Text Summarization using Sequence to Sequence

MSc Research Project
Data Analytics

Ramandeep Singh
Student ID: 21106053

School of Computing
National College of Ireland

Supervisor: Abdul Razzaq

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ramandeep Singh
Student ID:	21106053
Programme:	Data Analytics
Year:	2022/2023
Module:	MSc Research Project
Supervisor:	Abdul Razzaq
Submission Due Date:	15/12/2022
Project Title:	Text Summarization using Sequence to Sequence
Word Count:	5013
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	1st February 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Text Summarization using Sequence to Sequence

Ramandeep Singh
21106053

Abstract

The data has been available on the internet, which is massive in numbers. Hence, it becomes difficult for a reader to read long articles or blogs as he needs to go through all the text for understanding. Text summarization makes life easier for an individual by using Natural Language Processing(NLP), which has been designed for the researcher's community to get large information in a short time. Text document reduces the size of the source document and represents only the key information without changing the text's actual meaning. The researchers have shown their research in extractive abstraction, but now the researcher approach has been broadened and they transform into abstractive methods. In this research paper, a novel approach has been used by combining extractive and abstractive methods. The extractive method has been implemented, including word frequency-based sentence feature extraction by using a graph-based TextRank algorithm. On the other hand, the deep artificial neural network approach has been involved in the abstractive phase, consisting of a sequence-to-sequence encoder-decoder model, which is a neural network of Long short-term memory (LSTM). The three approaches have been followed, sequence to sequence using LSTM encoder and decoder, sequence to sequence using attention mechanism, and text summarization using BERT, GPT-2, and NLTK. ROUGE metrics have measured their accuracies. In the results, the LSTM has achieved the highest accuracy on the news dataset by ROUGE-1(0.40), ROUGE-2(0.09), and ROUGE-L(0.39), which resulted in generating a concise summary without changing the original meaning of a text.

1 Introduction

Recently, a large number of data is available on various platforms such as the internet, blogs, and news articles. In these platforms, information is available excessively for an individual to access that information. With the help of text summarization, a user can save time and consumes a large information in a short span of time summarized data. During the process of summarizing data, the article has been gone through all by understanding the essence of the text and following the key points to generate summaries. On the other hand, if data will be summarized manually, it will take a plethora of time (Sethi et al.; 2017). Text summarization does not change the actual meaning of a text(Allahyari et al.; 2017). Generating the text in the form of short summary, the information can be retrieved and processed effectively as well as efficiently. (Maynez et al.; 2020)There are two types of text summarization extractive and abstractive. In abstractive text summarization, it takes core sentences or words to generate summary data. These words are not necessary it will be present in the text. In other words, the abstractive method has the tendency to create a summary from itself, which results in reducing the grammatical inconsistency of data. On the other hand, extractive summarization takes the key phrases from the original data, followed by combining them to generate a summary. It builds as per the defined metric without any changes to the text. In this research paper, the research has been done by text summarization using sequence to sequence approach.

Many enhancements have been done in the past, especially in the abstractive summarization using sequence to sequence (Baykara and Güngör; 2022). The data can be enhanced in saliency, semantics, and fluency basis by implementing sequence to sequence. As a consequence, the data will be readable with maintaining the flow of data. In Gupta's research team (Gupta and Gupta; 2019), a deep learning approach has been employed by him by using an encoder-decoder framework of recurrent neural network (RNN), especially to overcome the challenges occurring during the phase of implementation. (LeCun et al.; 2015) During this method, the input has been passed from the encode by an internal representation. In the distinct fields of speech recognition, image captioning, machine translation, and video captioning, the former approach has been implemented successfully. As a consequence, by using an encoder-decoder deep learning technique, the latest work has been employed directed towards abstractive summarization. Significant contributions have been employed in this approach by various authors (Nallapati et al.; 2016), (Zhou et al.; 2017),(Shi et al.; 2019). Along similar lines, the long short-term memory(LSTM) was implemented by the Weston group(Rush et al.; 2015) to address the exploding and vanishing gradient problem of the regular RNN network. However, the various negative consequences were discussed by different authors (See et al.; 2017), (Li et al.; 2017). The key problem was found in the training phase in which it was unable to generate a good representation of input documents in their framework, resulting in the wrong factual information and producing summaries repetition (See et al.; 2017).

This research paper has addressed the following question:-

- How can we generate better summaries by using various techniques?

In this research paper, the following three approaches have been done:-

1. Sequence to Sequence by using attention
2. Sequence to Sequence by using encoder decoder model
3. Summarizing using GPT2, XLNET, BERT

The summary model has been proposed in this research paper, which incorporates the extractive and abstractive framework for tasks related to summarization. The contribution of this research has been defined following:-

- The state-of-the-art framework has been explored to identify the appropriate model for summarizing a text.
- The models have been implemented, which capture the best semantic context and show a better input document.
- Implemented a model which can understand the different topics of text and result in generating better summaries.
- The distinct data mining techniques have been investigated, resulting in producing non-repetitive and cohesive summaries.

The research objective is to capture a good representation of the input documents and generate a non-repetitive summary.

