A Transformer based Multi Model Ensemble Approach for Bengali Crime Report Summarization

By
Kazi Md Tanzil Islam
221-15-5084

Farhana Symoom
221-15-5263

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Md Ashraful Islam Talukder

Lecturer
Department of Computer Science and
Engineering Daffodil International University

Co-Supervised by

Ms. Umme Ayman

Lecturer
Department of Computer Science and
Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 17, 2025

APPROVAL

This Project titled **A Transformer based Multimodel Ensemble Approach for Bengali Crime Report Summarization**, submitted by **Kazi Md Tanzil Islam** and **Farhana Symoom** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **09-12-2025**.

BOARD OF EXAMINERS

____

Name
Board Chairman
Designation, Department of CSE, FSITDaffodil International University

____

Name
Internal Examiner 1
Designation, Department of CSE, FSITDaffodil International University

____

Name
Internal Examiner 2
Designation, Department of CSE, FSITDaffodil International University

____

Name
External Examiner
Designation, Department of CSE, FSITDaffodil International University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Name of the Supervisor**, **Supervisor's Designation**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Ashraful Islam Talukder
Lecturer
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by:

Ms. Umme Ayman
Lecturer
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

Kazi Md Tanzil Islam
Student ID: 221-15-5084
Department of Computer Science and EngineeringDaffodil International University

Farhana Symoom
Student ID:221-15-5263
Department of Computer Science and EngineeringDaffodil University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project(FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Md Ashraful Islam Talukder**, **Lecturer**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Natural Language Processing** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Com- puter Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Text summarization is one of the notable task in the field of Natural Language Processing. Bengali text summarization is a under research filed in the area of Natural Language Processing, which creates a research gap . Bengali crime report summarization is a very challenging phase due to its unique sentence structure and complex narrative structure which creates a scarcity of annotated resources, and the critical need for accurate information extraction in safety-sensitive contexts. By addressing these gaps this study developed a novel dataset of 4000 instances and developed a ensemble model by merging two state - of - the art transformer model BanglaT5 and an mT5-based Bengali Summarizer model . This ensemble incorporates extractive fusion, semantic rank-based selection, voting-based n-gram consensus, transformer-guided fusion, and hybrid aggregation mechanisms, with the final summary chosen through many metrics, BERT score, Rogue score, SBERT score, Bleu,Coverage, CER and Length. Based on this metrics the based summary is selected. These method outperformed the standalone transformer model, BanglaT5 and mT5 based Bengali Summarizer model with BERT score of 0.82, Rogue score of 0.46 and SBERT score of 0.70. The results shows the capability of the fusion model in the area of text summarization.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter demonstrates the the introduction, motivation , objectives of the study. It highlights the problem statement and outlines the methodology adopted for the Bengali Crime Report Summarization.

1.1 Introduction

Text summarization is one of the most essential areas of research are in the field of Natural Language Processing(NLP).Text summarization refers to generating a short,concise para from a large paragraph, keeping the main and important features and information of the main text [1] . In this digital era, text summarization is a very important tool for quick understanding of the large paragraphs . Automatic summarization tools reduces the overload of the information, enabling quick understanding of the the topic . In some fields such as journalism, research and legal proceedings which are highly information dense, text summarization can reduces work overload by summarizing the information in short.One of the field like this is crime report summarization, summarizing large crime reports in short and concise reports helps the journalists and lawyers to swiftly grasp the whole idea of the report.

There is significant progress in text summarization for high-resource language like English [2] , Arabic [3] , Urdu [4] , but there is a significant research gap in the low-resource languages like Bengali. The Bangla language presents multiple challenges like complex sentence structure, morphological and phonological complexity, large vocabulary etc, due to this complexities and also due to the unavailability of proper resources less research works have been conducted in this field . Furthermore , crime report summarization is much more challenging than summarizing normal Bengali text because crime reports contain emotional intensity, legal policies, details about criminals ,suspects, and victims. For this , crime report summarization is not like normal summarizations, crime report summarizers should be able to include the information about, legal policies , details about crime and criminals in the summary, which makes it a difficult and challenging research area. There is no published research works about Bengali crime reports summarizer to the best of author's knowledge .

Our study addresses these gaps by proposing a transformer based Bengali crime report summarizer model. For this framework two pretrained mode BanglaT5 [5] and and mT5 [6] based Bengali summarizer model were fine tuned and trained on a newly curated dataset of 4000 instances of Bengali crime news collected from different news portal, then the both models were integrated through a multi strategic ensemble mechanism to overcome the traditional limitations of the text summarization models .The new ensemble model will generate five candidate summaries following different approach of summarization, extractive, voting , ranking , transformer and hybrid. Among these summaries the model will select the best summary based on several metrics such, Bleu, CER, Coverage, ROGUE, BERT,SBERT score. This ensemble model ensures, lingual consistency, proper length ration, contextual accuracy demonstrating the significant improvements over traditional transformer models. This study not only contributes novel dataset but also proposed a metric guided ensemble approach advancing the state of Bengali text summarization in the domain of Natural Language Processing.

1.2 Motivation

The motivation behind this research is to contribute in the Bangla NLP. Bengali is still a very low resource language for which researchers don't work with this language that much . In this modern world information is everything, for this automatic summarization tools are very important to swiftly understand large texts. Crime reporter, Lawyers , journalists even general peoples also read newses and reports about crimes daily for their own purposes, these reports are too long which consumes a lot of time to understand the whole news, this is where a automatic summarizer is needed to summarize the long report in short highlighting important information about crime type, victim suspects name etc . The motivations behind this study is given below,

1. To address the research gaps in the Bengali text summarization for crime news .
2. To address the gap in resource for Bengali crime report summarization.
3. The absence of domain specific trained model for Bengali crime report summarization.

1.3 Objectives

The main objective of the study is to create a automatic Bengali text summarization in domain of crime news which can generate short ,concise,accurate and informative summaries from long reports . Our objectives are,

1. To develop a novel domain specific dataset for Bengali Crime Report Summarizer.

2. To develop a domain specific trained model to translate the Bengali crime reports to short and concise form.

3. To demonstrate the practical applicability in real world task.

1.4 Research Questions

This paper looks into how transformer-based models can be adapted and fused to create accurate, coherent, and semantically faithful summaries of Bangla crime reports. Crime reports are difficult to summarize as they lack a generic structure and limited annotated data.

To achieve this ultimate aim, we study the following research questions: How well do individual models BanglaT5 and mT5 perform when they are fine-tuned over domain-specific Bangla crime-news data? What limitations do these models have when they function in isolation with respect to semantic retention, factual consistency, linguistic fluency? Can the ensembles extractive, ranking-based, voting based, transformer fusion and hybrid produce better quality summaries as compared to single models? In conclusion, which combination of evaluation metrics best reflects summary quality in low-resource Bangla NLP scenarios? Also, how can the evaluation metrics be instrumental in selecting the most reliable summary for each input text? These questions together form the basis of research on a mention-aided universal and contextual summarization framework for Bangla crime news.

1.5 Methodology

The study follows a structured approach for methodology which includes data collection, data processing , model implementation and result. For this study, we curated a noble dataset of 4000 instances , sourced from different bengali news portals such as Prothom Alo, Doinik Jay Jay DIn , Doinik Ittefaq etc. Then the collected data was preprocessed to ensure data quality . Two transformer models BanglaT5 and mT5 were trained on the dataset and then ensemble together for better results . the model generates five distinct summaries extractive, rank-based, voting, transformer-guided, and hybrid and selects the best one among them based on several metrics . To evaluate the models performance we have used several metrics, ROGUE,BERT,SBERT and CER . Tis structured approach ensures semantic and linguistic accuracy.

1.6 Project Outcome

The outcome of this study is both theoretical and practical, contributing to Bengali Natural Language processing in the field of text summarization. The most important outcome of this study is a system capable of generating short, precise and concise summaries from large Bengali Crime Reports . The system leverages an ensemble model combining two transformer models(BanglaT5 and mT5) achieving good results , accuracy and high contextual alignment then single base transformer model.

Another major outcome of this study is a novel dataset consisting of 4000 Bengali crime newses. The dataset summaries was human authored.The dataset is one of the first large scale corpora for text summarization of Bengali crime report, which can contribute to lot in future research works in the similar domain.

The impact of this body of work on continued research in the field of Bangla text generation, news analytics, and AI driven journalism is that it makes a substantial contribution to the academic field of study and its practical use in both digital media and law-enforcement settings.

1.7 Organization of the Report

This thesis has been organized into six chapters; each chapter includes a major section of the research study and collectively that present a comprehensive study of Bangla crime-report summarization using transformer-based ensemble models. Chapter 1 discusses the background, motivation, aims and objectives, research questions, and methodology of the research. Furthermore, chapter 1 shows that transformer-based ensemble technique is relevant to use to overcome the challenges of Bangla crime-news summarization . Chapter Two lays the theoretical foundations by surveying existing literature on text summarisation, transformer architectures and Bangla NLP. It also highlights related applications and identifies important research gaps that motivate this work. Chapter 3 provides a detailed account of the research methodology The different processes that involve dataset construction, preprocessing, two-phase model fine-tuning, ensemble summarization strategies, and evaluation framework are presented in detail. The system-design and data-flow diagrams depict the complete workflow of the proposed system. In Chapter 4, implementation is being discussed and experimental results are being reported. We are comparing the performance of banglaT5, mT5, and the ensemble model on several metrics. Subsequently, the practical application of the same is shown with an analysis of its strengths and weaknesses. In Chapter 5, you will review some of the engineering standards (software, hardware, communication, etc.) related to the AI-based text summarization system. You will also analyse the ethical, social, environmental, and sustainability impact of the AI-based text summarization system. Moreover, you will also look at project-management and design issues. In the end, findings presented in Chapter 6 encapsulate the thesis. They show the contribution and outcomes of this thesis. Further, the chapter points out future methods to extend and improve the entire system. Thus, more advanced systems can be developed using the current one. Overall, the chapters are coherent, rigorous, and systematically organised in the exploration of research problem from conceptual grounding to practical realisation and final reflections

Chapter 2
Background

The chapter presents the history of text summarization theory, which includes the development of traditional statistical systems as well as current transformer-based neural systems. It critically evaluates the current research on Bangla and multilingual summarization, especially on T5, mT5, and BanglaT5, and analyzes the applications of such in automated news summarization. The evaluation of the existing domain-specific summarization of crime report in the Bengali language demonstrates that there are considerable loopholes in the system, which drives the creation of the proposed transformer-based system ensemble.

2.1 Introduction

The Natural Language Processing (NLP) has become one of the most revolutionary fields of artificial intelligence, which allows computers to perceive, analyze, and produce human language. The automatic text summarization is one of the numerous uses that this technology has, and it plays a special important role in the problem of information overload that plagues our world today. With the growing volume of digital content, particularly news portals, online newspapers, and social media, there is a growing demands to create concise, accurate and meaningful summaries that enable users to gather the necessary information in a very short time.

Text summarization is generally categorized into two, namely, extractive and abstractive. Extractive summarization works on the principle of recognizing the most significant sentences or phrases in the original text and putting them in a compressed form. Extractive methods are less expensive to compute, and usually less challenging to implement, but they frequently fail to present semantic relationships, or produce coherent narratives, about complex texts or high-context texts.

Abstractive summarization, on the contrary, attempts to replicate the human summarization process by rephrasing and reducing the length of the source phrase. This requires a high level of understanding of semantics, contextual knowledge and the ability to produce natural language. Conventional abstractive approaches that utilized rule-based approaches or classical machine learning struggled to obtain good outcomes. Deep learning transformed the discipline in a large manner particularly the Sequence-to-Sequence (Seq2Seq) model that incorporated Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) units and attention models. Although they were helpful, with long-range dependencies, and efficient with the computers, these models had a problem.

The presentation of the Transformer architecture by Vaswani et al. (2017) [7] was a paradigm shift in NLP. Transformers use a self-attention mechanism that can be able to model relationships in an entire text sequence at the same time, which results in better performance on a wide variety of tasks, such as machine translation, question answering, and abstractive summarization. Their parallel processing and contextual depth coupled with scaling have formed the basis of the modern language models, including BERT, GPT, T5, and mT5.

Though transformer-based models have recorded impressive success in high-resource languages such as English, the advancement and use of the models in Bengali is relatively under researched. The morphologically rich language of complex syntax and having fewer annotated datasets, fewer NLP resources, and linguistic constructs that are unique to Bangali, present a challenge to existing multilingual models. The news about crime in particular may include specialized vocabulary, different journalistic approaches, and descriptions that are quite emotional described which makes it even more difficult to summarize. The factors have shed some light on the relevance of creating a domain-specific, transformer-based summarization model specific to Bengali crime reports. This system development is also filling a distinct research gap, and it serves larger-scale interests in the field of developing Bangla NLP to real-world uses including digital journalism, crime analytics applications, and a public information system.

2.2 Literature Review

An analysis of the literature available indicates that there has been significant developments in automatic summarization of high resource languages such as English, Chinese and French, but there have been minimal developments in terms of low resource languages such as Bangla. Simple early summarization systems mostly used statistical or graph-based approaches including TextRank, LexRank, and the algorithm of Luhn, which used term frequency, word co-occurrence, and sentence position properties. Although these methods performed at an acceptable level in extractive summarization, they had no semantic maturity and did not produce semantically coherent abstractive summaries.

The performance of summarization has increased significantly with the deployment of deep neural networks, especially, Sequence-to-Sequence (Seq2Seq) models based on the Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) frameworks. Nevertheless, these models had weaknesses of long-term dependence and contextual prerogatives. Transformer architecture that was first suggested by Vaswani et al. (2017) [7] in the groundbreaking article titled Attention Is All You Need consists of a novel self-attention mechanism, which is able to perform parallel processing and contextual learning.

As an extension of the transformer architecture, the T5 (Text-to-Text Transfer Transformer) model by Google, integrated various NLP tasks into a single architecture through a text-to-text format. This was later expanded by mT5 (Multilingual T5) to 101 languages to allow cross-lingual learning and cross-lingual transfer on low-resource languages like Bangla. A study by Xue et al. (2021) [8] has shown that multilingual transformers such as mT5 are of great benefit compared to monolingual models in cross-lingual transfer applications.

Over the past few years, a language adapted form of the T5 language, called BanglaT5, that was trained on large-scale Bengali corpora has been presented to enhance contextual sense and fluency in generating text in Bengali. BanglaT5 preliminary studies have provided promising results when used in the general summarization and translation tasks, which suggests that this model can be further fine-tuned to accommodate more specific fields. Nevertheless, although there have been these developments, there has never been a previous study that specifically concentrates on summarization of Bangla crime news, an area with a high level of linguistic diversity and area specificity.

In addition to these fundamental developments, a number of additional researchers have investigated text summarization in another language, domain and format model architecture. The literature that is presented below gives a deeper analysis of methodologies, datasets and experimental outcomes that are related to the current study.

Masri et al[9] concentrate on constructing a higher level of text summarization system in Arabic books of biology in 11 th and 12 th grade of the Palestinian curriculum. The vision is to help students and educators learn complex science material easier as well as promote the Arabic Non-Language Processing and educational technology. Biology textbooks were gathered into a special dataset so that they could be accurate and context-illuminating. Some of the models considered in the study are mBART50, AraBART, mT5, as well as AraT5. The best of them was AraBART which has the highest evaluation score of 8.409, confirmed by checking the evaluation by experts and ROUGE-based measures, making it effective in summarizing Arabic scientific work.

The purpose of the article by Radha et al[10] is to make academic research more efficient by proposing an automated approach to the summarization of academic literature. It also employs an optimized Text-To-Text Transfer Transformer (T5) model, coupled with such techniques as Named Entity Recognition (NER), Latent Dirichlet Allocation (LDA), term frequency analysis and TF-IDF, to retrieve essential information in the research articles. In summarization, the model gets a high score of F1 93.5% which is a strong score. The methodology does not specify the dataset, though, as far as it is aimed to enhance the current techniques and implement more advanced NLP methods aimed to create more efficient and informative summaries.

