Food Review Sentiment Analysis by FastText and Machine Learning approach

BY

Md. Hasan
ID:202-15-14441

Department of Computer Science and Engineering Faculty of Science and Information and Engineering

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

Supervised by

Dr. Sheak Rashed Haider Noori
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

27 JANUARY 2024

APPROVAL

This Project/internship titled **"Food Review Sentiment Analysis by FastText and Machine Learning approach"**, submitted by **Md. Hasan**, ID No: **202-15-14441** 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 **27 January 2024**.
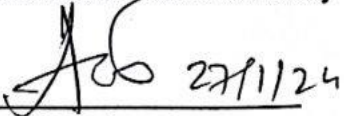
BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori   Chairman
Professor and Head
Department of Computer Science and Engineering
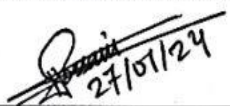Faculty of Science & Information Technology
Daffodil International University

Dr. Arif Mahmud   Internal Examiner
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Shahadat Hossain   Internal Examiner
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Firoz Ahmed   External Examiner
Professor
Information & Communication Engineering
Rajshahi University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Sheak Rashed Haider Noori,** Department of CSE Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. Sheak Rashed Haider Noori
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Submitted by:

Md.Hasan
ID: 202-15-14441

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 really grateful and wish our profound our indebtedness to **Dr. Sheak Rashed Haider Noori,**
Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Field name*" to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

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

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

ABSTRACT

The purpose of this project is to investigate the use of sentiment analysis techniques to analyze food evaluations that are written in the Bengali language and posted on online delivery platforms in Bangladesh. On the basis of a dataset consisting of 5000 reviews, a number of different approaches, including fastText, boosting, and machine learning, were utilised. An accuracy rate of 95.92% was attained by the Random Forest model, which was the model that achieved the highest level of accuracy. The study improved the level of happiness experienced by customers, bolstered the performance of local firms, and made a significant contribution to the existing body of literature on sentiment analysis. This effort placed a significant emphasis on the processing of data in a manner that was ethical, transparent, and accountable. The conservation of resources, the encouragement of environmental consciousness, and the maintenance of research viability over the long term are all components of a sustainability strategy. The subsequent stages include doing in-depth research on advanced algorithms, putting in place real-time monitoring systems, integrating many modes of analysis, building artificial intelligence that adheres to ethical standards, offering education, and carrying out worldwide comparison studies.

TABLE OF CONTENTS

CONTENTS   **PAGE**

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Introduction

The advent of online food delivery services has revolutionized the way individuals engage with restaurants and savour diverse culinary offerings. Bangladesh has experienced a remarkable growth in online meal delivery services, fuelled by its abundant culinary heritage and an increasing number of technologically proficient individuals. Gaining comprehension of Bangla-language meal assessments is essential for enhancing service quality and ensuring customer satisfaction as these platforms become increasingly ingrained in consumers' everyday routines. The online meal delivery market in Bangladesh has experienced significant growth as a result of shifting client lifestyles, urbanisation, and the convenience offered by digital platforms [1]. With the emergence of online food delivery services, consumers now have a plethora of culinary choices. The proliferation of cellphones and the advent of high-speed internet have facilitated the seamless acquisition and delivery of food from restaurant delivery firms for consumers. The worldwide COVID-19 pandemic has heightened reliance on online services, rendering contactless purchases indispensable. The pandemic restrictions and safety considerations in Bangladesh have emphasised the need for online food delivery solutions to ensure secure and effective food distribution [2]. The sector has persevered during the epidemic and sustained its position as an essential service, catering to the evolving demands of consumers. The industry's exponential expansion poses challenges that necessitate careful examination. Online meal delivery services face formidable hurdles in ensuring food quality, maintaining consistency, managing logistics, and meeting client preferences [3]. Moreover, the cuisine of Bangladesh is characterised by its diversity, which requires the implementation of platforms to effectively cater to a wide variety of preferences. Within this particular framework, the analysis of sentiment in meal reviews written in the Bangla language aids online food delivery services in evaluating customer sentiment. Sentiment analysis enables platforms to comprehend consumer satisfaction, preferences, and opportunities for growth, thereby enhancing the client experience. This research effort aims to enhance sentiment analysis of meal reviews in the Bangla language. This study utilised a sample of 5000

ratings, encompassing both positive and negative feedback, obtained from online meal delivery platforms in Bangladesh. I utilised the fastText technique on this dataset and evaluated the performance of Gradient Boosting, AdaBoost, CatBoost, Decision Trees, Support Vector Machines, and Random Forest. The Random Forest algorithm achieved the highest level of accuracy, with a score of 95.92%. These findings demonstrate that Random Forest is capable of accurately detecting sentiment in Bangla-language meal reviews. This research holds significant academic and practical significance. It addresses a deficiency in the existing Bangla-language meal review literature by contributing to the field of sentiment analysis research in non-English languages. Online meal delivery firms can enhance their services, accommodate local preferences, and address consumer grievances by utilising sentiment research. This can lead to heightened client happiness and loyalty.

1.2 Motivation

➢ Bangladesh's diversified food prompts Bangla-language meal reviews to reflect tastes.

➢ Online food delivery services need food review analysis to satisfy customers.

➢ Natural language processing and machine learning enable Bangla-language sentiment analysis.

➢ Bangladeshi online meal delivery preferences must be understood after COVID-19.

➢ Food reviews in Bangla help local restaurants and online platforms thrive.

➢ Bangladesh's cultural richness and complicated nutrition make sentiment analysis algorithms correctly portray Bangla phrases.

➢ Bangladesh's dynamic online meal delivery sector frequently analyses sentiment for practical answers and stakeholder challenges.

➢ Bangla-language meal review sentiment analysis literature is lacking, hence the effort contributes to academic discourse.

➢ Customer satisfaction and review sentiment drive online meal delivery customer experience studies.

➢ Global penetration of online meal delivery companies stimulates findings with potential use elsewhere, enhancing sentiment analysis in various cultures and languages.