2 Related Work

NLP, especially in transfer learning has been witnessed to be one of the effective and enabled state-of-the-art consequences in major varieties of tasks. (Devlin et al.; 2018) The pretraining topic, which is a language model, is proficient in enhancing knowledge of task agnostic by different pretraining objectives, followed by transferring that information to downstream tasks, resulting in accomplishing a depth understanding of the natural language. On the other hand, those tasks that need the knowledge of natural language generation such as text summarization and machine translation might not take advantage of the pre-trained encoder models to that extent, which leads to pretrained sequence-to-sequence models. A masked Seq2Seq generation model (MASS) can reproduce any part of a sentence, especially in the scenario of the remaining parts have been provided (Song et al.; 2019). UNIfied pretrained Language Model (UNILM) engages the synchronized training on the modelling objectives such as bidirectional, unidirectional, and sequence to sequence (Dong et al.; 2019). The distinct denoising objectives were followed in the BART (Lewis et al.; 2019) for corrupting an input text before reconstructing it by using an autoencoder. A generalized text-to-text framework was introduced by T5 (Raffel et al.; 2020), which was capable of handling different kinds of tasks in NLP in the scenarios of sole text in the form of input and output.
In abstractive sentence summarization, a Selective Encoding for Abstractive Sentence Summarization (SEASS) model was proposed by (Zhou et al.; 2017) in which a selective encoding model was used, especially to extend the sequence-to-sequence framework. It is made up of a selective gate, decoder, and encoder. The model consisted of a unidirectional GRU decoder and a bidirectional GRU encoder. The datasets were used from Gigaword, MSRATC, and DUC 2004, especially for testing and training. Interestingly, the best search was implemented to choose the best target word. The ROUGE1, ROUGE2, and ROUGE-L were recorded by 36.15, 17.54, and 33.63 respectively.

The encoder-decoder and sequence-to-sequence problems are considered into abstractive text summarization, which solves these problems. The pioneer of implementing an encoder-decoder architecture was (Rush et al.; 2015) by using a neural network language model (NNLM), especially in addressing the problem of title generation tasks considered as a part of the abstractives ummarization issue. Moving further, the NNLM with recurrent neural networks(RNNs) was replaced by (Chopra et al.; 2016). The state of the art gained momentum recently,which is based on deep learning for summarizing the text. The Sum- maRuNNER was proposed by [8]. It is based upon a GRU-RNN-based sequence model, which can be trained abstractively and extractively for generating summaries. The main aim is to remove redundancy. The work involved two architectures selector and classifier. These two architectures consist of GRU-RNNS, especially for extractive summarization.The DUC 2002 dataset and CNN/DailyMail dataset obtained state-of-the-art performance. To capture the document structure, (Nallapati et al.; 2016) various novel models such as a switching pointer-generator model, a bidirectional LSTM-based encoder-decoder with an attention mechanism, and a hierarchical encoder decoder were introduced. The major contribution of (Hermann et al.; 2015) was transforming the CNN/DailyMail dataset into a format for summarizing the text. Addressing the issue of word repetition, the pointer-generator model (See et al.; 2017) was improved, allowing the repetition of words into copy words from a coverage mechanism and the source document.

(Fitrianah and Jauhari; 2022) Preserving the crucial information without changing the real meaning of a text, the text size was reduced for summarization. The Latent Dirichlet Allocation (LDA) approach was applied for topic modelling. PyLDAvis web-based interactive visualization tool was implemented, especially to visualize particular topics, providing an overreaching on the particular topics. Interestingly, a similarity was shown between the topics during the topic modelling. The limitation was found in the emerging term during evaluation as it was taking longer. The reason behind this is that larger text was being used for analysis during the approach of topic modelling.

The CNN framework has additional competence compared to RNN seq2seq models in the proposed research of (Zhang et al.; 2019). One notable point is that hidden or rare words were enough capable with the model. At the same time, it contains the key sentences and keywords in modelling the hierarchical attention mechanism. The combination of the clustering technique and NLU technique was implemented by using a single document (Nada et al.; 2020). The evaluation was done by ROUGE, resulting in an F-measure of 0.5 scores. Various limitations were to be found. For example, the model relies on the sentence boundary. In the longer texts, the coverage accuracy was decreased, leading to the misunderstood summary due to a linguistic expression. A selective reinforced sequence-to-sequence attention model was proposed by (Liang et al.; 2020),especially for social media. Confirming the decision of using the selective gate for retaining or removing information, before encoding the hidden layer was added. The gate selection method witnessed better performance with the grouping of various methods such as word hidden state, global semantics, and word embedding. The combination of textual and visual information, and summarizing social events were one of the limitations found in an unsupervised way. A sequence-level contrastive(SeqCO) learning model was implemented (Xu et al.; 2022) to reduce the distance from the document to generated summaries in the process of training. This was utilized as an experiment on the three datasets in which it improved the Seq2Seq text generation model. There were also some negative consequences in the multiple contrastive objectives performance. In the abstractive text, there were a lot of improvement was made with sequence-to-sequence (Seq2Seq) (Baykara and Güngör; 2022). XLM, GPT, BERT, and SEq2Seq models such as BART and T5 are one of those examples. This kind of enhancement has resolved the issues occurring in neural summarization, improving semantics, fluency, and saliency to generate quality summaries. The large vocabulary and teacher-forcing techniques were used in a simple RNN to build a Seq2seq method (Abolghasemi et al.; 2022). As a consequence, it was not performed as well as expected, so the attention mechanism was implemented in the model. By changing this approach, better results were achieved. This is one of those models, that took additional time to train.

