

**A STUDY TO ANALYZE THE REVIEWS OF E-COMMERCE BUSINESS IN
BANGLADESH USING MACHINE LEARNING TECHNIQUES.**

BY

**MD. JANNAT-UL NAIM
ID: 0242220005103038**

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

Supervised By

Abdus Sattar
Assistant Professor & Coordinator MSc
Department of CSE
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JANUARY 2024**

APPROVAL

This Project titled **A Study to Analyze the reviews of e-commerce Business in Bangladesh using Machine**", submitted by Md. Jannat-UL Naim, ID No: 0242220005103038, 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 (MSc.) and approved as to its style and contents. The presentation was held on 27/01/2024.

BOARD OF EXAMINERS

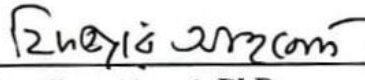


27.01.2024

Chairman

Professor Dr. Sheak Rashed Haider Noori
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



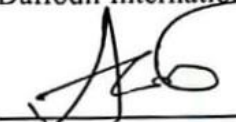
27.01.24

Internal Examiner

Dr. Fizar Ahmed, PhD

Associate Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



27.01.24

Internal Examiner

Dr. Arif Mahmud, PhD

Associate Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



27.01.24

External
Examiner

Dr. Firoz Ahmed, PhD

Professor

Department of Information & Communication Engineering
Rajshahi University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Abdus Sattar**, Assistant Professor & Coordinator MSc, 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:



Abdus Sattar
Assistant Professor & Coordinator MSc
Department of CSE
Daffodil International University

Submitted by:



Md. Jannat-UL-Naim
ID: 0242220005103038
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 **Abdus Sattar, Assistant Professor & Coordinator MSc**, 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

In the fast-changing landscape of e-commerce, establishing trust has become essential for sustainable development, particularly in regions like Bangladesh where customer confidence is poor. This study analyzes customer reviews in the Bangladeshi e-commerce industry using few of the Machine Learning approaches, especially utilizing Natural Language Processing (NLP). The primary focus is on identifying patterns and sentiments in Bengali reviews in order to educate businesses about consumer preferences as well as enhance services, ultimately building trust. This research work evaluates effectiveness of very well-known Machine Learning technique call natural language processing for processing a huge number of Bengali e-commerce reviews while considering specific linguistic and contextual differences. The study also intends to derive useful insights from consumer feedback analysis in order to improve products and services, enhance customer satisfaction, and achieve a competitive edge in the Bangladeshi e-commerce environment. Additionally, the analysis of sentiment findings is expected to be beneficial to the growth and development of Bangladesh's e-commerce sector, with an eye on possible challenges with implementing data-driven initiatives that depend on consumer feedback. The research methodology is based on collecting data from Kaggle and efficiently pre-processing text using Natural Language Processing. To gain more precise outcomes of my models, unigram, bi-gram, and trigram features will be integrated into the Linear Support Vector Machine, Kernal Support Vector Machine, Random Forest Classifier, Decision Tree Classifier, Naive Bayes Classifier and Logistic Regression techniques. The research study has achieved 90.76% accuracy using the Random Forest algorithm utilizing Bigram feature.

TABLE OF CONTENTS

CONTENTS	PAGE
Approval Page	II
Declaration	III
Acknowledgements	IV
Abstract	V
CHAPTER 1: INTRODUCTION	1-6
1.1 Introduction	1
1.2 Motivation	2
1.3 Research Objective	3
1.4 Research Question	4
1.5 Report Layout	5
1.6 Research Timeline	6
CHAPTER 2: BACKGROUND	7-11
2.1 Introduction	7
2.2 Related Works	8
2.3 Scope of the Problem	10
2.4 Challenges	11
CHAPTER 3: RESEARCH METHODOLOGY	12-22
3.1 Introduction	12
3.2 Proposed Methodology	12
3.3 Dataset Description	13
3.4 Data Pre-processing	13

3.4.1 Data Handling	13
3.4.2 Data Balancing	14
3.4.3 Text Pre-processing	15
3.4.4 Data Summarization	15
3.5 Machine Learning Models	16
3.5.1 Random Forest	16
3.5.2 Decision Tree	17
3.5.3 Multinomial Naive Bayes	18
3.5.4 Support Vector Machine	19
3.5.5 K-Nearest Neighbors	20
3.5.5 Logistic Regression	20
3.6 Natural Language Processing	21
3.8 Summary	22
 CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	 23-29
4.1 Introduction	23
4.2 Experimental Results	23
4.3 N-gram Language Models	24
4.4 Result Comparison of Machine Learning Approaches	25
4.4.1 Unigram Feature	25
4.4.2 Bigram Feature	26
4.4.3 Trigram Feature	26
4.5 Accuracy and F1 Score Comparison with Bar Chart	27
4.6 Conclusion	29

CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY	30-32
5.1 Impact on Society	30
5.2 Impact on Environment	31
5.3 Ethical Aspects	32
5.4 Sustainability Plan	32
CHAPTER 6: CONCLUSION AND FUTURE WORK	33-34
6.1 Conclusion	33
6.2 Future Scope	33
REFERENCES	35-36
ABBREVIATIONS	37
PLAGARISM REPORT	38

LIST OF FIGURES

FIGURES	PAGE NO
Figure 1.1: Bangladesh e-commerce sales to more than double by 2026	2
Figure 2.1: Number of reviews customer read before buying something	7
Figure 3.1: Proposed Methodology of the research study	12
Figure 3.2: Data distribution of each review class	14
Figure 3.3: Text cleaning and Pre-processing example of each class of Review	15
Figure 3.4: Data Summarization of the Dataset	17
Figure 3.5: Working procedure of Random Forest Classifier	18
Figure 3.6: Working procedure of Decision Tree Classifier	20
Figure 3.7: Working procedure of Support Vector Machine	21
Figure 3.8: Working procedure of KNN Classifier	21
Figure 3.9: Working procedure of Logistic Regression	21
Figure 3.10: Working procedure of Natural Language Processing	21
Figure. 4.1 Result comparison using Unigram Feature	27
Figure. 4.2 Result comparison using Bigram Feature	28
Figure. 4.3 Result comparison using Trigram Feature	28

LIST OF TABLES

TABLES	PAGE NO
Table 1.1: Research Timeline of the Research Work	12
Table 3.1: Data Summarization of the Dataset	14
Table 4.1: Result comparison using Unigram Feature	25
Table 4.2: Result comparison using Bigram Feature	26
Table 4.3: Result comparison using Trigram Feature	26

CHAPTER 1

INTRODUCTION

1.1 Introduction

In 21st century, dependence of individuals on technology have been entrenched to their everyday routines, with online shopping arising as a key component of this revolution in technology. In Bangladesh, the rapidly growing e-commerce sector is experiencing an increase in interest among consumers, indicating an unprecedented change in how people buy things. However, preserving confidence in the e-commerce domain remains a significant issue for Bangladesh, despite all of this progress. Certain e-commerce businesses [12] instances of fraud and unethical conduct have left a lasting impression on consumer confidence, resulting in significant financial losses.

A few well-known [11] e-commerce platforms are Daraz, Chaldal, Pickaboo, and others. They are currently dominating Bangladeshi e-commerce environment. Several Bangladeshi e-commerce businesses often disregard the most essential part that consumer feedback plays in building and strengthening business operations. The industry's potential for growth is hampered when the potential impact of customer reviews [14] on client satisfaction and business development is neglected. Bangladesh has already faced fraud [15] issues from few famous ecommerce including Evaly, E-orange, Qcoom, Dalal Plus etc. Setting priorities for aspects like product quality, affordable prices, brand publicity and recognition, rapid customer service, and sensitivity to customer requirements and desires are necessary for e-commerce businesses to succeed in overcoming these challenges. Establishing a relationship with customers through reliable assistance and high-quality products is vital for e-commerce success since it drives customers to buy again.

This research work intends to address these issues by analyzing and deciphering customer reviews in the Bangladeshi e-commerce market using modern Machine Learning approaches, especially Natural Language Processing (NLP). The objective is not only to analyze prevailing behaviors and attitudes in Bengali reviews but also to provide significant information that might recommend businesses about customer preferences and motivate service delivery improvements.

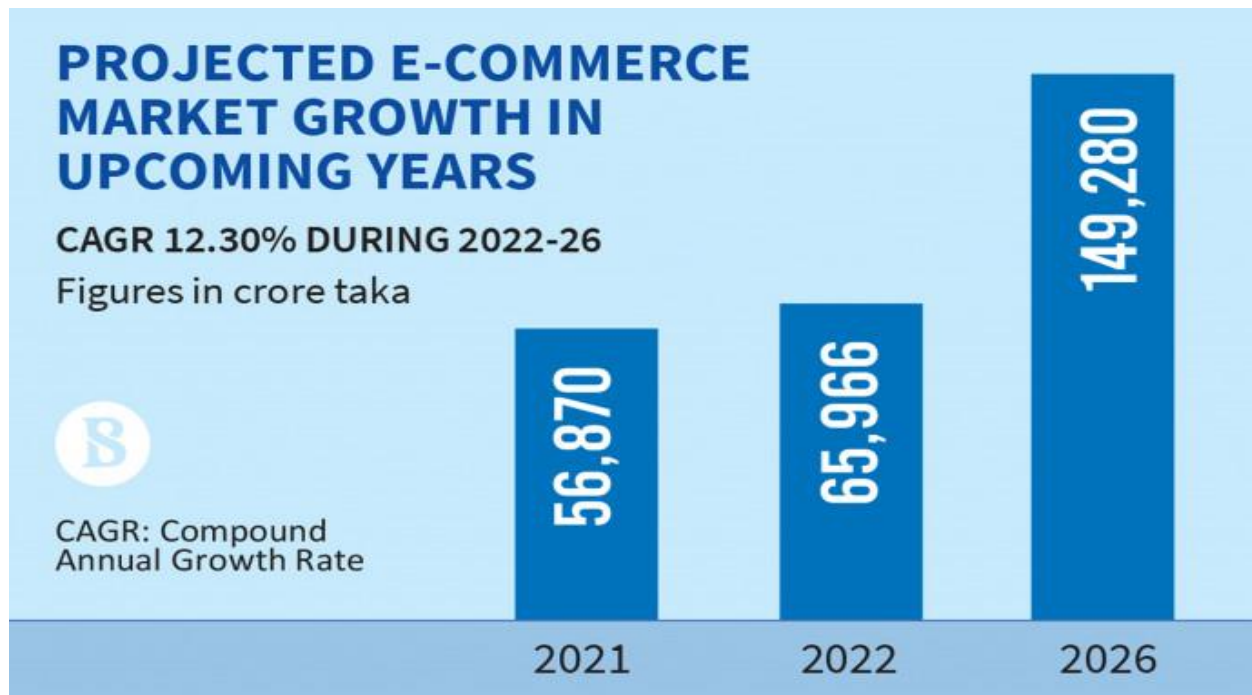


Figure 1.1: Bangladesh e-commerce sales to more than double by 2026 [\[13\]](#).

