International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Abstractive Text Summarization Using Transformer Based Approach

Karishma Shukla[1], Kartik Barange[2], Prajakta Shahabade[3], Akanksha Pandey[4], Bavkar Dnyaneshwar Madhukar[5]

[1, 2, 3, 4]*B.E. Student, Department of Computer Engineering, Terna Engineering College, Navi Mumbai, Maharashtra, India*
[5]*Assistant Professor, Department of Computer Engineering, Terna Engineering College, Navi Mumbai, Maharashtra*

***Abstract*—The amount of textual data on the Internet has risen in an exponential manner over the past decade. Information utilization has become an expensive and time-consuming activity since data expands in large quantities at a time and includes irrelevant content or noise. One such common technique that helps in dealing with enormous data is Text summarization. Automatic text summarization is a renowned approach that is used to reduce a document to its main ideas. It operates by preserving substantial information by creating a shortened version of the text. Automatic Text Summarization helps in creating a short, coherent, and fluent summary of a longer text document and involves outlining of the text's major points using Natural language processing (NLP). Automatic Text Summarization is the process of creating a condensed form of a text document which maintains significant information and the general meaning of source text. This project aims to build a web-based text summarization application using the abstractive text summarization technique to get the most precise and useful information from a document and eliminate irrelevant data. This is achieved by using T5 (Text-to-Text Transfer Transformer) which is a state-of-the-art model that converts every language problem into a text-to-text format along with the CNN DailyMail dataset which is used as the training dataset.**

***Index Terms*—Text Summarization, T5 (Text-to-Text TransferTransformer), Natural Language Processing.**

I. INTRODUCTION

The process of utilizing software to condense a text docu- ment into a concise summary that highlights the main concepts from the original text is known as text summarization the process of text summarization involves using software to reduce a text document in order to provide an abstract or summary of the original material automatic text summariza- tions objective is to extract the crucial information from a large body of text however because the process is time- consuming and more information is being created every day this is becoming increasingly challenging. There are two for text summarization namely Extractive Summarization and Abstractive Summarization.

Extractive Summarization: By selecting the key phrases or sentences from the original text and piecing together chunks of the content to create a condensed version, extractive meth- ods try to summarise articles. The summary is then created using these extracted sentences..

Abstractive Summarization: Contrary to extraction, this method depends on the ability to condense and paraphrase portions of a document utilising sophisticated natural language approaches. Considering that abstractive machine learning algorithms can produce fresh words and phrases to accurately reflect the content of the source text. The correct application of such abstraction in deep learning issues can help overcome grammatical errors.

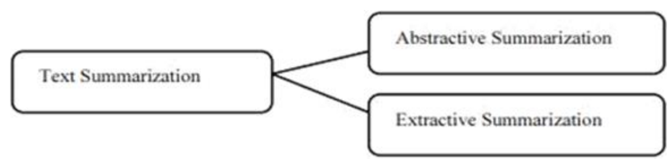

Fig. 1. Text Summarization Techniques

The goal of summarising in this project is to condense a lengthy document or a news article without losing any infor- mation. Deep Learning has become more and more popular in recent years for generating text summaries. Understanding the principles of natural language processing and developing a tool for text summarization are the goals of this research. The need for automatic summarization is growing significantly, eliminating the need for manual labor. The project focuses on developing a tool that summarises the document automatically. This project's primary goal is to extract or produce a summary from the text provided as input. Examples of the range of this online application's functionality include writing headlines for news channels, briefings, note-taking, making highlights for speeches, etc.

II. LITERARY SURVEY

The paper [1] This research [1] illustrates how the volume of online text data is continuously increasing, making text summarization critical for modern recommender and text clas- sification systems. Extractive and abstractive summarization are two ways for creating summaries. Extractive summariza- tion picks relevant information. Sentences from the original content, whereas abstractive summarization analyses the text to construct a summary. The research evaluates and compares the summaries generated by machine learning models using the BBC news dataset with human-written summaries. The authors used pre-trained transformer-based language models to summarise text and discovered that carefully tailored models provided fluent summaries. They used ROUGE ratings to compare the models and found that the T5 model performed the best. A model hybrid could improve the summaries' correctness, fluency, and coherence.

The authors implemented pre-trained transformer-based language models for text summarization and found that finely- tuned models produced fluent summaries. They evaluated the models using ROUGE scores and concluded that the T5 modelperformed the best. A hybrid of the models could improve the accuracy, fluency, and coherence of the summaries.

