Food Review Sentiment Analysis In Bangla Language: A Comparative Analysis Of Boosting And Traditional Machine Learning Algorithm

BY

Md. Asmaul Haque
ID: 191-15-12622

AND

Samiul Islam Abir

ID: 191-15-12559

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

Supervised By
Lamia Rukhsara
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By
Md Azharul Islam Tazib
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This project titled "**Food Review Sentiment Analysis in Bangla Language: A Comparative Analysis Of Boosting and Traditional Machine Learning Algorithm**", submitted by Md Asmaul Haque, ID:191-15-12622 and Samiul Islam Abir, ID: 191-15-12559 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 23 January, 2023.
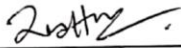
BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Zahid Hasan
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Fahad Faisal
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
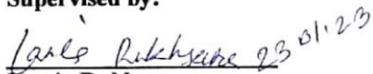Daffodil International University

External Examiner

Dr. Ahmed Wasif Reza
Associate Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this thesis has been done by us under the supervision of **Lamia Rukhsara, Senior Lecturer, Department of CSE**, and co-supervision of **Md Azharul Islam Tazib, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Lamia Rukhsara 23 01.23

Lamia Rukhsara
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Md Azharul Islam Tazib
Lecturer
Department of CSE
Daffodil International University

Submitted by: *Asmaul*

Md Asmaul Haque
ID: 191-15-12622
Department of CSE
Daffodil International University

Samiul Islam Abir

Samiul Islam Abir
ID: 191-15-12559
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for his divine blessing makes us possible to complete the final year project internship successfully.

We are really grateful and express our earnest indebtedness to **Lamia Rukhsara**, Senior Lecturer, Department of CSE Daffodil International University, Dhaka, Bangladesh. Profound Knowledge & intense interest of our supervisor in the field of "Machine Learning & Deep Learning" make our way very smooth to carry out this thesis. Her remarkable patience and dedication, scholarly guidance, continual encouragement, vigorous motivation, direct and fair supervision, constructive criticism, valuable advice, and great endurance during reading many inferior drafts and correcting the work at all stage have made it possible to complete this thesis.

We would like to express our gratitude wholeheartedly to **Prof. Dr. Touhid Bhuiyan**, Professor, and Head, Department of CSE, for his kind help to finish our thesis and also to other faculty members and the staff of the CSE department of Daffodil International University.

We would like to express thankfulness to fellow student of Daffodil International University, who took part in this discussion during the completion of this work.
We would like to express our immense thanks to the Different food application to visible us user original reviews as a result we collected raw data to make our work possible.
We would also like to thank the people who provide the done by us to collect the market real information.

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

ABSTRACT

Even in this day and age of advanced internet technology, businesses that deliver meals via the internet in Bangladesh are doing quite well. Even after the COVID-19 pandemic, the majority of people living in urban areas of Bangladesh are unfazed when it comes to placing meal orders with a variety of food delivery services. The expansion of the food supply has made it possible for consumers to experience previously unattainable levels of convenience. People frequently consult food reviews and ratings before making decisions about what to eat. Ratings are not the only factor that goes into determining the caliber of a dish. The menu, the delivery monitoring program and the attitude of the delivery person are all elements that go into the rating. As a consequence of this, the reader is required to read every single comment pertaining to food. However, this process does require some patience, mainly due to the fact that it is difficult to read all of the comments on the different dishes. As a consequence of this, we attempted to develop a system for providing insightful feedback regarding meals. Around two thousand Bangla phrases have been compiled by our team for use in a variety of online meal delivery applications. Throughout the course of our investigation, we made utilized a number of different methods for analyzing sentiment and classifying data. Some of these traditional machine learning methods include DT, RM, and LR. Compare with Boosting classifier algorithms. We made utilized an algorithm that not only gave us an F1 score but also gave us the highest possible level of accuracy in our prediction.

TABLE OF CONTENTS

CONTENTS	PAGE

CHAPTER	PAGE NO

LIST OF TABLE

TABLE	PAGE NO.
Table 3.1 Tokenization Table	13
Table 4.1 Traditional Algorithm Accuracy	19
Table 4.2 Boosting Classification Algorithm Accuracy	20

LIST OF FIGURE

CHAPTER 1

INTRODUCTION

1.1 Introduction

Significant growth in the number of online food businesses has been observed in Dhaka and a few other cities in the surrounding region. The market for online food supplies now has a daily turnover of 1 million takas, which is a significant increase from the almost nonexistent level seen in 2013. In 2013, HungryNaki, a local startup, became a member of the group. Subsequently, the company developed an online meal service platform for the country of Bangladesh. [1] It makes it possible for people living in Dhaka to look up nearby restaurants and place orders for food from those restaurants. Since 2013, a great number of companies that will deliver meals to your doorstep via the internet have emerged. The year before, we saw an increase that was 400 percent, which helped us expand to other cities, where people from those other cities are anticipated to respond positively. [2] It is estimated that 20,000 orders are currently being distributed across Dhaka on a daily basis [2], which provides a better understanding of this imperialism. An online meal delivery company reports that the average cost of each delivery is approximately 50 taka, which is equivalent to approximately 11800 dollars in the United States in Dhaka alone. On the basis of the explanation that has just been provided, one may reach the conclusion that in the not-too-distant future, there will be a significant demand for an online meal delivery service. As a consequence of this, any automated system will function as an essential component of this system.

In order to fulfill this requirement, we developed an intelligence method that can automatically classify the opinions expressed in Bangla food reviews. In this project, we used a combination of natural language processing and machine learning. The TFIDF approach was the one that we went with for the conversion of text to the numeric value. Following that, we assigned a classification to each sentiment using a number of different machine-learning approaches. We utilized data test utilization rates of varying degrees both for training and testing. In addition, the percentage range has been established as being between 30 and 70 percent. When the test data was reduced to 30%, our model performed

significantly better in the majority of cases. We went through all of the available options and settled on the one that would give us the best combination of accuracy and an f1 score. Following the selection of the most appropriate model, we validated it using fresh data on which it had not been previously trained.

1.2 Motivation

The number of users accessing the food-sale website in Bangladesh increases each day. The vast majority of people report feeling at least some level of comfort when purchasing food over the internet. This occurrence is becoming more commonplace every day among people who are passionate about food as a direct result of the epidemic of coronary disease. However, when people buy food online, they are subject to a number of restrictions that they must adhere to. The evaluation of meals is the primary focus of this investigation. Reading up on different cuisines is a smart way to get a comprehensive understanding of any food. But right now is the time to put in a significant amount of effort. As well as that, the purchaser could find each comment to be tiresome at times. As a result of the conversation that just took place, it is clear that something needs to be done to find a solution to this problem so that individuals can save time and easily obtain their preferred meal.

We are putting in effort because of these two limitations. In addition to this, we need to make a decision regarding whether or not to develop an intelligent system that is able to analyze the comment and place it into appropriate positive or negative categories. Ratings are now a distinct topic, and they make it much simpler to choose the best meals from a website. In actual practice, a different parameter is chosen. This is the overall quality of this company's business, as seen by the quality of the books, the delivery time, and the manner in which the delivery man conducted himself, for example. Choosing food using this method is not the best strategy for the same reason. In the end, we decided to address the issue by applying NLP and machine learning techniques. The algorithm is aware that the string cannot be decoded in any meaningful way. To get things started, we have to turn the string into its numerical representation. Within the context of this example, we utilized the TFIDF method. In addition, we used a technique called machine learning to categorize

any sentence. We utilized a variety of settings for each distinct algorithm. In addition to this, we have selected the settings that are the most efficient.