Above mentioned figure shows the predicted growth of Bangladeshi E-Commerce sector for upcoming years. The research presented here attempts to close the gap between what consumers want and online retailer goods through the use of on consumer feedback evaluation, hence improving the industry's growth in Bangladesh. In simple terms, the success of the country's e-commerce business depends on an ongoing commitment to quality services, on time customer engagement, and an effective utilization of Machine Learning approaches for an in-depth understanding of client behaviors.

1.2 Motivation

The research is driven by the necessity of addressing the significant challenges affecting Bangladesh's e-commerce industry. A beneficial but unsafe environment has been created by Bangladeshi consumers' increasing interest in online shopping in addition to persistent issues with a lack of trust, fraudulent company practices, and an absence of active customer-focused approaches in the industry as a whole. The Understanding that customer trust is crucial to the long-term sustainability of the e-commerce sector is what inspired researchers to begin this study. False incidences and a lack of respect for client feedback have led to significant financial losses in along with hampering the development of the sector. A comprehensive

comprehension of consumer perceptions is urgently needed as the business community suffers with these kinds of problems.

A novel approach for studying and evaluating customer reviews in the local environment has been introduced via the application of machine learning approaches, for example Natural Language Processing (NLP). The objective is to utilize the potential of data-driven knowledge gathered from Bengali customer reviews to provide guidance and information to e-commerce businesses. Organizations can adjust their approach in order to improve the level of service, establish trust, and ultimately establish an atmosphere for e-commerce that is more consumer-friendly by recognizing the possibilities found in feedback from customers. This research work is fueled by the prospect of using modern technologies in order to bridge the gap between customer expectations and the actual state of e-commerce in Bangladesh today. Guided by an ambition for the study to bring about constructive change in this industry, the study's intended outcomes are a growth of trust, the enhancement of customer satisfaction, and the encouragement of long-term progress for Bangladesh's e-commerce sector. This study's goal with this research work is to enhance this e-commerce industry of the country substantially as well as to add to the amount of knowledge in the research sector.

1.3 Research Objective

Using advanced Machine Learning approaches, the research identified in this work aims to systematically tackle all of the challenges facing Bangladeshi e-commerce sector. The primary objective is the proper use the analysis of Bengali reviews from customers to get practical information that would encourage positive changes in the commerce sector. The following are the specific study goals:

Sentiment Analysis of Bengali Customer Reviews: The present study investigates the common sentiments found in Bengali customer reviews of Bangladeshi e-commerce businesses. This research will be investigating all emotional tones used in the communication of concepts. Customers and the business alike need to understand how these affect choices made by customers.

Effectiveness of Natural Language Processing: This study explores Natural Language Processing's (NLP) ability to analyze a significant number of Bengali e-commerce reviews. It

solves issues that are caused by contextual and linguistic variances in Bengali, ensuring the validity of sentiment analysis results for an accurate understanding of consumer views in the e-commerce environment.

Influence of Customer Reviews on Business: Evaluate how feedback from consumers affects e-commerce trust in Bangladesh. Also, develop plans that mitigate negative feedback and increase customer confidence.

Strategic Insights Integration for E-commerce Excellence: The primary objective of this research study is to boost general satisfaction and service quality by generating useful knowledge from customer feedback. Businesses in Bangladesh's e-commerce market could establish an edge over their competitors by using data-driven strategies. It also analyzes the analysis of sentiment results for sectors growth contribution, bringing out problems and proposing solutions. In addition, the study investigates the effects of linguistic variances and proposes changes to machine learning approaches to gain accurate sentiment classification in Bangla e-commerce reviews.

1.4 Research Questions

This research study aims to answer the following questions as follows:

- ✓ What are the common type of Bangla reviews of Bangladeshi e-commerce businesses and how they are influencing customer choice on buying any product?
- ✓ Considering the challenges of language subtleties and contextual distinctions, can Natural Language Processing (NLP) approaches efficiently process and analyze an extensive amount of Bengali e-commerce reviews?
- ✓ How do feedback from consumers affect trust and credibility in Bangladesh's e-commerce marketplace, and what determines can businesses use to minimize disappointment and increase consumer trust?
- ✓ What insights can business entities get from studying consumer feedback in order to improve their products and services, increase customer satisfaction, and gain an advantage in the Bangladeshi e-commerce scenery?

- ✓ How might sentiment analysis results assist to the growth and development of Bangladesh's e-commerce industry, and what potential issues may occur when implementing initiatives that are based on the customer feedback?
- ✓ How do variations in language affect sentiment analysis accuracy in terms of Bangla review, and how can machine learning approaches handle the changes?

1.5 Report Layout

The following is a list of the chapters and the subsections that together make up this research work:

- ✓ The first chapter includes an introduction to the research work, research motivation, research aims, research questions, research schedule, and predicted research findings.
- ✓ The second chapter addresses the background study and related research works on e-commerce sentiment, as well as a study overview of previous studies done by other publishers, background of the issue, and then challenges during the research work.
- ✓ The third chapter discusses a number of topics for this study, including the methodology of this research project, the techniques used to gather data for the research, statistical analysis of the data groups, different approaches, and feature implementations associated with the study.
- ✓ In the fourth chapter, I will address the experiment evaluations along with related concepts, as well as the graphical and numerical outcomes of the research study by describing the results of using machine learning techniques.
- ✓ In the fifth chapter, I address the implications of this study for many sectors of society. How it affects many aspects of our environment. The rest of this chapter describes why the work being done is important and how it is going to persist in this industry.
- ✓ The sixth chapter I have presented an overview all of the work throughout this study. I also highlighted the research work in the future implementation and the limitations of my current study as well.

1.6 Research Timeline

SN.	Task Title	Task Short Description	Completed Date
1	Background Study	Research the study topic's background.	15 th Aug 2023
2	Literature Review	Studied several relevant study articles then analyzed their research work along with gaps.	1 st Sep 2023
3	Submit Project Proposal	Submitted the proposal to thesis supervisor.	15 th Sep 2023
5	Implementing the Machine Learning Techniques	Finding and working on a few machine learning approaches that are most suited for my Study	30 th Oct 2023
9	Writing the Full Report	Completed the writing of the full report.	30 th Nov 2023
10	Submitted Final Report	Submitted the final project report to my thesis supervisor.	18 th Dec 2023
11	Attended Seminar Defense	Attended Seminar Defense and presented my research work.	22 nd Dec 2023

Table 1.1: Research Timeline of the Research Work.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

This research study offers a comprehensive analysis of unique research approaches focused on addressing crucial difficulties in the constantly changing environment of Bangladesh's establishing e-commerce industry. The initiative integrates latest functionalities, including Natural Language Processing (NLP) for sentiment analysis, to enhance comprehension and increase consumer trust. This introduction emphasizes the practical relevance of particular research approaches, focusing on the application of Logistic Regression, Decision Tree Classifier, and Support Vector Machine Random Forest Classifier, models, each enhanced with 3 different features call unigram, bigram, and trigram. NLP, which is major part, enables excellent data preprocessing and emotion extraction from Bengali customer reviews, which corresponds to the study's fundamental objectives.