In this paper, Alipour et al[11] observe abstractive text summary in the Turkish language through the multilingual mT5 transformer model. Because the majority of previous studies are conducted on English, the purpose of the research is to narrow the gap in relation to low-resource languages such as Turkish, by fine-tuning mT5 using both the large-scale MLSUM Turkish news dataset and a 1,010-article custom academic dataset of Dergipark. In the long text, long texts were cut by BERT-based extractive summarization and then processed by mT5. The optimal step was the MLSUM dataset trained using the learning rate parameter of 0.00004, which resulted in ROUGE-1 of 58.76, ROUGE-2 of 52.98, and ROUGE-L of 58.45 indicating a high ability to apply multilingual transformers to summarization processes in Turkish.

Abujar et al[12] dwell on devising a text summarization algorithm in the Bengali version as research in this sphere did not reach the same level as in English. It applies a word2vector model of the representation of the words as vectors to allow finding important content in Bengali texts in a more efficient way. Although the dataset and particular accuracy measures are not discussed, the paper highlights the theoretical background and many possibilities of using word2vector in the summarization of Bengali. It further observes that although the recurrent neural networks (RNNs) can be competitive in general, the core innovation is to adapt the vector based summarization to Bengali, so as to achieve the gap of the existing research in this language.

Tanjila et al. [13] introduce Bengali ChartSumm, the first chart-to-text summarization benchmark dataset in Bengali, to enable the low-resource language research. The data comprises 4,100 marked chart pictures that contain human written summaries in English version. Three LLMs (mT5, BanglaT5, and Gemma) were finetuned on this dataset, where mT5 was best overall: BLEU: 0.2779, CER: 0.5295 and WER: 0.7189. As opposed to BanglaT5 (0.0678), mT5 performed better in terms of ROUGE-1 balance. The present work is the basis that is to be used in further research on multilingual and Bengali-specific summarization study.

Singha et al. [14] suggest an abstractive machine summarizing Bengali text based on Seq2Seq model which is an attention-based model and contains LSTM architecture as encoder and decoder with word2vec embeddings. The model will be trained on a Kaggle dataset of Bengali news, which responds to the absence of tools to summarized Bengali text. TensorFlow with RMSprop and early stopping gave the highest accuracy of ROUGE-1: 0.66, ROUGE-2: 0.45 and ROUGE-L: 0.44 with high overlaps with reference summaries. The paper is a deep learning baseline of Bengali abstractive summarization in a low resource environment.

Shahariar et al. [15] introduce a ranking-based framework that is used to leverage Bengali text summarization through the choice of the most appropriate summary among Bengali text summaries of four pre-trained transformer models (mT5 XLSum, mT5 CrossSum, SciBERT, and mT5 by Shahidul). Summaries are ranked by the similarity of perspective on human references by using TextRank with the embeddings of BangaliBERT. The approach was found to be most successful when using the XL-Sum Bengali subset and 5 K-sample datasets of Bangali summarization, with BERTScore (F1): 0.749, ROUGE-1 F1: 0.249 and BLEU-3: 0.0783 being the highest performance, compared to single models, and showing the usefulness of summary ranking to low-resource NLP.

Morshed et al. [16] describe an extractive summarizing approach to Bengali based on the WGSS algorithm, which compares the similarity of a sentence by using word-paired Gaussian distribution, spectral clustering and ranking by TF-IDF. Tested with four Bengali datasets such as Kaggle Prothom Alo and the performance comparing WGSS to such models as BenSumm and LexRank was 0.49, 0.43, and 0.48 with ROUGE-1, ROUGE-2, and ROUGE-LCS. It also yielded good results on Hindi, Marathi and Turkish data sets showing that it is an effective and adjustable tool to low-resource languages.

Automated News Summarization Using Transformers by Gupta et al[17] is an attempt to automate news article summarization using transformer-based pre-trained models with the goal of abstractive summarization to produce summary which is human and fluently summarized. The authors compared such models as BART, T5, PEGASUS and pipeline-based BART using the BBC News Dataset composed of 2,225 articles. The T5 model was the most successful in terms of ROUGE scores (ROUGE-1: 0.47, ROUGE-2: 0.33, ROUGE-L: 0.42). The paper also reveals that old and prior methods of ML (such as TF-IDF and LSTM-based Seq2Seq) are insufficient, and transformer-based models prove to be effective in creating high-quality summaries.

Chaves et al[18], in their Automatic Text Summarization of Biomedical Text Data: A Systematic Review states that recent systems that summarize biomedical texts were reviewed as of 2014-2022 and included models, datasets, and evaluation metrics. It draws attention to such typical datasets as PubMed, MEDLINE, and NTUH-iMD. There are such models as statistic (e.g., TF-IDF, Bayesian), machine learning (e.g., BERT, GPT-2), and hybrid methods. A ROUGE-1 score of 78.86 with Bayesian extractive model proved to be the most accurate and GPT-2 and AlphaBERT were also highly accurate.

Table 2.1: Summary of Literature Reviewed.

Author (s)	**Year**	**Title**	**Methodology**	**Key Findings**
Masri et al[1]	2024	Transformer Models in Education: Summarizing Science Textbooks with AraBART, MT5, AraT5, and mBART	• Created a domain specific dataset: Arabic 11 th -12 th -grade biology textbooks. • The models that were evaluated: mBART50, AraBART, mT5, AraT5. • Expert evaluation + ROUGE metrics were used.	• AraBART got the highest score (8.409) • Successful for Arabic scientific summarization
Radha et al[2]	2024	AI-Driven Summarization of Academic Literature using Transformer Model	• Proposed automated academic summarization using optimized t5 model. • Combined Approach-NER, LDA Topic Models, TF, TF- IDF for extracting key information. • Model Trained Based on Hidden Academic Dataset	• Achieved F1 = 93.5% • Demonstrated better academic text summarization
Alipour et al[3]	2025	Abstractive summarization using multilingual text-to-text transfer transformer for the Turkish text	• Turkish abstractive summarization using mT5. • fine-tuning on MLSUM Turkish dataset and 1,010-article Dergipark dataset. • Extractive summarization is based on BERT and	• Best setting: LR=0.00004 • ROUGE-1: 58.76, ROUGE-2: 52.98 • Performance in summarization of the Turkish language is

			was used to reduce the length of long input. • Experimented on different learning rates.	good
Abujar et al[4]	2019	An Approach for Bengali Text Summarization using Word2Vector	• Proposed word2vec based Bengali summarization technique. • Focused on vector representations in order to detect important sentences. • Conceptual/theoretical approach; Dataset not clearly described.	• Noticed possible high quality of the vectors based on the Bengali summarization • No dataset/metrics reported
Tanjila et al. [5]	2025	Bengali ChartSumm: A Benchmark Dataset and study on the feasibility of Large Language Models on Bengali Chart to Text Summarization	• Introduced Bengali ChartSumm, a chart to text dataset containing 4100 chart images + human summaries. • Fine-Tuned mT5, BanglaT5 and Gemma on the dataset. • Evaluated using BLEU, CER, WER and ROUGE.	• mT5 best overall (BLEU=0.2779,CER= 0.5295,WER = 0.7189) • Bangali chart-to-text dataset Foundation dataset
Singha et al. [6]	2023	Bengali Text Summarization with Attention-Based Deep Learning	• Trained an attention-based Seq2Seq model which has LSTM encoder-decoder model with word2vec encodings. • Kaggles Bengali news are trained. • RMSprop, early stopping, and TensorFlow have been implemented.	• ROUGE-1: 0.66, ROUGE-2: 0.45, ROUGE-L: 0.44 • Substantial abstractive summarization, Bengali base.
Shaharia r et al. [7]	2024	Rank Your Summaries: Enhancing Bengali Text Summarization Via a Ranking-Based Approach	• Presented a ranking-based super ensemble summarization system. • Outputs of 4 transformer models used: mT5 XLSum,	• Outperformed single models • BERTScore (F1): 0.749, ROUGE-1 F1: 0.249

			mT5 CrossSum, SciBERT, mT5 (Shahidul). • Ranked summaries with TextRank and BanglaBERT score derived on the similarity to reference summaries.	and BLEU-3: 0.0783 being the highest performance
Morshed et al. [8]	2024	A Novel Word Pair-based Gaussian Sentence Similarity Algorithm For Bengali Extractive Text Summarization	• WGSS algorithm proposed based on word-pair Gaussian distributions to calculate sentence similarity. • Ranking based on applied spectral clustering + TF-IDF. • Evaluated on 4 datasets of Bengali language.	• ROUGE-1: 0.49, ROUGE-2: 0.43, ROUGE-LCS: 0.48 • Competitive with LexRank, BenSumm • Publications in several low languages.
Gupta et al[9]	2021	Automated News Summarization Using Transformers	• Comparing to transformer models (BART, T5, PEGASUS) of news summarization • Used BBC News DataSet (2,225 articles) • Concentrated on abstractive summaries.	• T5 best (R-1: 0.47, R-2: 0.33, R-L: 0.42) • Transformers are superior to previous ML/Seq2Seq.
Chaves et al[10]	2022	Automatic Text Summarization of Biomedical Text Data: A Systematic Review	• Systematic review about biomedical text summarization (2014-2022). • Examined datasets (PubMed, MEDLINE, NTUH-iMD) and different types of models: statistical, ML, and hybrid models.	• Bayesian extractive ROUGE-1 = 78.86. • Also good showings are GPT-2 and AlphaBERT.

2.3 Gap Analysis

A gap analysis is made to compare the level and constraints of the existing summarization studies on various languages and fields. These observations are as follows and can be tabulated as shown below to bring out the areas that the proposed system is special in covering.

Table 2.2 : Gap Analysis Table

Features	Masri et al.	Radha et al.	Alipour et al.	Abujar et al.	Tanjila et al.	Singha et al	Shahariar et al.	Morshed et al.	Gupta et al.	Chaves et al.	Proposed System (Bangla Crime News Summarizer)
Summarization in Bangla	No	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Summarization for Crime News Domain	No	No	No	No	No	No	No	No	No	No	Yes
Domain-specific dataset creation	Yes (Biolo-gy)	Not specifi-ed	Yes (News + Dergip-ark)	No	Yes (Charts)	No	Yes (Rank-ing dataset)	Yes (Kaggle news)	Yes (BBC News)	No	Yes (Bangla Crime News Dataset)
Use of Transformer Models (mT5 or T5 or BART or PEGASUS)	Yes	Yes	Yes(mT5)	No	Yes (mT5, BanglaT5, Gemma)	No (LSTM)	Yes	No (WGSS)	Yes	Yes(rev-iew)	Yes
Evaluation with ROUGE or BLEU or BERTScore	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Support for Low-Resource Language Challenges	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Model Comparis	Yes	Yes	Yes	No	Yes	Yes	Yes (ranki	Yes	Yes	Yes	Yes

on or Benchmarking							ng approach)				
Application to News Text	No	No	Yes	No	No	Yes	Yes (News subset)	Yes	Yes	No	Yes
Combining Multiple Models or Hybrid Approaches	No	Yes	Yes (extractive + abstractive)	No	No	No	Yes (ranking)	Yes (WGSS hybrid)	Yes	Yes (systematic review)	Yes

2.4 Summary

This chapter includes a general review of the history, current state, and open research challenges in automatic text summarization with particular interest in the multilingual and low-resource languages, e.g., Bangla. Early methods of summarization were based on statistical and graphical methods like TF-IDF, TextRank, LexRank, and Luhns' algorithm, whose extractive performance was reasonable but lacked semantic comprehension and could not involve lovely abstractive summaries. The introduction of deep learning, in particular, the RNN-based models as well as the LSTM-based Seq2Seq models, improved the contextual representation but were restricted in their capabilities to consider long-term dependency and do sequential processing.

The situation began to change with the emergence of transformer architectures, beginning with Attention Is All You Need (Vaswani et al., 2017), and their prospects of parallel processing and deep contextual learning. Subsequent models, such as T5 and mT5, directed the field a step further, with the usage of a more universal text-to-text model that supports 100 languages, and making them available in cross-lingual and low-resource applications. The recent developments, such as BanglaT5, demonstrate that increasingly more models are being made available in languages adapted to the Bangla language, in particular.

The chapter also reviews the domain-specific and language-specific work on summarizing the Arabic, Turkish, and Bengali languages. Models like AraBART, mT5, and T5 that are transformer-based have proven to be effective in scientific, academic, as well as news summarization. The Bengali summarization research has presented promising developments in the form of different methods, including Seq2Seq LSTM and vectors, extractive methods, including WGSS, ranking models that involve multiple transformer models, and a charts-to-text dataset, i.e. Bengali ChartSumm. No matter how far it has gone, the earlier work suffers from limited coverage, especially in the area of specialization of the fields and coverage of the news genres in the Bangla language in totality.

Generally, the literature section indicates that abstractive summary is a mystically studied area on behalf of Bangla crime news, which is a linguistically diverse area with context variation and domain-specific trends of narratives. The existing Bengali summarization systems are quite general and not crime-specific. This makes it lighter to the need of an individual, transformer-based model of summaries that may generate coherent, contextual, and domain-adaptive summaries of the crime-related news articles in the Bangla language. The proposed system will eliminate this gap and add a novel contribution to the under-resourced summarization research.

Chapter 3

Research Methodology

This chapter describes about the overall methodology, model architecture of the proposed model. This chapter also includes the UI/UX of the developed system based on this study demonstrating its capability in real world application. The chapter also includes requirement analysis, data-flow descriptions.

3.1 Methodology

This section describes the overall working flow of this study along with the architecture of the proposed model. This section includes data preparation, transformer model fine-tuning, ensemble fusion strategies, and a metric-driven evaluation mechanism. Both functional and non-functional requirements of the system are analyzed to ensure model reliability , consistency over predicting good summaries.

3.1.1 Overview

The over all methodology follows a five step structure , beginning with the construction of the dataset with over 4000 instances of data, then data preprocessing which includes steps such as punctuation removal english alphabet removal etc. , then come the model implementation process we have trained two separate model transformer model on the curated dataset and strategically assembled these two models and making a ensemble model , The model merges the two output from the each summarizer model and generate 5 distinct candidate summary and choose the best based on several metrics. Then a weighted evaluation module selects the best summary which mostly aligns with the original text contextually ans semantically.

3.1.2 Proposed Methodology/ System Design

The Bangla Crime Report Summarization System is structured to work through different multi-stages. It will start from collecting the datasets and will end with creating an optimized summary. The latter will be evaluated through an ensemble-based approach. The first step of the process is Dataset Collection. Here, crime-related Bangla news articles and their summaries have been collected from various reliable news portals to create a large corpus. Then, the text undergoes a thorough processing pipeline that improves its linguistic consistency and reduces noise. The pipeline involves removing punctuations, handling null values, removing numbers and English characters and normalizing Unicode to easily decode the Bangla script. After cleaning, the dataset is divided into training (70%), testing (20%), and validation (10%) subsets to allow for controlled model tuning and a rigorous evaluation process.

After preparing the data, the model implementation phase begins, in which the two transformer-based architectures, BanglaT5 and mT5, are fine-tuned independently using the same training parameters. We train each model for 20 epochs with a learning rate of 1e-3 and weight decay of 0.01, save checkpoints every 5000 steps, and logging after each step to monitor the model behaviour. The multi-model reasoning of the system is based on fine-tuned models. Outputs from both are passed to an ensemble model that harnesses their strengths. The ensemble utilizes five fusion techniques, which are extracting fusion, rank fusion, vote fusion, transformer-assisted fusion, and hybrid fusion. All the fusion methods interpret and combine the model outputs in their own way: pick the most salient sentences, compute their semantic similarity, reconstruct the most frequent n-grams or build a summary using an extra prompt to the transformer.

Candidate summaries are evaluated using a comprehensive metric evaluation framework consisting of BLEU, ROUGE, BERTScore, SBERT similarity, coverage, length ratio, CER, vocabulary complexity, and diversity. The metrics look at the vocabulary level, semantic level, factual level, grammar, and structure. The output of every metric is weighted and integrated in a scoring mechanism to obtain the best summary. The final output of the system is the best summary, chosen based on the candidate with the highest overall performance on all the metrics. The systematic approach guarantees the robustness, scalability and quality of the summarization pipeline in terms of producing high-quality informative summaries for Bangla crime news articles.

