Abstractive Text Summarization

Namratha Sampath Kumar
Department of Applied Data Science
namratha.sampathkumar@sjsu.edu

Nandini Sreekumran Nair
Department of Applied Data Science
nandini.sreekumarannair@sjsu.edu

Pooja Manjunatha
Department of Applied Data Science
pooja.manjunatha@sjsu.edu

Priya Varahan
Department of Applied Data Science
priya.varahan@sjsu.edu

***Abstract*—In the domain of document summarization, manual efforts are notably time-consuming and labor-intensive. The integration of deep learning techniques within Natural Language Processing (NLP) frameworks presents a viable solution to enhance the efficiency of summarization tasks. The primary objective is to utilize advanced deep learning models like Long Short-Term Memory (LSTM), Pointer-Generator Network (PGN), Bidirectional Long Short-Term Memory (BiLSTM), and Bidirectional Encoder Representations from Transformers (BERT) to create an abstract from extensive documents with minimal human intervention, thereby ensuring rapid comprehension which is crucial for processing lengthy documents effectively.Among the models evaluated, the BERT model demonstrated notable performance, achieving a high BLEU score indicating that BERT not only generates summaries that are more fluent but also captures better semantic content of the source documents.The conclusions and observations derived from these evaluations of the models are critical for iterating on model design and deployment strategies, aiming to optimize both accuracy and relevance in generated summaries.**

***Index Terms*—Summerization, NLP, LSTM, BiLSTM, BERT**

I. Introduction

In the contemporary digital era, the voluminous data available online presents significant challenges in the manual summarization of extensive texts, which is both time-consuming and prone to inaccuracies. This project proposes a robust solution leveraging advanced deep learning technologies to enhance the abstractive summarization of lengthy documents. Utilizing datasets with long articles and their respective summaries our approach aims to efficiently extract and condense key information with minimal human intervention.

Sophisticated models, Long Short-Term Memory (LSTM), Pointer-Generator Network, Bidirectional Long Short-Term Memory (BiLSTM), and Bidirectional Encoder Representations from Transformers (BERT) are chosen for their respective capabilities to handle nuances and continuity in long texts which traditional neural networks and rule-based methods fail to capture. LSTMs are employed to identify long-range dependencies, BiLSTMs analyze texts from dual perspectives, Pointer-Generator Networks address out-of-vocabulary words and repetitive patterns, and BERT optimizes contextual embeddings for superior summary generation.

Each component of the system is fine-tuned through a series of preprocessing steps tailored to the specific attributes of the algorithm, including tokenization, noise reduction, vocabulary creation, and data augmentation. The efficacy of our system is quantitatively assessed using performance metrics such as Rouge-1, Rouge-2, Rouge-L, and BLEU. We establish the T5 by Raffel et al. (2020) as our baseline for measuring advancements over existing state-of-the-art summarization techniques.

The results anticipate that this integrated approach will significantly enhance the accuracy and relevance of automated text summarization, offering profound implications for applications in news digestion, report generation, and other areas requiring rapid comprehension of extensive documents. Implementation of the Graphical User Interface (GUI) enhances the usability, accessibility, and overall user experience. The aim of the project is to devolop a system that surpasses the performance of existing summarization systems.

II. Related Work

The field of deep learning has seen significant advancements in text summarization as evidenced by a range of studies utilizing various datasets and models. Nallapati et al. (2016) employed the Gigaword and DUC corpus along with the CNN/Daily Mail corpus to train models such as the Encoder-Decoder RNN, Generator-Pointer, and Feature-rich Encoder, with the Feature-rich Encoder showing superior performance on the ROUGE-1, ROUGE-2, and ROUGE-L scores. Similarly, Chopra et al. (2016) utilized the Gigaword corpus and DUC-2004 data to evaluate models like the Bag-of-Words, Convolutional TDNN, and RAS-Elman, with the latter outperforming others in terms of ROUGE scores.

Karmakar et al. (2021) focused on Indian regional languages using datasets like News please library, kaggle hindi text corpus, and github marathi news. They found that Attention-based LSTM models performed best, showcasing the effectiveness of attention mechanisms in handling language nuances. Song et al. (2018) studied the performance of LSTM-CNN-based models using the CNN and DailyMail datasets, achieving impressive ROUGE scores, highlighting the synergy between LSTM and CNN architectures for deep learning tasks.

Jiang et al. (2021) explored hybrid models combining Bi-LSTM with attention mechanisms using the LCSTS short-text corpus and TTnews long-text corpus, where complex models integrating multiple strategies showed the best results. In a more focused study, Shahin et al. (2020) analyzed the efficacy of Bi-directional LSTM on Bangla news articles, showing a significant improvement over traditional ANNs and LSTMs in terms of accuracy.

Other notable studies include Mahadevaswamy et al. who achieved a high accuracy with Bidirectional LSTM on the Amazon Product Review dataset, and Pabbi et al. (2021) who demonstrated the effectiveness of Bi-directional LSTM for summarizing Amazon food reviews. Moreover, Darapaneni et al. (2023) and Gupta and Patel (2021) utilized BERT for advanced text summarization tasks, showing that contextual understanding significantly affects summarization quality.

See et al. (2017) introduced a Pointer-Generator Network for the CNN/Daily Mail dataset, which was particularly notable for its ability to blend extraction and abstraction to enhance summarization outcomes. Similarly, Boutkan et al. (2019) and Shobana and Murali (2021) expanded on this work by integrating attention mechanisms and pointer-generator models, further pushing the boundaries of what's achievable in abstractive text summarization.