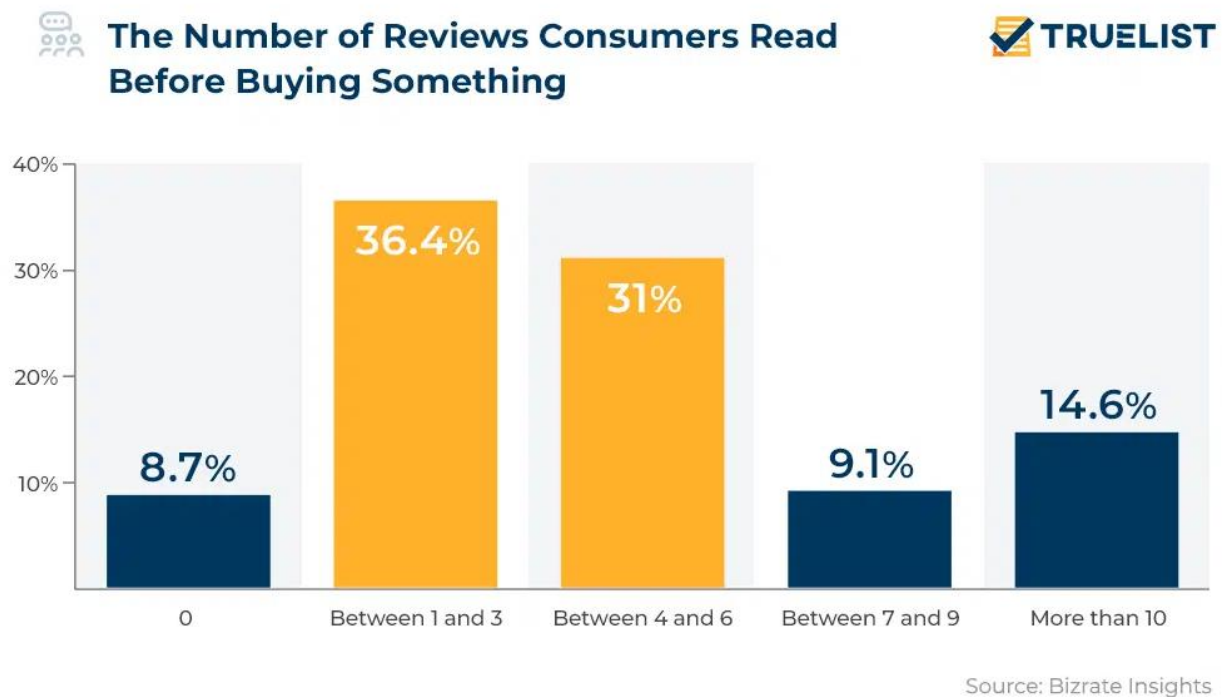


Figure 2.1: Number of reviews customer read before buying something [\[16\]](#).

This above-mentioned figure displays that almost 92% of the consumer worldwide check and analyze reviews before the buy any product. The chosen machine learning models are experts in categorization of texts, which is essential for understanding customer opinions. These approaches enable scalability, consistency, and impartiality by tackling the effectiveness of NLP in analyzing thorough reviews. They establish a strong foundation for extracting important information, maneuvers improvements in e-commerce services, and supporting sustained development in Bangladesh's vibrant e-commerce marketplace.

2.2 Related Work

Shafin et al. [1] offered a study on Product Review Sentiment Analysis by utilizing NLP Natural Language Processing in the Bangla review data, where they used a multiple machine learning classifiers including Decision Tree, KNN, Support Vector Machine, Logistic Regression, and Random Forest. They investigated this particular problem during a COVID circumstance and gathered 1000 customer review data from popular e-commerce sites. Among the models Support Vector Machine gain the highest accuracy of 88.81% among all machine learning techniques.

Munna et al. [2] presented Sentiment Analysis on the Product Review Categorization in an online shopping-based service on customer feedback in the Bangla language. They have focused on two domains: sentiment analysis also the review classification. They employed Deep Neural Network (DNN) and Natural Language Processing (NLP), and their accuracy rates were 84% and 69%, respectively.

Hossain et al. [3] proposed a study on sentiment analysis by analyzing three separate categories of languages, including Bangla, English, and Romanized Bangla, and utilizing 6 different machine learning techniques. They applied a different dataset, and the dataset columns have been classified into five different categories: Negative, Positive, Slightly Positive, Slightly Negative, and Normal. They gain highest accuracy with SVM for the on Bangla dataset, and the highest accuracy with the Random Forest classifier for the other two datasets, English and Romanized Bangla, respectively.

Das et al. [4] proposed an approach with machine learning that uses Natural Language Processing to classify also analyze sentiment in products with Bangla review dataset. They

employed five different machine learning algorithms for evaluating 6000 product reviews, including Decision Tree, Random Forest, Support Vector Machine, KNN, and Logistic Regression. With the Support Vector Machine algorithm, they achieved the highest accuracy of 94.78%.

Zulfiker et al. [5] proposed a research study on Bangladeshi E-Commerce Sentiment Analysis with different Machine Learning Approaches. They have applied six different machine learning approaches, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Stochastic Gradient Descent (SGD), Multinomial Naive Bayes (MNB), and Support Vector Machine (SVM), to focus on top Bangladeshi e-commerce platforms. They implemented various features and optimization strategies on their review dataset and achieved a highest accuracy of 90.68% using the support vector Machine approach.

Hossain et al. [6] introduced a machine-learning strategy to discover two different types of sentiment polarity for Bengali book reviews: positive and negative. A combined collection of 2000 textbook reviews have been gathered from various sources for example Facebook, Blogs sites and online shopping platforms. In the research they conducted, they utilized the SVM, Naive Bayes, Logistic Regression, and SGD algorithms, and concluded that Multinomial Naive Bayes had the highest level of accuracy of 84%.

Soumik et al. [7] proposed A sentiment analysis model based on machine learning was applied for classifying reviews of Bengali-language Google Play applications with three different categories. They gathered a massive number of 10,000 reviews from different applications including Daraz, Bkash, Ubar etc. After that they applied machine learning techniques, including Logistic Regression, Naive Bayes, Gradient Boosting Accuracy Support Vector Machine, as well as the ABP. They achieved the highest accuracy with the gradient boosting accuracy with 76.95%.

Zhang et al. [8] applied Support Vector Machine (SVM) to analyze sentiment in smartphone reviews from customers. In order to represent the review texts, they have focused on feature selection and extraction strategies such as n-gram features and term frequency-inverse document frequency (TF-IDF). Their findings from experiments revealed SVM's efficiency in sentiment classification, with high precision and F1 scores. Other classification techniques,

besides Support Vector Machine (SVM), have been explored for sentiment analysis of smartphones based on reviews.

Khan et al. [9] established an machine learning approach that included collecting review data, cleaning the data, and extracting features. They utilized Naive Bayes Support Vector Machine, and Decision Tree (DT) algorithms to classify customers opinion. This model assisted researchers, others relevants simultaneously. When the model was examined using the airline's dataset, then the outcome proved that Support Vector Machine gain 90.3% accuracy, which is significantly higher than other utilized approaches.

Rumman et al. [10] proposed a machine learning approach for assessing the sentiments expressed in Bengali film reviews. The dataset for this test was manually gathered and categorized from different social media comments also publicly available post and reviews. This model achieved the highest accuracy of 88.90% on the test set using the Support Vector Machine algorithm, then the highest accuracy of 82.42% using the long short-term memory (LSTM).

2.3 Scope of the Problem

The main scope of this study is defined by the pressing challenges within Bangladesh's dynamic e-commerce industry, necessitating innovative solutions for sustained growth. With a focus on enhancing consumer trust, the study integrates advanced technologies as example Natural Language Processing for sentiment analysis. The chosen research methodologies, including Decision Tree Classifier, and Support Vector Machine (SVM), Logistic Regression, Random Forest Classifier, enriched with unigram, bigram, and trigram features, align with the study's core objectives. The main theme of the research study is the sentiment analysis and its impact on the ecommerce business based on Bangla review from the customers. The exploration of related works underscores the significance of these methodologies, showcasing their scalability, consistency, and objectivity. The research's scope extends to extracting valuable insights, catalyzing improvements in e-commerce services, and contributing to the continuous development of Bangladesh's vibrant e-commerce landscape.

2.3 Challenges

While performing this research using Bangla review datasets from online source and processing them with various machine learning techniques, I discovered several challenges that need to be addressed:

Language Variations in Bengali Customer Reviews: The diverse language environment in Bengali reviews is a considerable difficulty. Variations in languages, words, and writing styles can affect sentiment analysis efficiency, requiring strong algorithms to account for linguistic variations.

Quality of the Data and Diversification: The availability and quality of different datasets for Bengali e-commerce reviews can fluctuate, improving machine learning model strength. Obtaining representative and impartial data becomes challenging, reducing the conclusions' generalizability.

Effective Feature Selection: Identifying significant features, particularly when dealing with the context of unigram, bigram, and trigram properties, needs thoughtful consideration. This thing plays an essential role for the improvement of machine learning model accuracy and conquering other dimensionality restriction.

Selection and Optimization of Machine Learning Techniques: Selecting the most appropriate machine learning algorithms and optimizing them for their best performance is a challenging process. Balancing the accuracy, speed, and scalability is challenging, particularly when working with large amounts of e-commerce review data.

Imbalanced Dataset: The dataset used to train all the Machine Learning models may be imbalanced, with a large number of positive, negative, and neutral reviews. Class imbalance can lead to biased model training and hamper the model's ability to categorize sentiment precisely. Techniques such as oversampling and under sampling may be utilized to deal with this issue and balance the dataset properly.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Data analysis techniques involve gathering data, cleaning data, using software tools for data analysis, and retrieving accurate information in accordance with the given data requirements. In my investigation, I intend to apply text analysis research approaches to evaluate online reviews in order to better understand what customers sentiment and improve the experience of customers. Various machine learning approaches and features will be used to get a reliable and optimized output for the study work. It will use various algorithms for machine learning to enhance accuracy and evaluate the outcomes achieved from each model.