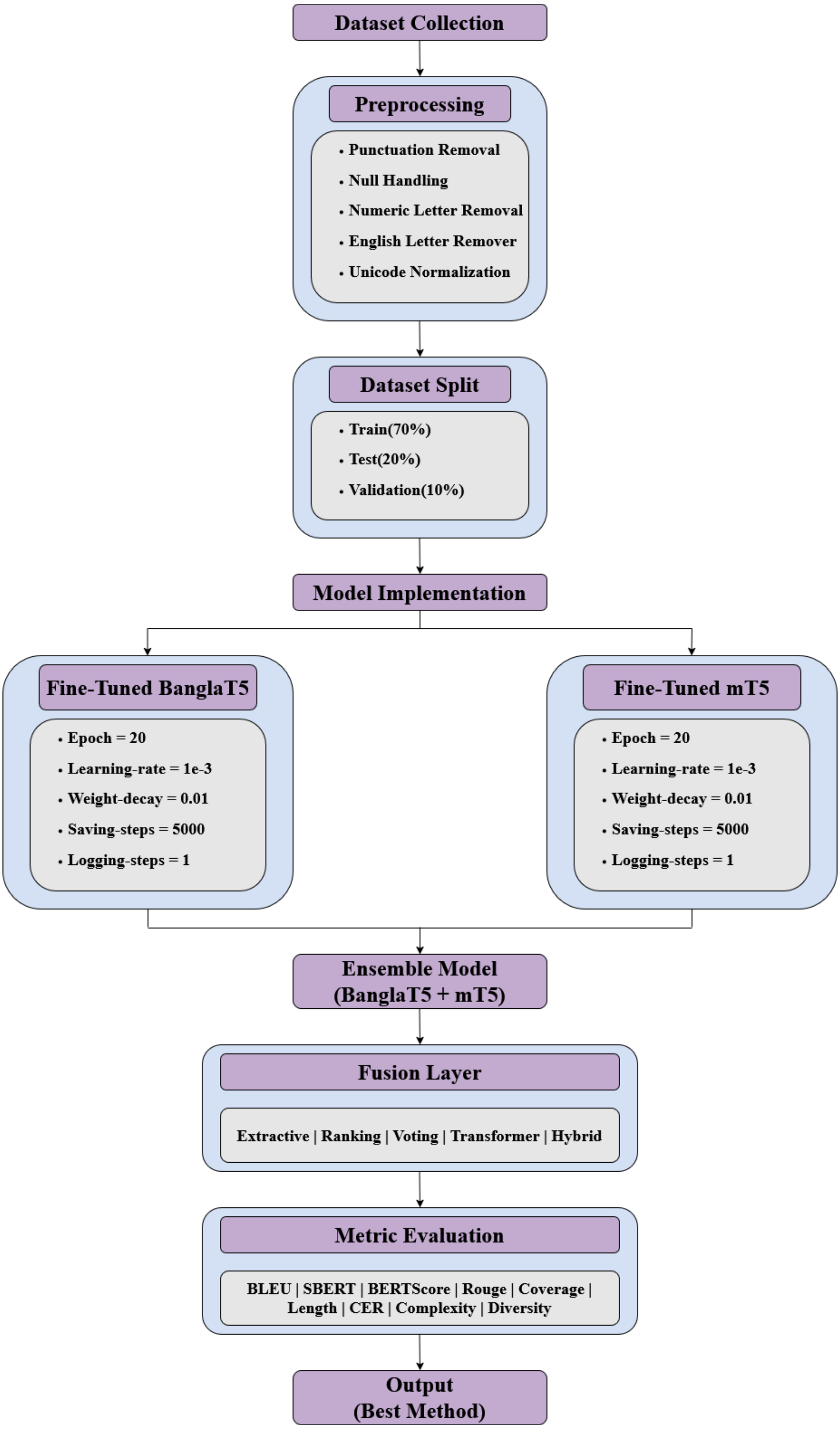

Figure 3.1: Overall Methodology of The system

3.1.3 Functional and Nonfunctional Requirements

The Bangla Crime Report Summarization system is based on a list of functional and non-functional requirements that, in turn, form the scope of its operations and performance expectations.

Functional requirements: These are the functional requirements that define the core activities that Bangla Crime Report Summarization system must perform in an attempt to fulfill its purpose. The system should have the capacity to receive the articles on Bengali crime-news and their abstract and conduct the necessary preprocessing tasks, i.e. cleaning, normalization, and tokenization. It has to facilitate fine-tuning of both BanglaT5 and mT5 models under the two organized training processes, whereby the models learn gradually through the dataset. The system must derive summaries of each transformer model and then perform an ensemble fusion mechanism that yields five summary variants namely extractive, rank-based, voting based, transformer-guided and hybrid. It needs a detailed metric evaluation module to calculate the important indicators of performance, including BLEU, ROUGE-L, BERT Score, SBERT similarity, Character Error Rate (CER), coverage and length ratio. According to such measures, the system needs to choose the best summary using a weighted scoring system. Also, the system must be able to visualize performance trends where comparative graphs among models and fusion strategies can be drawn.

Nonfunctional requirements: The nonfunctional requirements explain the wider quality characteristics that make the system robust, useful, and applicable over a long period of time. Performance is an important factor, the system must produce summaries with ease, and must be capable of managing large volumes of documents or long documents containing crime-news without degradation. A deterministic and score-based process of evaluation that ensures a consistent summary selection should be maintained as a form of reliability. To be maintainable, a modular code structure is necessary, which allows adding, updating, or incorporating other components with minimum effort in the future. The concept of usability is to generate human-readable, consistent and contextually relevant summaries that are acceptable to the practical application of journalism or analysis. Finally, security is a critical requirement to prevent a breach of service by unauthorized individuals or corruption of the dataset, model checkpoints, system outputs during development and deployment lifecycle.

3.1.4 Data Flow Diagram

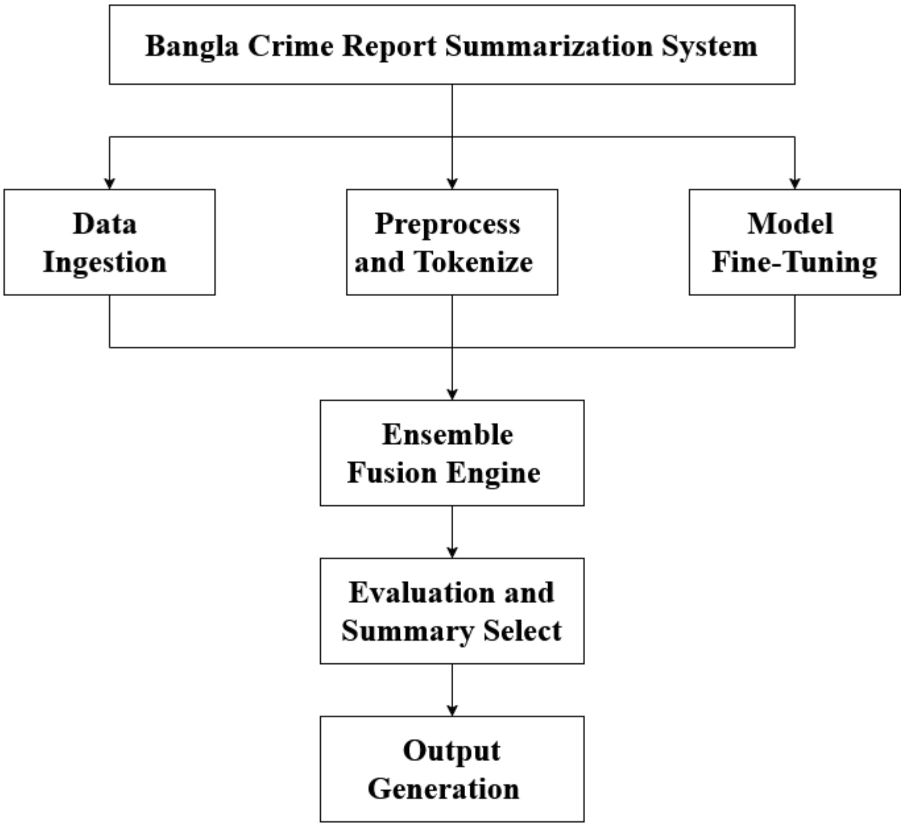

Figure 3.2 : Data Flow Diagram

The following diagram demonstrates the data flow diagram of the system.

3.1.5 UI Design

This section shows the UI design of the implementation of the system . We have build a website to show the implementation of the system in real world situations . The UI pages are shown bellow,

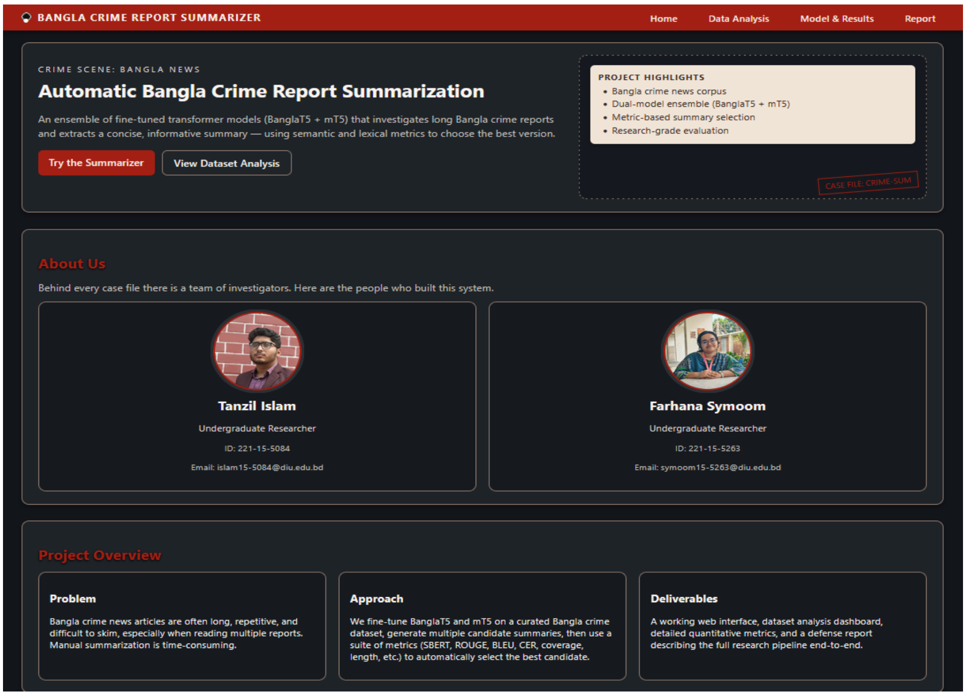

Figure 3.3: Home page

The image portrayed on Figure 3.2 shows the homage of the system. The home page Includes a hero section consisting the title and a short description about the project. Then in the About Us section contains information about us and in the last section a short project overview.

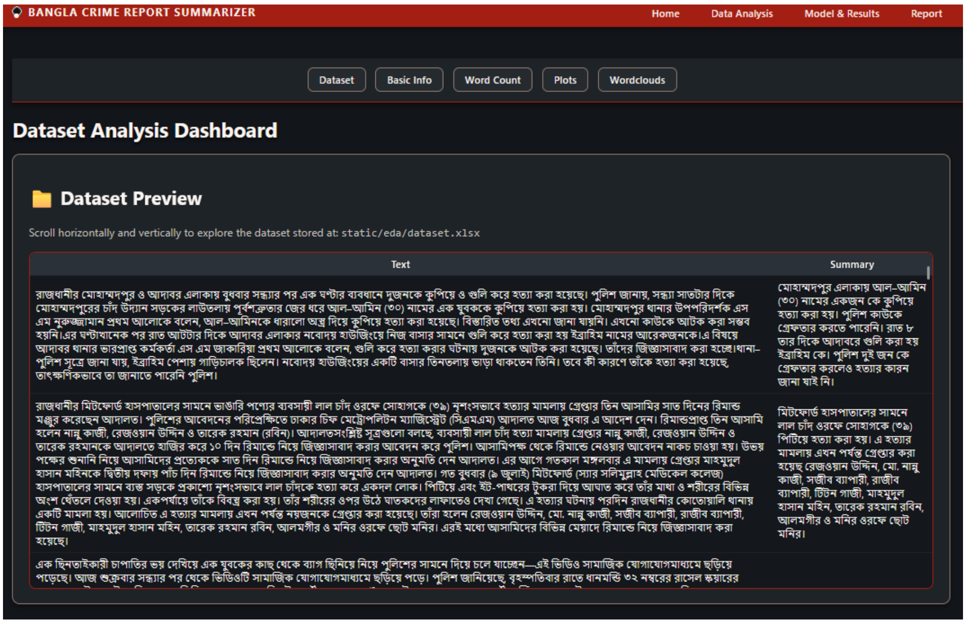

Figure 3.4 : Data Analysis Page(Section 1)

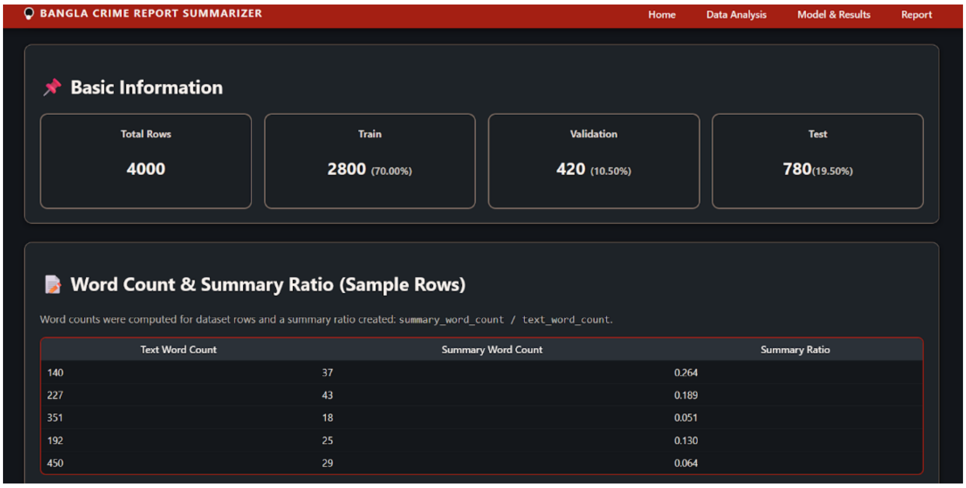

Figure 3.5 : Data Analysis Page(Section 2)

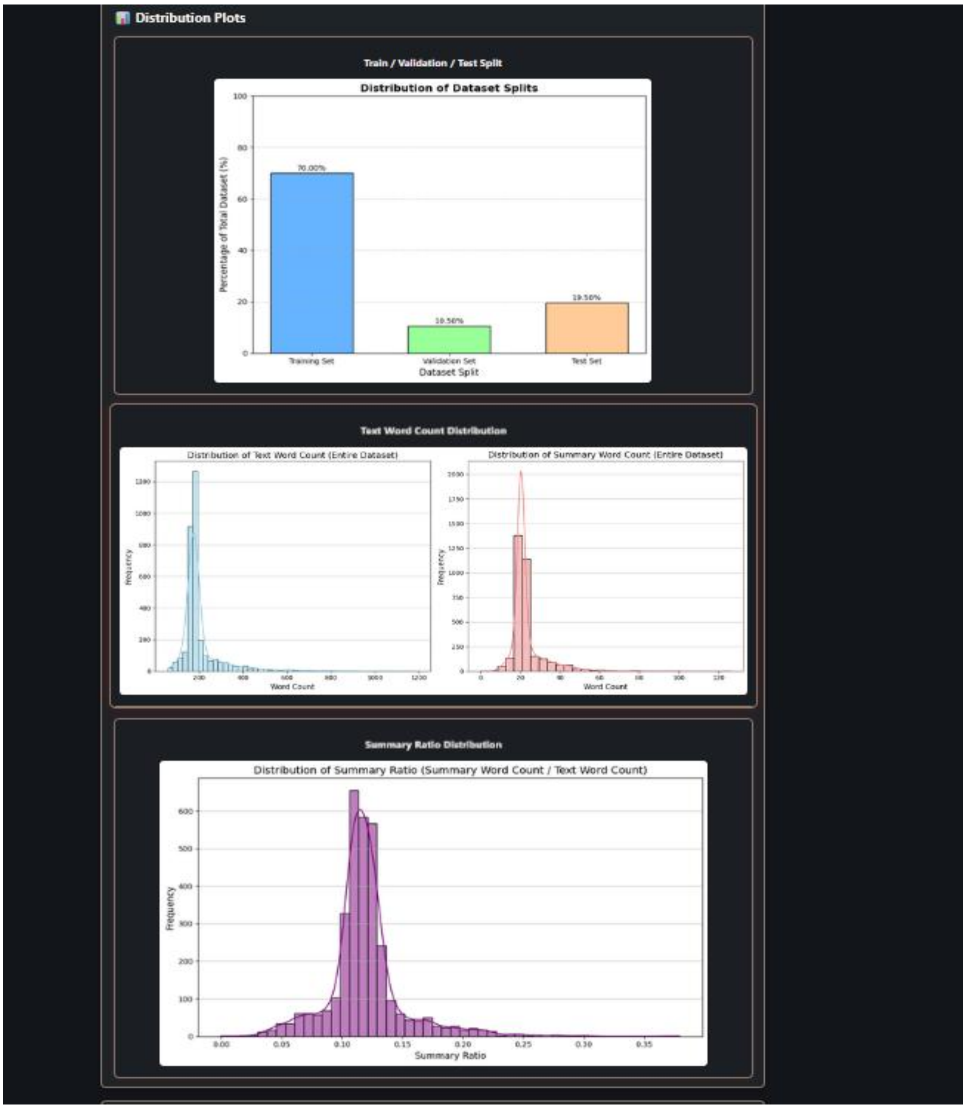

Figure 3.6 : Data Analysis Page(Section3)

