INTERPRETABLE SENTIMENT ANALYSIS OF BANGLA E-COMMERCE COMMENT USING EXPLAINABLE AI TECHNIQUES

BY

Mst. Sumaiya Afrin Nahar
ID: 203-15-14564

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

Dr. S.M. Aminul Haque
Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Sharun Akter Khushbu
Lecturer (Sr. Scale)
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Interpretable Sentiment Analysis of Bangla E-commerce Comment Using Explainable AI Techniques", submitted by Mst. Sumaiya Afrin Nahar 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 13 July 2024.

BOARD OF EXAMINERS

Dr. Md. Taimur Ahad — **Board Chairman**
Associate Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Fizar Ahmed — **Internal Examiner 1**
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Rahmatul Kabir Rasel Sarker — **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Sadat Hasan — **External Examiner**
Data Scientist (Senior Principal Officer)
Risk Management Division
BRAC Bank

DECLARATION

I hereby declare that this project has been done by me under the supervision of **Dr. S.M. Aminul Haque, Professor and Associate Head,** Department of Computer Science and Engineering, Daffodil International University. I 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:

Dr. S.M. Aminul Haque
Professor and Associate Head
Department of CSE
Daffodil International University

Co-Supervised by:

Sharun Akter Khushbu
Lecturer (Sr. Scale)
Department of CSE
Daffodil International University

Submitted by:

Sumaiya Afrin

Mst. Sumaiya Afrin Nahar
ID: 203-15-14564
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty for His divine blessing making it possible for me to complete the final year project successfully.

I am grateful and wish my profound indebtedness to **Dr. S.M. Aminul Haque, Professor and Associate Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of Natural Language Processing (NLP) 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.

I would like to express my heartiest gratitude to the **Dr. Sheak Rashed Haider Noori, Professor and Head**, Department of CSE Daffodil International University, Dhaka, for his kind help in finishing my project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

I would like to thank my entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, I must acknowledge with due respect the constant support and patience of my parents

ABSTRACT

This research focuses on sentiment analysis of Bengali e-commerce comments using various machine learning algorithms. The primary objective is to develop and evaluate models that can accurately classify comments into positive or negative sentiments. I experimented with LR, DT, RF, MNB, KNN, SVM, and SGD. My dataset, collected through web scraping, underwent extensive preprocessing to ensure quality and relevance. The performance of each model was assessed using metrics such as accuracy, precision, recall, and F1-score. Among all the models, SVM exhibited the best performance, achieving a training accuracy of 89.6%, validation accuracy of 87.6%, and a model accuracy of 86.7%. This superior performance indicates SVM's robustness and effectiveness in handling high-dimensional data and non-linear decision boundaries in the context of sentiment analysis for Bengali text. In addition to model training and evaluation, I implemented a web application for practical deployment, designed using HTML, CSS, JavaScript, and the Flask framework. This application facilitates user-friendly interaction and real-time sentiment analysis, ensuring scalability and reliability. The integration of Explainable AI (XAI) techniques further enhances the interpretability of the model's predictions, providing insights into the factors influencing sentiment classification. Overall, my study demonstrates the potential of SVM in achieving high accuracy and reliability in sentiment analysis of Bengali e-commerce comments, paving the way for more advanced applications in natural language processing for underrepresented languages.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

NLP	Natural Language Processing
SA	Sentiment Analysis
ML	Machine Learning
DL	Deep Leering
XAI	Explainable Artificial Intelligence
LR	Logistic Regression
DT	Decision Tree Classification
RF	Random Forest
MNB	Multinomial Naïve Bayes
KNN	k-nearest neighbors
SVM	Support Vector Machine
SGD	Stochastic Gradient Descent
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
Bi-LSTM	Bidirectional Long Short-Term Memory
LIME	Local Interpretable Model-Agnostic Interpretation

CHAPTER 1
Introduction

1.1 Overview

Sentiment analysis (SA) or opinion mining, serves as an important tool for understanding public opinion and extracting insights from textual information obtained from various online platforms [1]. SA is an important technique for analyzing consumer feedback to improve business operations and customer happiness. The rapid growth of e-commerce in Bangladesh and the increasing adoption of online shopping platforms has led to an increase in information on user-generated content such as product reviews, ratings, and comments in the Bengali language to determine customer sentiment towards products, services, and the overall shopping experience [2]. The Bangla language's unique linguistic qualities and cultural complexities make sentiment analysis much more important. Day by day increasing number of Bangla-speaking consumers engaged in online purchasing, there is a need to create powerful sentiment analysis models that are suited to the nuances of the Bangla language [3]. By precisely assessing customer sentiment in Bangla e-commerce comments, businesses can acquire a better knowledge of customer preferences, identify areas for development, and modify marketing strategies to better fit the demands of Bangla-speaking customers [4]. SA can be approached through various methods and techniques; each proposal has its own advantages and applicability depending on the specific context and requirements of the analysis. A common method among these is ML based sentiment analysis where algorithms are trained to classify text into different sentiment categories like positive and negative through labeled datasets. Deep learning models such as Support Vector Machine (SVM) Logistic Regression, K-Nearest Neighbors (KNN) Naïve Bias, Decision Tree, Random Forest, Stochastic Gradient Descent (SGD) and Recurrent Neural Network (RNN) Convolutional Neural Network (CNN) or Transformer Learning Model are often used for this purpose. Another approach involves dictionary-based sensory analysis, where sensory scores are assigned to words or phrases based on predefined sensory lexicons or dictionaries [5]. Text data is analyzed by combining sensory scores to determine the overall feelings expressed in the text data. In this study, I followed a machine learning-based approach. To improve the interpretability and transparency of the sentiment analysis model, explainable artificial intelligence (XAI) methods have been used in this study. XAI techniques help users understand how

sentiment analysis models arrive at their predictions. It also provides transparency into the decision-making process of sentiment analysis models, revealing which features or words contribute most significantly to the predicted sentiment. It can uncover biases in sentiment analysis models by revealing which features or contexts influence the model's predictions [6]. In this research I design and implement a user-friendly website that accepts single or multiple user comments at a time as input and output produce the sentiment polarity (positive or negative) of each comment. Additionally, the system will offer detailed explanations highlighting the words or phrases responsible for the sentiment classification, thereby facilitating a deeper understanding of the sentiment conveyed in user comments of Bangla text. By meeting these goals, the project seeks to contribute to the development of sentiment analysis techniques in Bangla text, in addition to building transparency and interpretability in automated sentiment classification systems.

1.2 Background and Present State

The evolution of SA, has been driven by the exponential growth of digital data and the growing need for businesses to understand public opinion. SA involves analyzing the text to determine the feelings expressed, usually categorized as positive, negative, or neutral. This field has gained significant traction due to its wide range of applications in areas such as marketing, customer service, and social media monitoring.

Background: The origins of SA can be traced back to the early 2000s, with early research focusing on English text. Early methods were largely rule-based, relying on a predefined lexicon of positive and negative terms. However, this approach had limitations in terms of managing the complexity and variability of human language. With the advancement of NLP and ML, more sophisticated techniques have emerged. ML algorithms such as NB, SVM and DT to outperform rule-based systems by learning from large labeled datasets.

Present State: Despite these advances, SA remains challenging for low-resource languages such as Bengali. Bengali, spoken by more than 230 million people worldwide, has unique linguistic features, including complex morphology and syntax. This complexity requires specially trained preprocessing techniques and models on the Bengali text. However, the lack of large, labeled datasets in the Bengali language is a significant barrier. The integration of XAI techniques into SA models marks a significant advance. XAI aims to make the decision-making process of the AI model

transparent and understandable. Techniques such as LIME are increasingly being applied to sentiment classifiers. These methods provide insight into which features or terms most influence the model's predictions, thus increasing trust and enabling actionable insights for the business.

1.3 Problem Statement

The significant growth of e-commerce in Bangladesh has led to an overwhelming increase in user-generated content such as product reviews, ratings, and comments in the Bengali language. These comments provide valuable insight into customer opinions and preferences, which can significantly impact business strategy and customer satisfaction. However, the functional analysis of this textual data poses several challenges. First, the unique linguistic features of the Bengali language, including its rich morphology, complex syntax, and diverse vocabulary, complicate the process of SA. Traditional SA models, primarily developed for English and other resource-rich languages, often fail to capture these nuances, leading to optimal performance when applied to Bangla text. Furthermore, the lack of large, labeled datasets for Bengal exacerbates this problem, limiting the training and evaluation of robust SA models. Second, although advanced ML and DL models have shown promise in increasing SA accuracy, their inherent complexity and opacity make them difficult to interpret. For businesses that rely on these models to make decisions, understanding the reasoning behind a model's predictions is crucial. Without explainability, stakeholders may be reluctant to trust and act on the insights provided by these models.

1.4 Objectives

The expected outcome of this research project is a robust SA framework customized for Bangla text, which is designed to accurately classify users' comments received from e-commerce platforms such as Daraz [24] website. This framework will incorporate XAI methodologies, including Lime, to offer transparent insights into the sentiment dynamics present in Bangla user comments. By enhancing interpretability, these XAI techniques will enable businesses to gain deeper understanding and actionable insights from the sentiment analysis results. Additionally, a user-friendly website interface will be developed and implemented, allowing users to input single or multiple comments, with the output providing the sentiment polarity (positive or negative) of each comment.

This comprehensive framework will contribute to improved decision-making processes and enhanced customer satisfaction in the Bangla e-commerce domain.

1.5 Scope and Limitations

The scope of this research includes the development and evaluation of a SA framework developed for Bengali e-commerce comments. The primary focus areas are creating a new data set. Apply various ML and DL algorithms will be employed to classify sentiment as positive or negative accurately. To increase the explainability of SA models, XAI techniques such as LIME will be integrated. It will help to understand how models make predictions and which features contribute most significantly to these predictions. A user-friendly web application will be created using HTML, CSS, JavaScript and Flask Framework. This app will allow users to input single or multiple comments and get the sentiment polarity output along with the explanation. Despite the wide scope, this research has encountered several limitations like the lack of large, labeled Bengali datasets is a significant challenge. The unique linguistic features of Bengali, including its rich morphology and complex syntax, increase the difficulty in developing effective SA models. There is often a trade-off between model explainability and performance. High precision models, such as deep learning architectures, may be less interpretable than simpler models. Balancing these aspects to meet the needs of the business is a significant challenge. Creating a user-friendly web application that can handle multiple comments at the same time while providing real-time analysis and interpretation. The research is conducted within certain resource constraints including computational power and time.

1.6 Report Organization

The proposed report is structured with a broad format to guide readers through the research process, findings, and implications. Each chapter serves a specific purpose, which contributes to the overall coherence and depth of the document.

In the first chapter or the introductory chapter, the context of the study is set by highlighting the importance of emotion analysis in Bangla text and emphasizing the importance of Explainable Artificial Intelligence (XAI) in emotion analysis. It describes the motivation behind the study, explains the research questions, explains the objectives and presents an overview of the study structure.

Chapter 2 (Literature Review) provides a comprehensive examination of the relevant

literature and previous research efforts. This section explores in depth the theoretical foundations, methods, and technological advances associated with emotion analysis, machine learning algorithms, deep learning algorithms, and interpretive artificial intelligence. (XAI). By providing a comprehensive overview of existing knowledge, it establishes relevant frameworks for current studies and also identifies gaps in the current understanding of the subject.

The third chapter (Methodology) explains the strategy employed to conduct the study by taking a quick look at the research design, data collection techniques, model architecture, user friendly website design, and the rationale behind selecting specific methodologies.

In the fourth chapter (Implementation) explain about implementation overview, how to train the model and how model evaluation.

Chapter 5 (Result and Analysis) demonstrates experimental results from applying deep learning methods to machine learning, emotion prediction. It provides comprehensive analysis of the performance of various algorithmic methods, which is complemented by visual representations and statistical metrics to prove the results and show the performance of the website

The Impact on Society, Environment and Sustainability in chapter 6 explores the impact of research findings within the field of e-commerce and consumer behavior. It explains how advanced sentiment analysis techniques for e-commerce comments can positively impact business strategy, customer satisfaction, and market dynamics. Emphasizing the importance of using advanced artificial intelligence techniques to improve the Bengali e-commerce sector, the discussion expanded to potential advances in technology and their social impact.

Chapter 7: Conclusion and Future Work This final chapter serves as a synthesis of the entire report. It summarizes the main results, reiterates the conclusions drawn from the research, and provides recommendations for practical application and further study. Implications for future research are discussed, providing a roadmap for researchers interested in expanding current research.

This structured framework provides a coherent progression of information, which takes the reader from the introduction through the research campaign to the larger implications and recommendations. It provides a thorough understanding of the research methodology and its potential implications in both academic discourse and

practical application in sentiment analysis for Bangla e-commerce comments using machine learning and deep learning algorithms.

1.7 Summary

The report begins by contextualizing sentiment analysis or opinion mining within the realm of understanding public opinion through textual data from online platforms Section 1.1. It underscores SA's crucial role in enhancing business strategies and customer satisfaction, particularly in the context of Bangladesh's burgeoning e-commerce sector. With the proliferation of Bangla-speaking consumers engaging in online shopping, there's a pressing need for robust SA models tailored to the nuances of the Bengali language Section 1.1, 1.2. The study identifies key challenges in adapting traditional SA methods to Bengali, including the language's intricate morphology and syntax, compounded by the scarcity of large, labeled datasets Section 1.3, 1.5. To address these challenges, the research focuses on employing machine learning techniques, such as Support Vector Machine, Logistic Regression, and Deep Learning models like Convolutional Neural Network and Recurrent Neural Network Section 1.1. A significant contribution of the study lies in integrating XAI techniques, particularly LIME, to enhance the transparency and interpretability of SA models Section 1.1, 1.4. This approach not only improves trust in AI-driven decisions but also identifies biases and key features influencing sentiment predictions in Bengali text data Section 1.4. The practical implementation of the research culminates in the development of a user-friendly web application. This application facilitates real-time sentiment analysis of single or multiple Bengali comments, providing clear explanations for each sentiment classification Section 1.1, 1.4, 1.5.

CHAPTER 2

Literature Review

2.1 Overview

The primary topics of this research include the essential foundations needed for effective sentiment analysis on Bengali e-commerce text, integrating XAI techniques. Sentiment analysis, or opinion mining, involves extracting and categorizing emotions and opinions expressed in text, focusing here on user comments from Bangla e-commerce platforms. This task is challenging due to the complex morphology and syntax of the Bangla language, which requires tailored preprocessing and analytical methods. Advanced machine learning and deep learning algorithms will be employed to accurately predict emotions. To increase the transparency and interpretability of these models, XAI techniques such as LIME or SHAPE will be used. These techniques will help to explain how the models arrive at their predictions, making the results more understandable and actionable for businesses. This research aims to create a comprehensive framework that not only achieves high accuracy in sentiment classification but also provides clear insight into the decision-making process of the model, thereby facilitating better business strategy and customer satisfaction in the Bengali e-commerce sector.

2.2 Related Works

Sentiment analysis (SA) has been studied extensively across languages and fields using a variety of methods, from traditional machine learning techniques to advanced deep learning models. In this section, I have discussed the significant works that contributed to the development of sensory analysis framework, especially focusing on Bangla text and other low-resource languages, as well as the application of XAI in sensory analysis. Bhowmik et al. research introduced a hybrid approach combining dictionary-based SA and DL models to increase the accuracy of sentiment analysis in Bangla text. Their approach extends existing lexicon dictionaries, integrates deep learning algorithms such as CNN or RNN. They reported accuracy values of 78.52%, 80.82% and 84.18% respectively [7]. Tuhin et al. aimed to improve sentiment classification amidst noisy data with missing values, demonstrating that traditional algorithms like Naive Bayes are less effective. Their method focused on classifying emotions at the sentence level, achieving over 90% accuracy with a temporal approach [8]. Cao et al. addressed

sentiment analysis in the context of agricultural product reviews using an enhanced BERT model, achieving an F1 score of 89.86%. This study illustrated the importance of domain-specific adaptations in SA models [9]. The need for interpretability in AI models has led to the development of XAI techniques. Abdelwahab et al. and Atabuzzaman et al. explored XAI for Arabic sentiment analysis, incorporating models like BLSTM and CNN-BLSTM with noise layers to explain specific predictions. They emphasized the importance of understanding model decisions in low-resource languages [10]and [11]. Rao and Swathi and Barik et al. proposed deep learning frameworks for analyzing sentiment in e-commerce reviews. These studies utilized models like CNN-WDE and LSTM-DGWO, demonstrating improved accuracy and efficiency in processing large datasets of customer feedback [12] and [13]. Haque et al. tackled the issue of inadequate performance in Bengali SA due to language idiosyncrasies and resource constraints. They proposed a CNN and LSTM-based classification for multi-class SA in Bengali social media comments, achieving an accuracy of 85.8% and an F1 score of 0.86. Their model was particularly effective in classifying comments into sexual, religious, political, and acceptable categories [14]. Shanto et al. focused on creating a balanced dataset for Bengali sentiment analysis in e-commerce, contributing valuable resources for future research in this area [15].