3.2 Proposed Methodology

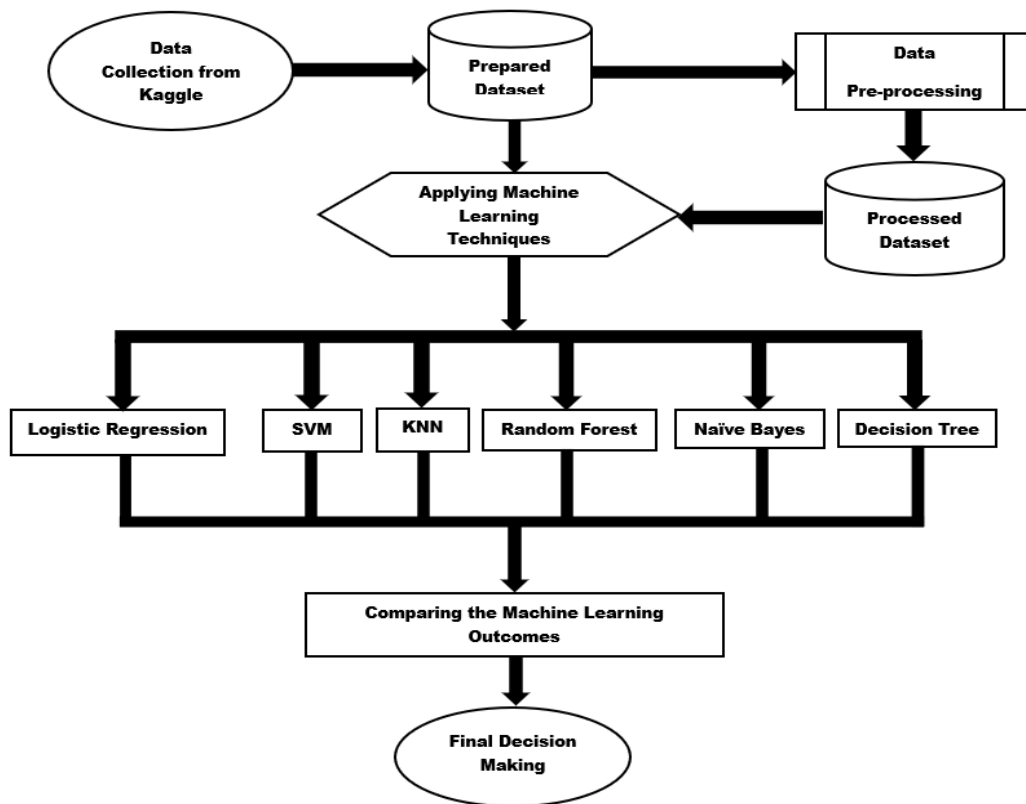


Figure 3.1: Proposed Methodology of the research study.

3.3 Dataset Description

In this research, the dataset utilized for the analysis of e-commerce reviews in Bangladesh was sourced from Kaggle, a renowned online platform for datasets. The dataset encompasses a diverse array of customer reviews pertaining to various products and services available on e-commerce platforms. The data collection process involved systematic scraping and curation, ensuring the inclusion of opinions spanning both positive and negative sentiments. The dataset is structured with two key columns: "Comment" and "Tag." The "Comment" column contains textual representations of customer reviews, expressing sentiments and opinions regarding their e-commerce experiences. Concurrently, the "Tag" column serves as the ground truth for machine learning models, categorizing each comment into one of three classes: 'Good,' 'Bad,' or 'Neutral.' This classification reflects the overall sentiment conveyed in the respective review.

3.4 Data Pre-processing

In this section, I have refined Kaggle-sourced e-commerce review dataset for optimal machine learning analysis. This involves essential steps such as text cleaning, removing extraneous characters and stopwords (17), tokenization for breaking down sentences, and vectorization to convert text into numerical vectors. These processes enhance the dataset's compatibility with machine learning algorithms, ensuring effective sentiment analysis. The transformed dataset, cleaned and structured, forms the basis for training our models.

3.4.1 Data Handling

This is the beginning phase of pre-processing, in which the null values are removed from the data being collected. I investigated my data set to see if there were any null values, values that were absent, or unsuitable values. I obtained a few columns that were null and had incorrect values while cleaning the data sets. I also removed those tiny reviews that were only one word which made no sense.

3.4.2 Data Balancing

This section demonstrates the distribution of each data class in the dataset that I have applied for the research work. One of the most significant factors in achieving a reliable outcome is data balancing. As previously mentioned, that our Bangla review data of ecommerce business site is classified into three distinct groups: good, bad, and neutral. Considering a from total of 3280 Bangla review data machine learning features has removed 524 small reviews, remaining 2756 review data are distributed as follows:

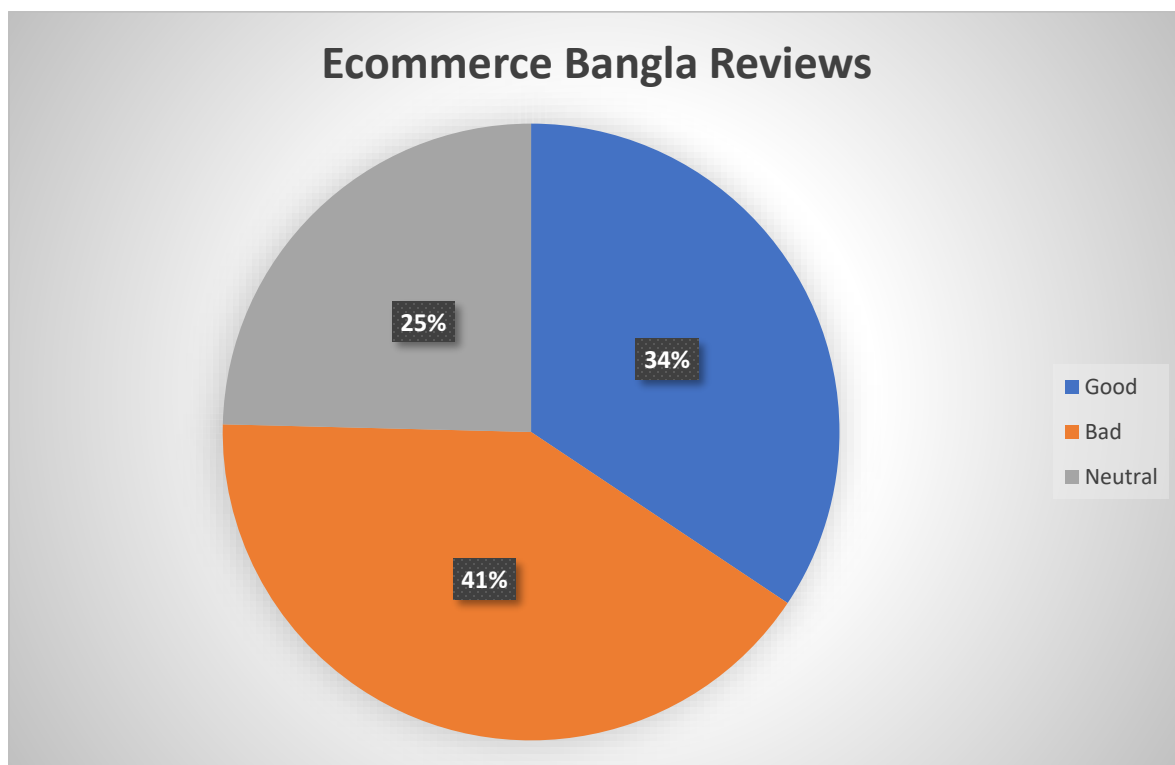


Figure 3.2: Data distribution of each review class.

The diagram above presents the dataset review classes in percentage format, and we can clearly see that Neutral reviews are slightly lower as compared to the other two balanced classes. Due to the fact that most reviews in ecommerce are categorized as positive or negative, the Neutral class encounters fewer reviews than the other categories.

3.4.3 Text Pre-processing

The Kaggle-sourced e-commerce review dataset is carefully pre-processed for machine learning analysis during this essential phase of the study approach. Several essential techniques are used to guarantee that the dataset is clean and appropriate with machine learning algorithms. In the beginning, text cleaning is performed, with the objective of removing unnecessary characters and stopwords [17].

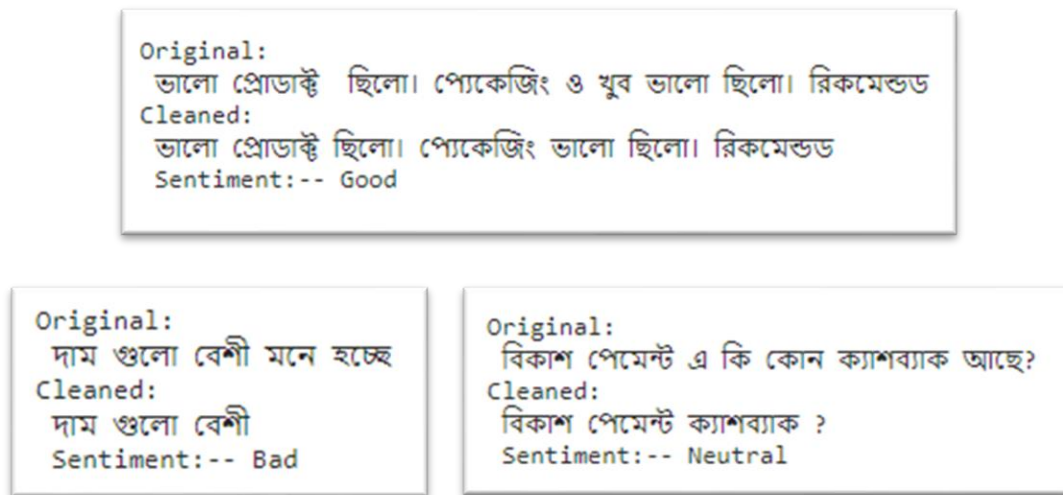


Figure 3.3: Text cleaning and Pre-processing example of each class of Review.

The image above represents the text pre-processing for all three types of review data. The dataset then gets tokenized, which breaks down its sentences into individual words, and then vectorized, which transforms the text into numerical vectors. Collectively, these essential features improve the dataset, making it appropriate for sentiment analysis. The modified dataset, which has become clean and organized is going to be used as an initial input of our machine learning algorithms.

3.4.4 Data Summarization

Data summarization is a data mining strategy by which we extract data from a dataset to drive any choice and show the data based on the requirements. In this section, I will calculate the number of documents, positive and neutral words, unique values, and the top 3 unique words in the data I have gathered.