The unique approach was taken by (Fang et al.; 2022). In this approach, automatic summarization was combined with text classification. The TextCNN and the pointer generator network were implemented for generating abstracts and generative summaries respectively. One notable point is that few deficiencies were shown in the proposed research, resulting in unable to combine of both the automatic summarization and text classification datasets because they were independent parts and lack of datasets. The main focus was on the original document by (Qiu and Yang; 2022) in which the internal structure of semantic features was neglected. Thus,it resulted in problems of grammatical structure errors and semantic deviation, especially in the extractive text summary existing models. The multi-head self-attention and the soft attention mechanism models have been implemented in which the latter mechanism was in the coding stage, which results in achieving maximum weight by enabling accurate summary syntax and semantics information. Therefore, it was witnessed coherent and precise summary. The point network model was also used at the same time, resulting in the repetitive and out-of-vocabulary problems enhanced during generating abstracts.

The different methods have been discussed in this research paper such as the graph-based and tf-idf method, which had been employed to generate better summarization data. Hence, on the basis of the literature, it can be seen that no single model yielded the expected consequence of making a better summary of data. Therefore, the novel model of the combination of extractive and abstractive models has achieved the state of the art results. The main advantage of this is to handle uncommon and repetitive words, constructing human-readable summaries in a concise way. The LSTM encoder-decoder and LSTM with an attention mechanism have been implemented in this research paper. A better understanding of the data has been achieved by the three layers as it was stacked in the encoder, especially before producing the abstractive summary.

3 Methodology

This section illustrates the approach, which has been used as a methodology to accomplish the research on text summarization. In this research paper, text summarization is the prominent part of the machine learning issue in which Natural Language Processing (NLP) is one of the important tasks in building an efficient and working model. To achieve this, distinct data analytics frameworks can be followed, which are defined in the processes to deliver a successful product. Various methodologies are available such as Sample, Explore, Modify, Model, Assess (SEMMA), Knowledge Discovery in Databases (KDD), and Cross Industry Standard Process for Data Mining (CRISP-DM) (Azevedo and Santos; 2008). In these methodologies, the KDD is the best approach for the research purpose. This is because it covers all the essential steps, which are needed for the proposal. The KDD is one of those methodologies that emphasize the execution of tasks, starting from the data selection, followed by building a model and evaluation. It can help in retrieving essential information from repositories and databases. Figure 1. illustrates the steps that have been followed in the KDD approach for the proposed model.

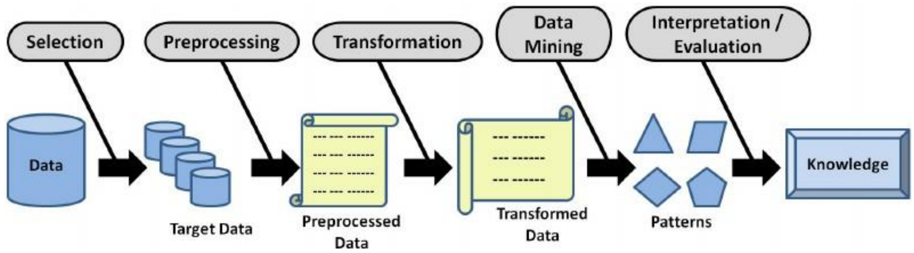

Figure 1: Methodology

3.1 Data Collection

In this research, the two datasets have been collected from Kaggle website. Firstly, the News dataset, contains the author's name, headlines, complete article, short text, and URL of the article. The information has been extracted into a CSV file format. This dataset contains a total of 4515 examples. Secondly, CNN DailyMail News dataset, which consists of news articles and their corresponding highlights. It contains 13,368 validation pairs, 11,490 test pairs, and 287,113 training pairs.

3.2 Data Preparation and Pre-processing

In the preprocessing step, after the collection of input data, the data pre-processing steps have been followed. It removes noisy data, irrelevant data, or unnecessary data from the samples. In other words, noise data have negative results, especially in the scenarios of building the model and generating outputs. In the beginning step, the required attributes have been taken from the data samples. In this scenario, both news datasets text and reference summary has been taken. Following preprocessing steps, the duplicate and NA values, symbols, punctuations, short words, stop words, and special characters have been removed. Although contraction mapping and everything has been converted into lowercase on the specific fields from the sample. These data cleaning tasks have been performed by using Python programming language.

3.3 Feature Extraction

The feature extraction part is needed after the preprocessing steps. Specifically, it is used to improve the features from the raw data which are relevant. Various machine learning models have been implemented. Term Frequency-Inverse Document Frequency(TF-IDF) features were extracted to generate phrases and common words. It was generated by the Scikit library in the python programming language.