The paper [2] proposes a model for web document summarising that makes advantage of social media data such as user comments and tweets. Unlike previous research that used feature engineering, this model uses transformers to fully use the data's contextual component. To enrich the information, relevant user postings are matched with sentences, and a convolution neural network layer is added for categorization. Experiment findings on two English datasets show that this model outperforms other models that do not use transformers in summarising single documents.

The paper [3] describes The explosion of data on the internethas created a demand for solutions that convert raw data into valuable data that humans can understand. Text summariza- tion, which entails condensing a document to its key ideas, is one method for dealing with massive amounts of data. There are two types of text summarization: extractive and abstractive. sentences from the original text to create a subset of mean- ingful sentences, whereas abstractive approaches construct a shorter version of the text while retaining significant informa- tion. NLP academics are particularly interested in extractive approaches, which use linguistic and statistical properties to calculate the implications of phrases. This research examines both extractive and abstractive strategies for summarising textsand provides insight into methods that result in less repetition and a more condensed summary. The future scope of automatic text summarization is to answer these issues and make the technology easier and more feasible to apply. The search for the optimal model that can generate a summary like a humanis ongoing.

The paper reviewed various research papers on abstractive, extractive, and hybrid techniques for text summarization, as well as learning methods such as supervised, unsupervised, and reinforcement. Each of these techniques has its own set of challenges that need to be addressed, including evaluating the quality of the summaries, obtaining labeled data, dealing with anaphora and cataphora problems, and others. The future scope of automatic text summarization is to resolve these challenges and make the technology easier and more feasible to implement. Research on automatic text summarization is ongoing to find the perfect model that can generate a summarylike a human.

The paper [4] investigates several strategies for producing text document summaries, including Abstractive (ABS) and Extractive (EXT) summaries. Query-based summarising ap- proaches are also presented. The research article focuses on structured-based and semantic-based approaches to summaris-ing text documents. The summaries produced by these models were tested using a variety of datasets, including the CNN corpus, DUC2000, and single and multiple text documents. The paper also analyses previous work, achievements, trends, and the future scope of text summarization in many domains. These algorithms can generate several sorts of summaries, and their accuracy can be compared using metrics like as ROGUE and TF-IDF scores. However, the generated summaries are not always perfect and may be irrelevant to the original content. The field of text summarization is still evolving, and numerous efforts have been made to improve it. Because there is no single model that produces the best summaries, the discussed models can be updated in the future utilising GANs and transfer learning to produce more accurate summaries. This can result in the development of new and improved text summarising ideas.

These methods can produce different types of summaries, and their accuracy can be compared using metrics such as ROGUE and TF-IDF scores. However, the generated sum- maries are not always perfect and can be irrelevant to the original document. The field of text summarization is ongoing, and various works have been done to improve it. There isno specific model that generates the best summaries, so the discussed models can be modified using GANs and transfer learning for more accurate summaries in the future. This can lead to the development of new and enhanced ideas for text summarization.

The paper [5] explores the significance of text summariza- tion in light of the increasing amount of text data available online. Extractive and abstractive summarization are two ways for producing summaries. The former pulls relevant sentences from the original material, while the latter interprets the text to construct the summary. The research compares transformer-based pre-trained models for text summarization using the BBC news dataset as an example. Human-generated sum-maries are used to evaluate and compare machine learning model summaries. This study describes the implementation of pre-trained language models for summarization tasks using the transformer architecture. Finely tuned transformer-based language models achieved great results and delivered fluent summaries for a particular text material, according to the researchers. They used ROUGE ratings to compare the per- formance of several models and discovered that the T5 model outperformed all others. Future research should focus on developing more robust models that can generate summaries of varied lengths and be applied to multi-document summarising, according to the researchers. They also propose that a hybrid ofthe models be employed to increase the summaries' accuracy, fluency, and coherence.

