Sentiment Analysis of Customer Reviews on Food Ordering Portals of Bangladesh using Natural Language Processing

by

Priom Deb
23341092
Asibur Rahman Bhuiyan
23341095
Farhan Ahmed
20141015
Md. Rakib Hossain
20101315
Habiba Mahrin
20301339

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
September 2023

Declaration

It is hereby declared that

1. The thesis submitted is my/our original work while completing my degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material that has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Asibur Rahman Bhuiyan
23341095

Priom Deb
23341092

Md. Rakib Hossain
20101315

Farhan Ahmed
20141015

Habiba Mahrin
20301339

Approval

The thesis/project titled "Sentiment Analysis of Customer Reviews on Food Ordering Portals of Bangladesh Using Natural Language Processing" submitted by

1. Priom Deb (23341092)
2. Asibur Rahman Bhuiyan (23341095)
3. Farhan Ahmed (20141015)
4. Md. Rakib Hossain (20101315)
5. Habiba Mahrib (20301339)

of Fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on January 18, 2024.

Examining Committee:

Supervisor:
(Member)

Mr. Dewan Ziaul Karim
Lecturer
Department of Computer Science and Engineering
Brac University

Co-supervisor:
(Member)

Mr. Faisal Ahmed
Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, Ph.D.
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, Ph.D.
Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

In recent years, online food ordering services have gained popularity by providing customers with suitable and user-friendly platforms for ordering food from restaurants and receiving doorstep delivery. Foodpanda Bangladesh and HungryNaki have been anticipated to make significant contributions to the expansion and development of the online food delivery market during this period. This study aims to forecast the attitudes of Bangladeshi consumers toward digital platforms for food ordering, with a particular focus on Foodpanda Bangladesh and HungryNaki. To achieve this goal, an online review sentiment analysis will be implemented. A dataset of customer reviews from the company's website will be compiled. The data will undergo preprocessing techniques to filter out unnecessary and irrelevant information and refine the features and characteristics of the data. Subsequently, natural language processing (NLP) techniques will be applied to conduct sentiment analysis. The objective of this research is to determine the prevailing customer opinions regarding restaurants and food delivery platforms in Bangladesh. This includes their future assessments of delivery schedules, meal quality, and customer service on the platform. The results of this research should shed light on the future of Bangladesh's food-ordering portals from the perspective of their users. The research will help the platform enhance its reputation and competitiveness in the online food delivery market.

Keywords: Sentiment analysis, Restaurant reviews, Food ordering portal, Customer reviews, Data analysis, Neural Network, Rating.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis has been completed without any major interruption.
Secondly, to our supervisor Mr. Dewan Ziaul Karim sir, and our co-supervisor Mr. Faisal Ahmed sir for their kind support and advice in our work. They helped us whenever we needed help.
And finally to our parents since it may not be possible without their support. With their kind support and prayer, we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, restaurant food delivery applications have emerged as a transformative force in the culinary industry, fundamentally altering the way consumers experience dining. These digital platforms serve a crucial purpose in our fast-paced and convenience-oriented society, offering a seamless connection between consumers and their preferred restaurants.

Globally, there has been a significant increase in demand for and usage of food delivery applications in recent years. The primary objective of restaurant food delivery applications is straightforward: to provide customers with a hassle-free means of accessing a wide array of culinary offerings from the convenience of their homes, workplaces, or virtually any location with internet connectivity. Whether it involves ordering a general meal from a renowned restaurant, obtaining a quick and hearty takeout option, or even procuring a special treat for a celebratory occasion, these applications have become the preferred solution for fulfilling culinary desires. Typically, these applications offer consumers a convenient and cost-effective means of food delivery. Food is often ordered through online platforms, such as mobile applications and websites, with a commitment to timely delivery. A substantial number of people prefer making online payments for these services. In Bangladesh, the utilization of food delivery applications is experiencing considerable popularity. Notably, Foodpanda, HungryNaki, PathaoFood, and ShohozFood are emerging as the most renowned applications, garnering extensive recognition across the nation.

Several compelling factors underpin the increasing demand for online food ordering from restaurants. Firstly, the busyness of our modern lives places a premium on convenience. Secondly, the outbreak prompted more people to start ordering food online, underscoring the importance of contactless dining solutions and reducing the necessity for in-person restaurant visits. Safety concerns and lockdown measures further emphasize the significance of these platforms. Moreover, these apps offer a wide variety of food options, including different types of cuisine and restaurants, which suit the diverse preferences of customers. This variety, along with easy-to-use menus and helpful features like reviews and ratings from other customers, makes the whole dining experience even better. Food delivery services are now a big part of how we experience and enjoy our favorite meals, changing the way we approach dining.

Sentiment analysis is a substantial area of research within the field of natural language processing (NLP) that aims to identify underlying sentiments or emotions expressed in various forms of language. This concept holds a multitude of potential applications, including the analysis of consumer feedback, the management of brand reputation, the enhancement of business operations, and the advancement of market research. Employing sentiment analysis in restaurant reviews can raise awareness of the overall level of customer satisfaction and enable the evaluation of specific areas in need of improvement.

This research aims to analyze text-based customer reviews on the online platforms of food ordering portals in Bangladesh, including Foodpanda and HungryNaki.

The primary objective of this study is to analyze diverse patterns within restaurant website reviews provided by various individuals and determine whether the sentiments expressed are positive, negative, or neutral, utilizing unique databases. This approach will facilitate improvements in food delivery services and restaurants to better align with customer preferences. In addition to ensuring customer satisfaction and meeting their expectations, this study has the potential to yield positive outcomes for the business landscape of Bangladesh.

In conclusion, the primary objective of this research is to utilize natural language processing for conducting an in-depth analysis of text-based review evaluations from Bangladeshi online food delivery platforms. The aim is to draw insights into the reliability of restaurant services.