Title	Good	Bad	Neutral
Number of Reviews	961	1154	641
Number of Words	6544	9010	4849
Number of Unique Words	1117	1675	639
Top 3 Frequent Words	ভালো প্রোডাক্ট ধন্যবাদ	অর্ডার ডেলিভারি টাকা	বাংলাদেশ অভিনন্দন অফার

Table 3.1: Data Summarization of the Dataset.

3.4.4 Word Cloud

Here the following figure represents the word cloud of my dataset word that have been utilized during the research work.



Figure 3.4: Word Cloud representation of Dataset Words.

3.5 Machine Learning Models

I have utilized 6 distinct types of Machine Learning Approaches for the research study. In this section, I will go over the mechanisms and behaviors of each Machine Learning Approach separately.

3.5.1 Random Forest

The Random Forest is a renowned machine learning approach which comes under the category of supervised learning methods. It is used in problems related to regression and classification in the field of machine learning. It utilizes the ensemble learning principle to combine many classifiers to address complicated challenges that enhance model performance and accuracy. The greater the number of trees in the forest, the higher the accuracy, which decreases the issue of overfitting. The following diagram explains how Random Forest (RF) algorithm functions:

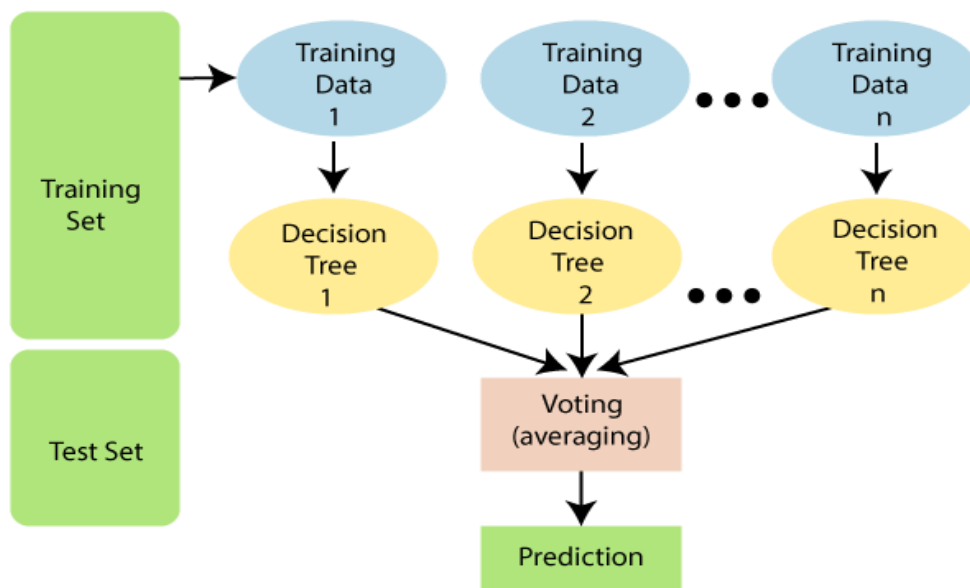


Figure 3.5: Working procedure of Random Forest Classifier [\[18\]](#).

Above figure demonstrates the way how Random Forest algorithm works. Random Forest excels in performing Classification and Regression tasks, handling large, high-dimensional

datasets, and boosting model accuracy while preventing overfitting. However, its suitability for Regression tasks is comparatively limited, representing a notable disadvantage.

3.5.2 Decision Tree

A decision tree features a flowchart-like structure where essential nodes represent features, branches represent rules, and leaf nodes represent algorithm outcomes. It is an effective supervised learning method that can potentially used for both classification and regression purposes. The aim is to find the property that maximizes information gain or impurity reduction after the splitting.

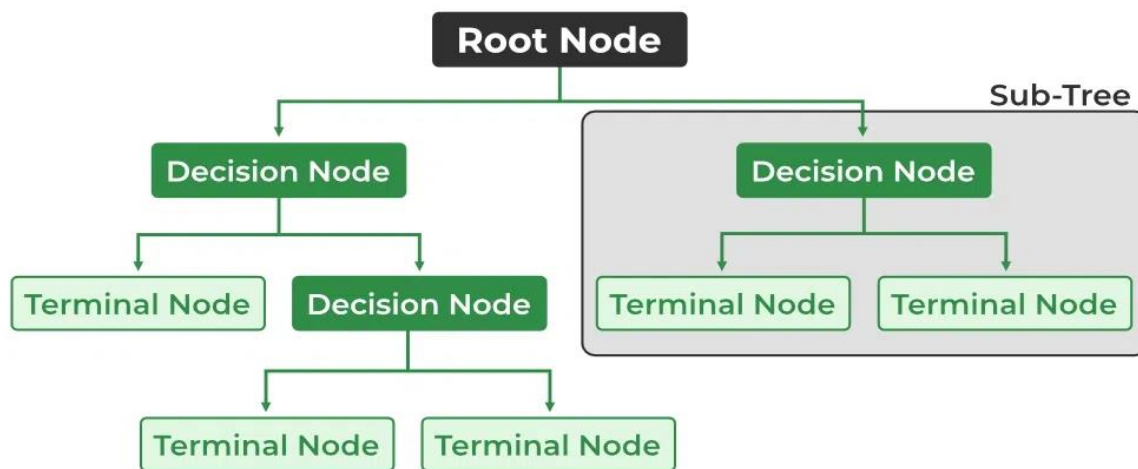


Figure 3.6: Working procedure of Decision Tree Classifier [19].

The corresponding diagram demonstrates the workflow used by the Decision Tree algorithm, which predicts classes through investigating dataset properties via frequent node comparisons. Despite its deep layer's complexity, the decision tree reproduces human decision-making, remaining comprehensible while implementing real-life decision process.

3.5.3 Multinomial Naive Bayes

Multinomial Naive Bayes is frequently used for classifying texts based on statistical analysis of their material [20]. According to the Bayes theorem, the probability $P(c|x)$, here x denotes the data provided occurring which must be divided into groups. The equation of MNB goes as follows:

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

It is mostly used in NLP area and predicting the text/audio label. The dataset I applied for this research study features two key columns: comments or reviews and the tag of that review. In this scenario, it will classify an emotion as good, bad, or neutral reviews, and we need to compute the following:

$P(\text{Good} | \text{My Dataset})$ represents the probability that a review tag will be suitable when the term "My Dataset" is utilized.

$P(\text{Bad} | \text{My Dataset})$ represents the probability that a review tag will be Negative while the term "My Dataset" is utilized.

$P(\text{Neutral} | \text{My Dataset})$ represents the probability that a review tag will be Neutral when the term "My Dataset" is utilized.

Another thing to keep in mind is that before we use the algorithm, we must remove stopwords and perform stemming for this approach.

3.5.4 Support Vector Machine

The Support Vector Machine is an advanced machine learning approach that can be used for linear or nonlinear classification, regression, and additionally identifying outliers [\[21\]](#). Text classification, image classification, spam detection, handwriting recognition, face detection, and anomaly detection belong to the operations that SVM are capable of performing. The following represents the workflow of this algorithm.

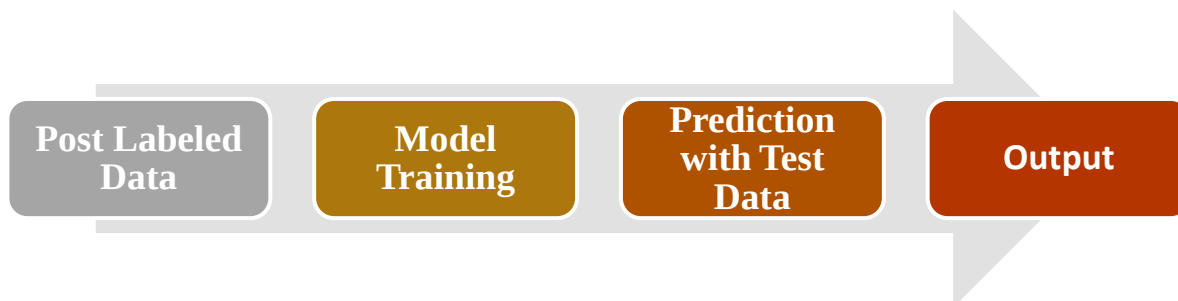


Figure 3.7: Working procedure of Support Vector Machine [\[21\]](#).

Its algorithm's principal objective are the pinpoint of the most suitable hyperplane in an N-dimensional space which can divided datapoints into distinct categories in the features. Support Vector Machines (SVM) are classified into two types based on the functionalities.

Linear SVM: Linear SVMs separate data points into classes by using a linear decision boundary.

Non-Linear SVM: Non-Linear SVM is used to classify data when a straight line divided it into more than two classes.

3.5.5 K-Nearest Neighbors

K-Nearest Neighbors is a simple and significant method for classification in Machine Learning Sector. KNN predicts the label or value of a new data point through analyzing its K nearest neighbors in the training dataset, applying the idea of similarity. Assume the following table of items of data with two distinct features:

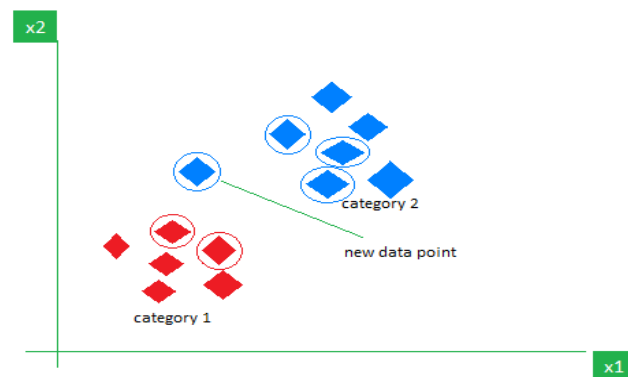


Figure 3.8: Working procedure of KNN Classifier [\[22\]](#).