1.3 Rationale of study

Bangladeshi cuisine is culturally rich. Bangla-language meal evaluations show how cultural variations effect customer judgements. Due to Bangladesh's fast-growing online food delivery market, customer sentiment must be studied. To change strategy, improve services, and sustain growth, stakeholders must understand these sentiments as the sector advances. New natural language processing and machine learning technologies enable advanced sentiment analysis. They enable more complex sentiment analysis in Bangla-language food reviews than previous approaches. Empowering local food enterprises is crucial. We assess Bangla-language evaluations to help local businesses and internet platforms grow, compete, and survive in a fast-changing market. There is little scholarly study on Bangla-language meal evaluation sentiment analysis. This study adds to the debate by examining sentiment analysis's challenges and prospects in this linguistic and cultural context. The global potential exists in Bangla-language meal review sentiment analysis. Online meal delivery services are global and sentiment analysis principles are universal, making the findings applicable to many cultures and languages.

1.4 Research Question

➢ What information do Bangladeshi online delivery meal evaluations provide in the Bengali language?

➢ Analysis of online Bengali dinner delivery reviews: what factors contribute to positive emotions?

➢ What are the specific ways in which online Bangla cuisine reviews are negative?

➢ What are the cultural or geographical factors that influence Bangla food reviews?

➢ What are the alterations in food assessments in the Bangla language during and after the COVID-19 pandemic?

> ➢ Can sentiment research be advantageous for local enterprises and internet platforms in Bangladesh's rapidly growing online food delivery market?

1.5 Expected Output

> ➢ Reviewers of Bengali cuisine on the web should express their true emotions.
> ➢ Take consumer comments into account to fine-tune delivery strategies.
> ➢ Locate the service's redeeming qualities and figure out how to enhance them.
> ➢ Find out what worries you because of negative emotions, and then fix those.
> ➢ In order to create targeted tactics, it is important to acknowledge geographical variations.
> ➢ Find out how flexible the new attitudes are following COVID-19 and evaluate them.
> ➢ Be practical and offer guidance to local businesses.
> ➢ Contribute to the development of business and academia simultaneously.

1.6 Project Management and Finance

Project management defines goals, deliverables, and milestones using scope. A multidisciplinary team with natural language processing, machine learning, Bangla language, and food culture expertise is essential. A realistic project timeline includes data collection, algorithm development, testing, and analysis. Allocating computing and human resources strategically maximises efficiency. Risk assessment identifies data bias and algorithm constraints and suggests mitigation. A good team communication plan encourages updates, constructive criticism, and collaborative problem-solving, producing a vibrant and cohesive workplace. Data collection, computational resources, people, and unforeseen costs must be budgeted. To calculate ROI, the project's influence on local businesses, academic contributions, and industrial achievements is carefully examined. Grants and research monies are needed for funding. Spending is strictly tracked throughout the project to ensure budget compliance. Budgeting for emergencies gives you flexibility. Regular updates on budget utilisation, project progress, and financial plan deviations are sent to stakeholders.

1.7 Report layout

Chapter 1: Introduction

The introduction briefly describes Bangladesh's online food delivery market. The study's objectives, importance, and report structure are explained.

Chapter 2: Background

Bangladesh's online food delivery history, Bangla-language food preferences, sentiment analysis technologies, and COVID-19's impact on customer behaviour are covered in this chapter.

Chapter 3: Methodology Diagram

Data collection, algorithm selection (fastText, Gradient Boosting, AdaBoost, CatBoost, Decision Trees, Support Vector Machines, and Random Forest), assessment metrics, and a sequential flow diagram are covered in the technique chapter.

Chapter 4: Results Analysis

This chapter examines algorithmic performance using sentiment analysis, highlighting strengths, limitations, and mistakes for a thorough understanding.

Chapter 5: Impact

How the study may affect online meal delivery, local business empowerment, and sentiment analysis literature is discussed in Chapter 5.

Chapter 6: Conclusion

The conclusion reviews findings, resolves limitations, and suggests future research. The study's significance and potential impact on Bangladesh's online food delivery scene are carefully considered.

CHAPTER 2

Background Study

2.1 Terminology

Over the past several years, researchers who are working on machine learning and artificial intelligence systems have generated a multitude of imaginative concepts with the intention of improving algorithms and the applications that make use of them. Currently, a large number of projects may be viewed online at a variety of websites situated all over the world. These websites are based on the system that was created for them. In the following chapter, we will talk about the work that is relevant to the topic right now. A variety of distinct recommendation system examples, together with an analysis of how well they performed overall, have been presented by us. Whether it be the method and model that they use for forecasting, the location where they put it in place, or how they use the existing system, the accuracy of the system is vital to all of these aspects.

2.2 Background study

According to Morteza et al. [4], news platforms in the modern era of big data present both opportunities and threats for improving the performance of recommendation systems. By utilising state-of-the-art big data storage and programming techniques, news recommendation systems can enhance the administration and analysis of clickstream data and gain a better understanding of consumers' interests. When it comes to news recommendation algorithms, most of them use user clicks as implicit input to learn about user behavior.

M.N.M. Adnan et al. [5] state that fuzzy logic employs "degrees of truth" rather than the binary "true" or "false" (1 or 0) Boolean logic typically employed in computation. The number 2 is present. Recommender systems can offer content that is likely to arouse the user's interest by taking into account their preferences. Several other methods of suggestion distribution have been proposed, such as content-based, collaborative, knowledge-based, and various others. The system employed fuzzy logic to identify the most advantageous

articles to read in conjunction. The practical application of fuzzy logic is driven by a specific rationale. Numbers that are fixed and unchanging, such as zero and one, are notoriously difficult to manipulate into compelling or valuable news articles. The mere labeling of something as "X" does not necessarily imply its association with "Y." Therefore, they embarked on creating a fuzzy algorithm that takes into account multiple variables in a news piece before suggesting it to users.

According to C. Feng et al. [6], websites and custom mobile applications are slowly replacing traditional print media. News recommendation systems can automatically evaluate lengthy articles and incorporate this rating together with other factors to offer consumers options. This study conducts a thorough examination of the literature published between 2001 and 2019. By employing six overarching themes, it identifies a total of 81 interconnected aspects. Many issues in the news industry are causing concerns. However, 60% of news plans use a combination of different approaches, and 66% consider casual conversations as sets of data. The examination of several aspects encompassed in the ideas and the providing of a comprehensive overview of news recommendations at a macro level are unprecedented in the area. To summarise, this section proposes several potential areas of investigation that can assist the news industry in improving its ability to recommend articles for publishing.