Wijayanti et al. (2021) compared performances of BertSum and Pointer Generator Network with NeuralSum which is the aim of the paper. The dataset used is an Indonesian news article. ROGUE is the evaluation metric used and the results show that BertSum had a low ROGUE score and struggles in generating abstract summaries. PGN performs the best but it has context issues and NeuralSum offers faster training but can produce a blank summary when no sentence is classified as important.

Raman and Kambli (2022) proposed a methodology for text summarisation using Long Short-Term Memory Model and Transformer with Attention. The optimizer used for training and evaluation is ADAM optimizer.The dataset used is CNN/Daily Mail. The results show that the Transformer Model performs better than LSTM with a greater evaluation metrics which is accuracy. The highest evaluation accuracy achieved by the Transformer Model is 85.12% after 1500 steps and for the sequential model it is 65.27% achieved after 12 steps and then the accuracy drops. It is concluded that LSTM model performs better with small datasets.

Qiu and Jin (2022) proposed an extractive text summarisation model using fine tuned BERT and k-means clustering. ROUGE is the evaluation metrics used and a human based evaluation method is used as well. BERT is used to generate sentence embeddings and is given as input to the k-means method. BERT based models perform better and have a better ROUGE score compared to non-BERT based. Fine tuned BERT model performs the best with a mean ROUGE-L score of 4.9%.

Jeyakarthic and Leoraj(2024) built a Knowledge-based BERT summariser using the ROUGE evaluation Metrics. The dataset used is CNN/Daily Mail which consists of news articles. BERT is used to generate embeddings after data collection, pre-processing and corpus building. A Knowledge Graph is built using these BERT Embeddings to provide a contextual summarisation. This method leverages a corpus to enhance text summarisation. ROUGE-1 score is 0/913, ROUGE-2 is 0.892, ROUGE-L is 0.932, F-1 score is 0.92 which shows that the model performs well.

These studies collectively illustrate the vibrant and evolving landscape of text summarization in deep learning, emphasizing the role of innovative architectures and complex datasets in pushing forward the capabilities of AI in understanding and processing human language.

III. METHODOLOGY

A. Dataset Description

a) Data Collection: CNN /Daily dataset(528 MB), New York Times dataset (3 GB), Wikihow dataset(590MB) are the 3 dataset that can be used for the Abstractive Text Summarization Project. The datasets contain 3 main fields, a unique id, article and summary from the news as shown in Figure 1.

	A	B	C
1	id	article	highlights
2	0001d1afc246a7964130f43ae940af6bc6c57f01	By . Associated Press . PUBLISHED: . 14:11 E	Bishop John Folda, of North Dakota, is taking time off after being diagnosed .
3	0002095e55fcbd3a2f366d9bf92a95433dc305ef	(CNN) -- Ralph Mata was an internal affairs	Criminal complaint: Cop used his role to help cocaine traffickers .
4	00027e965c8264c35cc1bc55556db388da82b07f	A drunk driver who killed a young woman in a	Craig Eccleston-Todd, 27, had drunk at least three pints before driving car .
5	0002c17436637c4fe1837c935c04de47adb18e9a	(CNN) -- With a breezy sweep of his pen Pre	Nina dos Santos says Europe must be ready to accept sanctions will hurt
6	0003ad6ef0c37534f80b55b4235108024b407f0b	Fleetwood are the only team still to have a 1	Fleetwood top of League One after 2-0 win at Scunthorpe .
7	0004306354494f090ee2d7bc5ddbf80b63e80de6	He's been accused of making many a fashior	Prime Minister and his family are enjoying an Easter break in Lanzarote .
8	0005d61497d21ff37a17751829bd7e3b6e4a7c5c	By . Daily Mail Reporter . PUBLISHED: . 01:1	NBA star calls for black and Hispanic communities to get tested .
9	0006021f772fad0aa78a977ce4a31b3faa6e6fe5	By . Daily Mail Reporter . This is the moment	London Midland service had been pulling into Telford Station in Shropshire .
10	00083697263e215e5e7eda753070f08aa374dd45	There are a number of job descriptions waiti	Tony Pulis believes Saido Berahino should look up to Darren Fletcher .
11	000940f2bb357ac04a236a232156d8b9b18d1667	Canberra, Australia (CNN) -- At first glance,	Black box data from Flight 370 could be analyzed at a laboratory in Australia .
12	0009ebb196751174162992 6ef9f5faea2bb6be24	By . Ellie Zolfagharifard . Take a look at a maj	The distortion is the result of the Mercator map which was created in 1596 to
13	000c835555db62e319854d9f8912061cdca1893e	Two lawyers representing a woman who . cla	Alan Dershowitz has filed defamation suits against two other U.S lawyers .
14	000ca3fc9d877f8d4bb2ebd1d6858c69be571fd8	It's the moment every pet owner dreads - wh	Sarah Ernhart, the owner of Sarah Beth Photography in Minneapolis, created
15	000cd1ee0098c4d510a03ddc97d11764448ebac2	Louis van Gaal said he had no option but to s	Manchester United beat Southampton 2-1 at St Mary's on Monday night .
16	000e009f6b1d954d827c9a550f3f24a5474ee82b	(CNN) -- One can hardly read the news thes	U.S. corporations merge with foreign companies, move their headquarters .
17	001097a19e2c96de11276b3cce11566ccfed0030	For most people, it has become a travel ess	Half of Brits admit to checking work e-mails while on holiday, while a third
18	0010c870d3fc53ea7f2a4a50f6496dc2df17e02f	By . Margot Peppers . Nigerian and Cameroc	Dencia's comment is hypocritical considering she recently courted controver
19	00120f91cfcab17bac165f7a4719019a628a9db3	Britain and the West must brace themselve	Ex-military chiefs suggested .
20	00128f1ba30d5e9e0f17df83285a1bc2072e2f01	A woman has been charged with reckless ma	Claudia Yanira Hernandez Soriano, 25, and Juan Francisco Martinez Rojas,
21	00142065a6c92f788eacd40c9023184808a7e2d1	Beirut (CNN) -- Syria carried out an airstrik	Airstrike kills nine Syrians in refugee camp, state media reports .
22	0015194573f9b4430319683cde41e4aa17092a9d	An Australian citizen, who has awaited trial	Roman Shilov has been detained by Russian authorities without trial since
23	00156d9892fb27f1d2e100cbdd8a3997f8273781	Everton are still looking to add two new play	Everton are looking to add two new players to their squad .
24	001732b374f362d3961a510da315601e4b5e7e84	By . Daily Mail Reporter . Last updated at 1:2	Police brought in over Twitter threats believed to be made by man in Russia .
25	001789cf9b865dcac3d9fc032a6b1533e3318eda	Last updated at 3:31 PM on 19th July 2011 .	Amish population in New York has grown by a third in the past two years .
26	00189f37b1c8bdc2b132b40270bb28ffcc622af1	Glen Johnson looks destined to leave Anfielc	Full back scored late winner as Liverpool beat Stoke 1-0 on Saturday .
27	001a6162391594e2a8607fba135bdfa154e57904	Would you believe that these breathtaking p	Adventurer Tomasz Furmanek photographs Norwegian fjords from a kayak .
28	001adf6209be103cb198b8599f236b4d5760a5fe	By . Daily Mail Reporter . PUBLISHED: . 19:4	Ashleigh Davis, 32, arrested last year after a fight with another woman .