The figure above demonstrates the KNN algorithm in workflow. Unclassified points are indicated as White in this above diagram. KNN is easier to implement because the algorithm's complexity is less, but we must be confirming data scaling right before starting the process of Machine Learning.

3.5.6 Logistic Regression

Logistic regression is a method based on supervised machine learning that is utilized to determine the probabilities of a particular case belonging to a specific class of data. It influences in decision-making processes through its functionalities. It transforms projected values to probabilities applying the sigmoid function [23]. The basis for my study is Logistic Regression, which is classified into Binomial, Multinomial, and Ordinal varieties, with a particular focus on Multinomial Logistic Regression with three options available: Good, Bad, and Neutral. The following figure shows the diagram of Logistic regression the way it works:

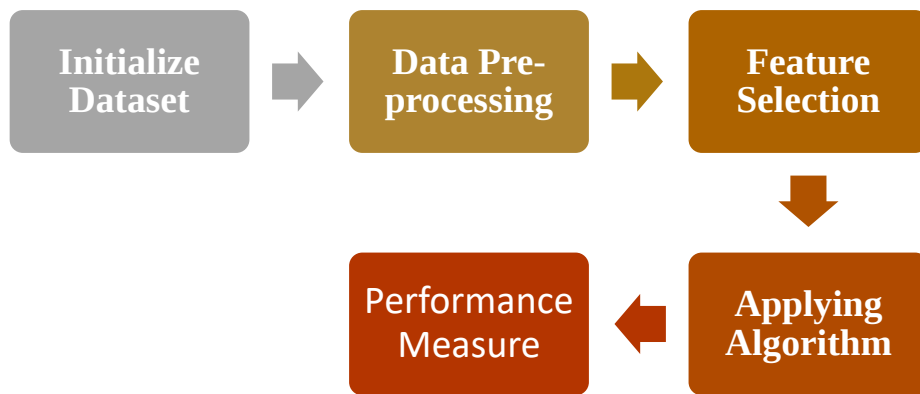


Figure 3.9: Working procedure of Logistic Regression [23].

The odds define the ratio of an event occurring towards not occurring, as compared to probability, which describes the ratio of an event occurring against all possible instances.

3.6 Natural Language Processing

Natural Language Processing is commonly referred as NLP. Which the branch of AI, which allows machines to fully understand and analyze human languages. Human languages may adopt form in the form of text or audio. In my research task, I am working on text-based Bangla e-commerce reviews. Natural Language Processing works in the following steps, as shown in the figure:

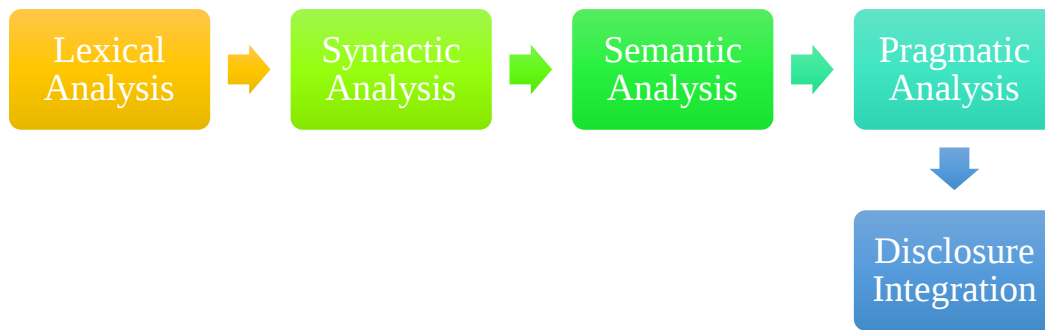


Figure 3.10: Working procedure of Natural Language Processing

NLP is extraordinarily fast and efficient. It enables the evaluation of data from both structured and unstructured resources [24]. Although NLP was created for a particular, specific task, it makes it possible for users to ask queries about every topic and receive an immediate response in milliseconds.

3.7 Summary

The research approach is described in a systematic format, starting with data analysis procedures and the application of text-analyzing research methodologies to determine the customer sentiment. Also demonstrates the proposed approach, which displays the application of multiple algorithms for machine learning to assure accuracy, precision, F1 Score and recall. The dataset section is clearly described, with a special focus on customer reviews' sentiment categorization as 'Good,' 'Bad,' or 'Neutral.' To achieve maximum machine learning compatibility, data pre-processing includes accurate text cleaning, tokenization, and vectorization and other processes. On the next handling the data, balancing the data, and text pre-processing the data with the objective to enhance the dataset for sentiment analysis through Bangla eCommerce review. In addition, significant insights are provided by data summarization, and machine learning methods such as Multinomial Naive Bayes, Support Vector Machine, K-Nearest Neighbors, Random Forest, Decision Tree, and Logistic Regression have been applied and discussed briefly. The performance, capabilities, and limits of each model are provided, resulting in an in-depth understanding of the study tactics.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

In this Chapter, I will focus on the results of using various algorithms for machine learning which on my utilized dataset, that consists 3280 Bangla reviews collected from top and popular e-commerce websites around the country. The significantly accomplished Kaggle dataset includes different views among consumers, classifying reviews as 'Good,' 'Bad,' or 'Neutral'. After that Decision Tree, Multinomial Naive Bayes, Support Vector Machine, Random Forest, K-Nearest Neighbors, and Logistic Regression are all parts of our machine learning process for the sentiment classification. The accuracy, precision, recall, and F1 score of these algorithms which employed to measure the performance. Accuracy examines overall accuracy, precision, recall, and sensitivity, as well as the F1 score. The following discussion analyzes each machine learning model's effectiveness in comprehending the complex nature of Bangla e-commerce reviews, putting light on their different strengths and areas of development. This research will study helps to acquire an deep knowledge of the models' performance in acquiring and classifying different opinion in the dataset.

4.2 Experiment Result

In this section of my research study of a large dataset of 3280 Bangla e-commerce customer reviews, categorized into 'Good,' 'Bad,' or 'Neutral,' shows a few interesting results as well. Using the Random Forest algorithm with a Bigram feature, my research study produced an outstanding accuracy of 90.76%. This machine learning approach required precise management of data, balancing, and text pre-processing, all of which enhanced the dataset's quality for sentiment evaluation. Particularly, Random Forest beat other algorithms in effectively categorizing reviews. The efficacy of the approach can be due to the collaborative learning principle, which efficiently integrates multiple classifiers to meet challenging requirements. Taking advantage of Bigram features increased the algorithm's performance even further, capturing advanced language patterns in Bangla reviews of ecommerce. This result demonstrates Random Forest's success in recognizing sentiment nuances within the

dataset, proving its ability to deal with the particular linguistic characteristics of Bangla e-commerce reviews.

4.3 N-gram Language Models

The continuous sequence [25] of n items from a specific portion of text or audio is defined as an N-gram. N-grams frequently come from a text or audio database. In my research study, I employed Unigram, Bigram, and Trigram to compare outcome and discovered that Trigram outperformed them with greater accuracy. From the given example here, I am explaining all three of them.

“চাওয়া মাএ পাইলাম সব কিছু ঠিকঠাক ধন্যবাদ দারাজ”

Unigram: An n-gram of size 1 is referred to as a Unigram. Here is an example from a review from the dataset I have utilized in this research work:

চাওয়া	মাএ	পাইলাম	সব	কিছু	ঠিকঠাক	ধন্যবাদ	দারাজ
--------	-----	--------	----	------	--------	---------	-------

Bigram: An n-gram of size 2 is referred to as a Bigram. Here is an example from a review from the dataset I have utilized in this research work:

চাওয়া মাএ	পাইলাম সব	কিছু ঠিকঠাক	ধন্যবাদ দারাজ
------------	-----------	-------------	---------------

Trigram: An n-gram of size 3 is referred to as a Trigram. Here is an example from a review from the dataset I have utilized in this research work:

চাওয়া মাএ পাইলাম	সব কিছু ঠিকঠাক	ঠিকঠাক ধন্যবাদ দারাজ
-------------------	----------------	----------------------

4.4 Result Comparison of Machine Learning Approaches

Here in this section, I will be comparing the outcomes provided by the Machine Learning Algorithms using all three types of n-Gram features. For each n-gram feature I have compared Accuracy, Precision, F1 Score, and Recall respectively. Then I have highlighted the highest value from each column and its respective Machine Learning model individually.

4.4.1 Unigram

Models	Accuracy	Precision	Recall	F1 Score
DT	88.22	88.71	88.22	88.27
RF	90.58	90.72	90.58	90.58
LR	84.78	85.00	84.78	84.76
MNB	83.33	83.46	83.33	83.27
Liner SVM	78.80	82.20	78.80	78.23
Karnel SVM	81.16	83.33	81.16	80.80
KNN	78.26	80.85	78.26	78.35

Table 4.1: Result comparison using Unigram Feature.

*Highest Accuracy achieved by **Random Forest** at = 90.58*

*Highest F1-Score achieved by **Random Forest** at = 90.58*

*Highest Precision Score achieved by **Random Forest** at = 90.72*

*Highest Recall Score achieved by **Random Forest** at = 90.58*

4.4.2 Bigram

Models	Accuracy	Precision	Recall	F1 Score
DT	88.22	88.52	88.22	88.24
RF	90.76	90.85	90.76	90.77
LR	89.67	90.18	89.67	89.62
MNB	87.50	87.43	87.50	87.44
Liner SVM	78.80	82.98	78.80	78.22

Karnel SVM	83.70	85.73	83.70	83.34
KNN	72.46	79.72	72.46	72.99

Table 4.2: Result comparison using Bigram Feature.