2.3 Comparison between existing works

In SA, particularly for low-resource languages like Bengali, various methodologies have been explored to improve accuracy and efficiency. This comparative analysis explores into the multifaceted landscape of SA in Bengali text, assessing the efficacy of diverse methodologies employed across several studies. From rule-based systems like the Bengali Text Sentiment Score to DL models and traditional classifiers like Random Forests and SVMs, the research scrutinizes the strengths and limitations of each approach. While rule-based systems offer simplicity and interpretability, hybrid DL models demonstrate significant accuracy improvements, emphasizing the importance of advanced techniques in Bengali SA. Additionally, the integration of XAI techniques enhances model interpretability and trust, crucial for real-world applications.

TABLE 2.3.1: COMPARATIVE ANALYSIS WITH PREVIOUS WORK.

Study	Used Algorithm	Best Accuracy with Algorithm
Abdelwahab, Youmna et al. [10]	LSTM Classifier and LIME Algorithm	LSTM = 79%
Atabuzzaman, Md et al [11]	BiLSTM, CNN-BiLSTM Classifiers and LIME	CNN-BiLSTM =88%
Haque, Rezaul et al. [14]	SVA, SGD, RF, LR, MNB, DT, BiGRU, LSTM, CLSTM	CLSTM = 85%
Hossain, et al. [16]	SVM, LR, CNN, BERT	BERT = 68%
Ashik, Md Akther-Uz-Zaman et al. [17]	LSTM	LSTM = 79%
Palash, Partha Protim Das et al. [18]	VAE	VAE = 53%
Tripto, Nafis Irtiza et al. [19]	CNN and LSTM	LSTM = 66%
k. Sarkar, [20]	LSTM	LSTM = 55%
Islam; Md Ashiqul et al. [21]	NB with unigram, NB with bigram	NB with bigram = 77%
Jain, Rachna et al. [22]	VADER and LIME	VADER = 69%
Patel, Aksh et al. [23]	DT, RF, BERT	BERT = 83%

TABLE 2.3.1 shows that comparative analysis of studies based on their best accuracy achieved with different algorithms reveals a spectrum of performance across different text classification tasks. LSTM models were featured prominently with accuracy ranging from 55% to 79% in various studies. Study by Abdelwahab et al. and Ashik, etc. LSTM has highlighted its consistent performance, achieving 79% accuracy, while Atabuzzaman et al. CNN-BiLSTM has achieved a maximum accuracy of 88% with the models, which indicates their effectiveness in capturing complex data patterns. Additionally, the CLSTM-like models used by Houk et al. demonstrated 85% competitive accuracy, which underscores the effectiveness of the convolutional LSTM architecture. Conversely, the BERT implementation, as by Hossain et al. and Patel et al., achieved 68% and 83% accuracy, respectively, demonstrating the power of BERT in understanding natural language but also its variability based on task specificity.

2.4 Open Issues

The research faces several significant challenges inherent to sentiment analysis in Bangla language e-commerce data. Firstly, the lack of labeled datasets presents a formidable obstacle, necessitating extensive manual annotation efforts to acquire high-quality training data for machine learning models. Secondly, Bangla's linguistic

complexity, characterized by intricate syntax, morphology, and semantics, poses challenges in capturing linguistic nuances and cultural references essential for accurate sentiment analysis. Thirdly, user comments on e-commerce platforms often feature domain-specific vocabulary, slang, abbreviations, and emojis, requiring robust algorithms capable of handling diverse language variations and informal expressions while maintaining sentiment accuracy. Fourthly, the variability in comment length, tone, and content adds complexity to sentiment analysis, demanding techniques for processing unstructured text data and extracting meaningful features while preserving context and sentiment polarity. Addressing these challenges is crucial for advancing sentiment analysis in Bangla language e-commerce data and unlocking its potential for informing decision-making processes effectively.

2.5 Summary

In exploring sentiment analysis for Bangla language e-commerce data, this research confronts significant challenges and draws upon a breadth of methodologies and studies. The review of related works 2.2 underscores the evolution from traditional machine learning to advanced deep learning models like CNN-BiLSTM, emphasizing their effectiveness in capturing complex data patterns in Bengali text. Comparative analysis 2.3 reveals varying accuracies across studies, highlighting LSTM and CNN-BiLSTM as standout performers, albeit with considerations of model interpretability. Meanwhile, the discussion on open issues 2.4 illuminates' critical hurdles such as data scarcity, linguistic complexity, and the need for robust algorithms capable of handling informal language and diverse content. These insights underscore the necessity of tailored approaches and XAI techniques 2.1 to enhance accuracy, interpretability, and scalability in sentiment analysis for Bengali e-commerce, aiming to inform strategic decisions and improve customer engagement effectively.

CHAPTER 3
Methodology

3.1 Overview

The study "Interpretive Sentiment Analysis of Bengali E-Commerce Comments Using Interpretive AI Techniques" focuses on the application of advanced artificial intelligence methods to enhance the understanding and interpretation of customer feedback in Bangla text. This research specifically targets the domain of e-commerce, where analyzing user comments accurately is crucial for improving customer experience and business strategies. As Bangla is a morphologically rich and syntactically complex language, presents unique challenges for sentiment analysis. This research aims to address challenges using XAI techniques, which provide transparent insights into the decision-making process of AI models. Interpretive AI techniques such as LIME are employed to make sentiment analysis models more explainable and convincing. These methods help in understanding how different parts of a comment contribute to the overall sentiment score, making the results more actionable for stakeholders. This study involves a comprehensive analysis of various AI techniques, comparing their effectiveness and interpretability in the context of Bangla e-commerce comments. By doing so, it seeks to identify the best practices and models that balance accuracy with interpretability, ensuring that the AI-driven insights are both reliable and understandable. Basically, this research is located at the intersection of artificial intelligence, linguistics and e-commerce, which aims to advance the analysis of sentiment for Bengali language comments.

3.2 Proposed Methodology

The process begins with data collection from Daraz [24] website through web scraping techniques, followed by manual annotation of the dataset. Data preprocessing involves removing null values and duplicate entries to enhance data quality and mitigate bias. Non-character words, non-Bengali characters, and stop-words are eliminated to reduce noise, while tokenization breaks down the text into individual units. Stemming is applied to reduce the original size of the word, thereby reducing the size of the vocabulary. In addition, short and long sentences are filtered based on a predefined threshold to standardize the length of the text. In addition to deep learning models such

as CNN, RNN, GRU, various machine learning algorithms such as LR, DT, RF, Naive Bayes, KNN, SVM, SGD are applied and trained on pre-processed datasets.

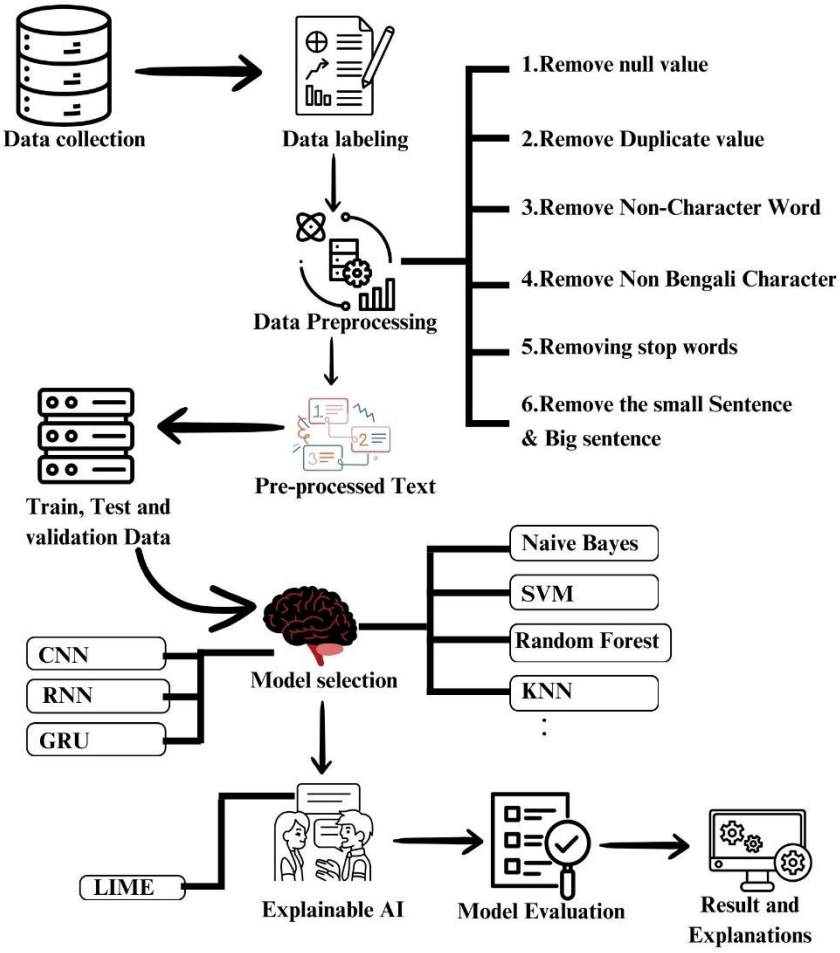


Figure 3.2.1: Proposed Methodology for Integrating XAI and Machine Learning Algorithms.

The model evaluation is conducted using appropriate metrics such as accuracy, precision, recall and F1-score with the help of XAI algorithm to interpret the predictions and understand the influencing factors. The best-performing model is selected based on accuracy and other relevant metrics, taking into account computational efficiency, interpretability, and scalability. Subsequently, a front-end website is designed using HTML, CSS, JavaScript, other framework like bootstrap, Font Awesome etc. and backend developed using the Flask framework for intuitive and user-friendly interaction. Finally, the trained model and web application are deployed on a local server, ensuring scalability and reliability of the deployed system.

Data Collection Procedure:

A. Datasets

Creating a dataset was one of the most challenging tasks in this study. Data was collected from Daraz [24] website using a web scraping technique. The workflow

involved navigating to the website and selecting products whose user comments I aimed to collect. After identifying the desired products, I copied their URLs and seamlessly integrated them into my web scraping code. To facilitate the data extraction process, I leveraged a suite of Python libraries. Selenium was used for web automation, allowing for dynamic interaction with web pages to handle JavaScript-heavy content. BeautifulSoup was employed for parsing the HTML content, enabling the extraction of specific elements from the web pages. By running the web scraping script, I retrieved the first 100 comments associated with each selected product. After that I manually annotated the data set 0 for Negative and 1 for Positive comment then verified this dataset by a Bangla language expert.

Text	Sentiment
সবই ভালো ছিল কিন্তু এর স্টান টা ভাঙ্গা ছিল😞😞😞	0.0
খুব সুন্দর একটা প্রোডাক্ট চাইলে আপনারা ও নিতে ...	1.0
চমৎকার 👍	1.0
★ দারাজের জঘন্য প্রোডাক্ট লিস্টে আরো ৩টা প্রোড...	0.0
ঘড়ি টা ভালোই সবাই নিতে পারেন	1.0
এখন পর্যন্ত দারাজ থেকে যত গুলি পণ্য নিয়েছি বল...	1.0
হাতে পেতে একটু লেট হলেও সেলার কে অনেক অনেক ধন্...	1.0
যাদের বাড়িতে চার্জিং সকেট এর নিচে ফোন রাখার ক...	1.0
ছবির সাথে প্রডাক্টের কোন মিল নাই। আমি দারাজের...	0.0
চমৎকার পণ্য. আমি আরো পণ্য কিনব	1.0

Figure 3.2.2: Sample Dataset Overview

Figure 3.2.2 provides an overview of the sample dataset, showcasing customer reviews written in Bengali alongside their respective sentiment ratings. The dataset encompasses diverse feedback on products and services, reflecting the customers' experiences and opinions. Each review's sentiment is quantified with a numeric value, where 0.0 signifies negative sentiment and 1.0 denotes positive sentiment. This dataset serves as a valuable resource for sentiment analysis work, providing insight into consumer perceptions and preferences across different product categories. Analyzing this dataset can aid in understanding customer sentiments, identifying trends, and improving products or services to better meet customer expectations.

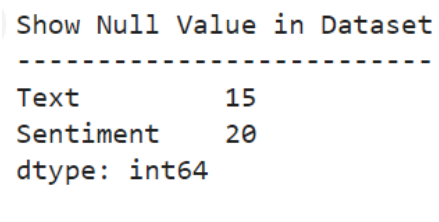
B. Process of Experiments

The experimental process begins with data preprocessing, which includes cleaning, remove duplicate value and more. The dataset is then balanced using techniques like SMOTE. Next, feature extraction methods such as TF-IDF vectorization and Word Embedding (Word2VEC) convert the text into numeric features. The dataset is divided

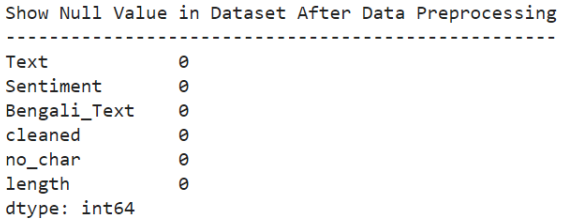
into training, testing and validation sets, followed by model training using algorithms from neural networks called LR, DT, RF, Naive Bayes, KNN, SVM, SGD or CNN, LSTM and GRU. The model is evaluated using metrics such as accuracy and F1-score. The best-performing model is selected and placed on the website.

Data Preprocessing:

The data pre-processing stage involves cleaning the collected comments to remove noise and irrelevant information. This included tasks such as removing null values, special characters, and stop words. Finally, very large and small sentences have been removed.



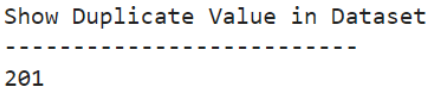
(a) Null Value Before Data Preprocessing



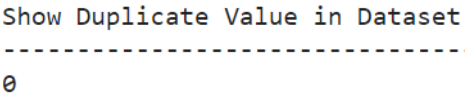
(b) Null Value After Data Preprocessing

Figure 3.2.3: Null Value Before and after Data Preprocessing.

In Figure 3.2.3 (a) show that there is total 15 null value in my Text feature and 20 Null value in my Sentiment Feature. On the other hand, Figure 3.2.3 (b) show that there is no null value after data preprocessing.



(a) Duplicate Value Before Data Preprocessing



Duplicate Value After Data Preprocessing.

Figure 3.2.4: Duplicate Value Before and after Data Preprocessing.

Figure 3.2.4 show that there are 201 duplicate values before data preprocessing and there is no duplicate value after data preprocessing.

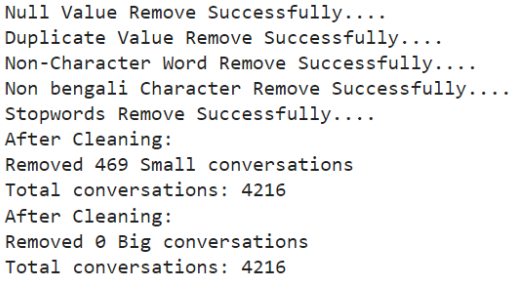


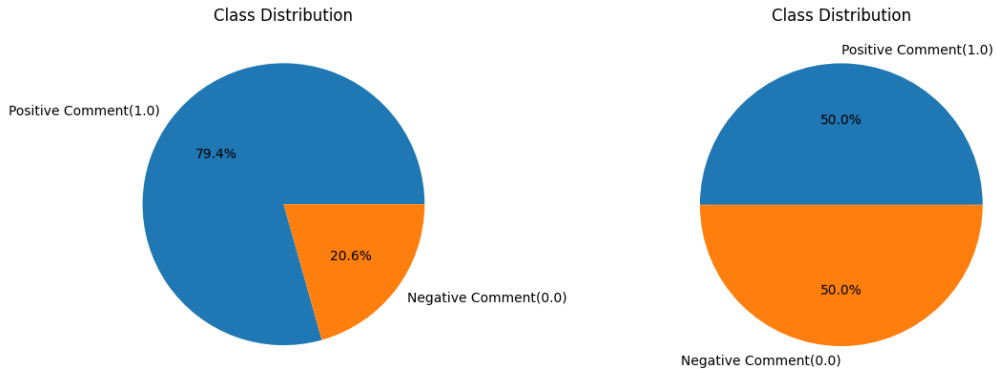
Figure 3.2.5: All Data Preprocessing Steps.