3.4 Data Mining

The two types of stages have been involved abstractive and extractive. The former results were consequently fed as input, especially to the second phase, followed by generating the final result. In the abstractive phase, a sequence-to-sequence RNN-LSTM model has been implemented by using Keras and TensorFlow. The extractive summary served as an input to this phase and then the data has been split into test and train by using sklearn library. The training has been implemented by fine-tuning hyperparameters to accomplish a model, which yields the best review representation as well as proficiency in making predictions accurate.

3.5 Interference and Evaluation

Finally, the most prominent part is to make predictions after the training phase, especially unseen data and accuracy. The probability distribution has been produced by this model for each token in the results. In other words, the model has been predicted for each summary because the maximum amount of words occurring in the summary would be produced as an array by that. Thus, the specific word will be shown by probability. The decoding algorithm needs to be implemented for the probability distribution (Wilt et al.; 2010). The Greedy search decoder is considered to predict the highest likelihood words for each prediction and combined all the predicted words to generate the final output results. However, the Beam search decoder is not known to consider likelihood words for each prediction. This is because it considers top k words with high probabilities. As a result, it does not give an output sequence as compared to a Greedy search. On the basis of performance, the Greedy search gives better results, especially with computational resources. Next, the accuracy of the model was calculated after making interferences from the predictions. Between the models generated summary and reference memory, the accuracy of the model has been achieved by measuring the overlapping of words. Ideally, the model's accuracy was achieved by recall and precision. But these metrics are not capable of predicting model summary. Therefore, RecallOriented Understudy for Gisting Evaluation (ROUGE) metrics has been considered for measuring the accuracy of the model. The accuracy has been compared on the distinct levels of granularities by using ROUGE-1, ROUGE-2 and ROUGE-L.

4 Design Specification

4.1 Abstractive Approach

The best approach to building unstructured data is the artificial neural network. The decision to implement Recurrent Neural Networks (RNN) or Convolutional Neural Networks (CNN) depends on the current task. In other words, CNN is best recommended for image and video deep-learning tasks by researchers. Face recognition or image captioning are examples of CNN. On the other hand, RNN is best recommended in the scenarios of taking a sequence of words as an input and in the results producing a sequence of words. Considering all this, the RNN has been implemented in this research paper and the schematic of the encoder-decoder model is shown in Figure 2. It is an encoder-decoder framework in which the encoder extracts the text length from the new data whereas the decoder gives translation by generating input text representation. On the contrary, the RNN is the one which is more proficient in short sequences of words, but it has been seen suffering in the longer sequences from the exploding gradients and vanishing problem. Therefore, the extension of RNN has been implemented, which is capable of handling long sequences of data and specifically developed to encounter the issue of vanishing gradient during the training of RNN. The attention mechanism has been inspired by (Bahdanau et al.; 2014), especially in the RNN-LSTM text of the model. Therefore, the RNN LSTM with an attention-based mechanism has been considered in the abstractive approach.

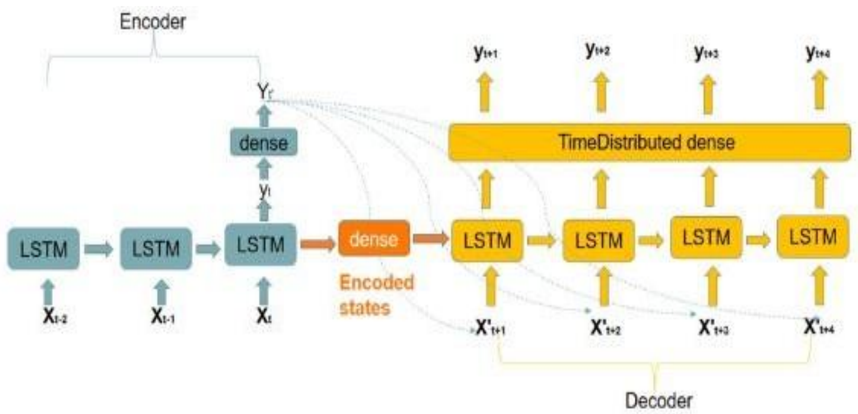

Figure 2: Sequence-to-Sequence by using Encoder-Decoder Model

4.2 Extractive Approach

It is one of the beginning steps to generate an extractive summary. The unsupervised learning strategy was performed by it. It chooses the main sentences. Preprocessing is the prominent part of any dataset as the data should be preprocessed in a required format. This is because the other phases of the model performance depend upon the preprocessing. In this stage, the irregularities and inconsistencies have been eliminated in the data, resulting in the final consequence. The total size of the news summary dataset was 102915. So, the missing values, stop words, and punctuations have been removed as they did not contribute to the text character. Lastly, the abbreviated and lowercase words have been converted to the base format by performing contraction mapping. In feature extraction, the features have been extracted from the raw form of the data. Next, the machining learning algorithm has been recalled in which data has been represented by numerical values. Extracting the word tokes is the essential step and the word tokens' frequency has been computed. One notable point is that less memory has been taken by the extracted features.

