iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

TEXT SUMMARIZATION SYSTEM USING ABSTRACTIVE METHODS IN NATURAL LANGUAGE PROCESSING

Adejumo, Samuel Olujimi
Alade, Samuel Mayowa
Ezuruka Evelyn Ogochukwu
Ruphin, Kusinza Byamungu
Akano, Olatunde Ayodeji,
Nwafor, Jacinta Chinecherem
[1]Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria
[2]Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria
[4]Department of Computer Science, Olivia University, Bujumbura, Burundi
[5]Department of Computer Science, Abiola Ajimobi Technical University, Ibadan, Nigeria
[3]Department of Software Engineering, Nnamdi Azikiwe University, Awka, Nigeria
so.adejumo@unizik.edu.ng, sm.alade@unizik.edu.ng, eo.ezuruka@unizik.edu.ng, rkusinza@oliviauniversity.ed.bi, Olatunde.akano@tech-u.edu.ng, jessienwafor156@gmail.com

ABSTRACT
With the current growth of digital contents across social media networks, education and research platforms, the availability of text summarizing systems has evolved as a crucial and vital tool for users, organizations and corporations. It can analyze and process large amounts of text effectively. However, there are issues raised while reading and analyzing such textual information using manual methods. Problems such as inefficiency, inaccuracy, incoherence, downtime, and delays in generating concise summaries of long texts are common with manual methods. The development of the summarization system involved collecting and cleaning data, applying Bi-directional and Auto-Regressive Transformer models for summarization. The system was implemented in Dart Programming language. The Hugging Face Transformer library was used for model integration, while Flutter Software Development Kit was employed to design an interactive interface where users can input text or upload documents. Performance of the system was evaluated using the ROUGE metric for summarization. The ROUGE evaluation yielded a favorable result, with respondents' responses yielding 79% consistency and accuracy rate after using the developed system. This indicates that when users used the chat summarization system, it was deemed acceptable and good enough to meet the requirement specified as well as demonstrating that the summaries were coherent and representative of the original raw text data. The paper concluded that deploying a proposed summarization system saves time and effort while ensuring more accurate summaries with enhanced user satisfaction and increased productivity.

Keywords:
Bi-directional and Auto-Regressive Transformer models, Summarization, Document, Abstractive Summarization, Coherence, Natural language Processing, Deep Learning.

INTRODUCTION
The quantity of textual information available has increased dramatically in the current digital era. Considering articles in the news, scholarly articles, social media posts, admissible court documents and business reports, managing such an enormous volume of information can be overwhelming, especially for students and managers. For instance, an executive manager might be required to review tens or even hundreds of reports before making a strategic decision on the affairs of the organization. This process is not only time-consuming but also mentally exhausting, often leading to decision-making errors due to the overwhelming quantity of data that must be processed (Luhn, 1958). To mitigate this challenge, a Text Summarization System can be adopted, as it significantly reduces the time required to comprehend lengthy documents or reports. This allows managers to focus on critical tasks while relying on summaries to capture the essential information needed for informed decision-making [1].
Summarization is also crucial in academic environments. Skimming a document or textbook may lead to missing vital information, which could result in severe consequences. For example, a university student preparing for a

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

history exam might skim a textbook due to time constraints and miss intricate details and connections between historical events that are essential for a deep understanding of the subject matter [2]. Summarization condenses large texts into shorter versions while retaining the critical points, thus helping individuals grasp essential information more effectively.

Text summarization has evolved from early manual techniques to sophisticated automated methods. Initial computational approaches were extractive, identifying key sentences based on statistical methods. However, extractive summarization often failed to capture the deeper meaning of the text [3]. This limitation led to the rise of abstractive summarization methods, which uses deep learning and advanced natural language processing methods to create new sentences that can effectively express the original text's main idea [4].

Bidirectional and Auto-Regressive Transformers (BART) is a cutting-edge method for abstractive summarization model which employs both sequence-to-sequence learning and deep transformer networks. BART excels in generating human-like summaries by reconstructing corrupted text, thereby producing coherent and concise outputs [5]. This has made BART a powerful tool in overcoming the challenges posed by the abundance of textual data in today's information-driven world.

Incorporating BART into a Text Summarization System enables users to efficiently extract meaningful insights from large bodies of text, making it an invaluable tool in academic and managerial settings alike. Its ability to produce high-quality summaries reduces the risk of missing crucial information and supports more informed decision-making processes [6].

Text summarization, particularly automated summarization systems, provides a solution by significantly reducing the amount of text while retaining the essential information [7]. These systems allow users to quickly identify key points and make informed decisions with minimal effort. In organizations, this can streamline workflows, allowing managers to focus on strategic tasks while leaving routine text processing to the summarization system [8].