Figure 3.2.5 shows that all steps during data preprocessing. If any text length is less

than two, I remove those sentences on the other hand if any sentence length is greater than two hundred, I also remove this sentence.

Dataset Balanced:

To ensure that the dataset was balanced, the collected comments were analyzed for sentiment distribution. As an imbalance in my dataset was detected, I applied SMOTE techniques to the majority class or oversampling to the minority class. This step was crucial to avoid biased model training and to increase the model's ability to accurately classify feelings into different categories.



(a) Before Data Balance. (b) After Data Balance.

Figure 3.2.6: Class Distribution of the Datasets before and after Data Balance.

Figure 3.2.6 (a) presents a pie chart showcasing the class distribution of the dataset. The chart is divided into two sections: one representing negative sentiment reviews (0.0) and the other representing positive sentiment reviews (1.0). The size of each section corresponds to the proportion of reviews belonging to that sentiment class in the dataset. The pie chart visually indicates the imbalance between negative and positive sentiment reviews, with one section noticeably larger than the other, suggesting a prevalence of negative sentiments in the dataset. Figure 3.2.3 (b) shows a pie chart that represents the class distribution of the dataset after applying the synthetic minority over-sampling technique. (SMOTE). The chart shows two equal categories, one for negative sentiment (0.0) and one for positive sentiment (1.0) indicating that the dataset is now balanced. Smoot has successfully increased the number of positive emotional reviews in line with the number of negative reviews. This balanced distribution is essential for developing effective sentiment analysis models, as it prevents bias towards the previously dominant class and enhances the model's ability to accurately predict sentiment across both categories.

Feature Extraction:

Feature extraction involves converting pre-processed text data into numerical representations suitable for machine learning models. Techniques such as term frequency-inverse document frequency (TF-IDF) and word embedding Word2VEC were employed. In addition, linguistic features specific to the Bengali language, such as linguistic expressions and cultural references, were considered to enhance feature richness.

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (i)$$

$$TF(t,d) = \frac{\text{Number of times term } t \text{ appears in document } d}{\text{Total number of terms in document } d} \ldots\ldots\ldots\ldots\ldots\ldots (ii)$$

$$IDF(t) = \log\left(\frac{\text{Total number of documents}}{\text{Number of documents containing term } t}\right) \ldots\ldots (iii)$$

Dataset Split:

The dataset was divided into training, validation, and testing sets. A common split ratio of 70% for training, 15% for validation, and 15% for testing was used. This section ensures that the model can be effectively trained and evaluated on invisible data to assess its generalizability.

Model Training:

Various machine learning and deep learning models were trained on the processed datasets. For traditional machine learning, algorithms such as LR, DT, RF, Naive Bayes, SVC and SGD were used. For deep learning, models such as CNN, LSTM and GRU were implemented using TensorFlow and Keras libraries.

Model Evaluate:

The trained models were evaluated using standard metrics such as accuracy, precision, recall, and F1-score. In addition, the Confusion Matrix was developed to provide a detailed understanding of model performance across different sentiment classes. Explainable AI (XAI) techniques, such as LIME and SHAPE, were used to interpret model predictions and ensure transparency in the sentiment analysis process.

Model Selection:

Based on the evaluation metrics, the best-performing model was selected. This model demonstrated the highest accuracy and robustness in classifying Bangla e-commerce comments. The selected model not only performed well in terms of traditional metrics but also provided meaningful and interpretable explanations for its predictions for the integration of XAI techniques.

Model Deployment:

The final step is to deploy the selected model in a web application. Using Visual Studio code, the model was integrated into a Flask web framework for the backend, while HTML, CSS, and JavaScript were used for the frontend design. This setup allows users to input Bangla comments and receive sentiment analysis results along with an explanation of the model's predictions, thus making the tool accessible and user-friendly.

3.3 Hardware and Software Requirement

The implementation of this study involves certain requirements to ensure the successful implementation of the study with a machine learning based sentiment detection and interpretation. The requirements of this implementation include both hardware and software components as well as considerations regarding data collection and model evaluation.

Hardware Requirements:

- **High-performance computing resources:** Access to computational infrastructure with sufficient processing power and memory to handle the complexity of deep learning and natural language processing tasks.
- **Graphics Processing Unit (GPU):** Utilization of GPUs for accelerating the training of deep learning models, especially for large-scale datasets and complex neural network architectures.

Software Requirements:

- **Machine learning and deep learning frameworks:** Deploying deep learning frameworks such as TensorFlow for building and training neural network models and using the Scikit-Learn library for sentiment analysis**.**
- **Natural language processing libraries:** Integration of NLP libraries like NLTK to preprocess and analyze Bangla text data, including tokenization, and stemming.
- **Explainable AI tools:** Adoption of XAI techniques and libraries such as LIME to interpret and explain the predictions of sentiment analysis models.
- **Web development frameworks:** Implementation of web-based interfaces for visualizing and interacting with the sentiment analysis results, using frameworks like Flask is used for backend development and HTML, CSS, and JavaScript for frontend design.

The implementation requirements for this research encompass hardware, software, data collection, model training and evaluation. Hardware requirements include access to high-performance computing resources and GPUs for efficient model training. Software requirements involve the deployment of deep learning frameworks, NLP libraries, explainable AI tools, and web development frameworks.

3.4 Project Management and Financial Analysis

The project management and financial aspect of this research is very important to ensure the successful implementation of the study within the allocated resources and time frame. Effective project management involves several key elements, including schedule planning, resource allocation, and risk management. Figure 3.4.1 show a detailed project timeline will be created outlining the different stages of the research process, from data collection and pre-processing to model development, evaluation, and implementation. This timeline will serve as a roadmap, guiding the research team in meeting deadlines and milestones.

Figure 3.4.1: Gantt chart of the project timeline.

Figure 3.4.1 show that this project comprises several interdependent subsystems, including Data Collection, Data Preprocessing and Visualization, Model Selection and

Implementation, Model Evaluation, and Website Creation. The project timeline spans from mid-October to the end of November, during which the research domain and title are selected. From mid-November to the end of January, reference collection takes place concurrently with paper reviewing, followed by dataset collection from mid-January to the end of April. This is followed by data pre-processing and statistical analysis, which requires about 7 weeks. Subsequently, various machine learning and deep learning models are applied, which takes about 7 to 8 weeks. Then I evaluate the model performance. Based on accuracy I implement the model into a user-friendly website.

Financial planning is essential for budgeting and allocating funds to cover expenses related to data collection, software licenses, hardware infrastructure, and personnel costs. A comprehensive budget will be created, accounting for both fixed and variable expenses throughout the duration of the project. Efforts will be made to optimize costs and maximize the value without compromising on research quality.

TABLE 3.4.1: APPROXIMATELY COST CALCULATION.

SN	**Components**	**Estimated Cost (BDT)**
01	Data collection	5000 - 8000
02	Data annotation	8000 - 10000
03	Domain name registration and hosting	5000 - 7000
04	Website Design	8000 - 12000
05	Custom built website	10000 – 15000
06	Database Integration	7000 - 10000
07	User Testing	5000 - 8000
08	Others	15000-20000
09	Hardware	70000 – 80000
10	Learning new technology	5000-7000
Total Estimated Cost		138000 - 177000

The approximate cost calculation for the research encompasses various components, including data collection, annotation, website design, hosting, hardware, and learning

new technology. The total estimated cost ranges from BDT 138,000 to 177,000, with the majority allocated to hardware expenses. Other significant costs include website development and database integration. Additionally, expenses for data annotation, and learning new technologies contribute to the overall budget. Overall, the estimated cost provides an overview of the financial requirements for conducting the research.

3.5 Summary

This study focuses on enhancing understanding and interpretation of customer feedback on Bangla text within the e-commerce domain. This study addresses the complexities of sentiment analysis in Bengali, using advanced AI methods such as LIME to provide transparent insights into model predictions. Key elements include data collection through web scraping from Daraz [24], manual annotation, rigorous data preprocessing to handle linguistic nuances, and creating balanced datasets using SMOTE. The study used a variety of machine learning and deep learning models, evaluated based on accuracy metrics and complemented by XAI for interpretability. A user-friendly web interface has been developed for real-time sentiment analysis and model interpretation using Flask, HTML, CSS, and JavaScript. Project management includes detailed scheduling and resource allocation, ensuring timely completion of milestones. The financial plan accounts for costs across hardware, software, and operational costs, with the goal of optimizing budget allocations while maintaining the quality of the research. Overall, the aim of this study is to advance the capabilities of sentiment analysis in Bengali e-commerce, providing valuable insights for improving business strategy and customer satisfaction.

CHAPTER 4
Implementation

4.1 Overview

The experiments were conducted in the Google CoLab environment. Python was employed as the primary programming language for the study, taking advantage of its extensive library and framework suitable for deep learning and data analysis tasks. The study was divided into two parts. In the first part, which involved data analysis, model selection, and the implementation of XAI techniques, TensorFlow was used for its robust support for complex neural network tasks. Google CoLab provided the necessary computational infrastructure, offering cloud-based resources including a Tesla GPU and up to 16GB of RAM. This setup ensured efficient handling of deep learning tasks, particularly the training and inference of sentiment analysis models. For deep learning models like CNN, LSTM, and GRU, the Keras library within TensorFlow was utilized. In addition, Scikit-learn libraries were employed for machine learning algorithms such as LR, DT, RF, Naive Bayes, SVC, and SGD. In the second part of the study, which involved placing the best-performing model on a website, development was conducted in a Visual Studio Code environment. For the front-end design of the website, HTML, CSS, and JavaScript were used to ensure a user-friendly interface. For the back-end design and model deployment, the Flask web framework was employed, facilitating the integration of the trained sentiment analysis model into a web application. Overall, the instrumentation of this study involved a combination of Python, TensorFlow, Keras, scikit-learn, and the Google CoLab environment for the development and training of the models, and Visual Studio Code, Flask, HTML, CSS, and JavaScript for the deployment of the model onto a website. This comprehensive approach ensured efficient model training and practical application, enhancing the accessibility and usability of the sentiment analysis system.

4.2 Train Model

Statistical Analysis:

In this study, statistical analysis plays an important role in verifying the effectiveness of various sentiment analysis models and understanding the underlying patterns within the dataset. Following the successful collection of data from Daraz [24], a comprehensive analysis was conducted using Python alongside libraries such as

matplotlib and wordcloud. This analysis encompassed several key facets, including the average number of characters versus sentiment, average text length categorized by sentiment, class distribution, and identification of the top 100 words, as well as the most frequently used negative and positive words. These examinations provided invaluable insights into the dataset's characteristics and potential trends, forming the foundation for robust model evaluation and enhancement.

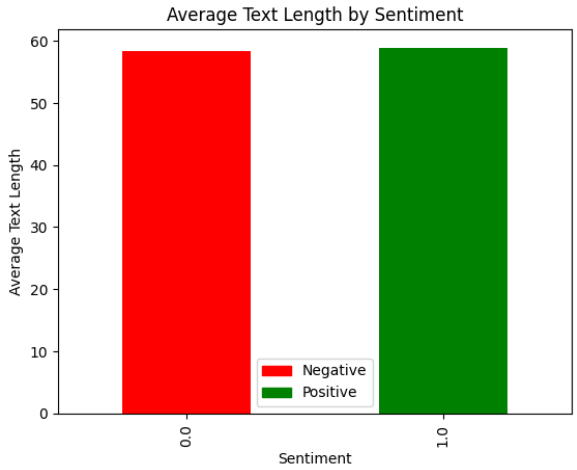


Figure 4.2.1: Average Characters Length vs Sentiment of my Dataset.

Figure 4.2.1 presents a bar plot illustrating the relationship between the average number of characters in comments and their corresponding sentiment scores within the dataset. The y-axis denotes the average characters per comment, while the x-axis represents the sentiment, categorized as positive, or negative. This suggests that the sentiment expressed in comments is not inherently influenced by the length of the text.

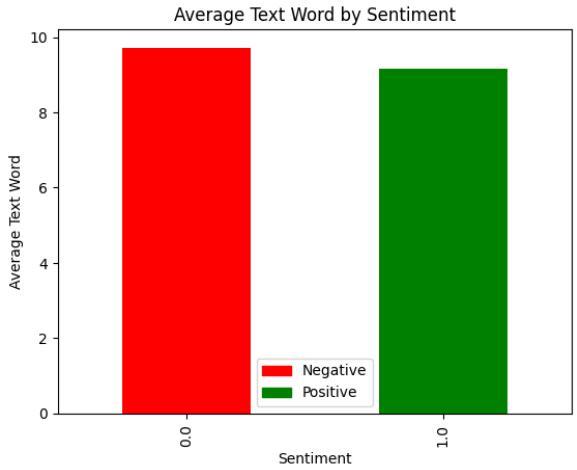


Figure 4.2.2: Average Word Length by Sentiment of my Dataset.

Figure 4.2.2 presents a bar chart depicting the average word length of text in comments categorized by sentiment within the dataset. The x-axis represents different sentiment

categories, including positive, and negative, while the y-axis indicates the average text length. Each sentiment category is represented by a distinct colored bar, allowing for easy comparison of text lengths across sentiments.

Figure 4.2.3: Top Used Word in the Dataset.

Figure 4.2.3 displays the most frequently used words within the dataset. more frequently in the dataset are displayed in larger fonts, while less common words are smaller. This visualization offers a quick and intuitive overview of the most prevalent terms used in the dataset.

Figure 4.2.4: Top Used Positive Word in the Dataset.

Figure 4.2.4 presents a visual representation of the most frequently occurring positive words extracted from the dataset. This visualization offers valuable insights into the

prevailing positive sentiments among users, shedding light on the aspects of products or services that are well-received and appreciated by customers.

Figure 4.2.5: Top Used Negative Word in the Dataset.

Figure 4.2.5 provides a visual depiction of the most frequently occurring negative words extracted from the dataset collected from the Daraz website. Through a word frequency analysis, this figure captures the prevalent negative sentiments expressed within user comments, showcasing words associated with dissatisfaction, criticism, or unfavorable experiences. These negative words are represented in different font sizes, large fonts indicate a high frequency of occurrence.

Logistic Regression:

Logistic Regression (LR) is a foundational machine learning algorithm that is widely employed for binary classification tasks, including sentiment analysis of textual information in the context of your research. LR is governed by modeling the probability that a given input belongs to a particular class, usually represented as a binary result such as positive or negative sentiment. Mathematically, LR applies the logistic function (sigmoid) to a linear combination of input characteristics, which is expressed as:

$$P(y = 1|\boldsymbol{x}; \boldsymbol{w}) = \frac{1}{1 + e^{-(\boldsymbol{w}^{T}\boldsymbol{x}+b)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (iv)$$

The LR model is a cost function, often trained by minimizing logistic losses or cross-entropy losses, that punishes deviations between predicted probabilities and true labels in training data. Once trained, the LR model can make predictions by setting predictable probability thresholds, assigning positive class instances above a certain threshold and negative class instances below. Overall, LR provides a transparent and interpretable

approach to sentiment analysis, making it well-suited for your research objective of interpreting sentiment in Bangla e-commerce comments using Explainable AI techniques.

Naïve Bayes:

Naive Bayes is a probabilistic machine learning algorithm based on Bayes' theorem, which assumes that attributes are conditionally independent due to class labels. Despite its simplicity, Naive Bayes often performs well in text classification tasks, which makes it suitable for sentiment analysis. Mathematically, the Naive Bayes classification predicts the probability of a class labeling y given a set of properties using Bayes' theorem:

$$P(y|x_1, x_2, \dots, x_n) = \frac{P(y)\cdot P(x_1, x_2, \dots, x_n|y)}{P(x_1, x_2, \dots, x_n)} \dots\dots\dots\dots\dots\dots\dots\dots (v)$$

Naive Bayes simply assumes that all properties are given the conditionally independent class label y. Despite this simplification, Naive Bayes can be remarkably useful in practice, especially for text classification tasks, where the presence or absence of certain words can provide strong evidence for or against certain classes.

Support Vector Machines:

In the field of sentiment analysis for Bengali e-commerce comments, Support Vector Machine (SVM) stands out as a powerful classification algorithm that is known for its effectiveness in managing the boundaries of high-dimensional data and nonlinear decisions. SVM works by mapping the input data points to a high-dimensional feature space and finding the optimal hyperplane that separates the data points into different classes. This hyperplane is chosen to maximize the margin, which is the distance between the hyperplane and the nearest data point of each class, known as the support vector. Mathematically, the decision boundary of an SVM can be represented as:

$$f(x) = w^T x + b \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (vi)$$

In the case of sentiment analysis, SVM aims to classify Bangla e-commerce comments into positive or negative sentiments based on the extracted features. The algorithm learns to find the optimal hyperplane that separates the feature space into distinct feeling classes.