5 Implementation

The CNN/Daily Mail dataset has been used for sequence-to-sequence encoding decoding, which has 13,368 validation pairs, 28,7226 training pairs, and 11,490 test pairs. In this dataset, two 256-dimensional LSTM has been used for the bidirectional encoder as well as same for the decoder. As per the information guide network, the encoding keywords approach is similar to the encoder. Moreover, fifty thousand words of vocabulary have been used for both source and target and also word embeddings have not been pretrained because they have been learned from the beginning during the phase of training. The train has been done by using Adagrad, following the learning rate of 0.15 and 0.1 value for the initial accumulator. The size of the batch was set to 16.

On the other hand, in the attention mechanism, the decoder is mainly responsible for stepping during the process of output time steps where words have been extracted one by one. The issue with the seqseq model is the numerical summary of an input sequence in which information received from the encoder to the decoder is considered in the hidden state. Therefore, the expectation from the decoder is to utilize only one representation of vector for a long input text as an output translation. The attention is used to both translate and align. It proactively finds the relevant input sequence, which is relevant to each word in the output. With this relevant information, it picks the right output.

Finally, in the third approach, text summarization has been done by using BERT, GPT-2, and NLTK on the news summary dataset. In the BERTmodel, it took the context from both directions. In other words, it provides feasibility and ease, especially in finetuning. It has been trained on the vast batches. The GPT has been the successor from the first release of GPT. It has been scaled up version of GPT by the parameters and data size ten times. XLNET has been built on the basis of GPT and BERT shortcomings. Transformer-XL has been employed by this unsupervised learning method as its core architecture. Specifically, BERT accomplishes the best performances as compared to pretraining approaches on the basis of autoregressive language modelling.

Table 1: CNN Daily Mail News

	Train	**Dev**	**Test**
Pairs	287,227	13,368	11,490
Article Length	751	769	778
Headline Length	-	-	-
Summary Length	55	61	58

Table 2: CNN-Daily Mail News and NewsSummary Datasets ROUGE F scores

Datasets	**ROUGE-1**	**ROUGE-2**	**ROUGE-l**
NewsSummary	0.40	0.09	0.39
CNN-DailyMail News	0.14	0.01	0.13

6 Evaluation

The experiments have been carried out with evaluation and interference of models' consequences. In the research paper, the three approaches have been followed. Firstly, sequence to sequence using an encoder and decoder. Secondly, Sequence to sequence with an attention mechanism. Lastly, text summarization using GPT-2, BERT, and XLNET. Once the implementation was done, the performance measure needed to consider. So, the ROUGE metrics had been considered for evaluation metrics. It is used to measure the performance of NLP models, which considers standard metrics. A direct comparison has been done between two generated summaries such as a reference summary and model generated summary. The ROUGE scores have been divided into distinct variations: ROUGE-N, ROUGE-L, ROUGE-S, and ROUGE-W. The unigram, trigram, bigram, and higher order n-gram overlap have been measured by the ROUGE-N. Therefore, ROUGE-1 and ROUGE-2 are used to capture the bigram and unigram overlap between summaries. The longest matching words sequence has been measured by ROUGE-L.

6.1 Experiment 1

In the first experiment, the two datasets have been implemented by using an LSTM encoder and decoder. As per ROUGE, the news summary generated good results as shown in Table 2. In this dataset, 44 epochs have been used in Figure 4. On the other hand, the results for CNN/Daily Mail News dataset results were not up to the mark as compared to the news summary dataset. In the former dataset, 10 epochs have been implemented. Interestingly, the grammatical errors can be seen in the generated summaries in Figure 3.

6.2 Experiment 2

In the second experiment, the attention mechanism has been implemented to emphasize the essential part of input sequences, especially before generating the output on the news summary dataset. The stopping was done to stop the training to monitor validation loss. The ROUGE results are shown in Figure 7. The ROUGE-2 result is not good due to capturing the bigram words. The model has achieved 0.45 accuracy. This is because the majority of cases machine summary is not able to overlap with human-generated summaries, especially while comparing two at one time. One notable point is that it

1	**News:** pm narendra modi sunday paid tribute indian soldiers fought first world war war india directly involved yet soldiers fought world cause peace said india reiterates commitment towards world peace pledges work atmosphere harmony added **Original summary:** start pm modi pays tribute indian soldiers fought wwi end **Predicted summary:** start pm modi pays homage indian soldiers war war end
2	**News:** western railway ded deploy women ticket examiners coaches mumbaiahmedabad shatabdi express pilot basis sion taken two women checkers employed train trial basis following suspension two male ttes malpractices helping improve revenue collection aaa1 **Original summary:** start mumbai shatabdi allwomen ticket checkers board end **Predicted summary:** start railways train women apply coaches end
3.	**News:** amazon indias flagship campaign apni connect customers remotest parts country facing copyright issue name apni ravi jains domain registered trade apni going request amazon remove references apni jain said **Original summary:** start amazon india faces copyright issue apni campaign end **Predicted summary:** start amazon launches amazon legal notice amazon end

Figure 3: Sequence to Sequence by encoding and decoding text summarization on the news summary dataset

summarized the data, but some repetitions of words were there in the results of the summary in Figure 9. This might be unable to handle uncommon words. In other words, during the process of uncommon word found, the model might replace that word with a common word.