The introduction of neural network-based models such as the Bidirectional and Auto-Regressive Transformers (BART) has revolutionized text summarization. Unlike traditional models, BART combines the strengths of both extractive and abstractive techniques by utilizing a denoising autoencoder architecture. This enables the model to reconstruct corrupted text, making it capable of handling various types of summarization tasks more effectively [9]. BART's ability to process text bidirectionally and autoregressively allows it to better capture the context of a document, resulting in more accurate and fluent summaries [10]. Furthermore, it can handle diverse inputs ranging from short news articles to lengthy research papers, making it highly versatile [11]. Moreover, BART’s pretrained language model offers an advantage in summarizing domain-specific documents, such as medical reports, legal documents, and technical papers, which require a deep understanding of the context [9].

Additionally, text summarization systems, powered by models like BART, are becoming increasingly vital in areas such as legal and healthcare industries. For example, in legal contexts, lawyers must review thousands of documents to prepare for cases. A robust summarization system could significantly reduce the time and effort needed to process these documents, allowing legal professionals to focus on critical aspects of case analysis and strategy [12]. Similarly, in healthcare, summarization systems can assist doctors by condensing patient records, research papers, and clinical trials into concise summaries, enabling them to stay informed without reading full reports, thus improving patient care and decision-making [13].

Overall, text summarization using BART and other advanced models is an indispensable tool in today's information-rich environment. Whether it is in the academic field or the corporate world, summarization systems offer a way to efficiently handle vast amounts of information, ensuring that critical points are not overlooked [14]. As technology continues to improve, the integration of text summarization into various industries will likely expand, further demonstrating its role in enhancing productivity and informed decision-making.

OBJECTIVES

This study is aimed at developing an efficient and adaptable text summarization system. The objectives of the research are to examine the text summarization approaches employed in social media platforms, design and develop the summarization system using pretrained transformer-based model, implement the adopted model using Dart programming language tool and its libraries and evaluate the performance of the system using several performance metrics. Therefore, the paper presents a text summarization system for conversation application.

The organization of the paper is as follows: Section 2 provides a review of the literature or related works, discussing approaches, techniques and identifying existing gaps. Section 3 outlines the methodology used to achieve the stated objectives of the text summarization system including data collection, preprocessing, model training and development. The experiment was conducted and the results including model performance evaluation using recall,

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

precision, F1-score and ROUGE score metrics discussed in Section 4. Finally, Section 5 provided the conclusion and introduced a further improvement of the research topic.

RELATED WORKS

[15] published an article titled "A survey on natural language processing (NLP) -based Text Summarization for Summarizing Product Reviews." This paper provides a comprehensive overview of various NLP techniques applied to product reviews, highlighting the challenges posed by the growing volume of online reviews and the need for effective summarization methods. They conducted a comparative analysis of different techniques, emphasizing the importance of extracting valuable insights efficiently. The paper concluded by evaluating the effectiveness of these methods and suggesting future research directions to enhance summarization techniques further.

[16] collected a large dataset of legal texts and preprocessed the data by tokenization, stopword removal, and lemmatization. Latent Semantic Analysis (LSA) was applied to identify the underlying semantic structure, scoring sentences based on their relevance. The highest-scoring sentences were then extracted to form a summary. The approach focused on selecting and concatenating key sentences from the original text, typical of extractive summarization methods.

[17] presented a method for summarizing large documents by identifying and extracting key sentences using the SpaCy library. Highlighting the growing need for effective summarization techniques due to the increasing volume of digital text, the authors introduce extractive summarization to create coherent summaries while preserving the original meaning. The method is noted for its effectiveness and suitability for applications such as news aggregation, academic research, and information retrieval, with suggestions for future work to explore hybrid models that combine extractive and abstractive techniques for improved summarization quality.

[18] explored a technique for text summarization that combines K-means clustering and TF-IDF (Term Frequency-Inverse Document Frequency). This method focuses on identifying key sentences in a document to generate concise summaries. The researchers addressed the growing need for efficient text summarization due to the overwhelming amount of information available online and the lack of coherence and readability of traditional extractive summarization methods. The authors propose a hybrid approach that leverages the strengths of K-means clustering and TF-IDF to produce more accurate and meaningful summaries. This method aims to improve the selection of representative sentences by considering both the frequency of terms and the structural organization of the text.

[19] introduced a model that combines Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) to generate coherent summaries by rephrasing text. This hybrid model leverages LSTM's sequential modeling and CNN's feature extraction capabilities, outperforming traditional methods. The research highlights the potential for improved summaries with more complex architecture and larger datasets in future work.

[20] explored the use of neural networks for text summarization. The paper highlights the need for efficient summarization techniques due to the enormous amount of information available. Discussing the limitations of traditional methods, the authors proposed neural networks as a solution, demonstrating that neural networks can produce accurate and coherent summaries.