CNN:

In sentiment analysis for Bengali e-commerce comments, convolutional neural networks (CNNs) are employed to capture complex patterns and characteristics of textual information. CNNs, originally designed for image processing tasks, have proven

to be highly effective for natural language processing (NLP) tasks, including sentiment analysis. They are adept at identifying local patterns such as word combinations and phrases, which are crucial for understanding feelings.

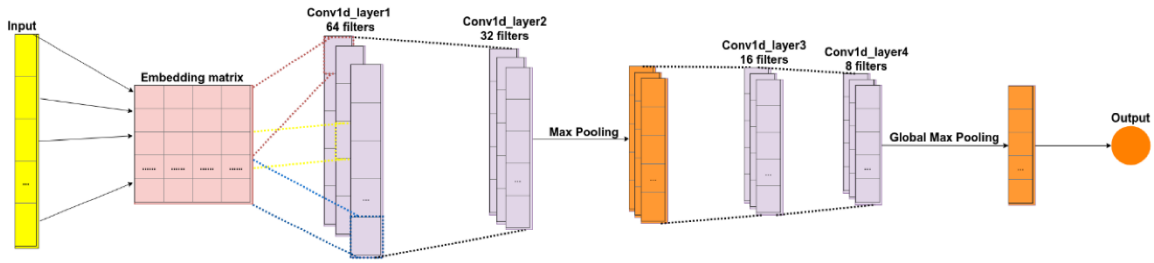


Figure 4.2.6: Working Flow of Convolutional Neural Network on Text.

Input level: Text data (e-commerce comments) is first tokenized and converted into a sequence of word embeddings. These embeddings are dense vector representations of words, containing semantic meanings.

Convolutional layer: The convolutional layer implements multiple filters (kernels) in the input embedding. Each filter is a matrix that slides over the input matrix to create the feature map. Mathematically, a convolution operation can be represented as

$$h_i = f(w \cdot x_{i:i+k-1} + b) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (vii)$$

Pooling layer: After convolution, a pooling layer (usually max-pooling) is applied to reduce the dimension and retain the most important properties. can be represented as a maximum-pooling operation:

$$p = \max(h_{i:i+m-1}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (viii)$$

Output level: The final output level generates a probability distribution on the positive or negative of the emotion class using a softmax activation function. The softmax function is given by:

$$P(y = j|z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (ix)$$

CNNs are utilized to analyze Bangla e-commerce comments by identifying key phrases and contextual information that indicate sentiment. The architecture is designed to handle the morphological richness and linguistic complexity of the Bangla language.

LSTM:

LSTM networks are used to capture temporal dependencies and contextual information in sequential text data. LSTM is a type of recurrent neural network (RNN) that is specifically designed to overcome the limitations of standard RNNs, such as the invisible gradient problem, which enables them to effectively learn long-term dependencies.

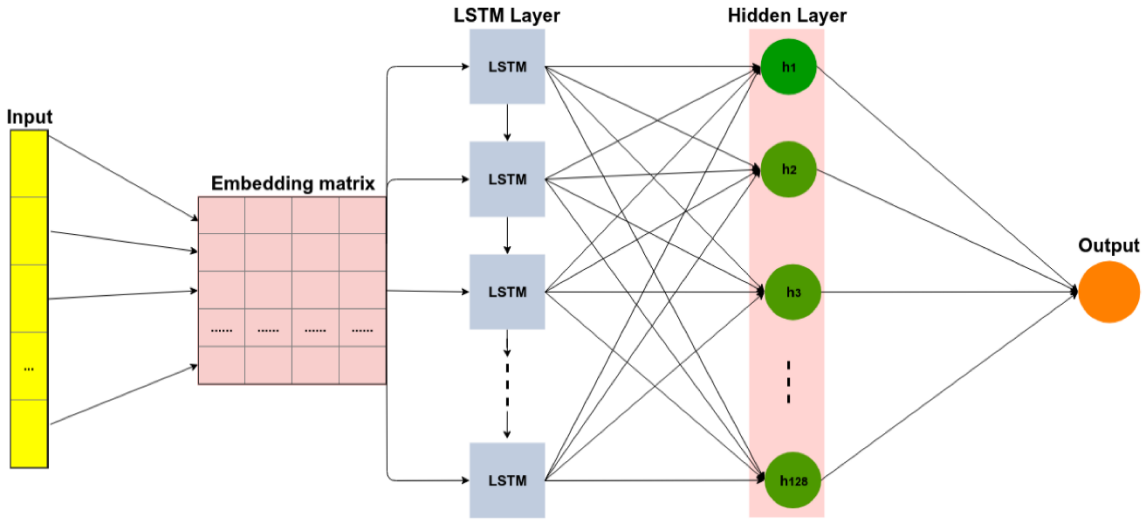


Figure 4.2.7: Working Flow of LSTM on Text.

LSTM networks are utilized to analyze Bangla e-commerce comments by understanding the sequential nature of text and capturing long-term dependencies. The architecture is tailored to handle the linguistic complexities and nuances of the Bangla language, making it suitable for sentiment analysis. By leveraging LSTMs, the model can effectively learn from annotated datasets, providing robust insights into customer feedback.

LIME:

LIME is an explanation technique used to understand the predictions of complex machine learning models. In the context of sentiment analysis for Bangla e-commerce comments, LIME helps in interpreting how the model makes decisions for individual predictions, offering insights into which features (words or phrases) contribute to the sentiment classification. LIME is used to provide explanations for the predictions made by sentiment analysis models applied to Bangla e-commerce comments. By generating interpretable explanations, LIME helps in understanding which words or phrases in a comment are driving the sentiment prediction (positive, or negative). This transparency is crucial for validating the model's performance and building trust among users and stakeholders. LIME's explanations enable researchers and practitioners to fine-tune models, identify potential biases, and ensure the model's decisions align with human expectations, ultimately improving the reliability and acceptance of sentiment analysis systems in real-world applications.

4.3 Model Evaluation

Accuracy:

Accuracy is a fundamental criterion for evaluating the effectiveness of the sentiment analysis model, especially in the context of Bengali e-commerce comments. It measures the ratio of correctly categorized instances to total instances and provides a simple way to evaluate the overall effectiveness of a model. The mathematical formula for accuracy is given by:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (x)$$

The models provide a clear and intuitive measure of how well Bangla e-commerce comments categorize into positive or negative.

Precision:

Accuracy is an important criterion in evaluating the effectiveness of the sentiment analysis model, especially in the context of Bengali e-commerce comment analysis. It measures the accuracy of positive (or negative) predictions made by the model, providing insight into its ability to avoid false positive (or negative) classifications. The mathematical formula for accuracy is given by:

$$\text{Precision} = \frac{TP}{TP + FP} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (xi)$$

Recall:

Recall is a fundamental measure in evaluating the effectiveness of the sentiment analysis model, especially in the context of Bengali e-commerce comment analysis. It measures the model's ability to identify all relevant instances of positive (or negative) feelings, thereby providing insight into its ability to avoid false negative classifications. The mathematical formula is given to remember:

$$\text{Recall} = \frac{TP}{TP + FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (xii)$$

F1-score:

The F-1 score is an important measure in evaluating the overall performance of the sentiment analysis model, especially in the context of Bengali e-commerce comment analysis. It combines both accuracy and recall into a single value, which provides a balanced measure of a model's accuracy. F1 - The mathematical formula for the score is given:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (xiii)$$

However, there is a limit to how closely the model can simulate the training set of data before it loses the ability to generalize. The performance of an algorithm can be evaluated using a specific table format called Confusion Matrix. (CM). CM is used to illustrate important predictive parameters such as recall, specificity, accuracy and precision. Confusion matrices are useful because they provide a direct evaluation of values such as TP, FP, TN, and FN. The following sections answer the research question of this study based on the research question.

4.4 Summary

In sections 4.1 to 4.3, this study presents a comprehensive approach to developing and evaluating sentiment analysis models for Bengali e-commerce comments. Using Python, TensorFlow, Keras, Skit-Learn, and Google Collab, the study applied traditional algorithms such as Logistic Regression, Naïve Bayes, and SVM, as well as deep learning models such as CNN, LSTM, and Bi-LSTM. This instrumentation facilitated efficient model training and deployment on a website using Flask, HTML, CSS, and JavaScript. Statistical analyses using Matplotlib and WordCloud provide insight into dataset characteristics, including sentiment distribution and sound frequency. Evaluation metrics including accuracy, precision, recall, and F1-score were employed to evaluate the effectiveness of the model, which detailed the predictive outcomes supported by the Confusion Matrix analysis. By integrating these approaches, the study enhances the understanding of sentiment dynamics in Bengali e-commerce, ensuring robust model functionality and practical applicability in real-world settings.

CHAPTER 5
Result and Analysis

5.1 Overview

The experimental setup for this study involves a systematic configuration of hardware, software, and data resources designed to conduct a comprehensive investigation of sentiment analysis. High-performance computing resources, including a powerful central processing unit (CPU) and potentially powerful graphics processing unit (GPU), are essential for accelerated processing tasks. Software requirements encompass deep learning frameworks like TensorFlow, along with libraries such as scikit-learn for machine learning algorithms, all integrated within a Python programming environment. Additionally, specialized libraries for NLP in the Bangla language are utilized for text preprocessing and feature extraction. The experimental dataset, comprised of Bangla e-commerce comments, is meticulously collected using web scraping techniques and undergoes preprocessing to remove noise, handle imbalances, and extract relevant features. Various machine learning and deep learning models including logistic regression, decision trees, naive bias, k-nearest neighbors, support vector machines, iterative neural networks, convolutional neural networks are applied and trained on the processed dataset. Evaluation metrics such as accuracy, precision, recall, F-1 score, and explanatory score are employed to comprehensively evaluate the effectiveness of the model. Ethical considerations, including data confidentiality and compliance, are carefully adhered to throughout the vetting process. Detailed documentation of code, model configuration, hyperparameters, and experimental results are maintained to facilitate transparency, reproducibility, and further research collaboration in the field of Bengali sentiment analysis.

5.2 Experimental Result

Result of Experiment 1: Machine Learning algorithms

This section conducts a comparative analysis of seven machine learning algorithms: LR, DT, RF, multinomial naive bias, KNN, SVM, and SGD. The comparison begins with an assessment of their classification performance, followed by a review of the overall metrics of these models. The collection of each algorithm, the description, the possible causes of their performance and the areas of improvement of the results are thoroughly examined.

TABLE 5.2.1: ACCURACY FOR CLASSIFICATION OF INDIVIDUAL MACHINE LEARNING MODEL.

Algorithm	Training Accuracy	Validation Accuracy	Model Accuracy
LR	85.6%	83.3%	83.2%
DT	99.6%	81.2%	82.6%
RF	99.6%	86.1%	85.7%
MNB	52.2%	50.9%	52.4%
KNN	66.7%	59.3%	57.9%
SVM	89.6%	87.6%	86.7%
SGD	87.1%	83.4%	83.3%

TABLE 5.2.1 presents the comparative analysis performance of various machine learning algorithms based on training, validation, and test accuracy. LR shows training, validation, and test accuracies of 85.6%, 83.3%, and 83.2%, respectively. DT achieving 99.6% training accuracy and 82.6% test accuracy. RF performs well, with high training 99.6%, validation 86.1%, and test accuracies 85.7%. MNB and KNN struggles. MNB achieving 52.2% training and 52.4% test accuracy. KNN struggles with low accuracies across the board training: 66.7%, test: 57.9%. SVM excels with high training 89.6%, validation 87.6%, and test accuracies 87.7%, SGD shows good training performance 87.1% and test accuracy of 83.3%. SVM is the best performer, while MNB significantly underperforms.

TABLE 5.2.2: PRECISION, RECALL, F1-SCORE, AND SUPPORT FOR MACHINE LEARNING ALGORITHMS.

Logistic Regression				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.80	0.89	0.84	552
Positive	0.87	0.78	0.82	551
Accuracy			0.83	1103
Macro Avg	0.84	0.83	0.83	1103
Weighted Avg	0.84	0.83	0.83	1103
Decision Tree				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.81	0.86	0.83	552
Positive	0.85	0.79	0.82	551
Accuracy			0.83	1103
Macro Avg	0.83	0.83	0.83	1103

Weighted Avg	0.83	0.83	0.83	1103
Random Forest				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.84	0.88	0.86	552
Positive	0.88	0.83	0.85	551
Accuracy			0.86	1103
Macro Avg	0.86	0.86	0.86	1103
Weighted Avg	0.86	0.86	0.86	1103
Multinomial NB				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.91	0.05	0.10	552
Positive	0.51	0.99	0.68	551
Accuracy			0.52	1103
Macro Avg	0.71	0.52	0.39	1103
Weighted Avg	0.71	0.52	0.39	1103
K Neighbors Classifier				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.60	0.95	0.74	552
Positive	0.88	0.37	0.52	551
Accuracy			0.66	1103
Macro Avg	0.74	0.66	0.63	1103
Weighted Avg	0.74	0.66	0.63	1103
SVM				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.87	0.87	0.87	552
Positive	0.87	0.86	0.87	551
Accuracy			0.87	1103
Macro Avg	0.87	0.87	0.87	1103
Weighted Avg	0.87	0.87	0.87	1103

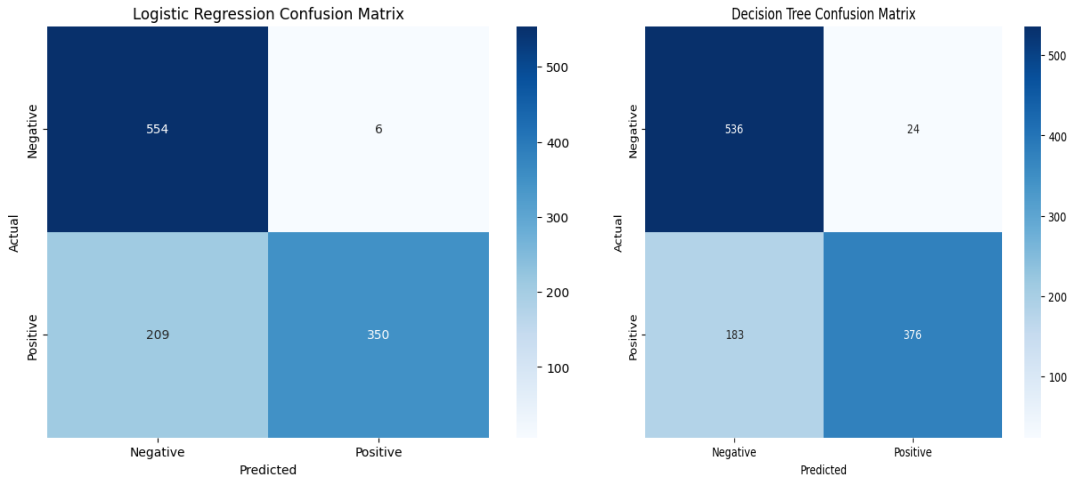
SGD				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.81	0.88	0.84	552
Positive	0.87	0.79	0.83	551
Accuracy			0.83	1103
Macro Avg	0.84	0.83	0.83	1103
Weighted Avg	0.84	0.83	0.83	1103

TABLE 5.2.2 provides a comparative analysis of various classification models applied to a dataset with 1103 samples, divided almost evenly into 552 Negative and 551 Positive labels. The models evaluated include LR, DT, RF, MNB, KNN, SVM, and SGD. Among these models, the SVM and RF exhibit the highest performance, with accuracies of 0.87 and 0.86, respectively, and both achieving a balanced macro average F1-score of 0.87 and 0.86. LR, DT, and SGD models each attain an accuracy of 0.83, displaying similar performance metrics with a macro average F1-score of 0.83. KNN shows moderate performance with an accuracy of 0.66 and a macro average F1-score of 0.63. MNB model demonstrates the weakest performance, with a notably low accuracy of 0.52. Although it achieves a high recall of 0.99 for the Positive class, its precision for this class is significantly low at 0.51, resulting in a poor macro average F1-score of 0.39. Overall, the SVM and RF models are identified as the most effective for this dataset, providing high accuracy and balanced precision, recall, and F1-scores across both classes. Other models like LR, DT, and SGD also perform well, whereas the Multinomial Naive Bayes model's performance suggests it is less suitable for this specific task.

