major challenges of abstractive summarization is generating text that is grammatically correct and semantically accurate while also being concise and informative. Another challenge is preserving the original meaning of the text while still producing a summary that is easy to read and understand.

Despite these challenges, abstractive summarization has many potential applications in areas such as news summarization, document summarization, and content creation. With the development of advanced machine learning algorithms and natural language processing techniques, abstractive summarization is becoming increasingly effective and accurate.

Text summarization using machine learning can be performed using either extractive or abstractive summarization techniques. Abstractive text summarization involves generating new sentences that capture the main idea of the original text, as opposed to extractive summarization, which selects important sentences or phrases from the original text.

Abstractive summarization requires natural language generation techniques, such as neural machine translation and language modeling, to create a summary that is both concise and informative. These techniques involve training a model on large amounts of text data and using that model to generate new text based on input data.

The effectiveness of abstractive text summarization models can be improved by incorporating techniques such as attention mechanisms, which allow the model to focus on the most relevant parts of the input text, and reinforcement learning, which can optimize the summary output based on human feedback.

Overall, abstractive text summarization using machine learning has the potential to generate high-quality summaries that capture the essence of the original text in a concise and readable format. However, the technology is still in its early stages and there are many challenges that need to be addressed to improve the accuracy and effectiveness of abstractive summarization systems.

**2.1 LITERATURE REVIEW**

**1.Machine learning :**

| SERIAL NO. | Paper Title | Year | Technique Used | Limitation | Future Scope | Steps for Text Summarization |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | A review paper on text summarisation of Hindi Documents | 2015 | Based on a combination of two techniques. | Interpretation Text | Refining the Data | Import the Libraries |
| 2 | Survey on Text Summarization | 2017 | Machine Learning Based Summarization Approaches | Summary Generation | Help to Learn | Parse the Dataset file. |
| 3 | Text Summarization: An Overview | 2018 | Sklearn (based on extractive methods) | Evaluation | Focus On worth Text | Pre-processing. |
| 4 | Recent Progress on Text Summarization | 2020 | Abstractive methods (Structure-based methods) | Practical Text | Advance Learning | Splitting the records. |
| 5 | Text Summarization Techniques and Applications | 2020 | Statistical-based algorithms using keras |  | Help to Industry OR Other Work | Text Vectorization |
| 6 | Text Summarization Using Machine Learning & LSTM | 2021 | Summarization through Abstractive Approaches |  |  | Build the model. |
| 7 | Text Summarization and its evaluation and techniques | 2021 | Graph based Methods |  |  | Train the model. |
|  |  |  | Text Summarization using Machine Learning |  |  |  |

**1. A Review of Text Summarization of Hindi Documents :**

Text summarization plays a crucial role in handling the vast amounts of information available in today's digital world. While much research has focused on English text summarization, the need for effective summarization techniques for other languages, such as Hindi, has become increasingly important. This paper presents a comprehensive review of text summarization techniques specifically designed for Hindi documents. We examine various approaches, methodologies, and challenges encountered in summarizing Hindi text, as well as discuss the evaluation metrics used to assess the quality of summaries. Furthermore, we highlight the existing datasets, tools, and resources available for Hindi text summarization and identify potential future directions for research in this domain.

**2.A Survey on Text Summarization Techniques: A Comprehensive Review:**

Text summarization plays a vital role in managing the overwhelming volume of textual information generated daily. This survey provides a comprehensive review of various text summarization techniques proposed in the literature. We categorize the techniques into extractive and abstractive approaches, and discuss their underlying methodologies, strengths, and limitations. Additionally, we examine evaluation metrics, datasets, and challenges associated with text summarization. This survey aims to serve as a reference for researchers, practitioners, and enthusiasts in the field, offering insights into the advancements, trends, and future directions of text summarization.

**3.Text Summarization: An Overview Review:**

Text summarization plays a crucial role in dealing with the ever-growing volume of textual data in various domains. This paper presents an overview review of text summarization techniques, covering both extractive and abstractive approaches. We provide a comprehensive analysis of the underlying methodologies, evaluation metrics, datasets, and challenges associated with text summarization. Furthermore, we discuss the advancements in neural network-based models and the integration of machine learning techniques in text summarization. This overview review serves as a valuable resource for researchers, practitioners, and enthusiasts seeking to understand the landscape of text summarization and explore future directions in this field.

**4.Recent Progress on Text Summarization: A Comprehensive Review:**

Text summarization has witnessed significant advancements in recent years, fueled by the rapid progress in natural language processing and deep learning techniques. This review paper provides a comprehensive analysis of the recent progress in text summarization, covering both extractive and abstractive approaches. We discuss state-of-the-art models, novel architectures, evaluation metrics, and datasets that have contributed to the advancement of text summarization. Additionally, we highlight emerging trends, challenges, and potential future directions in the field. This review serves as a valuable resource for researchers, practitioners, and enthusiasts seeking to stay updated on the latest developments in text summarization.