[21] discussed the development of a system that uses Natural Language Processing (NLP) to automatically summarize texts. The system is designed to address the challenges of manual summarization, such as time consumption and bias, by automating the process and ensuring that the essential meaning of the text is retained. The paper emphasizes the system's effectiveness in generating concise summaries and suggests its application in areas like news aggregation and academic research, with recommendations for further improvements.

[22] developed an automated system that uses NLP approaches to produce succinct and logical text summaries with the view of addressing the difficulties presented by the large volume of textual data in areas such as academic research and news aggregation. The conclusion emphasizes how well the system produces accurate summaries and offers suggestions for further development.

[23] applied a neural network-based approach to generate abstractive summaries of sentences, focusing on attention mechanisms to enhance the production of concise and coherent summaries. Rush outlines the complexity of generating summaries that are not mere extractions of the original text, stating the difficulty in contributing to the improvement of text summarization. The system study focused on creating computer models that can understand and produce text based on both images and written descriptions. This research is significant for text summarization because it helps in developing systems that can generate meaningful summaries by understanding and synthesizing information from different types of inputs, like combining what they see in an image with the

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

associated textual information, making smarter computer programs that can look at pictures, read text, and then create summaries or new text based on that combined understanding.

[24] presented a model called CTRLsum that allows users to control aspects of a summary, such as focus and length, by using keywords and prompts. This approach addresses the limitations of traditional summarization methods which often produce inflexible summaries. The model shows potential in generating more personalized and contextually relevant summaries, with future work suggested to refine control mechanisms further.

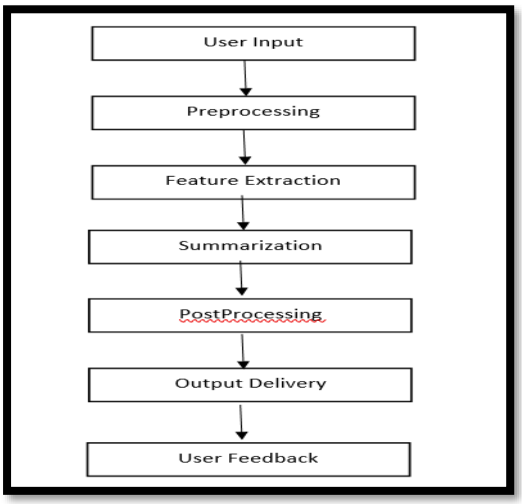
Study conducted by [25] on Urdu language text has seen limited advancements in text summarization within NLP. This paper provided insight and experiment on abstractive summarization using a labeled dataset, achieving a ROUGE-1 score of 25.18 with a transformer model. A similar study by [26] addresseed data scarcity in deep learning in clinical domains. The authors explored the capabilities of large language models like T5 and BART using the CHARDAT dataset. In an attempt to achieve the objective of their study, the authors employed ChatGPT for data augmentation, rephrasing training instances, and compared it with other methods like Easy Data Augmentation (EDA) and An Easier Data Augmentation (AEDA). The results show that ChatGPT augmentation outperformed back-translation, with the BART model achieving ROUGE-1 scores of 52.35.

METHODOLOGY

This section presents the design, analysis and implementation of a text summarization system using the BART model, integrated with a user-friendly interface built in Flutter and Dart. The phase is composed of a variety of resources and methods for evaluating the system's performance, dependability, and robustness, such as ROUGE score.

3.1 Proposed Model

The methodology adopted in the development of this system is the Object-Oriented Analysis and Design methodology (OOADM): a software engineering methodology for assessing and developing an application, system, or business utilizing object-oriented programming and visual modelling throughout the software development process to influence stakeholder communication and product quality. The proposed system begins by acquiring text data from users or external sources whether raw text or image containing texts. It preprocesses the text through tokenization, normalization, and stopword removal to prepare it for analysis. Relevant features are extracted, including keyword identification and sentence scoring to assess significance. The summarization phase involves selecting key sentences for extractive summarization or generating new sentences using deep learning techniques for abstractive summarization. After this, the summary generated undergoes postprocessing to refine clarity, formatting, and redundancy removal. Finally, the output is delivered to the user in an accessible format, with options for saving or sharing. User feedback on the quality of the summary can be utilized to further refine the summarization algorithm. This systematic approach allows for quick condensation of large text volumes while maintaining essential information, making text summarizers valuable tools for obtaining quick insights in an information-rich environment. The system architecture was designed to give the ideal representation that describes the structure and views of the system as shown in Figure 1. Database development tool used was local SQLite database to store user preferences and the history of summarized texts



iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Figure 1. Proposed Text Summarization System

3.2 Data Acquisition and Preprocessing

This research utilizes a dataset comprising chat room text data records related to social media platforms. After obtaining the dataset, several preprocessing steps were carried out to prepare the chat data for further analysis. These steps include:

(i) Data Cleaning: the stage where data cleaning and preprocessing operations are created to ensure that the data is in the correct format for further analysis.

