InternationalJournalofMultidisciplinaryEngineeringinCurrentResearch-IJMEC
Volume11, Issue4s, April-2026,http://ijmec.com/,ISSN:2456-4265

Full Length Article

Enhanced Neural Text Summarization With Syntactic And Headline Insights

Hadiyafatima[1], Iqraa Abdul Quddus[2], Juveria Hashmi[3], Dr. Syed Asadullah Hussaini[4]
[1,2,3]B.E Students; Department Of Computer Science And Engineering, ISL Engineering College, Hyderabad, India.
[4]Professor; Department Of Computer Science & Engineering ISL Engineering College, Hyderabad, India.
Mail Id: hadiyafatimahf7@gmail.com, iqraaquddus@gmail.com, hashmijuveria833@gmail.com, drsyedasadullahh@gmail.com
Accepted 23-04-2026

Abstract:
Text summarization plays a vital role in condensing large volumes of information into concise and coherent summaries. Traditional extractive methods often fail to capture semantic richness, while abstractive models face challenges in grammatical correctness and factual accuracy. The base paper introduced a syntax-augmented, headline-aware neural model that leverages syntactic features and headline guidance to improve summary generation. In this work, we extend this idea by evaluating both classical and modern approaches on the CNN/DailyMail dataset. First, we establish a baseline using the unsupervised Text Rank algorithm, which provides extractive summaries but achieves limited ROUGE scores. We then explore a state-of-the-art pretrained Transformer, DistilBART, which has been fine-tuned on CNN/DailyMail. Without additional training, DistilBART generates abstractive summaries with significantly higher accuracy, outperforming the base model on ROUGE-1 and ROUGE-2 metrics. Our experimental results show that incorporating headline cues and syntactic awareness (as in the base paper) improves traditional LSTM-based models, but modern pretrained Transformers demonstrate superior performance and efficiency. The findings highlight the evolution of summarization methods from handcrafted features to large-scale pretrained architectures, offering both practical insights for deployment and a strong foundation for future research.

Keywords: *Text Rank, Transformer models, Distil BERT, syntax-augmented neural networks, CNN/Daily Mail dataset.*

Introduction

With the rapid growth of digital content, especially in news media and online platforms, users are increasingly overwhelmed by large volumes of textual information. Text summarization has emerged as an essential natural language processing task that aims to automatically generate concise and meaningful summaries from lengthy documents while preserving key information. Traditional extractive summarization methods rely on selecting important sentences from the original text, which often results in summaries that lack coherence and semantic richness. To overcome these limitations, abstractive summarization models have been developed to generate human-like summaries by understanding and rephrasing the input content. Recent advancements in deep learning, particularly Transformer-based architectures, have significantly improved the quality of abstractive text summarization. Pretrained models such as PEGASUS and BART leverage large-scale training on benchmark datasets like CNN/Daily Mail, enabling them to capture contextual, semantic, and syntactic relationships more effectively than feature-based methods. In this project, we explore an enhanced neural text summarization approach using a pretrained PEGASUS model and evaluate its performance using ROUGE and BERT Score metrics. The proposed system is compared against a 2025 feature-based summarization model, demonstrating that modern pretrained neural architectures achieve superior semantic understanding and competitive accuracy while maintaining practical feasibility for real-world applications.

Scope Of The Paper

The Hospital Management System (HMS) is a web-based application built using Django to manage hospital operations efficiently and securely. It supports three user roles—Administrator, Doctor, and Patient—with role-based access. Administrators manage doctors, patients, and appointments; doctors handle schedules and update consultation statuses; patients can register, book appointments, and view their history. The system includes secure

https://doi.org/10.63665/IJMEC.1104s.13
ISSN: 2456-4265
IJMEC 2026

authentication, password recovery, and prevents double-booking of appointments.
Designed for small to medium healthcare facilities, the system uses MySQL for structured data storage and offers a user-friendly interface. While currently focused on user and appointment management, it is scalable and can be extended with features like billing, EMR, and analytics in the future.