1.3 Rationale of The Study

In today's modern society, the field of natural language processing includes a subfield known as machine learning, which plays an extremely important role. The development of the NLP industry will be aided by machine learning because it can detect a variety of aspects at an earlier stage. We took one step at a time in order to overcome the difficulty of the task. The most important challenge was collecting relevant information. We compiled this data using a wide range of Bangla food blogs, food delivery apps, and restaurant websites. The second problem is that the data were not presented in an orderly fashion, and there was a great deal of background noise. As a consequence of this, we need to clean them very carefully, as the algorithms that perform machine learning are very good at this. In order to acquire the necessary knowledge, we investigated the entire dataset after it had been preprocessed. Before beginning the program's training, we finished by converting our dataset into a numerical representation.

1.4 Research Question

The following is a list of some of the topics that are discussed in this thesis:

- How should we go about collecting the necessary information?
- Is the classification scheme appropriate for the project at hand?
- What method are you going to use to label the data?
- What advantages would people have from participating in this project?
- Which approach do you think is the best for classifying different kinds of restaurant reviews?
- Can this concept actually be implemented in the world we live in?

1.5 Expected Output

Our research technique refers to the approach we plan to take when carrying out our inquiry. This chapter explains how we plan to address a variety of issues, including data collection methods, statistical analysis, participant observations, and more. Within this section of our study report, we have compiled the most recent Data Set, as well as data cleaning and selecting characteristics and feature selection procedures. After the model has been given some training, a classification strategy is used to determine the outcome (SVC, Logistic Regression, RF, SGD, MLP).

1.6 Research Objectives

The following is a list of our primary goals:

➢ To investigate how a particular classification scheme could be utilized to evaluate or categorize Bangla meal reviews.

➢ You should set up a discussion board where people can post their positive and negative comments about their meals.

➢ Give some thought to the theoretical studies that have been done on the classification of the Bangla language by classifier algorithms.

➢ Recognize reviews in multiple categories, such as excellent and terrible meals, using techniques such as machine learning and natural language processing.

1.7 Report Layout

Chapter 1: In this chapter, we will discuss the following topics: introduction, motivation, problem definition, research question, research technique, and the anticipated completion date of our program. In this section, we will also discuss the rationale behind our decision to carry out this research.

Chapter 2: In the second chapter, we investigate the background of the study, as well as other studies that are analogous to it and the current state of events as seen from Bangladesh's vantage point. In it, you'll find both an examination of the context as well as a concise description of the work.

Chapter 3: The methodology of the study will be discussed in Chapter 3. This chapter delves deeply into the process, whether it be the technique or the workflow. In this section, we will discuss the methodology that was used to carry out the research.

Chapter 4: The performance of the suggested model will be analyzed in this chapter with the help of a classification report and an accuracy table.

Chapter 5: The last chapter of the report is titled "Chapter 5," and it summarizes the findings of the previous chapters. The results of the model are broken down and discussed in this section. This section also includes a comparison of the accuracy of the information. Additionally, the web implementation and output of the model are discussed in this section. An analysis of the shortcomings of the work serves as the final topic of discussion in this chapter. In addition to that, it contains information that is embedded about upcoming projects.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

Researchers working on machine learning and artificial intelligence systems have developed a plethora of innovative ideas in recent years with the goal of enhancing algorithms and the applications that use them. Based on their intended system, a great number of projects are currently viewable online at a variety of websites located all over the world. This chapter will discuss the work that is pertinent to the discussion. We have provided a number of different recommendation system examples, along with their overall performance. The accuracy of the existing system is essential to how they use it, the method and model they use for forecasting, as well as the location where they put it in place.

2.2 Related Works

Utilizing statistical and machine learning methodologies, recent years have seen significant advancements made in the process of identifying and recommending newspapers based on historical data. CPD has been the subject of research in the fields of data mining, statistics, and computer science for a number of years at this point. In the following sections, we will go over the majority of the previous research that has been done on recommendation engines in a variety of fields. This research makes use of machine learning, shallow learning, and deep learning.

Morteza et al. [1] introduce a system that, in today's environment of big data, news platforms highlight both the challenges and opportunities for increasing the functioning of recommendation systems. Improved management and analysis of clickstream data, as well as a better understanding of users' preferences, are two benefits that can accrue to news recommendation systems that make use of cutting-edge big data storage and programming techniques. The vast majority of currently available algorithms for news recommendation make use of the clicks made by users as implicit input in order to study user behavior.

In contrast to the binary "true" or "false" (1 or 0) Boolean logic typically used in computation, "degrees of truth" are employed in fuzzy logic, as defined by M.N.M. Adnan

et al. Recommender systems take into account user tastes and make recommendations for content to read or view accordingly. Suggestion delivery strategies ranging from content-based to collaborative to knowledge-based and beyond have been proposed. Their system relied on fuzzy logic to find a group of articles that could be read in tandem and recommended to a reader. There is a clear purpose for utilizing fuzzy logic in practice. Zero and one are absolute numbers, and they are notoriously difficult to turn into engaging or informative news articles. They can't just say that because something is labeled "X," it must be related to "Y." That's why they attempted to create a fuzzy algorithm that takes into account multiple aspects of a news article to decide whether or not it should be recommended to users.

According to C. Feng et al. [2], traditional newspapers are being replaced by digital media such as websites and custom mobile applications. It has been demonstrated that news recommendation systems can evaluate lengthy articles automatically and make recommendations to users based on those evaluations and other parameters. From 2001 to 2019, this work examines the entire body of writing and nominates 81 interconnected factors, which are then discussed under six main headings. Many worries from a long list of issues in the news arena are addressed, and 60% of news proposal frameworks use a hybrid approach, while 66% consider tiny conversations almost datasets. This is the first work in the field to provide such a comprehensive overview of news recommendations at a macro level and to analyze the various metrics contained within the musings. The final section outlines some possibilities for a long-term study that could improve the news industry's ability to propose news stories.

In their paper, Walter et al. [3] present a social network model for a trustworthy recommendation system. According to the concept behind the model, agents gain access to data via their social network and filter it via their trust links. To learn how agent trust dynamics affect the system's success, they compared its performance to that of a frequency-based recommendation system. In addition, they found that knowledge sparsity, preference heterogeneity among agents, and network density all play crucial roles in the system's performance. The system self-organizes into a state where its performance is nearly

optimal; this global performance is an emergent property of the system, arising from the interactions of individual agents rather than from any premeditated effort to achieve it.

H.Tan and his fellow employees. [4] E-learning recommendation systems can assist students in making decisions, which is an extremely important skill to have in this day and age of overwhelming amounts of information when students lack the appropriate personal experience with the possibilities. In this study, they focused primarily on the user-based collaborative filtering technique as a recommended method related to distance learning. The workflow of an internet-based e-learning recommendation system consists of five stages: data collection, data extraction, transformation, and loading (ETL), model construction, and strategy setup and supply. The architectural layout does double duty by preparing the ground for future expansion. There are a total of seven parts that make up this architecture. However, only four of them are essential for proper functioning. A database of recommendation models, a database of recommendation systems, a recommendation administration module, and a data/model management module are the four components that make up this set of modules.

According to Goynai et al. [5] in today's information age, when learners lack sufficient first-hand experience with the available options, an e-learning recommendation system can help them decide how to proceed with their education. Their primary interest is in the collaborative filtering process, which is based on the input of the users of the system. Gathering data, extracting and transforming data, building a model, configuring a strategy, and delivering the service are the five steps in the process flow for an online learning recommendation system. The provided structure can also serve as a stepping stone to accommodate future growth. The suggestion models databases, the recommendation systems databases, the recommendation management, and the data/model management are the four primary modules that make up their architecture.

2.3 Comparative Analysis Summary

Research associates from a variety of study institutions undertake research into the studies that have been carried out in the subject of historical assessment. Our company is performing at an exceptionally high degree of efficiency right now. It is thought that as a result of the divide, it will be easier to evaluate the finer nuances of acquiring various goods, which might make it a more intelligent choice. Despite the fact that the referendum was not completely peaceful, this is the result.