1.1 Research Problem

Due to the rapid growth of online food ordering platforms, people heavily rely on online reviews before ordering food from any restaurant. Ordering food from restaurants online has become incredibly convenient, but it's not without its challenges for customers. One of the primary issues they face is uncertainty about the quality of food and delivery service. Without being physically present at the restaurant, customers heavily rely on reviews to make rational decisions. However, not all reviews are reliable, as some may be biased or unrepresentative. To address this issue, they need to sift through a substantial volume of reviews along with their respective ratings.

The solution to this problem was leveraging the power of sentiment analysis in reviews. By applying natural language processing techniques, the sentiments expressed in customer reviews were systematically analyzed and categorized. Sentiment analysis involves a series of essential steps. To begin, feedback from customers was gathered, originating from online platforms. Subsequently, the data underwent preprocessing to eliminate noise, including irrelevant information and special characters, while also standardizing text through tasks such as lowercasing and removing stopwords. Following this, text analysis took center stage, employing natural language processing techniques such as tokenization, lemmatization, and part-of-speech tagging to dissect the textual content. The core of sentiment analysis lay in sentiment classification, where various machine learning algorithms were utilized to categorize the sentiment expressed in each review. A score was then assigned to each piece of text to indicate its positivity, neutrality, or negativity. Finally, data visualization techniques were applied to present the results of sentiment analysis in an accessible format using graphs, charts, and word clouds.

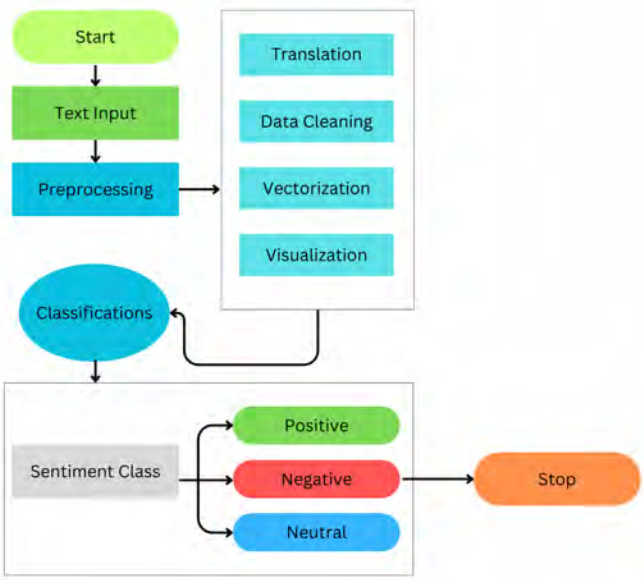

Figure 1.1: General framework of sentiment analysis process

All these processes will allow us to determine the overall satisfaction level of customers and identify specific areas of concern, such as delivery speed, food quality, or customer service. Armed with this data-driven insight, restaurants, and food delivery services will be able to make targeted improvements to address these issues and enhance the overall customer experience. Moreover, it will empower customers to make more informed choices by providing them with a more accurate representation of what to expect when ordering food online, ultimately making the overall experience more enjoyable and reliable.

In Bangladesh, people are increasingly using food delivery platforms, and the number of users is growing day by day. The concept of online food delivery from restaurants in Bangladesh has gained significant popularity in its early stages. Foodpanda, HungryNaki, Shohoz Foods, PathaoFood, and UberEats are now the top contenders in the food delivery industry. This paper [1] informs about how many customers are using this service and their perspective on it, considering factors such as delivery and restaurant service pricing, quality, time, and management.

