

**OPTIMIZING CUSTOMER TRUST AND SATISFACTION IN E-COMMERCE WITH
MACHINE LEARNING TECHNIQUES**

BY

KAMAL UDDIN

ID: 0242220005103007

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

Supervised By

Dr. S. M. Aminul Haque

Professor & Associate Head

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JUNE 2024

APPROVAL

This Thesis titled “**Optimizing Customer Trust And Satisfaction In E-Commerce With Machine Learning Techniques**”, submitted by **Kamal Uddin**, ID No: 0242220005103007 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 **M.Sc. in Computer Science and Engineering** and approved as to its style and contents. The presentation has been held on 06-07-2024.

BOARD OF EXAMINERS



Chairman

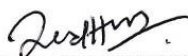
Dr. Sheak Rashed Haider Noori, PhD

Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



Internal Examiner

Dr. Md. Zahid Hasan, PhD

Associate Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



Internal Examiner

Dr. Arif Mahmud, PhD

Associate Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



External Examiner

Dr. Md. Zulfiker Mahmud

Professor

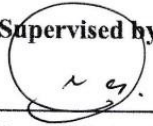
Department of Computer Science and Engineering

Jagannath University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. S. M. Aminul Haque, Professor & Associate Head, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



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

Submitted by:



Kamal Uddin
ID: 0242220005103007
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I would like to express my heartiest thanks and gratefulness to almighty Allah for His divine blessing makes us possible to complete the final year project successfully.

I really grateful and wish my profound our indebtedness to **Dr. S. M. Aminul Haque, Professor & Associate Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of *Natural Language Processing and Machine Learning* 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 stage have made it possible to complete this project.

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

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

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

ABSTRACT

Establishing trust is vital for sustainable success in a constantly changing e-commerce marketplace, particularly in locations such as Bangladesh where trust among customers is usually lacking. This research study focuses on assessing the sentiment of customer reviews in Bangladesh's e-commerce industry. It employs several Machine Learning approaches, with a primary emphasis on Natural Language Processing (NLP). The major purpose is to evaluate patterns and sentiments in Bengali review in order to give organizations with insights into customer preferences and enhance the quality of their offerings, ultimately establishing trust. This study investigates the usefulness of Natural Language Processing in managing a substantial number of Bengali e-commerce reviews, while taking into consideration particular linguistic and contextual distinctive characteristics. The study attempts to acquire key insights from consumer opinions in order to better products and services, enhance customer happiness, and achieve an advantage over others in the e-commerce industry of our country. Also, the anticipated outputs of the sentiment analysis could serve to boost the accomplishment and progress of Bangladesh's e-commerce business by overcoming potential barriers associated to data-driven initiatives that depend mostly on feedback from users. The study methodology is made up acquiring data from Kaggle and successfully preparing the text employing natural language processing techniques. To increase the statistical results of the models, the Decision Tree Classifier, K-nearest Neighbor approach, Random Forest Classifier, Logistic Regression, and Naive Bayes Classifier methods use unigram, bigram, and trigram data. The study produced an accuracy rate of 80.59% by applying the Logistic Regression algorithm by incorporating the trigram feature.

TABLE OF CONTENTS

CONTENTS	PAGE NO
Approval Page	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
Table of contents	v
List of Figures	vii
List of Figures	viii
 CHAPTER 1: INTRODUCTION	 1-5
1.1 Introduction	1
1.2 Motivation	2
1.3 Research Objective	3
1.4 Research Question	4
1.5 Report Layout	5
 CHAPTER 2: BACKGROUND	 6-10
2.1 Introduction	6
2.2 Literature Review	7
2.3 Scope of the Problem	9
2.4 Challenges	9
 CHAPTER 3: RESEARCH METHODOLOGY	 11-20
3.1 Introduction	11
3.2 Proposed Methodology	11
3.3 Dataset Description	12
3.4 Data Preprocessing	12

3.5 Machine Learning Models	15
3.6 Natural Language Processing	19
CHAPTER 4: RESULTS AND DISCUSSION	21-29
4.1 Introduction	21
4.2 Experimental Results	21
4.3 N-gram Language Models	22
4.4 Result Comparison of Machine Learning Approaches	22
4.5 Confusion Matrix	24
4.6 Bar Chart Comparison of Accuracy and F1 Score	27
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY	30-31
5.1 Impact on Society	30
5.2 Impact on Environment	30
5.3 Ethical Aspects	31
5.4 Sustainability Plan	31
CHAPTER 6: CONCLUSION AND FUTURE WORK	32
6.1 Conclusion	32
6.2 Future Scope	32
REFERENCES	33-34
PLAGARISM REPORT	35

LIST OF FIGURES

FIGURES	PAGE NO
Figure 1.1.1: Seles Prediction of Bangladesh e-commerce by 2026	2
Figure 2.1.1: Buying Decision Based on Number of Review Reading	6
Figure 3.1.1: Proposed diagram of the research method	11
Figure 3.4.1: Review dataset class distribution	13
Figure 3.4.2: Data pre-processing examples	13
Figure 3.4.3: Dataset Word Cloud from review words	15
Figure 3.5.1: Random Forest Workflow	16
Figure 3.5.2: Decision Tree Workflow	16
Figure 3.5.3: K Nearest Neighbor Workflow	18
Figure 3.5.4: Logistic Regression Workflow	19
Figure 3.6.1: Natural Language Processing Workflow	19
Figure. 4.5.1 Confusion Matrix of Decision Tree	25
Figure. 4.5.2 Confusion Matrix of Random Forest	25
Figure. 4.5.3 Confusion Matrix of Logistic Regression	26
Figure. 4.5.4 Confusion Matrix of Naïve Bayes	26
Figure. 4.5.5 Confusion Matrix of K-Nearest Neighbor	27
Figure. 4.6.1 Comparing result using Bar Chart (Unigram)	27
Figure. 4.6.2 Comparing result using Bar Chart (Bigram)	28
Figure. 4.6.3 Comparing result using Bar Chart (Trigram)	28

LIST OF TABLES

TABLES	PAGE NO
Table 3.4.1: Dataset Summarization with Example	14
Table 4.1.1: Algorithm Result Comparison (Unigram)	23
Table 4.2.2: Algorithm Result Comparison (Bigram)	23
Table 4.3.3: Algorithm Result Comparison (Trigram)	24

CHAPTER 1

INTRODUCTION

1.1 Introduction

In the twenty-first century, technology has been firmly incorporated in people's everyday lifestyle, with online shopping appearing as a significant feature of them revolution in technological aspect. In our country's perspective, the rapidly expanding e-commerce industry is experiencing a spike in customer interest, suggesting a radical shift in buying patterns. Despite this growth, maintaining confidence in the e-commerce field remains a critical problem for Bangladesh. Instances of fraud and unethical behavior by certain e-commerce firms have greatly affected customer confidence, causing to considerable financial losses [1].

Major e-commerce platforms such as Chaldal, Daraz, Pickaboo, and others which are currently dominate the Bangladeshi e-commerce industry [2]. However, many Bangladeshi e-commerce enterprises typically underestimate the necessary role that consumer feedback plays in improving and refining corporate operations. Neglecting the potential impact of the ecommerce customer reviews on client satisfaction and business growth limits the industry's growth prospects [26]. Bangladesh has already suffered fraud difficulties from well-known e-commerce enterprises as Evaly, E-orange, Qcoom, and Dalal Plus, among others [3]. To overcome these issues, e-commerce enterprises must need to focus issues like product quality, competitive price, brand marketing, social awareness, rapid customer service, and responsiveness to client wants and their choice. Building a solid relationship with customers through smooth service and high-quality items is vital for e-commerce success, as it drives repeat purchases.

The research study endeavors to navigate these challenges by analyzing and understanding customer evaluations in the Bangladeshi e-commerce business using contemporary Machine Learning techniques, particularly Natural Language Processing. The purpose is not only for research and current stage and attitudes indicated in Bengali ecommerce reviews but also to give significant data that can benefit businesses in understanding what customers want and generate improvements to how services are provided.

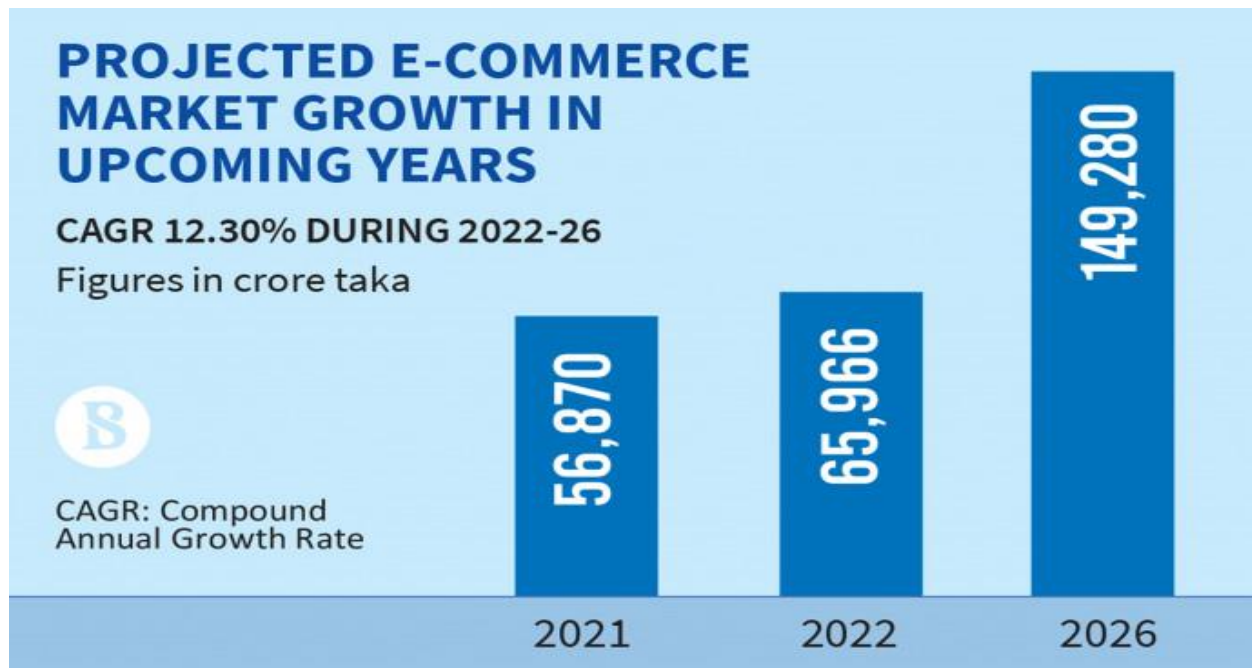


Figure 1.1.1: Sales Prediction of Bangladesh e-commerce by 2026 [4].