III. EXISTING SYSTEM

Our proposal employs the T5 (Text-to-Text Transfer Trans- former) model, which is a more sophisticated and success-ful method to text summarization than the present system's TextRank algorithm. Transformers outperform the TextRank algorithm for text summarization because they can capture the contextual relationships and dependencies between words in a sentence or document. Transformers, unlike TextRank, are trained on large-scale data to learn representations of languagethat capture the intricacies of natural language. This allows transformers to generate more human-like and informative summaries. Transformers can also be fine-tuned f or certain domains or jobs, something TextRank does not allow. Transformers, unlike TextRank, are trained on large-scale data to learn representations of language that capture the intricaciesof natural language. This allows transformers to generate more human-like and informative summaries. Transformers can alsobe fine-tuned F urthermore, t ransformers o utperform LSTM and RNN for text summarising because they can learn associations between words in a sentence or document more effectively, resulting in more accurate and insightful summaries. In general, the T5 model is a better option for abstractive text summarization. 1.Transformers can also be fine-tuned Subash Voleti, Chaitan Raju, and Teja Rani Mugada Swetha : Text summarization project developed utilising natural language processing and Google text-to-speech API. It is extractive (no paraphrasing) and employs the TextRank algorithm, which are both negatives. TextRank employs graph-based techniques and heuristics, whereas T5 is a transformer-based model.T5captures contextual linkages and dependencies between words in a sentence or page better than TextRank.T5 is trained on large-scale data to understand linguistic representations, whereas TextRank depends on heuristics. TextRank cannot befine-tuned f or s pecific do mains or ta sks, wh ereas T5 ca n. T5 is a more advanced and successful text summarising method than TextRank. 2. Pooja Batra; Sarika Chaudhary; Kavya Bhatt; Saloni Varshney; Srishti Verma : I started a project. A Review of Abstractive Text Summarization Techniques Using Natural Language Processing. The disadvantage was that they were performing abstractive summarization using LSTM and RNN. Transformers outperform LSTM and RNN for text summarising because they can more efficiently understand re- lationships between words in a sentence or document, making them more accurate and capable of producing more insightfulsummaries.

1.Subash Voleti, Chaitan Raju, Teja Rani M. Swetha: Developed a project Text summarization using natural language processing and google text-to-speech API. The disad- vantages where it is extractive (no paraphrasing) and it uses the TextRank algorithm. While T5 is a transformer-based model, while TextRank uses graph-based approaches and heuristics.
T5 captures contextual relationships and dependencies be- tween words in a sentence or document better than Tex- tRank.T5 is trained on large-scale data to learn representations of language, whereas TextRank relies on heuristics. T5 canbe fine-tuned f or s pecific do mains or ta sks, wh ile TextRank cannot. T5 is a more sophisticated and effective approach to text summarization than TextRank.

2. Batra, Pooja; Chaudhary, Sarika; Bhatt, Kavya; Varshney, Saloni; Verma, Srishti : Created an project A Review: Abstractive Text Summarization Techniques using NLP. The disadvantage was They are doing abstractive summarization but with LSTM and RNN. Transformers are better than LSTM and RNN for text summarization because they can learn relationships between words in a sentence or document more effectively, making them more accurate and able to generate more informative summaries.

IV. COMPARATIVE STUDY

Comparison between T5 Transformer, BART, Pegasus, and TextRank algorithm for Abstractive Text Summarization basedon their Rouge Score.

T5 Transformer has been shown to outperform both BART and Pegasus models for abstractive text summarization tasks. The T5 model has achieved a RougeL score of 41.2 on the CNN/Daily Mail dataset, which is considered to be a good score. This is because the T5 model has the ability to generate summaries with higher semantic coherence and accuracy by using its powerful pre-training and fine-tuning algorithms.

T5 Transformer has been proven to outperform both the BART and Pegasus models for abstractive text summarization tasks. On the CNN/Daily Mail dataset, the T5 model earned a RougeL score of 41.2, which is considered a good score. Because of its sophisticated pre-training and fine-tuning tech- niques, the T5 model can provide summaries with improved semantic coherence and correctness.

However, it is worth noting that the TextRank algorithm,an unsupervised approach for text summarization, has demon-strated good performance on some datasets. The TextRank al- gorithm selects key sentences from the input text using graph- based ranking methods. While the RougeL score obtained by the TextRank algorithm is lower than that of the T5, BART, and Pegasus models, it has demonstrated good performance on particular datasets.

In conclusion, T5 Transformer has proven to be the best model for abstractive text summary tasks because to its high RougeL score and ability to generate summaries with higher semantic coherence and accuracy. While BART and Pegasus models have also performed well, they fall short of T5. Furthermore, while the TextRank algorithm, an unsupervised approach for text summarization, has demonstrated good per- formance on some datasets, it falls short of the performanceof supervised models such as T5, BART, and Pegasus.

V. PROBLEM STATEMENT

Over the last decade, the volume of textual material on the Internet has grown at an exponential rate. Because data expands in vast numbers at once and includes useless content or noise, information utilisation has become an expensive and time-consuming task. Text summarization is a frequent technique for dealing with large amounts of data. Automatic

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

text summary is a well-known method for distilling a document's primary points. It works by compressing the text and preserving important information. summarising can help many downstream applications by compressing enormous amounts of information into brief summaries, such as making news digests, report generation, news summarizing and headlinedevelopment.