Figure 3.7 : Data Analysis Page(Section3)

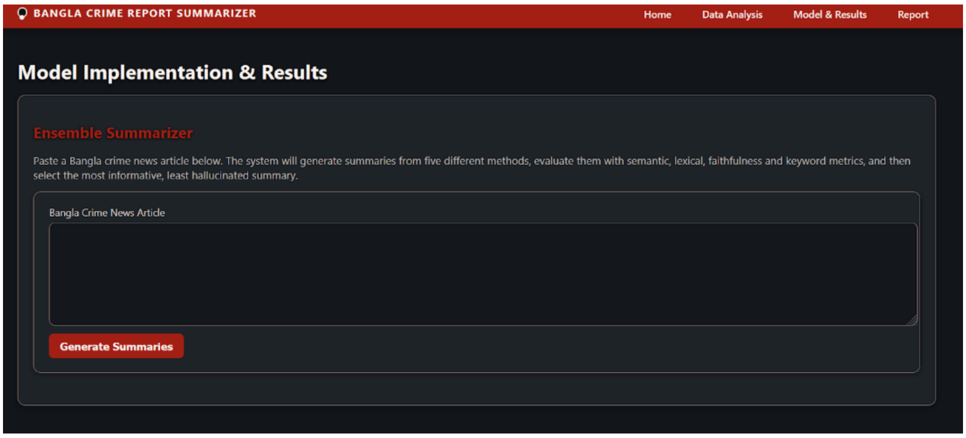

Figure 3.8 : Model Implementation and Results

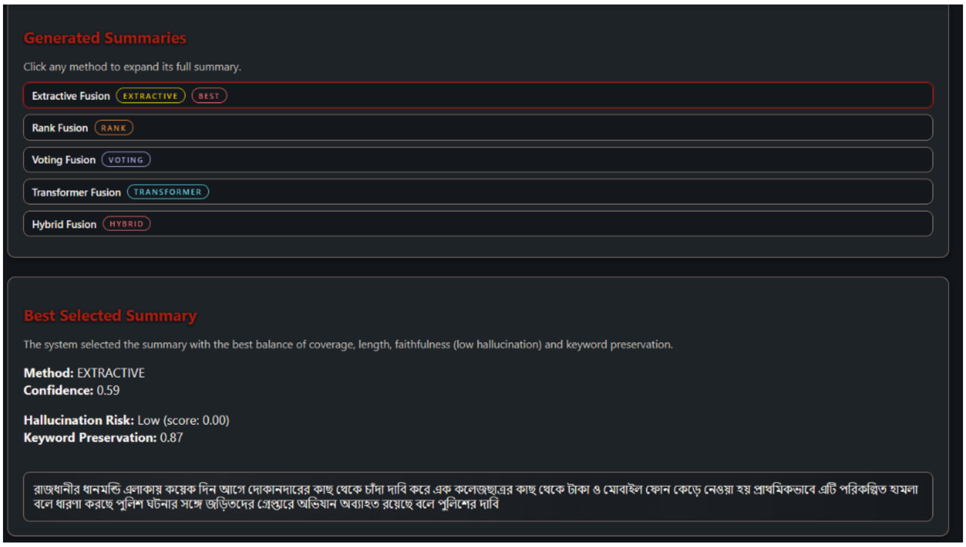

Figure 3.9 : Generated Summaries

Figure 3.10 : Results of the generated model

The above images are for the data analysis page of the system which contains three main section,

First section is for dataset preview where user can preview the whole dataset used for this study. Instead of loading the entire dataset in to a static mode, the system loads the dataset in a scrollable viewport to properly see the dataset while saving page space.

Second section contains basic overview of the dataset such as dataset size , what is the split percentage for train , test and validation.

Third section demonstrates plots and graphs which gives a over view about the dataset. The plots used in this section are, train - test -validation split graph , text distribution over both columns and summary ration distribution plot. We also added word clouds for both of our columns which shows the most recurred words.

The Third page is for model implementation and result where user can upload crime reports to the text field and the model will generate summary with the results and scores that the generated summary how much accurate.

The system is build with HTML,CSS for frontend and Flask for backend. This system is the real life application of the ensemble transformer model .

3.2 Detailed Methodology and Design

The method of this research can be described as a comprehensive multi-step pipeline developed for creating concise coherent and semantically salient summaries of raw bangla crime-news articles. All the stages designed in small tasks to solve a specific problem of Bangla text summarization, such as the challenge of linguistic varieties, unavailability of sufficient training data, and the need for domain-adapted representation learning. In the subsequent subsections, we will describe all major components of the system, including dataset construction, preprocessing, model fine-tuning, ensemble construction and evaluation metrics.

1. **Dataset Construction** : There is no currently publicly available Bengali Crime Report dataset, though there is dataset available for news data containing newses of all kind there is lackings of domain specific crime news dataset. We collected crime-related articles from Prothom Alo, BDNews24, Jugantor, Kaler Kantho, and other regional Bangla news portals. To maintain linguistic diversity, representation from all parts of Bangladesh, and variability in journalistic style, these sources were chosen. Table 3.1 portrays the sample dataset used in this study to build the Bengali Crime Report Summarizer.

   When collecting articles, we emphasized not just on complete crime stories. Also, we chose articles written by different journalists. Also, we included articles from different districts for regional dialectal influences. The articles were paired with the corresponding editorial summaries to create high-quality text-summary pairs for supervised learning. The last dataset contains 4000 full-length crime reports corresponding to human-written summaries. Report article-length is between 150–450 words while summaries are 15–50 words. Dataset is divided into Train(2800 rows) , Test(780 rows) and Validation (420 rows) (Figure 3.11) . When the fine-tuning dataset contains reports with significantly different lengths, the model gets to see many types of reporting styles from the data.

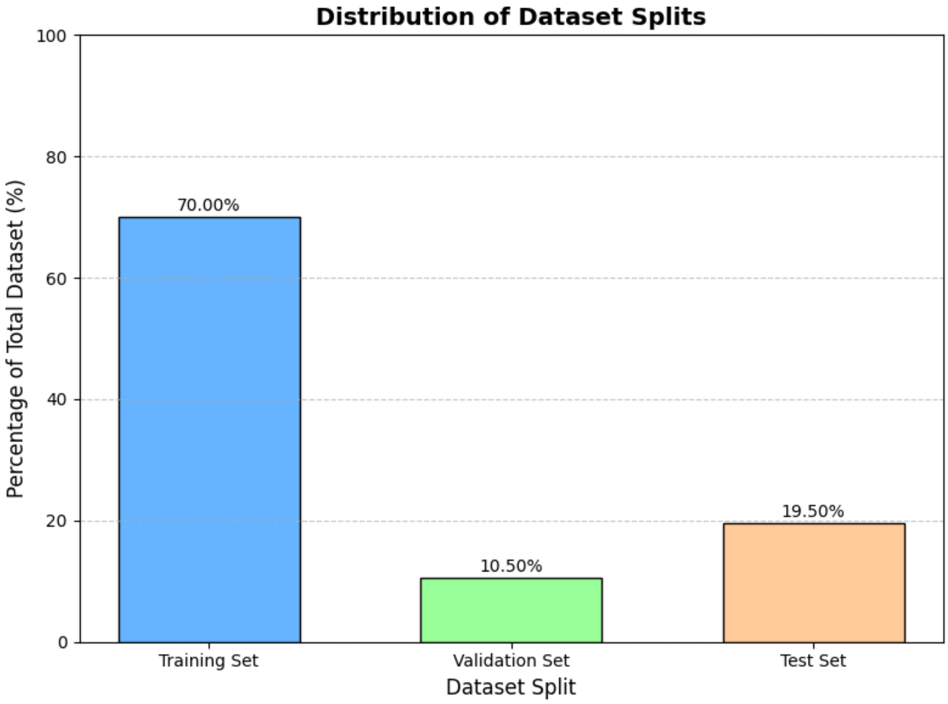

Figure 3.11: Train , Test and Validation Split of the dataset

Figure 3.12 : Wordcloud for Text and Summary Column

The above images illustrates the most recurrent words of the both column through wordcloud.

Table 3.1 : Sample Dataset

Text	Summary
রাজধানীর মোহাম্মদপুর ও আদাবর এলাকায় বুধবার সন্ধ্যার পর এক ঘণ্টার ব্যবধানে দুজনকে কুপিয়ে ও গুলি করে হত্যা করা হয়েছে। পুলিশ জানায়, সন্ধ্যা সাতটার দিকে মোহাম্মদপুরের চাঁদ উদ্যান সড়কের লাউতলায় পূর্বশত্রুতার জের ধরে আল–আমিন (৩০) নামের এক যুবককে কুপিয়ে হত্যা করা হয়। মোহাম্মদপুর থানার উপপরিদর্শক এস	মোহাম্মদপুর এলাকায় আল–আমিন (৩০) নামের একজন কে কুপিয়ে হত্যা করা হয়। পুলিশ কাউকে গ্রেফতার করতে পারেনি। রাত ৮ তার দিকে আদাবরে গুলি করা হয় ইব্রাহিম কে। পুলিশ দুই জন কে গ্রেফতার করলেও হত্যার কারন জানা যাই নি।

এম নুরুজ্জামান প্রথম আলোকে বলেন, আল–আমিনকে ধারালো অস্ত্র দিয়ে কুপিয়ে হত্যা করা হয়েছে। বিস্তারিত তথ্য এখনো জানা যায়নি। এখনো কাউকে আটক করা সম্ভব হয়নি।এর ঘণ্টাখানেক পর রাত আটটার দিকে আদাবর এলাকার নবোদয় হাউজিংয়ে নিজ বাসার সামনে গুলি করে হত্যা করা হয় ইব্রাহিম নামের আরেকজনকে।এ বিষয়ে আদাবর থানার ভারপ্রাপ্ত কর্মকর্তা এস এম জাকারিয়া প্রথম আলোকে বলেন, গুলি করে হত্যা করার ঘটনায় দুজনকে আটক করা হয়েছে। তাঁদের জিজ্ঞাসাবাদ করা হচ্ছে।থানা–পুলিশ সূত্রে জানা যায়, ইব্রাহিম পেশায় গাড়িচালক ছিলেন। নবোদয় হাউজিংয়ের একটি বাসার তিনতলায় ভাড়া থাকতেন তিনি। তবে কী কারণে তাঁকে হত্যা করা হয়েছে, তাৎক্ষণিকভাবে তা জানাতে পারেনি পুলিশ।	
রাজধানীর মিটফোর্ড হাসপাতালের সামনে ভাঙারি পণ্যের ব্যবসায়ী লাল চাঁদ ওরফে সোহাগকে (৩৯) নৃশংসভাবে হত্যার মামলায় গ্রেপ্তার তিন আসামির সাত দিনের রিমান্ড মঞ্জুর করেছেন আদালত। পুলিশের আবেদনের পরিপ্রেক্ষিতে ঢাকার চিফ মেট্রোপলিটন ম্যাজিস্ট্রেট (সিএমএম) আদালত আজ বুধবার এ আদেশ দেন। রিমান্ডপ্রাপ্ত তিন আসামি হলেন নান্নু কাজী, রেজওয়ান উদ্দিন ও তারেক রহমান (রবিন)। আদালতসংশ্লিষ্ট সূত্রগুলো বলছে, ব্যবসায়ী লাল চাঁদ হত্যা মামলায় গ্রেপ্তার নান্নু কাজী, রেজওয়ান উদ্দিন ও তারেক রহমানকে আদালতে হাজির করে ১০ দিন রিমান্ডে নিয়ে জিজ্ঞাসাবাদ করার আবেদন করে পুলিশ। আসামিপক্ষ থেকে রিমান্ডে নেওয়ার আবেদন নাকচ চাওয়া হয়। উভয় পক্ষের শুনানি নিয়ে আসামিদের প্রত্যেককে সাত দিন রিমান্ডে নিয়ে জিজ্ঞাসাবাদ করার অনুমতি দেন আদালত।এর আগে গতকাল মঙ্গলবার এ মামলায় গ্রেপ্তার মাহমুদুল হাসান মহিনকে দ্বিতীয় দফায় পাঁচ দিন রিমান্ডে নিয়ে জিজ্ঞাসাবাদ করার অনুমতি দেন আদালত।গত বুধবার	মিটফোর্ড হাসপাতালের সামনে লাল চাঁদ ওরফে সোহাগকে (৩৯) পিটিয়ে হত্যা করা হয়। এ হত্যার মামলায় এখন পর্যন্ত গ্রেপ্তার করা হয়েছ রেজওয়ান উদ্দিন, মো. নান্নু কাজী, সজীব ব্যাপারী, রাজীব ব্যাপারী, টিটন গাজী, মাহমুদুল হাসান মহিন, তারেক রহমান রবিন, আলমগীর ও মনির ওরফে ছোট মনির।

(৯ জুলাই) মিটফোর্ড (স্যার সলিমুল্লাহ মেডিকেল কলেজ) হাসপাতালের সামনে ব্যস্ত সড়কে প্রকাশ্যে নৃশংসভাবে লাল চাঁদকে হত্যা করে একদল লোক। পিটিয়ে এবং ইট-পাথরের টুকরা দিয়ে আঘাত করে তাঁর মাথা ও শরীরের বিভিন্ন অংশ থেঁতলে দেওয়া হয়। একপর্যায়ে তাঁকে বিবস্ত্র করা হয়। তাঁর শরীরের ওপর উঠে ঘাতকদের লাফাতেও দেখা গেছে। এ হত্যার ঘটনায় পরদিন রাজধানীর কোতোয়ালি থানায় একটি মামলা হয়। আলোচিত এ হত্যার মামলায় এখন পর্যন্ত নয়জনকে গ্রেপ্তার করা হয়েছে। তাঁরা হলেন রেজওয়ান উদ্দিন, মো. নান্নু কাজী, সজীব ব্যাপারী, রাজীব ব্যাপারী, টিটন গাজী, মাহমুদুল হাসান মহিন, তারেক রহমান রবিন, আলমগীর ও মনির ওরফে ছোট মনির। এরই মধ্যে আসামিদের বিভিন্ন মেয়াদে রিমান্ডে নিয়ে জিজ্ঞাসাবাদ করা হয়েছে।	

2. **Data Preprocessing :**

The scraped Bangla text included a significant amount of noise and inconsistency. Therefore, we implemented a strict preprocessing pipeline to clean and make it transformer-compatible. Bangla text can have encoding issues, punctuation errors, unnecessary HTML codes, and spacing errors. All of these can severely affect pasting text into the tokenizer and downstream models.

Unicode normalization was applied to various glyph different representations of Bangla characters in the dataset. Processes for removing noise removed HTML tags, navigation text, emoji, ads, etc. from the data. Issues with duplicated delimiters (commas and dots) (space) at the beginning or end of phrases in various languages or punctuation (dash) were solved by standardizing punctuation. We then performed segmentation of sentences using the Bangla matching punctuation markers (।,?,!,.) but taking care of the exception for example a numerical number or abbreviation etc.

The dataset was converted into tokens using the SentencePiece tokenizer of mT5 and the BanglaT5 tokenizer for model ingestion. To prevent memory overflow and computation issues, the input sequences were limited to 512 tokens and summaries to 200 tokens. The dataset which was cleaned was then split to train 80% validation and 10% testing and the 10% used for systematic tuning.