Fig. 1. Sample of Dataset.

b) Exploratory Data Analysis: The dataset consists of 3211971 rows and 3 columns. The columns contain no missing values as shown in Figure 2.

```
Missing Values:
 id
article        0
highlights     0
dtype: int64

Combined data shape: (311971, 3)
```

Fig. 2. Shape and Missing Values in the Dataset.

The histogram in Figure 3 shows the range of length of words. The words in the article column are in the range of 400 to 500 words. The words in the highlights column fall in the range of 70 - 80 words. This helps in understanding the maximum sequence length while defining hyper parameters for the respective models.

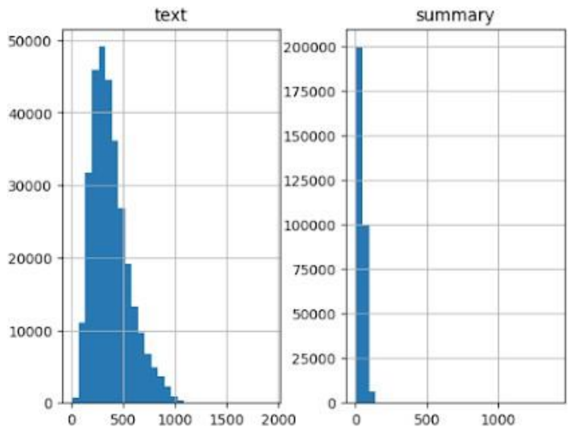


Fig. 3. Histogram to show the length of words

c) Data Pre-processing: The text data has to be pre-processed for Model Training. The text before and after pre-processing is shown in Figure 4 and Figure 5 respectively. The steps are:

Expanding Contractions: Contractions are short form of words. This is created by removing certain letters or sounds. This is an important step while performing NLP tasks. For example, "can't" is a contraction is "cannot" and these words are expanded for clarity in the text.

Lemmatization: This is the process of reducing the words to it's root word or base form. This reduces the dimensionality and complexity of the data. It involves understanding the context of words. For example, "The Horse is running slowly" is converted to "The Horse is running slow" after being Lemmatized.

Lower Casing: Converting all text to lowercase as it maintains consistency and reduces the vocabulary size. For example, "Orange" and "orange" are considered to be the same after Lower Casing

Removing Punctuation: Punctuation marks do not help the model understand the meaning of a sentence while performing NLP tasks. Therefore, they are removed to simplify the data. Commas, Periods, and Exclamation Marks are all removed from the text. For example, "I am good!" becomes "I am good" after the removal of punctuation.

Removal of Stopwords: Stopwords do not carry any meaning and are removed to focus more on the meaningful words. They include articles. prepositions, conjunctions and words like "and", "the" and "is". This can help reduce noise and improve the performance of the model. For example, "This is an example of preprocessing" becomes "example text preprocessing" after the removal of Stopwords.

d) Data Transformation: Some Transformation steps have been applied to the pre-processed text. Figure 6 shows text before and after transformation. The texts are converted to integers as the models take numeric inputs. The transformation steps are described below.

Tokenization: The process of words or phrases being split

```
cleaned_text[:2]
```

```
['associated press published est october updated est october bishop fargo catholic diocese north dakota exposed pot
entially hundreds church members fargo grand forks jamestown hepatitis virus late september early october state hea
lth department issued advisory exposure anyone attended five churches took communion bishop john folda fargo cathol
ic diocese north dakota exposed potentially hundreds church members fargo grand forks jamestown hepatitis state imm
unization program manager molly howell says risk low officials feel important alert people possible exposure dioces
e announced monday bishop john folda taking time diagnosed hepatitis diocese says contracted infection contaminated
food attending conference newly ordained bishops italy last month symptoms hepatitis include fever tiredness loss a
ppetite nausea abdominal discomfort fargo catholic diocese north dakota bishop located',
 'ralph mata internal affairs lieutenant miami dade police department working division investigates allegations wro
ngdoing cops outside office authorities allege year old longtime officer worked drug trafficking organization help
plan murder plot get guns criminal complaint unsealed district court new jersey tuesday accuses mata also known mil
k man using role police officer help drug trafficking organization exchange money gifts including rolex watch one i
nstance complaint alleges mata arranged pay two assassins kill rival drug dealers killers would pose cops pulling t
argets shooting according complaint ultimately decided move forward murder plot mata still received payment setting
meetings federal prosecutors said statement complaint also alleges mata used police badge purchase weapons drug tra
ffickers mata according complaint used contacts airport transport weapons carry luggage trips miami dominican repub
lic court documents released investigators specify name drug trafficking organization mata allegedly conspired says
organization importing narcotics places ecuador dominican republic hiding inside shipping containers containing pal
lets produce including bananas organization distributing narcotics new jersey elsewhere complaint says authorities
arrested mata tuesday miami gardens florida immediately clear whether mata attorney police officials could immediat
ely reached comment mata worked miami dade police department since including directing investigations miami gardens
working lieutenant unit miami international airport according complaint since march working internal affairs divisi
on mata faces charges aiding abetting conspiracy distribute cocaine conspiring distribute cocaine engaging monetary
transactions property derived specified unlawful activity scheduled appear federal court florida wednesday convicte
d mata could face life prison cnn suzanne presto contributed report']
```