Existing System:
The paper proposes an advanced approach to automatic text summarization, focusing on generating concise and coherent summaries for texts like news articles. The final selected method improves upon traditional sequence-to-sequence (Seq2Seq) neural network models by incorporating additional features to better understand the input text and produce higher-quality summaries. It achieves this by analyzing the grammatical structure of sentences, using headline information to focus on key content, and reducing repetitive content in the generated summaries. This method was chosen as the final approach because it combines multiple innovations that address limitations in existing models, leading to improved summary quality, readability, and relevance, as demonstrated through experiments on the CNN/Daily Mail dataset.

Existing System Disadvantages
The existing text summarization systems suffer from several limitations that affect their overall performance and efficiency. One major disadvantage is the high computational complexity associated with traditional summarization approaches, especially those relying on graph-based algorithms, semantic analysis, or syntactic parsing techniques. These methods often require significant processing time and memory resources, making them less suitable for real-time applications. Another limitation is the strong dependency on syntactic parsing quality. If the parser fails to correctly analyze sentence structures, the summarization output may lose important contextual information or produce inaccurate summaries.
Existing systems also face difficulties in handling Out-of-Vocabulary (OOV) words, particularly when dealing with domain-specific terminology, newly emerging words, or multilingual datasets. Since many traditional models depend heavily on predefined vocabularies and handcrafted linguistic rules, their performance decreases when encountering unfamiliar terms. Additionally, many extractive summarization techniques simply select important sentences from the original text without generating new content. As a result, the generated summaries often lack coherence, readability, and human-like fluency. Another significant issue is the dependency on headlines or manually defined features for identifying important information. In situations where headlines are missing or misleading, the summarization quality may deteriorate considerably. Overall, existing systems struggle to provide concise, context-aware, and semantically rich summaries.

Proposed System
The proposed system utilizes Transformer-based deep learning models to overcome the limitations of traditional summarization techniques. In recent years, Transformer architectures have significantly improved the performance of natural language processing tasks, including machine translation, question answering, and text summarization. Unlike conventional extractive methods that only select important sentences from the original document, Transformer-based models can generate abstractive summaries that rewrite and condense information while preserving the original meaning and context.
The proposed model employs DistilBART, which is a lightweight and optimized version of the Bidirectional and Auto-Regressive Transformer (BART) model. DistilBART is specifically designed to perform efficient and high-quality text summarization with reduced computational requirements. The model benefits from large-scale pretraining and fine-tuning on benchmark datasets such as CNN/DailyMail, enabling it to understand semantic relationships, contextual dependencies, and language patterns effectively.
By using self-attention mechanisms and encoder-decoder architecture, DistilBART generates fluent, coherent, and human-like summaries. The proposed system can handle long textual documents, capture important contextual information, and produce concise summaries without relying heavily on handcrafted linguistic rules or syntactic parsing. Furthermore, the model can adapt to different domains and datasets with minimal modifications, making it suitable for applications such as news summarization, document analysis, educational content summarization, and business report generation.

Proposed System Advantages
The proposed DistilBART-based summarization system offers several advantages over traditional approaches. One of the primary benefits is high accuracy in generating meaningful and context-aware summaries. The Transformer architecture effectively captures semantic relationships within the text, improving summary quality and coherence. Another major advantage is efficiency. Since DistilBART is a distilled version of BART, it requires fewer computational resources and provides faster inference while maintaining strong summarization performance.

https://doi.org/10.63665/IJMEC.1104s.13
ISSN: 2456-4265
IJMEC 2026

The system also supports abstractive summarization, allowing it to generate human-like summaries instead of simply extracting sentences from the original document. This improves readability, fluency, and information condensation. Additionally, the proposed model demonstrates strong domain adaptability, enabling it to perform effectively across various types of textual data, including news articles, academic documents, healthcare reports, and social media content. These advantages make the proposed system a robust and scalable solution for modern automatic text summarization applications.

