Group ID: 32

Ensemble Text Summarization Algorithm

Final Year Project Presentation

9th May, 2025

Team Members

Aritra Ghosh

Registration Number: 211040100110105
Enrollment Number: 12021002002137
Stream: CSE
Section: A
Roll Number: 74

Subhojit Ghosh

Registration Number: 211040100110130
Enrollment Number: 12021002002160
Stream: CSE
Section: B
Roll Number: 97

Project Mentor

Dr. Anupam Mondal

Table of Contents:

Outcomes:

Paper Title	Publication Venue	Indexing	Date
An Analytical Study of Text Summarization Techniques	IEEE IEMTRONICS 2024	IEEE, Scopus, Springer	May 2024
Performance Analysis of Transformer Models in Text Summarization	IRTM 2024	Scopus	December 2024
Factuality-Aware Stacked Ensemble Summarization	ISACC 2025, Assam University	IEEE, Scopus	March 2025
Factuality-Aware Stacked Ensemble Summarization: Combining Meta-Learning with Feature-Weighted Transformer Gating	Arabian Journal for Science and Engineering (AJSE) --Pending	Journal	May 2025

Abstract

In the evolving digital landscape, the surge in textual data has amplified the need for effective automatic text summarization. This presents a **Stacked Ensemble Text Summarization** model that integrates the strengths of four architectures—**RoBERTa, BERT, LSTM, and PEGASUS**—within a dynamic, meta-learning framework to overcome limitations of individual models such as factual hallucinations and inconsistent generalization. The proposed system operates in two tiers. **Level-0** comprises fine-tuned base models, where RoBERTa and BERT perform extractive summarization, while LSTM and PEGASUS handle abstractive summarization. **Level-1** features a lightweight Transformer meta-learner inspired by the Feature-Weighted Linear Stacking (FWLS) approach. It leverages both base model outputs and meta-features such as ROUGE scores, factuality metrics (QAFactEval, CoCo), semantic distances, and inter-model variance to produce adaptive, instance-specific weightings for summary generation. A **confidence-based early-exit** mechanism further enhances efficiency by selecting a base model's output when it meets high factuality and certainty thresholds. Evaluations on **CNN/DailyMail** and **XSum** datasets show the model outperforms baselines with up to **+1.3 ROUGE-L** and **+0.08 QAFactEval** improvements while reducing inference time by **40%**. The system offers a scalable, explainable, and factually consistent summarization solution for real-world applications across domains.

Introduction

Summarization is the process of condensing a larger piece of text into a shorter version while still capturing the essential information. It allows readers to quickly grasp the main points and key details of a document without having to read the entire text.

- Extractive summarization involves selecting and combining important sentences.
- Abstractive summarization generates new sentences that capture the essence of the original text.

Challenges in Summarization

Summarization research faces hurdles in dealing with dataset limitations, the computational demands of training large models, ensuring coherence and fluency in the generated summaries, and the challenges of evaluating model performance.

Real-World Applications

Summarization finds crucial applications in various domains, including news aggregation, scientific research, and social media content analysis, facilitating efficient knowledge acquisition.

Background Study

Automatic text summarization has evolved significantly with the advent of deep learning and pre-trained language models. While individual models like BART, PEGASUS, and T5 have achieved strong performance on benchmark datasets, their outputs often vary in informativeness, fluency, and factual accuracy. To harness the complementary strengths of diverse models, ensemble methods have emerged as a powerful strategy. Feature-Weighted Linear Stacking (FWLS) is a meta-learning approach that combines multiple base models using learned weights derived from meta-features—such as ROUGE scores, compression ratios, and semantic similarity. However, conventional stacking approaches often fail to account for the nuanced relationship between model behavior and factual reliability. This gap motivates the integration of factuality-aware metrics like QAFactEval into the FWLS pipeline. By leveraging transformer-based meta-learners to dynamically compute instance-specific weights based on these features, the proposed framework aims to generate summaries that are not only fluent and concise but also factually grounded with respect to the source document.