2.4 Scope of The Problem

The scope of a project to analyze the sentiment of customer evaluations of food goods or restaurants would include doing an analysis of customer reviews in order to ascertain the general sentiment that is conveyed in the reviews. This might entail assessing the general degree of satisfaction that is stated in the evaluations as well as finding particular components of the food or service that consumers liked or disliked. The project could also require employing methods for natural language processing in order to extract important information from the evaluations, such as the precise types of food or menu items that were mentioned. In addition, the scope may include the visualization of the data in order to acquire insights and the identification of patterns in the sentiment of customers over time.

2.5 Challenges

When it comes to doing a study of sentiment based on restaurant reviews, there are a few potential obstacles that may arise:

How to deal with subjective language: Reviews may be quite subjective, which makes it difficult to precisely assess the feeling that is being communicated.

Taking care of sarcasm and irony: Some evaluations may be written in a sarcastic or ironic tone, which might cause sentiment analysis algorithms to provide inaccurate results.

How to deal with misspellings and slang: Reviews may include misspellings, slang, and informal language, all of which may make it challenging for natural language processing algorithms to interpret the content.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes several data collection and analysis strategies that were used that are pertinent to the study. Techniques of analysis are discussed in Chapter 3, which contributes to the development of the mathematical approaches utilized in this work. In addition to that, this chapter demonstrates how the Machine Learning classifier works its way through the operation. There is access to raw data, data that has already been preprocessed, data processing, tokenization, algorithms for the classifier, and all of the other stages that are required.

3.2 Data Collection Procedure

The entirety of the market has transitioned online in some capacity or another over the course of the past decade. Millions of comments and reviews on various food products are posted online every day by customers who have either purchased or used the item in question. Each review will contain the author's individual point of view and opinion. Only reviews written in Bangla were considered for this project because that was the only language in which we were competent to work. After that, we gathered Bangla reviews that were acceptable and easy to comprehend from a range of sources, including websites for restaurants, websites and apps for meal delivery, food review groups on Facebook, and other online resources. The 2000+ reviews of our research data have been successfully finished by our team.

Preprocessing

In the beginning, we performed some preliminary processing on the raw data by extracting the Bangla punctuations and stop words. This allowed us to get a head start on the analysis. After that, we made the necessary edits to the reviews, getting rid of the emoji that had been included. We have come to the conclusion that this information should be added at the very end of the review because it is impossible for us to reliably separate the emoji from the article because it is an integral part of the sentiment. This will make it possible for an accurate assessment of the appropriateness of the sentence to be carried out.

Analysis

In accordance with the evaluation criteria that were presented to us, we partitioned the dataset into two distinct categories (positive and negative). Since algorithms are unable to comprehend string data in an explicit manner. As a consequence of this, we are compelled to convert all of the data that we have processed into binary format. In order for us to accomplish this objective, we made use of a piece of software referred to as the TF-IDF algorithm. This abbreviation stands for the term frequency-inverse document frequency.

$$W_{i,j} = tf_{i,j} \times log\left(\frac{N}{df_i}\right) \qquad (1)$$

(1) The score of the TF-IDF, which can be found in the table, is denoted by the letters $W_{i,j}$. The variable $tf_{i,j}$ is shorthand for the overall frequency of occurrence of I in j. A conclusion can be drawn about the total number of documents based on the value of N. And df_i proves the entire number of papers, which includes I as part of its evidence.

3.3 Statistical Analysis

Table 3.1 Tokenization Table

Raw Data	Type	Tokenized data
খাবারটি গরম ছিলো	Positive	"খাবারটি"," গরম"," ছিলো"
যা অর্ডার করেছিলাম তাই দিয়েছে	Positive	"যা"," অর্ডার "," করেছিলাম "," তাই"," দিয়েছে"
এমন প্যাকেজিং করছে সব খাবার একসাথে হয়ে গেছে	Negative	"এমন"," প্যাকেজিং"," করছে"," সব"," খাবার"," একসাথে"," হয়ে"," গেছে",
ছবি দেখায় এক দেয় আর এক খাবার	Negative	"ছবি"," দেখায়"," এক"," আর"," এক"," খাবার"
খাবারটি অনেক মজা ছিলো	Positive	"খাবারটি"," অনেক"," মজা"," ছিলো"

By employing a strategy known as tokenization, we have separated individual words from a term that is necessary for the analysis of the phrase. This separation was necessary to complete the investigation. It was always our goal to present the tokenization system in Table 3.1. The tokenization system was going to be displayed there. Tokenizing each string is the first step that must be taken before employing the TFIDF methodology. As a direct consequence of this change, the punctuation, as well as the stop word, have been removed. Following that, in order to tokenize our raw data, we resorted to porter stemmer, which is a tokenization function that is utilized by a lot of people.
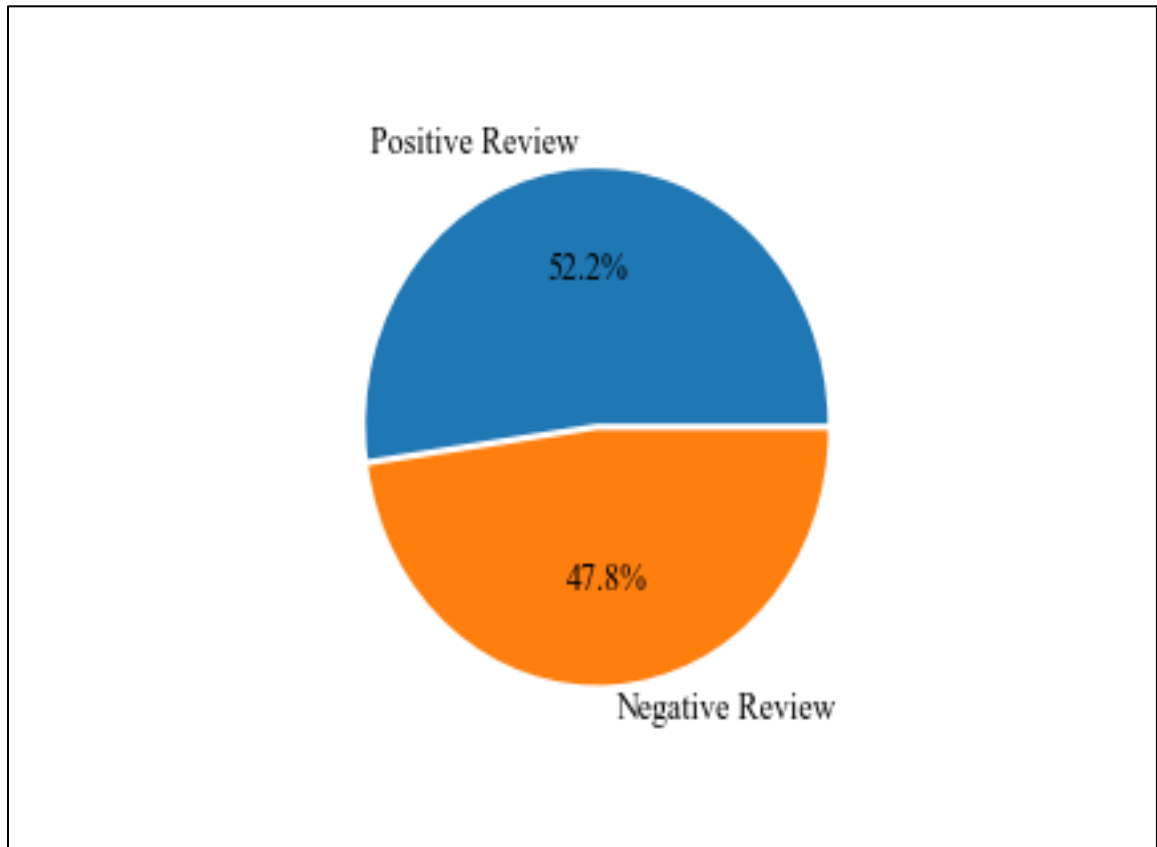
Classification