*Highest Accuracy achieved by **Random Forest** at = 90.76*

*Highest F1-Score achieved by **Random Forest** at = 90.77*

*Highest Precision Score achieved by **Random Forest** at = 90.85*

*Highest Recall Score achieved by **Random Forest** at = 90.76*

4.4.3 Trigram

Models	Accuracy	Precision	Recall	F1 Score
DT	88.77	89.13	88.77	88.79
RF	90.40	90.48	90.40	90.40
LR	89.86	90.34	89.86	89.79
MNB	87.50	87.45	87.50	87.45
Liner SVM	79.53	83.63	79.53	79.03
Karnel SVM	84.24	86.28	84.24	83.93
KNN	70.83	78.35	70.83	71.25

Table 4.3: Result comparison using Trigram Feature.

*Highest Accuracy achieved by **Random Forest** at = 90.4*

*Highest F1-Score achieved by **Random Forest** at = 90.4*

*Highest Precision Score achieved by **Random Forest** at = 90.48*

*Highest Recall Score achieved by **Random Forest** at = 90.4*

4.5 Accuracy and F1 Score Comparison with Bar Chart

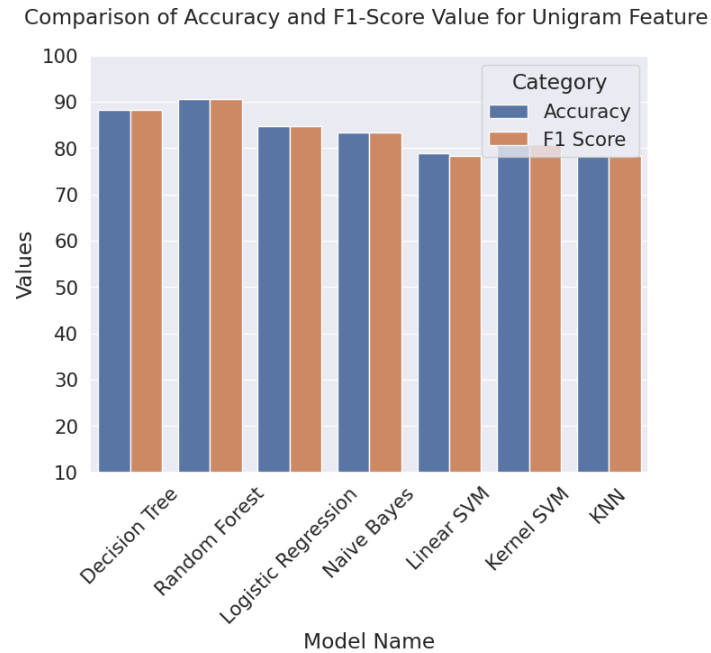


Figure 4.1: Result comparison using Unigram Feature.

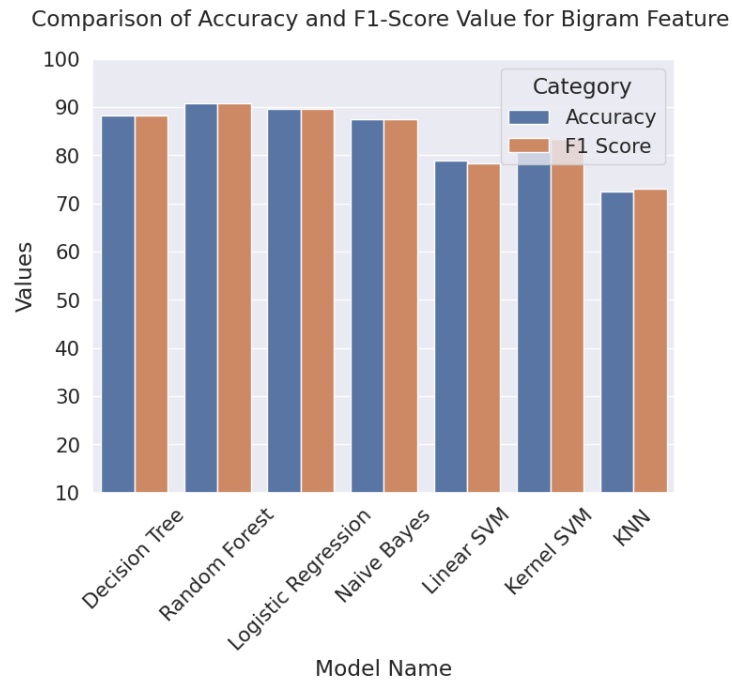


Figure 4.2: Result comparison using Bigram Feature.

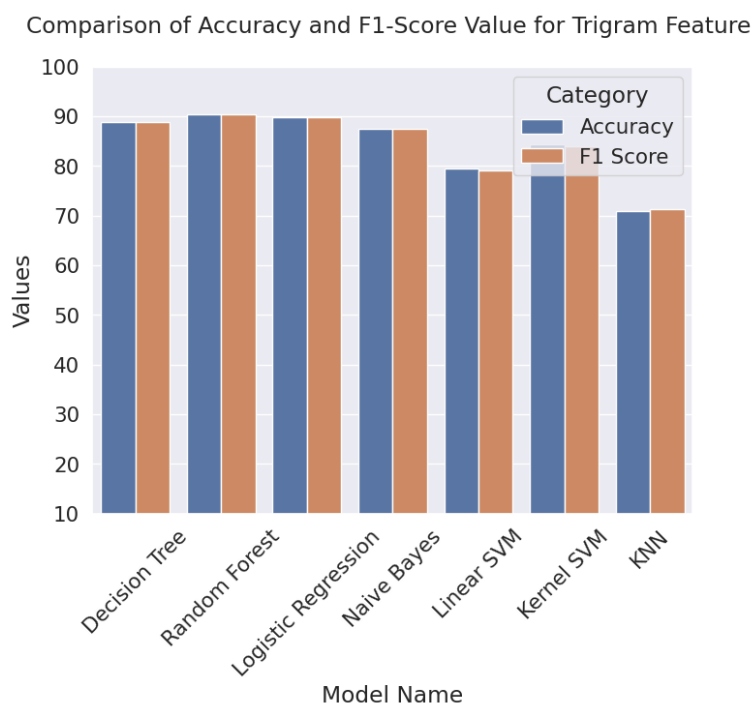


Figure 4.3: Result comparison using Trigram Feature.

Above mentioned figures represent the Bar Chart of Unigram, Bigram, and Trigram. There it displays values for each model that I have used during Machine Learning operation.

4.6 Conclusion

This chapter demonstrates the outstanding efficacy of seven different machine learning algorithms on a massive set of Bangla e-commerce reviews. With an impressive 90.76% accuracy, the Random Forest algorithm, especially applying a Bigram feature, succeeded. The addition of Bigram features enhanced the algorithm's ability to capture complex linguistic patterns in Bangla reviews. These results indicate Random Forest's capability for identifying sentiment variations as well as its ability to deal with the particular language characteristics of Bangla e-commerce reviews. Also, a comprehensive comparison across unigram, bigram, and trigram analyses regularly endorses Random Forest as a reliable and consistent substitute for sentiment analysis in the environment of Bangla e-commerce reviews, outperforming in accuracy, precision, recall, and F1 score.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

This chapter demonstrates the outstanding efficacy of seven different machine learning algorithms on a big set of Bangla e-commerce reviews. The Random Forest algorithm, especially when using a Bigram feature, obtained an impressive 90.76% accuracy.

Encouraging Customer Choices: One of the most important ways to promote informed customers in Bangladesh is to utilize machine learning algorithms for evaluating e-commerce reviews. From sentiment analysis of reviews, customers can find out a great deal about their contemporaries' experiences and then apply that knowledge to make better decisions when visiting e-commerce sites. As an outcome, the need for high-quality services and goods has risen and the customer base which become more sophisticated and powerful as well.

Customer Satisfaction: Employing machine learning approaches for reviewing e-commerce reviews in Bangladesh provides strongly to affecting accurate choices made by customers. People receive helpful insights on their peers' experiences via the extraction of sentiments associated with the reviews, encouraging greater understanding for making smart choices while using e-commerce platforms to buy something. This changing environment produces an informed and effective customer community, driving a rise in the market for high-quality products and services.

Building Trust: Trust is a vital component in everywhere in the natural world. The information acquired from machine learning analysis assist in developing and sustaining trust between customers and e-commerce organizations. Transparency the processing of customer comments, fixing complaints, and emphasizing positive experiences every assist in the development of a trustworthy and reputable shopping atmosphere.

Impact on Local Businesses: The impact of the study extends to local businesses involved in the market for online shopping. Local businesses can adapt and change their product lines to better coordinate with what customers want when the views of customers are investigated and understood. Along with in consequence, supports the growth of local businesses,

promoting economic development inside the boundaries of Bangladesh's larger e-commerce sector.

Influence on Market Trends: The study assists to recognizing and influence business developments by identifying sentiments conveyed in e-commerce reviews. Employers are able to modify their strategies depending on common behaviors that are ensuring that they are compatible with the needs of customers. This adaptability not only enhances the competitive ability of individual businesses, but it also contributes to molding the overall destiny of the Bangladesh online shopping industry.

Marketing Strategies: Marketers might better understand the viewpoints and interests of their customers through applying sentiment analysis. By discovering positive attitudes, marketers can emphasize key features and benefits that are appealing to consumers in their marketing campaigns. In the same way, marketers might develop tactics for resolving complaints from the customers, improving product quality, and efficiently providing solutions by tackling negative perceptions. As a result, marketing campaigns are more successful, improving customer engagement and revenue.

5.2 Impact on Environment

The study's adverse effect on the environment derives from e-commerce's basic characteristics, especially the decrease in sales. Improving logistics and transportation are likewise the result of supply chain process enhancement made possible by machine learning insight from this research work. This efficiency decreases the overall impact on the environment of product delivery, which not only connects with sustainability goals but also reduces the environmental impose of the e-commerce sector. Also, shifting to online shopping facilitates paperless transactions by replacing the need for paper bills along with additional documents, which cuts the amount of paper needed and avoids deforestation as well. The study's overall findings show the opportunity for online retailers to uphold environmentally friendly company procedures and promote sustainable operations when it is supplied using machine learning observations.