The figure above depicts the expected diversification of the Bangladeshi e-commerce business in the next years. This research is aimed at closing the gap between customer want and the merchandise given by online shops by employing user feedback evaluations. By doing so, it wants to accelerate the development of the manufacturing sector in our country's perspective. Essentially, the continued prosperity of the online shopping industry in the country relies on a consistent commitment to deliver outstanding products and services, connecting with clients swiftly, effectively and employing Machine Learning techniques to get a proper understanding of customer behaviors.

1.2 Motivation

The concept for this research comes up from the critical necessity to overcome the significant challenges facing Bangladesh's e-commerce business. The increasing enthusiasm of Bangladeshi consumers in online shopping has developed a promising yet potentially hazardous environment, strengthened by frequent issues such as lack of trust, fraudulent company operations, and the lack of effective customer-centric efforts all over the industry. Understanding that client trust is vital for the long-term success of e-commerce, this research endeavor was begun to analyze these problems in depth. The sector has suffered tremendous

financial failures and hampered progress due to increasing cases of fraud and lack of respect for client feedback. Thus, the deep grasp of customer perceives is vital as the business community grapples with these kinds of problems.

To solve these difficulties, this research provides a unique methodology for analyzing and evaluating customer feedback within the main domain, applying algorithms for machine learning such as NLP. The primary objective is to harness the strength of data-driven insights from Bengali customer feedback to empower and lead e-commerce businesses. By identifying the opportunities embedded in client feedback, organizations can adapt their strategies in order to improve service quality, develop trust, and build a more consumer-friendly e-commerce atmosphere. This research is driven by the possible ability of emerging technologies for reducing the gap among consumer needs and the actual situation of e-commerce in our country. The purpose underpinning this study is to effect positive change inside the industry, with targeted objectives include greater trust, greater satisfaction among consumers, and the advancement of sustainable growth for Bangladesh's online shopping sector. This research aspires not only to considerably enhance the country's e-commerce sector but also to deliver beneficial insights towards the larger research community.

1.3 Research Objective

This research study leverages modern Machine Learning methodologies in order to carefully address the challenges impacting the e-commerce industry in our country's perspective. The main objective is to employ sentiment analysis of Bengali customer reviews to extract practical knowledge that might encourage constructive changes within the commerce environment. The specific objectives of this research project are as follows:

Bengali Customer Reviews Sentiment Analysis: The present research investigates the most common opinions of Bengali customers when evaluating Bangladeshi e-commerce sites. The research aims to assess the various emotional complexities involved in the communication of concepts, offering both customers and businesses an in-depth comprehension of how this sentiment influence consumer decision-making of buying.

Effectiveness of NLP: This research study examines the proficiency of Natural Language Processing (NLP) methods in analyzing significant quantities of Bengali e-commerce

customer reviews. This study focuses on resolving challenges arising from contextual and linguistic anomalies in the Bengali language. The objective is to increase the reliability of sentiment analysis consequences, enabling a more precise comprehension of consumer opinions in the context of e-commerce of our country.

Influence of Customer Reviews: It examines the influence of customer feedback on trustworthiness within the e-commerce business in Bangladesh. Its objective is to discover approaches to mitigate the impact of negative feedback and enhance confidence among consumers.

Strategic Insights: The primary intent of this research is to boost overall customer pleasure and quality of service by producing vital insights from the opinions of customers. By employing data-driven strategies, businesses in Bangladesh's online market can earn an edge over their competition. The research also examines the impact of sentiment analysis conclusions on industry development, identifying weaknesses and suggesting fixes. In addition, it examines the importance of language differences and proposes refinements to machine learning algorithms for accurate sentiment categorization in Bengali e-commerce customer's feedback.

1.4 Research Questions

My research study aims to address the bellow pivotal questions:

- ✓ What types of Bangla reviews dominant the Bangladeshi e-commerce business, and how do they impact consumer buying decisions?
- ✓ Are NLP approaches efficiently analyzing a wide number of Bengali e-commerce reviews, incorporating linguistic nuances and ecological variations?
- ✓ How does buyer feedback effect trust and authenticity in Bangladesh's e-commerce the marketplace?
- ✓ What practical insights can Bangladeshi e-commerce enterprises extract from user feedback to improve merchandise?
- ✓ How could sentiment analysis improve Bangladesh's e-commerce development of cells and what hurdles can buyer feedback efforts confront?

1.5 Report Layout

The first chapter describes the research endeavor, addressing motivation, aims, questions, plan, and expected outcomes. The second chapter dives into the background research and investigation on e-commerce sentiment, evaluates prior investigations, discusses the issue's background, and discusses research barriers. The third chapter includes the research technique, gathering information methodologies, analysis of statistical information, numerous approaches, and feature implementations associated with this study. In Chapter Four, I have discussed the experiment evaluations, associated terms, and the graphical and statistical results of the research study, explaining the effects of employing machine learning strategies. The fifth chapter analyzes the study's societal ramifications and ecological consequences, underlining its importance and everlasting significance in the industry. In the Final Chapter, I summarized the entire research, suggested potential implementation of the research, and acknowledged the study's limitations.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

This thesis presents a detailed evaluation of distinctive research approaches dealing with crucial difficulties in Bangladesh's flourishing e-commerce industry. It integrates modern capabilities like Natural Language Processing for sentiment analysis to boost understanding and confidence among customers. This introduction illustrates the practical value of the chosen approaches, particularly the employment of Logistic Regression, Decision Tree Classifier, Naive Bayes, KNN, and Random Forest Classifier models, adjusted with unigram, bigram, and trigram feature. NLP takes a vital role in this investigation by providing superior data preparation and sentiment analysis from Bengali customer evaluations, aligning with the study's key objectives.

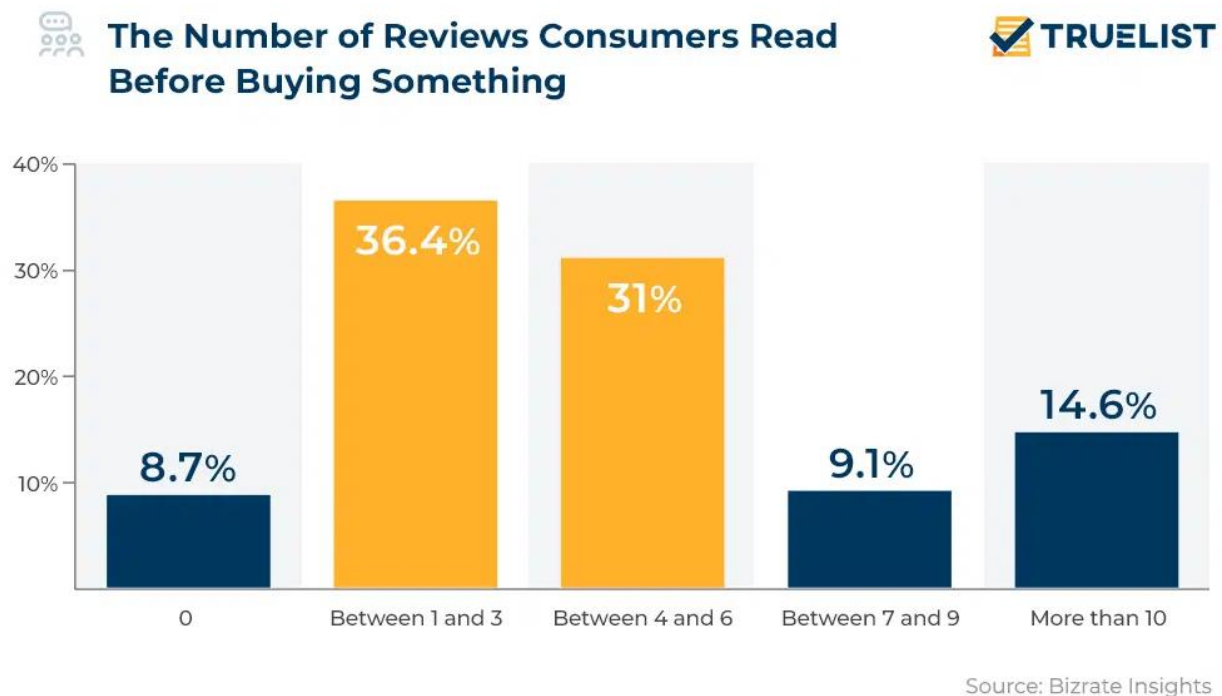


Figure 2.1.1: Buying Decision Based on Number of Review Reading [5].

Over 90% of worldwide consumers review reviews while purchasing things. Machine learning algorithms excel in text categorization, and analyzing feedback from consumers.

These ideas leverage Natural Language Processing for scalability, consistency, and objectivity, boosting e-commerce services and cultivating growth in our country's growing industry.

2.2 Literature Review

Shafin et al. [6] performed a research on Sentiment Analysis on Product Review by implementing the NLP approach employing Bangla ecommerce review dataset. They utilized different machine learning algorithms such Support Vector Machine, Decision Tree, Random Forest, K-Nearest Neighbors, and Logistic Regression to analyze feedback from site users during the COVID situation. The SVM model achieved an outstanding accuracy rate of 88.81%, overcoming other mentioned approaches.