Figure 3.1: Classification of Dataset

We developed two separate sorts of Bangla reviews, and as a direct consequence of this, our entire dataset was split into two categories: positive and negative, which are represented by the numbers 0 and 1, respectively. There were approximately four hundred Bangla reviews handed out in each of the classes. An explanation of the categorization approach that was taken for the complete dataset is provided in figure 3.2. In addition, the percentage of each type of data that belongs to each type has been clarified. In this particular categorization graph, the overall amount of data that is positive is equal to 52.2%, while the overall amount of data that is negative is equal to 47.8%. It is evidence that our data are well-balanced in the way that you specified. It demonstrates how exact our dataset is and ensures that it will be suitable for the following tasks because of its precision.
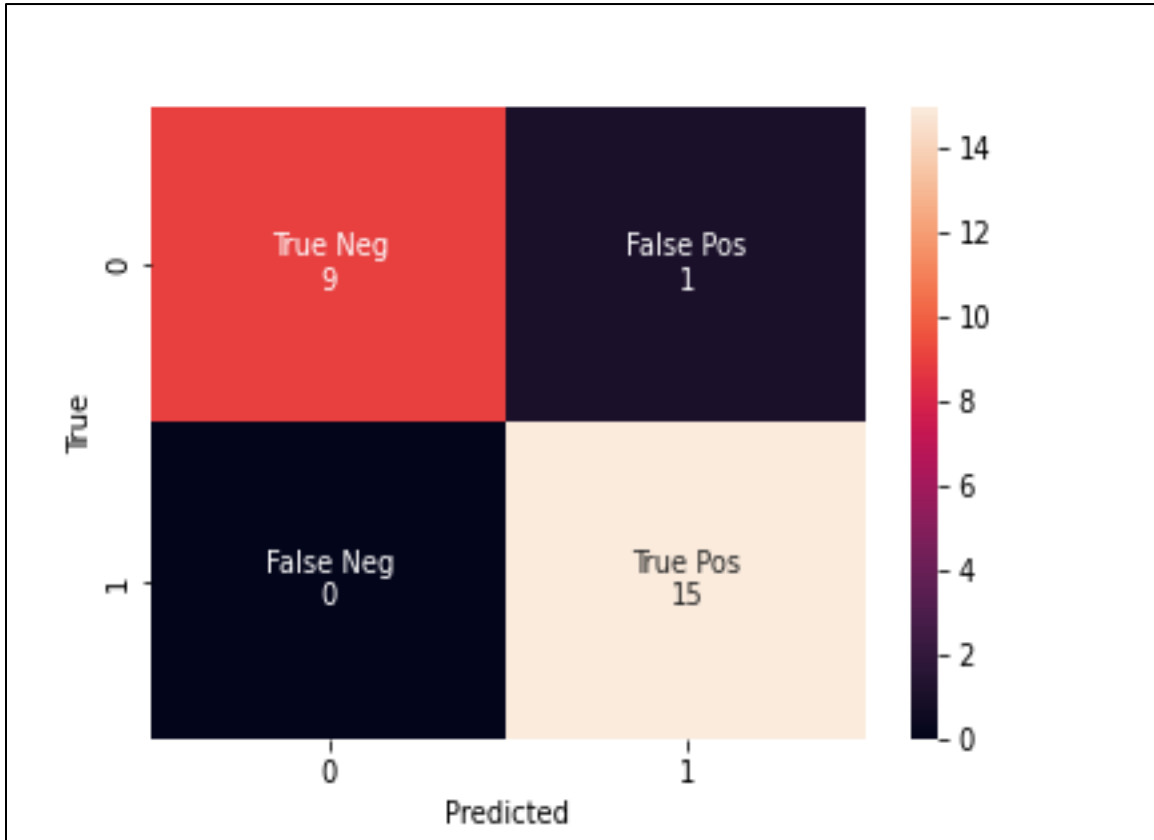
Figure 3.2: Confusion Matrix

We were successful in making an estimate of the Logistic Regression algorithm by employing the use of a confusion matrix. We used 25 newly collected primary data contains for the purpose of validation. These primary data contain included both positive and negative reviews of the product. The following are the results obtained from our algorithm: 15 appropriate positives, nine appropriate negatives, one inappropriate positive, and no inappropriate negatives were found. As a result, we are in a position to determine the precision-recall ratio and validate the model for the purpose of the prediction of Bangla food reviews.

$$\text{Accuracy} = TP + TN / TP + FP + FN + TN$$

$$= 15 + 9 / 15 + 1 + 0 + 9 = 0.93 * 100 = 93\%$$

$$\text{Error} = 1 - 0.93 = 0.07 * 100 = 7\%$$

The recall rate for positive:

$$TP/ TP + FN$$

$$= 15 /15+0$$

$$= 1*100$$

$$= 100\%$$

The recall rate for Negative:

$$TN /TN+FP$$

$$= 9 /9 + 1$$

$$= 0.90*100$$

$$= 90\%$$

The calculation that was just shown allows us to draw the conclusion that our model is more accurate than other models when it comes to the prediction of favorable reviews. This is due to the fact that a rate of positive memory is preferable to a rate of negative recall. In figure 3.6, we have included examples of all of the phenomena.

3.4 Proposed Methodology

Its term "research technique" refers to the practices or procedures that are utilized in identifying, selecting, managing, and assessing data relevant to a topic. This data can then be utilized to draw conclusions about the issue. The reader is given the ability to more correctly evaluate the general legitimacy and consistent quality of the consideration after reading the section of a research paper devoted to the methodology. The method includes a total of seven stages, all of which bring to a successful conclusion the research that is discussed in section 3.1. The procedures are outlined in the following list.
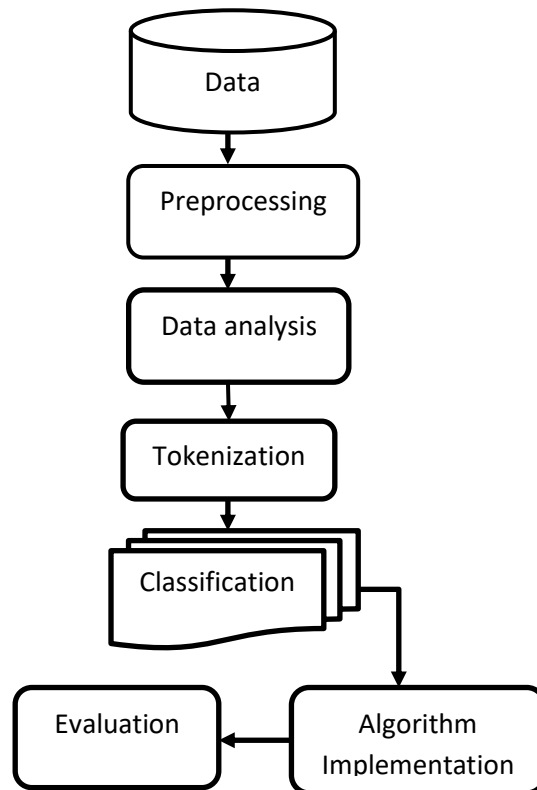

Figure 3.3: Methodology Diagram

3.5 Implementation Requirement

- Hardware:
    Processor: Core I3
    Ram: 8 GB
- Software:
    Google Colab
    Language: Python 3.8

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

The empirical data and descriptive research are going to be the main topics of this chapter. When we look into it, the first thing that we do is think about the findings. Constructing the Implications section in such a way that the findings are presented without the need for awareness or scrutiny is the goal of this part's construction. Under the area that discusses research papers, you'll find some recommendations. Documentation of the findings, as well as the exam, will be provided in this chapter.