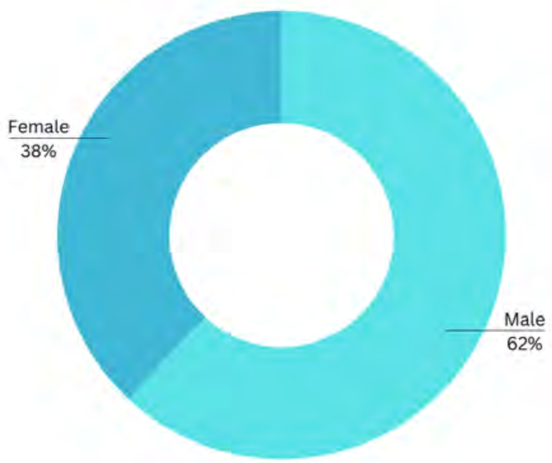

Figure 1.2: Gender distribution of online food orders

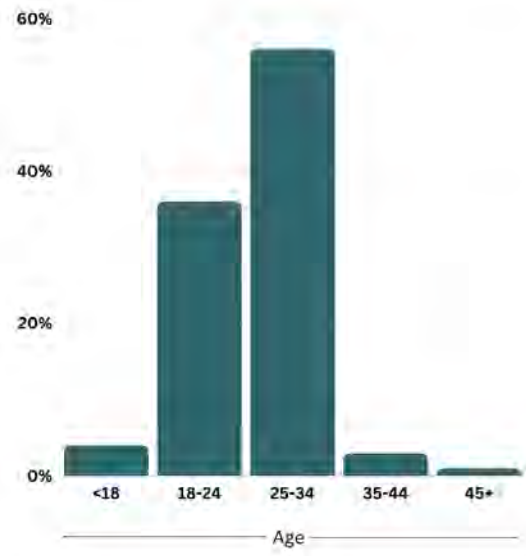

Figure 1.3: Age group preferences in online food ordering

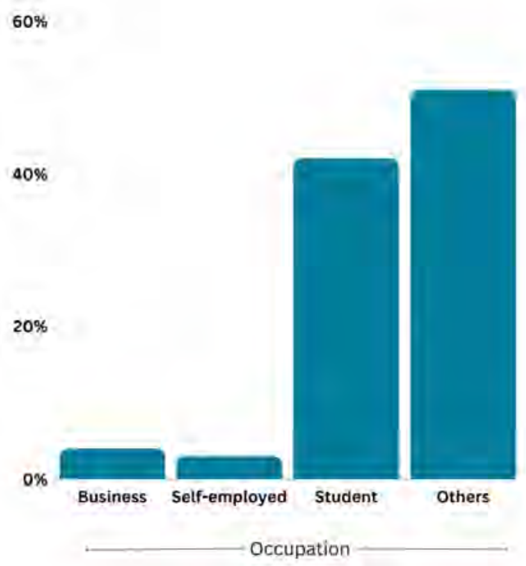

Figure 1.4: Occupational patterns in online food ordering

Our goal is to resolve this everyday challenge by establishing a model capable of identifying customer reviews and conducting sentiment analysis to categorize them into three groups: positive, negative, or neutral. Due to the current scarcity of resources and datasets in the Bengali language for this purpose, a new dataset needs to be created. The author has gathered reviews from various online food ordering platforms and classified them based on their sentiment. Initially, the focus was on collecting reviews from platforms such as Foodpanda and HungryNaki. Our findings are intended to be presented through an application or website. This tool allows consumers to access reliable reviews of their preferred restaurants within seconds quickly.

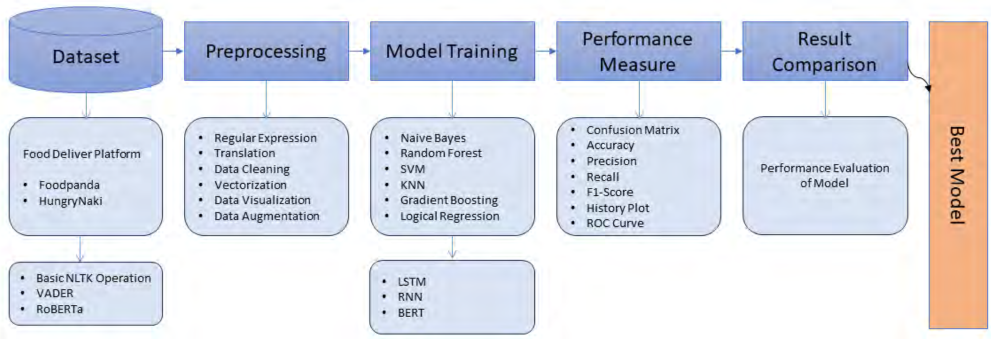

Figure 1.5: Workplan

Raw Data	Preprocessed Data	Class
I'm impressed by both the food and the delivery service. This restaurant never disappoints!"	i'm impressed by both the food and the delivery service. this restaurant never disappoints	Positive
বাসি খাবার ছিল, পুরোনো তেলে ভাজা।	there was stale food, fried in old oil.	Negative
Delivery was on time. No complaints in food, but nothing outstanding either.	delivery was on time. no complaints in food, but nothing outstanding either.	Neutral

Figure 1.6: Sample dataset before preprocess and after preprocess with class

1.2 Research Objective

Our research aims to ascertain the sentiments of people in Bangladesh regarding restaurants and online food delivery services, focusing on aspects such as delivery time, food quality, and customer service.

1. Understanding the expectations and requirements of customers in Bangladesh regarding restaurants and food delivery services.

2. Analyzing customer reviews on websites where they can order food.

3. Developing a model capable of distinguishing between positive, negative, and neutral reviews.

4. Converting the model into a web application to provide customers with comprehensive reviews, enabling quicker decision-making and well-informed choices when using these services.

1.3 Thesis Organization

1. Introduction:
    - Background of the Research: Provided an overview of the online food delivery industry and the demand for it in Bangladesh.
    - Research Problem: Defined the research problem, highlighting the limitations and work plan of the research.
    - Research Objective: Presented the main goals and objectives of the research, specifying the contributions in the field.
2. Literature Review:
    - Summarized related works in the chosen field, emphasizing key methodologies, results, and datasets.
3. Model Specification:
    - Described the machine learning models and deep neural network architectures used, highlighting their design principles and advantages in addressing the research problem.
4. Description of the Data:
    - Data Collection Method: Explained the sources and methods used to collect relevant data for the research.
    - Data Preprocessing Method: Describes the preprocessing steps undertaken to clean and prepare the data for model training and evaluation.
    - Data Augmentation Method: Discussed any data augmentation techniques employed to enhance the diversity of the dataset.
    - Data Visualization: Presented visualizations that aid in understanding the characteristics of the dataset, providing insights into patterns.
5. Performance Analysis:
    - Result Evaluation: Analyzed the results obtained from both machine learning and deep neural network models, and the behavior of different models in various scenarios.
    - Comparison of Results: Compared the performance metrics of different models, showcasing their relative strengths and limitations in addressing the research problem.
6. Website Implementation:
    - Provided a detailed description of the implemented website, outlining its features.
7. Conclusion
8. References

Chapter 2

Literature Review

In this section, most of the previous works conducted in this field have been summarized. The study of each paper has resulted in our enrichment with various ideas and methods of implementing a model, ways of data collection, different challenges, along the major insights of several aspects of a subject. In the following, the summaries of different papers in the field of their study have been enlisted.

In this paper [2], they have analyzed reviews from nearly one thousand restaurants, collected from various Facebook groups and public papers. These reviews were used to train the system. To create the corpus, an English benchmark dataset was translated into Bengali due to a lack of datasets and resources in Bengali. Various machine-learning approaches were employed, resulting in an 80.48% accuracy using the multinomial Naive Bayes classifier, which is the highest accuracy compared to other approaches. Their suggested approach can categorize 84% of positive reviews and 74% of negative reviews. While a small amount of inconsistency may exist in the manually collected dataset, this paper could significantly impact future research in the field. In summary, future research could employ enhanced methods to establish more meaningful relationships between words, improving the model's ability to accurately understand sentiments. Additionally, increasing the dataset size is essential for achieving higher accuracy.