Munna et al. [7] researched on sentiment analysis with evaluations of products categorization in online shopping business, particularly from a Bengali language perspective, using Natural Language Processing and Deep Neural Networks techniques. They accomplished 84% accuracy for the sentiment analysis category and 69% for the review classification, demonstrating the promise of advanced predictive machine learning algorithms in comprehending and categorizing feedback from customers in regional languages like Bengali.

Hossain et al. [8] did a study on sentiment analysis utilizing three language categories: English, Bangla, and Romanized Bangla. They employed 6 different machine learning approaches and grouped the data set into five sentiment categorizations. The Bengali dataset had the best accuracy using the SVM classification algorithm, while the Romanized Bangla and English datasets showed the highest accuracy using the Random Forest classifier. The study demonstrates the efficiency of different classifiers and the refined performance of machine learning approaches in sentiment evaluation.

Zulfiker et al. [9] performed an investigation on sentiment analysis in the Bangladeshi online shopping industry utilizing six diverse techniques of machine learning: Random Forest, Logistic Regression, Stochastic Gradient Descent, Decision Tree, Multinomial Naive Bayes, and Support Vector Machine. They have gathered dataset from prominent platforms and

employed various feature selection and optimization methodologies. The Support Vector Machine approach produced the highest rate of 90.68% accuracy.

Hossain et al. [10] suggested a machine-learning methodology to identify differences in opinion positive and negative in Bengali book reviews. They generated a dataset of 2000 reviews from textbook varied sources such as Facebook, online buying platforms, and blog sites. In the course of their study, they applied numerous approaches, including Logistic Regression, Support Vector Machine, Naive Bayes, and Stochastic Gradient Descent. Their findings suggested that the Multinomial Naive Bayes (MNB) approach had the maximum accuracy, with an achievement rate of 84%.

Soumik et al. [11] established a model for sentiment analysis by leveraging machine learning for sorting reviews of Bengali users Google Play Store into 3 unique categories. They acquired a substantial dataset encompassing 10,000 bangla reviews from various applications such as Uber, Daraz, and Bkash. Afterwards, they applied many machines learning methods, including SVM, Logistic Regression, and Naive Bayes, and Gradient Boosting. Among these instances, the gradient boosting approach achieved the highest accuracy, with a rate of 76.95%.

Rumman et al. [12] introduced a new machine learning model for analyzing sentiments communicated in Bengali film reviews. The dataset employed for this investigation was thoroughly assembled and classified from a range of social media comments, as well as publicly accessible postings and reviews. This model displayed outstanding efficacy, on the test set it obtained an excellent accuracy level of 88.90% with the SVM algorithm. Additionally, this model attained a significant accuracy of 82.42% by leveraging the LSTM network as well.

Khan et al. [13] Constructed a machine learning framework for processing and interpreting review data utilizing Support Vector Machine (SVM), Decision Tree algorithms, Naive Bayes. The algorithm efficiently classified feedback from consumers, with an excellent accuracy rate of 90.3% achieved with an airline dataset. This model proved beneficial for academics and consumers in the industry, with results outperforming the performance of other algorithms used during the research study.

2.3 Scope of the Problem

The principal focus of this research analyzes the substantial challenges inside Bangladesh's very quickly developing online shopping sector, underlining the requirement for new solutions to ensure long-term development. This research investigates sentiment influence on e-commerce firms, focusing on Bengali customer reviews. It generates important insights, enhances services, and contributes to Bangladesh's dynamic e-commerce business development. The information above illustrates that over 90% of consumers globally study and assess reviews before buying any kind of product. The chosen machine learning models perform in the categorization of texts, which is crucial for evaluating customer feedback. These approaches make use of the power of natural language processing to analyze significant Bengali reviews, enabling scalability, consistency, and objectivity. They offer a comprehensive framework for obtaining essential information, developing e-commerce services, and encouraging continuous growth in Bangladesh's strong e-commerce industry.

2.3 Challenges

During the research study employing Bengali review datasets collected online platform and processed using different machine learning approaches, many significant difficulties occurred that demanded emphasis:

Language Diversity in Bengali Customer Reviews: Bengali reviews have numerous language variations, including vocabulary, dialects, and writing techniques, considerably impact sentiment analysis productivity, necessitating powerful algorithms for successful addressing.

Data Quality and Diversity: The accurateness of machine learning models in Bengali e-commerce evaluations is impacted by the parameter availability and quality of datasets, which might be challenging to obtain.

Effective Feature Selection: Identifying key features, particularly in unigram, bigram, and trigram qualities, is crucial to improving the machine learning accuracy of models and reducing dimension barriers.

Selection and Optimization: Optimizing all machine learning methods for maximum efficacy is a hard task, especially when confronted with vast amounts of Bengali e-commerce review data.

Imbalanced Dataset: Machine learning models based on excessively positive and negative reviews can lead to biased training, requiring approaches like oversampling and under sampling to ensure perfectly balanced datasets.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Data analysis techniques include the procedures of acquiring, cleaning, and analyzing data using specialized software tools to obtain trustworthy information depending on unique data demands. In my research study, I intend to use text-to-text sentiment analysis methods to evaluate internet reviews, intending to get deeper insights into client attitudes and increase the entire customer experience. To generate dependable and ideal outcomes, I will utilize several machine learning approaches and characteristics. The research will contain multiple machine learning methods to boost accuracy for every utilized machine learning models.

3.2 Proposed Methodology

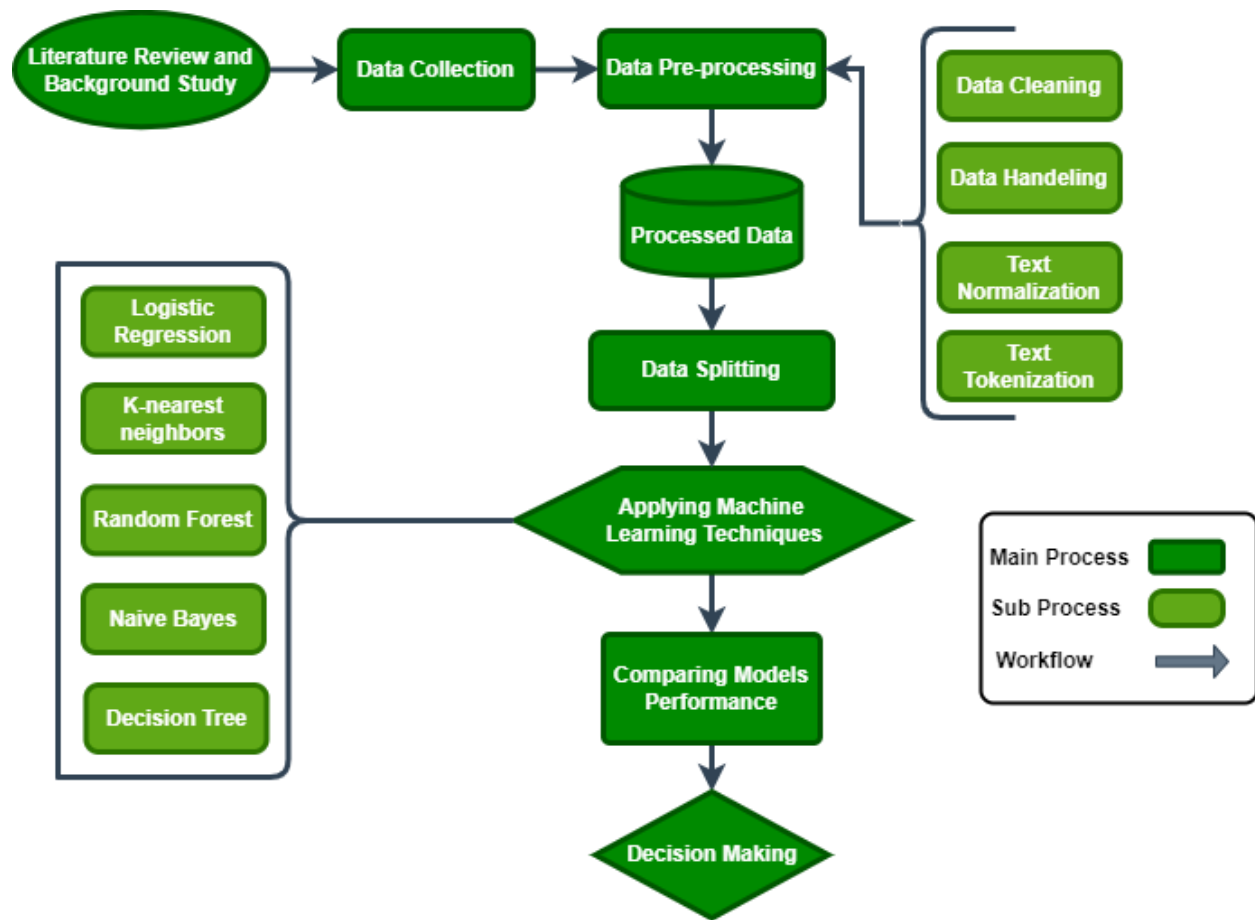


Figure 3.1.1: Proposed diagram of the research method.

3.3 Dataset Description

In this research work, the dataset employed for sentiment analysis of e-commerce reviews in our country's perspective was collected from the popular online source called Kaggle, a well-known platform for datasets. This dataset provides an immense number of e-commerce user reviews connected to various items and services offered on online shopping sites. The data collection procedure involves carefully scraping and curation to achieve a balanced representation of both positive and negative sentiments reviews. The dataset is structured into two primary columns: "Reviews" and "Category". The "Reviews" column provides the textual content of consumer reviews, expressing their opinions and experiences with e-commerce, comprising a total of 1205 entries. At the same time the "Category" column offers an understanding of data for machine learning algorithms by classifying all feedback into one of two categories: Positive or Negative. Which shows the overall attitude defined in each Bengali e-commerce review.