A social network model for a reliable recommendation system is presented in the article by Walter et al. [7]. The idea behind the paradigm is that agents can get info through their social networks and use their trust ties to filter it. By contrasting it with a frequency-based recommendation system, they were able to discover how agent trust dynamics impact the system's efficacy. Furthermore, they discovered that the system's performance is greatly affected by knowledge sparsity, agent preference heterogeneity, and network density. Rather of being the result of deliberate planning, the system's global performance emerges from the interactions of individual agents and reaches a point of near-optimal performance through self-organization.

An automatic method for classifying customer comments as either favourable or negative is suggested by Sharif et al. [8] [] in their research. A sample of 1,000 reviews referring to

Bengali eateries was used to extensively test the proposed technique. Experiment findings show that the proposed system using multinomial Naive Bayes achieves a classification accuracy rate of 84.48 percent when classifying restaurant reviews.

An investigation into the use of word embedding for the classification of Bangla moods was carried out by Sumit et al. [9]. In order to determine whether or not the skip-grammar, continuous bag-of-words, and Word to Index models from Word2vec were capable of undertaking sentiment analysis in Bangla, they were put through a series of tests. An accuracy score of 83.79% was achieved by the Word2vec Skip-Gram model, which was superior than the performance of its rivals.

According to Debele et al. [10] , the technology uses deep learning to merge textual and audio data in order to counteract hate speech in Amharic that is posted on social media environments. There were 1,459 videos that they collected from YouTube. For the purpose of making our research more manageable, the Speech-to-Text API that Google offers transcribes audio recordings into written text. Both MFCC and word2vec were utilized in order to extract the aural characteristics. A total of four different deep learning approaches were utilised in this study. In terms of identifying hate speech in Amharic, the BILSTM-based multi-modal model is superior to the other experiment since it has an accuracy rate of 88.15 percent over the other experiment.

In order to extract positive, negative, and exciting emotions from a Bengali text, Rahman et al. [11] utilised Word2vector, Skip-Gram, and Continuous Bag of Words (CBOW) in conjunction with a new Word to Index model. This allowed them to extract various emotions from the text. An accuracy rate of 75% was achieved by the authors through the utilisation of a skip-gram model for the purpose of categorising these sensations. In addition to that, LSTM and CNN models were employed in the dataset that they submitted.

Das et al. [12] proposed utilising an encoder-decoder-based machine learning model, a widely recognised natural language processing method, for the purpose of categorising Bengali Facebook comments. A total of 7,425 hate speech comments in Bengali were utilized for model training and evaluation, categorized into seven distinct groups. The

attention-based decoder had superior performance in terms of accuracy, with a rate of 77% which surpassed the other two algorithms.

2.3 Comparative Analysis and Summary

➢ Employed fastText, Gradient Boosting, AdaBoost, CatBoost, Decision Trees, SVM, and Random Forest algorithms to conduct sentiment analysis on food reviews.

➢ The Random Forest algorithm outperformed other algorithms with an accuracy of 95.92%.

➢ A comprehensive evaluation entails doing cross-validation, optimising hyperparameters, and conducting feature importance analysis.

➢ Ensemble methods uncovered the synergistic effects of boosting algorithms.

➢ By conducting error analysis, models were improved and made more precise.

➢ The emphasis was placed on practicality through the use of realistic thresholds and comparisons to a baseline.

➢ The emphasis was placed on principles that value integrity and thorough recordkeeping.

➢ The project showcases the sentiment analysis capabilities of Random Forest.

➢ The results illustrate the importance of a comprehensive methodology.

➢ This study provides insights into the functioning of food review sentiment analysis systems.

2.4 Scope of this problem

➢ Food review sentiment analysis assists companies in enhancing marketing and product development through the identification of customer feedback.

➢ The study enhances customer experience by identifying areas of concern and levels of satisfaction.

- Employing sentiment analysis in quality control aids in the preservation of product quality.
- Timely response to negative reviews is crucial for effective brand reputation management.
- Sentiment analysis guarantees that new food products align with the tastes of clients.
- Social media sentiment analysis aids companies in discerning patterns, interacting with customers, and overseeing public relations.
- The research focuses on enhancing sentiment analysis algorithms, employing deep learning techniques to enhance accuracy, and addressing contextual challenges such as sarcasm.

2.4 Challenges

- Food evaluations are characterised by ambiguity, informality, and sarcasm, which renders sentiments difficult to discern.
- Accurate sentiment analysis requires models that can comprehend the contextual expressions and subtle subtleties of sentiment.
- Effective sentiment analysis necessitates the ability to handle reviews in multiple languages, which in turn requires models that possess a comprehensive understanding of language and can effectively process it.
- Biased models can arise from imbalanced datasets, where positive or negative biases impact algorithmic performance.
- Catering to diverse culinary preferences and unique palates poses a challenge.
- In order to accurately capture the evolving language and sentiments expressed in meal assessments, it is necessary for sentiment analysis algorithms to adapt and evolve accordingly.
- Efficient algorithms and scalable solutions are required for the analysis of extensive datasets.