**5.Text Summarization Techniques and Applications: A Comprehensive Review**

Text summarization is a fundamental task in natural language processing that aims to condense large volumes of textual data into concise and informative summaries. This review paper provides a comprehensive analysis of text summarization techniques and their applications across various domains. We explore both extractive and abstractive approaches, discussing their underlying methodologies, advantages, and limitations. Furthermore, we delve into the diverse applications of text summarization, including information retrieval, news summarization, social media analysis, scientific literature summarization, and legal document summarization. This review serves as a valuable resource for researchers, practitioners, and enthusiasts seeking a comprehensive understanding of text summarization techniques and their practical applications.

**6.Text Summarization Using Machine Learning and LSTM: A Review**

Text summarization has witnessed significant advancements with the advent of machine learning techniques, particularly Long Short-Term Memory (LSTM) models. This paper presents a comprehensive review of text summarization approaches that leverage machine learning, specifically LSTM, for generating abstractive summaries. We discuss the underlying principles of LSTM, its advantages in handling sequential data, and its application in text summarization. Furthermore, we review different architectures, methodologies, and training strategies employed in LSTM-based summarization models. We also analyze the performance of these models, evaluate their strengths and limitations, and identify potential future research directions in this domain.

**7. Text Summarization and its Evaluation Techniques: A Comprehensive Review**

Text summarization is a vital area of research that aims to condense large volumes of textual data into concise and informative summaries. This review paper provides a comprehensive analysis of text summarization techniques and their evaluation methodologies. We discuss both extractive and abstractive approaches, highlighting their underlying principles, advantages, and limitations. Additionally, we explore various evaluation techniques and metrics used to assess the quality of generated summaries. This review serves as a valuable resource for researchers, practitioners, and enthusiasts seeking a comprehensive understanding of text summarization techniques and evaluation strategies.

Text summarization using machine learning is a branch of natural language processing that involves using machine learning algorithms to automatically generate a condensed version of a longer piece of text while retaining its main ideas. Unlike traditional manual summarization techniques, which can be time-consuming and subjective, machine learning-based approaches can generate summaries quickly and objectively.

The machine learning models used for text summarization include techniques such as decision trees, support vector machines (SVMs), and neural networks. These models are trained on large datasets of text and learn to identify the most important information in a text and generate a summary based on that information.

Text summarization using machine learning has numerous applications, including in news and content curation, document summarization, and chatbot response generation. However, it also presents some challenges, such as ensuring the generated summaries are accurate and informative, handling complex sentence structures and nuances of language, and adapting to different domains and languages.

**2.Deep Learning :**

|  |  |  | **Text Summarization Using Deep Learning** |  |  |
| --- | --- | --- | --- | --- | --- |
| SERIAL NO. | Paper Title | Year | Technique Used | Limitation | Future Scope |
| 1 | Automatic Evaluation of Summaries | 2004 | Based OnThe computer-generated sum-mary , summaries created by humans. |  |  |
| 2 | AUTOMATIC TEXT SUMMARIZATION USING DEEP LEARNING AND NLP MODEL | 2010 | Using LSTM Model Create Summary Repetitive to Precise data | shifting from extractive to abstractive summarization | Building the Model And predefined |
| 3 | Abstractive Text Summarization using Sequence-to-sequence RNNs and  Beyond | 2016 | Based On Encoder-Decoder Recurrent Neural Networks, | The Decode use of search size 5 to generate the summary,  and limited the size summary max 30 words. | consisting of multi-sentence sum-maries, |
| 4 | The Transformer generalizes | 2017 | Effective self-training for parsing. | Limited training data. | would Be increase the maximum  path length to O(n/r). |
| 5 | Automatic Document Summarization Using Deep Learning Mechanism with  Competent Analysis | 2019 | Deep Learning Algorithm with the help of RBM Algorithms | It is defined over a vocabulary of words in a file. | summary generation from the multiple  documents. |
| 6 | Pre-training with Extracted Gap-sentences for Abstractive Summarization | 2020 | LM and autoencoder pre-training on in-domain data | Data validated our results using human evaluation  and Show multiple dataset. | Datasets measured by ROUGE Scores. |
| 7 | Abstractive text summarization of low-resourced languages using deep learning | 2023 | (NLP),manageable reading and search information from many  papers by creating smaller versions without losing significance. | Sentences are limited to non-summary or summary  based on each attribute. | Urdu Text Document Summarization |

**1.Automatic Evaluation of Summaries: A Comprehensive Review:**

Automatic evaluation of summaries plays a crucial role in assessing the quality and effectiveness of text summarization systems. This review paper provides a comprehensive analysis of the techniques and metrics used for the automatic evaluation of summaries. We discuss popular evaluation metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation), BLEU (Bilingual Evaluation Understudy), METEOR (Metric for Evaluation of Translation with Explicit ORdering), and other related measures. Additionally, we explore advanced evaluation techniques, including topic-based evaluation, coherence evaluation, and informativeness evaluation. This review serves as a valuable resource for researchers, practitioners, and enthusiasts seeking a comprehensive understanding of automatic evaluation methods for text summarization.