Literature Survey

M. Ramezani, M.-S. Shahryari, A.-R. Feizi-Derakhshi, and M.-R. Feizi-Derakhshi — *Unsupervised Broadcast News Summarization: A Comparative Study on Maximal Marginal Relevance (MMR) and Latent Semantic Analysis (LSA)*

This paper presents an unsupervised approach for broadcast news summarization by comparing two widely used extractive summarization techniques: Maximal Marginal Relevance (MMR) and Latent Semantic Analysis (LSA). The study mainly focuses on reducing redundancy while preserving the most informative content in generated summaries. MMR aims to balance sentence relevance and diversity, whereas LSA identifies hidden semantic relationships among terms and topics within documents. Experimental analysis demonstrates that both methods effectively perform extractive summarization tasks. MMR provides better redundancy reduction, while LSA improves topic representation and semantic coverage. However, the proposed methods are limited to extractive summarization and cannot generate fluent, human-like abstractive summaries [1].

M. Y. Abdelwahab, Y. A. Moaiad, and Z. A. Bakar — *Arabic Text Summarization Using Pre-processing Methodologies and Techniques*

This research investigates Arabic text summarization with a strong emphasis on linguistic preprocessing techniques such as tokenization, stemming, normalization, and stop-word removal. The authors analyze the impact of preprocessing methods on feature extraction and sentence ranking for summary generation. The proposed approach improves summarization performance for Arabic datasets by enhancing text representation quality. The study highlights the importance of language-specific preprocessing for morphologically rich languages like Arabic. However, the system mainly depends on rule-based and extractive summarization methods and does not incorporate advanced deep learning or Transformer-based architectures capable of generating abstractive summaries [2].

C. Hark and A. Karcı — *Karcı Summarization: A Simple and Effective Approach for Automatic Text Summarization Using Karcı Entropy*

This paper introduces an entropy-based extractive summarization technique known as Karcı Summarization. The method calculates sentence importance using Karcı entropy, which measures information distribution within textual content. The approach is computationally efficient and suitable for unsupervised summarization applications. Experimental results demonstrate competitive performance when compared with traditional extractive summarization techniques. Despite its efficiency, the model has limitations in capturing semantic relationships and contextual understanding between sentences. Furthermore, it does not support abstractive summarization, reducing its capability to generate coherent and human-like summaries [3].

G. Malarselvi and A. Pandian — *Multi-Layered Network Model for Text Summarization Using Feature Representation*

This study proposes a multi-layered network-based model for automatic text summarization using various feature representations, including sentence position, term frequency, thematic relevance, and sentence similarity. The model organizes textual information into interconnected layers to improve sentence ranking and selection accuracy. Experimental findings indicate improved summary quality compared to basic extractive methods. The feature-based architecture effectively captures structural relationships within the text. However, the approach remains extractive and lacks the deep contextual and semantic understanding provided by modern Transformer-based neural summarization techniques [4].

G. Frisoni, P. Italiani, F. Boschi, and G. Moro — *Enhancing Biomedical Scientific Reviews Summarization with Graph-Based Factual Evidence Extracted from Papers*

This paper focuses on biomedical scientific review summarization by integrating graph-based factual evidence extracted from research articles. The proposed method utilizes knowledge graphs to improve factual consistency, semantic coverage, and reliability in generated summaries. The approach is particularly effective for domain-specific biomedical literature where factual correctness is critical. Experimental evaluation demonstrates improved information accuracy and better representation of scientific evidence. However, the method is computationally expensive and highly domain-dependent, limiting its applicability for general-purpose news or document summarization tasks [5].