4.2 Experiment Result & Analysis

Three conventional machine learning algorithms and three boosting classification techniques have both been employed. The decision tree, random forest, and logistic machine learning algorithms are used in this context. The GradientBoosting classifier, AdaBoosting, and CatBoostingClassificaiton are the other three boosting techniques. We compared these algorithms using accuracy and accuracy score. According to the accuracy table, Random Forest and CatBoostingClassificaiton acquire height accuracy of 99.70% and 99.10 while using only 30% of the input. Table 4.1 show the traditional machine learning algorithm and table 4.2 show the boosting classifier algorithms' accuracy.

Table 4.1: Traditional Algorithm Accuracy Table

Test Data usage rate	Algorithms		
	Decision Tree	*Logistic regression*	*Random Forest*
30%	97.34%	99.10%	99.70%

Table 4.2: Boosting Classification Algorithm Accuracy Table

Test Data usage rate	Algorithms		
	GradientBoosting	**AdaBoosting**	**CatBoosting**
30%	98.56%	96.99%	99.01%

From the analysis of the two tables, it can be said that the parts of the boosting algorithm and the traditional algorithm are almost the same.

Result Analysis

In this section, we take three metrics to present our algorithms more transparently, which are Recall, Precision, and F1 score. Here, we compare our main accuracy with these three metrics to see which algorithm performs more accurately.
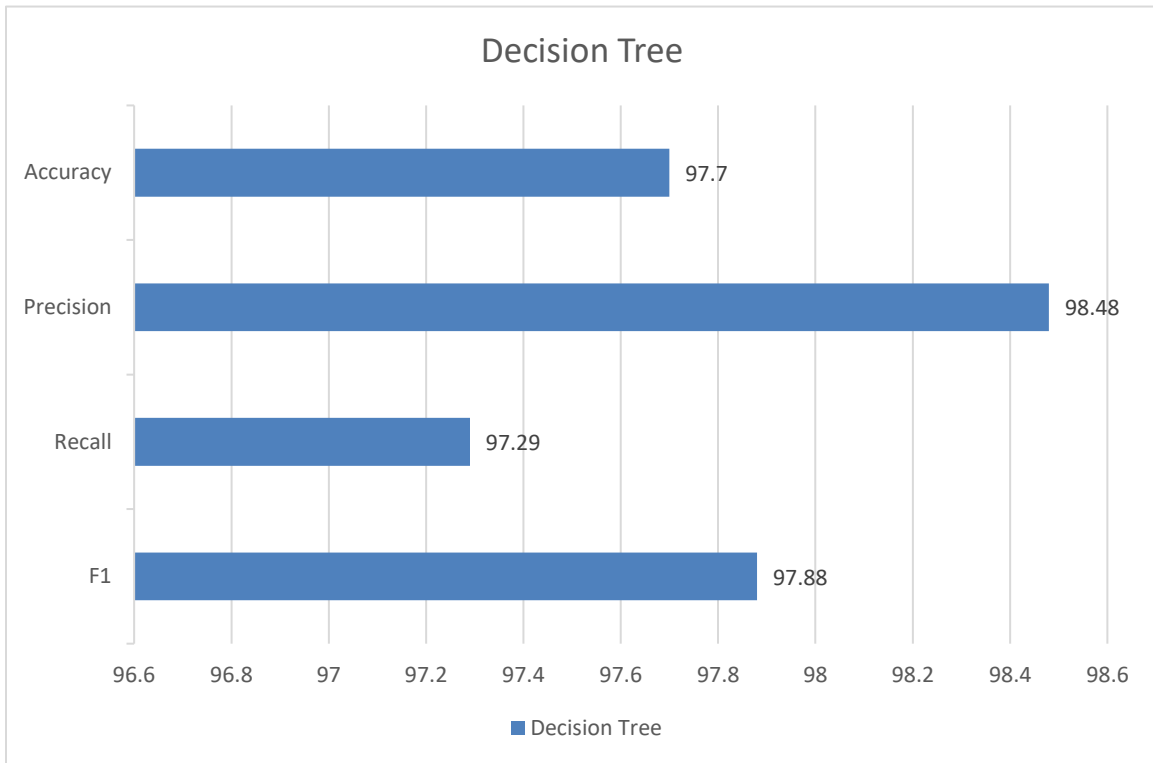
Decision Tree

Figure 4.1: Show the accuracy of the Decision Tree

Figure 4.1 show the accuracy of the DT algorithm. Here we compare three matrices with our accuracy. In this graph there the F1 value is 97.29, the Recall value is 97.88 and the precision is gaining the height value that is 98.48. The test data of 30% where our accuracy gain in DT is 97.7%, Which is similar to our three matrixes. we do not take this algorithm for our next step. It shows much lower than other algorithms.
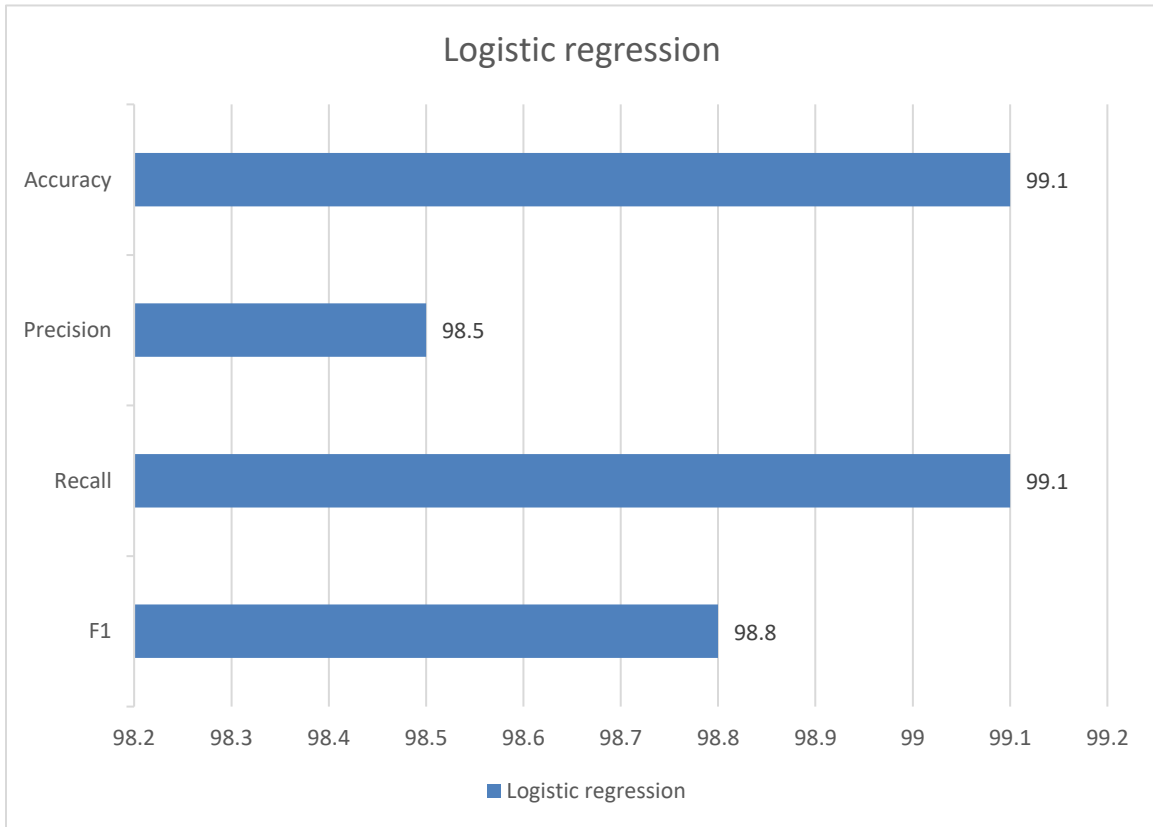
Logistic regression