3.4 Data Pre-processing

During this study, I have prepared an e-commerce review dataset through Kaggle for machine learning study findings, that include text tokenization, cleaning, and vectorization. These precise plans boosts the dataset's compliance with the models, facilitating effective analysis of user sentiment. The cleaned and structured data serves as the starting point for training of the chosen machine learning strategies.

3.4.1 Data Handling

This is one of very first stage of data preprocessing involves deleting null entity from the acquired dataset. I have extensively analyzed the dataset for discovering any missing, absent, or incorrect values. During this cleaning process, I identified several columns having null entries and inaccurate data. Additionally, I filtered out brief reviews consisting of a single word, as they gave no meaningful information.

3.4.2 Data Balancing

This section shows the distribution of data classes within the dataset utilized for this research. Ensuring data equilibrium is vital to accomplishing reliable outcomes. As indicated

before, the Bangla feedback obtained from the e-commerce business site has been divided into two distinct classes: positive and negative. Out of a total of 1205 Bangla reviews, machine learning characteristics recognized and eliminated 24 short reviews, keeping a total of 1181 reviews. The distribution of these 1181 reviews is as follows:

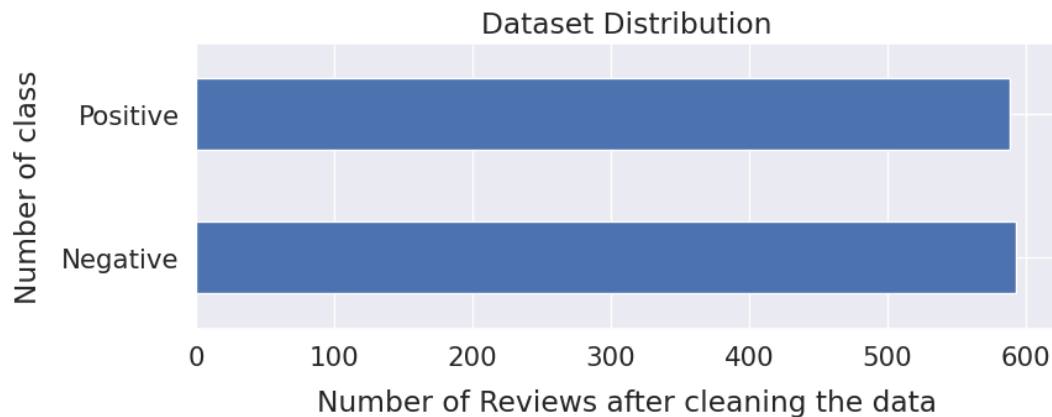


Figure 3.4.1: Review dataset class distribution.

The diagram above depicts the dataset's review classifications in the number of entity format. It is evident that both reviews are almost equal which reflects the perfect distribution resulting the best outcome from machine learning models.

3.4.3 Text Pre-processing

During that essential part of the research, the e-commerce review dataset collected from Kaggle receives comprehensive preprocessing in order to prepare it for machine learning analysis. Various necessary procedures are performed to make sure that the review dataset is clean and appropriate for use with machine learning models. In the beginning, text cleaning is conducted to erase unnecessary letters and done by utilizing stopwords.

Original Review: আজ ডেলিভারি করা হলো। আলহামদুলিল্লাহ, অনেক সুন্দর হয়েছে। Cleaned Review: ডেলিভারি হলো। আলহামদুলিল্লাহ সুন্দর হয়েছে। User Satisfaction:---> (Positive)	Original Review: এটা সত্যিই খারাপ Cleaned Review: সত্যিই খারাপ User Satisfaction:---> (Negative)
--	---

Figure 3.4.2: Data pre-processing examples.

The figure above depicts the text pre-processing applied to all two forms of review data. Initially, the dataset undergoes tokenization, a process that breaks down phrases into individual words. Following tokenization, the text is vectorized, translating the words into numerical vectors. These important processes increase the dataset's effectiveness, rendering it ideal for sentiment analysis. The modified the data set, now clean and well-structured, which will serve as the main source of data for the machine learning models.

3.4.4 Data Summarization

Data summarization is a fundamental mining technique of data where applied to extracted important insights from a set of data, supporting informed decision-making and tailored data display. Here in the below table, I will demonstrate the numbers of documents, identify positive and negative phrases, ascertain unique values, and highlight the top two distinctive words within the obtained data.

Table 3.4.1: Dataset Summarization with Example.

Dataset Summarization		
	Positive	Negative
Total Reviews	588	593
Total Words	3505	3706
Total Unique Words	1381	1607
Top two Frequent Words	ভালো, প্রোডাক্ট	অঁড়ার, ডেলিভারি

3.4.4 Word Cloud

The following figure illustrates the word cloud generated from the dataset words used throughout this research study.



Figure 3.4.3: Dataset Word Cloud from review words.

3.5 Machine Learning Models

For this research study, I employed five diverse Machine Learning methodologies. This section will detail the mechanisms and behavior of each Machine Learning approach individually.

3.5.1 Random Forest

The Random Forest technique is a well-known machine learning method that falls under the category of supervised learning approaches, which applied for addressing both regression and classification problems in the machine learning arena. This method exploits the ensemble learning principle by merging many classifiers to solve complicated problems, hence enhancing model performance and accuracy. Increasing the number of trees in the forest often boosts outcome which mitigates the risk of overfitting. Following graphic demonstrates the functioning of the algorithm:

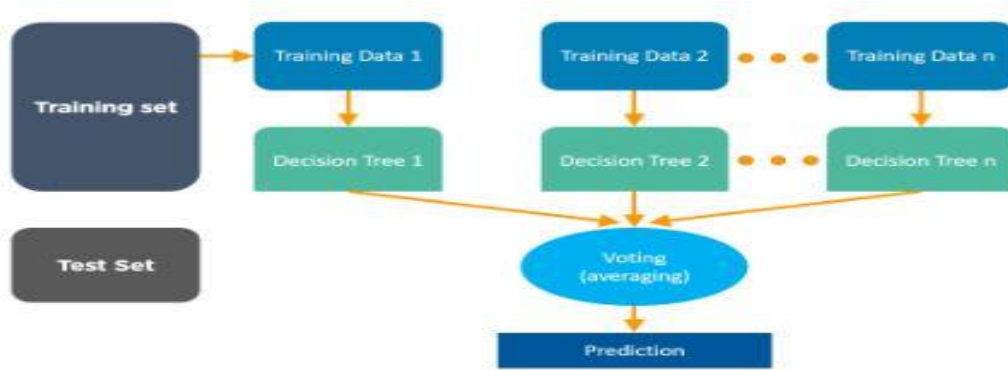


Figure 3.5.1: Random Forest Workflow [29].

The figure above depicts the working of the Random Forest algorithm. Recognized for its effectiveness in both regression and classification tasks, Random Forest is particularly adept at managing immense, high-dimensional datasets. It increases model accuracy and successfully mitigates the risk of overfitting. Despite these features, its implementation in regression problems has various boundaries, which might be considered a serious drawback.

3.5.2 Decision Tree

A decision tree is defined by some type of flowchart-like shape in which key nodes symbolize features, branches denote decision rules, and leaf nodes signify the outputs of the algorithm. This method serves as an efficient supervised learning strategy suitable to both regression and classification tasks. The basic aim is to discover the feature which optimizes information acquisition or minimizes impurity following the split.

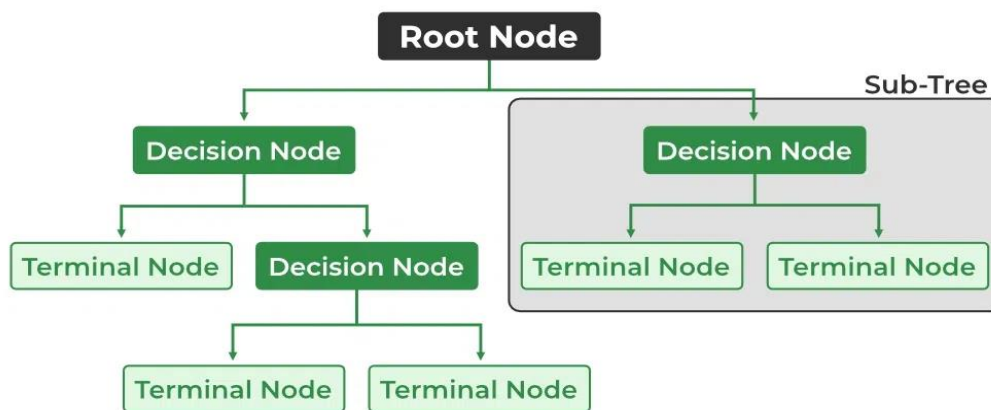


Figure 3.5.2: Decision Tree Workflow [30].

The corresponding Figure demonstrates the workflow accomplished by the Decision Tree algorithm, which classifies data by accessing dataset attributes through repeated compares at separate vertices. Despite its various layers, the decision tree replicates human decision-making, keeping clarity while implementing real-world decision-making procedures.

3.5.3 Multinomial Naive Bayes

Multinomial Naive Bayes is a widely utilized method for text classification, relying on statistical analysis of the content [31]. Bayes theorem states that, the probability $P(c|x)$ where x represents the provided data that needs to be categorized into distinct groups can be calculated. The Multinomial Naive Bayes equation is expressed as follows:

$$P(c|x) = P(x|c) * P(c) / P(x)$$

This study predominantly focuses on the application of NLP for predicting audio and text labels. The dataset employed in this research comprises two primary columns: user comments or reviews and their corresponding tags. The objective is to classify the emotions expressed in the reviews as negative or positive. To achieve this, the following probabilities need to be calculated:

$(P(\{\text{Positive}\} | \{\text{Review Dataset}\}))$: The probability that a review will be classified as positive when the term "Review Dataset" is present.

$(P(\{\text{Negative}\} | \{\text{Review Dataset}\}))$: The probability that a review will be classified as negative when the term "Review Dataset" is present.

It is necessary to preprocess the data through eliminating stopwords and applying filtering before deploying the algorithm to make sure optimal performance.