(ii) Tokenization and stopword removal which involved splitting the sentence into individual tokens thus preparing the text for further analysis. During preprocessing, the text sentences from the paragraphs culled from the chat room history is converted into single string. Moreso, the tokenization works in such a way that it involves bringing together many paragraphs that make up the single input string with varying sentence length. Successively, this is reliant on the argument where irrelevant whitespace is removed before and after each sentence in order to standardize the string text. Afterwards, a detailed string denoting a data type of string is formed by concatenating each phrase. The clean string was further divided into sentences using NLTK tokenizer, which contains stop words, to form the sentences into lists. To facilitate tokenization later on the sentences, stop words in sentences are indicated using punctuation to compare the quantity of tokens with the BART tokens having 1,024 tokens or sub-words. Moreover, when the chat text input's token counts above 1,022 excluding two special tokens, which is the BART model's maximum of 1,024 tokens, the text inputs start to segment. This feature enables inputs to go beyond the limitations set by the BART model's long text. Here, a list of sentences is iterated as part of the chunking process in the input text document: Chunk, Chunk list, and Total value (with a default value of zero), which are equally the three variables that the procedure will initially generate. Each chunk in the chunks list has a minimum of 1,024 tokens and is made up of an array of string data types.

(iii) Feature extraction, where important word features are extracted, including keyword identification and sentence scoring to assess significance key words.

3.3 Model Development and Training

During the development stage of the text summarization system, generation of good summaries were carried out using pretrained Transformer model known as BART. The summarization phase of the proposed model involves generating new sentences using deep learning techniques, particularly using the BART model for abstractive summarization. The bidirectional encoder and the autoregressive decoder are the two main components of the BART summarization phase. The bidirectional encoder of BART sequentially took chunks of the chat text that have been preprocessed and tokenize them using the BART tokenizer based on the maximum token criteria. During this stage, several tasks such as token masking, token deletion, infilling, sentence permutation and reconstructions were accomplished. The Token masking was achieved by randomly generating token, which is hidden with a mask, as a selection process of strings. The abstractive then create a representation based on the tokenization of the bidirectional encoder since the mark of the masked token would help the model in making predictions throughout the decoding phase of the BART text summarization.

Subsequently, the input is stripped of a random selection of tokens in such a way that the model has to find the point where the token was removed and identify which locations need missing inputs. In addition to the task of token deletion, the BART model by observation examined the overall number of tokens present and the lack of specific tokens, thereby allowing the model to learn about the content of this text. Tokens that are not consecutive are removed, and a single mask token is placed in their place. As a result of the bidirectional function of the BART model, this process establishes how to comprehend the model during the text infilling.

Furthermore, at the sentence permutation, the model having a clear understanding of how to choose appropriate sentences for a summary randomly select and combine sentences which divide the input chat text collection into sentences using periods or full stops. The document's text collection is methodically rotated to start with the designated token. At this stage, the model learns how to recognize the preamble of the document. Following tokenization, the dictionary of elements contains two keys: attention masks, which indicate which texts should and should not be used, and input IDs, which contain the token ID for the text's location, marking the end of the encoding process, hence making the decoding phase to produce the tokenization needed for the summarization.

The tokenizing algorithm substitutes a token sequence and sentence length that closely mimics the original text for each word in each input text. Through text completion, the model is trained to predict the number of tokens missing from a given sequence. The decoder, which must indicate the start position of the prediction and produce a new word,

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

receives words from the encoder, including the mask position. The model pre-trains on tokens "x" by examining "y" tokens from the previous context to ascertain the current token on the decoder. Meanwhile, token "y" is the reference used to forecast token "x," and token "x" is the token that is produced into a continuous word from the preceding token. This guarantees that the generated word preserves the original text's context.

The models are trained on extensive corpora and can be fine-tuned for specific tasks. They have set new benchmarks in various NLP tasks, including text summarization. After this, the summary generated undergoes postprocessing to refine clarity, formatting, and redundancy removal. Finally, the output is delivered to the user in an accessible format, with options for saving or copying. Similarly, user feedback on the quality of the summary can be utilized to further refine the summarization algorithm. This systematic approach allows for quick condensation of large text while maintaining essential information, making text summarizers valuable tools for obtaining quick insights in an information-rich environment.

3.3 System Design

The proposed summarization system was designed using Unified Modelling Language (UML) tools namely, Use case and Activity diagrams, as shown in Figures 2 and 3, respectively. The Use Case diagram describes the corresponding actors, that is, the students, linguists, users, lecturers, Director, Human Resource personnel and the Administrator, and the roles the actors perform such as uploading text, editing text and image, and naming text. Similarly, Figure 3 describes the different activities that a user can perform when he logs into a summarization system. The activities include input text or image containing text, analyse text or scan the image, tokenize, generate summaries and output summarized text. Using BART model API for summarization, a pretrained transformer model that employs both sequence-to-sequence learning and deep transformer networks, coherent and concise outputs are produced.