Confusion Matrix Machine Learning Algorithms:

A confusion matrix is a performance measurement tool used in ML and statistical classification to assess the accuracy of a classification algorithm. This is a table that compares the actual values to the predicted values produced by the model. A matrix usually has four components: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These components allow different performance metrics to be calculated, such as accuracy, precision, recall, and F1-score. Provides a detailed breakdown of prediction results, helps identify how well the confusion matrix model distinguishes between different classes, and highlights where the model can make mistakes.

LR and DT Confusion Matrix:

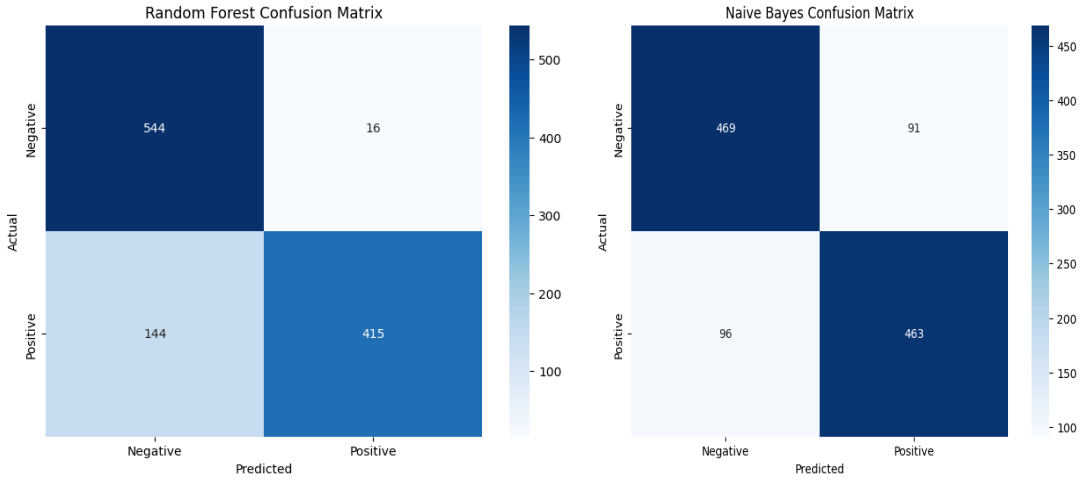


(a) CM of LR Algorithms (b) CM of DT Algorithms

Figure 5.2.1: Confusion Matrix for LR and DT Algorithm.

Figure 5.2.1 (a) show that LR predict 904 sentences accurately. But 215 sentences predict wrong and figure (b) show that DT predict 912 sentences accurately but for 207 sentence it makes wrong prediction.

RF and Confusion Matrix:



(a) CM of RF Algorithms (b) :CM of NB Algorithms

Figure 5.2.2: Confusion Matrix for RF and NB Algorithms.

Figure 5.2.2 (a) show that RF predict 959 sentences accurately, 160 sentence wrong predicted and figure(b) shows that NB 932 predict sentence accurately and 187 sentence make wrong prediction.

KNN and SGD Confusion Matrix:

(a) CM of KNN Algorithms (b) :CM of SGD Algorithms

Figure 5.2.3: Confusion Matrix for KNN and SGD Algorithms.

Figure 5.2.3 (a) show that KNN 662 predict sentence accurately but for 471 sentences it makes predict wrong. In (b) show that SGD predict 900 sentences accurately but 219 sentence predict wrong.

SVM Confusion Matrix:

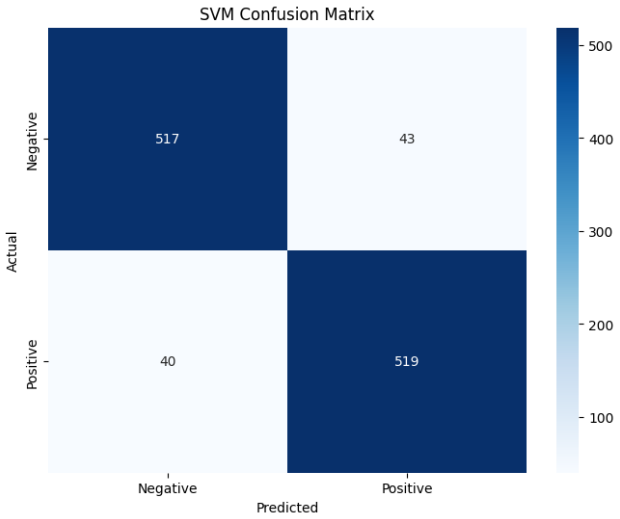


Figure 5.2.4: Confusion Matrix for SVM Algorithms.

In Figure 5.2.4 shows that SVM predict 1036 sentences successfully only 83 sentences make wrong prediction.

Explanation Sentence with LR model:

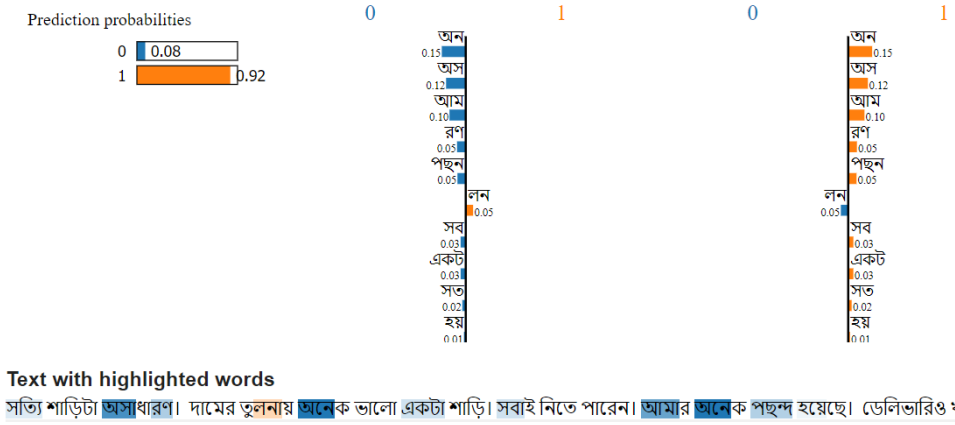


Figure 5.2.5: Sentence Explanation using LIME with LR Model.

Figure 5.2.5 shows the LIME explanation for this prediction. Each row in the table corresponds to a word in the sentence, and the value in the table represents the contribution of that word to the model's prediction. Words with higher values contributed more to the model's prediction of a positive sentiment. For example, the word "অনেক" has a value of 0.15, which is the highest value in the table. This means that this word had the strongest positive influence on the model's prediction. Other words with positive contributions include "অসাধারণ", "ভালো" and "পছন্দ".

Explanation Sentence with DT model:

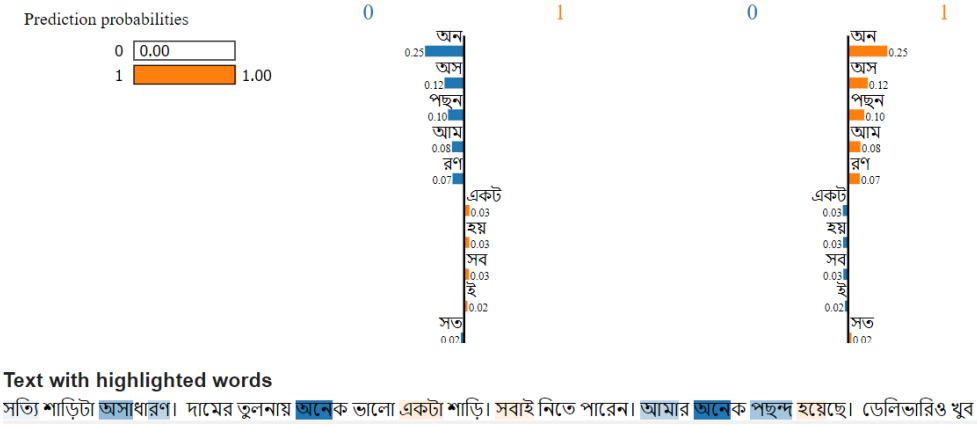


Figure 5.2.6: Sentence Explanation using LIME with DT Model.

Figure 5.2.6 show that "অনেক" word has more contribution (0.25) to make this sentence positive. Other words are "অসাধারণ", "পছন্দ" and "আমার" respectively 0.19, 0.10 and 0.05.

Explanation Sentence with RF model:

Figure 5.2.7 show that "অনেক" , "আমার" , "অসাধারণ" and "পছন্দ" words are plays important role to make this sentence positive.

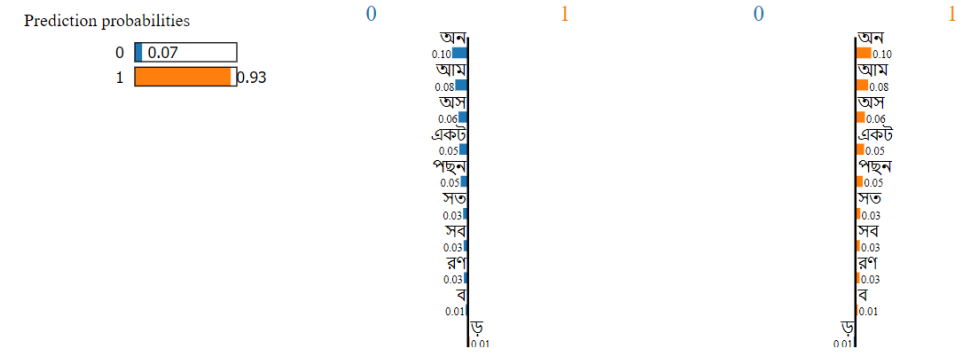


Figure 5.2.7: Sentence Explanation using LIME with RF Model.

Explanation Sentence with KNN model:

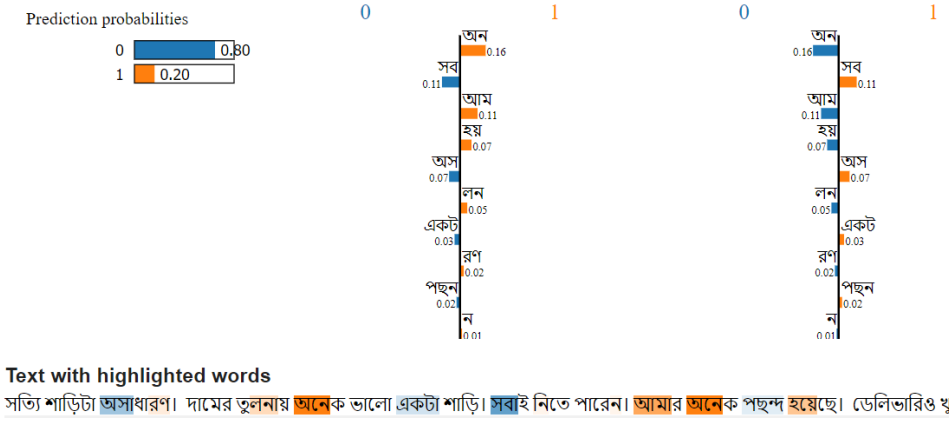


Figure 5.2.8: Sentence Explanation using LIME with KNN Model.

Figure 5.2.8 show that "অনেক", "আমার" words are plays important role to make this sentence negative. That's mean when I train my dataset using KNN model these words are train for negative sentence. In KNN model it is biased and give us wrong prediction. That's why I use explainable AI technique so that I can identify which word plays important role for making prediction.

Explanation Sentence with NB model

Figure 5.2.9: Sentence Explanation using LIME with NB Model.

Figure 5.2.9 show that "অনেক", "অসাধারণ" and "পছন্দ" words are plays important role to make this sentence positive.

Explanation Sentence with SGD model

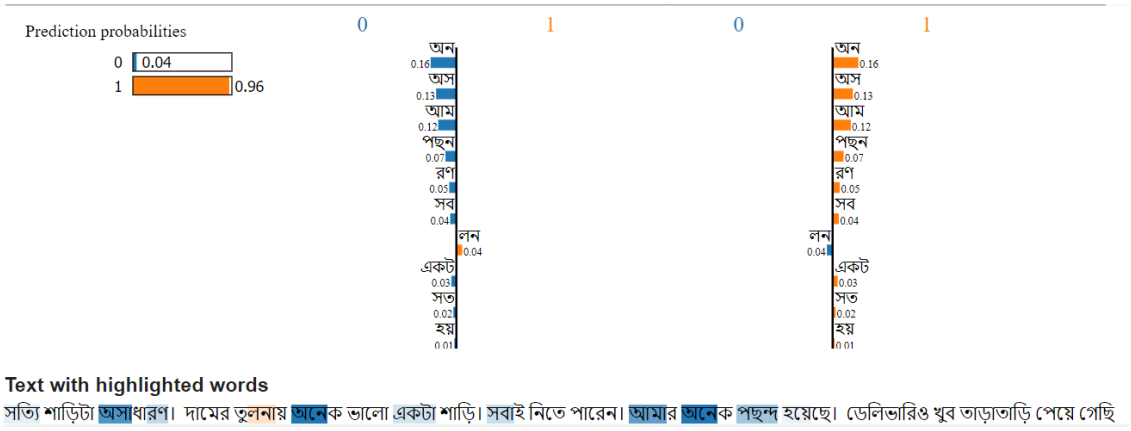


Figure 5.2.10: Sentence Explanation using LIME with SGD Model.

Figure 5.2.10 show that "অনেক" , "অসাধারণ" , "আমার" and "পছন্দ" words are plays important role to make this sentence positive respectively 0.16, 0.13, 0.19 and 0.07

Result of Experiment 2: Deep Learning:

In this section, a comprehensive comparative analysis is done on three deep learning models namely CNN, LSTM, and Bi-LSTM. The evaluation begins with an in-depth examination of the classification performance of each model. Subsequently, an overall review of the overall measurements for these models is presented, which covers key performance indicators such as accuracy, precision, recall, and F1 score.

TABLE 5.2.3: ACCURACY FOR DEEP LEARNING CLASSIFICATION.

Architecture	Training Accuracy	Validation Accuracy	Model Accuracy
CNN	99.4%	84.6	88.9%
LSTM	99.2%	83.4	85.6%
Bi-LSTM	99.4%	85.1	87.3%

Table 5.2.3 summarizes the performance metrics of four different neural network architectures: CNN, LSTM and Bi-LSTM. Each architecture is evaluated based on its training accuracy and model accuracy. CNN achieves a remarkably high training accuracy of 99.4%, indicating its capability to effectively learn from the training data. However, its model accuracy on unseen data is 88.9%, suggesting a moderate level of overfitting but still demonstrating strong generalization. LSTM and Bi-LSTM also exhibit high training accuracies of 99.2%, similar to CNN. However, their model accuracies on the test data are slightly lower, with LSTM at 85.3% and Bi-LSTM at

87.3%. This indicates that both LSTM and Bi-LSTM models experience some degree of overfitting and may not generalize as effectively as CNN

TABLE 5.2.4: PRECISION, RECALL, F1-SCORE, AND SUPPORT RESULT OF DEEP LEARNING MODEL.

CNN				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.81	0.58	0.68	144
Positive	0.90	0.96	0.93	551
Accuracy			0.88	695
Macro Avg	0.85	0.77	0.80	695
Weighted Avg	0.88	0.88	0.88	695
LSTM				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.65	0.50	0.57	144
Positive	0.88	0.93	0.90	551
Accuracy			0.84	695
Macro Avg	0.77	0.72	0.74	695
Weighted Avg	0.83	0.84	0.83	695
Bi-LSTM				
	Precision	**Recall**	**F1-Score**	**Support**
Negative	0.68	0.63	0.65	144
Positive	0.91	0.92	0.91	551
Accuracy			0.86	695
Macro Avg	0.79	0.78	0.78	695
Weighted Avg	0.86	0.86	0.86	695

Table 5.2.4 provides performance metrics for three deep learning models CNN, LSTM, and Bi-LSTM evaluated on a dataset comprising 695 samples, with 144 labeled as Negative and 551 as Positive. The CNN model demonstrates the highest overall accuracy at 0.88, with particularly strong precision 0.90 and recall 0.96 for the Positive class, resulting in a high F1-score of 0.93. However, its performance for the Negative class is lower, with a recall of 0.58 and an F1-score of 0.68. The LSTM model achieves an accuracy of 0.84, also performing well for the Positive class with a F1-score of 0.90, but showing the lowest performance for the Negative class, with a precision of 0.65, recall of 0.50, and F1-score of 0.57. The Bi-LSTM model strikes a balance between these two, achieving an overall accuracy of 0.86, with improved metrics for the Negative class precision of 0.68, recall of 0.63, and F1-score of 0.65 compared to the

LSTM, while maintaining high performance for the Positive class F1-score of 0.91. The macro average F1-scores for CNN, LSTM, and Bi-LSTM are 0.80, 0.74, and 0.78, respectively.