3. **Transformer Fine Tuning and Implementation :**

It is crucial to design the methodology effectively for low-resource languages like Bangla while fine-tuning transformer based seq-to-seq models so that convergence happens stably and overfitting, underfitting or catastrophic forgetting do not take place. To deal with the concerns, a two-phase incremental fine-tuning was used for BanglaT5 and mT5 models in this study. The design permitted the models to learn Bangla linguistics, domain-specific words, and crime-report story-telling patterns in a progressive manner. Experiments were conducted with the same Training Arguments to ensure consistency, reproducibility and controlled optimization behavior in experiments.

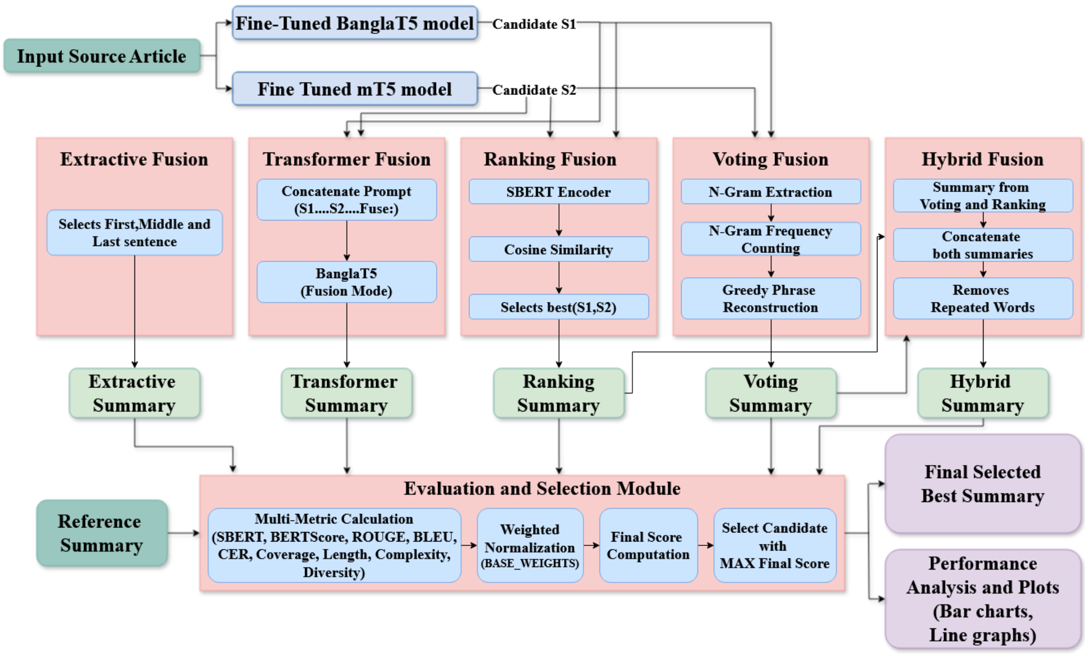

Figure 3.13 : Model Architecture of the proposed model

Phase 1: Initial Fine-Tuning on the First 2,000 Samples : In the first stage, each model was trained on the first 2,000 samples of the dataset. As a first step, the models were trained in this way which helped them to learn important structure and meaning related information essential for Bangla crime reporting. The training took place over 20 epochs. The per-device batch size was 10 and gradient accumulation step was 8. This helped size without moving beyond hardware limitations.

To quickly adapt at the beginning of learning for a domain, we chose 1e-3. The cosine-with-restarts scheduler was used to improve stability and exploration when training the agent. It restarts the learning rate so that the learning agent can escape shallow local minima. An additional 100 warmup steps were applied to stop new gradients from being applied all at once. Weight decay (0.01) was also used on the model to avoid overfitting.

Phase 2: Fine Tuned to Saved Checkpoints : In the second stage, training was resumed from the best checkpoint of Phase 1 and models were fine-tuned on the other 2,000 samples.This part gave the models more types of language. As a result, they could more easily apply what they learned to different kinds of writing. For example, news stories, fairy tales, and crime reports.

To maintain method consistency, the same TrainingArguments configuration was used. During the subsequent 20 epochs, the models improved their internal representations, learned long-range dependencies, and became more sensitive to details from context. Using the cosine-with-restarts learning rate schedule allowed more in-depth exploration of the loss landscape and avoided premature convergence.

4. **Ensemble Summarizer Framework :**

To make generated summaries more reliable, robust and contextually accurate, this study adopts a multi-strategy ensemble summarization model incorporating the strength of BanglaT5 and mT5. To avoid a single transformer model which sometimes generates incomplete, inconsistent, or heavily compressed summaries, the ensemble framework generates five candidate summaries based on different fusion mechanisms. These two approaches allow for broader semantic capturing, greater diminishing of model-specific biases, and an increased chance of choosing the most coherent, precise and contextually relevant final summary.

A. **Extractive Fusion:** Ensemble has Extractive Fusion as the baseline method. This method does not create new phrases like the abstract methods, rather it selects sentences directly from the article . Crime-report stories usually have a set way of telling.

- The first sentence introduces the main event or crime type.
- The middle sentence gives important information about who was victimized, where it occurred or when.
- The last sentence summarizes what happened as a result of the execution of a search warrant by police or the current state of the investigation.

Through the choice of the above positional sentences, Extractive Fusion will create a summary that is true and faithful to the text. Extractive summaries may not be the best written statements but serve to aid in providing coherence and supporting robustness, particularly in the face of abstract models that encounter ambiguity or noise.

B. **Rank Fusion :** Rank Fusion adds a semantic evaluation layer by analyzing the two abstractive summaries BanglaT5 and mT5 generates. This method operates in three primary steps.The summary are made on its own from both models.We encoded each summary and complete article into vector representations of high-dimensional spaces using Sentence-BERT (SBERT).Calculating cosine similarity of the embedding of the article with the embedding of each summary.

Rank Fusion output is the summary which has the highest degree of semantic similarity compared to the original document. This strategy does a great job of ensuring that selected summary reproduce the meaning, detail and semantic structure of the original text. It corrects for instances of hallucinations in one model or omission of an important detail.

C. **Voting Fusion:** By conducting n-gram analysis, Voting Fusion ascertains the common semantic core of BanglaT5 and mT5 outputs. It retrieves the consecutive two, three, and four-wording from the summaries and counts the frequency in both the summaries and gets the maximum frequency. Both models found this same information to be important, as indicated by the recurring n-grams. The ordering and alignment algorithm is then applied to build a meaningful sentence from the selected phrases. This enhances robustness by emphasizing consensus-driven messages while down-weighting model-specific discrepancies.

D. **Transformer Fusion :** Transformer Fusion helps put together the powers of BanglaT5 and mT5 for producing an improved summary in a more coherent way. In this approach, the summary output from both the models is combined and fed back into BanglaT5 using a structured prompt.

S1: <BanglaT5 Summary>

S2: <mT5 Summary>

Fuse:

BanglaT5 reads both of the summaries and creates a new summary by combining the best parts from both. As a result, the model can fix its errors, remove contradictory information, and generate a better-sounding summary. So, the final output from Transformer Fusion is usually more fluent and well-organized, and captures the overall meaning of the original article better than either summary's output.

E. **Hybrid Fusion :** Hybrid Fusion takes the complementary strengths of Rank Fusion and Voting Fusion. It combines the semantically aligned summary from Rank Fusion with the phrase-consensus summary from Voting Fusion. To enhance readability, duplicate vocabulary and overlapping information are eliminated, and the content is reorganised into smooth and coherent sentences. The outcome is a balanced abstraction that maintains semantic integrity while preserving key narrative elements.

Table 3.2 : Summary of the Strengths of The Summarizer methods

Methods	Strength
Extractive Method	Captures the main sentences, ensuring accuracy
Ranking Method	Highest semantic fidelity via SBERT similarity.
Voting Method	Captures shared meaning from multiple models.

Transformer Method	Creates clear and precise summaries.
Hybrid Method	Balance semantic correctness and coverage of contents

5. **Evaluation Metric and Summary Evaluation** : For the ensemble framework to select the best summary among many, we need a reliable and mathematical evaluation mechanism. Since the system produces five different summaries for each article, it must objectively evaluate which summary has maximum fidelity to the source text while maintaining coherence and readability. To achieve this, a set of metrics is designed to jointly evaluate the lexical fidelity, structural similarity, semantic adequacy, linguistic quality, and content preservation of the translations. The chosen metrics were widely used in the summarization lexicon for their ability to capture varied dimensions of summary quality. These metrics include BLEU, ROUGE, BERT Score, SBERT similarity, Character Error Rate (CER), Coverage, Length Ratio, Vocabulary Complexity, and Diversity .

3.3 Project Plan

The Bangla Crime Report Summarization System project was designed to follow a good sequence of project planning to enable the systematic progress of the project in the timely, methodological, and quality outcome. The plan was divided into many phases that depended on the success of the previous phase. Gradual implementation allowed for clear tracking, reduced risk, continuous enhancement, and coherent integration of all components of the system. Each phase of this narrative outlines the main activity, objectives and transition that guided the project from inception to the end.

Phase 1: Determine Summary and Understood Statement,

At the start of the project, a comprehensive requirement analysis was conducted to identify the nature of the crime-report texts in Bangla, system requirements, and prove the problem. At this point of the inquiry, the descriptive features of crime news - event sequence, legal terminology, narrative reporting - were analysed to find their summarisation implications. While doing the literature survey, we reviewed related works to take a look at what has been done in this research area. As these insights emerged, both functional and non-functional requirements were defined, finalizing the implementation decision of BanglaT5 and mT5 as the two core summarization models. This step was crucial for all the technical steps thereafter.

Phase 2 : Collection and Preprocessing of Dataset,

After we agreed on the requirements, we focused on developing a very large and high-quality dataset. From authentic online news portals, several thousands of Bangla crime-news articles were collected. The articles were picked to make them diverse in writing style, area coverage and types of crime. A comprehensive preprocessing pipeline was applied to each document since scraped text often has noise. We stripped out any HTML tags, normalized the Unicode characters, filtered out English letters and unnecessary punctuation, numeric patterns, space issues, and so on. The data was segregated for training, validation, and testing. (11) The cleaning and preparation of the data for fine-tuning the transformer models took place in this phase.

Phase 3: Training models with two-phase fine-tuning,

Phase 1 of the project was the heart of the project where BanglaT5 and mT5 were fine-tuned using a two-stage strategy in an incremental manner. Both models were trained initially on a subset of 2000 samples to learn basic structure of crime-report and nomenclature. The training settings consisted of 20 epochs, an effective batch size enlarged by gradient accumulation, a step size of 1×10^{-3}, and a cosine-with-restarts scheduler to aid stability. In the second stage, I further trained the saved checkpoints on the remaining 2,000 samples to help the models build generalization over more styles of writing without catastrophic forgetting. The robustness of models and performance of summaries improves significantly.

Phase 4: Ensemble Framework Development,

After the model was fine-tuned, the next step in the project was to design and develop the ensemble summarization framework. Integrate any five of the many summarization methods such as extractive fusion, ranking fusion, voting fusion, transformer fusion, hybrid fusion, etc. The Extractive Fusion took main sentence from sources (like in Extractive Summarization). The Rank Fusion with SBERT similarity based semantic similarity measure (the cosine similarity) for … The Vote Fusion created the summary from repeated n-grams in all summaries. Further, The Transformer Fusion synthesized the multiple summaries with a prompt that specifies the target summary length with constraint. Finally, the Hybrid Fusion uses the phrase level consistency with a modified semantic similarity. For the construction of a flexible and highly adaptive ensemble mechanism, the coordination of the model outputs, reconstruction algorithms and fusion logic was harmonized.

Phase 5: Calculate Metrics and Evaluation Pipeline,

We designed an exhaustive evaluation pipeline to objectively select the best summary among the ensemble outputs. During this phase, different quantitative measures that feed into the model were integrated. Included measures in the design are BLEU, ROUGE-L, BERT Score, SBERT cosine similarity, CER, Coverage, Length Ratio, Complexity, Diversity, Hallucination Score, Faithfulness Score, Keyword Preservation.

Summary evaluations of each candidate were done numerically. The metric was normalized and was also weighted based on their importance. For each methodology a final composite score was calculated to find out the optimal summary. The goal of this phase was to ensure fairness, transparency and scientific validity of the selection of summary.

Phase 6: Integration of System and Development of User Interface,

Once the summarization engine and evaluation framework were ready, the next step was to assemble all components into a web-based application. A Flask backend was created to send requests for summary generation to the ensemble algorithms, compute metrics, and organized results. The user interface allows the user to input text, view summaries, inspect metrics breakdown and visualize performance graphs. The research model must be transformed into a working tool in this phase. This tool should be usable, clear and responsive.

Phase 7: Test, Validate, and Optimize,

As after the assembling of the system adequate testing is performed for reliability and correctness. Functional tests evaluated the accuracy of each ensemble method, while stress tests assessed performance against long-string inputs. The tests of faithfulness and hallucinations guarantee that the summaries are factually and article-consistent. The hyperparameters along with weights were refined according to their respective behaviors to make the scoring more reliable. A more stable and user-friendly system was developed in this phase.

Phase 8 is the preparation of documentation,

report preparation and finalization. The thesis report was written the last phase along with drawing diagrams, formatting chapters, writing algorithms, and compiling results. Following academic formats, we prepared detailed descriptions on methodology, experiment workflow, evaluation metrics, and system architecture. We made our final defense using slides and materials we created.

3.4 Task Allocation

This section provides the detailed task allocation of the implementation of the Bengali Crime Report Summarizer system. The purpose of this plan is to proper distribution of the time , resources and workload.

The following table illustrates the major activities and timeline of the project extending from Week 12 to Week 48 .

Table 3.3 Project Timeline Table

Tasks	**Weeks**																		
	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48
Data collection phase																			
Data Preprocessing and Cleaning																			
Model Fine-Tuning and Training																			
Ensemble Integration and																			

Evaluation																			
System Testing and Optimization																			
UI Development and Web Deployment																			
Documentation and Final Report																			

3.5 Summary

This chapter describes in detail the method of designing the Bangla Crime Report Summarization System. The research design and the preparation of the dataset have been discussed first in this paper. All steps (i.e., data collection, data cleaning and normalization, tokenization) have tailored to the characteristics of Bangla text. The BanMis and MultiLoss fine-tuning techniques are discussed next. The former is adopted for the fine-tuning of BanglaT5/MultiT5 models. The latter assists in better domain adaptation through optimized application of rescaling and loss choices. A framework was built using Extractive, Rank, Voting, Transformer, and Hybrid Fusion to improve robustness and contextual relevance. To assess the quality of the summaries, we used a range of metrics which include BLEU, ROUGE-L, BERTScore, SBERT similarity, CER, Coverage, Length Ratio, Complexity, and Diversity. These metrics are quantitative metrics but they were fed into a weighted scoring function to give us the result. The chapter has also explained the entire project's plan and allocation of tasks to ensure the systematic execution of the project from dataset construction to deployment of model. Overall, the methodology was provided as a basis for the implementation, assessment, and real-life applicability of the summarization system which is elaborated further in the next chapter on implementation and results.

Chapter 4

Implementation and Results

This chapter describes the practical implementation, evaluation and performance of the proposed Bangla Crime Report Summarization System. It describes how the developed methodology was turned into an operational framework, i.e. technical environment, tools, and their configurations that were used in model training, ensemble construction, and web deployment. The chapter also performs the testing and performance analysis where multiple models and ensemble techniques are compared on quantitative metrics such as BLEU, ROUGE-L, BERTScore, SBERT similarity, CER, etc. Moreover, it discusses a variety of experiments used for taking the findings. It also compares and interprets visual performances. The chapter ends with a summary of key insights and improvements observed through the ensemble summarization architecture.

4.1 Environment Setup

For training, evaluation, and deploy the model, a computation and software environment was configured during the implementation of the Bangla Crime Report Summarization System . Given that the transformer-based architectures like BanglaT5 and mT5 are highly complex and heavy models , the environment was created in such a way that the different stages of execution involved a trade-off between computational performance, memory-wise efficiency, and reproducibility. The hardware specifications, software dependencies and framework configurations used in the project are discussed in the following sub-section.