6.3 Experiment 3

Finally, three transformer models have been implemented. GPT-2 is the advanced version of GPT and created a good summary in Figure 8. The BERT is being utilized by Google, especially in its search algorithm and produced effective summaries. The XLNET is being implemented to enhance BERT and integrate Transformer-XL into the current model.

6.4 Discussion

In the research work, various techniques were implemented, especially to generate a con- cise summary, which will result in saving reading time for readers. While summarizing the text, the main object of the research was to achieve a representation of summarizing data before training. This can help to build factual and grammatical error-free summar- ies. In the first experiment, the salient sentences were captured from the input sequences. Unfortunately, this model failed to produce grammatical errors in most occurrences. Al- though the model was capable to identify information salient from the input sequence, poor performance was seen in generating a long output sequence. To address this issue, the attention mechanism was extended. It had the ability to generate long sequences of words from the input and emphasise the necessary sentences from the input text. In this scenario, the model was learning to a significant level and was able to fit the uncommon words in the correct places. On the contrary, this experiment was being produced repetitive summaries. The news summary dataset has shown better results as per ROUGE metrics in Table 2. The state-of-the-art methods have been explored successfully in the text summarization and were able to accomplish a summary model.

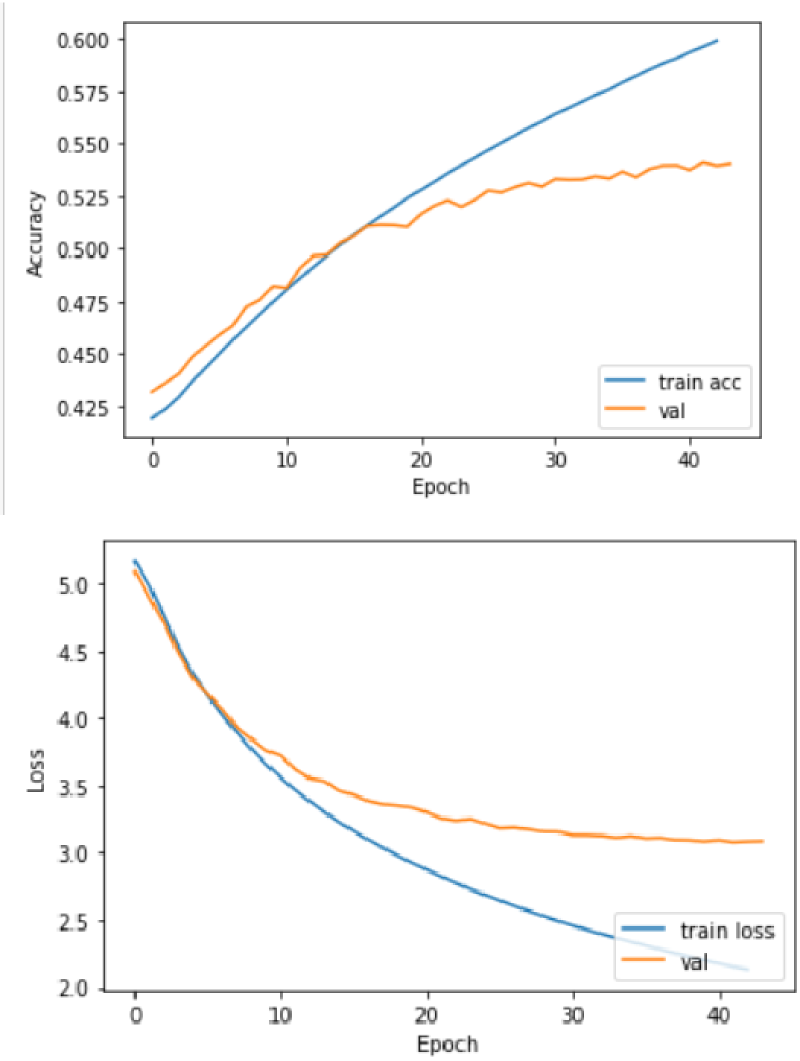

Figure 4: Epochs on Sequence to Sequence by encoding and decoding text summarization on the news summary dataset