VI. METHODOLOGY

This project employs Natural Language Processing for automatic text summarization. We use the CNN DailyMail Dataset trained on the T5 transformer model. The CNN DailyMail Dataset is an English-language dataset including over 300,000 unique news stories authored by CNN and the Daily Mail journalists. T5, or Text-to-Text Transfer Transformer, is a Transformer-based architecture that works with text. The dataset is separated into three sections: training, testing, and validation. The training set has 287,113 articles, the testing set has 11,490 articles, and the validation set has 13,368 articles.

The dataset's characteristics are article and highlight, where article is a string holding the body of the news story and highlights is a string containing the article author's highlight (considered the summary).Before developing the model, the dataset is pre-processed. In this phase, we tokenize, lowercase, and remove stop words and quotations.

The training data is used to train the T5 model. We utilise Rouge Score as a performance metric to evaluate the model's performance. The Rouge Score is a set of criteria used to assess the quality of artificial text summarising systems, notably abstractive text summarization.

The training phase is subsequently followed by the testing phase, in which we employ the reserved text data for testing. Following testing, the model is used to construct a REST API that is delivered to the web application and allows any user to generate summaries from textual input.

VII. PROPOSED SYSTEM

The project proposes developing a web-based text summarising program that employs the abstractive text summarization technique to extract the most precise and useful information from a document while removing extraneous material. This is accomplished by combining the CNN Daily Mail dataset with T5 (Text-to-Text Transfer Transformer), a cutting- edge model that turns every language difficulty into text-to- text format. The collection contains 3,11,971 new articles. The dataset is separated into train, test, and validation data. Before constructing the model, the dataset is pre-processed. In this phase, we tokenize, lowercase, and remove stop words and quotations. The T5 model is trained using training data and assessed using the Rouge Score. The training phase is subsequently followed by the testing phase, in which we employ the reserved text data for testing. This model is further used to create the web application, using which any user can generate summaries from textual data.

VIII. DESIGN

A. Data Flow Chart (Application)

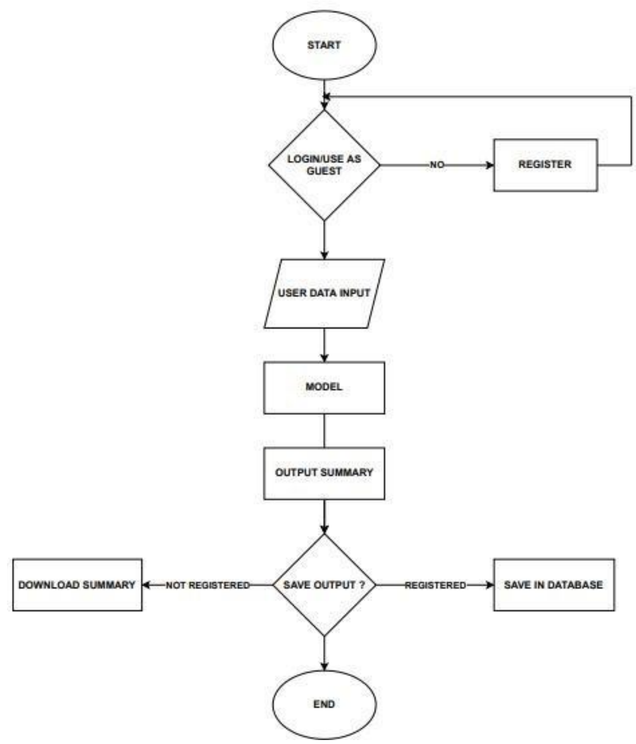

Fig. 2. Data Flow Chart (Application)

As shown in Fig.3 The purpose of this process is to extract a summary that focuses only on the most important points from a large amount of data. To begin, the user must either login or proceed as a guest. Input can be provided in one of three forms: text or document. The model will then process the input and provide a summary of the given information as output. If the user has not registered, the output can be downloaded. Otherwise, it can be saved in the database upon completion of registration. Once the process is complete, the user can create as many text summaries as desired and save them in the database if they are logged in as a registered user or download them directly in .txt format.