Hardware Configuration

All experiments were conducted in Google Colab Pro+ using NVIDIA GPUs for testing and training. The computation resources were dynamically allocated, generally using NVIDIA Tesla T4 or A100 GPUs based on availability.

The key hardware parameters are summarized below.

- **Processor:** Intel Xeon 2.30 GHz
- **GPU:** NVIDIA Tesla T4 (16 GB VRAM) / A100 (40 GB VRAM)
- **Runtime:** CUDA-enabled environment for accelerated model fine-tuning

This setup ensured that large transformer models could be trained efficiently with optimized batch sizes and parallelized GPU computations.

Deployment Setup

The system deployed as a flask web app for practical implementation of the idea. The web interface allowed users to enter raw Bangla text, initiate the summarization process and view results directly in the browser. The Flask server used the trained models for ensemble summarization, calculating metrics, and generating plots in real time. Matplotlib plots were carried out in base64-encoded images for seamless integration in UI.

Summary

There are GPUs that perform high-level computations and a software environment that comes with modular software, like PyTorch or the Hugging Face Stack and evaluation toolkits. With help of this set up, we were able to train our system, experiment with it and deploy it conveniently. In the next section, an adaptation of the document will show the comparative study in the Testing, Evaluation and Comparative Study of Individual Models and Ensemble Models with the help of various Evaluation Matrices.

4.2 Testing and Evaluation/Performance/ Comparative Anallysis

After the completion of development and deployment of the Bangla Crime Report Summarization System, proper testing and evaluation have been performed for making sure that the system works properly and efficiently without any inconsistency. The focus of this phase was not only to check the functional correctness but also to perform a detailed comparative performance analysis of the individual transformer models, BanglaT5 and the mT5, along with the proposed ensemble framework incorporating multiple summarization strategies. The purpose of this phase was to find out the best way to generate coherent, contextually relevant, and semantically accurate Bangla summaries.

Testing Approach

The system's testing was done in multilevel to check the reliability of the system as a whole. In the Beginning unit testing was done to test each independent module like data pre-processing, text cleaning, tokenization and sentence tokenization. It made sure that the input data pipeline works properly and that no information gets lost during text normalization or token encoding.

After that, functional testing was done on the summarization engine to ensure the proper functionality of the end-to-end workflow that is from user input to the ensemble summary. Each of extractive, rank, voting, transformer and hybrid technique was experimented upon using varied lengths and styles of input in order to confirm the adaptability of the summarizer with various crime reports.

To find out how effective each model and method were, performance and comparative testing were conducted using a fixed evaluation pipeline. This means that the authors ran all of the models on the same dataset and calculated the same metrics.

Evaluation Methodology

The evaluation of the summary quality utilized a multi-metric approach to capture various aspects of summary quality. We supplemented conventional measures as BLEU and ROUGE with more deep semantic measures as BERTScore and SBERT similarity. Also, we added some other measures as CER, Coverage, Length Ratio, Complexity, Diversity, Faithfulness, and Keyword Preservation to measure linguistic quality and factual accuracy.

Every summary that we generated was evaluated on two primary dimensions.

Lexical Accuracy, measuring with BLEU and ROUGE-L how aligned the generated entry was with the reference or actual article in terms of word overlap and sentence construction.

Semantic Fidelity evaluates meaning preservation and contextual relevance between an article and its summary using BERTScore and SBERT.

Further improvements took place through metrics such as the character error rate (CER), which quantifies intra-word distortions, and coverage, which indicates how much essential content was retained. The Length Ratio made sure the summary is not too short and not too long. On the other hand, the metrics Complexity and Diversity measured linguistic richness and reduced redundancy. Also, Faithfulness and Keyword Preservation verified the generated summary is factually accurate and the important named entities in the summary are preserved. These named entities include the victim name, suspect name, and crime type.

Each metric was normalized between 0 and 1 for compatibility. Furthermore, the composite scoring function assigned weights for each metric. The final weighted score determined the best ensemble method for each article's summary.

Performance Insights :

The graphical visualizations of the result of evaluation showed the same trend for the all test samples. The ensemble system scored high in semantic similarity, faithfulness, and coverage showing that it succeeded in getting the right key points without hallucinated/fake information. The Length Ratio and Diversity metrics confirm that the summaries are appropriately summative concise yet informative, and linguistically varied without excessive redundancy.

Since each model offset the disadvantages of the others, the multi-model ensemble proved advantageous. This was highlighted by the results. The hybrid approach gave the best overall performance. This shows us that such fusion of transformer-based models can yield accurate and stylistically natural summaries.

Discussion and Evaluation Findings :

The results of the evaluations show how important it is to have many evaluation metrics to grab all things good and bad about the summaries. Lexical metrics (e.g., BLEU, ROUGE) are useful to assess the degree of structural overlap between summarised and reference content. Semantic metrics (e.g., BERTScore, SBERT) are helpful in evaluating the extent to which meaning is effectively preserved. Ensuring Factual Accuracy and Linguistic Quality through CER, Coverage, Faithfulness, and Keyword Preservation was part of our goal.

The ensemble system which outperforms individual models shows that no one summarization model is the best one. However, the outputs of these models can be fused together. Then, these fused outputs can be evaluated using a multi-dimensional metric framework. This results in a strong balance of precision, readability, and factuality. The low values of CER and high values of Coverage along with high values of semantic score signify that the summary produced was grammatically rich and full of content.

4.3 Results and Discussion

In this chapter, we present the implementation, results, and comparative study of the Bangla Crime Report Summarization. This system combines two transformer-based architectures – BanglaT5 and mT5 – with an ensemble framework to improve accuracy, fluency and contextual relevance of Bangla crime report summarization. The following section show how the model was trained, evaluation metric and performance comparison with the help of visuals and tables.

The model training, evaluation, and ensemble testing results show the ensemble framework performed better than individual transformer models. The metrics evaluated in this work where limited to BLEU, ROUGE, BERTScore, SBERT similarity and Character Error Rate (CER). Also linguistic metrics like Coverage, Length ratio, Complexity and Diversity.

1. **Ensemble Model**

   Figures 4.1 and 4.2 show the methods vs average metrics for all the summarization methods and evaluation of the summarization methods over Final score respectively .

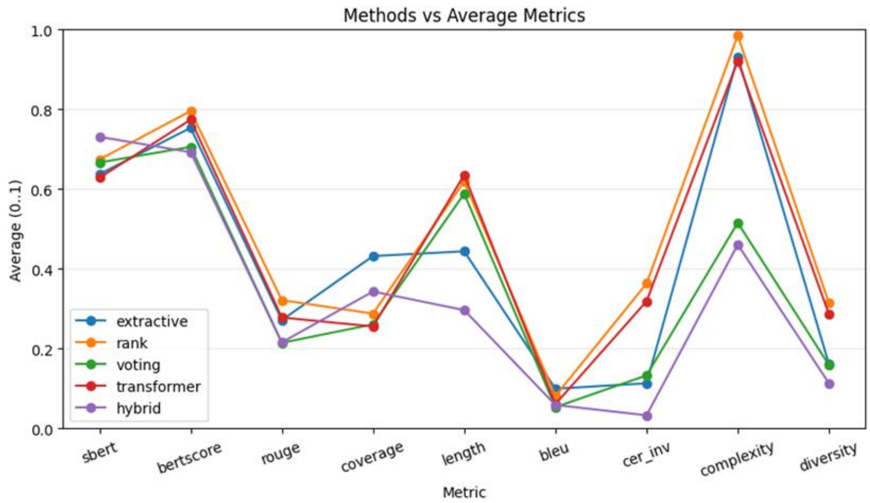

Figure 4.1: Methods evaluation

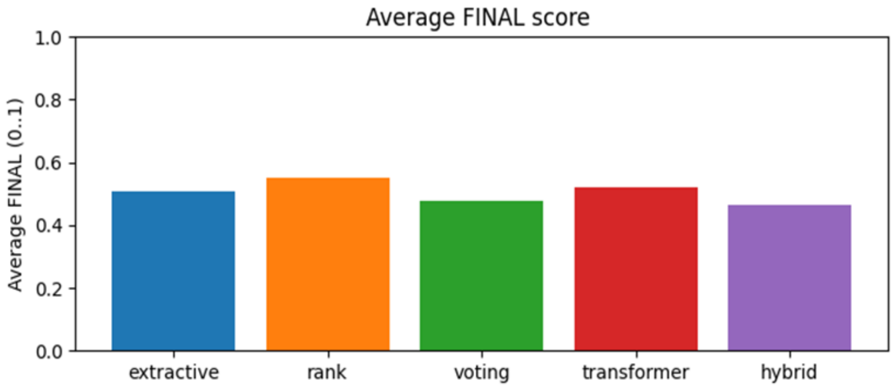

Figure 4.2 : Methods Evaluation for Final Score

Comparative Ensemble Performance:

The comparative performance graphs show the efficacy of each ensemble method, i.e., Extractive, Rank, Voting, Transformer, and Hybrid, as they are evaluated on all measures. Out of all the methods available, the Rank and Transformer combination offers the best performance in semantic similarity ranking. Although Hybrid method have achieved a good score but it performs poorly in fluency.

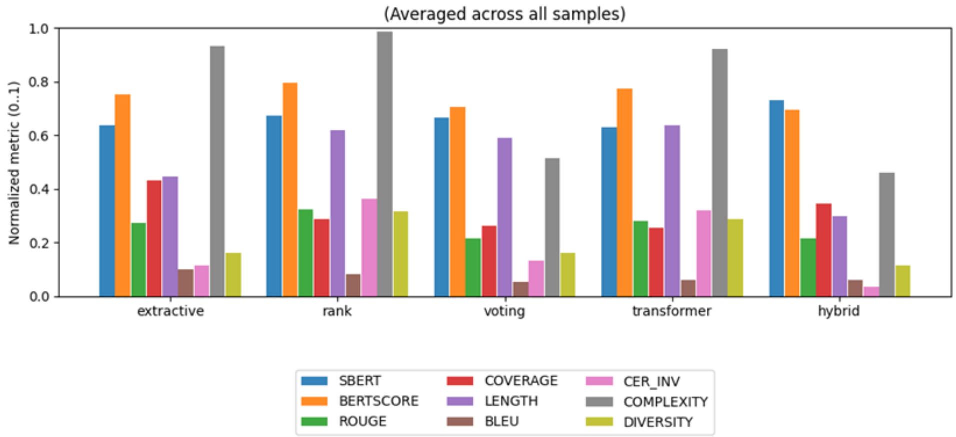

Figure 4.3 : Methods evaluation over all the metrics

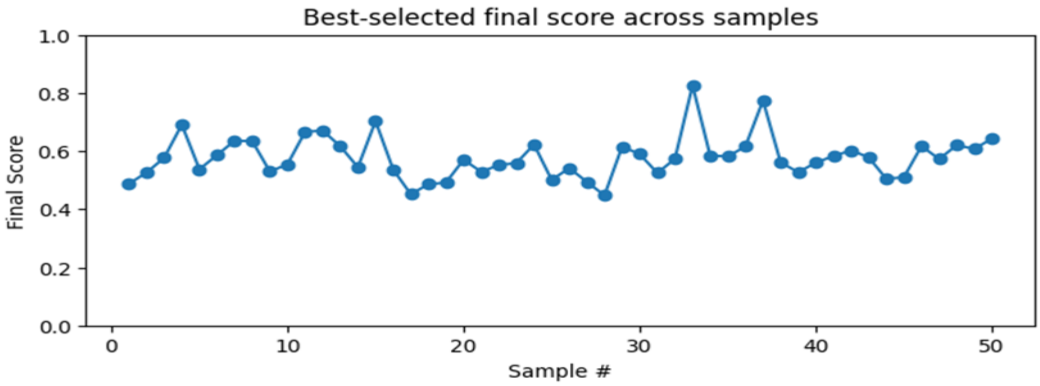

Figure 4.4 : Average Final Score on Test Sample

Figure 4.4 Illustrates the average final score throughout the all sample test set for this case it is 50.

2. **Traditional BanglaT5 and mT5 Model :**

According to the Figure shown (Figure 4.6), the BanglaT5 model follows a consistent downward trend of training loss within the first 50 training steps. This shows that the model quickly learns the language from the data. The training loss is close to zero at around 200 steps, meaning the model is capable of decreasing its reconstruction errors. The validation loss stays at some 0.1, staying on a similar trend in further epochs. The plateauing trend indicates that the model has found a sweet spot between underfitting and overfitting: it is learning sufficiently to generalise well to unseen data as the training and validation curves do not differ significantly. The loss values declining smoothly suggest that parameters have been updated correctly, and the model's pre-trained weights were successfully fine-tuned to features like criminal entities, incident descriptors, and temporal expressions. The ability of BanglaT5 to successfully fine-tune for monolingual Bangla summarization with minimal instability is established.

The training and validation loss of mT5 model is shown in Figure 4.7. mT5 also converges efficiently but comparatively slowly. In the beginning of the learning phase (up to ~100 steps), the training loss face is effectively constant but exhibits small perturbations (not visible in the chart) mid a learning phase motor through~

100 steps. This reflects that the model is learning Bangla language structure; which is not entirely present in its multilingual training set. After 250 steps, these two curves do stabilize around 0.05–0.08, which suggests they both converge uniformly. It also suggests better regularization due to the inclusion of beam search and weight decay. mT5 has lower overall loss values, which suggests the model has gone through better optimizations. That said, the validation curve of this model is quite flat compared to BanglaT5 and this indicates a less degree of language-specific generalization. This happens because mT5's cross-lingual embedding space is strong. But it can weaken performance for strictly monolingual as well as domain-specific tasks like Bangla crime reporting.

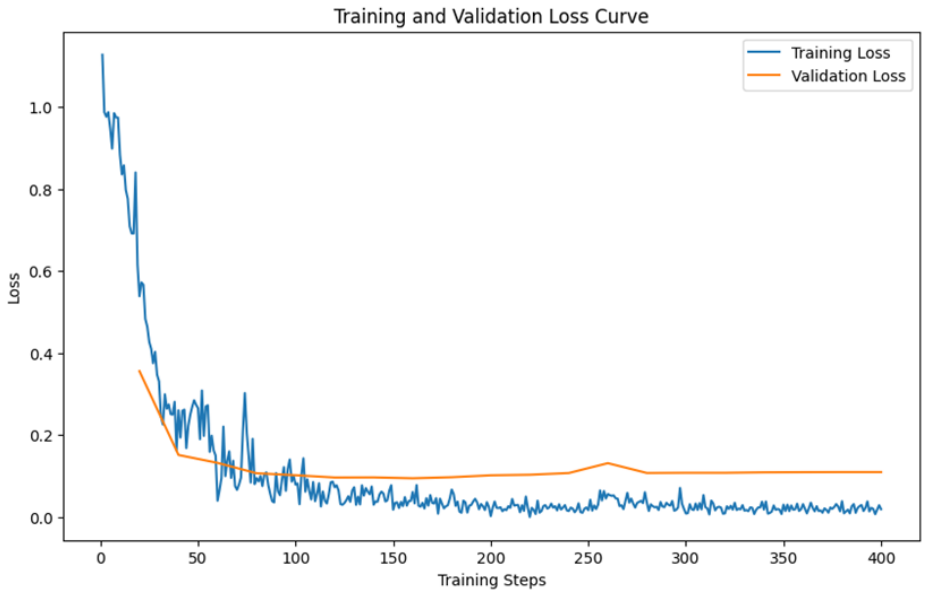

Figure 4.5 : Train and Validation Loss Curve For BanglaT5

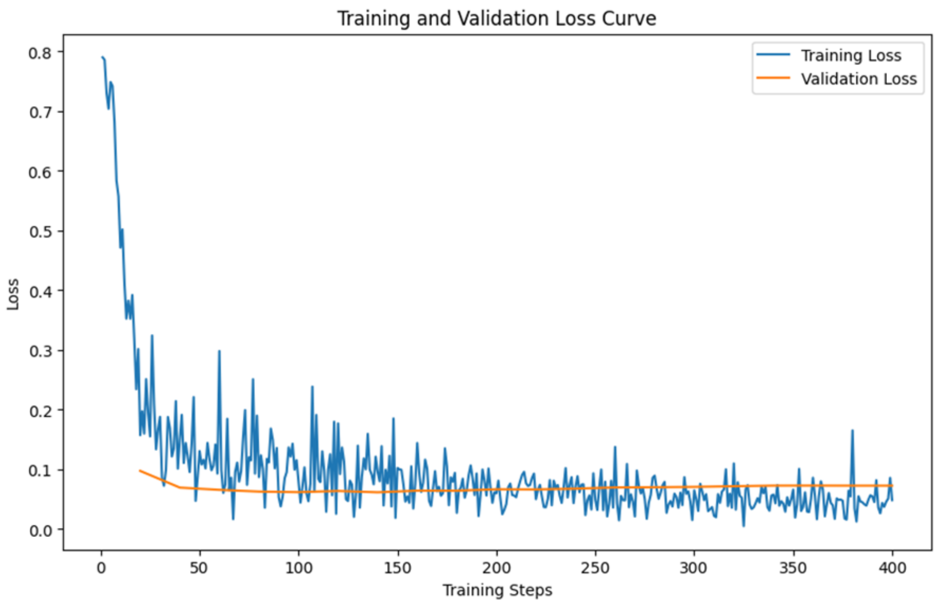

Figure 4.6 : Train and Validation Loss Curve for mT5

3. **Result Analysis :**

BanglaT5, mT5, and our ensemble model are compared in terms of performance which shows significant difference in terms of meaning, context, and structure. When we compare the standalone models, we see mT5 is slightly better than BanglaT5 on all metrics including a higher SBERT similarity of 0.6505 v. 0.6422, slightly higher BERTScore 0.7938 v. 0.7903, and higher ROUGE-L 0.3173 v. 0.2911. This means mT5 captures the relationship of the semantics more effectively and preserves key information during summarization. On the other hand, the ensemble model provides an improvement by taking advantage of both architectures. The ensemble's SBERT value is highest at 0.6852 indicating that this the closest semantic similarity to the original text, despite its ROUGE-L score of 0.3052 being slightly lower than mT5. The BERTScore (0.7924) is relatively stable across evaluations. To sum up, the findings reveal that advantageous as mT5 can serve as a standalone model; the ensemble outcome produces a summary that is more semantically faithful and contextually enriched.