Fig. 4. Text Before Pre-Processing


```
cleaned_text[:2]
```

```
['associated press published est october updated est october bishop fargo catholic diocese north dakota exposed pot
entially hundreds church members fargo grand forks jamestown hepatitis virus late september early october state hea
lth department issued advisory exposure anyone attended five churches took communion bishop john folda fargo cathol
ic diocese north dakota exposed potentially hundreds church members fargo grand forks jamestown hepatitis state imm
unization program manager molly howell says risk low officials feel important alert people possible exposure dioces
e announced monday bishop john folda taking time diagnosed hepatitis diocese says contracted infection contaminated
food attending conference newly ordained bishops italy last month symptoms hepatitis include fever tiredness loss a
ppetite nausea abdominal discomfort fargo catholic diocese north dakota bishop located',
 'ralph mata internal affairs lieutenant miami dade police department working division investigates allegations wro
ngdoing cops outside office authorities allege year old longtime officer worked drug trafficking organization help
plan murder plot get guns criminal complaint unsealed district court new jersey tuesday accuses mata also known mil
k man using role police officer help drug trafficking organization exchange money gifts including rolex watch one i
nstance complaint alleges mata arranged pay two assassins kill rival drug dealers killers would pose cops pulling t
argets shooting according complaint ultimately decided move forward murder plot mata still received payment setting
meetings federal prosecutors said statement complaint also alleges mata used police badge purchase weapons drug tra
ffickers mata according complaint used contacts airport transport weapons carry luggage trips miami dominican repub
lic court documents released investigators specify name drug trafficking organization mata allegedly conspired says
organization importing narcotics places ecuador dominican republic hiding inside shipping containers containing pal
lets produce including bananas organization distributing narcotics new jersey elsewhere complaint says authorities
arrested mata tuesday miami gardens florida immediately clear whether mata attorney police officials could immediat
ely reached comment mata worked miami dade police department since including directing investigations miami gardens
working lieutenant unit miami international airport according complaint since march working internal affairs divisi
on mata faces charges aiding abetting conspiracy distribute cocaine conspiring distribute cocaine engaging monetary
transactions property derived specified unlawful activity scheduled appear federal court florida wednesday convicte
d mata could face life prison cnn suzanne presto contributed report']
```


```
cleaned_summary[:2]
```

```
['bishop john folda of north dakota is taking time off after being diagnosed he contracted the infection through co
ntaminated food in italy church members in fargo grand forks and jamestown could have been exposed',
 'criminal complaint cop used his role to help cocaine traffickers ralph mata an internal affairs lieutenant allege
dly helped group get guns he also arranged to pay two assassins in murder plot complaint alleges']
```

Fig. 5. Text After Pre-Processing

```
array(['england become overcrowded major country europe population growth rapid four times many people soon
crammed france twice many germany england overtaken netherlands become second tiny malta densely populated
nation eu squeeze england overtaken netherlands become densely populated major nation eu next years gap widen
germany france holland either decline grow slowly house commons figures based data uk eu statistical agencies
show huge impact labour open door immigration policy estimated people squeezed every square kilometre england
compared tony blair took power las controls figures based data uk eu statistical agencies show huge impact labour
open door immigration policy revelations fuel debate immigration especially uk opening borders romanian
bulgarian workers new year day james clappison tory mp obtained figures said last labour government england
green pleasant land became england green crowded land number people expected living per square kilometre
country reasons never properly explained labour instigated policy massive expansion immigration fear must future
labour government would sir andrew green chairman migrationwatch think tank said per cent immigrants uk
headed england rapidly growing population density inevitable consequence labour mass immigration nearly four
million years added already see pressure maternity units primary schools less visible pressure housing already
crisis need build houses day next years simply new immigrants families house commons report says number
people living every square kilometre england rise astonishing people living square kilometre equivalent figure
france germany netherlands england also three times packed poland estimated one million arrivals labour
originated research raises concerns uk infrastructure cope increased pressure schools hospitals roads clash
liberal democrat business secretary vince cable left likened david cameron policies enoch powell notorious rivers
blood speech large numbers packed country also affect water power supplies increase pressure build green spaces
david cameron pressure confront electoral threat posed ukip changed law prevent eu migrants claiming benefits
first three months following arrival wake open door immigration policies deliberately pursued new labour england
crowded country europe twice many people crammed every square kilometre germany four times france three
decades unless significant tightening border controls almost people living square kilometre compared labour came
power vince cable colleagues left disgracefully suggest somehow racist worry immigration revelations show
nothing race everything schools social services nhs housing possibly expected cope unprecedented pressure
officials say want reduce pull factor uk last weekend tensions two coalition government parties boiled liberal
democrat business secretary vince cable likened tory policies enoch powell notorious rivers blood speech office
national statistics already warned britain must make room almost million people next years equivalent building city
even larger london increase mainly result immigration high migrant birthrates push numbers million home office
spokesman said immigration brought benefits uk welcome people want come contribute economy society however
important control immigration effect social cohesion public services jobs wages figures rest uk predicted scotland
per square kilometre wales northern ireland malta figure expected',
```