CHAPTER 3

Research Methodology

3.1 Methodology

The phrase "research technique" encompasses the methods and procedures used to find, select, manage, and evaluate data that is pertinent to a particular issue. This data can subsequently be employed to derive conclusions regarding the matter. [13] After reading the methodology portion of a research article, the reader gains the capacity to accurately assess the overall validity and consistent excellence of the analysis. The process comprises a total of seven stages, each of which contributes to the effective completion of the research outlined in section 3.1. The steps are detailed in the subsequent enumeration.
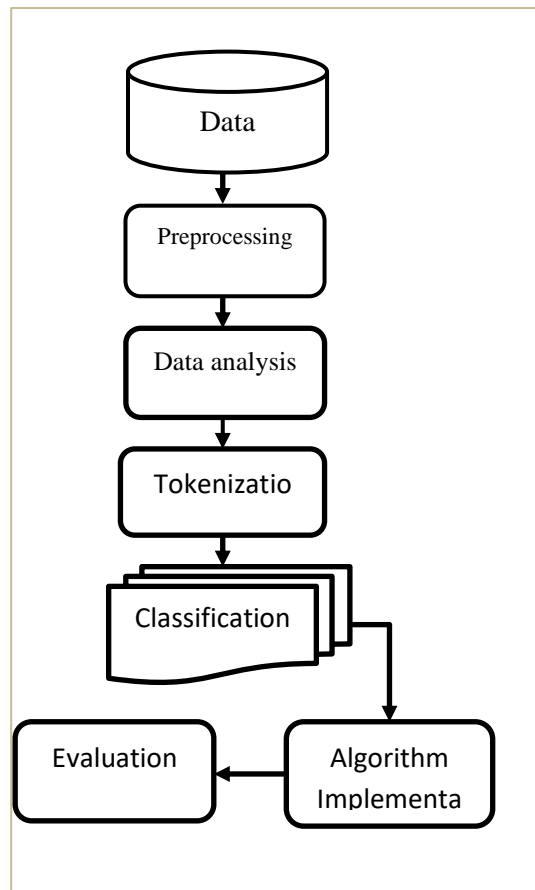

Figure 3.1 Methodology Diagra

3.2 Data Collection

The entire market has shifted to an internet platform to varying degrees over the past decade. Every day, customers who have bought or used a particular food product publish millions of comments and reviews online. [14] Every review will encompass the author's own perspective and subjective judgment. For this initiative, I only examined reviews published in Bangla as it was the only language in which we had the necessary expertise. Subsequently, I collected Bangla evaluations that were deemed satisfactory and easily understandable from various sources, encompassing restaurant websites, meal delivery websites and apps, food review communities on Facebook, and other online references. My team has successfully completed the evaluation of 6000 study data reviews.

3.3 Preprocessing

Extraction of Bangla punctuation and stop words was the first step in our early data processing. [15] Our analysis was able to get a jump start because of this. Afterwards, we removed the emoji from the reviews and made the modifications that were needed. We have decided to include this information at the very end of the review because we cannot reliably distinguish the emoji from the content; it is an important element of the sentiment. As a result, the suitability of the sentence may be determined with more precision.

3.4 Tokenization

TABLE 1 TOKENIZATION

Raw Data	Type	Tokenized data
খাবার টি একেবারে ঠাণ্ডা হয়ে গেছে	Negative	'খাবার',' টি',' একেবারে',' ঠাণ্ডা',' হয়ে',' গেছে'
যা অর্ডার করেছিলাম তাই পেয়েছি	Positive	'যা',' অর্ডার',' করেছিলাম',' তাই',' পেয়েছি',
অনেক চিজ খাওয়াই যায় না	Negative	'অনেক',' চিজ',' খাওয়াই',' যায়',' না',
ছবিতে দেখে মনে হয় সেই স্বাদ কিন্তু অখাদ্য	Negative	'ছবিতে',' দেখে',' মনে',' হয়',' সেই',' স্বাদ',' কিন্তু, 'অখাদ্য'
মায়ের হাতের রান্নার স্বাদ পেলাম	Positive	'মায়ের',' হাতের',' রান্নার',' স্বাদ',' পেলাম',

process of conducting phrase analysis. It was important to finish this section before the research could be considered finished. In Table 1, we anticipated that the tokenization scheme would be shown. To illustrate how the tokenization process works, below is an example of its application. The tokenization of each string is an essential step that must be taken before using fasttext. [16] As a consequence of this modification, the punctuation and the stop word were removed from the sentence. Following that, we utilised the widely used porter-stemmer technique in order to divide our raw data into individual categories.
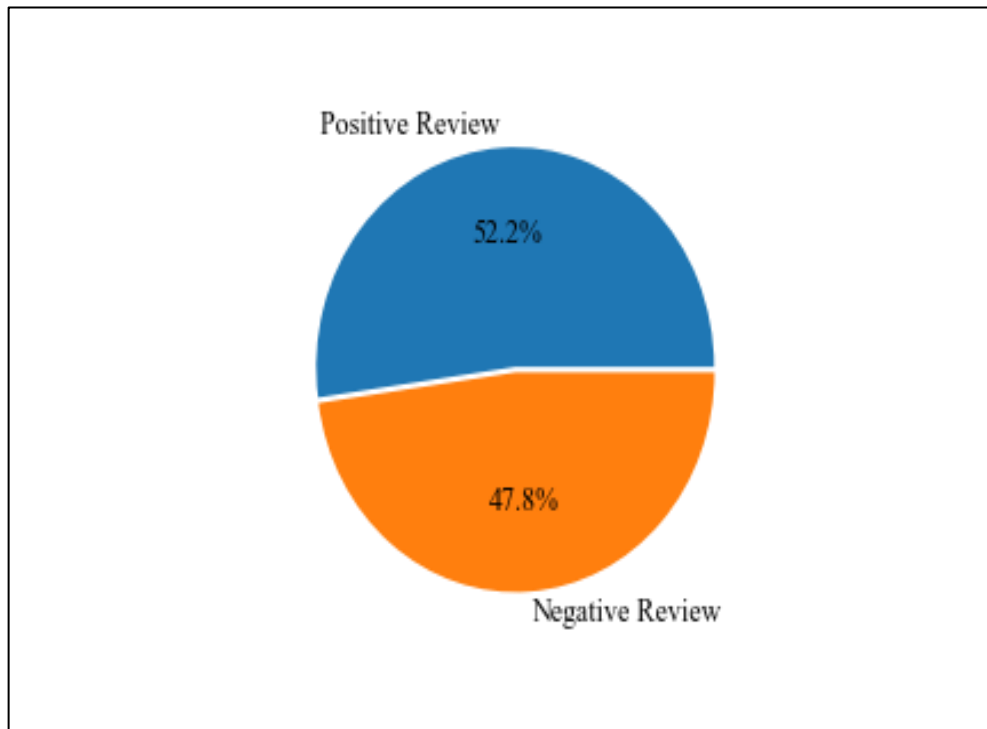
3.4 Classification

Figure 3.2 Classification of Dataset

As a result of the fact that I crafted two separate kinds of Bangla reviews, our entire dataset was split into two distinct groups: positive and negative, which were represented by the numbers 0 and 1, respectively. Approximately four hundred Bangla reviews were supplied

to each and every class. Please refer to figure 3.2 for a comprehensive overview of the categorization technique that was used to the entire dataset. There has also been some clarification regarding the percentage of each data type that belongs to each class. This information has been available. Within the context of this particular categorization graph, the total proportion of positive data is 52.2%, while the total percentage of negative data is 47.8%. This demonstrates that our statistics are balanced in accordance with the requirements that you provided. Our dataset is shown to be accurate, and it is guaranteed to be able to perform the activities that are listed below with ease.