Training and Validation Accuracy and loss of CNN:

Figure 5.2.11 displays the training and validation performance of a CNN over 20 epochs. The left plot shows the training loss (blue line) rapidly decreasing and stabilizing near zero, while the validation loss (orange line) initially decreases but then increases, indicating overfitting. The right plot shows the training accuracy (blue line) quickly reaching nearly 100%, whereas the validation accuracy (orange line) fluctuates around 85% without significant improvement, further suggesting that the model is overfitting to the training data.

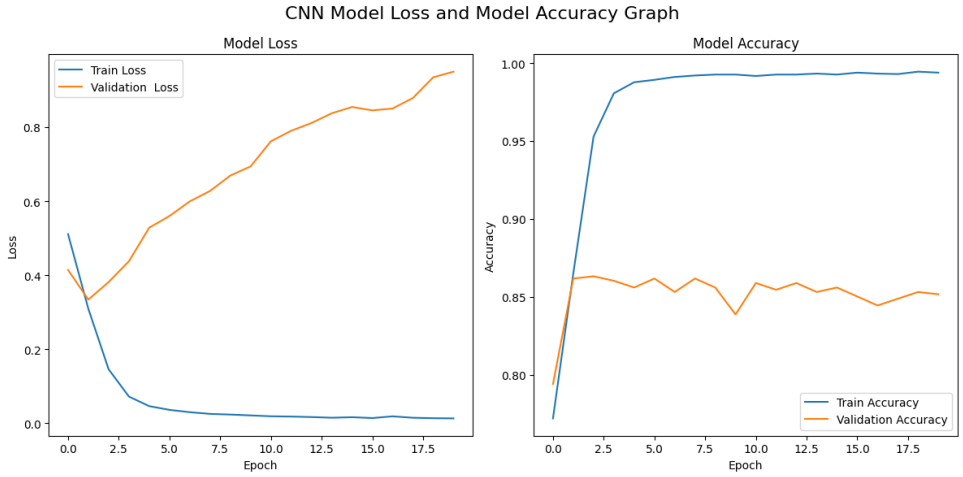


Figure 5.2.11: Training and validation accuracy and loss over the epochs CNN

Training and Validation Accuracy and loss of LSTM:

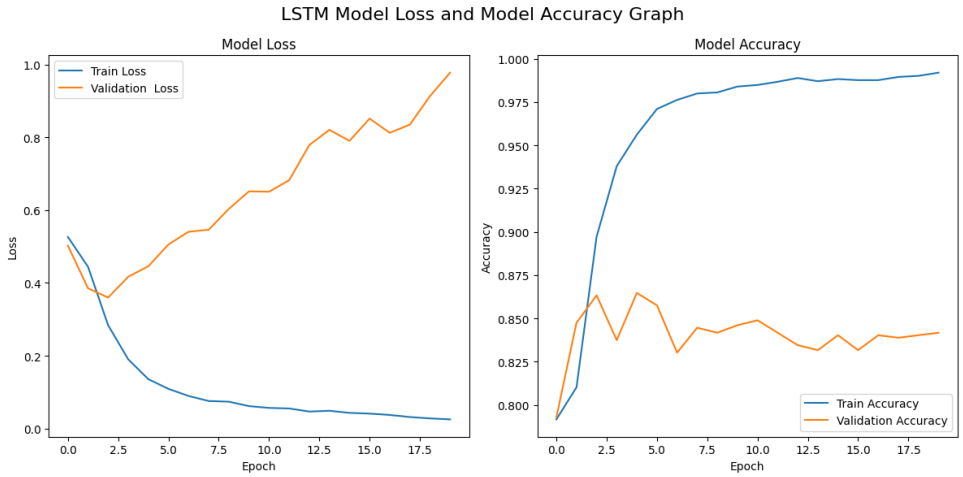


Figure 5.2.12: Training and validation accuracy and loss over the epochs LSTM.

Figure 5.2.12 displays the training and validation performance of a LSTM over 20 epochs. The left plot shows the training loss rapidly decreasing and stabilizing near

zero, while the validation loss initially decreases but then increases, indicating overfitting. The right plot shows the training accuracy quickly reaching nearly 100%, whereas the validation accuracy fluctuates around 83% without significant improvement.

Training and Validation Accuracy and loss of Bi-LSTM:

Figure 5.2.13 shows that the training and validation performance of a Bi-LSTM over 20 epochs. The left plot shows the training loss rapidly decreasing and stabilizing near zero, while the validation loss initially decreases but then increases, indicating overfitting. The right plot shows the training accuracy quickly reaching nearly 100%, whereas the validation accuracy fluctuates around 85% without significant improvement.

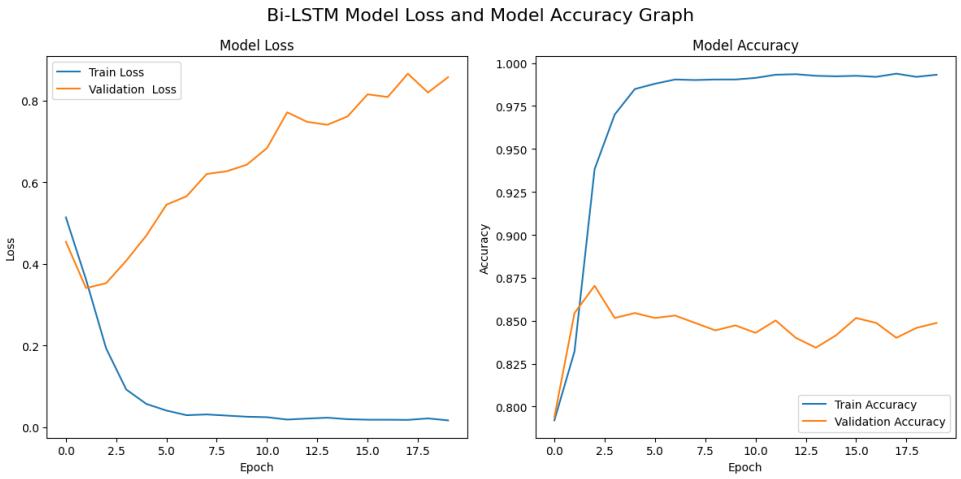
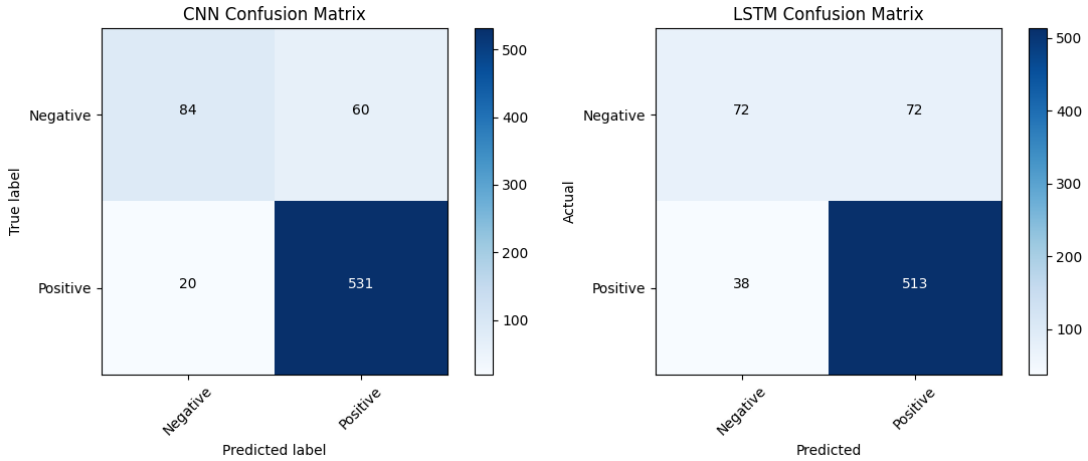


Figure 5.2.13: Training and validation accuracy and loss over the epochs Bi-LSTM.

Confusion Matrix of CNN and LSTM:

A confusion matrix is an important tool in evaluating the performance of a classification model, providing a detailed breakdown of predicted versus actual results across different classes. Figure 5.2.14 (a) show that CNN predict 615 sentences accurately. But 80 sentences predict wrong among 695 sentences. and figure (b) show that LSTM predict 585 sentences accurately but for 110 sentence it makes wrong prediction among 695 sentences.



(a) CM of CNN (b) : CM of LSTM

Figure 5.2.14: Confusion Matrix for CNN and LSTM Deep learning Model.

Confusion Matrix of Bi-LSTM:

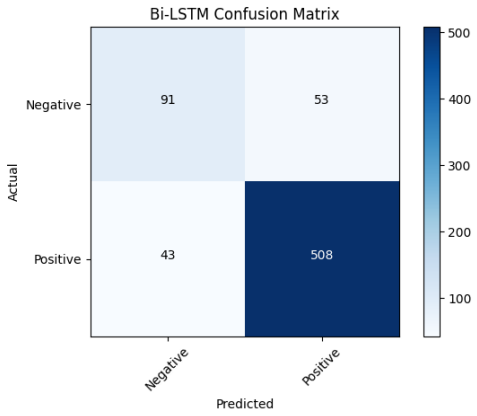


Figure 5.2.15: Confusion Matrix of Bi-LSTM Deep learning Model.

In Figure 5.2.15 show that Bi-LSTM predict 599 sentences accurately. But 96 sentences predict wrong among 695 sentences.

Web Site design specification:

The research aims to provide a user-friendly web interface that allows users to analyze the sentiment of Bangla text inputs efficiently. The design of the website is centered around simplicity and accessibility, ensuring that users can easily navigate through the features and functionalities. The homepage features a clean and intuitive form where users can enter Bangla text directly into a text area or upload a file containing the text. The form submission is facilitated by a prominently displayed "Analyze" button, which triggers a request to the server for real-time sentiment analysis. The results are dynamically displayed on the result page, ensuring a seamless user experience without the need for page reloads. The website's aesthetic is minimalistic yet visually appealing,

with a focus on readability and ease of use. The text area for input is sufficiently large to accommodate extended text entries, and the upload button is clearly labeled for users who prefer to analyze text files. The results page is designed to provide clear and concise feedback, highlighting the analyzed sentiment prominently.

Front-end Design

Font-end design is very important for a website where a user directly interact with the web site. So, I design my website as simple as possible:

Home page:

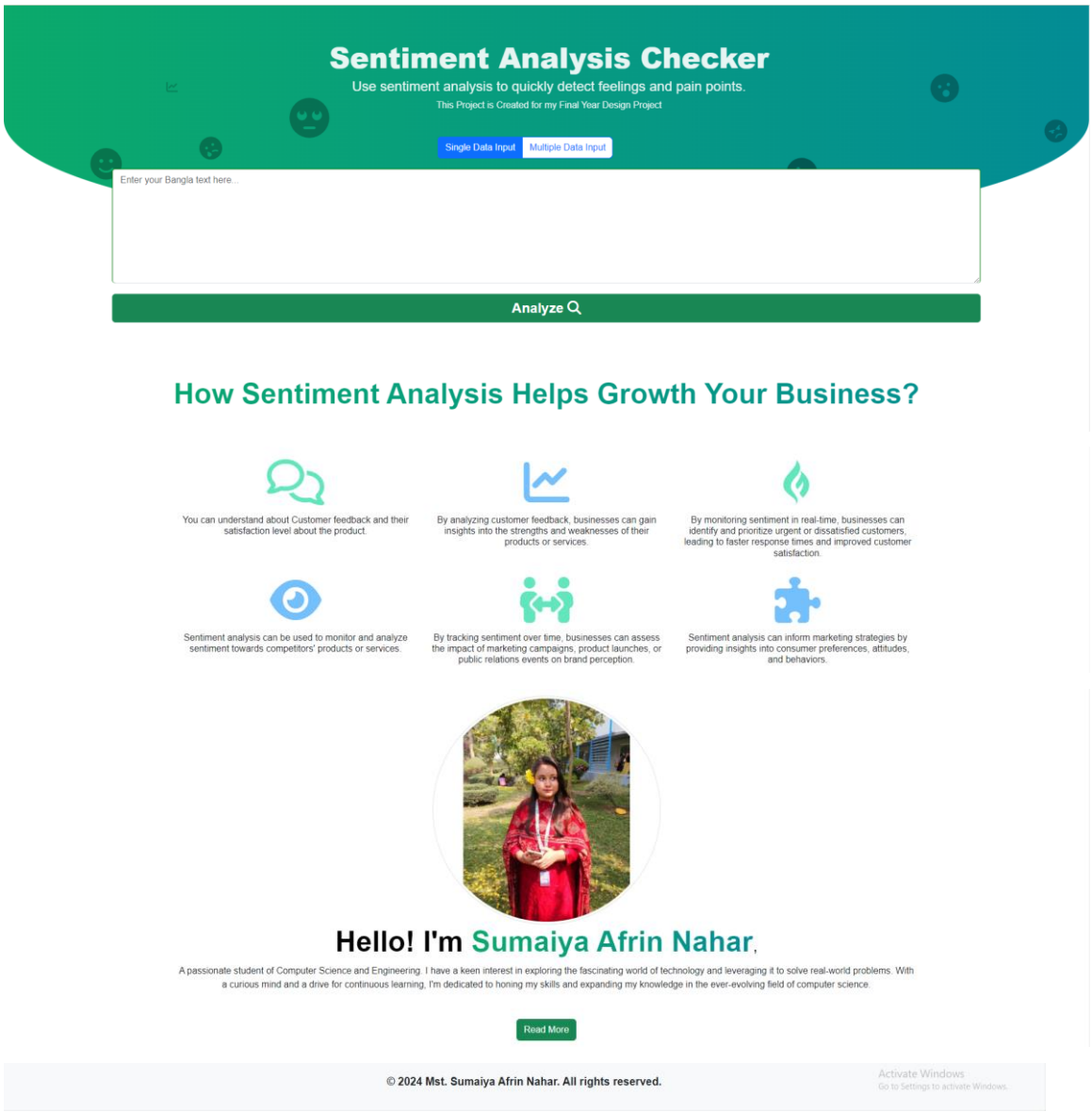


Figure 5.2.16: Home Page Design for user-friendly Website.

Figure 5.2.16 showcases the homepage of a Bangla sentiment analysis website. Users can input Bangla text, and the website will analyze its sentiment. The site can handle multiple text inputs, entire CSV files, and show a section on the importance of

sentiment analysis in business. Clicking "Read More" button reveals further details about the project.

Result Page:

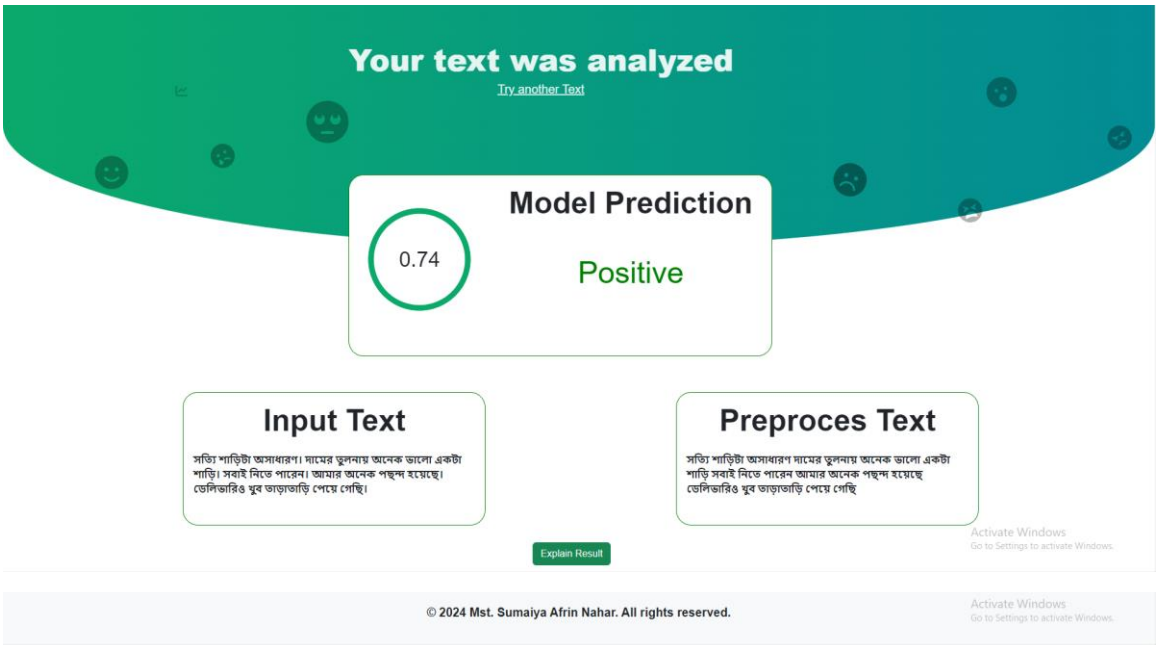


Figure 5.2.17: Model Result Page Design for single text input.