In another research [3], sentiment analysis has been used to analyze customers' reviews of the ten best restaurants in Surabaya. For collecting the data and reviews from the hotel's customers, the researchers have relied on online media resources such as TripAdvisor. A software named WebHarvy has been used to retrieve information from TripAdvisor, and Python has been used for further processing, such as removing punctuations, stopwords, etc. The researchers have evaluated the performance of the Naive Bayes classifier with Textblob, a Python module for natural language analysis. The outcome reveals that while the Naive Bayes classifier has achieved a greater accuracy of 72.06% and has indicated it is somewhat more efficient (2.94%) than TextBlob sentiment analysis, TextBlob has an accuracy of 69.12%. The strong point of this research is that consumer reviews on this paper's dataset are in text format, while the results of the collected data have been labeled positive and negative, which might be helpful for the customers as well as for the hotel owners.

In another paper [4], the author proposed a way for scoring reviews on a scale of 1 to 5, with each category depending on the sentiments expressed in the text. It has also suggested a way for judging the review of the foods and combining them with already available text analysis software. The author has used a dataset from Amazon for the research purpose containing the food reviews of around 3000. A classifier has been developed that can determine the level of the sentiment of the dataset and denote it as either positive, negative, or neutral of a specific review using machine learning techniques. Based on these weights, a sentiment score has been calculated for each review, and the dataset has been compared with several baseline methods that reflect the consumer's overall sentiment. This approach can be noisy and informal language, which the result reflects. As data that has been collected online might have some drawbacks like fake opinions, paid or biased reviews, etc., it has been more challenging to deal with that, but this paper has a strong point in this case. However, this approach mentioned in the research struggles with hidden sentiments and only performs well for subjective feelings, such as ratings or scores. It will combine the current methodology with prediction-based strategies in future studies, and more characteristics will be retrieved to carry out hidden sentiment analysis.

Furthermore, this paper [5] aims to provide a significant contribution to the domain of Sentiment Analysis by comprehending and scrutinizing the discrepancies in the outcomes of each method and ultimately determining the most appropriate approach. The chief objective of this particular machine learning model has been to evaluate whether or not the spectator has relished the movie by thoroughly scrutinizing their critique. The effectiveness of various techniques to analyze sentiments has been investigated by the authors based on a 2000-movie review dataset from the IMDb website. The present study has employed a dataset that comprises 1,000 affirmative and 1,000 negative appraisals of movies. The present investigation has involved the utilization of the Naive-Bayes algorithm to conduct sentiment analysis on film evaluations. To achieve this objective, the data has been partitioned into two distinct groups, namely the training set and the testing set. Additionally, the study has intended to compare the results obtained through the Naive-Bayes approach with those derived from a Rule-Based Approach utilizing the AFINN-111 sentiment dictionary. The Natural Language Toolkit (NLTK) library has been utilized within the proposed framework to carry out a sentiment analysis of the given data. Ultimately, the sentiment with the greatest likelihood has been ascertained as the conclusive overall sentiment of the movie critique. The highest level of precision has been observed when both the training and test data have been equally divided at 50%, according to findings. Following a thorough investigation of the findings, it has been observed that the model exhibits a predisposition to commit elementary mistakes. Remarkably, these errors are noticeably missing when a significant portion of the data is utilized for training purposes. Consequently, this specified ratio of training data and test data, which is 7:3, has been deemed the optimal partitioning arrangement between the training and test data. One of the limitations of this paper is that the dataset used for sentiment analysis is limited to only movie reviews. Additionally, the authors have used only one sentiment dictionary (AFINN-111) for the Rule-Based Approach, which may not be suitable for all types of textual data. The rule-based approach using the AFINN-111 sentiment dictionary is relatively simple and interpretable. Nonetheless, it can be a quick and effective approach for basic sentiment analysis tasks. The scope of future work is not explicitly mentioned.

Besides, in this paper [6] the objective has been to conduct sentiment analysis at the aspect level of e-commerce data utilizing customer reviews from Amazon. The paper has used Amazon customer review data for aspect-level sentiment analysis. This paper has utilized a strategy for distinguishing the grammatical classification of each word in every sentence by using Parts-of-Speech (POS) labeling. The study has further extracted commonly used words, eliminated unwanted or redundant terms, and conducted adjective extraction from the sentences. Subsequently, sentiment analysis has been performed via classification algorithms to categorize the sentiment as negative, neutral, or positive. The SentiWordNet tool has been utilized to obtain the positive, negative, and neutral scores for each lemma individually. The experimental findings have facilitated the assignment of Nouns, Pronouns, Verbs, adverbs, and Adjective tags to each lemma present in the reviews. Subsequently, the Apriori algorithm has been utilized to identify the most recurrent lemmas in the reviews. The study has conducted a comparison between SVM and Naive Bayes, evaluating their recall, f-measure, precision, and accuracy. The distinctiveness of this article resides in its implementation of aspect-level sentiment analysis to spot affirmative, pessimistic, and impartial attitudes expressed in consumer feedback on online marketplaces. This approach affords the paper a strong advantage in its examination of the information contained within such reviews. The weakness of the paper is that it has not provided a detailed explanation of the aspect extraction and ranking process. The paper's future work can focus on improving the aspect extraction and ranking process. Overall, I believe that this research paper has the potential to contribute to the existing knowledge in the field of sentiment analysis and e-commerce. By focusing on aspect-level sentiment analysis, it addresses a specific need in the industry and provides insights that can inform decision-making processes for businesses and help consumers make informed purchasing decisions.