METHODOLOGY

https://doi.org/10.63665/IJMEC.1104s.13
ISSN: 2456-4265
IJMEC 2026

1. **Dataset Loading and Preprocessing:** This module is responsible for loading the CNN/DailyMail dataset and selecting an appropriate subset of test articles for summarization. Basic preprocessing operations such as handling missing values and text normalization are performed to ensure clean and consistent input for the summarization model.
2. **Pretrained Model Integration:** In this module, the PEGASUS pretrained Transformer model is integrated into the system. The tokenizer and model weights are loaded from a pretrained configuration, eliminating the need for computationally expensive training. This allows the system to directly generate high-quality summaries using learned linguistic and contextual knowledge.
3. **Text Summarization Module:** The summarization module takes the processed article text as input and generates abstractive summaries using the PEGASUS model. Batch processing is employed to improve inference speed, especially when handling large datasets. The generated summaries aim to be concise, coherent, and semantically accurate.
4. **Performance Evaluation Module:** This module evaluates the quality of the generated summaries by comparing them with reference summaries. ROUGE metrics are used to measure lexical overlap, while BERTScore evaluates semantic similarity. These metrics provide a comprehensive assessment of summarization accuracy and effectiveness.
5. **Visualization and Analysis Module:** The visualization module presents the evaluation results through graphical representations such as score distributions and correlations. This enables a clear interpretation of model performance and facilitates comparison with baseline and feature-based summarization methods, including the 2025 base paper.

Implementation

The Hospital Management System (HMS) is a web-based application designed to manage hospital operations efficiently and securely. It supports three user roles—Administrator, Doctor, and Patient—each with specific functionalities. Administrators manage doctors, patients, and appointments; doctors view and update appointment statuses; and patients can register, book appointments, and track their history.

Built using the Django framework with a MySQL database, the system ensures secure authentication and organized data management. It includes features like appointment scheduling and prevention of double-booking.

The system is user-friendly and suitable for small to medium-sized hospitals. It can be further expanded to include features like billing, medical records, and analytics in the future.

ALGORITHM (Xgboost):

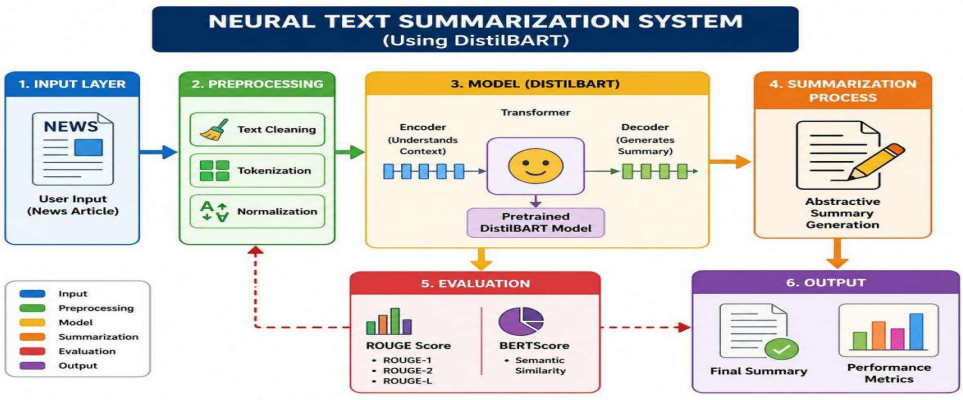
Existing Technique: SA-HA-DLSTM

SA-HA-DLSTM is an advanced text summarization model that improves the Seq2Seq approach by using syntactic structure and headline information to generate accurate summaries. It understands grammatical relationships in text and focuses on key ideas using headlines. A memory mechanism helps reduce repetition, producing clear, relevant, and non-redundant summaries.