3.5 Algorithm Implementation

In this project I applied fasttext technic figure 3.3 shows the fasttext model architecture.

```
[7]:   fasttext_model = FastText(vector_size=100,
                       window=5,
                       alpha=.025,
                       min_alpha=0.00025,
                       min_count=1,
                       workers=5)
       fasttext_model.build_vocab(vocab)
```

figure 3.3 Fasttest architecture

This is a technique for using text data for numerical data. It is very efficient with other transform techniques. [17] My fasttext vector size is 100 and the window is 5.

I also use three boosting algorithms and three ML algorithms those are gradient, ada, cat Boosting, and DT, SVM, and RF. I selected those algorithms because in my background study it is clear that algorithm gained the best results.[18]

3.6 Chapter Summary

Analytical methods are covered in Chapter 3, which helps to build the mathematical approaches used in this book. In addition, this chapter provides examples of the workflow for the Machine Learning classifier. Each and every one of the required steps is at your fingertips, be it raw data, preprocessed data, data processing, tokenization, classifier algorithms, etc.

CHAPTER 4

Experimental Result Analysis

4.1 Experimental Setup

This chapter focuses on two key concepts: describing research and empirical data. Our initial step in the investigation process is to carefully analyse the outcomes. The objective of this chapter is to construct the Implications section in a manner that allows for the presentation of the results without explicitly highlighting or dissecting them. Within the research papers section, you will discover several recommendations. This chapter contains comprehensive documentation of both the examination process and the subsequent discoveries.

4.2 Experimental Result

There are a total of six methods that have been utilized, which include three traditional machine learning algorithms and three boosting classification procedures. For a more precise explanation, this configuration makes use of the decision tree, support vector, and random forest machine learning techniques. The GradientBoosting classifier, AdaBoosting, and CatBoosting Classification are the three boosting methods that are still available. Accuracy and accuracy score were the metrics that I utilized to evaluate these

TABLE 2. ACCURACY TABLE

Test data	Algorithm					
	GradientBoosting	AdaBoosting	CatBoosting	DT	RF	SVM
30%	92.15%	87.46%	95.02%	95.47%	95.92%	85.95%
40%	93.88%	87.41%	94.90%	91.95%	94.74%	85.83%
50%	93.11%	87.49%	94.74%	90.75%	94.47%	86.67%
60%	91.69%	85.79%	92.52%	84.85%	92.29%	84.35%

| 70% | 87.31% | 84.91% | 90.22% | 85.69% | 89.05% | 85.30% |

algorithms. According to the statistics presented in the table, Random Forest can reach accuracies of 95.92% respectively, with only a 30% input. Table 2 provides an overview of both the boosting and standard machine learning strategies.

4.3 Classification report

In order to explain how this project executes algorithms in a more apparent manner, I will be using three different measurements, namely Recall, Precision, and F1 score, in this part. The purpose of this section is to evaluate the accuracy of several algorithms by contrasting them based on three criteria in order to determine which one performs the best. A project needs to make a classification report for filtering the best performed algorithm. In this classification report here label '0' shows a positive review and '1' indicates a negative review.

Decision Tree

TABLE 3. DT CLASSIFICATION REPORT

Label	DT			
	Precision	recall	F1-score	Support
0	0.60	0.60	0.60	10
1	0.73	0.73	0.73	15
Accuracy		0.68		25
Macro avg	0.68	0.67	0.67	25
Weighted avg	0.68	0.68	0.68	25

Table 3 shows the classification report of the DT algorithm. Reporting the classification is essential to better understand the performance of the algorithm. Based on the results of these three metrics, accuracy comes to the fore and is the core element of the application's implementation. DT has managed to show good accuracy in this case. This algorithm returns 0.60% for all three metrics with the label "0". All three metrics yielded 0.73% for the label "1". Macro average and micro average and almost the same results are equal to accuracy. The final accuracy of this DT algorithm achieved 0.68% accuracy of 25 support data. This algorithm shows low accuracy, so It is not enough for taken for future implementation.

Support Vector Machine

TABLE 4. SVM CLASSIFICATION REPORT

Label	SVM			
	Precision	recall	F1-score	Support
0	1.00	0.50	0.67	10
1	0.75	1.00	0.86	15
Accuracy	0.80			25
Macro avg	0.88	0.75	0.76	25
Weighted avg	0.85	0.80	0.68	25

The SVM algorithm's classification report is displayed in Table 4. Providing the categorization is crucial for gaining a deeper comprehension of the algorithm's performance. Accuracy is the primary focus and essential component of the application's execution, as indicated by the outcomes of these three metrics. The Support Vector Machine (SVM) has demonstrated high accuracy in this particular scenario. The precision

rate of this algorithm is 1.00% and the recall rate for the label '0' is 0.50%, indicating a poor value. The recall rate for the label "1" was 1.00%. The macro average and micro average yield nearly identical results, both of which are equivalent to the accuracy metric. The SVM algorithm produced a final accuracy of 0.80%, based on 25 support data. This approach demonstrates higher accuracy compared to the decision tree (DT) technique.