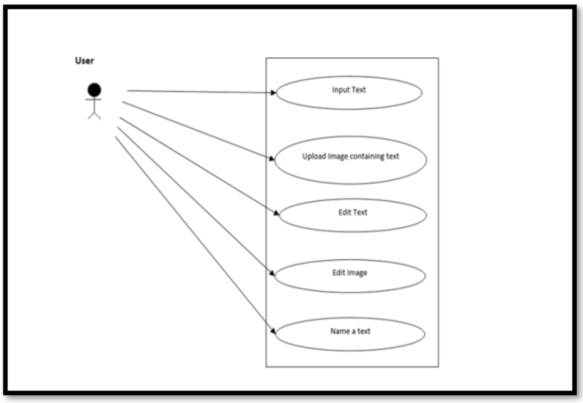


Figure 2. Use Case Diagram of the Summarization System

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

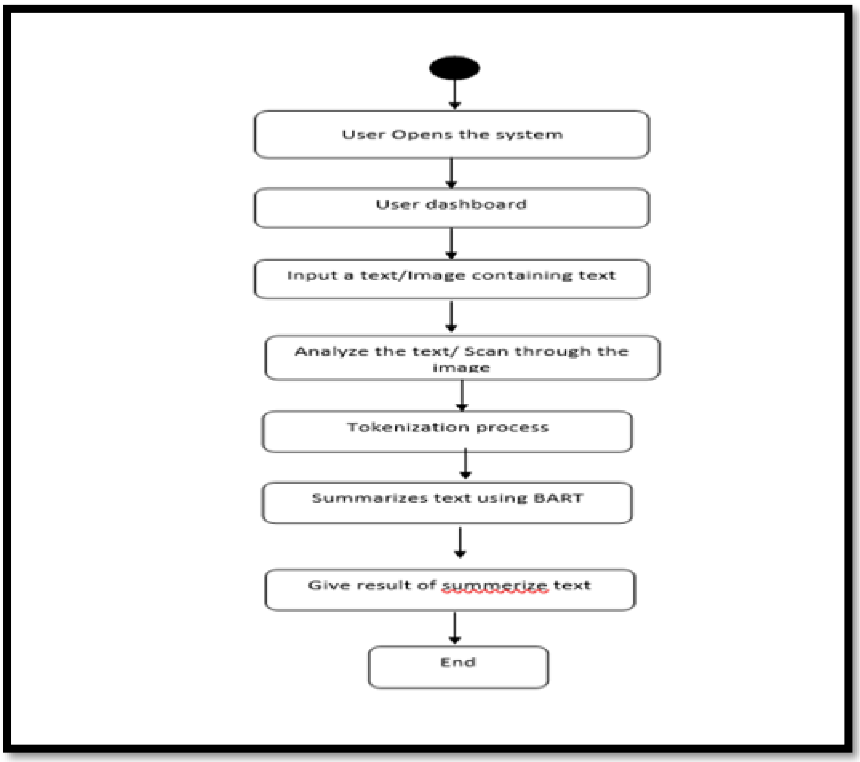


Figure 3. Activity Diagram of the Text Summarization System

3.4 System Evaluation

Metrics like Precision, Recall, F1-score, and the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) scores were used in this experiment to assess the quality of the generated summaries. According to equation (1), the precision score is calculated by dividing the total number of sentences in the reference and candidate (system) summaries by the total number of sentences in the candidate summary. Equation (2) calculates the recall score by dividing the total number of sentences in the reference and candidate summaries by the total number of sentences in the reference summary; Equation (3) displays the F1-score, which is the harmonic combination of precision and recall score. The ROUGE itself uses unigram and bigram to gauge how similar machine-generated summaries are to human-written ones. The fundamental concept of ROUGE is to determine the average number of units that overlap between system summaries and the reference summaries as shown in Equation (4).

$$Precision\ score = \frac{Overlapping\ number\ of\ n-grams}{Number\ of\ n-grams\ in\ the\ candidate} \quad (1)$$

$$Recall\ score = \frac{Overlapping\ number\ of\ n-grams}{Number\ of\ n-grams\ in\ the\ reference} \quad (2)$$

$$F1-Score = 2 * \frac{(recall \times precision)}{recall + precision} \quad (3)$$

$$Rouge_{-1}Score = Mean\ of\ F1-score \quad (4)$$

3.5 *Experiment*

The text dataset used was collected from the social network platform Chat-room. The dataset included conversation between users, particularly from students interacting in a chat room. The experiment was run on a PC containing 8GB RAM, 4 Intel cores (2.67GHz each), with basic preprocessing and other exclusive hybrid techniques where term

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

frequency and outlier removal were applied to the dataset. The cleaned data was fed into the pretrained model called BART and the tokenizer used was from Hugging Face. Meanwhile, when using the Hugging Face training process, BART was applied on all the dialogue data and on the summaries by initiating a function to process the data with the tokenizer which returns the input_IDs, attention_mask and labels for the passed data.