Model Explanation Page:

Figure 5.2.18: Model Explanation Page Design for single text input.

Figure 5.2.17 showcase that user give an input in home page then model predict the sentiment an show the result in the result page. Here there are three section first one is model prediction with confidence score, it also show that input text and preprocess text. Figure 5.2.18 showcase explanation of which word play important role to make prediction.

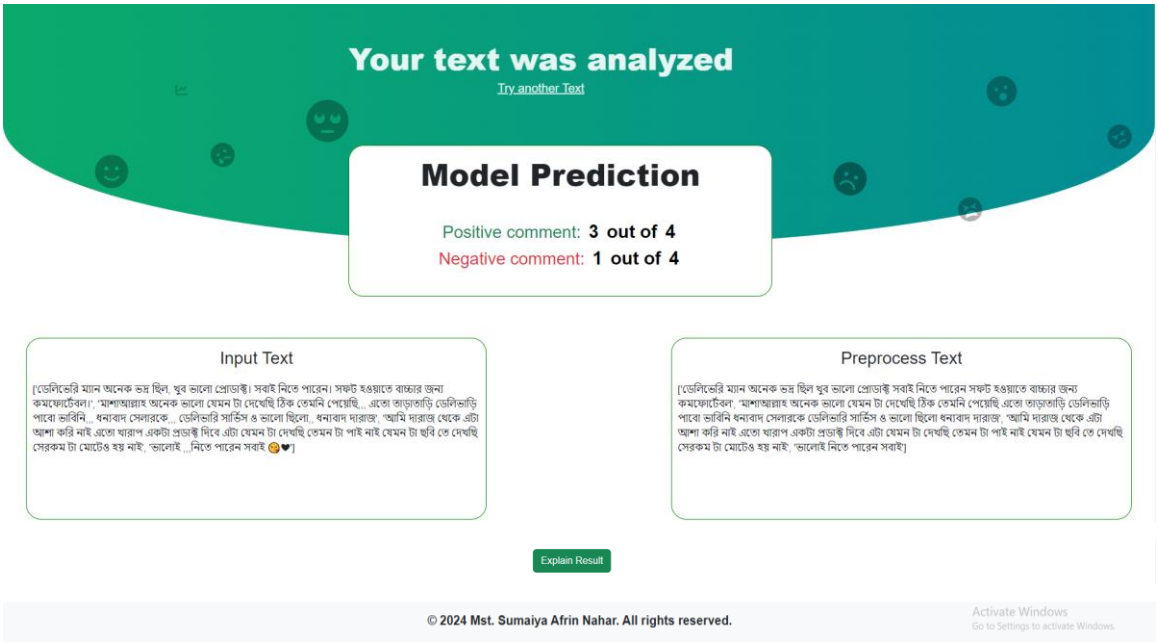


Figure 5.2.19: Model Result Page Design for single text input.

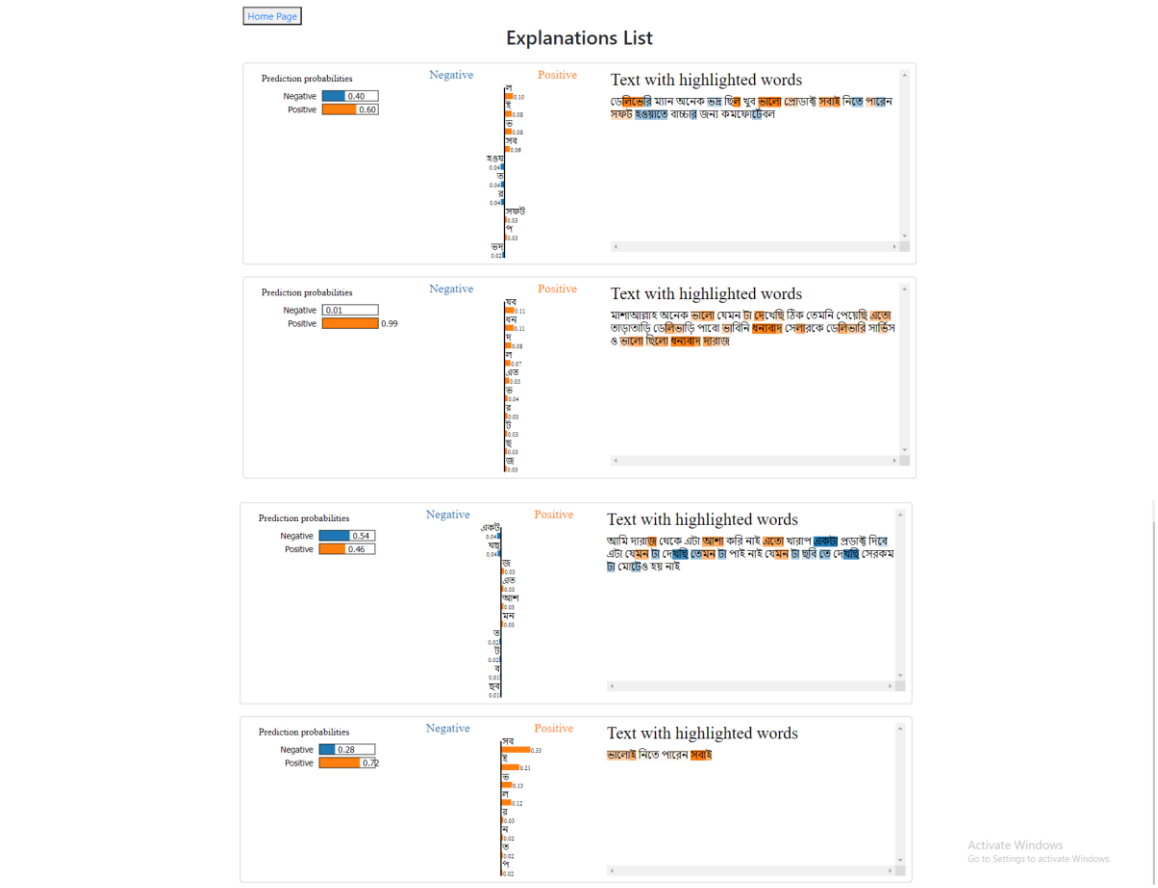


Figure 5.2.20: Model Explanation Page Design for multiple text input.

Figure 5.2.19 showcase that user give an input in home page as a list of data or they can upload a csv file that has user comment, then model predict the sentiment a show the result in the result page. Here there are three sections first one is how many comments are positive or negative, it also shows that input text and preprocess text. Figure 5.2.20

showcase explanation of which word play important role to make predictionComparative Analysis.

5.3 Comparative Analysis

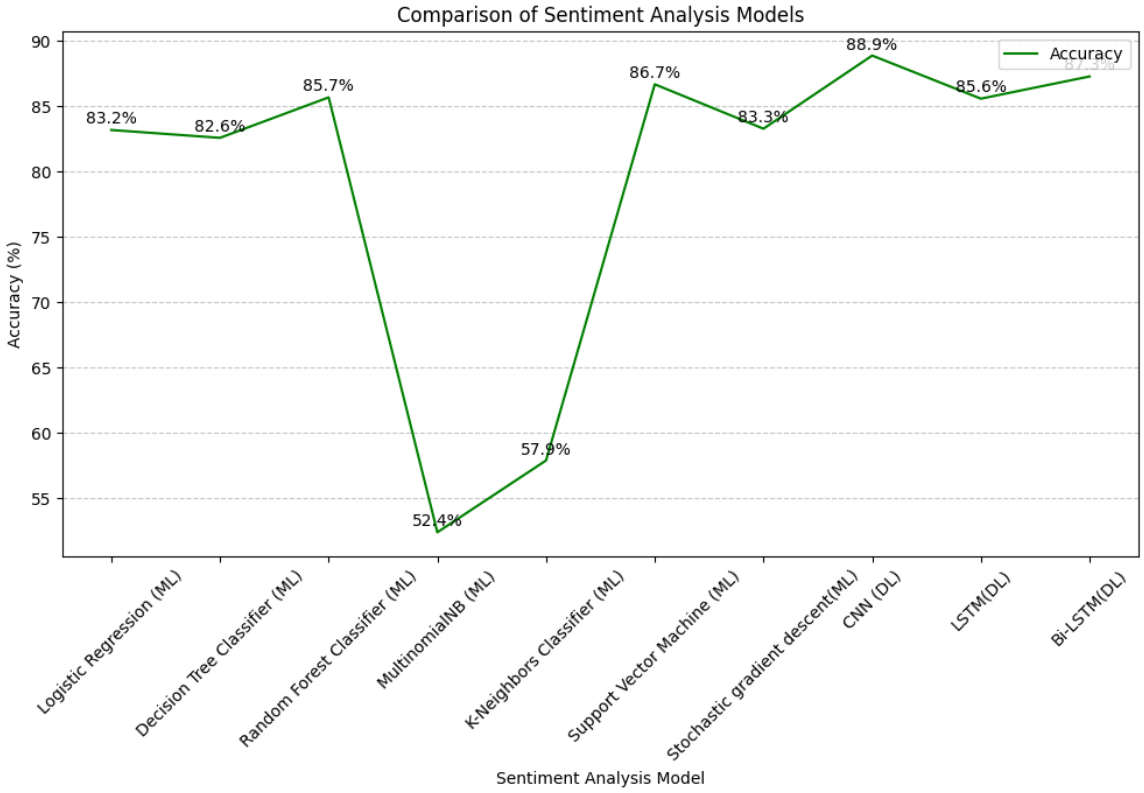


Figure 5.3.1: Accuracy comparison among all models of my study.

Figure 5.3.1 presents a comparative analysis of different sentiment analysis models, highlighting their accuracy in a specific dataset. Among the traditional Machine Learning models, the SVM outperformed all others with an impressive accuracy of 86.7%, demonstrating its effectiveness in high-dimensional space classification. The RF, DT and LR classifiers also showed strong performances with accuracies of 85.7% and 82.2% and 83.2% respectively, indicating their robustness and efficiency. In contrast, the MNB and KNN performed poorly with52.4% and 57.9%, likely due to high-dimensional data challenges. Deep Learning (DL) models, including CNN, exhibited competitive accuracy at 88.9%, suggesting their capability to capture complex patterns in text data. However, RNN -based models, such as LSTM and "Bi-LSTM, achieved lower accuracies of 85.6% and 87.34%, respectively. Overall, the analysis reveals that traditional ML models, particularly SVM, can outperform DL models in sentiment analysis for this dataset, underscoring the importance of algorithm selection based on data characteristics and suggesting potential for further optimization in DL approaches.

Comparison my model with other study:

TABLE 5.3.1: COMPARISON MY WORK WITH OTHER STUDY.

Study	Dataset	Method & Techniques	Results
Abdelwahab, Youmna et al. [10]	Arabic Tweets (4201)	LSTM Classifier and LIME XAI Algorithm	LSTM = 79%
Haque, Rezaul et al [14]	Secondary Data	SVC, SGD, RF, LR, MNB, DT, LR, BiGRU, BiLSTM, LSTM, CLSTM	CLSTM=85%
Ashik, Md Akhter-Uz-Zaman et al. [17]	Bengali News comments	LSTM	LSTM = 79%
Palash, Partha Protim Das et al. [18]	Newspaper comments	VAE	VAE = 53%
Tripto, Nafis Irtiza et al. [19]	YouTube comments	CNN and LSTM	LSTM = 66%
k. Sarkar, [20]	Bangla tweet dataset (SAIL content 2015)	LSTM	LSTM = 55%
Islam; Md Ashiqul et al. [21]	Facebook comments (2,500)	NB with unigram, NB with bigram	NB with bigram = 77%
Jain, Rachna et al. [22]	Secondary Data	VADER and LIME	VADER = 69%
Patel, Aksh et al. [23]	Secondary Data	Decision Tree, Random Forest, BERT,	BERT=83%
My Study	Dazar website user comment (primary data = 4887	LR, DT, RF, MNB, KNN, SVM, SGD, CNN, LSTM, Bi-LSTM	CNN = 88.9% SVM = 86.7% RF = 85.7%

Table 5.3.1 shows comparison to other studies in natural language processing, and my research on Dazar [22] website user comments demonstrate superior performance with CNN achieving 88.9% accuracy, outperforming SVM 86.7% and RF 85.7%.

5.4 Summary

Section 5.1 introduces the experimental framework designed for the evaluation of sentiment analysis in Bengali e-commerce comments, emphasizing the robust hardware configuration, the necessary software tools such as TensorFlow and Skit-Learn, and the specialized NLP library for Bengali text preprocessing. Datasets, generated through web scraping, undergo careful preprocessing to ensure the quality and relevance of the data. Various machine learning and deep learning models, including Logistic Regression, Decision Trees, Random Forest, SVM, CNN, LSTM, and Bi-LSTM, are

trained and evaluated using comprehensive metrics such as accuracy, precision, recall, and F1-score. Ethical considerations are prioritized throughout the study, ensuring compliance with data confidentiality and research transparency through detailed documentation of methods and results. In section 5.2, the experimental results of these models are presented, highlighting SVM and RF as the top performers among traditional ML models with 86.7% and 85.7% accuracy, respectively. CNN has emerged as the leading deep learning model with 88.9% accuracy, demonstrating its ability to capture complex patterns in Bangla text data. The detailed confusion matrix provides further insight into the predictive capabilities of models and areas for improvement, increasing the interpretability of their effectiveness in sentiment analysis tasks. Section 5.3 proposes a comparative analysis, benchmarking the results of the research against other NLP research efforts. Bengali e-commerce comments from the Dazar website highlighted the superior performance of the CNN model in achieving 88.9% accuracy, surpassing the SVM and RF models. This comparative analysis underscores the effectiveness of CNN in handling the nuances of sentiment analysis of Bangla text compared to other models used in similar studies.

CHAPTER 6

Impact on Society, Environment and Sustainability

6.1 Impact on Life

The impact of my research on daily life is profound, particularly within the realm of Bengali e-commerce and online consumer interactions. By developing and applying advanced machine learning and deep learning models to analyze sentiments expressed in Bangla language comments on platforms like the Dazar [24] website, the research enhances ability to understand consumer preferences and behaviors more accurately than ever before. This capability is crucial for businesses, enabling them to tailor their products, services, and marketing strategies based on real-time customer feedback. By achieving high accuracies, such as the SVM model's 86.7% accuracy, my research equips businesses with robust tools to make informed decisions swiftly and effectively. Furthermore, the ethical considerations embedded in the research, such as data privacy and transparency in AI methodologies, ensure that these advancements are implemented responsibly, fostering trust and reliability in AI-driven insights within the Bengali-speaking community. Overall, the impact extends beyond technological advancements to positively influence business strategies and enhance consumer experiences in the digital age.

6.2 Impact on Society & Environment

The impact of my research extends beyond individual consumer insights into broader social and environmental impacts. Within the context of society, the application of advanced sentiment analysis models in Bengali e-commerce comments facilitates an in-depth understanding of social trends, preferences, and consumer behavior. This knowledge can empower businesses to offer more relevant and targeted products and services, thereby potentially increasing economic growth and employment opportunities within the community. Furthermore, by promoting data-driven decision-making in business practices, my research contributes to a more efficient allocation of resources, reducing waste and environmental impacts associated with outdated or under-informed strategies. Ethically sound practices, including data privacy protection and transparent AI procedures, further enhance trust and social responsibility in using AI technologies.

6.3 Ethical Aspects

The ethical considerations of my research are paramount in ensuring responsible and transparent use of artificial intelligence in sentiment analysis of Bengali e-commerce comments. Key ethical aspects include data privacy and confidentiality, where stringent measures are employed to protect user information gathered from online platforms using web-scrapping technique. Adherence to ethical guidelines ensures that data is anonymized and aggregated to prevent identification of individual users without consent. Additionally, fairness and bias mitigation are critical; efforts are made to address biases in training data and algorithmic outputs to ensure equitable analysis across diverse demographic groups. Transparency in model development and deployment is maintained through detailed documentation of methodologies, model performances, and limitations, enabling stakeholders to understand and critique the AI-driven insights effectively. Continuous monitoring and updates to the AI models help mitigate risks associated with evolving data patterns and societal changes, ensuring that the technology benefits society without compromising ethical standards or privacy rights.