Common Methods of Summarization (Traditional Approaches)

METHOD	DESCRIPTION	ADVANTAGES	LIMITATIONS
TF-IDF	Ranks terms by frequency and document importance.	Simple and fast.	Ignores semantic meaning.
Graph-based Approaches	Constructs graphs of sentences ranked by algorithms like PageRank.	Effective for extractive tasks.	Computationally expensive for large datasets.
Pointer-Generator Model	Combines extractive and abstractive approaches for flexibility.	Balances strengths of both approaches.	May struggle with highly technical texts.
Transformer Models	Leverages self-attention to capture long-range dependencies (e.g., BERT, GPT).	High-quality outputs with contextual fluency.	Demands high computational resources.

Traditional Methods vs. LLMs in Text Summarization

Method	Description	Advantages	Limitations
TF-IDF	Ranks terms by frequency and document importance.	Simple and fast.	Ignores semantic meaning.
Graph-based Approaches	Constructs graphs of sentences ranked by algorithms like PageRank.	Effective for extractive tasks.	Computationally expensive for large datasets.
Pointer-Generator Model	Combines extractive and abstractive approaches for flexibility.	Balances strengths of both approaches.	May struggle with highly technical texts.
Transformer Models	Leverages self-attention to capture long-range dependencies (e.g., BERT, GPT).	High-quality outputs with contextual fluency.	Demands high computational resources.

- Traditional Methods: Rule-based/statistical (e.g., TF-IDF, TextRank); extractive, domain-specific, efficient, but limited in coherence and abstraction.
- LLMs (e.g., GPT, BART): Neural networks; both extractive and abstractive, domain-agnostic, fluent, semantically rich; resource-intensive and less interpretable.
- Summary: Traditional methods suit simple tasks; LLMs excel in handling complex texts.

New Factuality Evaluation Metrics

ROUGE

- Type: Lexical
- Focus: Measures informativeness through n-gram overlap
- Limitation: Surface-level; ignores meaning and factual correctness

BERTScore

- Type: Semantic
- Focus: Evaluates similarity using contextual embeddings
- Limitation: Can reward hallucinated but fluent content

QAFactEval

- Type: Factual
- Focus: Checks QA-based consistency between summary and source
- Limitation: Requires generation of QA pairs (resource-heavy)

CoCoScore

- Type: Factual
- Focus: Measures entailment and consistency using QA-based logic
- Limitation: Computationally intensive and slower to evaluate

Dataset Collection

Dataset Characteristics

Our dataset consists of 200,000 entries, encompassing various text styles, including formal, informal, and conversational, reflecting the diverse nature of real-world text.