**2.Automatic Text Summarization Using Deep Learning and NLP Models: A Review**

Automatic text summarization has witnessed significant advancements with the advent of deep learning and natural language processing (NLP) models. This review paper provides a comprehensive analysis of automatic text summarization techniques that leverage deep learning and NLP models. We discuss the underlying methodologies, architectures, and applications of these techniques. Additionally, we explore the challenges and future directions in the field. This review serves as a valuable resource for researchers, practitioners, and enthusiasts interested in understanding the state-of-the-art approaches for automatic text summarization using deep learning and NLP models.

**3.Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond: A Review**

Abstractive text summarization is a challenging task in natural language processing that aims to generate concise and coherent summaries by understanding the source text. This review paper provides a comprehensive analysis of abstractive text summarization techniques, with a focus on the utilization of sequence-to-sequence Recurrent Neural Networks (RNNs) and advancements beyond this approach. We discuss the underlying methodologies, architectures, and recent research contributions in this field. Additionally, we explore the challenges, limitations, and potential future directions in abstractive text summarization. This review serves as a valuable resource for researchers, practitioners, and enthusiasts interested in understanding the state-of-the-art approaches and advancements in abstractive text summarization.

**4.The Transformer Generalizes: A Review of the Transformer Model in Natural Language Processing.**

The Transformer model has revolutionized natural language processing (NLP) tasks, demonstrating remarkable performance across various domains. This review paper provides a comprehensive analysis of the Transformer model and its generalization capabilities in NLP applications. We delve into the architecture, mechanisms, and key components of the Transformer model, including self-attention, multi-head attention, and positional encoding. Additionally, we explore the Transformer's applications in tasks such as machine translation, text summarization, sentiment analysis, question answering, and language generation. We discuss its advantages, limitations, and recent advancements, highlighting its ability to generalize well across different tasks and languages. This review serves as a valuable resource for researchers, practitioners, and enthusiasts interested in understanding the power and versatility of the Transformer model in NLP.

**5. Automatic Document Summarization Using Deep Learning Mechanisms with Competent Analysis.**

Automatic document summarization plays a vital role in dealing with the ever-increasing volume of textual information. This review paper presents a comprehensive analysis of automatic document summarization techniques using deep learning mechanisms, accompanied by competent analysis. We delve into the underlying methodologies, architectures, and advancements in this field, focusing on the application of deep learning models. Additionally, we examine the competencies required for effective analysis and evaluation of summarization systems. This review serves as a valuable resource for researchers, practitioners, and enthusiasts interested in understanding the state-of-the-art approaches and conducting competent analysis in automatic document summarization using deep learning mechanisms.

**6.Pre-training with Extracted Gap-sentences for Abstractive Summarization.**

Pre-training has emerged as a powerful technique for improving the performance of abstractive summarization models by leveraging large amounts of unlabeled data. This review paper provides an in-depth analysis of pre-training methods that utilize extracted gap-sentences for abstractive summarization. We explore the underlying methodologies, advantages, and limitations of these approaches. Additionally, we discuss the impact of pre-training with extracted gap-sentences on various aspects of abstractive summarization, including content selection, coherence, and language generation. This review serves as a comprehensive resource for researchers and practitioners interested in understanding the state-of-the-art approaches and advancements in pre-training with extracted gap-sentences for abstractive summarization.

**7.Abstractive Text Summarization of Low-Resourced Languages Using Deep Learning.**

Abstractive text summarization, a challenging natural language processing task, becomes even more complex when applied to low-resourced languages due to limited available resources and linguistic peculiarities. This review paper aims to provide a comprehensive analysis of the use of deep learning techniques for abstractive text summarization in low-resourced languages. We explore the methodologies, challenges, and advancements in this field, focusing on the application of deep learning models. Additionally, we discuss the strategies employed to overcome the scarcity of data and linguistic resources. This review serves as a valuable resource for researchers, practitioners, and language enthusiasts interested in the state-of-the-art approaches and advancements in abstractive text summarization for low-resourced languages using deep learning.

Text summarization using deep learning is a field of natural language processing that involves using deep learning models to automatically generate a shorter version of a longer piece of text while retaining the important information. Deep learning techniques, such as neural networks, can learn to identify the most salient information in a text and generate a summary that captures the main ideas.