In another paper [7], the study has emphasized the value of sentiment analysis for the food delivery business and has shown how well various models have worked for this purpose. In this paper, 20,000 reviews have been collected from Foodpanda and Hungrynaki. They have taken 18,000 customer reviews from Hungrynaki, but just 2,000 from Foodpanda. Separately, they have gathered all the data in CSV formats, which they have then combined for analysis. Furthermore, they have used the BERT Pre-Training Approach (RoBERTa), AFINN, and DistilBERT to conduct the sentiment analysis. They have combined both machine learning and non-machine learning techniques. They have deployed a tool for sentiment analysis that is lexicon-based for non-machine learning approaches. For sentiment analysis, Convolutional Neural Networks (CNNs), Naive Bayes classifiers, Support Vector Machines (SVMs), and transformer models from Deep Learning have been utilized. With the use of the Google Cloud Translation API, they have converted the Bangla-romanized reviews. The author has addressed 20,000 reviews in all, and three distinct models—RoBERTa, AFINN, and DistilBERT - have been used to analyze them. Among all, the accuracy has been 74%, 73%, and 77%, accordingly. Initially using the Pandas library's assistance, the author has looked for any instances of missing information or empty values and has removed them. The author has also looked for data mistakes, such as inaccurate ratings. Additionally, they have removed the extra rows that were a result of the API review. It is typically important to tune the model on a large dataset and thoroughly analyze the way it has performed on multiple test scenarios to get high performance.

In this study [8], the author has proposed a solution for analyzing large datasets of Amazon Fine Food reviews, and they have achieved 80% accuracy. When using these techniques, the author has realized that NB and logistic regression are less effective than linear SVC. Five phases have constituted the strategy implemented in this study. These have included collecting datasets via data visualization and preprocessing, implementing machine learning classifiers via Spark MLlib, and evaluating models via train-test split using multiple binary classification metrics. Experiments have made use of the Fine Food dataset from Amazon. The dataset on Amazon has had a total of 568,454 reviews. The rating has been based on the number of users who have rated the review as useful or unhelpful, a description of the review, and the review's content. Multiple classifiers have been trained and tested; however, in this article, 3 classifiers with greater than 80% accuracy have been chosen. Linear Support Vector Classifier, Logistic Regression, and Naive Bayes have been the models used. It could be important to get a bigger and more varied dataset for delivery applications in Bangladesh. With the help of the businesses running the meal delivery apps, this may have been accomplished.

In another research [9], FoodPanda, HungryNaki, Pathao Food, and Shohoz Food's Facebook pages have been used as the source of data for a sentiment analysis of user comments. A thorough data pre-processing procedure has been performed, including steps like tokenizing, deleting stop words, and creating contractions. This has been achieved via the application of three distinct supervised classification methods: extreme gradient boosting, random forest classifier, decision tree classifier, and multinomial Naive Bayes. Three independent deep learning (DL) models have been utilized, including convolutional neural networks, long-term short-term memories, and recurrent neural networks. With an accuracy of 89.64%, the XGB model has outperformed all three machine learning (ML) techniques. Out of the three DL algorithms, LSTM has the greatest accuracy rate (91.07%). The LSTM DL model, which combines ML and DL, excels in predicting sentiment. On the other hand, Bengalis have just lately started teaching their language using computers. NLP is a strong technology with many benefits, but it nevertheless still has many restrictions and problems. To distinguish favorable and unfavorable reviews, the author has had to go through several stages, including dataset collection, data preparation, and model development. After using Word2Vec to calculate the vocabulary size and pad sequence, and the Bangla Natural Language Processing Toolkit (BNLTK) library to tokenize the text, they have prepared or gathered cleansed or purified text to go to the next level of their study endeavor. LSTM has the best accuracy rate out of the three DL algorithms, which is 91.07%.

This paper discusses [10] the purpose is to conduct a comparative sentiment analysis of user evaluations. The opinions of various smartphone users were compiled and categorized as positive-negative-neutral. It entails constructing a system to collect and analyze product reviews from multiple online stores. This paper's Data is an organized collection of product evaluations obtained from Amazon.com. In total, they gathered over 500 evaluations of products belonging to their main categories, including Mobiles, Computers, Flash drives, and Electronics. Over 3,2 million consumers posted these evaluations online regarding 10,001 products. The goal of this paper is to present product evaluations based on consumer feedback using opinion mining, text mining, and sentiments. Opinion mining is primarily the identification of sentiments that determine the evaluation of people's opinions and sentiments about products and services [11]. They can either be direct or comparative opinions. Subjective content that comprises at least one positive or negative word and has semantic meaning was extracted. The Parts of speech tagger used for this research is a Penn Treebank Project-developed max-entropy POS tagger. The evaluation matrix, including accuracy, precision, and recall, was not mentioned in the paper. One of the limitations of this paper is that the sentiment analysis dataset is restricted to smartphone reviews only. The authors have not evaluated the model on other forms of textual data or any other website source. In addition, the authors have not utilized any deep-learning model that may apply to all textual data categories. Future work may entail enhancing the accuracy of the sentiment analysis model by incorporating more sophisticated techniques and algorithms. By evaluating the performance and limitations of each method, the author can gain a better understanding of sentiment analysis techniques in the context of product reviews, which can inform future research and application in this area.