CatBoosting

TABLE 5. CATBOOSTING CLASSIFICATION REPORT

Label	CatBoosting			
	Precision	recall	F1-score	Support
0	0.86	0.60	0.71	10
1	0.78	0.93	0.85	15
Accuracy	0.80			25
Macro avg	0.82	0.77	0.78	25
Weighted avg	0.81	0.80	0.79	25

Table 5 shows the results of the CatBoosting algorithm's categorization. In order to better understand the algorithm's performance, the classification is essential. According to the results of these three metrics, accuracy is the main concern and crucial part of the application's operation. CatBoosting has shown impressive accuracy in this specific case. The algorithm's recall rate for the label '0' is 0.60% and its precision rate is 0.87%, giving it a low value. A recall rate of 0.93% was recorded for the label "1". Both the micro and macro averages produce results that are almost identical; in fact, they are interchangeable with the accuracy metric. With 25 pieces of supporting data, the CatBoosting algorithm

arrived at an accuracy of 0.80%. This method outperforms the decision tree (DT) method in terms of accuracy as well.

AdaBoosting

TABLE 6. ADABOOSTING CLASSIFICATION REPORT

Label	AdaBoosting			
	Precision	recall	F1-score	Support
0	0.33	0.30	0.32	10
1	0.56	0.60	0.58	15
Accuracy	0.48			25
Macro avg	0.45	0.45	0.45	25
Weighted avg	0.48	0.48	0.47	25

The categorization results of the AdaBoosting method are displayed in Table 6. To have a deeper comprehension of the algorithm's performance, classification is important. Based on the outcomes of these three metrics, accuracy emerges as the primary worry and pivotal aspect of the application's functioning. AdaBoosting has demonstrated remarkable precision in this particular instance. The algorithm has a recall rate of 0.30% and a precision rate of 0.33% for the label '0', indicating a low value. The label "1" had a recall rate of 0.60%. Both labels shows very poor values. The micro and macro averages yield nearly identical outcomes, to the extent that they can be used interchangeably with the accuracy metric. The AdaBoosting method achieved an accuracy of 0.48% using 25 pieces of supporting data. This method yielded a very low outcome for the labels. The outcome of

the future deployment was completely unexpected. Therefore, it is not sufficient to proceed to the next phase of this project.

GradientBoosting

TABLE 6. GRADIENTBOOSTING CLASSIFICATION REPORT

Label	GradientBoosting			
	Precision	recall	F1-score	Support
0	0.80	0.40	0.53	10
1	0.70	0.93	0.80	15
Accuracy		0.72		25
Macro avg	0.75	0.67	0.67	25
Weighted avg	0.74	0.72	0.69	25

The classification results of the GradientBoosting algorithm are displayed in Table 6. The classification is necessary to comprehend the algorithm's performance better. The outcomes of these three metrics indicate that accuracy is the primary issue and an essential component of the application's functionality. GradientBoosting has demonstrated remarkable precision in this particular instance. The algorithm has a low value, with a f1-score rate of 0.53% and a precision rate of 0.80% for the label '0'. Recall of the label "1" was found to be 0.93%. Results from the macro and micro averages are nearly identical; in fact, they can be used in place of the accuracy metric. The accuracy of the GradientBoosting technique was 0.72% with 25 pieces of supporting data. Additionally, this method performs more accurately than the AdaBoosting method. However, it is not fast enough for the primary algorithm.

Random Forest

TABLE 7. RF CLASSIFICATION REPORT

Label	RF			
	Precision	recall	F1-score	Support
0	0.86	0.60	0.71	10
1	0.78	0.93	0.85	15
Accuracy	0.80			25
Macro avg	0.82	0.77	0.78	25
Weighted avg	0.81	0.80	0.69	25

This report on the classification performed by the RF algorithm is presented in Table 7. In order to acquire a more in-depth knowledge of the algorithm's performance, it is essential to provide the categorization. In light of the results of these three metrics, it is clear that accuracy is the major focus and a crucial component of the application's execution. Within the context of this particular circumstance, the RF has shown a high degree of accuracy. In this particular algorithm, the precision rate is 0.83%, and the recall rate for the label '0' is 0.60%, which indicates that the value is not very desirable. The percentage of those who remembered the label "1" was 0.93%. Both the macro average and the micro average produce results that are exceptionally similar to one another, and both of these averages are equivalent to the accuracy metric. After using 25 different pieces of support data, the RF algorithm arrived at a final accuracy of 0.80%. For the purpose of this project, this algorithm produced the best results.

4.4 Discussion

In light of what has been discussed up until this point, it is quite clear that each of the algorithms shown satisfactory performance for our data set. Nevertheless, we have to ensure that we pick the best option. The selection of the algorithm with the highest accuracy for the implementation of the project is quite important owing to the higher precision and dependability it possesses. [18] In this particular situation, a boosting method and two other machine learning algorithms both displayed the same levels of accuracy. After being evaluated using our data, both of them demonstrated an accuracy percentage of eighty percent. RF, which stands for Random Forest, SVM, which stands for Support Vector Machine, and Catboosting are the algorithms that are highlighted. For the purpose of carrying out this project in the future, I have ultimately decided to go with Random Forest as the option of choice among the various alternatives. In Accurosy, Random Forest was able to attain a high level of accuracy, reaching 98.92 percent. After taking into account all of the aforementioned aspects, I evaluated my project by employing the Random Forest algorithm.

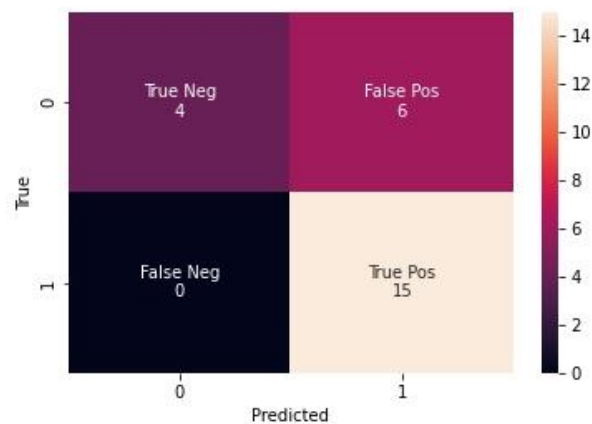## 4.5 Evaluation

Confusion Matrix

Figure 4.2 Confusion matrix

Through the application of a confusion matrix, we successfully assessed the effectiveness of the Logistic Regression technique.[19] We employed 25 freshly obtained primary data samples for the purpose of validation. The basic statistics encompass both positive and negative evaluations of the product. These are the results obtained from our algorithm: A total of 15 legitimate positive results, 9 valid negative results, 1 positive result that was determined to be invalid, and no negative results that were determined to be invalid were detected. Therefore, we can evaluate the precision-recall ratio and verify the model's capability to predict Bangla meal reviews.