Converted to:

```
array([[  136,   252, 14153, ...,     0,     0,     0],
       [10258, 29405,   109, ...,     0,     0,     0],
       [   23,   125,  1992, ...,     0,     0,     0],
       ...,
       [  849,  9688,   191, ...,     0,     0,     0],
       [23115, 45137,   109, ...,     0,     0,     0],
       [ 5560, 44202,   531, ...,     0,     0,     0]], dtype=int32
```

Fig. 6. Text after Transformation

into tokens is called Tokenization. It is an important step for Deep Learning Models as the models can now easily read the data.

Segment Embeddings: They are used to understand the different sentences in a text. They are added to token embeddings to show which sentence a token belongs to. This helps the model understand the context of the text and it's relationship.

Positional Encodings: They are used to give information about the position of the token to the model. This is specifically used in Transformers model since it is not sequence to sequence to understand the position of tokens.

B. Proposed Model Architecture

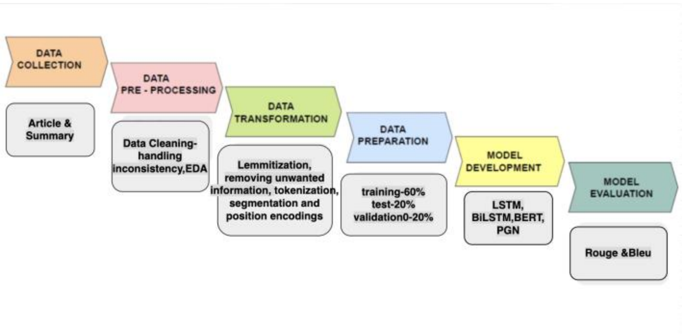
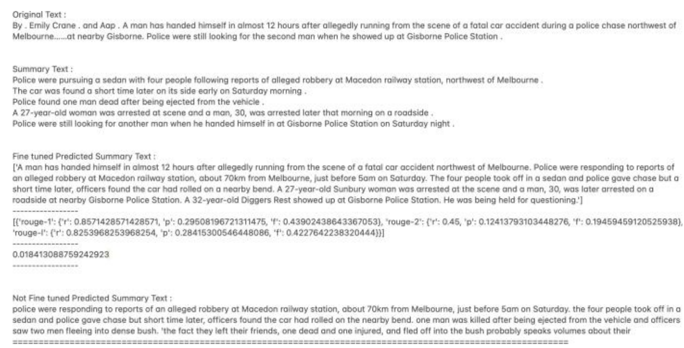


Fig. 7. High Level Data Flow Architecture

Figure 7 shows the high level architecture of the abstractive text summarisation model. The data is first collected and is pre-processed, transformed and split into training set, test set and validation set in a ratio of 60%, 20% and 20% respectively.Next steps inc;due Model Development and Model Evaluation which is explained in the next section. The Models include Flan-T5, Long Short Term Memory(LSTM),Bidirectional Long Short Term Memory(Bi-LSTM), Bidirectional Encoder Representations from Transformer(BERT) and Pointer-Generator Network. These Models are compared and evaluated using ROUGE and BLEU scores.

a) Pretrained T5: The Text-to-Text Transfer Transformer(T5) model performs Natural Language Processing(NLP) task by converting them into sequence-to-sequence tasks in the encoder-decoder variant. The text is used for encoder input and the decoder must generate the label as normal text.T5 architecture is the original Transformer architecture that is trained on the large crawled C4 dataset. Masked language modeling is used for training. The largest model of T5 class has 11 billion parameters that achieved State of The Art (SOTA) results on several benchmarks.The AutoTokenizer and AutoModelWithLMHeadclasses from the transformers library are being imported, which is used for creating an instance of a pre-trained transformer-based model and its tokenizer. Figure 8 demonstrates the input and output of the T5 Model.

b) LSTM: Long Short Term Memory is a Recurrent Neural Network type of model, as the name suggests it can learn and remember long sequences. It combats the vanishing



```
Original Text :
By . Emily Crane . and Aap . A man has handed himself in almost 12 hours after allegedly running from the scene of a fatal car accident during a police chase northwest of Melbourne......at nearby Gisborne. Police were still looking for the second man when he showed up at Gisborne Police Station .

Summary Text :
Police were pursuing a sedan with four people following reports of alleged robbery at Macedon railway station, northwest of Melbourne .
The car was found a short time later on its side early on Saturday morning .
Police found one man dead after being ejected from the vehicle .
A 27-year-old woman was arrested at scene and a man, 30, was arrested later that morning on a roadside .
Police were still looking for another man when he handed himself in at Gisborne Police Station on Saturday night .

Fine tuned Predicted Summary Text :
['A man has handed himself in almost 12 hours after allegedly running from the scene of a fatal car accident northwest of Melbourne. Police were responding to reports of an alleged robbery at Macedon railway station, about 70km from Melbourne, just before 5am on Saturday. The four people took off in a sedan and police gave chase but a short time later, officers found the car had rolled on a nearby bend. A 27-year-old Sunbury woman was arrested at the scene and a man, 30, was later arrested on a roadside at nearby Gisborne Police Station. A 32-year-old Diggers Rest showed up at Gisborne Police Station. He was being held for questioning.']
----------------
[{'rouge-1': {'r': 0.8571428571428571, 'p': 0.29508196721311475, 'f': 0.43902438643367053}, 'rouge-2': {'r': 0.45, 'p': 0.12413793103448276, 'f': 0.19459459120525938}, 'rouge-l': {'r': 0.8253968253968254, 'p': 0.28415300546448086, 'f': 0.4227642238320444}}]
----------------
0.018413088759242923
----------------

Not Fine tuned Predicted Summary Text :
police were responding to reports of an alleged robbery at Macedon railway station, about 70km from Melbourne, just before 5am on Saturday. the four people took off in a sedan and police gave chase but short time later, officers found the car had rolled on the nearby bend. one man was killed after being ejected from the vehicle and officers saw two men fleeing into dense bush. 'the fact they left their friends, one dead and one injured, and fled off into the bush probably speaks volumes about their
==========================================================================================
```