Furthermore, in this paper [12] the project's objective is to conduct sentiment analysis on e-commerce data using a large sample of online reviews for mobile phones. The objective has been to extract aspect terms from each review, both positive and negative. This data analysis has included anger, anticipation, revulsion, dread, happiness, sorrow, surprise, and trust. This study has collected a large dataset consisting of online evaluations from Amazon.com. The dataset has included over 400,000 evaluations for roughly 4500 mobile phones. The feature consisted of the product's name, brand, price, rating, reviews, and review votes. This research has contextualized unstructured data that has been filtered to remove distracting data. Then, stop words, punctuation marks, whitespaces, digits, and special symbols have been removed from the data using preprocessing. The 'tm' program has been used for text mining. In the third stage, pertinent features have been extracted through feature selection. Finally, a statistical analysis of the dataset has been performed to investigate the correlation between various features and to estimate the text's emotional tone. The labeled data has then been used to train and evaluate a Support Vector Machine (SVM) classifier, whose performance has been validated via 10-fold cross-validation. SVM's predictive accuracy has been determined to be 84.87%. The paper's strength is that it employs sentiment analysis to identify positive, negative, and impartial consumer feedback on e-commerce websites. This assists online retailers in understanding consumer expectations, enhancing the purchasing experience, and boosting sales. The shortcoming of the paper is that it lacks a comprehensive explanation of the implementation of alternative machine-learning models. Moreover, the paper can extend its analysis to other domains, such as healthcare, politics, and social media, beyond e-commerce.

In this study [13], a Hybrid CNN-LSTM Model has been proposed, which combines LSTM and a deep CNN model. It has utilized Word to Vector (Word2Vec) for initial word embeddings, allowing the conversion of text into numeric vectors, distance computation between words, and grouping of similar words. The model has incorporated features extracted by convolution and global max-pooling layers, along with long-term dependencies. Dropout, normalization, and rectified linear units have also been employed to enhance accuracy. The results indicate that the Hybrid CNN-LSTM Model has surpassed traditional deep learning and machine learning techniques in terms of precision, recall, f-measure, and accuracy. The model has demonstrated competitive performance on datasets such as the IMDB movie review dataset and Amazon movie reviews dataset when compared to state-of-the-art techniques.

In addition, this paper's [14] primary objective is to do an effective inspection of extracted Twitter messages and know people's opinions through opinion mining [11]. Using Hadoop, which is capable of processing vast quantities of data, it aims to create an autonomous model that anticipates the sentimentality of social media messages. TwitterAPI is used to retrieve and store tweet scores and timestamps from Twitter. Twitter users' public tweets are extracted. An application can be submitted by up to 5,000 Twitter user IDs via a single connection. Only publicly accessible Tweets were extracted using the API. The filter API facilitates navigation and provides an uninterrupted flow of Tweets that match the filter tag. Extracted Tweets then transferred into Hadoop. They were preprocessed utilizing map reduction. The model utilized here is the Bayes classification with a single term. This module received 20,00,000 already classified tweets from multiple sources, and its task is to train a classifier on the massive data set. Nltk is utilized to eradicate words with POS identifiers. Hadoop is used to retrieve the information from the data, while MapReduce is used to rapidly extract multiple terms along with their probabilities. As expected, the results were generally favorable. Due to the reported articles, a few tweets were classified as neutral and only a few as negative. This work is exceedingly beneficial to sentiment analysis-reliant individuals and industries. The first limitation is that the data was trained and classified using the same word probabilities. This work could be enhanced in the future by employing n-gram classification instead of uni-gram a classification that will help to filter on Hadoop. It may be possible to build.

This work [15] presents a cross-media analysis framework with built-in sentiment analysis capabilities. The study evaluates the efficacy of two sentiment analysis techniques—lexicon-based and machine learning—for identifying emotions in online forum discussions. Apache Hadoop and Stanford CoreNLP's Recursive Neural Tensor Network model are used to conduct experiments that predict sentiment. RNTN outperforms lexicon-based by 9.88% in overall accuracy, but lexicon-based performs better in classifying positive comments. The F1-score of lexicon-based is 0.16 higher than RNTN.

Additionally, this research [16] addresses challenges in the COVID-19 vaccination program, factors including the spread of new viruses and public mistrust. Commenters on social networking sites are known for their grammatical errors and the awkward blend of formal and informal vocabulary. The researchers developed CoVaxBD, a corpus of sentiment-annotated Facebook comments code-mixed in Bangla and English. They also propose a sentiment analysis model based on multilingual

BERT, achieving a high validation accuracy of 97.3% and a precision score of approximately 97.4%.

In another research [17], using feature-specific sentiment analysis, Subhabrata Mukherjee and Pushpak Bhattacharyya examined product reviews of the customers. The relationship between the features and the opinions connected to those features has been identified using a dependency parsing technique. They generated an approach that gathers opinion expressions defining various aspects and extracts possible features from reviews. To evaluate the effectiveness of their system, they conducted experiments using two distinct datasets, namely those by Lakkaraju and Hu and Liu. The outcomes of their experiments revealed an average accuracy of 80.98% for Dataset 1 and a 70% accuracy for Dataset 2.

This study [18] introduced a framework designed to discern product feature aspects and delve into the specific reasons expressed by consumers in online product reviews. Their proposed co-clustering algorithm plays a pivotal role in generating succinct summaries that encapsulate consumers' concerns across various facets of product features. Moreover, it assists in elucidating the rationales behind consumers' inputs and expectations, thereby providing concise insights that can guide product designers. To accomplish this, they leveraged conditional random fields as an integral part of their approach, enabling the simultaneous identification of product feature aspects and detailed consumer reasons.

Another study [19] showed that various platforms in the catering sector, such as Yelp, Open Table, and Zomato, provide users with essential details about restaurants, evaluation data, and recommendations. Despite the nascent stage of user and restaurant information, current research on restaurant recommendations predominantly relies on assessment data for neighbor identification. Additionally, there is limited research on how various types of input information influence the performance of recommender systems in this context.

In this research [20], a hybrid bidirectional recurrent convolutional neural network attention-based model called BRCAN has been developed. To achieve the objective of fine-grained text categorization, the model has combined word2vec, bidirectional long short-term memory, and a convolutional neural network with the attention mechanism. In their model, a bidirectional recurrent structure has been utilized to capture sentence dependencies over time and contextual information, and word2vec has been applied to automatically produce word vectors.