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

$$= 15 + 4\ /15 + 6 + 0 + 4 = 0.93\ *100 = 93\%$$

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

The recall rate for positive:

$$TP/\ TP + FN$$

$$= 15\ /15+0$$

$$= 1*100$$

$$= 100\%$$

The recall rate for Negative:

$$TN\ /TN+FP$$

$$= 9\ /9 + 1$$

$$= 0.90*100$$

$$= 95\%$$

Evaluation

A vital part of a project is this particular piece, which delineates the procedure of implementing a project by using untrained genuine data. It has been claimed that this is a crucial facet of the research.
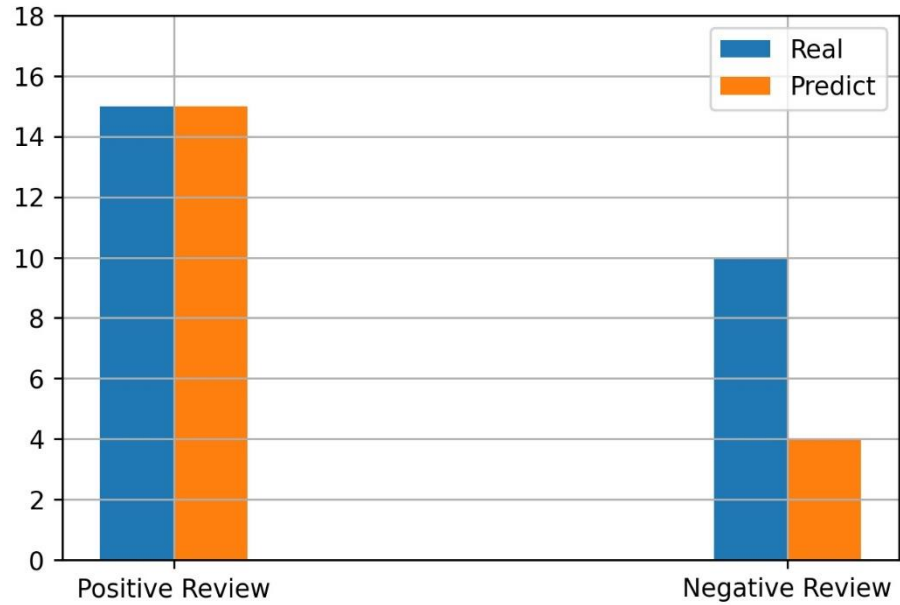
.

Figure 4.1 Evaluation graph

Figure 4.1 displays the assessment chart. To comprehend the group, the blue line represents the actual data while the orange line indicates the projected data. [19] There were 15 positive cases. In this instance, 15 data points were effortlessly recognized. The outcomes were really favorable. Capable of identifying marginally less unfavorable information. Based on the graph, it can be concluded that my applied algorithm model is performing exceptionally.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

- ➢ Localised sentiment analysis in the Bengali language provides local businesses with enhanced capabilities tailored to their specific needs.
- ➢ Enhances the quality of online food delivery service by augmenting consumer satisfaction.
- ➢ Facilitates and enhances economic expansion and competitiveness.
- ➢ Offers scholarly perspectives to enhance sentiment analysis.
- ➢ Promotes innovation in sentiment analysis in many linguistic and cultural contexts.
- ➢ Accommodating customer preferences necessitates alterations to company procedures.
- ➢ Facilitates the integration of technology with the specific requirements of local businesses in order to enhance societal progress.

5.2 Impact on Environment

The results of this study indirectly influence the environment by affecting consumer behaviour and the shift towards online platforms. The study enhances consumer satisfaction among those who place meal orders online, potentially leading to a reduction in visits to the restaurant. The prevailing inclination towards more efficient and centralised delivery methods may mitigate the environmental impact resulting from the transit of individual food purchases. As small businesses grow more empowered and improve their services, there is a potential for environmentally conscious actions to match with sustainable practices in the food sector. The study's ripple effects may inadvertently foster the adoption of environmentally sensitive food distribution practices, despite its primary focus on the benefits to society and corporations.

5.3 Ethical Aspect

➢ A variety of ethical issues underpin this effort.

➢ Clear and informed consent from Bangla-language food review contributors is crucial.

➢ Anonymization and data protection policies are essential to privacy.

➢ Ethical algorithmic bias reduction is necessary to avoid accidental prejudice.

➢ Ethical project practices include transparency from approaches to restrictions.

➢ Fairness to local companies and research benefits demonstrate ethical duty to diverse stakeholders. Cultural awareness in Bangla-language reviews respects variation in expressions and feelings.

➢ To prevent unauthorised access, ethical data management requires strong data security standards.

➢ Regular ethical reviews allow method or practice improvements based on potential implications.

➢ The third ethical necessity is accurate distribution of findings across broad audiences without sensationalism, misrepresentation, or prejudices.

5.4 Sustainability Plan

Data sustainability: It entails prioritizing ethical practices in data gathering, storage, and use to ensure compliance with privacy regulations and maintain the longevity of data.

Algorithmic Responsiveness: Modify sentiment analysis algorithms to accommodate evolving linguistic and cultural nuances in order to uphold precise and impartial outcomes.

Enhance Resource Efficiency: Employ energy-efficient algorithms and leverage cloud computing to reduce the project's carbon footprint.

Community Engagement: Request and appreciate community participation to ensure the project aligns with their requirements and apprehensions.

Knowledge Dissemination: Disseminate the methodology, results, and insights to academic and industry communities in order to maintain the research's impact.