The BART model was fine-tuned by using sequence lengths of 512 and 128 for the encoder and decoder, respectively , thus allowing us to run it quickly on Colab. This shows that the encoder and decoder texts are tokenized, padded, and special tokens are added by the preprocessor. The preprocessor moves the decoder text one position to the right in order to produce labels for auto-regressive training. This is carried out because the model is trained to anticipate the subsequent token at each timestep. During the training, the fine-tuned BART Model employed in the experiment on abstractive summarization is carried out using KerasHub and obtains summaries with the necessary parameters: Total parameters**:** 139,417,344 (4.15 GB), where the Trainable parameters are 139,417,344 (4.15 GB), and optimized using the Adam optimizer which has a linearly decaying learning rate. Of 0.00005 and weight entropy loss of 0.01 to compile the BART model. Afterwards, the scores in the chat data feedbacks data were generated. . For the experiments, SQLite was used for data storage, and Flutter SDK was used for the front end, and the Dart programming language was used to implement the model, render templates, and evaluate the proposed summarization system and comparisons.

The performance measures such as precision, recall, F-1 and RoUGE-1 scores were computed for each class label to analyse individual class performance. These values were then averaged to determine overall precision, recall, and F-measure. The calculations follow equations 1, 2, 3, and 4, respectively.

RESULTS AND DISCUSSION

The results of the summarization system using BART models are discussed in this section. The system attained a recall score of 0.6756, a precision score of 1, and an F1-score of 0.77 and a full ROUGE-1 score of 0.79, showing that the summaries were coherent and representative of the original raw text data. For instance, a chat message of about 80 words was effectively reduced to about 50 words while still retaining its main idea.

Comparison with Other Studies

[25] developed an abstractive text summarization system for Urdu language using NLP with a labelled dataset achieved a ROUGE-1 score of 25.18 with a transformer model. However, the result obtained in this experiment differs from the existing findings in that our findings achieved a score of 0.79. Again, [26] explored the capabilities of large language models like T5 and BART using the CHARDAT dataset to address data scarcity in deep learning, particularly in clinical domains. It uses ChatGPT for training instance rephrasing and data augmentation, and contrasts it with other techniques such as EDA and AEDA. The BART model achieved ROUGE-1, scores of 52.35, indicating that ChatGPT augmentation performed better than back-translation. However, this is not in consonance with the result obtained in the current work which is 79%. This is due to the fact that the data used in the current experiment is not the usual conventional standard textual corpus but a chat-room data which shows better performance as a result of higher unigram overlap, indicating a strong alignment with reference summaries.

4.1 Input and Output Design

The interface for the suggested text summarization system has been designed to enable users to effectively select their preferred method of input, set output preferences, and initiate processing. The input design specifies how users supply data to the system in an orderly and user-friendly manner. The first step for users is to choose their input method, which can be either pasting text straight into the system or uploading a document. Likewise, the user can type or paste the chat content directly into a huge chat text box that appears if the Paste Text option is chosen as depicted in Figure 4. Subsequently, the Output text, which is a summarized version of the input text, is generated by the BART model and is displayed to the user as shown in Figure 5.

ISSN: 2456-9348
Impact Factor: 8.232

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Figure 4. Input Format of the Text Summarization System