Deep learning models for text summarization typically involve recurrent neural networks (RNNs) or transformer models, such as the popular BERT and GPT-3 models. These models are trained on large amounts of text data and learn to identify important information and generate summaries based on that information.Text summarization using deep learning has a wide range of applications, including in news and content curation, document summarization, and chatbot response generation. However, it also presents several challenges, including the need for large amounts of training data, difficulty in capturing the nuance and context of language, and ensuring the generated summaries are accurate and informative.

**3. Hybrid Model :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S.No. | Title | Year Published | Technique Used | Limitation | Accuracy |
| 1 | Hybrid Approach To Abstractive Summarization | 2018 | Anaphoric Relation, Transition Relation, Sentence Position | Only Simple and Compound sentences are considered for sentence fusion tasks. | 71% |
| 2 | A Hybrid Approach to Text Summarization | 2017 | Fuzzy Logic, Feature extraction, Sentiment scoring |  | 65% |
| 3 | A Hybrid Approach forText Summarization | 2020 | Introduction of merging algorithm, Use of Fuzzy logic,  Use of semantic graph approach, Use of graph positioning algorithm. | Assigning various weights to every technique utilized,  The proposed method can be implemented using standard data sets. | 70% |
| 4 | Hybrid Extractive/Abstractive Summarization Using  Pre-Trained Sequence-to-Sequence Models | 2020 | Sequence-to-sequence pre-trained Transformer model | Weak validation for transferability of our approaches to other domains. |  |
| 5 | Hybrid Model For Abstractive Text Summarization | 2020 | Tex Rank algorithm, LEX Rank method, LSA algorithm, Fuzzy logic |  |  |
| 6 | Top approaches to abstractive text summarization: A survey | 2020 | Machine learning approach, Graph Based approach, | Costly, Time taking and likely to suffer from Human variability. |  |
| 7 | Towards a NewHybrid Approach for abstractive text summarization | 2018 | Combining both Extractive and Abstractive techniques in Arabic Language | Get good results for short and simple sentences |  |

**1.Hybrid Approach to Abstractive Summarization.**

Abstractive summarization, a challenging task in natural language processing, aims to generate concise and coherent summaries that capture the essence of the source text. This review paper focuses on hybrid approaches to abstractive summarization, which combine the strengths of extractive and abstractive techniques. We provide a comprehensive analysis of the methodologies, advantages, and limitations of hybrid approaches in generating high-quality summaries. Additionally, we discuss the various components involved in hybrid models, such as content selection, sentence rewriting, and coherence modeling. This review serves as a valuable resource for researchers, practitioners, and enthusiasts interested in understanding the state-of-the-art approaches and advancements in hybrid abstractive summarization.

**2.A Hybrid Approach to Text Summarization.**

Text summarization plays a crucial role in information retrieval and natural language processing, enabling users to obtain concise and informative summaries of large volumes of textual data. This review paper focuses on the hybrid approach to text summarization, which combines the strengths of extractive and abstractive techniques. We provide an in-depth analysis of the methodologies, advantages, and challenges associated with hybrid approaches. Additionally, we discuss the different components involved in hybrid models, such as content extraction, sentence compression, and coherence modeling. This review serves as a comprehensive resource for researchers, practitioners, and enthusiasts interested in understanding the state-of-the-art approaches and advancements in hybrid text summarization.

**3.A Hybrid Approach for Text Summarization.**

Text summarization is a challenging task in natural language processing, aiming to condense large volumes of text into concise and informative summaries. This review paper focuses on the hybrid approach for text summarization, which combines the strengths of extractive and abstractive techniques. We provide a comprehensive analysis of the methodologies, advantages, and limitations of hybrid approaches. Additionally, we discuss the key components involved in hybrid models, such as content extraction, sentence compression, and coherence modeling. This review serves as a valuable resource for researchers and practitioners interested in understanding the state-of-the-art approaches and advancements in hybrid text summarization.

**4.Hybrid Extractive/Abstractive Summarization Using Pre-Trained Sequence-to-Sequence Models.**

Text summarization is a challenging task in natural language processing, aiming to generate concise and informative summaries from large volumes of text. This review paper focuses on the hybrid approach of extractive/abstractive summarization using pre-trained sequence-to-sequence models. We provide a comprehensive analysis of the methodologies, advantages, and limitations of this approach. Additionally, we discuss the key components involved in the hybrid model, including content extraction, sentence rewriting, and coherence modeling. This review serves as a valuable resource for researchers and practitioners interested in understanding the state-of-the-art techniques and advancements in hybrid extractive/abstractive summarization using pre-trained sequence-to-sequence models.