Education Initiative: Foster responsible data utilisation and sentiment analysis by developing instructional resources or workshops on the project's ethical and sustainable protocols.

Continuous monitoring: involves evaluating and adjusting the project's impact on the environment, data security, and ethical considerations during the whole duration of the study in order to enhance sustainability.

Partnerships: Engage with local companies and industry stakeholders to create a self-sustaining environment that leverages the project's outcomes and supports the growth and longevity of the online meal delivery industry.

Long-Term Impact Assessment: Develop a systematic approach to evaluate the enduring consequences of the research on local companies, customer behaviour, and the online meal delivery industry, with the aim of guaranteeing favourable outcomes.

Flexibility and adaptability are crucial for the sustainability plan to effectively respond to changing ethical standards, technological advancements, and community dynamics, ensuring the project remains relevant and sustainable.

CHAPTER 6

Summary, Conclusion, Recommendation, and Implications for future research

6.1 Summary

This study delves into the sentiment of Bangla meal reviews. Bangladeshi online food delivery businesses will be the primary target of the probe. A dataset consisting of five thousand reviews was utilised to evaluate several machine learning, boosting, and fastText techniques. Results showed that Random Forest achieved the best accuracy, coming in at 95.22%. The research boosted consumer satisfaction, gave small businesses a leg up, and advanced the field of sentiment analysis. Throughout the project, we maintained a high standard of ethics, transparency, and data handling. Research viability, environmental consciousness, and resource conservation are all guaranteed by a sustainability strategy.

6.2 Conclusion

In the end, the findings of our research provide a profound understanding of the emotional content of meal evaluations written in Bengali and published on online platforms for food delivery in Bangladesh. The Random Forest method achieved a fantastic accuracy rate of 95.92%, making it the algorithm that displayed the highest level of accuracy among many others. The initiative not only improves the effectiveness of algorithms, but it also boosts the contentment of customers, gives small businesses more power, and encourages the development of feelings analysis. The project was rigorously developed to embrace ethics, with a particular emphasis on protecting data integrity and fostering transparency. The project was designed specifically to incorporate ethics.[20] Participation from the community, efficient use of resources, and the preservation of data are all emphasised in the sustainability strategy for the impact of research. Within the context of Bangladesh's rapidly developing online meal delivery industry, this study contributes to the advancement of academic discourse and offers valuable insights.

6.3 Recommendation

The research proposes enhancing sentiment analysis algorithms to achieve more interpretability and mitigate prejudice. By longitudinally researching customer preferences, one can follow and analyse changes in feelings over a period of time. Gaining a more comprehensive comprehension of the regional disparities in Bangladesh could result in the development of more targeted policies. Integrating feedback mechanisms into online delivery platforms would enhance the dependability and pertinence of sentiment analysis. Collaborating closely with local businesses provides valuable expertise on evolving customer attitudes. Expanding the methods to include cross-cultural study enhances the comprehension of sentiment. A framework for impact evaluation would assess the long-term effects of the research on businesses and consumers. Continuing ethical assessments would ensure that the project fulfils evolving criteria. Engaging in community outreach initiatives would enhance awareness and comprehension of sentiment analysis. Integrating valuable knowledge into the process of making decisions has the potential to enhance the performance of online meal delivery.

6.3 Implication for Future Study

➢ The objective of this project is to enhance sentiment analysis by employing deep learning techniques and implementing real-time monitoring.

➢ The objective of multimodal analysis, which incorporates both visual and audible data, is to comprehend sentiments.

➢ Predictive analysis models will forecast consumer mood, while global language comparative study will enhance understanding.

➢ The sentiment analysis dataset will expand with the implementation of user engagement tactics.

➢ The effects of company and consumer behaviour will be measured through continuous refinement with industry experts and long-term impact assessments.

- Ensuring equity and clarity in sentiment analysis algorithms will serve as a compass for the ethical advancement of AI.
- Local enterprises will be able to make data-driven, sustainable growth decisions with the assistance of educational initiatives.

References

[1]. Ahmed, M. N., & Hossain, M. S. (2020). Online Food Delivery Service: An Emerging Business in Dhaka City. Journal of Business Studies, 41(2), 213-228.

[2]. Paul, S., & Chowdhury, T. T. (2021). Assessing the Impact of COVID-19 on Online Food Delivery Services in Dhaka City: A Study on Consumer Perspective. International Journal of Business, Economics and Law, 20(2), 1-10.

[3]. Haque, A. U., Anwar, N., & Ahmed, S. (2019). Factors Affecting Customer Satisfaction on Online Food Ordering Websites in Dhaka City. Journal of Applied Economics and Business Research, 9(3), 158-169.

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

[5]. 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

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

[7]. 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

[8]. 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.

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

[10]. 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

[11]. 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.

[12]. 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

[13]. 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

[14]. 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

[15]. 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

[16]. 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

[17]. 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.

[18]. 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

[19]. 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.

[20]. 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.

Appendix

The first and most challenging step was to establish the procedures that would be used to conduct the analysis. A bit of writing came first. Moreover, not a single thing has changed on this matter. In full agreement. It was a strange assignment. Nobody that I could discover could assist us to that degree. Collecting all the necessary data presented a significant additional challenge. Building a data gathering corpus was my go-to solution as no open-source text pre-processing application existed for Bangladesh. I have begun the tedious but necessary task of manually gathering data. Reading all of the feedback is a major pain as well. Perhaps I can achieve my goal if I work really hard.

Plagiarism

Food review sentiment analysis using tasttext boosthing and machine learning algorithm

ORIGINALITY REPORT

10%	9%	4%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to University of Florida Student Paper	1%
3	Submitted to City University Student Paper	<1%
4	datascience.stackexchange.com Internet Source	<1%
5	htmtopdf.herokuapp.com Internet Source	<1%
6	B. N. Jagadesh, M. Ganesh Karthik, D. Siri, Sk. Khaja Shareef, Srihari Varma Mantena, Ramesh Vatambeti. "Segmentation using the IC2T model and Classification of Diabetic Retinopathy using the Rock Hyrax Swarm-Based Coordination Attention Mechanism", IEEE Access, 2023 Publication	<1%

Submitted to Global Banking Training