Fig. 8. Input and Output of the T5 Model

gradient problem. The model consists of three gates, input gate, forget gate and output gate that controls the flow of information. The architecture of the abstractive text summarisation model can be seen in Figure 9. This is an encoder decoder model for texts with a long sequence. It integrates attention layer and embedding layers for the input and output sequence. Every word in the input is mapped to 100 dimensions in the embedding layer. The output layer captures the context and dependency of the input data. The role of attention layer is to calculate the weights and scores to enhance the performance of the model. This layer allows the model to focus on important parts of the input sequence. The output of final LSTM layer and attention layer is concatenated. The final layer is the time distributed dense layer to generate probabilities and predict the next words to produce accurate summaries.

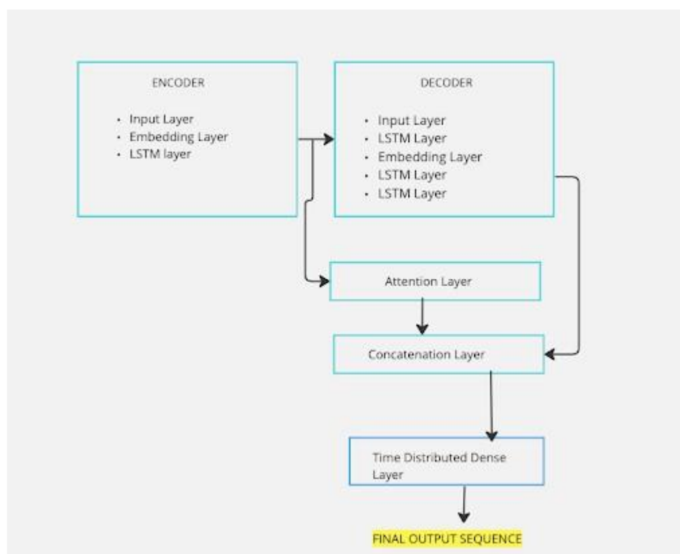


Fig. 9. LSTM Architecture

c) Bi-LSTM: The Bi-Directional Long Short Term Memory is also an encoder and decoder based model for long sequence texts. The architecture is the same of LSTM and contains the same layers as shown in Figure 9 and can be seen in Figure 10. Attention layers and Embedding Layers are used here as well. This model performs better and produces better results as compared to LSTM as it is Bi-Directional and can remember longer sequences compared to the latter.

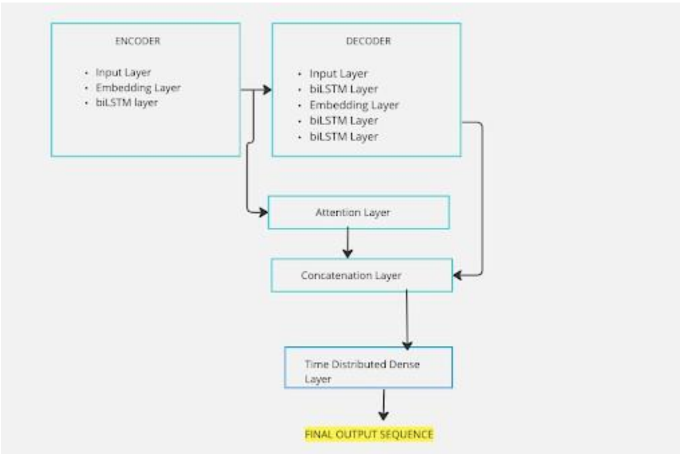


Fig. 10. Bi-LSTM Architectire

d) Pointer Generator Network: The Pointer-Generator Network (PGN) architecture is designed to produce abstractive summaries by leveraging a dual mechanism that allows it to either copy words directly from the input text or generate new words from a predefined vocabulary. This approach effectively handles out-of-vocabulary (OOV) words and enhances the overall quality of the summaries. The model architecture consists of encoder with four-layer LSTM with 256 hidden units per layer, utilizing a tanh activation function and a dropout rate between 0.2. It processes the input text, tokenized into words, and converts it into a sequence of hidden states, capturing the contextual information of 500 words. The attention mechanism is a single-layer attention module enabling the decoder to focus on the most relevant parts of the input text. The decoder is also a four-layer LSTM with 256 hidden units per layer and a tanh activation function, generating the summary word-by-word based on the hidden states from the encoder and the attention distribution.

The pointer- generator mechanism employs a Softax activation function to decide at each decoding step whether to generate a word from the predefined vocabulary or to copy a word from the input text, effectively handling out-of-vocabulary words. The coverage mechanism tracks the attention distribution over time with a coverage vector serving as an additional input to the attention mechanism, helping to prevent the model from repetition of the source text and thus ensuring comprehensive and concise summaries. Additionally the model employs the Adam optimize for handling sparse gradients. The learning rate is initially set to 0.001, which provides a good balance between convergence speed and training stability.