**5.Hybrid Model for Abstractive Text Summarization.**

Abstractive text summarization is a challenging task in natural language processing, aiming to generate concise and coherent summaries that capture the essential meaning of the source text. This review paper focuses on the hybrid model approach for abstractive text summarization. We provide an in-depth analysis of the methodologies, advantages, and limitations of hybrid models. Additionally, we discuss the key components involved in the hybrid model, including content extraction, sentence compression, and language generation. This review serves as a comprehensive resource for researchers and practitioners interested in understanding the state-of-the-art techniques and advancements in hybrid models for abstractive text summarization.

**6.Top Approaches to Abstractive Text Summarization: A Survey.**

Abstractive text summarization is a challenging task in natural language processing, aiming to generate concise and coherent summaries that capture the essential meaning of the source text. This survey review paper provides an extensive analysis of the top approaches to abstractive text summarization. We present an overview of various methodologies, techniques, and models used in state-of-the-art approaches. Additionally, we discuss the advantages, limitations, and recent advancements in each approach. This survey serves as a comprehensive resource for researchers and practitioners interested in understanding the landscape of abstractive text summarization and identifying the most promising techniques.

**7.Towards a New Hybrid Approach for Abstractive Text Summarization.**

Abstractive text summarization is a challenging task in natural language processing, aiming to generate concise and coherent summaries that capture the essence of the source text. This paper proposes a new hybrid approach for abstractive text summarization, combining the strengths of extractive and abstractive methods to overcome the limitations of existing techniques. We present an overview of the proposed approach, including content extraction, sentence compression, and language generation. Additionally, we discuss the potential advantages and challenges of the new hybrid model and propose future research directions. This paper serves as a starting point for developing innovative approaches that push the boundaries of abstractive text summarization.

Text summarization using a hybrid model is an approach that combines both extractive and abstractive summarization techniques to generate a summary that captures the most important information from a longer piece of text. This approach attempts to address the limitations of using only one type of summarization technique, such as the lack of coherence and readability in purely extractive summarization or the difficulty of generating novel and informative sentences in purely abstractive summarization.

Hybrid models typically involve using machine learning algorithms, such as neural networks or decision trees, to identify the most important sentences or phrases in the original text, followed by a natural language processing (NLP) model that generates a summary by combining and rephrasing the selected information.

Text summarization using hybrid models has numerous applications, including in news and content curation, document summarization, and chatbot response generation. Hybrid models can improve the accuracy and quality of the generated summaries, making them more useful and informative for users. However, they also present some challenges, such as the need for manual tuning of the rule-based methods and the difficulty in integrating the two approaches seamlessly.

**2.2 DRAWBACKS OF EXISTING SYSTEM.**

Text summarization includes **topic identification, interpretation, summary generation, and evaluation of the generated summary.** Most practical text summarization systems are based on some form of extractive summarization.

Existing systems of text summarization, particularly abstractive text summarization, have several drawbacks. Here are some of the key limitations:

1. **Quality and Coherence**: Abstractive text summarization systems often struggle with generating high-quality and coherent summaries. They may produce summaries that are grammatically incorrect, ambiguous, or semantically inconsistent with the original text. This can lead to summaries that are difficult to understand or convey inaccurate information.
2. **Information Loss**: Existing abstractive summarization systems may fail to capture all the relevant information from the source text. They may overlook important details or omit crucial aspects of the original content, resulting in incomplete or biased summaries. This can be problematic when the goal is to provide a comprehensive and unbiased summary.
3. **Subjectivity and Bias**: Abstractive summarization systems can introduce subjectivity and bias into the generated summaries. The system's training data or underlying algorithms may contain biases, leading to summaries that favor certain perspectives or overlook contrary viewpoints. This can undermine the objectivity and fairness of the summarization process.
4. Lack of Control: Existing abstractive text summarization systems often lack user control over the summary generation process. Users may have specific requirements or preferences regarding the summary length, style, or content coverage. However, current systems typically provide limited options for customization, making it challenging to tailor the summaries to individual needs.
5. Data Dependency and Generalization: Many abstractive summarization models require large amounts of labeled training data to perform well. This can be a significant drawback when summarizing domain-specific or low-resource texts where annotated training data is scarce. Furthermore, existing systems may struggle to generalize effectively to diverse types of text, leading to subpar performance when summarizing unfamiliar or complex content.
6. Evaluation Challenges: Assessing the quality of abstractive text summarization is a challenging task. Traditional evaluation metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation) are primarily designed for extractive summarization and may not fully capture the linguistic quality and content coverage of abstractive summaries. As a result, it can be difficult to objectively compare and benchmark different abstractive summarization systems.
7. Ethical Considerations: Abstractive text summarization raises ethical concerns related to the potential for misinformation or manipulation. If the system is not adequately trained or supervised, it may generate summaries that distort or misrepresent the original content, leading to the spread of inaccurate or misleading information.

Addressing these drawbacks is an ongoing research challenge in the field of abstractive text summarization, with the aim of improving the quality, reliability, and user control of the summarization process.