Additionally, this research [21] indicates that social media platforms enable human interaction and textual communication by incorporating emojis and written text. A noteworthy area of investigation that has emerged involves various forms of analysis, with emotional analysis being the most prevalent.

The paper [22] has a primary focus on the automation of customer opinion analysis regarding a restaurant's services through the utilization of Natural Language Processing (NLP). It employs NLP techniques such as opinion gathering, data collection, word extraction, opinion mining, and classification to automatically process and comprehend these reviews. Manual inspection by 10 volunteers further validates the automated analysis, revealing a higher accuracy rate which is approximately 86% in identifying positive and negative sentiments compared to manual observation.

The study [23] has delved into sentiment analysis, which involves determining whether a given text conveys positivity, negativity, or neutrality. This becomes particularly crucial when dealing with a multitude of reviews. During the assessment stage, it was found that the Support Vector Machine attained the highest accuracy, reaching 95%, utilizing TF-IDF vectorization. Following closely behind was Logistic Regression with an accuracy of 91%. The performance assessment considered metrics like Precision, Recall, and F1 Score.

A recent study [24] has concentrated on the assessment of customer reviews within the realm of online food delivery services in Bangladesh. Due to their often unstructured nature, these reviews present difficulties for manual analysis, leading the researchers to create an automated system for comprehending customer behavior and responses. The findings revealed that the CNN with an attention mechanism achieved the highest accuracy at 98.45%, outperforming baseline CNN and LSTM models.

The study [25] has outlined a research project carried out in Bangladesh that specifically examines online shopping feedback in the Bangla language during the COVID-19 pandemic. More than 1000 reviews were gathered, and sentiment analysis was performed using different algorithms like KNN, Decision Tree, SVM, Random Forest, and Logistic Regression. Among these, SVM demonstrated the highest accuracy, reaching 88.81%.

Another research [26] conducted sentiment analysis using various analyzers, including TextBlob, SentiWordNet, and W-WSD, on tweets gathered from public accounts. Out of 100,000 tweets, only 6,250 remained after preprocessing. TextBlob had the highest rate of positive sentiment tweets that is 3,380 or 54.08%, while SentiWordNet had the highest rate of negative sentiment tweets (3,054 or 48.86%). To validate the sentiments, a Naïve Bayes classifier was applied to a training set of 1,690 tweets from each sentiment analyzer. W-WSD had the highest accuracy, with 316 correct instances and 79% accuracy. TextBlob was second with 304 correct instances and 76% accuracy, while SentiWordNet had 219 correct instances and 54.75% accuracy. Additionally, 5,000 tweets were validated using the SVM classifier, resulting in TextBlob and W-WSD having similar sentiment accuracy at around 62%. Overall, TextBlob and W-WSD outperformed SentiWordNet in analyzing election sentiments and making accurate predictions based on the experimental results.

Moreover, this paper [27] gives a thorough analysis of the research procedure and progress in text sentiment analysis, covering various methods such as sentiment analysis based on dictionaries, conventional machine learning algorithms, and deep learning algorithms. Further, it talks about the sentiment analysis of Chinese and special texts, as well as the difficulties and future development directions in text sentiment analysis. The paper also discusses the challenges and limitations in sentiment analysis, such as accurately calculating sentiment similarity between words, dealing with the meaning of words in context, and analyzing interrogative sentences. These insights can help researchers identify and address practical difficulties in sentiment analysis. It emphasizes the breakthroughs achieved by deep learning algorithms in text vectorization and the accuracy of sentiment analysis, particularly with the use of recurrent neural network models like LSTM. It highlights the influence of sentiment analysis in the field of natural language processing and its impact on politics, economy, and social sciences.

The paper [28] applies the Support Vector Machine (SVM) machine learning algorithm for sentiment analysis of product reviews. Several datasets are used for training and testing the SVM algorithm to classify the sentiments and texts in the reviews. The SVM algorithm is used to compute the polarity of ambiguous sentiments or reviews. The performance of the resulting models is tested to measure the accuracy of the SVM learning algorithm. The paper also mentions the use of supervised machine learning algorithms for classifying movie reviews and a hybrid approach combining Word2Vec and sentiment-emotion information for sentiment prediction. The Support Vector Machine (SVM) classification algorithm achieved a higher accuracy of 89.98% compared to other algorithms used in the study. The accuracy of the SVM algorithm can be further improved by including more sentence forms. The SVM algorithm was found to behave well and was considered a better classification algorithm than others for sentiment analysis of product reviews.

Furthermore, in another paper [29], to address the imbalanced data issue, this work suggests a hybrid solution that combines the Support Vector Machine (SVM) algorithm with Particle Swarm Optimization (PSO) and several oversampling methods. SVM is by enhancing the machine learning classification algorithm, used to anticipate the feelings of reviewers dataset that includes various reviews of various Jordanian restaurants. Information was gathered from Jeeran, a popular social network for reviews in Arabic. The weights of the PSO are optimized using their various oversampling methods, including the Synthetic Minority Oversampling, which is included in the features.SMOTE, SVM-SMOTE, ADASYN, and borderline-SMOTE are examples of techniques, that were investigated to create an optimal dataset and address the dataset's imbalance issue.

The paper provides [30], the importance of fine-grained sentiment analysis for online evaluations is growing across a wide range of applications. Here, effective multi-grained aspect extraction, finding related opinions, and categorizing sentiment polarity are the important methods. Although several topic models have been suggested in recent years to handle some of these jobs, there hasn't been much effort done on efficient sentiment analysis. In this study, the authors provide a joint aspect-based sentiment topic (JABST) model for jointly extracting multi-grained aspects and opinions by modeling aspects, opinions, sentiment polarity, and granularities. They offer a maximum entropy-based JABST model that makes use of supervised learning to further improve accuracy and performance while extracting views and attributes.