6.4 Sustainability Plan

Creating a sustainability plan for my research involves ensuring that the benefits of AI-driven sentiment analysis of Bengali e-commerce comments are long-lasting and impactful. Key elements of sustainability planning include ongoing research and development efforts to increase the accuracy and efficiency of sentiment analysis models. This includes staying updated with advances in natural language processing and machine learning techniques, constantly refining algorithms to adapt to changing linguistic patterns and user behavior on online platforms.The sustainability plan emphasizes educational initiatives aimed at raising awareness of AI technologies and their social impact. These include workshops, seminars, and outreach programs designed to educate stakeholders about ethical considerations, potential biases, and limitations of AI-driven sentiment analysis. Through informed discussion and the promotion of critical thinking, the research aims to foster a responsible and sustainable deployment of AI technologies in e-commerce and beyond.

6.5 Summary

My research on AI-driven sentiment analysis of Bengali e-commerce comments has significant implications across various dimensions. In terms of its impact on life, the research enhances user experience by providing accurate sentiment analysis, thereby aiding informed decision-making and improving consumer satisfaction. This contributes to a more efficient and personalized online shopping experience. Societally, the research fosters technological advancement and innovation in the domain of natural language processing, setting benchmarks for sentiment analysis in lesser-explored languages like Bengali. This not only supports local language preservation but also promotes digital inclusion and accessibility. Ethically, the research emphasizes transparency, fairness, and accountability in AI applications, addressing concerns such as bias mitigation and privacy protection. It underscores the importance of ethical AI development and deployment practices, ensuring responsible use of technology for societal benefit. The sustainability plan focuses on continuous research and development, knowledge sharing, and educational initiatives to advance AI capabilities responsibly. This includes fostering collaboration, disseminating findings, and engaging in educational outreach to promote a sustainable and ethical AI ecosystem. Overall, my research aims to not only advance technological frontiers but also to ensure that these advancements are ethically sound, sustainable, and beneficial to society at large. It represents a commitment to leveraging AI for positive societal impact while navigating the complexities of ethical considerations and sustainability in technology innovation.

CHAPTER 7
Conclusion and Future Work

7.1 Conclusions

In conclusion, this research has delved into the realm of sentiment analysis for Bengali e-commerce comments using a variety of machine learning and deep learning models. Through rigorous experimentation and analysis, the study has highlighted several key findings. The highest accuracy was achieved by the SVM model at 86.7%, showcasing its robust performance in capturing intricate patterns in text data. On the other end, the Multinomial Naive Bayes model exhibited the lowest accuracy at 52.4%, indicating challenges in handling the high-dimensional nature of the dataset effectively. A comparative analysis across various models revealed that traditional machine learning algorithms, particularly SVM with an accuracy of 86.7% and Random Forest with 85.7%, outperformed deep learning models such as LSTM and Bi-LSTM, which achieved accuracies of 85.6% and 87.3%, respectively. This underscores the importance of selecting appropriate algorithms based on dataset characteristics and performance metrics desired. Furthermore, the research contributes to the broader field of natural language processing by demonstrating the applicability of AI-driven sentiment analysis in underrepresented languages like Bengali. It emphasizes the potential of such technologies to enhance user experience, support decision-making processes, and foster digital inclusivity.

7.2 Further Suggested Works

Building on the findings and methodologies of this research, several avenues for further exploration and improvement emerge. Firstly, investigating ensemble methods that combine the strengths of both traditional machine learning and deep learning models could potentially enhance classification accuracy and robustness in handling diverse datasets. Additionally, exploring advanced deep learning architectures such as transformers and their variants like BERT for Bengali sentiment analysis could leverage pre-trained language models for improved performance. Moreover, integrating more sophisticated techniques for interpretability and explainability, beyond LIME, could provide deeper insights into model predictions and enhance trust in AI-driven decision-making processes. Furthermore, expanding the scope to include a larger and more

diverse dataset encompassing varied domains beyond e-commerce could offer insights into generalizability and applicability across different domains of Bangla text analysis.

7.3 Limitations

Despite its contributions, this research has several inherent limitations that warrant acknowledgment. Firstly, the dataset used, while comprehensive for Bangla sentiment analysis in e-commerce comments, may not fully represent the diversity of sentiments and language nuances across all Bangla text genres. This could affect the generalizability of the models to other domains or dialectical variations. Secondly, the performance metrics, such as accuracy and F1-score, though robust for evaluating model effectiveness, do not always capture the full complexity of sentiment analysis. Sentiments can be subjective and context-dependent, which may require more nuanced evaluation metrics or domain-specific adjustments. Furthermore, while efforts were made to ensure data preprocessing and model training were rigorous, challenges such as data imbalance, noisy data, and domain-specific vocabulary may still influence model performance. Addressing these issues comprehensively could further enhance the reliability and applicability of the sentiment analysis models. Additionally, the interpretability techniques employed, such as LIME, provide insights into model predictions but may not fully explain complex decisions made by deep learning models. Exploring more advanced and domain-specific interpretability methods could offer deeper insights into model behavior and enhance trust in AI-driven analyses.

References

[1] N. R. Bhowmik, M. Arifuzzaman, M. R. H. Mondal and M. S. Islam, "Bangla Text Sentiment Analysis Using Supervised Machine," *Atlantis Press B.V.,* Vols. 1(3-4), pp. 34-45, 2021.

[2] H. HUANG, A. A. ZAVAREH and MUMTAZBEGUMMUSTAFA, "Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and," *DOAJ,* vol. 11, p. 90367 – 90382, 2023.

[3] N. Tabassum and M. I. Khan, "Design an Empirical Framework for Sentiment Analysis from Bangla Text using Machine Learning," in *Electrical, Computer, and Communication Engineering (ECCE)*, Cox's Bazar, Bangladesh, 2019.

[4] A. Adak, B. Pradhan and N. Shukla, "Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review," *Foods,* vol. 11, no. 10, p. 1500, 2022.

[5] M. Wankhade, A. C. S. Rao and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review,* vol. 55, no. 7, pp. 5731-5780, 2022.

[6] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion,* vol. 58, pp. 82-115, 2020.

[7] N. R. Bhowmik, M. Arifuzzaman and M. R. H. Mondal, "Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms," *Array,* vol. 13, p. p. 100123, 2022.

[8] R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter and A. K. Das, " An Automated System of Sentiment Analysis from Bangla Text using Supervised," in *IEEE 4th International* Conference *on Computer and Communication Systems (ICCCS)*, 2019 .

[9] Y. Cao, Z. Sun, L. Li and W. Mo, "A study of sentiment analysis algorithms for agricultural product reviews based on improved BERT model," *Symmetry,* vol. 14, no. 8, p. 1604, 2022.

[10] Y. Abdelwahab, M. Kholief and AhmedAhmedHeshamSedky, "Justifying Arabic Text Sentiment Analysis Using Explainable AI (XAI): LASIK Surgeries Case Study," *Information,* vol. 13, no. 11, p. 536, 2022.

[11] M. Atabuzzaman, M. Shajalal, M. B. Baby and A. Boden, "Arabic Sentiment Analysis with Noisy Deep Explainable Model," 2023.

[12] N. S. RAO and C. SWATHI, "Weakly-Supervised Deep Learning for customer Review," Journal *of Engineering Sciences,* vol. 14, 2023.

[13] K. Barik, S. Misra, A. K. Ray and A. Bokolo, "LSTM-DGWO-Based Sentiment Analysis Framework for," *Computational Intelligence and Neuroscience,* vol. 2023, p. 19, 2023.

[14] R. Haque, N. Islam, M. Tasneem and A. K. Das, "Multi-class sentiment classification on Bengali social media comments using machine learning," *International Journal of Cognitive Computing and Engineering (Int. J. Cogn. Comput. Eng.),* vol. 4, pp. 21-35, 2023.

[15] S. S. Shanto, Z. Ahmed and A. I. Jony, "Mining User Opinions: A Balanced Bangla Sentiment Analysis Dataset for E-Commerce," *Malaysian Journal of Science and* Advanced *Technology*, pp. 272-279, 2023.

[16] M. Z. Hossain, M. A. Rahman, M. S. Islam and S. Kar, "Banfakenews: A dataset for detecting fake news in Bangla," *arXiv preprint arXiv:2004.08789* , 2020.

[17] M. A.-U.-Z. Ashik, S. Shovon and S. Haque, "Data set for sentiment analysis on Bengali news comments and its baseline evaluation," in *International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2019.

[18] M. H. Palash, P. P. Das and S. Haque, "Sentimental style transfer in text with multigenerative variational auto-encoder," in *International Conference on Bangla Speech and Language Processing*, 2019.

[19] Tripto, N. Irtiza and a. M. E. Ali, "Detecting multilabel sentiment and emotions from Bangla YouTube comments," in *International Conference on Bangla Speech and Language Processing*, 2018.

[20] K. Sarkar, "Sentiment polarity detection in Bengali tweets using LSTM recurrent neural networks," in *2019*, Second International Conference on Advanced Computational and Communication Paradigms (ICACCP).

[21] M. S. Islam, M. A. Islam, M. A. Hossain and J. J. Dey, "Supervised approach of sentimentality extraction from Bengali Facebook status," in *19th International Conference on Computer and Information Technology (ICCIT)*, 2016.

[22] R. Jain, A. Kumar, A. Nayyar, K. Dewan, R. Garg, S. Raman and S. Ganguly, " Explaining sentiment analysis results on social media," *Multimedia Tools and Applications*, pp. 1-17, 2023.

[23] A. PATEL, P. OZA and S. AGRAWAL, "Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model," *Procedia Computer Science* , vol. 218, pp. 2459-2467, 2023.

[24] "Daraz," [Online]. Available: https://www.daraz.com.bd/. [Accessed 2024 June 25].

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title: Interpretable Sentiment Analysis of Bangla E-commerce Comment Using Explainable AI Techniques.
Student ID: 203-15-14564

CO Description for FYDP

CO	CO Descriptions	PO
	Phase -I	
CO1	Integrate knowledge of natural language processing, machine learning, and XAI to identify and formulate the problem of sentiment analysis in Bangla e-commerce comments.	**PO1**
CO2	Analyze various aspects of sentiment analysis techniques and set clear, measurable goals for adapting these techniques to Bangla text data.	**PO2**
CO3	Conduct a comprehensive literature review of existing sentiment analysis methodologies, including traditional machine learning and deep learning approaches, and perform a comparative analysis to identify gaps and opportunities.	**PO4**
CO4	Evaluate the economic implications of implementing sentiment analysis in Bangla e-commerce, estimate the costs, and employ effective project management procedures to ensure timely and within-budget project completion.	**PO11**
	Phase -II	
CO5	Design and develop a robust sentiment analysis framework tailored to Bangla e-commerce comments, ensuring compliance with relevant health, safety, cultural, socioeconomic, and environmental factors.	**PO3**
CO6	Apply appropriate engineering and methodologies, including natural language processing, machine learning algorithms, and explainable AI techniques, to address the complex problem of sentiment analysis in Bangla text.	**PO5**
CO7	Analyze the societal, health, safety, legal, and cultural considerations of implementing sentiment analysis in the Bangla e-commerce sector, ensuring adherence to ethical principles and professional standards.	**PO6**
CO8	Evaluate the sustainability and environmental impact of the sentiment analysis framework, considering the long-term benefits and potential drawbacks for the e-commerce industry and society at large.	**PO7**
CO9	Implement ethical principles and adhere to professional standards in data collection, analysis, and interpretation, ensuring transparency and accountability throughout the research process.	**PO8**

CO10	Demonstrate effective teamwork and leadership skills, working collaboratively with peers, supervisors, and stakeholders to achieve project objectives.	**PO9**
CO11	Communicate effectively with the engineering community and broader society, presenting research findings, methodologies, and implications clearly through written reports and oral presentations.	**PO10**
CO12	Recognize the importance of lifelong learning and continuously seek to update and expand knowledge in the fields of natural language processing, machine learning, and e-commerce analytics, adapting to technological advancements and emerging trends.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	**K1** and **K3** were achieved by demonstrating my prior understanding of sentiment analysis, interpretive artificial intelligence (XAI) website design, machine learning, and deep learning derived from previous courses.	**Page no:** [1-2, 11-17] **Section:** [1.1, 3.2]
				I demonstrate proficiency in **K2**, particularly in the statistical analysis of my datasets and in the formal aspects of computer and information science. This expertise is demonstrated within the domain of statistical analysis, where I analyze dataset statistics and apply appropriate methods based on these results.	**Page no:** [21-24] **Section:** [4.2]
				K4, K5 and **K6,** acquired through special engineering, focuses on data collection, website design and model implementation, web scraping techniques to collect relevant information.	**Page no:** [12-13, 30-46] **Section:** [3.2, 5.2]

				In the related work section, I completed **K8**, where I identified the key findings, sources of information, results, and scope / limitations of existing research.	**Page no:** [7-9] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses EP-2 by recognizing the challenges of sentiment analysis in Bengali e-commerce comments, including the limitations of traditional sentiment analysis techniques and the complexity of adapting these techniques to Bengali text. through the comparative analysis of different sentiment analysis methods.	**Page no:** [4] **Section:** [1.5]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses EP-3 by subtly comparing the results of different sentiment analysis methods with a focus on evaluating their effectiveness in analyzing Bengali e-commerce comments. Machine learning is the significant solution to enhance sentiment analysis in Bangla text through experimentation and analysis, providing valuable insights into the most effective approach among multiple possible approaches.	**Page no:** [30-46] **Section:** [5.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	In the data collection phase, I enlisted the assistance of a Bangla language expert to validate the dataset. Since I'm not proficient in Bangla myself, their expertise ensured the accuracy and reliability of the collected data.	**Page no:** [12-13] **Section:** [3.2]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A

6.	**EP6: Extends of stakeholders involved and conflicting**	Yes	CO8	At the data collection stage, I enlisted the help of a Bengali language expert to verify the dataset. Since I am not proficient in Bangla myself, their expertise has ensured the accuracy and reliability of the data collected.	**Page no:** [12-13] **Section:** [3.2]
7.	**EP7: Interdependence**	Yes	CO5	The project's reliance on various interconnected sub-systems underscores the interdependence of its components in achieving the overarching goal of sentiment analysis in Bangla e-commerce. From data collection to model training, detection, and integration into a user-friendly interface, each sub-system plays a crucial role in the project's success. This interdependence highlights the need for seamless coordination and collaboration across different teams and domains throughout the project lifecycle.**EP-7**.	**Page no:** [12-13, 21-24, 30-46] **Section:** [3.2, 4.2, 5.2]

Addressing CO11 with Complex Engineering Activities (EA):

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	The project leverages a spectrum of resources encompassing computational infrastructure, deep learning frameworks, annotated datasets, and ethical guidelines. These resources play an instrumental role in conducting systematic research and driving innovation in Bangla text analysis.	**Page no:** [17-18] **Section:** [3.3]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project revolutionized the Bengali e-commerce sentiment analysis, promoting user engagement and trust. It promotes informed decision-making while ensuring social benefits and environmental sustainability through ethical information practices and resource-efficient computing.	**Page no:** [49-50] **Section:** [6.2, 6.3, 6.4]

5.	**EA-5: Familiarity**	Yes		This study, based on previous research, explores innovative methods in Bengali textual sentiment analysis, which is evident through early terminology and thorough comparative analysis, enriching understanding in this field.	**Page no:** [7-9] **Section:** [2.1, 2.2,2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project ensures efficient project management and financial control (CO4). It involves thorough planning, allocating resources wisely, and accurately estimating budgets for optimal resource usage throughout the research process.	**Page no:** [18-20] **Section:** [3.4]
2	CO9	Yes	This project demonstrates ethical principles (CO9) by safeguarding patient privacy, securing informed consent, and transparently documenting the research process. It ensures responsible dissemination of knowledge and societal well-being through ethical use of advanced healthcare technologies.	**Page no:** [50] **Section:** [6.3]
3	CO10	No	I Individually work on this project.	N/A
4	CO12	Yes	This project demonstrates commitment to ongoing learning by engaging with current research, refining methodologies, and embracing emerging trends in Bangla text analysis and e-commerce, fostering adaptability and growth within the field.	**Page no:** [11-18, 21-27, 30-46] **Section:** [3.2, 3.3, 4.2, 5.2]

Interpretable Sentiment Analysis of Bangla E-Commerce Comments Using Explainable AI Techniques

ORIGINALITY REPORT

16%	13%	8%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
2	Submitted to Daffodil International University Student Paper	1%
3	ieomsociety.org Internet Source	<1%
4	dropsofai.com Internet Source	<1%
5	arxiv.org Internet Source	<1%
6	Submitted to Higher Education Commission Pakistan Student Paper	<1%
7	machinelearningmodels.org Internet Source	<1%
8	www.mdpi.com Internet Source	<1%