B. Data Flow Chart (Model)

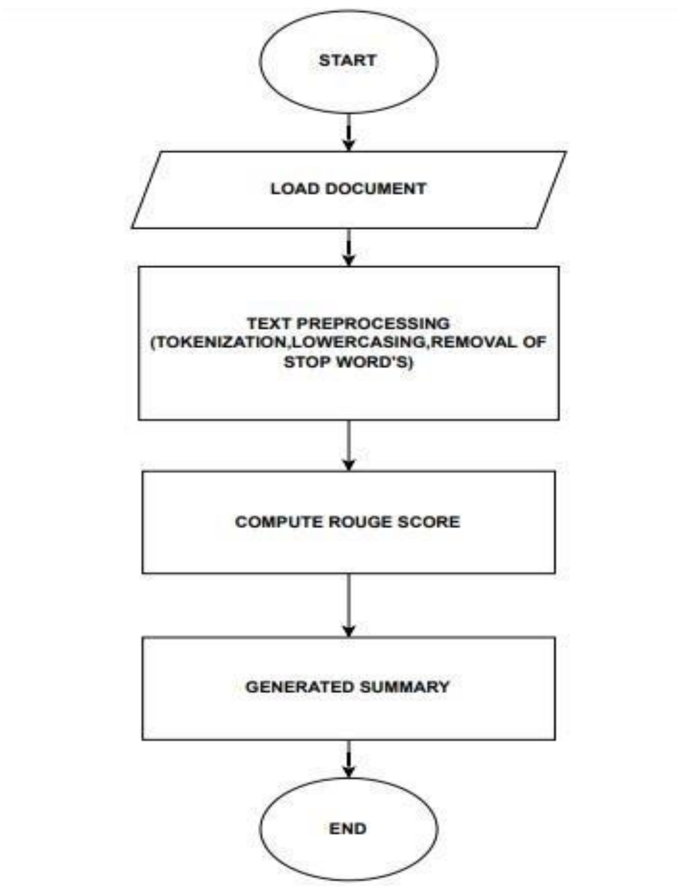

Fig. 3. Data Flow Chart (Model)

As shown in Fig.4 To begin the process, input the data into the provided box in the form of a document or text. The next step is to preprocess the input data by performing the following operations:(i) Tokenization: This involves splitting the entire data into tokens, meaning splitting paragraphs into sentences and then those sentences into words. This step is crucial for the subsequent parsing process of NLP. (ii) Lowercasing: Converting uppercase letters to lowercase helps to standardize the text and make it easier to compare and analyze. (iii) Removal of stop words: Stop words like articles are removed to create a summary that focuses only on important points. This helps to improve the efficiency and accuracy of the model. Our model works based on the Rouge score, Rouge Score is a performance metric used for evaluating automatic summarization by comparing produced summary against a set of references (human-produced). After completing all the steps in the procedure, the desired output is generated: the most accurate summary, covering all the important points. Thismarks the end of the process.

A use case diagram illustrates the step-by-step process of how a user can interact with our web application and how the system will handle various input data scenarios to provide proper output in the form of summarized text to the user. This diagram visually represents the different use cases of the application, which includes a user's ability to login, register, upload input from a PDF, save output as a .txt file, or simply download the generated summary.

The use case diagram is an important tool used in software development to help identify and define the application's functional requirements. It helps developers understand the different actions a user can take in the system and how the system should respond to those actions. The diagram provides a clear overview of the application's functionality, making it easier for developers to design, develop, and test the system.

In the context of a web application, a use case diagram can help identify potential user scenarios and streamline the user experience. For example, if the user's ability to login is a key use case, the diagram can help identify any potential issues with the login process and allow developers to optimize it fora smoother user experience.

Overall, a use case diagram is an important tool for under- standing and designing software systems. It helps ensure that the application meets user needs and requirements and pro- vides a clear understanding of how the system will function.

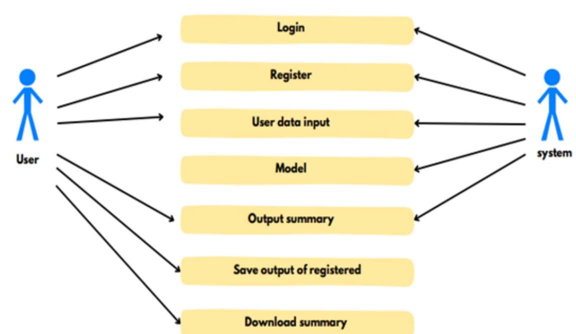

Fig. 4. Use Case Diagram

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

IX. RESULTS

A. Main Page

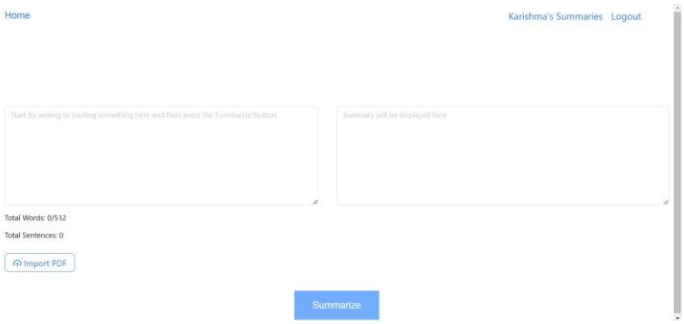

Fig. 5. Main Page

Users can login or register using the Main Page.