3.5.4 K-Nearest Neighbors

The K-Nearest Neighbors algorithm is a fundamental and effective technique for classification within the field of machine learning. This technique finds the label or value of a new data point by examining its closest K neighbors inside the training the data set,

applying the principle of similarity. Consider the following table of data items characterized by two unique features:

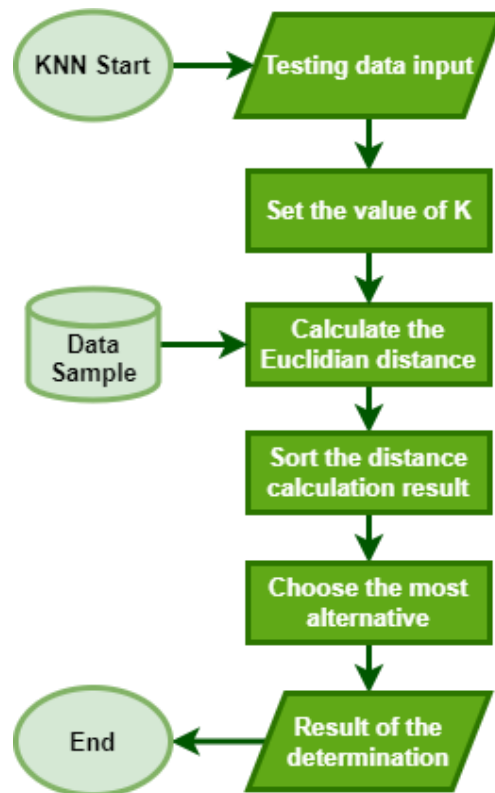


Figure 3.5.3: K-Nearest Neighbor Workflow [32].

The diagram above demonstrates the process of the algorithm known as KNN. The KNN method is relatively easy to implement because to its lesser complexity. However, it is important to establish correct data scalability before commencing the process of machine learning.

3.5.5 Logistic Regression

This is a machine learning approach under supervised learning, is used to estimate the probability that this particular instance belongs to a specific data category. This technique considerably enhances decision-making processes via its capabilities. By applying the sigmoid function, logistic regression transforms predicted values into probabilities [33]. My work is concerned with Logistic Regression, which may be divided into Multinomial, Binomial, and Ordinal varieties. This research mostly emphasizes

on Multinomial Logistic Regression, which provides two possible outcomes: Positive and Negative. The following figure reflects the functioning diagram of the model:

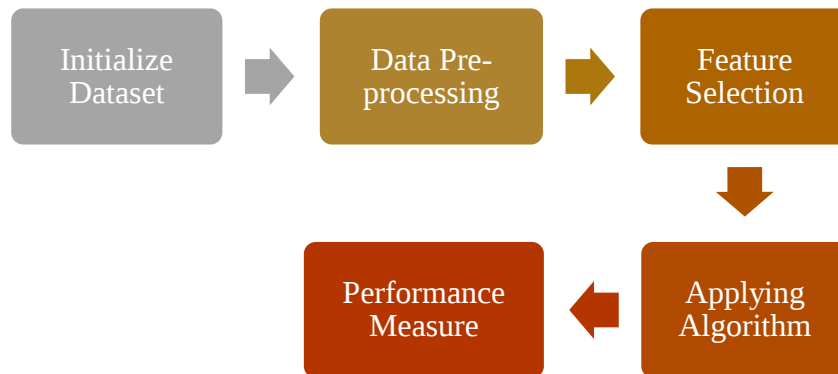


Figure 3.5.4: Logistic Regression Workflow.

Odds indicate the ratio of the probability of an event happening to it not happening, while probability reflects the ratio of an event happening to the total number of potential occurrences.

3.6 Natural Language Processing

Natural Language Processing, typically recognized as NLP, is the field within artificial intelligence that allows machines to comprehend and evaluate human languages. Each of these languages can be represented in text or audio formats. My research study particularly targets text-based reviews in Bangla e-commerce feedbacks. The process of NLP requires several phases, which are depicted in the figure that follows:

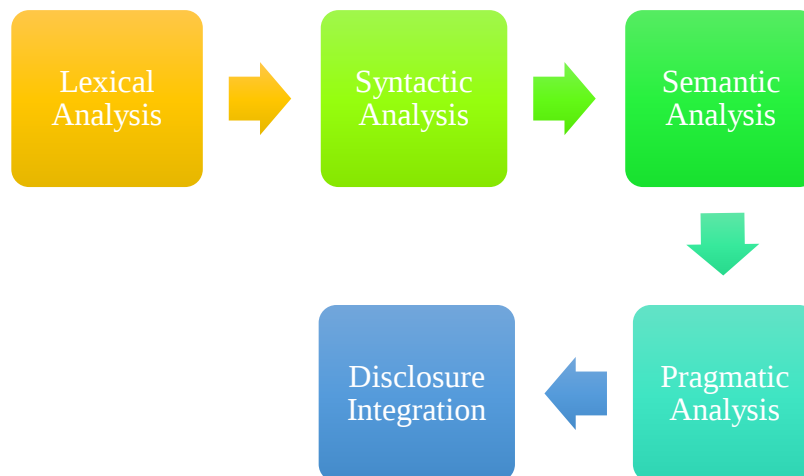


Figure 3.6.1: Natural Language Processing Workflow

Natural Language Processing is surprisingly rapid and effective, simplifying the analysis of data from both organized and unorganized data sources [34]. Initially developed for specific jobs, it now allows individuals to inquire about questions on any topic and receive rapid results within milliseconds.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, I will demonstrate the outcomes of applying various machine learning algorithms to a dataset of 1205 Bengali reviews, obtained from top e-commerce websites across the country. The well-curated Kaggle dataset covers varied consumer viewpoints, dividing reviews into Positive and Negative. We applied numerous machine learning methods, including Decision Tree, Random Forest, K-Nearest Neighbors, Multinomial Naive Bayes, and Logistic Regression, for sentiment classification. The analysis determines the efficacy of machine learning models for comprehending Bangla e-commerce reviews using measures including accuracy, precision, recall, and F1 score, underlining their strengths and areas for enhancement offering a full summary of their success. My work seeks towards providing an in-depth analysis of the models' capabilities in gathering and identifying distinct viewpoints within the dataset.

4.2 Experiment Result

In this section, I discuss the findings from analyzing a huge dataset of Bangla e-commerce customer evaluations, divided into positive and negative feelings, revealing numerous unexpected insights. Employing the Logistic Regression technique with Trigram features, the study produced a remarkable accuracy of 80.59%. This machine learning method requires diligent data management, balancing, and text preprocessing, all of which boosted the dataset's quality for sentiment analysis. Notably, Logistic Regression surpassed other algorithms in effectively classifying the reviews. The effectiveness of this strategy can be due to the collaborative learning principle, which efficiently integrates numerous classifiers to suit complex requirements. Leveraging Trigram characteristics substantially increased the algorithm's performance, catching more complicated linguistic patterns in Bangla e-commerce reviews. Logistic Regression accurately recognizes sentiment variations in Bangla e-commerce reviews, proving that it has the ability to manage unique language characteristics.

4.3 N-gram Language Models

An N-gram [35] is a sustained series of n things derived from a specific stretch of text or audio. The sequences are often extracted from a text or audio database. In my investigation, I used Unigrams, Bigrams, and Trigrams to assess their results and discovered that Trigrams had more precise results. To make things clear, I'll describe each of these 3 categories in the example following.

“আমি তাদের পরিষেবা নিয়ে অত্যন্ত সন্তুষ্ট”

Unigram: In the context of this research, an n-gram of size 1 is known as a Unigram. For instance, here is an example:

আমি	তাদের	পরিষেবা	নিয়ে	অত্যন্ত	সন্তুষ্ট
-----	-------	---------	-------	---------	----------

Bigram: A 2-gram, also known as a Bigram, consists of pairs of consecutive words. For instance, here is an example:

আমি তাদের	পরিষেবা নিয়ে	অত্যন্ত সন্তুষ্ট
-----------	---------------	------------------

Trigram: A trigram is an n-gram consisting of three elements. For instance, here is an example:

আমি তাদের পরিষেবা	নিয়ে অত্যন্ত সন্তুষ্ট
-------------------	------------------------

4.4 Result Comparison of Machine Learning Approaches

This part analyzes Machine Learning algorithms outputs employing each three n-gram feature types, evaluating and contrasting measures including Accuracy, F1 Score, Precision, and Recall. The highest number in each metric column is displayed along with the associated Machine Learning model name.

4.4.1 Unigram

Table 4.4.1: Algorithm Result Comparison (Unigram).

Unigram Feature				
	Accuracy	Precision	Recall	F1 Score
DT	73.42	82.62	73.42	71.36
RF	77.64	82.77	77.64	76.70
LR	78.48	79.07	78.48	78.36
MNB	78.48	78.82	78.48	78.41
KNN	67.51	70.73	67.51	66.15

The study examined the performance of multiple Machine Learning models, with the Logistic Regression model having the highest accuracy score of 78.48%. The Random Forest approach scored 82.77% in precision, whereas the Logistic Regression model got 78.48% for recall. The Naïve Bayes model generated the greatest F1-Score of 78.41%, proving the model's efficacy across multiple criteria for evaluation.

4.4.2 Bigram

Table 4.4.2: Algorithm Result Comparison (Bigram).

Bigram Feature				
	Accuracy	Precision	Recall	F1 Score
DT	73.42	82.62	73.42	71.36
RF	75.53	81.59	75.53	74.26
LR	79.32	80.46	79.32	79.12
MNB	78.06	78.25	78.06	78.02
KNN	62.87	74.62	62.87	57.74

The Logistic Regression model achieved an excellent accuracy of 79.32%, while the Decision Tree model obtained the highest precision of 82.62%. The Logistic Regression model similarly got the best recall score at 79.32% and the highest F1-Score at 79.12%. These findings indicate the efficiency of the Logistic Regression model across multiple performance variables, while also proving the remarkable precision of the Decision Tree approach.