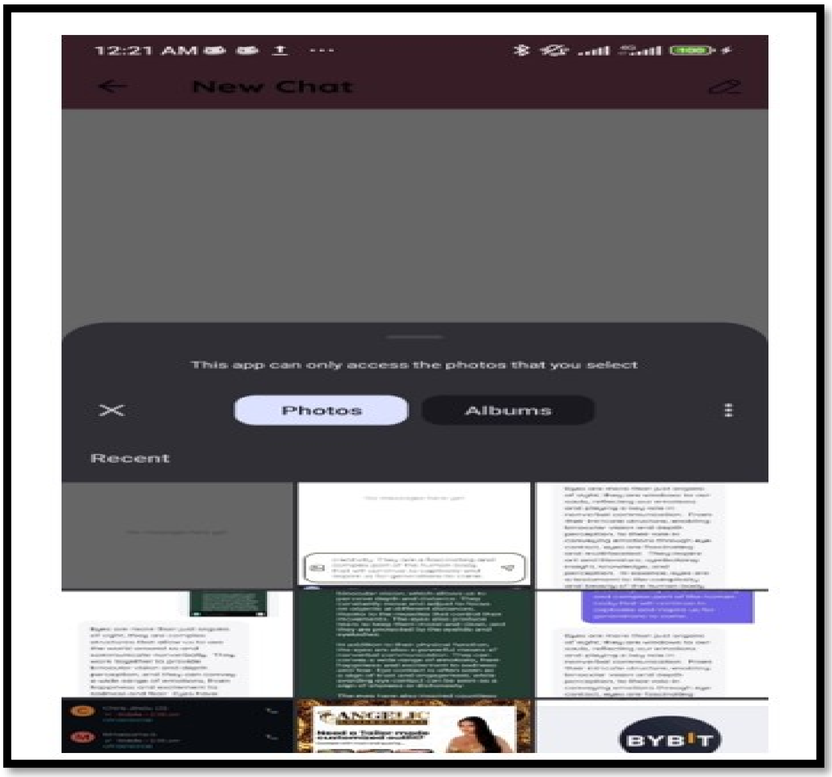


Figure 5. A Summarized Input Text Generated by BART Model

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

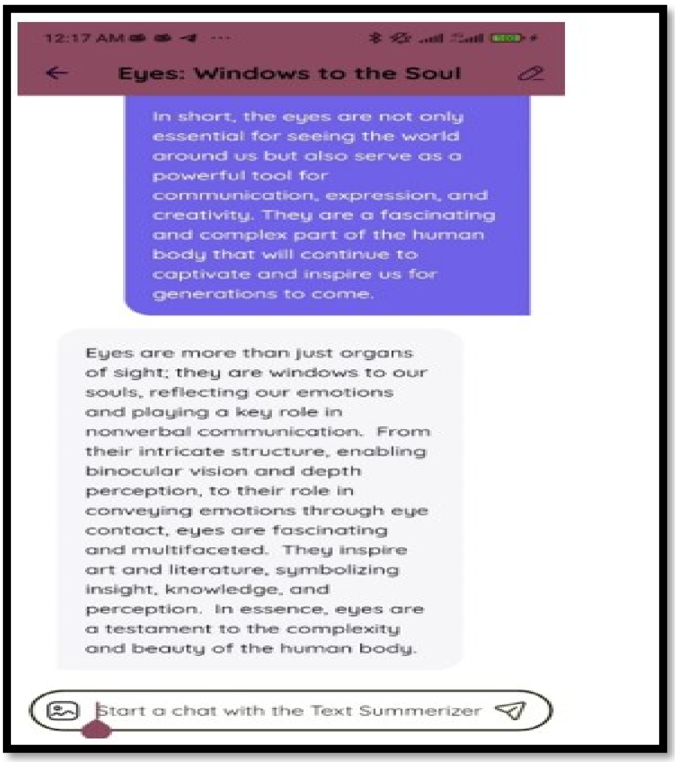


Figure 6: Output Result of Text Input

CONCLUSION

In this paper, technologies like Flutter, Dart, and BART were utilized to create a web-based text summarization system for generating summaries for public use. The text summarizer using the BART model is a vital tool for individuals and organizations that need to process large volumes of text quickly and efficiently. The system's design and implementation provide a fast, reliable, and accurate solution for summarizing text, transforming traditional manual summarization tasks into an automated, user-friendly experience. The text summarization system offers high performance and ease of use across different platforms. In conclusion, the text summarizer achieves its goal of providing an effective tool for extracting key information from text, streamlining workflows, and enhancing content management. As the need for quick information extraction grows, leveraging AI-powered tools like the BART-based text summarizer is essential for improving productivity. For researchers, students, content creators, and professionals in various fields, adopting such a tool can save time and effort while ensuring more accurate summaries. Users are encouraged to integrate this system into their daily workflows, particularly in environments where large amounts of text are processed. Finally, a tool that takes text as input and produces it as a concise summary output was built. The system shows that advanced NLP models like BART can run offline while remaining effective and accessible, providing a foundation for future improvements.

Future studies are urged to improve model generalizability and cultural relevance by using more varied and region-specific datasets, particularly from African communities.

REFERENCES

1) Mani, I. (2001). *A textbook on Automatic summarization* John Benjamins Publishing Company, the MITRE Corporation and Georgetown University, Amsterdam, Philadelphia.,
2) Gupta, V., & Lehal, G. S. (2010). A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence, 2*(3), 258-268. .https://doi.org/10.4304/jetwi.2.3.258-268.
3) Widyassari, A. P., Rustad, S., Shidik, G. F., Noersasongko, E., Syukur, A., Affandy, A., & Setiadi, D. R. I. M. (2022). Review of automatic text summarization techniques & methods. Journal of King Saud University-Computer and Information Sciences, 34(4), 1029-1046. https://doi.org/10.1016/j.jksuci.2020.05.006.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