B. Generated Summary

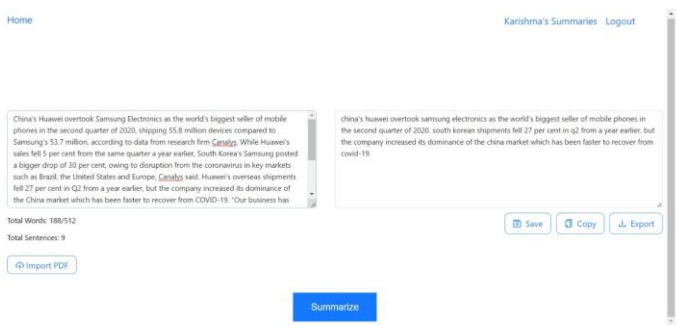

Fig. 6. Generated Summary

The user's inputted data get processed by our model and it generates a summary.

C. Exported File

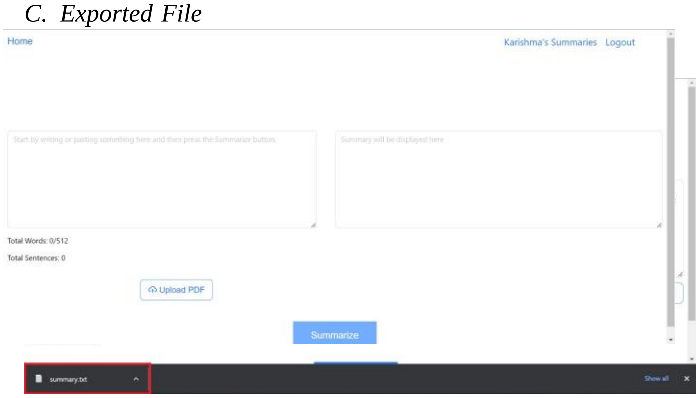

Fig. 7. Exported File

The user's inputted data get processed by our model and it generates a summary. Users can download the generated summary in the form of a .txt file.

D. Saved Summaries

Fig. 8. Saved Summaries

The user's inputted data get processed by our model and it generates a summary. Users can download the generated summary in the form of a .txt file. If the user has registered,he can save his generated summaries in the database.

E. Imported PDF

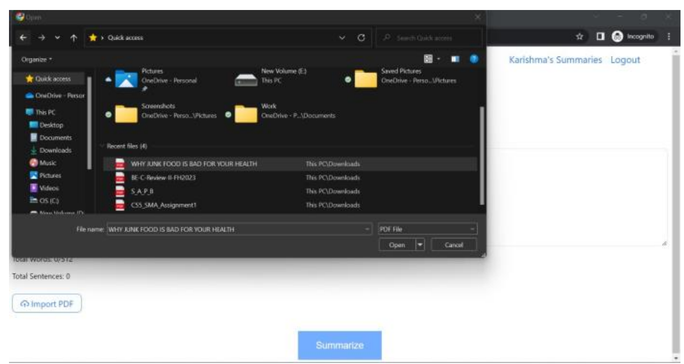

Fig. 9. Imported PDF

The user can import the document file which needs to be processed to generate summaries.

F. Upload PDF

Fig. 10. Upload PDF

the user can upload the document file as PDF which needs tobe processed to generate summaries. if the user has registered he can save his generated summaries in the database. Users can download the generated summary in the form of a .txt file.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

G. Summary from PDF

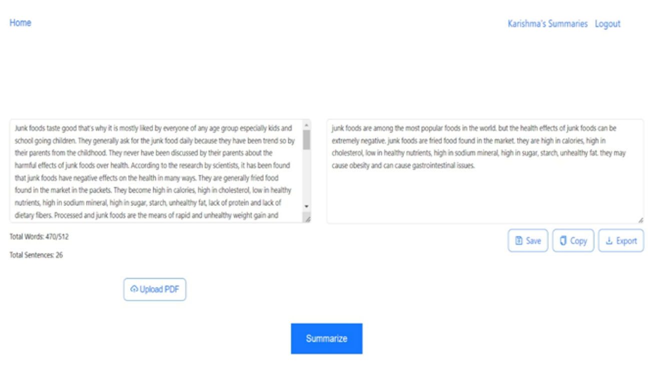

Fig. 11. Upload PDF

the user can upload the document file as PDF which needs to be processed to generate summaries. After uploading the PDF, we can generate a summary of that PDF too.