4.4.3 Trigram

Table 4.4.3: Algorithm Result Comparison (Trigram).

Trigram Feature				
	Accuracy	Precision	Recall	F1 Score
DT	73.42	82.62	73.42	71.36
RF	76.37	81.52	76.37	75.34
LR	80.59	81.67	80.59	80.41
MNB	78.06	78.25	78.06	78.02
KNN	60.34	72.95	60.34	53.90

The analysis of Machine Learning algorithms determined that the Logistic Regression model had the maximum accuracy with a score of 80.59. The algorithm for Decision Trees similarly outscored others with a score of 82.62. The Logistic Regression model likewise succeeded in recall with an accuracy score of 80.59. The Random Forest algorithm showed greater performance with an F1-Score of 80.41.

4.5 Confusion Matrix

The confusion matrix is an essential machine learning tool for assessing classification algorithms efficiency, which includes True Positives, True Negatives, False Positives, and False Negatives. It assists in identifying crucial factors including accuracy, precision, recall, and F1 score, allowing fine-tuning for increased accuracy in predictions.

➤ **Decision Tree Confusion Matrix:**

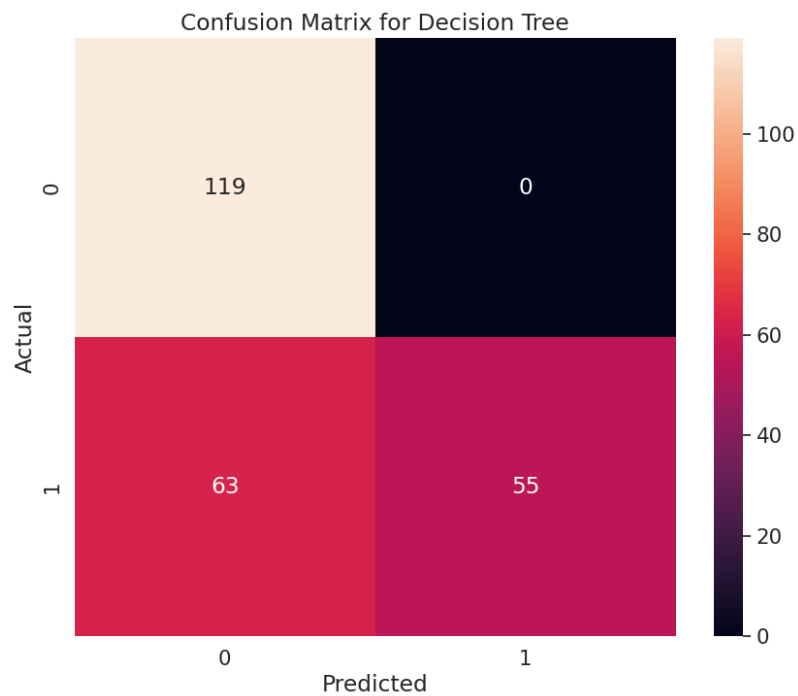


Figure 4.5.1: Confusion Matrix of Decision Tree.

➤ **Random Forest Confusion Matrix:**

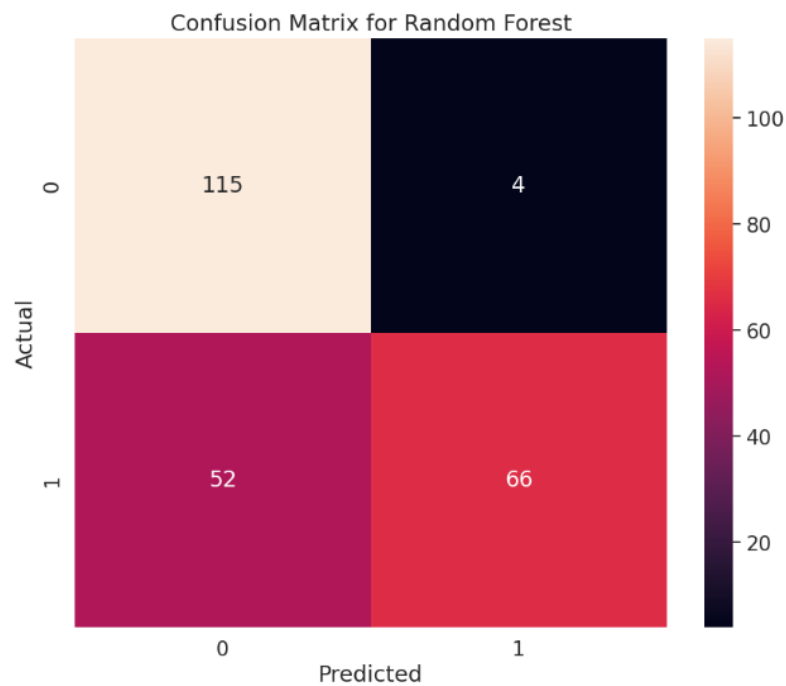


Figure 4.5.2: Confusion Matrix of Random Forest.

➤ **Logistic Regression Confusion Matrix:**

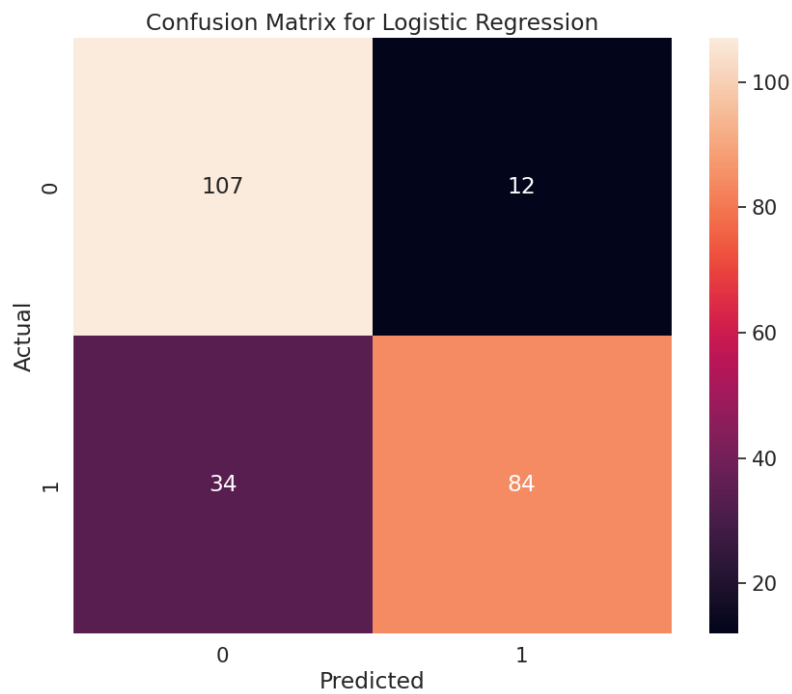


Figure 4.5.3: Confusion Matrix of Logistic Regression.

➤ **Naïve Bayes Confusion Matrix:**

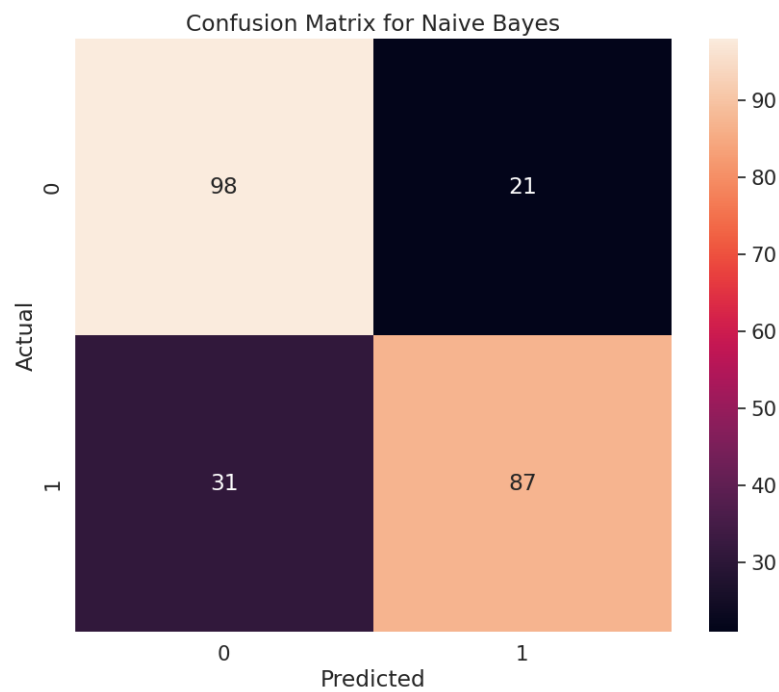


Figure 4.5.4: Confusion Matrix of Naïve Bayes.

➤ **K-Nearest Neighbor Confusion Matrix:**

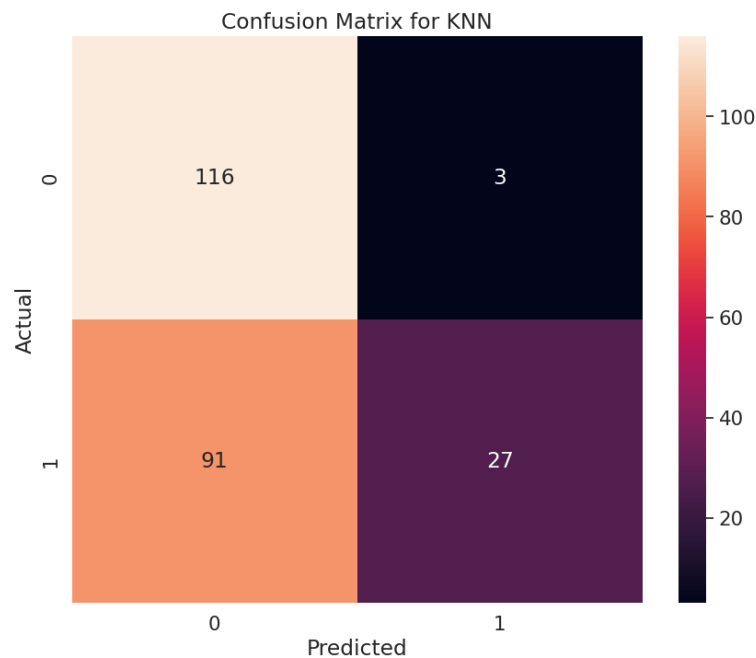


Figure 4.5.5: Confusion Matrix of K-Nearest Neighbor.