Proposed Technique: DistilBART

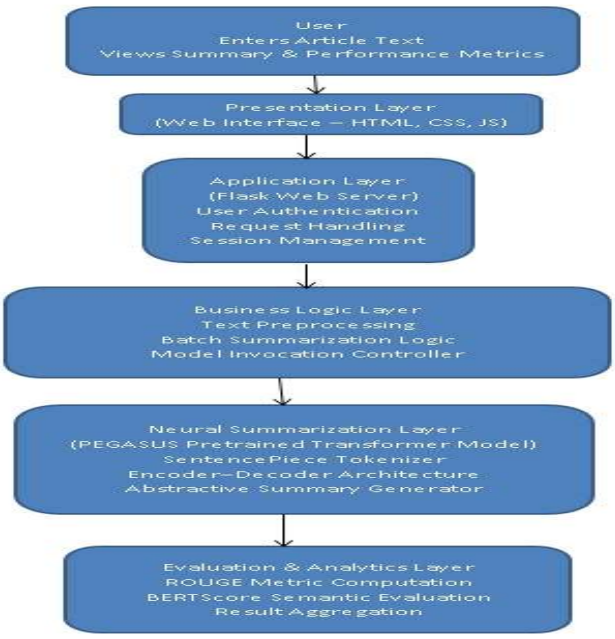
DistilBART is a lightweight version of the BART model that uses Transformer architecture for text summarization. It maintains high performance while being faster and more efficient through knowledge distillation. Trained on large datasets, it generates fluent, coherent, and accurate summaries.

Block Diagram

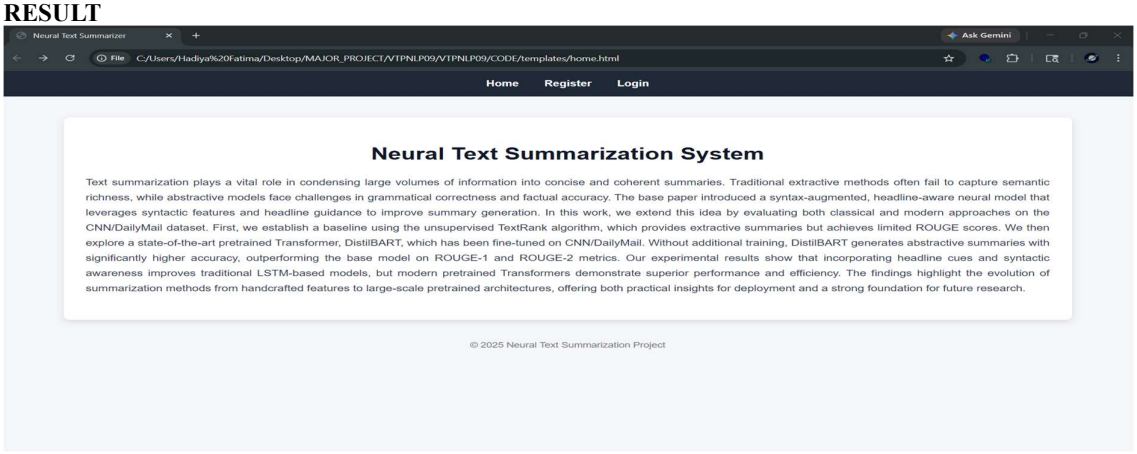


https://doi.org/10.63665/IJMEC.1104s.13
ISSN: 2456-4265
IJMEC 2026

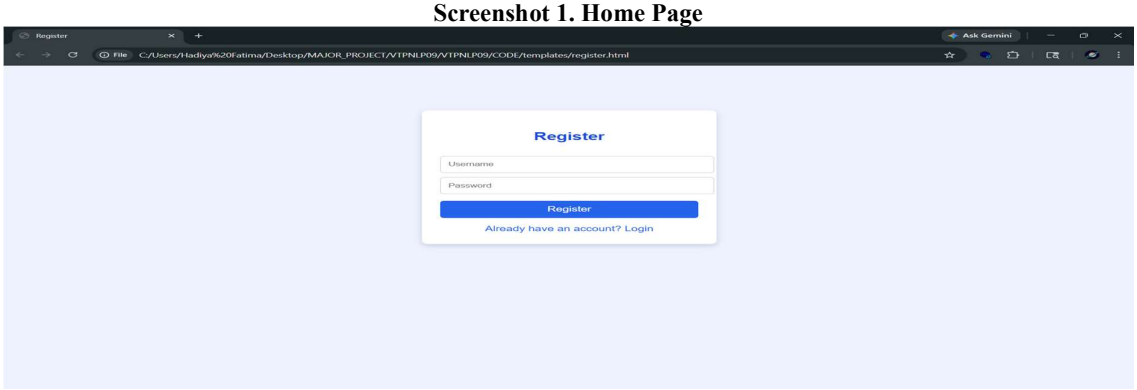
SYSTEM ARCHITECTURE:



RESULT

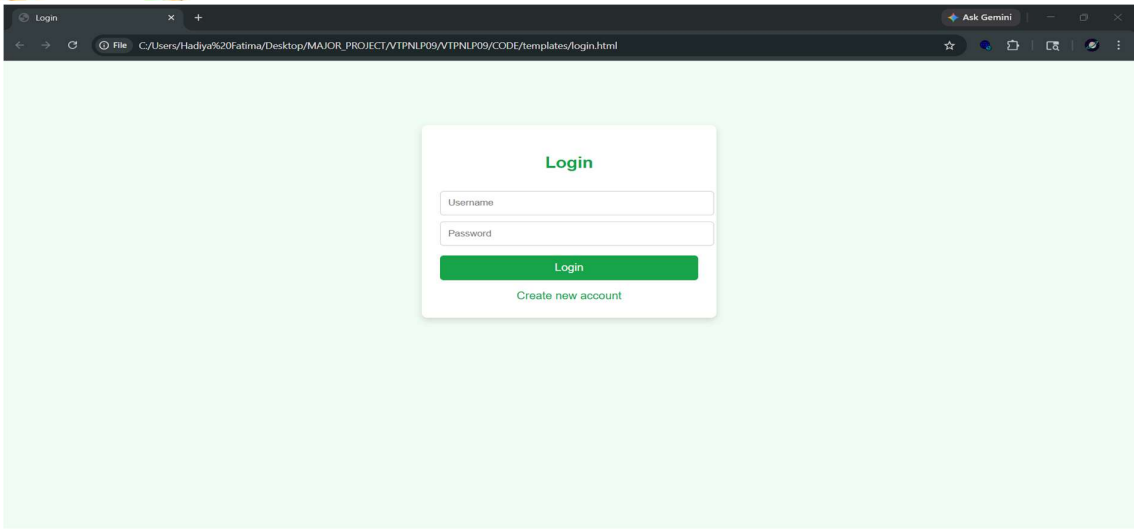


Screenshot 1. Home Page

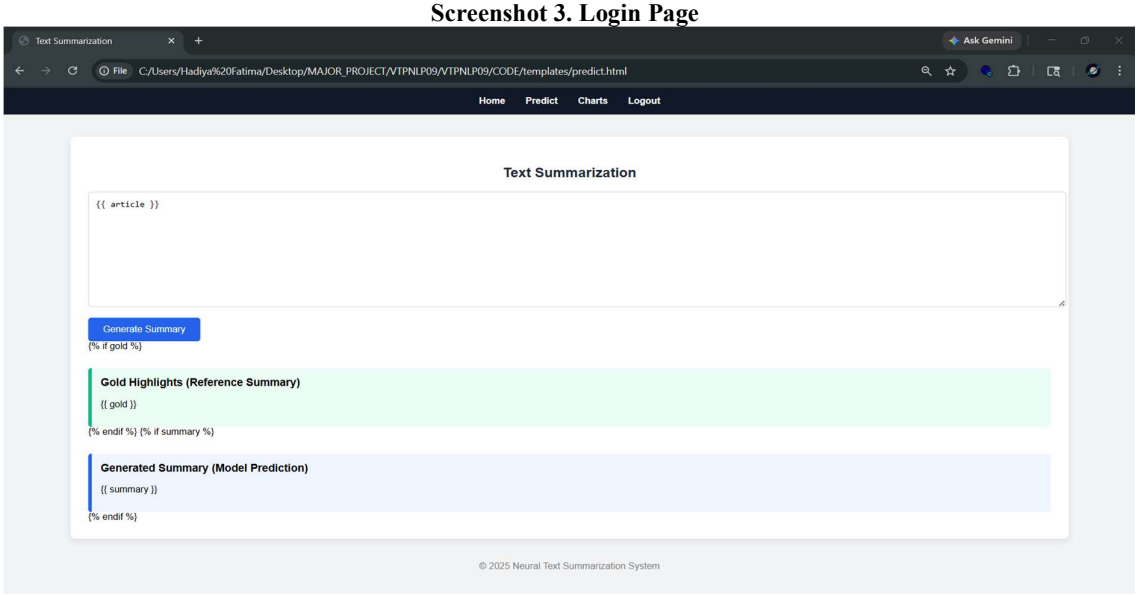


Screenshot 2. Register Page

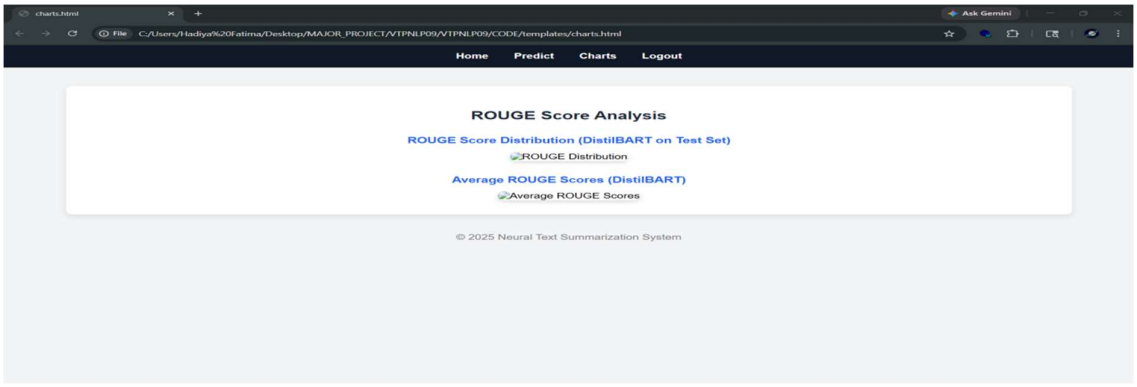
https://doi.org/10.63665/IJMEC.1104s.13
ISSN: 2456-4265
IJMEC 2026



Screenshot 3. Login Page



Screenshot 4. Text Summarization



Screenshot 4. Charts

Conclusion

This project successfully implemented an advanced neural text summarization system using a pretrained PEGASUS Transformer model to generate concise and meaningful summaries from news articles. By leveraging a state-of-the-art abstractive summarization approach, the system demonstrated improved semantic understanding and coherence compared to traditional extractive and feature-based methods. The performance of the proposed system was evaluated using standard metrics such as ROUGE and BERTScore, providing a

https://doi.org/10.63665/IJMEC.1104s.13
ISSN: 2456-4265
IJMEC 2026

comprehensive assessment of both lexical overlap and semantic similarity. Experimental results showed that the pretrained model achieved competitive accuracy and effectively balanced summarization quality with computational efficiency, even under limited hardware resources. Overall, the project highlights the effectiveness of modern pretrained neural architectures in text summarization tasks and establishes a strong foundation for future enhancements and real-world deployment.