X. EVALUATION METRICS

Rouge (Recall-Oriented Understudy for Gisting Evaluation) is a set of metrics used to evaluate the quality of automatic text summarization systems, particularly in abstractive text summarization. Rouge scores measure the similarity between the generated summary and a set of human-generated reference summaries.

There are several variations of Rouge scores, such as Rouge-N, Rouge-L, and Rouge-S. Rouge-N measures the n-gram overlap between the generated summary and the reference summary, while Rouge-L computes the longest common sub- sequence between them. Rouge-S is a variant of Rouge that focuses on sentence-level similarity.

In Rouge-S, the system-generated summary is compared to the reference summaries based on sentence-level skip-bigram co-occurrence statistics. The skip-bigram co-occurrence statis- tics measure the degree of similarity between sentences based on the number of overlapping skip-bigrams (i.e., pairs of non-contiguous words) in the sentences. Rouge-S is particularly useful for evaluating abstractive summarization systems that generate new sentences rather than just selecting sentences from the source text.

To compute Rouge-S, the system-generated summary and the reference summaries are first tokenized and then the skip-bigrams are extracted from each sentence. The skip-bigrams are then compared between the system-generated summary and each of the reference summaries. The Rouge-S score is the average of the F1 scores for all the reference summaries.

Rouge scores are widely used in the field of natural language processing to evaluate the quality of text summarization sys- tems. They provide a quantitative measure of the similarity between the system-generated summary and the reference summaries, allowing researchers and developers to compare different systems and to track improvements over time.

ROUGE-1 ROUGE-1 Precision and Recall compare the similarity of uni-grams between reference and candidate sum- maries. By uni-grams, we simply mean each token of com- parison is a single word.

For instance, we can express the uni-grams of ‘John loves data science’ as a Python list of tokens: [’John’,’loves’,’data’,’science’]

RECALL By Recall, we simply refer to the proportion of words in the reference summary captured by the candidate summary. Recall = TruePositives / (TruePositives + FalseNeg- atives)

PRECISION By Precision, we refer to the proportion of words suggested by the candidate summary that actually appear in the reference summary. Precision = TruePositives / (TruePositives + FalsePositives)

ROUGE-2 ROUGE-2 Precision and Recall compare the similarity of bi-grams between reference and candidate sum- maries. By bi-grams, we mean each token of comparison is 2 consecutive words from the reference and candidate summaries. For instance, the bi-grams of ‘John loves data science’ can be expressed as the following tokens: [’Johnloves’,’loves data’,’data science’]

ROUGE-L ROUGE-L Precision and Recall measures the Longest Common Subsequence (LCS) words between refer- ence and candidate summaries. By LCS, we refer to word tokens that are in sequence, but not necessarily consecutive.To understand this, let us look at a convoluted example with reference summary: John really loves data science very much and studies it a lot. And the candidate summary (italicised words represent LCS): John very much loves data science and enjoys it a lot.

Semantic Similarity Semantic Similarity, or Semantic Tex- tual Similarity, is a task in the area of Natural Language Processing (NLP) that scores the relationship between texts or documents using a defined metric. Semantic Similarity has various applications, such as information retrieval, text summarization, sentiment analysis, etc.

There have been a lot of approaches for Semantic Similarity. The most straightforward and effective method now is to use a powerful model (e.g. transformer) to encode sentences to get their embeddings and then use a similarity metric (e.g. cosine similarity) to compute their similarity score. The similarity score indicates whether two texts have similar or more different meanings.

T5 (Text-to-Text Transfer Transformer) is a state-of-the-art Transformer-based model that has achieved impressive results in many NLP tasks, including STS and text summarization. T5 is a text-to-text model, which means that it can be fine-tuned on a wide range of NLP tasks by framing them as text-to-text problems.

To use T5 for STS, the model is fine-tuned on a specific STS dataset, such as the STS-Benchmark dataset, which consistsof pairs of sentences with human-labelled similarity scores. During fine-tuning, the input to the model consists of two concatenated sentences separated by a special separator token. The model learns to predict a similarity score between 0 and 5 for each sentence pair.

T5 has several advantages over other STS models, includingthe ability to handle long input sequences, the ability to generate text as well as score text, and the ability to generate diverse and informative summaries.