Figure 4.2: Show the accuracy of the logistic Regressions

The accuracy of the LG algorithm is illustrated in figure 4.2. In this section, we evaluate the accuracy of three different matrices. The value of F1 is 98.8, the Recall value is 99.1, and the importance of precision gaining is 98.5 in this graph. The test data consists of thirty percent, in which our accuracy improvement in LG is 99.1 percent. This is comparable to our three matrices. This algorithm is not going to be the basis for our next step. It performs far better than earlier algorithms. However, it does not improve the accuracy of the height.
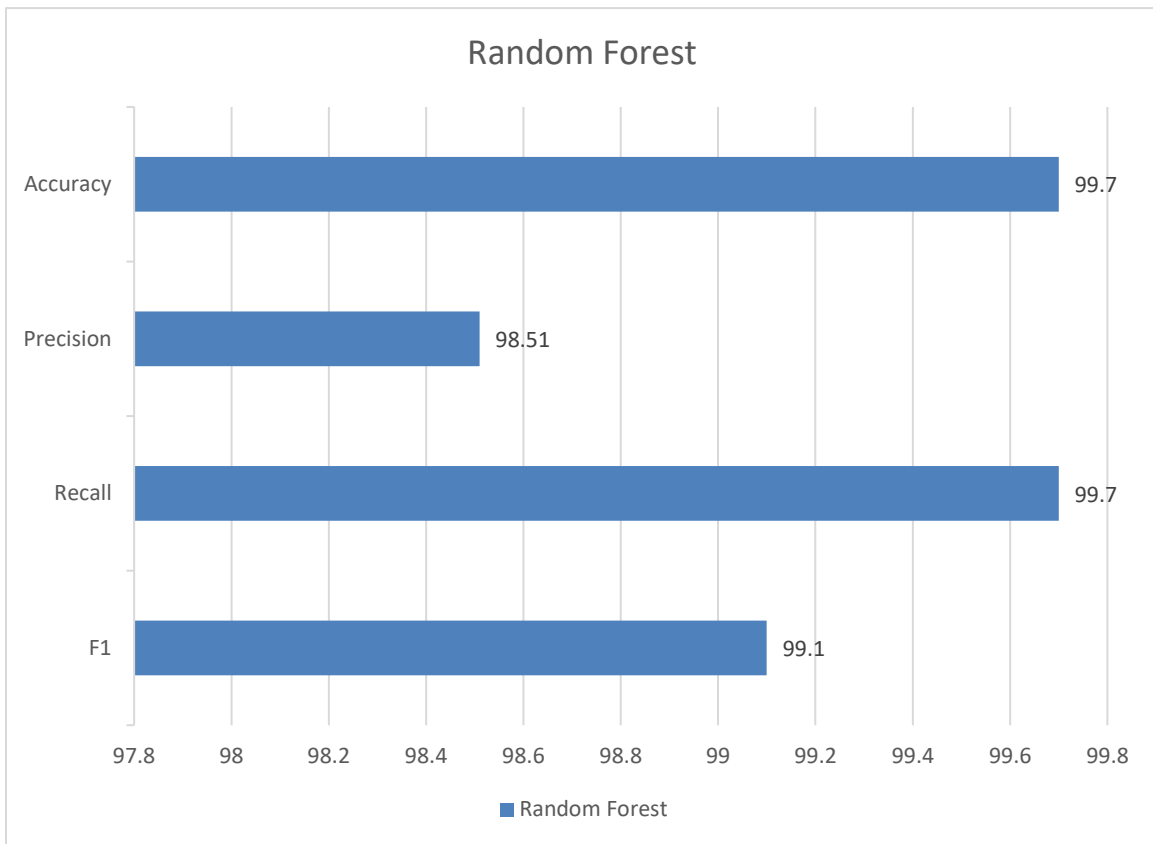
Random Forest

Figure 4.3: Show the accuracy of the Random Forest

Figure 4.3 provides a visual representation of how accurate the RF method is. In this part of the article, we are going to analyze the precision of three different matrices. This graph reveals that F1 has a value of 99.1, that Recall has a value of 99.7, and that the importance of precision gaining is 98.51. The data from the tests make up thirty percent, and our accuracy has improved by 99.7 percent since the beginning of the test. This is analogous to the three matrices that we have. We have decided to use this algorithm as the foundation for the next step in our process. It performs significantly better than any of the conventional machine learning methods that have been tried. On the other hand, it does make the highest degree of accuracy better.

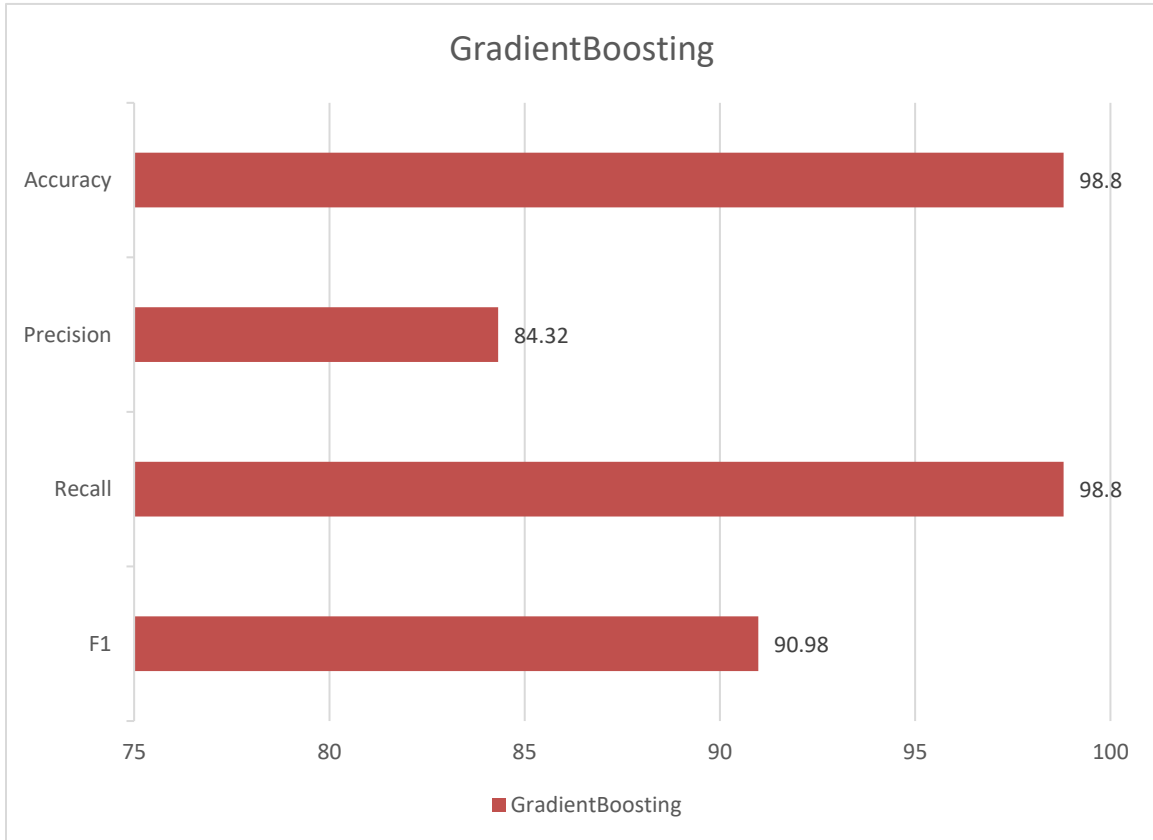Now apply boosting algorithm to compare our traditional algorithms.