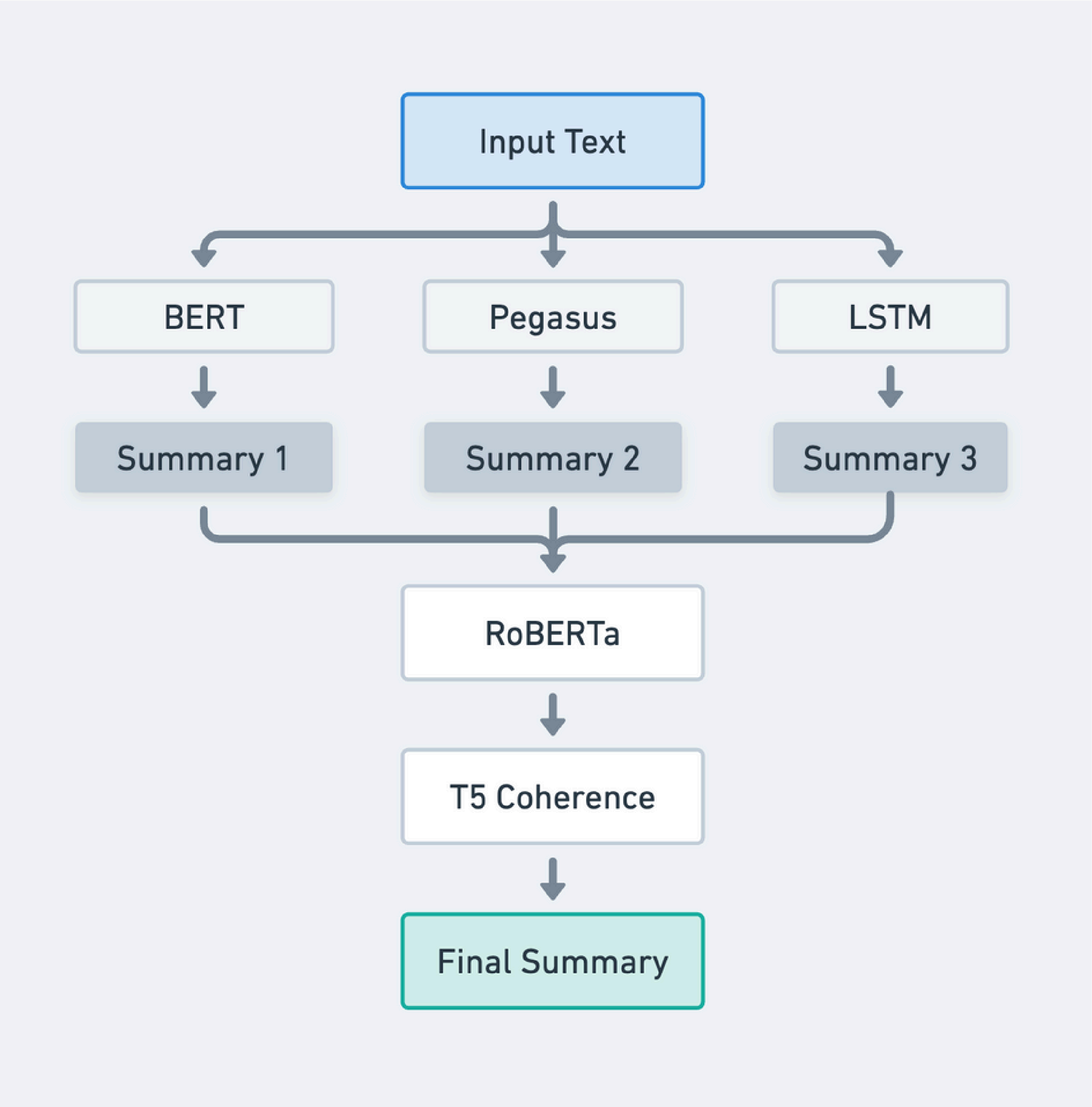
Challenges in Dataset Handling

Challenges in dataset handling include ambiguity in conversational data, domain-specific vocabulary, and data imbalance and redundancy, all requiring careful attention to ensure robust model training.

Improvements to the Dataset

We addressed these challenges by balancing data collection, standardizing the data format, and employing data augmentation techniques like paraphrasing and reordering, further enriching the dataset.

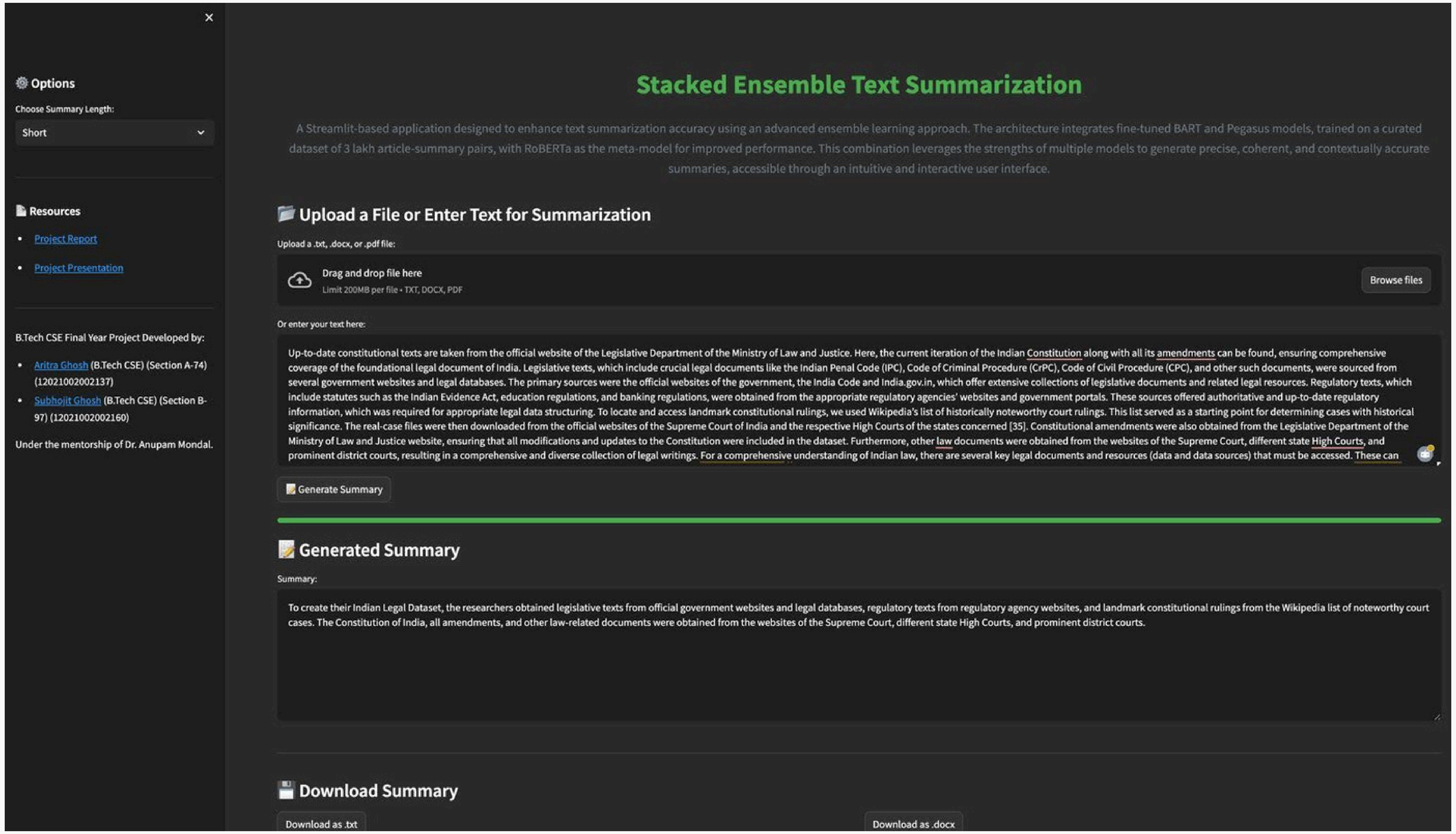
Methodology: Our Approach to Ensemble Summarization