**2.3 OBJECTIVE**

1. The goal of text summarization using machine learning is to reduce the size of the text, while preserving its important information and overall meaning.
2. Enhancing machine understanding of text by extracting the most important features of the text and facilitating further analysis or processing.
3. Improving the accessibility and usability of information by presenting key points in a clear and concise manner of whole / complete text
4. Supporting a wide range of applications, such as news aggregation, document summarization, and chatbot responses.
5. The main objective of text summarization using machine learning is to automatically generate a concise and coherent summary of a long document, while preserving its key information and meaning.

**CHAPTER 3: PROPOSED METHODOLOGY**

**3.1 PROPOSED METHODOLOGY**

Abstractive Text Summarization is the task of generating a short and concise summary that captures the salient ideas of the source text. The generated summaries potentially contain new phrases and sentences that may not appear in the source text.

1. Preprocessing - Preprocessing all the texts and converting them into presentable form.
2. Splitting the records - Split the dataset records into training and testing sets. We will be splitting in the 80:20 ratio where 80% record will be for training sets and 20% for testing sets.
3. Text Vectorization - We will convert our word into integer sequence using vectorization technique.
4. Build the model - We are using 3 Layers of LSTM. This will make our prediction much better. We will build it using an encoder and decoder model.
5. Train the model - We will initialize our Model class with input and output data from the encoder and decoder layers.
6. Inference Model - The inference model is used to test the new sentences for which the target sequence is not known.
7. Sentence selection - Set a threshold or select a specific number of sentences to include in the summary based on their scores. You can experiment with different thresholds or sentence limits to generate summaries of varying lengths.
8. Summary generation - Arrange the selected sentences in a logical order to form a coherent summary. Optionally, post-process the summary to ensure grammatical correctness and improve readability.

**3.2 Tools Used**

1. Seq2Seq Modelling
2. Encoder & Decoder Model
3. Attention Mechanism
4. Stemming

**3.3 Advantage & Disadvantage**

**Advantages:**

Abstractive summarization is considered to produce higher quality summaries compared to extractive methods. It can capture the essence of the original text and generate more coherent and fluent summaries. This leads to improved readability as abstractive summarization can rephrase the original text in a more natural and concise manner.

Furthermore, abstractive summarization is more flexible and can handle a wider range of text types, including documents with complex structures, technical jargon, and domain-specific language. It retains the meaning of the original text by generating a summary that conveys the essence of the original in a more concise way, unlike extractive summarization which simply selects and copies sentences.

Abstractive summarization is also capable of handling complex language, including idioms, colloquialisms, and metaphors, which can be challenging for extractive summarization. It has the ability to generate new sentences that are not present in the original text, resulting in a more coherent and readable summary.

Overall, abstractive summarization provides higher quality summaries, improved readability, flexibility in handling various text types, better retention of meaning, and the ability to generate new sentences.

**Disadvantages:**

Computationally intensive: Abstractive summarization is computationally intensive and requires more resources, including data and computing power, compared to extractive methods.

Risk of introducing errors: Abstractive summarization involves generating new sentences that are not present in the original text, which increases the risk of introducing errors or conveying unintended meaning.

Challenging to evaluate: Abstractive summarization can be challenging to evaluate since it involves generating new language, making it difficult to measure the quality of the summary objectively.

Difficulty in maintaining accuracy: Because abstractive summarization generates new sentences, there is a risk of the summary including inaccurate or misleading information.

Time-consuming and resource-intensive: Abstractive summarization requires significant computational resources and can be time-consuming to train and implement.

Requires large amounts of training data: Abstractive summarization models require a large amount of high-quality training data to learn how to generate accurate summaries. This can be challenging in fields with limited or specialized data.

Limited control over the output: Abstractive summarization models can be difficult to fine-tune or adjust to produce specific types of summaries, such as summaries with a specific tone or focus.

Lack of transparency: The process by which an abstractive summarization model generates its output can be opaque, making it difficult to understand how the model is making its decisions.

**CHAPTER 4: ANALYSIS AND DESIGN**

**4.1 ER diagram**

|  |
| --- |

**Fig.1**

**4.2 Class & Object Diagram**

|  |
| --- |

**Fig.2**

**4.3 Data flow diagram(DFD)**

|  |
| --- |

**Fig.3**

**CHAPTER 5: RESULTS AND DISCUSSION**

**5.1 DISCUSSION**

Text summarization is an important natural language processing and Machine Learning (NLP & ML) technique that aims to condense a longer piece of text into a shorter version while preserving its key information. It plays a crucial role in various applications, such as document summarization, news summarization, and even chatbot responses.

There are two main approaches to text summarization: extractive and abstractive summarization. Extractive summarization involves selecting the most important sentences or phrases from the original text and combining them to create a summary. This approach relies on identifying the most relevant information and does not generate new sentences. On the other hand, abstractive summarization involves generating new sentences that capture the essence of the original text. It uses advanced techniques like natural language generation and paraphrasing to produce coherent and concise summaries.