4.6 Bar Chart Comparison of Accuracy and F1 Score

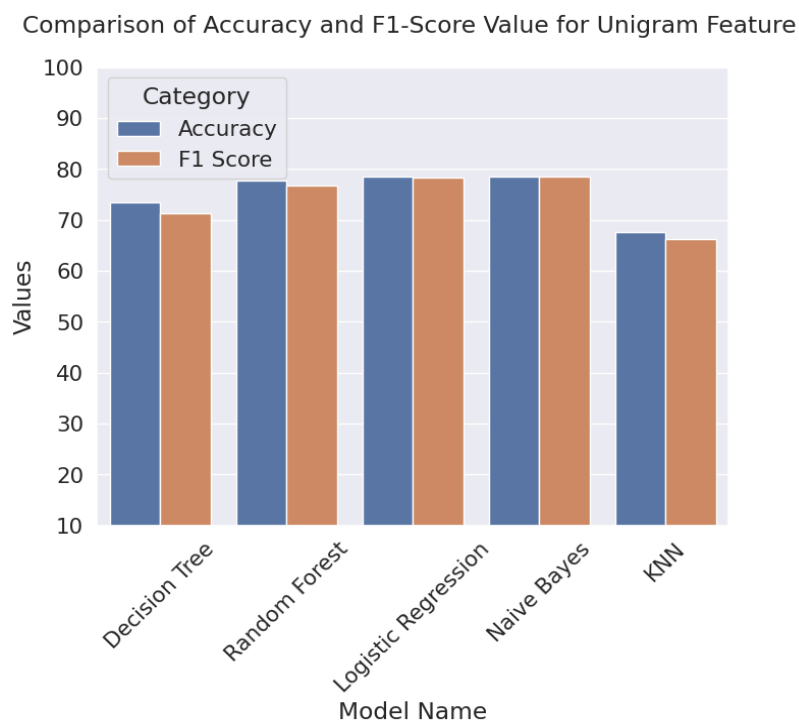


Figure 4.6.1: Result comparison using Unigram Feature.

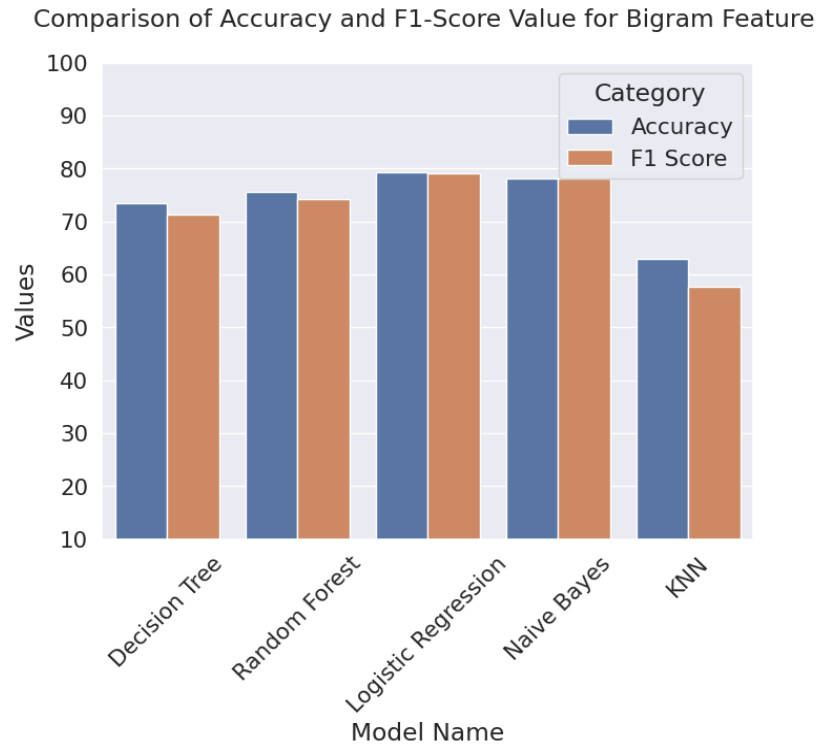


Figure 4.6.2: Result comparison using Bigram Feature.

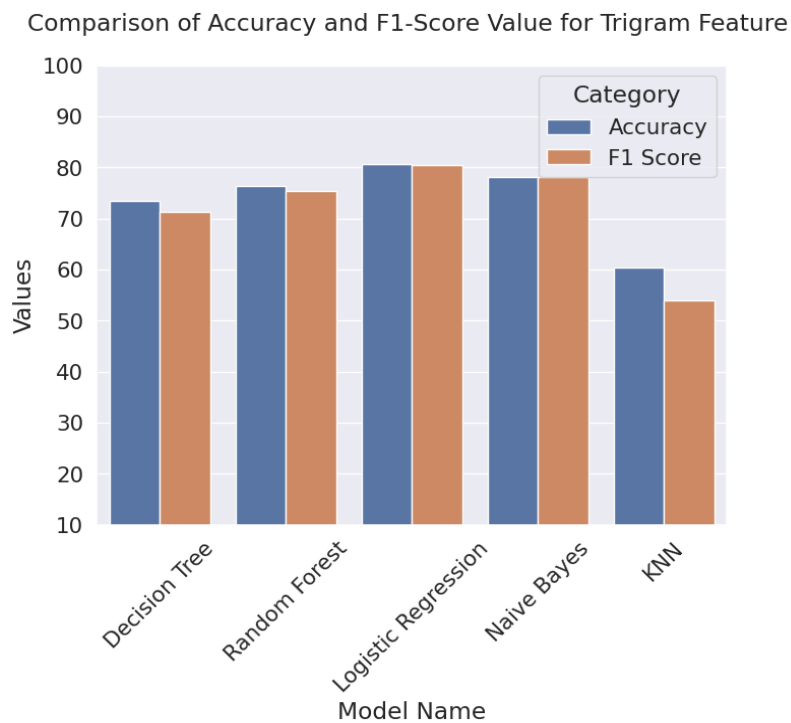


Figure 4.6.3: Result comparison using Trigram Feature.

The figures above show bar charts assessing Unigram, Bigram, and Trigram features performance in Machine Learning approaches. These visual comparisons offer an understandable overview of how each chosen model performs using distinct n-gram features, boosting the comprehension of the models.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

Here I have demonstrates the precise performance of five distinct machine learning model on a significant Bangla ecommerce review dataset, with the Logistic Regression approach accomplishing extraordinary accuracy.

Encouraging Customer Choices: Employing machine learning techniques to evaluate e-commerce evaluations is essential for developing informed consumer decisions in our country's perspective. Through sentiment analysis, consumers obtain vital insights into the experiences of other customers, enabling them to make informed decisions while engaging with e-commerce platforms. Consequently, this supports a need for higher-quality goods and services, contributing to a more sophisticated and empowered audience of the e-commerce sector.

Customer Satisfaction and Building Trust: Methods involving machine learning are being used to assess e-commerce evaluations in Bangladesh, improving consumer decision-making accuracy. This method accumulates user feedback, offering insights into other people's opinions, and fostering smarter online purchasing environment. This results in a knowledgeable customer community, and a growing need for quality products and services. Trust is essential in any marketplace, and machine learning analysis contributes to building and sustaining it.

Impact on Local Businesses: The findings of the research also affect local firms that engage in online shopping. By researching and interpreting feedback from consumers, small firms may improve their product offers to line more precisely with what customers want. This alignment fosters the establishment of local companies, supporting economic growth within our country's wider e-commerce sector.

5.2 Impact on Environment

The study's environmental consequences stem from the basic properties of e-commerce, especially the decline in physical sales. Enhancements in logistics and transportation,

prompted by machine learning insights, enhance supply chain activities, thereby lowering the environmental footprint of product delivery. This efficiency correlates with environmental desires, limiting the ecological effect of the e-commerce business. Additionally, switching to online purchase facilitates paperless transactions, eliminating the need for paper invoices and other documentation, consequently limiting paper use and preventing deforestation. Overall, the study discloses the potential for internet firms to apply eco-friendly practices and promote sustainable operations with machine-learning techniques.

5.3 Ethical Aspects

The proper implementation of machine learning for the examination of the e-commerce evaluates is the study's essential ethical challenge. It is necessary for maintaining the user privacy, security of data, and open knowledge of data management procedures. Ethical standards necessitate a neutral, fair its representation of a range of consumer input while limiting any chance of deception. The study effort is further examining how online shopping sites might ethically utilize machine learning observations, addressing moral conduct while using feedback from customers. These features support user confidence, protect equality of opportunity, and promote the moral evolution of machine learning algorithms within the e-commerce business environment.

5.4 Sustainability Plan

Sustainability strategy is essential for continuing beneficial effect, considering the projected implications of this study for the e-commerce business in Bangladesh. Creating an effective framework for handling data is vital for assuring the ethical employment of machine learning in sentiment analysis moving forward Integrity and transparency may be encouraged by working jointly with customer advocacy groups, government agencies, and e-commerce groups. Ethical application of technology will also be reinforced by promoting programs that increase understanding and offer skill development for people who participate in the e-commerce ecosystem. By ensuring that machine learning is in line with ethical principles and for the purpose of the community as a whole this sustainability strategy seeks to maintain the beneficial influence of machine learning upon the industry of e-commerce.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

In summary, this research discovered noteworthy discoveries with wide-ranging consequences by utilizing machine learning approaches to examine e-commerce reviews from a Bangladeshi perspective. This detailed sentiment analysis in reviews not only helps consumers to make informed selections but also dramatically impacts local business dynamics, consumer satisfaction, and trust-building within Bangladesh's online buying scene. Multiple machine learning algorithms examined to discover the most precise one. The study identified the Logistic Regression approach as particularly useful, obtaining a high accuracy rate of 80.59% when applying Trigram features. Additionally, the study reveals how Natural Language Processing (NLP) can efficiently extract emotions from any written input. This research also evaluated the effect of reviews on the e-commerce industry and discovered incidents of e-commerce fraud. The findings emphasize the revolutionary potential of machine learning in modifying the online shopping industry and impacting how consumers behave. Therefore, it is apparent that the e-commerce sector is likely supervised towards a more focused on customers, dependable, and innovative future.