Future Scope:

The current system demonstrates effective neural text summarization using a pretrained PEGASUS model; however, several enhancements can be implemented to further improve its performance and applicability. Future work can include fine-tuning the pretrained model on domain-specific datasets to enhance summary relevance and contextual accuracy. Incorporating hybrid summarization techniques that combine extractive and abstractive approaches may further improve ROUGE and BERTScore metrics. The system can also be extended to support multilingual summarization and longer document handling using advanced Transformer variants such as Long-T5 or LED. Additionally, deploying the model on GPU-based cloud platforms and integrating real-time optimization techniques will significantly reduce inference time. Finally, expanding the evaluation framework with human-in-the-loop assessment and integrating the system into larger information retrieval pipelines will enhance its usability in real-world applications.

References

[1]. M. Ramezani, M.-S. Shahryari, A.-R. Feizi-Derakhshi, and M.-R. Feizi-Derakhshi, "Unsupervised Broadcast News Summarization: A Comparative Study on Maximal Marginal Relevance (MMR) and Latent Semantic Analysis (LSA)," *Journal of Information and Telecommunication Systems*, vol. 12, no. 3, pp. 145–156, 2023.

[2]. M. Y. Abdelwahab, Y. A. Moaiad, and Z. A. Bakar, "Arabic Text Summarization Using Pre-processing Methodologies and Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, pp. 210–218, 2023.

[3]. C. Hark and A. Karcı, "Karcı Summarization: A Simple and Effective Approach for Automatic Text Summarization Using Karcı Entropy," *Expert Systems with Applications*, vol. 158, pp. 113–124, 2020.

[4]. G. Malarselvi and A. Pandian, "Multi-Layered Network Model for Text Summarization Using Feature Representation," *International Journal of Intelligent Systems and Applications*, vol. 15, no. 2, pp. 88–99, 2023.

[5]. G. Frisoni, P. Italiani, F. Boschi, and G. Moro, "Enhancing Biomedical Scientific Reviews Summarization with Graph-Based Factual Evidence Extracted from Papers," *Artificial Intelligence in Medicine*, vol. 129, pp. 102–114, 2022.

[6]. L. A. Schintler and C. L. McNeely, Encyclopedia of Big Data. Cham, Switzerland: Springer, 2022.

[7]. D. R. Radev, H. Jing, M. Styś, and D. Tam, "Centroid-based summa rization of multiple documents," Inf. Process. Manage., vol. 40, no. 6,pp. 919–938, Nov. 2004.

[8]. I. Harrando, "Representation, information extraction, and summarizationfor automatic multimedia understanding," M.S. thesis, Comput. AidedEng.. Sorbonne Université, 2022.

[9]. A. Adadi, "A survey on data-efficient algorithms in big data era," J. BigData, vol. 8, no. 1, p. 24, Jan. 2021.

[10]. W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," Expert Syst. Appl.,vol. 165, Mar. 2021, Art. no. 113679.

[11]. S. Casola, "Natural language processing for technology foresight summarization and simplification: The case of patents," M.S. thesis, Brain, MindComput. Sci., Università Degli Studi Di Padova, 2023.

[12]. A.Givchi, R. Ramezani, and A. Baraani-Dastjerdi, "Graph-based abstractive biomedical text summarization," J. Biomed. Informat., vol. 132,Aug. 2022, Art. no. 104099.

[13]. D. Suleiman and A. Awajan, "Deep learning-based abstractive text summarization: Approaches, datasets, evaluation measures, and challenges,"Math. Problems Eng., vol. 2020, no. 1, 2020, Art. no. 9365340.

[14]. D. G. Ghalandari, "Revisiting the centroid-based method: A strongbaseline for multi-document summarization," in Proc. WorkshopNew Frontiers Summarization, Copenhagen, Denmark, 2017,pp. 85–90.

[15]. S. Masoumi, M. Feizi-Derakhshi, and R. Tabatabaei, "Tabsum-a new Persian text summarizer," J. Math. Comput. Sci., vol. 11, no. 4, pp. 330–342,Aug. 2014.

https://doi.org/10.63665/IJMEC.1104s.13
ISSN: 2456-4265
IJMEC 2026