1. Input is sent to PEGASUS, BART, LSTM
2. Each model gives a full summary
3. Break summaries into sentences
4. RoBERTa scores and picks the best sentence at each position
5. Combine selected sentences into a rough summary
6. T5 refines this for flow and grammar
7. Output: Final Ensemble Summary

The input text is summarized separately by PEGASUS, BART, and a LSTM-based Seq2Seq model. Their outputs are evaluated line-by-line by a fine-tuned RoBERTa model, which selects the best sentences. This composite summary is then refined by T5 to improve grammar, flow, and coherence, resulting in a high-quality final summary.

Features of our model:



- Summary Length Selector: Choose from Short, Medium, or Long summaries via dropdown.
- Text Input: Upload .txt, .docx, or .pdf files (up to 200MB) or directly paste text.
- Generate Summary: Click to process input using a BART–PEGASUS ensemble with RoBERTa as meta-learner.
- Output Display: Summarized text appears instantly with factual and fluent content.
- Shows the ROUGE score of the generated text.
- Download Option: Save the summary as a .docx file.

Conclusion

This study presents a robust and adaptable Stacked Ensemble Summarization model that overcomes key limitations of standalone systems like hallucination, domain inconsistency, and inefficiency. By integrating RoBERTa, BERT, LSTM, and PEGASUS through a Transformer-based meta-learner, the framework intelligently fuses extractive and abstractive strengths.

The use of meta-features and a multi-objective loss ensures summaries are both informative and factually grounded. The addition of a confidence-based early-exit mechanism further optimizes runtime without compromising quality.

Extensive evaluation on CNN/DailyMail and XSum datasets demonstrates that the proposed model outperforms individual baselines and traditional ensembles, achieving up to +1.3 ROUGE-L and +0.08 QAFactEval gains with up to 40% lower inference time.

Overall, the system provides a scalable, explainable, and deployment-ready solution suitable for diverse applications like journalism, legal, and academic summarization.

Acknowledgement

We sincerely thank our supervisor, Dr. Anupam Mondal, for his expert guidance and constant encouragement. We are also grateful to the faculty of the CSE Department at IEM for their support. A special note of thanks to our Director, Dr. Satyajit Chakrabarti, for his visionary leadership and motivation. Lastly, we thank our families and friends for their unwavering support throughout this journey. This work would not have been possible without the collaborative spirit and academic environment provided by IEM, Kolkata.

Thank you