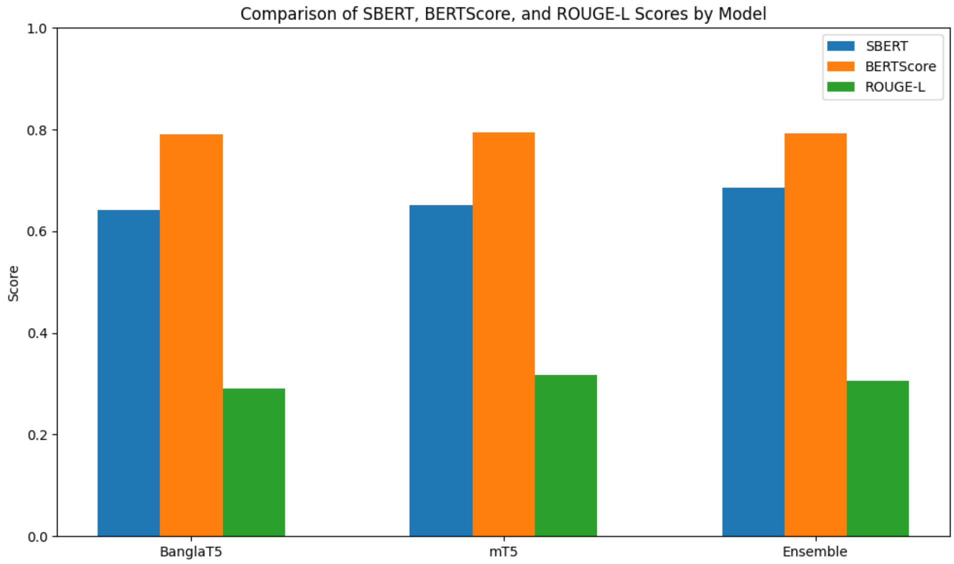

Figure 4.7 : Comparative Analysis of all the models

4.4 Summary

According to the results of this study, it has been observed that our proposed ensemble summarization framework outperforms BanglaT5 & mT5 individually in different metrics. Although the individual models produce good results on semantic similarity and lexical overlap, the overall balance and robustness as a result of combining different fusion approaches benefit from the ensemble. Quantitatively, the ensemble model shows the highest SBERT similarity (0.6852) which indicates better semantic preservation. Also, it maintains competitive BERT Score and ROUGE-L values. The plan that uses different strategies also helps reduce mistake rates, increase coherent in context and make everything a lot more reliable and robust. Training and validation curves also assure us of a stable combination for both BanglaT5 and mT5. Visual comparisons showed the ensemble method was consistently better. In general, the findings confirm that the ensemble of different abstractive and extractive summarization pathways is effective and therefore, the ensemble model is the most reliable and high-performing model for Bangla crime-report summarization.

Chapter 5

Engineering Standards and Design Challenges

The engineering standards, design considerations and constraints for developing the Bangladesh Crime Report Summarization System is discussed in this chapter. The project shall define and justify the software, hardware, and communication standards used in the project, whether they are made up or commercial. The project shall also discuss ethical, social and sustainability impacts. Evaluation of project management and costs and the mapping of the project work against the categorization of complex engineering problems used in the course.The objective of this chapter is to show that the system was designed according to accepted engineering principles while taking into account practical reality and social responsibility.

5.1 Compliance with the Standards

5.1.1 Software Standards

1. IEEE 830 / ISO/IEC/IEEE 29148 Software Requirements Specification (SRS) Standard.

Description:

Guidelines to collect, document, and validate software requirements. 4th section is for approval. 12 words. It ensures writing requirements in a clear verifiable and structured manner so that it gives appropriate functional non-functional interface and performance requirements. During the development of systems, it leads to uniformity and full coverage.

Rationale for Choice:

The summarization system consists of different components (data pipeline, model training, ensemble engine, evaluation metrics, and web interface). Using the IEEE/ISO SRS helped us in eliminating ambiguity in stating the expected behavior of each component. With multiple interacting modules present in an NLP system, it was also ensured that there would be traceability from requirements to implementation.

The lack of structure in alternative informal requirement formats could lead to inconsistencies during model evaluation.

Pros:

- Ensures clear and structured requirement documentation.
- Removes confusion among developers, evaluators and stakeholders.

- Supports formal verification and traceability.
- Widely recognized in academia and industry.

Cons:

- Preparing it takes more time and efforts than informal documentation.
- Might be seen as too rigid for experimental machine learning projects.
- Constant updates are necessary because research models are evolving.

2. PEP 8 Python Coding Standard

Description

PEP 8 is the official style guide for Python programming. It specifies how to indent, name, comment, add spaces, structure files, making it easy to read by others. It is widely recognized by research and industry worldwide.

Rationale for Choice

The project primarily relies on Python to train models, implement ensemble logic, and evaluate. Adhering to PEP 8 ensures that your code is readable. It's important to be consistent when working with many collaborators or when a project might be further developed.

We considered other custom coding standards but they weren't broadly adopted or very portable, so we rejected them.

Pros:

- Ensures uniform and readable code.
- It curbs mistakes from uneven indentation or naming.
- Enhances maintainability and collaboration.
- Well-supported and widely followed in the Python ecosystem.

Cons:

- Strict adherence can slow down rapid prototyping
- Developers unfamiliar with PEP 8 require a learning curve.
- Does not address performance optimization (only style).

5.1.2 Hardware Standards

1. IEEE 1016 – Recommended Practices for Software Design Descriptions (Applied to Hardware Configuration Documentation)

Description

IEEE 1016 discusses the system architecture with the architecture of hardware to be used. In machine learning projects, this standard preserves clarity in documentation of the GPU/CPU (graphics processing unit/central processing unit) configurations, memory, storage and execution environment requirements. Even though it was made for software components, one may apply the principles directly to the documentation of the computational infrastructure supporting model training.

Rationale for Choice

To train large transformer models such as BanglaT5 and mT5, a standardized and detailed hardware and configuration setup is required. This is because larger models require more powerful GPUs with the requisite VRAM or a faster clock speed for training that is faster and efficient. By following the rules of IEEE 1016, all hardware specifications were recorded in a consistent manner so that other researchers can reproduce it in their research environment such as Kaggle, Google Colab, Institution's own GPU servers, etc.

Pros

- It is essential to document computational environments clearly and systematically.
- Helps in reproducing the results of Machine Learning program
- Standardized logging helps identify performance issues.

Cons

- It only requires documentation standards, not hardware.
- Needs more effort to keep descriptions accurate.
- Could be less used if fast testing is needed.

2. Open Compute Project (OCP) Standards – GPU/Server Efficiency

Description

Guidelines for data-center hardware that is energy-efficient, scalable, and modular. The focus of OCP-aligned systems is on power, cooling and design. While the study does use GPUs in the cloud, these cloud providers are OCP compliant themselves.

Rationale for Choice

There is a lot of text transformation-based model generating which needs models and hardware for faster training and more energy-efficient usage. The NVIDIA Tesla T4, P100, and V100, which are utilized in this project, are Cloud GPUs that follow the OCP hardware principles. It ensures reliable operation and efficient heat dissipation to ensure consistent performance during training.

Pros

- Efficient computation is beneficial for deep learning workload.
- Less thermal throttling leads to longer training.
- Makes sure everything works with the standardized cloud.

Cons

- Researchers using cloud systems do not always have hardware-level control.
- Cloud based setups cannot have upgrades or personalization made to them.
- Some environments do not implement OCP which may create variance.

3. NVIDIA CUDA Architecture Standard

Description

NVIDIA's parallel computing platform and programming model, CUDA, lets you use a GPU to do computations. It tells a deep-learning framework like PyTorch, TensorFlow, everything from how it guides the GPU hardware to perform optimal matrix operations to how the model training should run.

Rationale for Choice

The Transformer operations (self-attention layers) of both BanglaT5 and mT5 are highly parallelizable. CUDA accelerates tensor computations to make them run more efficiently. Without CUDA, the training speed would be greatly reduced, experimentation would not be possible.

Pros

- Provides significantly faster performance for matrix and deep learning operations.
- Fully supported by PyTorch and Hugging Face Transformers.
- Creates a driver-level optimization to stabilize long training.

Cons

- Requires NVIDIA hardware no compatibility with AMD GPUs.
- CUDA version mismatches can cause installation conflicts.
- Due to the greater demands on GPU memory, low-budget setups may not be able to handle a large model size.

5.2 Impact on Society, Environment and Sustainability

The Bangla Crime Report Summarization System has larger implications than just its technical performance. Since it's an AI tool that processes, summarises, and rewrites text related to a crime, it must be analysed for social, ethical and environmental impact. This segment looks at the impact of the technology on users, communities, technological ecosystems as well as long-term sustainability. To international standards of responsible

AI development, engineering ethics, and sustainable computing principles, we direct the discussion.

5.2.1 Impact on Life

The scheme of the proposed system can improve the personal and working lives of the users. Specially for the users working on crime information and relevant areas. Investigative journalists, researches, intelligence analysts and law enforcement people as well as policy maker interfaces with extremely large size of crime reports, which are often lengthy, repetitive and difficult to analyse in an efficient manner. The system reduces the cognitive burden of users which helps in the faster and accurate processing of information by generating concise and coherent summaries encapsulating important entities like victims, suspect, type of crime, and incident context.

Educational users such as students or academic researchers will find it useful for understanding the complex narrative structures of Bangla crime journalism. This helps distil hard to read text and make them easier to read for better literacy. People who live in the communities benefits from better access to information about crime thus community awareness would increase and that would help people in making decision about safety and social issues.

While it is beneficial, it also raises serious questions about fairness and accountability. Machine-generated summaries should not exaggerate, misinterpret, or leave out essential facts as these could unintentionally mislead the reader's understanding of events. It is also important to be careful not to stigmatise vulnerable groups or reinforce negative stereotypes. Therefore, although the system may offer qualitative advantages in everyday life, its practice must be aware of ethics and continual observation.

5.2.2 Impact on Society & Environment

Impact on Society

At the level of society, the summarization system helps in the smooth flow of information in the media and public sector of Bangladesh. In Bangladesh, crime journalism is a core part of daily news and its high frequency often overburdens the reader and analyst. The system helps to make news reports more efficient and enhance investigative journalism. It also helps a governmental or legal authority review cases more quickly and accurately through the automatic generation of structured and concise summaries.

Additionally, this initiative pushes the development of Bangla Natural Language Processing (NLP), which is a field known for having fewer resources and technology. By creating a dataset for Bangla data and the implementation of advanced ensemble model, we can actively boost the ecosystem of Bangla AI applications. This will allow the Bangla-speaking communities to utilise digital and technological innovations in artificial intelligence.

Nevertheless, societal risks also require attention. Automated systems that sum up information may unintentionally put more focus on certain crimes than others. This can even lead to the removal of some important information that is sensitive in nature. Some ways people misuse it are creating wrong summaries and altering information for agendas. It is very important in the event of a failure of the system to stop its exploitation.

Impact on Environment

Even though it is a digital solution, we cannot ignore the environmental cost of training deep learning models. BanglaT5 and mT5, two transformer-based architectures, are very heavy on computation. Therefore, they require prolonged usage of GPU, which uses extensive energy and leads to carbon emission. As we do hyper parameter tuning again and again and again for model validation, the environmental cost piles on.

These concerns led to the adoption of numerous measures with sustainability in mind during development. We efficiently used cloud GPUs for training, which helped minimize idle resources and performing the computation as scheduled. To avoid unnecessary retraining, a two-phase fine-tuning strategy that uses the target dataset without supervision was implemented. Also, GPU utilization was reduced due to other optimization practices like gradient accumulation, early stop, and dataset refinement.

Once it is deployed, the system consumes very little energy as inference uses far less resources than training. Because it concedes to its carbon emissions during its development, the impact of the system over the long run is much smaller than that of industrial-scale AI models.

5.2.3 Ethical Aspects

The project heavily involves ethical issues because it deals with crimes and things related to crime. Crime reports usually contain sensitive information about victims, suspects, and minors. They also include details about the community where violence and other illegal conduct takes place. As such, the summarization system must be designed to follow responsible AI principles so that it doesn't cause harm or misrepresent real events.

1. Privacy and Data Protection:

Reports of crime submitted by users may be identifiable. To keep user information private, the system won't store user input and will summarize only the associated required information. For legal and ethical reasons relating to confidential data, only news articles that are available publicly were used to train the system.

2. Bias Minimization:

When machine learning models are trained on news sources, they can inherit the bias reflected in those news sources. To mitigate this risk.

- To ensure balanced representation data was collected from different Bangla news agencies.
- A team of models reduces the impact of any individual bias.
- Metrics like faithfulness and keyword preservation evaluate factual consistency.

3. Accuracy and Faithfulness:

Generating a summary must honour the essence of the original text. A common risk in NLP models is hallucination which is the tendency to invent details. To show that the semantic content of the summary and source text is aligned, we used SBERT similarity. Likewise, coverage and faithfulness metrics.

4. Responsible Deployment:

- The system is not designed for.
- Predicting criminality or behavioral traits.
- Profiling individuals.
- Making legal or judicial decisions.
- Roles for special datasets, oversight by law, and ethics around such applications. The system's scope is strictly informational.

Through the integration of the measures, the project is in line with global AI ethical frameworks like those of the UNESCO's AI Ethics Principles and the OECD Responsible AI Guidelines.

5.2.4 Sustainability Plan

The sustainability of AI is not just about the environmental impact but also about the robustness, reliable operation and positive development impact over time. The sustainability plan will allow the proposed summarization system to remain functional, flexible and relevant.

1. Technical Sustainability.

The system utilizes a modular architecture that allows for adding incremental updates to the datasets, pre-processing scripts, and transformer models without breaking the pipeline. The ensemble architecture allows for easy integration of future models making the system scalable with new Bangla NLP models.

2. Operational Sustainability.

It's inexpensive to use inference, so even a medium-end hardware or cloud server can use it.Cloud deployment cuts down maintenance costs and provides ongoing access. Execution is lightweight and cost-effective for education, government or journalism.

3. Societal Sustainability.

The system positively impacts journalism, analysis of public safety, and academic research by increasing access to data on crime and supporting data-driven decision-making. Moreover, Bangla NLP resources build up the technological infrastructure of the region.