Both approaches have their strengths and limitations. Extractive summarization is often favored for its ability to maintain the original context and coherence of the text. It tends to be more reliable in terms of factual accuracy since it relies on existing sentences. However, it may struggle with generating cohesive summaries, as it relies on stitching together disjointed sentences. Abstractive summarization, while more flexible in generating summaries, faces challenges in maintaining factual accuracy and ensuring the generated text is coherent and contextually appropriate.

Text summarization techniques heavily rely on NLP algorithms, machine learning models, and deep learning approaches. Popular models like transformer-based architectures. have significantly advanced the field of text summarization by leveraging large-scale pre-training on vast amounts of text data. These models can be fine-tuned for specific summarization tasks, improving the quality and coherence of the generated summaries.

Evaluating the quality of a text summary is a subjective task, as it depends on factors like relevance, coherence, and informativeness. Common evaluation metrics include ROUGE (Recall-Oriented Understudy for Gisting Evaluation) and BLEU (Bilingual Evaluation Understudy), which compare the generated summary with reference summaries or human-created summaries. However, these metrics have their limitations, and assessing the overall quality of a summary remains an ongoing research challenge.

As text summarization techniques continue to evolve, there are several areas of ongoing research and development. These include improving the abstractive summarization capabilities, addressing the challenge of generating factually accurate summaries, handling multiple document summaries, and incorporating domain-specific knowledge for specialized summarization tasks.

Overall, text summarization is a dynamic field that holds immense potential in information retrieval, content analysis, and aiding users in quickly digesting large volumes of text. Continued advancements in NLP and machine learning will likely drive further progress in this area, making text summarization an essential tool for efficiently processing and extracting valuable insights from textual data.

**5.2 RESULTS**

In this project, we have developed a Text Summarizer model which generates the summary from the provided review using the LSTM model and Attention Mechanism.

In the 1,00,000 dataset row we have used we have used 80% of the dataset for training and 20% dataset for testing to validate a correct result.

Abstractive Summarization: Abstractive methods, which generate new sentences to capture the essence of the text, have also made significant progress. Transformer-based models, such as LSTM and Seq2Seq, have shown impressive results in generating coherent and concise summaries.

Text summaries are essential for measuring the effectiveness of summarization techniques. have been widely used to compare generated summaries with human or reference summaries. These metrics have served as benchmarks for summarization systems and facilitated performance evaluation and comparison.

It's worth noting that while text summarization techniques have achieved impressive results, challenges such as maintaining factual accuracy, addressing information omission or redundancy, and ensuring coherence and fluency in the generated summaries still persist. Ongoing research aims to overcome these challenges and further advance the

field of text summarization, enabling more effective and accurate summarization of textual information across various domains and applications.

|  |
| --- |

**Fig.4**

**Final Output:-**

|  |
| --- |

**Fig.5**

**Model Loss**

|  |
| --- |

**Fig.7**

|  |
| --- |

**Fig.8**

**CHAPTER 6: CONCLUSION AND FUTURE SCOPE**

**6.1 FUTURE SCOPE**

Text summarization is a rapidly growing field, and there is enormous potential and future development. Here are some areas of growth of it:

1. Domain-specific summarization: Different industries and fields have their own unique language and terminology, making it challenging to summarize text accurately. There is an opportunity for text summarization systems to be developed that can summarize text for specific industries or domains.
2. Personalized summarization: Individuals have unique preferences and interests, and there is an opportunity for text summarization systems to be developed that can summarize text based on a person's interests and preferences.
3. Summarization of multimedia content: With the explosion of multimedia content, there is a growing need for text summarization systems that can summarize audio, video, and image content.
4. Real-time summarization: As the world becomes more fast-paced, there is a growing need for real-time text summarization systems that can quickly summarize text as it is generated.
5. Multilingual summarization: As the world becomes more interconnected, there is a growing need for text summarization in multiple languages. In the future, we may see advancements in multilingual summarization techniques that can accurately summarize text in multiple languages.

**6.2 CONCLUSION**

In conclusion, abstractive text summarization is a powerful technique in natural language processing that aims to generate concise and coherent summaries by understanding the meaning and context of the original text. It offers the advantage of flexibility and the ability to create summaries that go beyond simply extracting sentences from the source.

Abstractive summarization techniques, often based on advanced machine learning models such as transformer-based architectures, have made significant progress in recent years. These models leverage large-scale pre-training and fine-tuning to generate summaries that capture the essence of the original text while maintaining coherence and fluency.

However, abstractive summarization still faces challenges. Ensuring factual accuracy in the generated summaries remains a concern, as the models might generate information that deviates from the source or introduces inaccuracies. Additionally, generating concise and coherent summaries that are faithful to the original intent of the text can be a complex task.