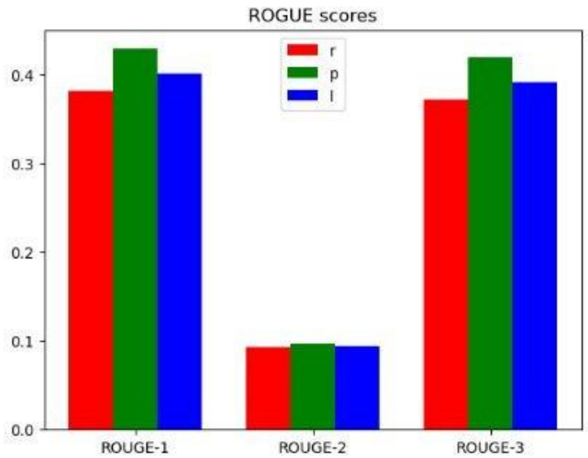

Figure 5: Encoder Decoder on News Summary Dataset

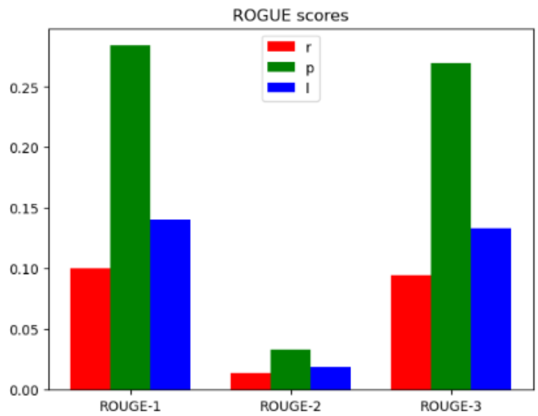

Figure 6: Encoder Decoder on CNN/Daily News Summary Dataset

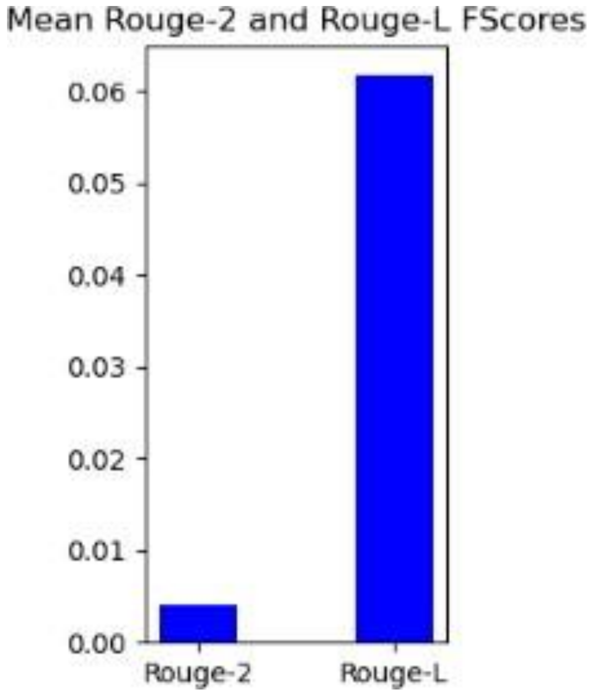

Figure 7: Encoder Decoder by attention mechanism on the News Summary Dataset

<table>
<tr>
<td>1</td>
<td>ORIGINAL TEXT:
India's food regulator Food Safety and Standards Authority of India (FSSAI) is planning to create a network to collect leftover food and provide it to the needy. It is looking to connect with organisations which can collect, store and distribute leftover food from weddings and large parties. It further added that all food must meet the safety and hygiene standards. The mother of Harshit Sharma, the class 12 Chandigarh boy who got a hoax job offer call from Google, said that the incident "devastated" his life. He got a call, after which he shared the information with the school principal, who sent out a press release. Harshit is hospitalised since Google denied giving him a job, his mother added.

BERT Summarizing Result:
India's food regulator Food Safety and Standards Authority of India (FSSAI) is planning to create a network to collect leftover food and provide it to the needy.

GPT-2 Summarizing Result:
India's food regulator Food Safety and Standards Authority of India (FSSAI) is planning to create a network to collect leftover food and provide it to the needy. Harshit is hospitalised since Google denied giving him a job, his mother added.

XLNet Summarizing Result:
India's food regulator Food Safety and Standards Authority of India (FSSAI) is planning to create a network to collect leftover food and provide it to the needy.</td>
</tr>
</table>

Figure 8: Summarizing a text by using BERT, GPT-2 and XLNET on the News Summary Dataset

1	**Pred:** man man found found found **Target:** mika singh daler mehndis elder brother amarjeet passes away **Pred:** man man found dead **Target:** sunfeast farmlite made aashirvaad atta 0 maida **Pred:** man kapoor shares pic **Target:** take tremendous pride woman colour priyanka **Pred:** delhi police arrested delhi delhi **Target**: swine flu toll rises 17 odisha **Pred:** india bans pak pak pak **Target:** rafale aircraft capable waiting iaf vice chief

Figure 9: Summarizing a text by using Sequence to Sequence with the attention mechanism

7 Conclusion and Future Work

The problem of text summarization has been solved in this problem. In most of the earlier research, the main aim was to generate a summary rather than omitting the point of meaningful, generating factual, and non-repetitive summaries. The three approaches have been followed in the overall research paper. The two data types have been compared in terms of ROGUE scores. In these scores, it was found that the news dataset performed the best as compared to the CNN-Dailymail news dataset in LSTM encoding and decoding. On the other hand, the attention mechanism by LSTM was also implemented in the second approach and it got 0.45 accuracy with repetitive words in the generated summary. This is because the model is unable to handle uncommon words as the model replaces that word with a common word while in the process of modelling. The third is followed by summarizing the data using GPT-2, BERT, and NLTK, which generated the best summaries. The major challenge part of text summarization is the lack of datasets available on public platforms. Moreover, the model takes much time on the local machine to build the model and run the epochs. The text summarization can be utilized in various other domains such as healthcare, Blogs, customer support, and journals for future work. This will result in minimizing the reader's time and efficiently retrieving the information. In the present scenarios, our model can be trained as well as validated by using different domains of the datasets.