4. Environmental Sustainability.

- Several eco-friendly approaches were utilized.
- Efficient GPU scheduling to avoid resource waste.
- Shared cloud computing to reduce dedicated hardware needs.
- Model optimization techniques to reduce energy consumption.
- All of these tactics help reduce machine learning research's impact on the environment.

5.3 Project Management and Financial Analysis

The project proceeded through the standard life cycle phase of software engineering from requirement gathering to implementation, testing and then deployment. The key phases include.

Requirement Analysis and Planning:

Identification of functional and non-functional requirements, dataset requirements, and model specifications.

Design and Architecture Development:

Module-level architectures are designed to perform preprocessing, model training, ensembling and evaluation.

Dataset Preparation and Preprocessing:

Gathering massive Bangla crime reports and preparing high-quality inputs for training models.

Model Training (BanglaT5 & mT5):

A two-phase fine-tuning strategy for training models

Ensemble Model Development:

Bringing together extractive, ranking, voting, transformer and hybrid fusion methods.

Evaluation and Optimization:

Using numbers to check and improve output quality.

Deployment and Demonstration:

The roles involved making a prototype interface and evaluate performance on unseen data.

The project's each phase was managed with objectives, timelines and deliverables to ensure smooth coordination and engineering standards

5.3.1 Cost Analysis :

The process of developing the Bangla Crime Report Summarization System did not bear any cash cost. Because this project used academic resources which are available free of charge, open-source tools, and infrastructure which is provided by the university. While you may not have spent any money, you should still report the implied or theoretical cost structure. This shows you're aware of the money involved in a machine learning project.

The project's entire workflow including dataset collection and preprocessing, training and evaluating models, and finally demonstrating it on a web-based interface was done on Google Colab, Kaggle and Google Drive for free as part of normal educational use. All the software framework (PyTorch, Transformers, Sentence-BERT, ROUGE, BLEU, Flask) were similarly available for free, open source.

Table 5.1 : Cost Analysis

Category	**Description**	**Actual Cost (USD)**
GPU Training	Google Colab Free Tier + Kaggle GPU	$0
Data Storage	Google Drive (free quota)	$0
Software/Libraries	All tools open-source	$0
Labor/Development	Student-led research	$0
Deployment	Local/Colab-based demo only	$0
Miscellaneous	None incurred	$0

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

In this section, provide a mapping with problem solving categories. For each mapping add subsections to put rationale (Use Table 5.1). For P1, you need to put another mapping with Knowledge profile and rational thereof.

Table 5.2: Mapping with Complex Engineering Problem.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✔	✔	✔	✔			✔

EP1: Depth of Knowledge:

Advanced knowledge is required in transformer-based NLP, semantic similarity modelling, ensemble fusion methods, and multi-metric evaluation techniques. Summarizing Bangla crime reports is not a trivial task. It needs a proper understanding of machine learning, Bangla language linguistic structure and model optimizers.

EP2: Range Of Conflicting Requirements :

The system must balance multiple conflicting goals at the same time for example,

- Making summaries short but still meaningful.
- Ensuring high accuracy without increasing computation time.
- Avoiding hallucination by remaining faithful to the text.
- Producing summaries understandable by humans that also satisfy evaluation metrics.

EP3: Depth of Analysis:

The system must balance multiple conflicting goals at the same time for example,

- Making summaries short but still meaningful.
- Ensuring high accuracy without increasing computation time.
- Avoiding hallucination by remaining faithful to the text.
- Producing summaries
- understandable by humans that also satisfy evaluation metrics.

EP4: Familiarity of Issues :

Crime-report summarization is a topic not usually known to the CSE students. The domain involves legal, forensic and narrative complexities so it requires the learning of new concepts and studying a different type of text. Thus the problem becomes unique and thinks exploratory.

EP7: Interdependence :

The architecture used for designing the proposed summarization system is modular. Each component in the proposed system performs a specific task and passes on the output to the next component in the system. To start with the Dataset Collection module collects Bangla crime-news articles and their summaries from the trusted websites. Next comes the Preprocessing Pipeline whose job is to clean the text. This is done through punctuation removal, handling of null, numeric and English letters, and Unicode normalization. After the preprocessing, the model development is completed in two modules: the BanglaT5 Training module and the mT5 Training module. These modules fine-tune their respective transformer model on the prepared dataset. In the Ensemble Fusion Layer, the outputs are combined according to various fusion strategies which yield diverse candidate summaries. The Metric Evaluation Module assess the candidates using measures like SBERT similarity, BERTScore, ROUGE, BLEU, CER. Lastly, the Summary Selection module selects the best summary based on weighted score. Since the system is built in a modular way, it works individually as well as clearly to create an efficient system.

Mapping with Knowledge Profile

In this section is we mapped the overall problem and EP1 ***(multiple between K3, K4, K5, K6, K8 for attaining EP1)*** to the Knowledge Profile.

Mapping with Knowledge Profile

This section is describes mapping of the the overall problem and EP1 ***(multiple between K3, K4, K5, K6, K8 for attaining EP1)*** to the Knowledge Profile.

Table 5.2: Mapping with knowledge Profile.

K1 Natural Science	K2 Mathematics	K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K7 Comprehension	K8 Research Literature
		✔	✔	✔	✔		✔

K3: Engineering Fundamentals :

The designing of system workflow, pipeline and deciding relationships between various component's like Preprocessing, Model Training, Ensemble Evaluation were all based on Engineering fundamentals. Key computing ideas like data structures, algorithms, and other things were vital for the system's scalability and reliability. These principles were obtained from the foundational CSE courses and formed a technical basis for building a complicated NLP system.

K4 – Specialist Knowledge.

This project consumed the expert knowledge in machine learning and Natural Language Processing (NLP). In order to adapt BanglaT5 and mT5 for domain-specific summarization, it was necessary to understand transformer architectures, tokenization strategies, attention mechanisms, fine-tuning procedures, and sequence-to-sequence modelling. My knowledge came from advanced classes, online research, technical papers, and trial-and-error with deep learning frameworks like PyTorch and Hugging Face Transformers. The limited set of domains improved model configuration and performance.

K5 – Engineering Design.

Architectural features adapted by engineering design principles to produce the summarizer that is logical, compact and modular design. The design of modules was such that they could work independently even though they contributed towards doing the same task. The decision to use a two-phase fine-tuning strategy, to use five fusion methods and to use weighted scoring were made deliberately with engineering practice in mind. These skills were gained through university project courses and applied research experiences with a focus on sequential design.

K6 – Engineering Practice.

The practice of engineering was demonstrated through hands-on implementation. Due to practical challenges in the real-world such as Noisy Bangla text, GPU memory issues, Training instability, and dataset format variations. Learning from coding labs, machine learning tasks and previous programming projects assisted greatly in making the implementation robust and efficient, and reusable. This hands-on knowledge also assisted with model training, runtime optimization, and deployment.

K8 – Research Literature.

Having knowledge of research literature helps with understanding what has been done in Bangla NLP, transformer-based summarization and ensemble frameworks. By reading academic papers, open-source documentation, and benchmarking studies, they made informed decisions regarding choice of model, evaluation metrics, and architecture. Due to the knowledge base of the project, the work will match research trends and create an ensemble framework to surpass classical summarization methods.

5.4.2 Engineering Activities

This section, provide a mapping with engineering activities.

Mapping with Complex Engineering Activities

This section shows the mapping of the overall problem and EA's ***(multiple).***

Table 5.4 : Mapping with Complex Engineering Activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✔	✔	✔		✔

EA1: Range of resources

The project makes use of a large number of relevant tools: transformer models, GPU resources, python libraries, datasets, evaluation metrics, web interface.

EA2 Level of Interaction

For this project we have to interact with our supervisors-supervisor and fellow class mates for helps.

EA3 Innovation

We proposed a transformer based ensemble project based on two transformers BanglaT5 and mT5.

EA4: Impact on Society and Environment

Our project have an great impact on the society, it can help people to learn about what is going around them easily , making information easier for them and make them aware.

EA5 : Familiarity.

Summarization of the crime-report in Bangla is an unknown, unusual area where superior engineering principles are applied.

5.5 Summary

To sum up, this project's engineering efforts show a considerable level of technical depth, interdisciplinary interaction, and social significance. The system needed the complex use of several computing resources including the large language model, GPU infrastructure, and evaluation libraries, thereby making strong use of EA1. The interaction between multiple subsystems comprising preprocessing modules, dual transformer models, a multi-strategy ensemble layer, and a metric-guided decision engine is indicative of EA2 integration. The project is also a realization of EA3 as it unveils a novel ensemble-based summarization architecture concerning Bangla crime reporting, for which there are no existing solutions. It also aligns with EA4 to summarize crime info; as well as to train the model environment. Finally, EA5 was fulfilled by requiring skills not typically encountered in academic experiences due to the challenges of dealing with crime-related textual data. All of these activities show that this project is a complex engineering activity and emphasize its technical, ethical and practical importance.

Chapter 6

Conclusion

In this chapter, the results and implications of the research are summarized, the limitations faced by the researchers are identified, and some future research opportunities are outlined. It outlines the most important lessons learned throughout the project and puts the proposed system in the context of the wider research.

6.1 Summary

The study is about developing a transformer-based ensemble architecture for the generation of abstractive summaries of Bangla crime news articles. In the previous chapters, the research problem, motivation and background were clearly articulated which reflected the increasing amount of digital Bangla crime reports, and the need for domain-specific tools for text summarization. The assessment of the existing literature illustrated that there was significant progress in the field of multilingual summarization but there was a big gap in the case of Bangla, especially in the area of crime reporting.

The project uses modern NLP techniques- particularly transformer-based models e.g. T5, mT5, and BanglaT5- to investigate an effective way to have domain-specific summarization. The methodology will involve preparing the dataset, preprocessing, experimenting and assessing the model with standard measurements.The work thus augments a systematic effort in filling an area of the Bengali NLP which has not been explored before and is the foundation of developing a useful and automated crime news summarizer.

6.2 Limitation

Regardless of the input of the work, there are a few constraints:

- Small Dataset Size: There is no public source of Bengali crime news dataset, and the datasets have to be collected manually, which will lead to a smaller dataset that might not be able to represent all linguistic variants.

- Computational Constraints: Transformer-based models are computationally intensive (high level of GPU resources); this limited the complexity of the models and their training.

- Abstractive Accuracy Difficulties: Bangla is morphologically rich and inconsistent in spelling rules that at times allow one to make mistakes in generated summaries.

- Lack of Human Evaluation: Because of time limitations, the evaluation is based mostly on automated measures instead of full expert or user-based

6.3 Future Work

Several improvements and extensions can be studied in the future:

- Growth of Dataset: Developing a more varied, larger, and balanced dataset of Bangla crime news, possibly by using crowdsourcing or media partners.

- Improved Model Architecture: Utilize advanced modeling techniques e.g. hybrid methods that integrate reinforcement learning, RAG methods, and prompt-based large language models to improve the accuracy of the summaries and depth of the context.

- Human Evaluation: Performing comprehensive user analysis or expert analysis to prove quality on summary measures over automated measures.

- Generalization as Mobile Application: Making the model an easily accessible tool that can be used by journalists, law enforcement analysts, or even the general population.

- Multimodal Crime News Summarization: Summaries can contain images, videos, or metadata to create more context-rich and detailed summaries. -humanize this part.

References

[1] O. Tas and F. Kiyani, “A Survey Automatic Text Summarization,” *Pressacademia*, vol. 5, no. 1, pp. 205–213, Jun. 2017, doi: https://doi.org/10.17261/pressacademia.2017.591.

[2] H. Zhang, P. S. Yu, and J. Zhang, “A Systematic Survey of Text Summarization: from Statistical Methods to Large Language Models,” *ACM Computing Surveys*, Apr. 2025, doi: https://doi.org/10.1145/3731445.

[3] Sumayya Afreen, S. S. Fatima, A. Begum, and A. Nuzha, “An Approach toward Abstractive Text Summarization for Urdu Language Using LLM (ATSUL),” pp. 319–337, Jan. 2024, doi: https://doi.org/10.1007/978-3-031-69336-6_14.

[4] M. Bani-Almarjeh and M.-B. Kurdy, “Arabic Abstractive Text Summarization Using RNN-based and transformer-based Architectures,” *Information Processing & Management*, vol. 60, no. 2, p. 103227, Mar. 2023, doi: https://doi.org/10.1016/j.ipm.2022.103227.

[5] A. Bhattacharjee, T. Hasan, W. U. Ahmad, and R. Shahriyar, “Findings of the Association for Computational Linguistics: EACL 2023, Pages 726–735 May 2-6, 2023 ©2023 Association for Computational LinguisticsBanglaNLG and BanglaT5: Benchmarks and Resources foEvaluating Low-Resource Natural Language Generation in Bangla,” *Findings of the Association for Computational Linguistics: EACL 2023*, pp. 726–735, May 2023.

[6] L. Xue *et al.*, “mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer,” *ACLWeb*, Jun. 01, 2021. https://aclanthology.org/2021.naacl-main.41/

[7] A. Vaswani *et al.*, “Attention Is All You Need,” *arXiv.org*, 2017 https://arxiv.org/abs/1706.03762

[8] L. Xue *et al.*, “mT5: A massively multilingual pre-trained text-to-text transformer,” Oct. 2020, doi: https://doi.org/10.48550/arxiv.2010.11934.

[9] S. Masri, Y. Raddad, F. Khandaqji, Ashqar, Huthaifa I, and M. Elhenawy, “Transformer Models in Education: Summarizing Science Textbooks with AraBART, MT5, AraT5, and mBART,” *arXiv (Cornell University)*, Jun. 2024, doi: https://doi.org/10.48550/arxiv.2406.07692.

[10] N. Radha, R. Swathika, K. R. Uthayan, and Mitul Krishna B, “AI-Driven Summarization of Academic Literature using Transformer Model,” vol. 4, pp. 359–364, Jun. 2024, doi: https://doi.org/10.1109/icici62254.2024.00065.

[11] N. Alipour and S. Aydın, “Abstractive summarization using multilingual text-to-text transfer transformer for the Turkish text,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 14, no. 2, p. 1587, Apr. 2025, doi: https://doi.org/10.11591/ijai.v14.i2.pp1587-1596.

[12]S. Abujar, A. K. M. Masum, M. Mohibullah, Ohidujjaman, and S. A. Hossain, "An Approach for Bengali Text Summarization using Word2Vector," *IEEE Xplore*, Jul. 01, 2019. https://ieeexplore.ieee.org/abstract/document/8944536/ (accessed Jan. 02, 2023).

[13]Nahida Akter Tanjila, Afrin Sultana Poushi, S. A. Farhan, Abu, M. A. Hossain, and Md. Hamjajul Ashmafee, "Bengali ChartSumm: A Benchmark Dataset and study on feasibility of Large Language Models on Bengali Chart to Text Summarization," *ACL Anthology*, pp. 35–45, 2025, Accessed: May 24, 2025. [Online]. Available: https://aclanthology.org/2025.chipsal-1.4/

[14]Anupam Singha and N. R. Rajalakshmi, "Bengali Text Summarization with Attention-Based Deep Learning," vol. 11, pp. 1–5, Aug. 2023, doi: https://doi.org/10.1109/asiancon58793.2023.10270772.

[15]G. M. Shahariar, T. Talukder, R. Alam, and T. Rouf, "Rank Your Summaries: Enhancing Bengali Text Summarization Via Ranking-Based Approach," *Lecture notes in networks and systems*, pp. 153–167, Jan. 2024, doi: https://doi.org/10.1007/978-981-99-8937-9_11.

[16]F. Morshed, M. A. Rahman, and S. Ahmed, "A Novel Word Pair-based Gaussian Sentence Similarity Algorithm For Bengali Extractive Text Summarization," *arXiv (Cornell University)*, Nov. 2024, doi: https://doi.org/10.48550/arxiv.2411.17181.

[17]A. Gupta, D. Chugh, Anjum, and R. Katarya, "Automated News Summarization Using Transformers," *arXiv (Cornell University)*, Jan. 2021, doi: https://doi.org/10.48550/arxiv.2108.01064.

[18]A. Chaves, C. Kesiku, and B. Garcia-Zapirain, "Automatic Text Summarization of Biomedical Text Data: A Systematic Review," *Information*, vol. 13, no. 8, p. 393, Aug. 2022, doi: https://doi.org/10.3390/info13080393.