6.2 Future Scope

The research findings provide the door for exciting new opportunities in research and development. In my study, I have focused on customer evaluation sentiment, dividing them into two primary types. Nonetheless, I feel the research might benefit from combining more precise information relevant to e-commerce and customer behavior along with adding more sentiment categories and subcategories. For this analysis, I made use of 1205 Bangla review data; raising the dataset size may increase the accuracy of my chosen models. Employing deep learning models may give a stronger understanding of sentiment and contextual details as well.

REFERENCES

- [1] O. Sharif, M. Hoque and E. Hossain, "Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naïve Bayes," in 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019.
- [2] R. ul Haque, P. Mehera, M. Mridha and M. Hamid, "A Complete Bengali Stop Word Detection Mechanism," in Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Spokane, WA, USA, 2019.
- [3] TBS, "Probes in e-commerce fraud cases move at snail's pace," [Online]. Available: <https://www.tbsnews.net/bangladesh/probes-e-commerce-fraud-cases-move-snails-pace-490538>. [Accessed 24 April 2024].
- [4] "Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism," in 4th MEC International Conference on Big Data and Smart City (ICBDSC), Muscat, Oman, 2019.
- [5] Branka, "Online Reviews Statistics – 2024," [Online]. Available: <https://truelist.co/blog/online-reviews-statistics/>. [Accessed 27 April 2024].
- [6] M. Shafin, M. Hasan, M. Alam, M. Mithu, A. Nur and M. Faruk, "Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language," in 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 2020.
- [7] M. Munna, M. Rifat and A. Badrudduza, "Sentiment Analysis and Product Review Classification in E-commerce Platform," in 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 2020.
- [8] M. Hossain, D. Joy, S. Das and R. Mustafa, "International Conference on Innovations in Science, Engineering and Technology (ICISSET)," in Sentiment Analysis on Reviews of E-commerce Sites Using Machine Learning Algorithms, Chittagong, Bangladesh, 2022.
- [9] S. Zulfiker, A. Chowdhury, D. Roy, S. Datta and S. Momen, "Bangla E-Commerce Sentiment Analysis Using Machine Learning Approach," in 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2022.
- [10] E. Hossain, E. Sharif and M. Hoque, "Sentiment Polarity Detection on Bengali Book Reviews Using Multinomial Naive Bayes," in Advances in Intelligent Systems and Computing (AISC), Singapore, 2021.
- [11] M. Jawad Soumik, S. Farhavi, F. Eva, T. Sinha and M. Alam, "Employing Machine Learning techniques on Sentiment Analysis of Google Play Store Bangla Reviews," in 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2019.
- [12] R. Das, M. Hossain, T. Ahmed, A. Devanath, S. Akter and A. Sattar, "Classification of Product Review Sentiment by NLP and Machine Learning," in Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Dhaka, Bangladesh, 2022.
- [13] X. Li, J. Zhang, Y. Du, J. Zhu, Y. Fan and X. Chen, "A Novel Deep Learning-based Sentiment Analysis Method Enhanced with Emojis in Microblog Social Networks," *Enterprise Information Systems*, vol. 17, no. 5, p. 2037160, 2023.
- [14] Li, X., Zhang, J., Du, Y., Zhu, J., Fan, Y., & Chen, X. (2022). A Novel Deep Learning-based Sentiment Analysis Method Enhanced with Emojis in Microblog Social Networks. *Enterprise Information Systems*, 17(5). <https://doi.org/10.1080/17517575.2022.2037160>
- [15] Chowdhury, R.R., Shahadat Hossain, M., Hossain, S., & Andersson, K. (2019). Analyzing Sentiment of Movie Reviews in Bangla by Applying Machine Learning Techniques. 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), 1-6.
- [16] N. Irtiza Tripto and M. Eunus Ali, "Detecting Multilabel Sentiment and Emotions from Bangla YouTube Comments," 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2018, pp. 1-6, doi: 10.1109/ICBSLP.2018.8554875
- [17] L. Nahar, Z. Sultana, N. Jahan and U. Jannat, "Filtering Bengali Political and Sports News of Social Media from Textual Information," 2019 1st International Conference on Advances in Science, Engineering

- and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934605
- [18] S. Phani, S. Lahiri, and A. Biswas, "Sentiment Analysis of Tweets in Three Indian Languages," ACLWeb, Dec. 01, 2016. <https://aclanthology.org/W16-3710/> (accessed Apr. 24, 2024)
- [19] A. P. Rodrigues and N. N. Chiplunkar, "Aspect Based Sentiment Analysis on Product Reviews," 2018 Fourteenth International Conference on Information Processing (ICINPRO), Bangalore, India, 2018, pp. 1-6, doi: 10.1109/ICINPRO43533.2018.9096796
- [20] M. A. Mirtalaie and O. K. Hussain, "Sentiment aggregation of targeted features by capturing their dependencies: Making sense from customer reviews," International Journal of Information Management, vol. 53, p. 102097, Aug. 2020, doi: <https://doi.org/10.1016/j.ijinfomgt.2020.102097>
- [21] Jitendra Kalyan Prathi, Pranith Kumar Raparathi, and M. Venu Gopalachari, "Real-Time Aspect-Based Sentiment Analysis on Consumer Reviews," pp. 801–810, Jan. 2020, doi: https://doi.org/10.1007/978-981-15-1097-7_67
- [22] S. Park, J. Cho, K. Park, and H. Shin, "Customer sentiment analysis with more sensibility," Engineering Applications of Artificial Intelligence, vol. 104, p. 104356, Sep. 2021, doi: <https://doi.org/10.1016/j.engappai.2021.104356>
- [23] J. Rana, L. Gaur and K. Santosh, "Classifying Customers' Journey from Online Reviews of Amazon Fresh via Sentiment Analysis and Topic Modelling," 2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 2022, pp. 1-6, doi: 10.1109/ICCAKM54721.2022.9990124
- [24] H. Shen, F. Namdarpour, and J. Lin, "Investigation of online grocery shopping and delivery preference before, during, and after COVID-19," Transportation Research Interdisciplinary Perspectives, vol. 14, p. 100580, Jun. 2022, doi: <https://doi.org/10.1016/j.trip.2022.100580>
- [25] K. Anshu, L. Gaur, and G. Singh, "Impact of Customer Experience on Attitude and Repurchase Intention in Online Grocery retailing: a Moderation Mechanism of Value Co-creation," Journal of Retailing and Consumer Services, vol. 64, no. 102798, Jan. 2022, doi: <https://doi.org/10.1016/j.jretconser.2021.102798>
- [26] S. Al-Natour and O. Turetken, "A comparative assessment of sentiment analysis and star ratings for consumer reviews," International Journal of Information Management, vol. 54, p. 102132, Oct. 2020, doi: <https://doi.org/10.1016/j.ijinfomgt.2020.102132>
- [27] S. Chowdhury and W. Chowdhury, "Performing sentiment analysis in Bangla microblog posts," 2014 International Conference on Informatics, Electronics & Vision (ICIEV), Dhaka, Bangladesh, 2014, pp. 1-6, doi: 10.1109/ICIEV.2014.6850712
- [28] S. Gohil, "Top 10 benefits of customer reviews in e-commerce," Meetanshi Blog - Learn Magento & Shopify, <https://meetanshi.com/blog/benefits-of-customer-reviews-in-e-commerce/> (accessed Apr. 20, 2024).
- [29] Machine learning random forest algorithm – javatpoint www.javatpoint.com. Available at: <https://www.javatpoint.com/machine-learning-random-forest-algorithm> (Accessed: 18 April 2024).
- [30] Decision Tree algorithm - GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/decision-tree/> (Accessed: 22 April 2024).
- [31] GeeksforGeeks - Applying multinomial naive Bayes to NLP problems, GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/applying-multinomial-naive-bayes-to-nlp-problems/> (Accessed: 23 April 2024).
- [32] K-Nearest Neighbor (KNN) algorithm - GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/k-nearest-neighbours/> (Accessed: 22 April 2024).
- [33] Logistic regression in machine learning - GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/understanding-logistic-regression/> (Accessed: 18 April 2024).
- [34] GeeksforGeeks - Natural language processing (NLP) tutorial, GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/natural-language-processing-nlp-tutorial/> (Accessed: 20 April 2024).
- [35] N-gram language modelling with NLTK - GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/n-gram-language-modelling-with-nltk/> (Accessed: 21 April 2024).

OPTIMIZING CUSTOMER TRUST AND SATISFACTION IN E-COMMERCE WITH MACHINE LEARNING TECHNIQUES

ORIGINALITY REPORT

15%

SIMILARITY INDEX

10%

INTERNET SOURCES

8%

PUBLICATIONS

7%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to University of Westminster

Student Paper

3%

2

[dspace.daffodilvarsity.edu.bd:8080](https://dspace.daffodilvarsity.edu.bd/8080)

Internet Source

3%

3

Submitted to Daffodil International University

Student Paper

2%

4

"Data Science and Security", Springer Science and Business Media LLC, 2024

Publication

<1%

5

Submitted to City University

Student Paper

<1%

6

quieora.ink

Internet Source

<1%

7

Mst. Tuhin Akter, Manoara Begum, Rashed Mustafa. "Bengali Sentiment Analysis of E-commerce Product Reviews using K-Nearest Neighbors", 2021 International Conference on Information and Communication Technology

<1%