References

Abolghasemi, M., Dadkhah, C. and Tohidi, N. (2022). Hts-dl: Hybrid text summarization system using deep learning, *2022 27th International Computer Conference, Computer Society of Iran (CSICC)*, IEEE, pp. 1–5.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B. and Kochut, K. (2017). Text summarization techniques: a brief survey, *arXiv preprint* *arXiv:1707.02268* .

Bahdanau, D., Cho, K. and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate, *arXiv preprint arXiv:1409.0473* .

Baykara, B. and Güngör, T. (2022). Turkish abstractive text summarization using pre-trained sequence-to-sequence models, *Natural Language Engineering* pp. 1–30.

Chopra, S., Auli, M. and Rush, A. M. (2016). Abstractive sentence summarization with attentive recurrent neural networks, *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pp. 93–98.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* .

Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J., Zhou, M. and Hon, H.-W. (2019). Unified language model pre-training for natural language understanding and generation, *Advances in Neural Information Processing Systems* **32**.

Fang, X., Xiao, T. and Liu, S. (2022). Research on text summarization generation and classification method based on deep learning, *2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, IEEE, pp. 757–762.

Fitrianah, D. and Jauhari, R. N. (2022). Extractive text summarization for scientific journal articles using long short-term memory and gated recurrent units, *Bulletin of Electrical Engineering and Informatics* **11**(1): 150–157.

Gupta, S. and Gupta, S. K. (2019). Abstractive summarization: An overview of the state of the art, *Expert Systems with Applications* **121**: 49–65.

Hermann, K. M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M. and Blunsom, P. (2015). Teaching machines to read and comprehend, *Advances in neural information processing systems* **28**.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning, *nature* **521**(7553): 436–444.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V. and Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, *arXiv preprint arXiv:1910.13461* .

Li, P., Lam, W., Bing, L. and Wang, Z. (2017). Deep recurrent generative decoder for abstractive text summarization, *arXiv preprint arXiv:1708.00625* .

Liang, Z., Du, J. and Li, C. (2020). Abstractive social media text summarization using selective reinforced seq2seq attention model, *Neurocomputing* **410**: 432–440.

Maynez, J., Narayan, S., Bohnet, B. and McDonald, R. (2020). On faithfulness and factuality in abstractive summarization, *arXiv preprint arXiv:2005.00661* .

Nada, A. M. A., Alajrami, E., Al-Saqqa, A. A. and Abu-Naser, S. S. (2020). Arabic text summarization using arabert model using extractive text summarization approach, *Int.*

J. Academic Inf. Syst. Res. **4**(8): 6–9.

Nallapati, R., Zhou, B., Gulcehre, C., Xiang, B. et al. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond, *arXiv preprint arXiv:1602.06023* .

Qiu, D. and Yang, B. (2022). Text summarization based on multi-head self-attention mechanism and pointer network, *Complex & Intelligent Systems* **8**(1): 555–567.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P. J. et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer., *J. Mach. Learn. Res.* **21**(140): 1–67.

Rush, A. M., Chopra, S. and Weston, J. (2015). A neural attention model for abstractive sentence summarization, *arXiv preprint arXiv:1509.00685* .

See, A., Liu, P. J. and Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks, *arXiv preprint arXiv:1704.04368* .

Sethi, P., Sonawane, S., Khanwalker, S. and Keskar, R. (2017). Automatic text summarization of news articles, *2017 International Conference on Big Data, IoT and Data Science (BID)*, IEEE, pp. 23–29.

Shi, Y., Meng, J. and Wang, J. (2019). Seq2seq model with rnn attention for abstractive summarization, *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*, pp. 348–353.

Song, K., Tan, X., Qin, T., Lu, J. and Liu, T.-Y. (2019). Mass: Masked sequence to sequence pre-training for language generation, *arXiv preprint arXiv:1905.02450* .

Wilt, C. M., Thayer, J. T. and Ruml, W. (2010). A comparison of greedy search algorithms, *third annual symposium on combinatorial search*.

Xu, S., Zhang, X., Wu, Y. and Wei, F. (2022). Sequence level contrastive learning for text summarization, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, pp. 11556–11565.

Zhang, Y., Li, D., Wang, Y., Fang, Y. and Xiao, W. (2019). Abstract text summarization with a convolutional seq2seq model, *Applied Sciences* **9**(8): 1665.

Zhou, Q., Yang, N., Wei, F. and Zhou, M. (2017). Selective encoding for abstractive sentence summarization, *arXiv preprint arXiv:1704.07073* .