GradientBoosting

Figure 4.4: Show the accuracy of the GradientBoosting

Figure 4.4 depicts how accurate the GradientBoosting method is in its application. In this part of the chapter, we will examine the correctness of three different matrices. This graph shows that F1 has a value of 90.98, that Recall has a value of 98.8, and that the relevance of precision gaining is 84.32. The percentage of test data is thirty percent, and the accuracy improvement achieved with the use of GradientBoosting is 98.8 percent. This is analogous to the way that we have three matrices. The following step that we do is not going to be based on this algorithm.
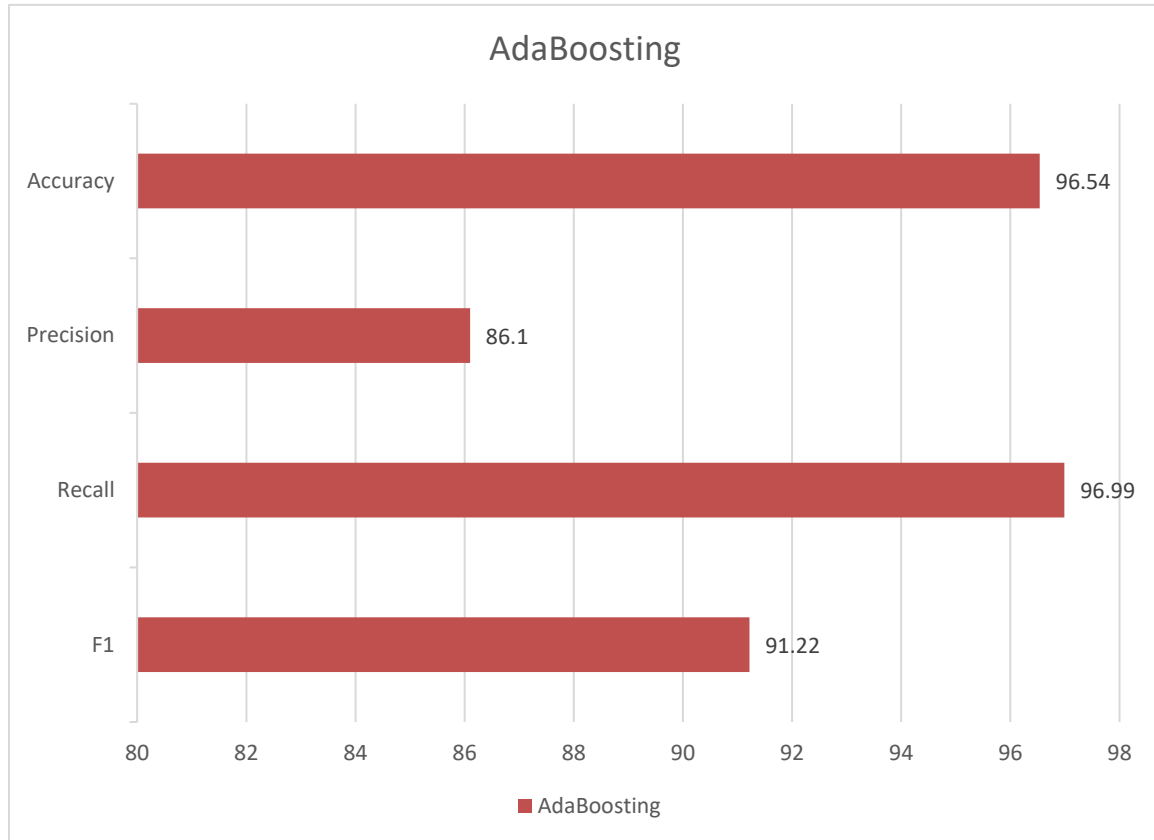
AdaBoosting

Figure 4.5: Show the accuracy of the AdaBoosting

The degree of precision that the AdaBoosting method exhibits when it is put into practice is illustrated in figure 4.5. In this section of the textbook, we are going to investigate the validity of three different matrices. It can be seen from this graph that the value of F1 is 91.22, the value of Recall is 96.99, and the relevance of precision gaining is 86.1. The percentage of data that was tested is thirty percent, and the improvement in accuracy that was obtained with the application of AdaBoosting is 96.54 percent. This is comparable to the manner in which we make use of three matrices. The next step that we do is not going to be predicated on this algorithm in any way, shape, or form.
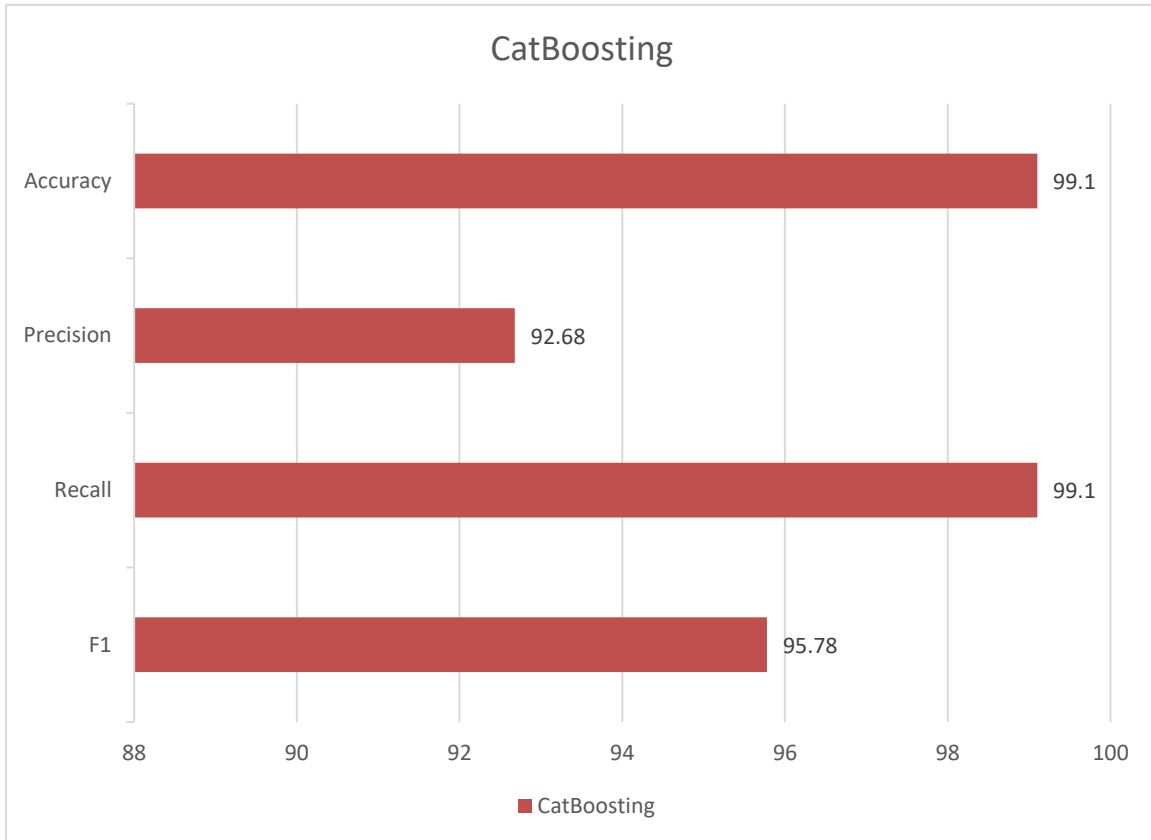
CatBoosting

Figure 4.6: Show the accuracy of the CatBoosting

The level of precision that the CatBoosting approach achieves is illustrated graphically in Figure 4.6. In the following section of the paper, we are going to investigate the precision of three distinct matrices. The information presented in this graph shows that F1 has a value of 92.68, that Recall has a value of 99.1, and that the importance of precision gaining is 95.78. Our accuracy has increased by 99.1 percent since the commencement of the test, and the data from the tests account for thirty percent of the total. This is comparable to the three matrices that we have available. Following careful consideration, we have reached the conclusion that the following stage of our procedure will be built upon the aforementioned algorithm. It outperforms every other approach of boosting categorization that has been investigated by a significant margin. On the other hand, the highest degree of precision is improved as a result of this.