Figure 11 shows the input long text and the summary generated by the PGN.

e) BERT: The model is built on a two-stage decoding process utilizing BERT's pre-trained contextualized language

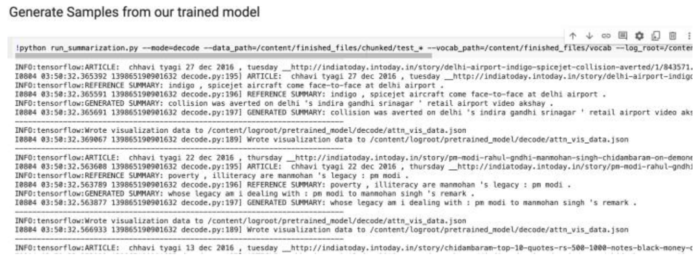


Fig. 11. Input Text and the Summary Generated by BERT

models. The integration of multi-head attention, fine-tuning, and regularization techniques improves summary representations by addressing the limitations of previous abstractive methods.

In the first decoder stage, the document embeddings generated by the BERT encoder are fed into a multi-head attention decoder with 12 layers of multi-head attention and feed-forward neural networks. Activation function ReLU maintains gradients for positive inputs, allowing the model to learn effectively during backpropagation. Draft summary tokens generated is masked one at a time and reprocessed through BERT decoder to generate contextualized embeddings. Another multi-head attention decoder with 12 layers then refines these embeddings to produce the final summary. Adam optimizer is used for training with a learning rate of 0.0001.

The model leverages Cross attention to utilize BERT's rich contextual information and employs Fine-tuning to optimize language model weights for high-quality summaries. It also incorporates Vocabulary handling mechanisms to manage out-of-vocabulary words, ensuring fluent and relevant summaries. Figure 12 given below represents the BERT Model overview. Figure 13 shows the input long text and the summary generated by the BERT Model.

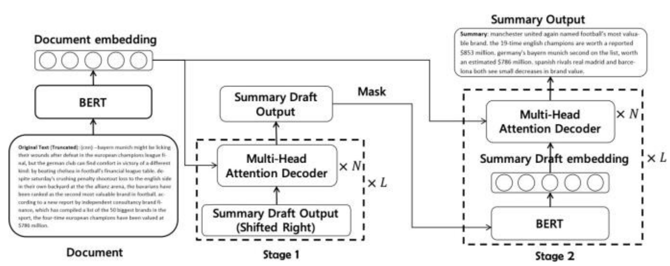


Fig. 12. BERT Model Overview

Fig. 13. Input Text and the Summary Generated by BERT

C. Experimental Setup

The experimental setup for abstractive text summarization model involves specific hyperparameters with a maximum of 10 epochs and a batch size of 16 to ensure effective learning and convergence. A dropout rate of 0.4 was applied to prevent overfitting. The Adam optimizer, known for its efficiency and adaptability, uses a learning rate of 0.0001. The activation function employed by LSTM, bi-LSTM and Pointer-Generator Mechanism of PGN is SoftMax, for word prediction. Tanh is used by Recurrent Layers of PGN and ReLU in their feed-forward layers of BERT to avoid the vanishing gradient problem. The model utilizes decoder start and end tokens to signal the beginning and end of the summary generation process like ‘summary’ and ‘eos_summary’ respectively.

The loss function Sparse Categorical Entropy, appropriate for handling sparse labels in the output sequence is used. Training is accelerated using a GPU, which facilitates faster computations. The model is designed to handle input articles of up to 512 tokens and generate corresponding summaries of up to 150 tokens, balancing comprehensiveness. software. The software packages used can be seen in Table 1.

TABLE I
PYTHON PACKAGES AND THEIR INFORMATION

Package	**Information**
NumPy	Comprises of mathematical functions for operation on arrays
Pandas	Manipulating time series and numerical tables
TensorFlow	Training and Deploying Models
Keras	Neural Network API
NLTK	Toolkit for performing NLP on texts.
Scikit-learn	Performs Data Analysis and Mining
Matplotlib	A plotting library for visualisations in Python
Tokenizer	Tool to convert text into tokens
re	Library for String Operations.
tensorflow.keras	Framework that build Deep Learning Models and is used for training and evaluatio.

IV. RESULTS

a) Evaluation: To assess the performance of the abstractive text summarization models, two primary evaluation metrics is utilized, Bilingual Evaluation Understudy(BLEU) and Recall-Oriented Understudy for Gisting Evaluation (ROUGE). These metrics are widely used in natural language processing tasks to evaluate the quality of generated text by comparing it with reference texts. BLEU scores range from 0 to 1, where higher scores indicate better quality. ROUGE scores range from 0 to 1, with higher scores indicating better quality. Table 2 shows the comparison of the values of BLEU and ROUGE received by the models.

There are a number of strengths the project possesses, such as robust model architectures, use of pre-trained models, comprehensive evaluation metrics and effective utilization of resources. On the other hand, it is also faced with difficulties like training time and computational resources required, the complexity of implementation, the dependence on data, the risk of overfitting, scalability, and the limitations of automated evaluation metrics.

TABLE II
TABLE COMPARING RESULTS OF MODELS

Model	ROUGE	BLEU	Excecution Time	GPU
T5(base model)	0.20	0.018	2.5 hours	T4
LSTM	0.24	0.05	3 hours	T4
BiLSTM	0.34	0.10	3 hours	T4
BERT	0.43	0.47	3.5hours	A100
PGN	0.41	0.38	2 hours	T4

V. CONCLUSION / FUTURE SCOPE

The project demonstrates, both the advantages and disadvantages of different models for abstractive text summarization. BERT excelled at understanding the semantics of the text and captures intricate relationships between word proved by the BLEU score. LSTM and BiLSTM efficiently handled sequential data but capturing summaries for long articles was an issue. While simultaneously providing fluent summaries, the Pointer-Generator Network was able to efficiently manage out of vocabulary words. Successfully implemented GUI using BERT model that accepts a long text and creates precise summary of the text.