Ongoing research in abstractive summarization focuses on improving the quality and accuracy of the generated summaries, incorporating domain-specific knowledge, and addressing challenges related to long document summaries and handling diverse types of texts.

Despite these challenges, abstractive text summarization holds immense potential in various applications, including news summarization, document summarization, and content summarization for chatbots and virtual assistants. It enables users to quickly grasp the main points of lengthy texts, facilitates information retrieval, and enhances overall efficiency in processing and understanding large volumes of textual data.

As advancements in NLP and machine learning continue, we can expect further improvements in abstractive text summarization techniques, making them increasingly valuable in extracting relevant information and providing concise summaries from complex and extensive texts.

**7. REFERENCES**

[**https://ieeexplore.ieee.org/document/9458057**](https://blog.floydhub.com/gentle-introduction-to-text-summarization-in-machine-learning/)

[**https://www.researchgate.net/publication/354291874\_Text\_Summarization\_Approaches\_Using\_Machine\_Learning\_LSTM**](https://blog.floydhub.com/gentle-introduction-to-text-summarization-in-machine-learning/)

[**https://blog.floydhub.com/gentle-introduction-to-text-summarization-in-machine-learning/**](https://blog.floydhub.com/gentle-introduction-to-text-summarization-in-machine-learning/)

[**https://towardsdatascience.com/a-quick-introduction-to-text-summarization-in-machine-learning-3d27ccf18a9f**](https://blog.floydhub.com/gentle-introduction-to-text-summarization-in-machine-learning/)

[**https://www.ripublication.com/ijaer19/ijaerv14n7\_41.pdf**](https://peerj.com/articles/cs-1176/)

[**https://peerj.com/articles/cs-1176/**](https://peerj.com/articles/cs-1176/)

[**https://www.researchgate.net/publication/363860140\_AUTOMATIC\_TEXT\_SUMMARIZATION\_USING\_DEEP\_LEARNING\_AND\_NLP\_MODEL**](https://peerj.com/articles/cs-1176/)

[**https://towardsdatascience.com/text-summarization-using-deep-neural-networks-e7ee7521d804**](https://peerj.com/articles/cs-1176/)

[**https://ieeexplore.ieee.org/document/9396855**](https://ieeexplore.ieee.org/document/9396855)

[**https://www.rgnpublications.com/journals/index.php/jims/article/view/760#:~:text=Text%20summarization%20is%20a%**](https://ieeexplore.ieee.org/document/9396855)

[**20process,fuzzy%20logic%20to%20select%20sentence.**](https://ieeexplore.ieee.org/document/9396855)

[**https://www.sciencedirect.com/science/article/pii/S1877050918322026**](https://ieeexplore.ieee.org/document/9396855)

[**https://www.sciencedirect.com/science/article/pii/S1877050918307701**](https://ieeexplore.ieee.org/document/9396855)

[**https://pub.towardsai.net/understanding-abstractive-text-summarization-from-scratch-baaf83d446b3**](https://pub.towardsai.net/understanding-abstractive-text-summarization-from-scratch-baaf83d446b3)

[**https://medium.com/globant/abstractive-text-summarization-bccb4bf5851c**](https://medium.com/globant/abstractive-text-summarization-bccb4bf5851c)

[**https://aclanthology.org/P19-1501/**](https://aclanthology.org/P19-1501/)

[**https://www.thesmartcube.com/ai-lab/experiments/7424/**](https://www.thesmartcube.com/ai-lab/experiments/7424/)

[**https://www.hindawi.com/journals/cin/2022/7132226/**](https://www.hindawi.com/journals/cin/2022/7132226/)

[**https://ceur-ws.org/Vol-2718/paper28.pdf**](https://ceur-ws.org/Vol-2718/paper28.pdf)

[**https://www.sciencedirect.com/science/article/pii/S1877050918307701**](https://www.sciencedirect.com/science/article/pii/S1877050918307701)

[**https://www.researchgate.net/publication/367139564\_Abstractive\_text\_summarization\_of\_low-resourced\_languages\_using\_deep\_learning**](https://www.researchgate.net/publication/367139564_Abstractive_text_summarization_of_low-resourced_languages_using_deep_learning)

[**https://www.semanticscholar.org/paper/Deep-Learning-Based-Abstractive-Text-Summarization%3A-Suleiman-Awajan/b7da726c244287748575ef404009609afde45bea**](https://www.semanticscholar.org/paper/Deep-Learning-Based-Abstractive-Text-Summarization%3A-Suleiman-Awajan/b7da726c244287748575ef404009609afde45bea)

[**https://dl.acm.org/doi/abs/10.5555/3524938.3525989**](https://dl.acm.org/doi/abs/10.5555/3524938.3525989)