4) Nallapati, R., Zhou, B., Dos Santos, C., Gulçehre, Ç., & Xiang, B. (2016, August). Abstractive text summarization using sequence-to-sequence rnns and beyond. In *Proceedings of the 20th SIGNLL conference on computational natural language learning* (pp. 280-290). https://doi.org/K16-1028.
5) Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7871-7880. https://doi.org/2020.acl-main.703/.
6) Kryściński, W., Keskar, N. S., McCann, B., Xiong, C., & Socher, R. (2019). Neural text summarization: A critical evaluation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 540-551). https://doi.org/D19-1051.
7) Gambhir, M. and Gupta, V. Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review,* **47**, 1–66 (2017). https://doi.org/10.1007/s10462-016-9475-9.
8) Radev, D., Hovy, E., & McKeown, K. (2002). Introduction to the special issue on summarization. *Computational linguistics*, *28*(4), 399-408. https://doi.org/J02-4001.pdf.
9) Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2020, July). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 7871-7880). https://doi.org/10.18653/v1/2020.acl-main.703.
10) Rothe, S., Narayan, S., & Severyn, A. (2020). Leveraging pre-trained checkpoints for sequence generation tasks. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2069-2081. https://doi.org/10.1162/tacl_a_00313/96450.
11) Zhang, X., Lapata, M., Wei, F., & Zhou, M. (2018). Neural latent extractive document summarization. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 779-784). https://doi.org/10.1162/tacl_a_00313.
12) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30. https://papers.neurips.cc/paper/7181-attention-is-all-you-need.pdf.
13) McDermott, M. B., Parikh, R. B., & Ghassemi, M. (2019). A comprehensive text summarization method for clinical data using BART. IEEE Transactions on Biomedical Engineering, 66(8), 2078-2090.
14) Gupta, S., Wadhwa, S., & Sharma, N. (2019). A hybrid text summarization approach using BART and neural networks. *Journal of Data Mining and Knowledge Discovery, 3*(4), 44-55.
15) Booorugu, N., & Ramesh, S. (2020). A survey on NLP-based text summarization for summarizing product reviews. Proceedings of the 2nd International Conference on Inventive Research in Computing Application, ICIRA, 2020.352-356. https://doi.org/10.1109/ICIRCA48905.2020.9183355.
16) Merchant, M., & Pande, S. (2020). NLP Based Latent Semantic Analysis for Legal Text Summarization. 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). https://doi.org/10.1109/ICACCI.2018.8554831
17) Jugran, S., Kumar, A., Sharma, Anand, V., TYAGI, B.S. & Jain, R. (2021). Extractive Automatic Text Summarization using SpaCy in Python & NLP. *International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2021, pp. 582-585. https://doi:org/ 10.1109/ICACITE51222.2021.9404712.
18) Khan, R., Qian, Y., & Naeem, S. (2019). Extractive based text summarization usinghttps://doi.org/ k-means and tf-idf. *International Journal of Information Engineering and Electronic Business*, *13*(3), 33-44.
https:// doi.org/10.5815/ijieeb.2019.03.05.
19) Song, S., Huang, H. & Ruan, T. (2019). Abstractive text summarization using LSTM-CNN based deep learning. *Multimed Tools Appl* **78**, 857–875. https://doi.org/10.1007/s11042-018-5749-3.
20) Gupta, V. K., & Sah, D. H. R. (2025). Advances in Text Summarization and Generation for Indian Languages: A Survey. *Journal of Advanced Applied Scientific Research*, *7*(3), 1- 33. https://doi.org/10.46947/joaasr7220251387.
21) Nabi, M. S., Mufaddhal, Z., Khant, K. L., Raufi, A. M., & Abdullahi, A. M. (2023). Precision in Communication: A Natural Language Processing Approach to Text Summarization. *Malaysian Journal of Information and Communication Technology (MyJICT)*, *8*(2), 1-7. https://doi.org/10.53840/myjict8-2-21.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

22) Prudhvi, A. B. Chowdary, P. S. R. Reddy, and P. L. Prasanna, titled "Text summarization using natural language processing," was published in 2021 within the *Advances in Intelligent Systems and Computing* series (Springer). Vol. 1171, pp. 535–547. https://doi.org/10.1007/978-981-15-5400-1-54.
23) Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 379-389. https://doi.org/10.18653/v1/D15-1044.
24) He, J., Kryściński, W., McCann, B., Rajani, N., & Xiong, C. (2022, December). Ctrlsum: Towards generic controllable text summarization. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 5879-5915. https://doi.org/10.18653/v1/2022.emnlp-main.396.
25) Raza, H &Shahzad, W. (2024). End-to-End Urdu Abstractive Text Summarization with Dataset and Improvement in Evaluation Metric. IEEE ACCESS vol 12, No.3, PP40311-40324. https://doi.org/1.1109/ACCESS.2024.3377463.
26) Latif, A & Kim, J. (2024). Evaluation and Analysis of Large language Models for Clinical Text Augmentation and Generation. IEEE ACCESS vol 12, No.4, pp. 48987-48996. . https://doi.org/1.1109/ACCESS.2024.3384496.