4.3 Discussion

As a result of the fact that this is essentially a multiclass problem, the classifier produced a confusion matrix. The confusion matrix mentioned above is used to determine the accuracy, F1, precision, and recall scores for this study. We discovered Traditional Random forest algorithm to be the best classification method after applying GradientBoosting, AdaBoosting, CatBoosting, Logistic Regression, Decision Tree and Random Forest algorithm.

CHAPTER 5
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

In our country Bangladesh, online food delivery website are increasing day by day. So, reviews plays a vital role in the online marketplace in general. During last pandemic different online shopping became the main platform for shopping as it is the safest way. So, more online product and food service providers makes it better for people but also raises the question of product quality and services. So it is easy for new customers to get scammed while ordering. In our work we solve this issue properly and it will make better impact on our society.

5.2 Impact on Environment

In our work we made a system that will analyze the customer feedback from online food delivery and provide a ratio of the positive and negative feedback written in Bangla language. In this case, if anyone ordering online and they implement ore using our system then they get better food. So in our work balanced society and environmental impact.

5.3 Ethical Aspects

Some online product and food service providers make it better for people but also raises the question of product quality and services. When customers and food services providers see our system then food services providers delivered better product and customers get better food. However, they are also maintaining ethical aspects properly.

5.4 Sustainability Plan

In last pandemic different online shopping became the main platform for shopping as it is the safest and easiest way. It also accelerated the business to come online. So, in future we want to work with large dataset and make it more sustainable.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLECATION FOR FUTURE RESEARCH

6.1 Summary of The Study

Making progress toward a machine learning-based method for identifying instances of Food Review is a significant focus of our research. When the random forest method was applied to both the validation data and the test data, it produced very good results, according to our observations and findings. By strictly adhering to all of the criteria of data preparation and hyperparameter adjustment, we were able to achieve this degree of precision in our calculations.

6.2 Conclusion

Within the scope of this study, a total of five distinct machine-learning approaches were used to identify instances of agoraphobia. In order to gather information, we will be conducting an online survey. The DT, RF, LG, and Boosting classification methods are some of the classification algorithms that are employed the most commonly. There are numerous different approaches to classification that can be utilized. DT The split of the data was as follows: 70% training data and 30% test data. According to the results of the tests, the random forest had the highest accuracy, coming in at 99.7%. This was the level of accuracy achieved by our tests. The validation was carried out in an extremely accurate manner. It makes exactly one mistake, despite the fact that there are 58 different bits of data that it cannot predict. The domain of negative detection for agoraphobia is one in which our method performs remarkably well. In order to reach this decision, we employed the use of a confusion matrix.

6.3 Implication For Further Study

➢ In the not-too-distant future, we are interested in working with a more comprehensive dataset.

➢ In the not-too-distant future, we are going to implement the use of a dataset that receives real-time updates. This indicates that the dataset will be updated on a continuous basis and that our model will be trained in real-time as it is used.

➢ We are going to create an application that is compatible with both the web and Android.

➢ In addition to this, we intend to make available an application programming interface (API) that can be integrated into any platform, such as a website dedicated to telemedicine or a mobile health care program.

➢ Given that our study is predicated on machine learning, we plan to incorporate a deep learning method in the not-too-distant future in order to enhance the utility of our investigation.

APPENDIX

The first was to outline the procedures for the analysis, which presented a number of difficulties. The report was the first. Furthermore, no progress has been made in this area previously. Indeed. It wasn't your typical job. We couldn't find someone who could help us that much. Another stumbling block was data collection, which proved to be a huge issue for us. We created a data gathering corpus because we couldn't locate an open source Bangladesh text pre-processing program. We've begun manually collecting data. Furthermore, classifying the various postings is a difficult task. We might be able to achieve it after a lengthy time of hard labor.

REFERENCES

[1].   Hossain, Mohammad Didar, et al. "Mental disorders in Bangladesh: a systematic review." BMC psychiatry 14.1 (2014): 1-8.

[2].   B. Nithya and V. Ilango, "Predictive analytics in health care using machine learning tools and techniques," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 492-499

[3].   Choudhury, Munmun De et al. "Predicting Depression via Social Media." ICWSM (2013).

[4].   Anu Priya, Shruti Garg, Neha Prerna Tigga, Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms, Procedia Computer Science, Volume 167, 2020,Pages 1258-1267

[5].   Arif M, Basri A, Melibari G, et al. (2020) Classification of Anxiety Disorders using Machine Learning Methods: A Literature Review. Insights Biomed Res 4(1):95-110.

[6].   Vijai, C; Wisetsri, Workamol. Rise of Artificial Intelligence in Healthcare Startups in India Vol. 14, Iss. 1, (Mar 2021): 48-52.

[7].   M. Riajuliislam, K. Z. Rahim and A. Mahmud, "Prediction of Thyroid Disease(Hypothyroid) in Early Stage Using Feature Selection and Classification Techniques," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021, pp. 60-64

[8].   T. Nguyen, D. Phung, B. Dao, S. Venkatesh and M. Berk, "Affective and Content Analysis of Online Depression Communities," in IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 217-226, 1 July-Sept. 2014.

[9].   R. Razavi-Far, E. Hallaji, M. Saif and L. Rueda, "A Hybrid Scheme for Fault Diagnosis with Partially Labeled Sets of Observations," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 2017, pp. 61-67

[10].   K-Nearest Neighbor(KNN) Algorithm for Machine Learning, available at << https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning >> last accessed on 08-09-2021 at 8PM

[11].   K. Taunk, S. De, S. Verma and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 1255-1260

[12].   Sri Krishnan, 6 - Machine learning for biomedical signal analysis, Editor(s): Sri Krishnan, Biomedical Signal Analysis for Connected Healthcare, Academic Press, 2021, Pages 223-264

[13]. JOUR , Lv, Zhihan, Zhang, Zhifei, Zhao, Zijian, Yeom, Doo-Seoung, Decision Tree Algorithm-Based Model and Computer Simulation for Evaluating the Effectiveness of Physical Education in Universities , 2020, SN - 1076-2787

[14]. Sarica A, Cerasa A and Quattrone A (2017) Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. Front. Aging Neurosci. 9:329.

[15]. M. Mehedi Hasan, M. Omar Faruk, B. Biswas Biki, M. Riajuliislam, K. Alam and S. Farjana Shetu, "Prediction of Pneumonia Disease of Newborn Baby Based on Statistical Analysis of Maternal Condition Using Machine Learning Approach," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 919-924

[16]. Sarang Narkhede, Understanding AUC - ROC Curve , available at << https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5 >> , last accessed on 05-07-2021 at 1 PM.

[17]. Simple guide to confusion matrix terminology, available at << https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/ >> March 25, 2014. last accessed on 06-07-2021 at 8AM.

PLAGIARISM REPORT

food review

ORIGINALITY REPORT

24%	22%	11%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	9%
2	Submitted to Daffodil International University Student Paper	5%
3	Israt Jahan Intia, Md. Mehedi Hasan, Khairul Alam, Khondoker Sangida Ferdous. "Prediction of Agoraphobia Disease Based on Machine Learning", 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2022 Publication	2%
4	Submitted to Jacksonville University Student Paper	2%
5	www.scilit.net Internet Source	1%
6	scholar.uwindsor.ca Internet Source	<1%
7	Submitted to Anna University Student Paper	<1%