Figure 14 below shows the GUI accepting text and Figure 15 shows the summary generated. a future work, the investigation of hybrid architectures that combine transformer-based models with pointer mechanisms has the potential to improve the handling of out-of-vocabulary terms while maintaining the semantic richness of the language. There is also the possibility that the implementation of few-shot learning could make it possible for the summarization model to adapt to niche domains that have little training data, hence increasing its adaptability and application.

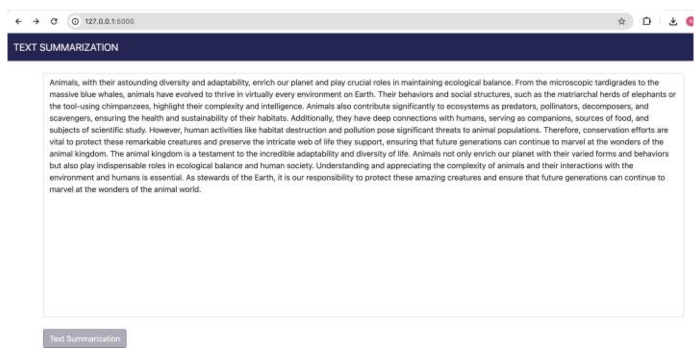


Fig. 14. GUI Accepting Input for Summarization

REFERENCES

[1] R. Nallapati, B. Zhou, C. N. D. Santos, C. Gulcehre, and B. Xiang, ”Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond,” *arXiv*, 2016. [Online]. Available: https://arxiv.org/abs/1602.06023.

Fig. 15. GUI Output Summary Generated

[2] S. Chopra, M. Auli, and G. Rushton, ”Abstractive Sentence Summarization with Attentive Recurrent Neural Networks,” in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, doi: 10.18653/v1/n16-1012.

[3] R. Karmakar, K. Nirantar, P. Kurunkar, P. Hiremath, and D. Chaudhari, ”Indian Regional Language Abstractive Text Summarization using Attention-based LSTM Neural Network,” in *2021 International Conference on Intelligent Technologies (CONIT)*, Hubli, India, 2021, pp. 1-8, doi: 10.1109/CONIT51480.2021.9498309.

[4] S. Song, H. Huang, and T. Ruan, ”Abstractive text summarization using LSTM-CNN based deep learning,” *Multimedia Tools and Applications*, vol. 78, no. 1, pp. 857–875, 2018, doi: 10.1007/s11042-018-5749-3.

[5] J. Jiang et al., ”Enhancements of Attention-Based Bidirectional LSTM for Hybrid Automatic Text Summarization,” *IEEE Access*, vol. 9, pp. 123660-123671, 2021, doi: 10.1109/ACCESS.2021.3110143.

[6] M. M. H. Shahin, T. Ahmmed, S. H. Piyal, and M. Shopon, ”Classification of Bangla News Articles Using Bidirectional Long Short Term Memory,” in *2020 IEEE Region 10 Symposium (TENSYMP)*, Dhaka, Bangladesh, 2020, pp. 1547-1551, doi: 10.1109/TENSYMP50017.2020.9230737.

[7] U. B. Mahadevaswamy and P. Swathi, ”Sentiment Analysis using Bidirectional LSTM Network,” *Procedia Computer Science*, vol. 218, pp. 45-56, 2023, doi: 10.1016/j.procs.2022.12.400.

[8] K. Pabbi and S. C., ”Opinion Summarisation using Bi-Directional Long-Short Term Memory,” in *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2021, pp. 256-259, doi: 10.1109/WiSPNET51692.2021.9419412.

[9] A. See, P. J. Liu, and C. D. Manning, ”Get To The Point: Summarization with Pointer-Generator Networks,” *arXiv*, 2017. [Online]. Available: https://doi.org/10.18653/v1/p17-1099.

[10] R. Wijayanti, M. L. Khodra, and D. H. Widyantoro, ”Single document summarization using BertSum and Pointer Generator Network,” *International Journal on Electrical Engineering and Informatics*, vol. 13, no. 4, pp. 916–930, 2021, doi: 10.15676/ijeei.2021.13.4.10.

[11] F. Boutkan, J. Ranzijn, D. Rau, and E. Van Der Wel, ”Pointless: More Abstractive Summarization with Pointer-Generator Networks,” *arXiv* (Cornell University), 2019. [Online]. Available: https://arxiv.org/pdf/1905.01975v1.

[12]] J. Shobana and M. Murali, ”Abstractive Review Summarization based on Improved Attention Mechanism with Pointer Generator Network Model,” *Webology*, vol. 18, no. 1, pp. 77–91, 2021, doi: 10.14704/web/v18i1/web18028.

[13] N. Darapaneni, R. Prajeesh, P. K. Dutta, V. K. Pillai, A. K. Karak, and A. R. Paduri, ”Abstractive Text Summarization Using BERT and GPT-2 Models,” in *2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)*, 2023, doi: 10.1109/iconscept57958.2023.10170093.

[14] H. Gupta and M. Patel, ”Method Of Text Summarization Using Lsa And Sentence Based Topic Modelling With Bert,” in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021, doi: 10.1109/icais50930.2021.9395976.

[15] M. Ramina, N. Darnay, C. Ludbe, and A. Dhruv, ”Topic level summary generation using BERT induced Abstractive Summarization Model,” in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, doi: 10.1109/iciccs48265.2020.9120997.

[16] Zhang, H., Cai, J., Xu, J., & Wang, J. (2019). Pretraining-Based Natural Language Generation for Text Summarization. *arXiv:1902.09243*. https://doi.org/10.18653/v1/k19-1074