5.3 Ethical Aspects

The proper use of machine learning for the analysis of the e-commerce reviews is the study's key ethical concern. It is critical for assuring the user privacy, security of data, and open knowledge of data management practices. Ethical norms require an impartial, fair representation of a variety of customer reviews while minimizing any possibility of fraud. The study work is further exploring ways online shopping sites could responsibly utilize machine learning observations, addressing moral behavior while using feedback from customers. These elements maintain user confidence, preserve equity, and encourage the moral development of machine learning applications in the e-commerce industry.

5.4 Sustainability Plan

sustainability plan is essential for continuous positive effect, given the possible effects of this study on the e-commerce industry in Bangladesh. Creating a solid structure for managing data is necessary for assuring the ethical use of machine learning in sentiment analysis moving forwards. It is essential to conduct regular checks and modification of algorithms to decrease biases while enhancing accuracy. Integrity and openness may be encouraged by working together with customer advocacy groups, government agencies, and e-commerce groups. Ethical utilization of technology will also be emphasized by supporting programs that promote awareness and supply skill development for those who are involved in the e-commerce environment. By ensuring that machine learning is in accordance with ethical standards and for the good of society as a whole this sustainability approach intends to preserve the beneficial impact of machine learning upon the industry of e-commerce.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

In conclusion, my study work uncovered significant findings with broad consequences through the use of machine learning techniques to analyze e-commerce reviews from the perspective of Bangladesh. Along with allowing consumers to make well-informed decisions, a comprehensive analysis of sentiments provided in reviews has had a significant effect on the local business motion, consumer satisfaction, and trust-building throughout the world of online shopping in Bangladesh. However, I've used several kinds of machine learning algorithms to figure out which is best. The study revealed the effectiveness of the Random Forest method, which comes out for its high accuracy of 90.76% while it uses Bigram features. The study also demonstrates how Natural Language Processing enables us to swiftly extract emotion from any written input. This research investigation also addressed the impact of the review on the e-commerce industry and included examples of e-commerce fraud. The study mentioned earlier reveals the incredible potential of machine learning to change the e-commerce sector and influence how customers act. So, at last, it is clear that the e-commerce sector could be shifting toward a future that is more focused on customers, trustworthy, and dynamic.

6.2 Future Scope

The findings of the research open up fascinating fresh possibilities for research and development. I have worked on customer reviews for this research study, and I have divided them into three categories. However, I believe that it would be more effective if I employed more particular information regarding e-commerce and customers. Following that, I used 3280 Bangla review data; machine learning algorithms would become more accurate if I was employed a greater quantity of dataset. At last, research that emphasizes subtle patterns of speech unique to Bangla reviews may enhance the models' adaptability. Applying deep learning techniques might potentially provide a more profound understanding of sentiment and environmental details.

REFERENCES

- [1] Shafin, M. A. (2020). Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language Learning | IEEE Conference Publication | IEEE Xplore. Doi: [10.1109/ICCIT51783.2020.9392733](https://doi.org/10.1109/ICCIT51783.2020.9392733)
- [2] Munna, M. H. (2020). Sentiment Analysis and Product Review Classification in E-commerce Platform. IEEE Xplore. Doi: [10.1109/ICCIT51783.2020.9392710](https://doi.org/10.1109/ICCIT51783.2020.9392710)
- [3] Hossain, M. J. (2022). Sentiment Analysis on Reviews of E-commerce Sites Using Machine Learning Algorithms Learning | IEEE Conference Publication | IEEE Xplore. Doi: [10.1109/ICISSET54810.2022.9775846](https://doi.org/10.1109/ICISSET54810.2022.9775846)
- [4] Das, R. (2022). Classification of Product Review Sentiment by NLP and Machine Learning | IEEE Conference Publication | IEEE Xplore. Doi: [10.1109/ICAECT54875.2022.9808003](https://doi.org/10.1109/ICAECT54875.2022.9808003)
- [5] Zulfiker, S.(2022). Bangla E-Commerce Sentiment Analysis Using Machine Learning Approach. | IEEE Conference Publication | IEEE Xplore. Doi: [10.1109/STI56238.2022.10103350](https://doi.org/10.1109/STI56238.2022.10103350)
- [6] Hossain, E. (2020). Sentiment polarity detection on Bengali Book Reviews using multinomial naive bayes. Doi: <https://doi.org/10.48550/arXiv.2007.02758>
- [7] Soumik, M. (2019). Employing Machine Learning techniques on Sentiment Analysis of Google Play Store Bangla Reviews. Learning | IEEE Conference Publication | IEEE Xplore. Doi: [10.1109/ICCIT48885.2019.9038348](https://doi.org/10.1109/ICCIT48885.2019.9038348)
- [8]. Li, X. (2022). A Novel Deep Learning-based Sentiment Analysis Method Enhanced with Emojis in Microblog Social Networks. Enterprise Information Systems, 1–22. Doi: [10.1080/17517575.2022.2037160](https://doi.org/10.1080/17517575.2022.2037160)
- [9] Dost, A. (2019). The Classification of Customers' Sentiment using Data Mining Approaches. Gssrjournal.com. Doi: [10.31703/gssr.2019\(IV-IV\).19](https://doi.org/10.31703/gssr.2019(IV-IV).19)
- [10] Chowdhuri, R. (2019). Analyzing Sentiment of Movie Reviews in Bangla by Applying Machine Learning Techniques Learning | IEEE Conference Publication | IEEE Xplore. Doi: [10.1109/ICBSLP47725.2019.201483](https://doi.org/10.1109/ICBSLP47725.2019.201483)
- [11] Top 10 e-Commerce Websites and Online Shopping Apps in Bangladesh (2023). Businesshaunt.com. <https://businesshaunt.com/top-10-e-commerce-website-in-bangladesh/>
- [12] Halder, S. (2023). Ecommerce Scam: Customers yet to get Tk 600cr back. The Daily Star. <https://www.thedailystar.net/business/news/ecommerce-scam-customers-yet-get-tk-600cr-back-3320106>
- [13] Karim, R. (2022). Bangladesh e-commerce sales to more than double by 2026: Research. The Business Standard. <https://www.tbsnews.net/economy/bangladesh-e-commerce-sales-more-double-2026-research-497134>
- [14] Top 10 Benefits Of Customer Reviews In E-commerce. (2020). Meetanshi.com. <https://meetanshi.com/blog/benefits-of-customer-reviews-in-e-commerce/>
- [15] Lion, A. (2022). Probes in e-commerce fraud cases move at snail's pace. The Business Standard. <https://www.tbsnews.net/bangladesh/probes-e-commerce-fraud-cases-move-snails-pace-490538>
- [16] Branka. (2023). Online Review Statistics. True List. <https://truelist.co/blog/online-reviews-statistics/>

- [17]. Haque, R. (2019). A complete Bengali stop word detection mechanism. IEEE Conference Publication | IEEE Xplore. Doi: [10.1109/ICIEV.2019.8858544](https://doi.org/10.1109/ICIEV.2019.8858544)
- [18]. Random Forest Algorithm. JavatPoint. <https://www.javatpoint.com/machine-learning-random-forest-algorithm>
- [19]. Decision Tree Algorithm. GeeksforGeeks. <https://www.geeksforgeeks.org/decision-tree/>
- [20]. Sanghi, D. (2023). Applying Multinomial Naive Bayes to NLP Problems. GeeksforGeeks. <https://www.geeksforgeeks.org/applying-multinomial-naive-bayes-to-nlp-problems/>
- [21]. Aswathiasidharan. (2023). Support Vector Machine (SVM) Algorithm. GeeksforGeeks. <https://www.geeksforgeeks.org/support-vector-machine-algorithm/>
- [22]. K-Nearest Neighbor (KNN) Algorithm. GeeksforGeeks. <https://www.geeksforgeeks.org/k-nearest-neighbours/>
- [23]. Logistic Regression in Machine Learning. GeeksforGeeks. <https://www.geeksforgeeks.org/understanding-logistic-regression/>
- [24]. Natural Language Processing (NLP) Tutorial. GeeksforGeeks. <https://www.geeksforgeeks.org/natural-language-processing-nlp-tutorial/>
- [25]. N-Gram Language Modelling with NLTK. GeeksforGeeks. <https://www.geeksforgeeks.org/n-gram-language-modelling-with-nltk/>

Abbreviations:

ML – Machine Learning

NLP – Natural Language Processing

LR – Logistic Regression

RF – Random Forest

SVM – Support Vector Machine

MNB – Multinomial Naive Bayes

KNN – K-Nearest Neighbor

DT – Decision Tree

TF-IDF – Term Frequency - Inverse Document Frequency

NLTK – Natural Language Toolkit

ABP – Adaptive Boosting Algorithm

GBA – Gradient Boosting Accuracy

A STUDY TO ANALYZE THE REVIEWS OF E-COMMERCE

ORIGINALITY REPORT

7 %	6 %	1 %	2 %
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1 %
2	academic-accelerator.com Internet Source	1 %
3	Submitted to University of Wolverhampton Student Paper	1 %
4	www.mdpi.com Internet Source	1 %
5	Submitted to University of New Mexico System Student Paper	1 %
6	www.ijraset.com Internet Source	<1 %
7	vdocuments.net Internet Source	<1 %
8	www.researchgate.net Internet Source	<1 %
9	Ajay Kumar, Seema Bawa, Neeraj Kumar. "iDP: ML-driven diabetes prediction framework	<1 %