In text summarization, T5 can be fine-tuned on a summarization the output is the summary. T5 has been shown to outperform previous state-of- the-art models in text summarization, achieving high scores on various evaluation metrics, including ROUGE.

Overall, the use of T5 in STS and text summarization has led to significant improvements in the quality of text summarization systems. By accurately measuring the semantic similarity between sentences, STS models can be used to select the most relevant and informative sentences for inclusion in a summary, resulting in more coherent and informativesummaries. T5's ability to generate diverse and informative summaries makes it a powerful tool for text summarizationand other NLP tasks.

XI. EXPERIMENT RESULT AND DISCUSSION

A. Result

T5 Transformer, an advanced text-to-text transformer model, was trained using CNN Daily Mail dataset for abstractive text summarization. The model achieved a ROUGEL score of 41.2 after training on 287,113 news items, which is a decent result. ROUGE-L is a popular text summarising evaluation metric that calculates the longest common subsequence of words betweenthe generated summary and the reference summary. A greater ROUGE-L score denotes a higher quality of the resulting summary. While a ROUGE-L score of 41.2 is commendable, there is always space for growth. Various strategies can be used to improve the score, such as using larger training datasets, including more complicated model architectures, and fine-tuning the model with domain-specific data. Overall, the T5 Transformer model's high ROUGE-L score on the CNN Daily Mail dataset suggests its ability in creating high-quality abstractive summaries.

B. Discussion

The T5 transformer model has emerged as the cutting- edge architecture for abstractive text summarising tasks. T5, unlike previous transformer models, is a unified text-to-text transformer that can be fine-tuned for a variety of purposes, including summarization. T5 has a large pre-training corpus, allowing it to master a wide range of language problems,making it extremely successful at abstractive summarization. Furthermore, T5 features a decoder-only architecture, allowing it to generate summaries in a single pass, which makes it faster and more efficient. T5 has also outperformed other state- of-the-art models on numerous benchmark datasets for ab- stractive summarization, attaining the highest ROUGE scores. Furthermore, T5's capacity to generate summaries based on prompts enables it to write summaries for certain domains or tasks, which is a useful feature for real-world applications. Overall, the T5 transformer model's remarkable performance, adaptability, and speed make it one of the finest models for abstractive text summarising tasks.

XII. CONCLUSION

Text summarization is a significant challenge since there is a large amount of information that grows every day, and collect-ing the important data from it is a time-consuming procedure. We created a web-based tool for summarising by implementing the pre-trained language model T5, which is based on the transformer architecture. Using a Deep Learning model, the output is more exact and accurate than other approaches. The abstractive approach has been demonstrated to be the most effective summarising strategy. Automatic text summarising assists users in extracting vital information from large amounts of data.Our research revealed that finely tailored transformers- based pre-trained language models produced excellent results, producing a sound and fluent summary of a given text material.We estimated ROUGE scores [21] for each model's predictionsfor comparison studies and determined that the T5 model beat all other models based on the transformer architecture for the job of summarization.

REFERENC
ES

[1] Anushka Gupta, Diksha Chugh, Anjum, Rahul Katarya, "Automated News Summarization Using Transformers", Delhi Technological Uni- versity, New Delhi, India 110042, 2021.

[2] Minh-Tien Nguyen, Van-Chien Nguyen, Huy-The Vu, Van-Hau Nguyen, "Transformer-based Summarization by Exploiting Social Information", The 2020 12th International Conference on Knowledge and Systems Engineering (KSE), 2021.

[3] Ishitva Awasthi, Kuntal Gupta, Sahejpreet Singh Anand, Prabjot Singh Bhogal, Prof. Piyush Kumar Soni, "Natural Language Processing (NLP) based Text Summarization - A Survey", Proceedings of the Sixth In- ternational Conference on Inventive Computation Technologies [ICICT 2021].

[4] Rahul, Surabhi Adhikari, Monika, "NLP based Machine Learning Ap- proaches for Text Summarization", Proceedings of the Fourth Interna- tional Conference on Computing Methodologies and Communication (ICCMC

2020).
[5] Manish Shinde, Disha Mhatre, Gaurav Marwal, “Techniques and Re- search in Text Summarization – A Survey”, 2021 International Confer- ence on Advance Computing and Innovative Technologies in Engineer- ing (ICACITE 2017).
[6] 5) 6) Subash Voleti, Chaitan Raju, Teja Rani,Mugada Swetha “Text Summarization Using Natural Language Processing And Google Text To Speech Api”, 2020 International Research Journal of Engineering and Technology (IRJET).