Chapter 3

Model Specification

3.1 Description Of the Model

Our research heavily relies on the user reviews of Bangladeshi online food ordering platforms. To navigate this vast sea of data effectively, modern machine learning and natural language processing models such as Naive Bayes, Logistic Regression, Random Forest, SVM, Gradient Boosting, KNN, LSTM, BERT, and RNN have emerged as invaluable tools. These models, each with its unique capabilities, enable us to gain deeper insights into customer sentiments expressed in reviews on food ordering and delivery portals of Bangladesh. In this context, we will know about the application of these models for sentiment analysis.

3.1.1 Machine Learning Models

Naive Bayes: Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem [31]. It is naive because it makes the assumption that the features used for classification are independent of each other, which simplifies calculations.

Naive Bayes calculates the conditional probability of a particular class (C) given a set of features (X) using Bayes' theorem:

$$P(C|X) = \frac{P(C) * P(X|C)}{P(X)} \tag{3.1}$$

Where:

- $P(C|X)$ is the posterior probability of class C given features X.
- $P(C)$ is the prior probability of class C.
- $P(X|C)$ is the likelihood of the features given class C.
- $P(X)$ is the marginal likelihood, which is the probability of observing features X.

The labeled dataset has been used to train a Naive Bayes classifier. The classifier determines the likelihood that a review belongs to each sentiment class by analyzing its characteristics (words). Using the trained Naive Bayes model, it determines the conditional likelihood that a new review falls into each sentiment class. The class with the highest probability becomes the predicted sentiment of the review. Suppose there are three sentiment classes: Positive (P), Negative (N), and Neutral (U). For example, certain words like excellent and happy are more likely to appear in positive reviews, while words like terrible and disappointed are more likely to appear in negative reviews.

Logistic Regression: Logistic Regression is a statistical model used for binary classification problems, where the outcome is a binary variable.

Logistic Regression models the probability using the logistic function (sigmoid function) [32]:

$$P(Y = 1|X) = \frac{1}{1 + e^{-\beta^T x}} \tag{3.2}$$

Where:

- $P(Y = 1|X)$ is the probability that the output Y is 1 given input features X.
- β represents the model parameters (coefficients).
- X represents the input features.
- e represents the base of natural logarithms

The sigmoid function is referred to as an activation function for logistic regression and is defined as:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.3}$$

Where:

- e represents the base of natural logarithms

A labeled dataset of reviews was first gathered, classifying each one as either positive, negative, or neutral, similar to Naive Bayes. The associations between feature words and emotion labels were then recognized by the model during training. The likelihood of a review falling into a specific category is represented using the logistic function (sigmoid function).

Random Forest: Random Forest combines multiple decision trees to make predictions. Each decision tree is trained on a different subset of the data and features, and the final prediction is determined by classification or regression of individual tree predictions [33].

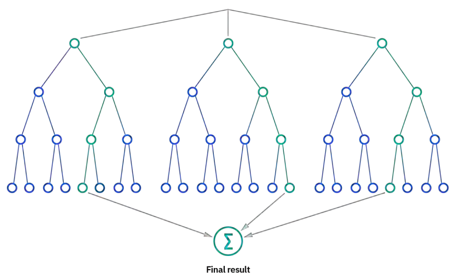

Figure 3.1: Random Forest diagram

At first, a Random Forest classifier was trained on the labeled dataset. During training, multiple decision trees were built by the Random Forest with different subsets of the training data. Overfitting was reduced, and the model was made more robust by the randomness introduced. The predictions from each decision tree were combined to formulate a final prediction. For multi-class sentiment analysis (positive, negative, neutral), a weighted average among the trees was employed to determine the final sentiment class.

Support Vector Machine (SVM): The goal of SVM is to locate a hyperplane that optimally divides data into its component classes [34]. The margin, defined as the distance between the hyperplane and the closest data points from each class, is maximized when selecting the hyperplane.

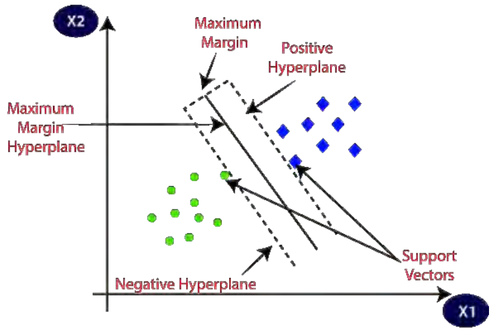

Figure 3.2: Support Vector Machine diagram [35]

Firstly, the SVM model has been trained on the labeled dataset. The hyperplane that best separates the reviews based on the chosen feature representation has been found by the SVM algorithm. To classify a new review, its preprocessed features have been input into the trained SVM model. The side of the hyperplane on which the review falls has been determined by the SVM model.

Gradient boosting classifier model: An ensemble of decision trees is constructed from labeled data and used to produce predictions about the sentiment of text documents using the Gradient Boosting Classifier.

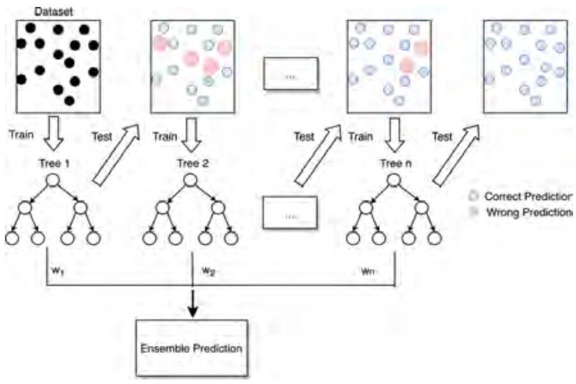

Figure 3.3: Flow diagram of Gradient Boosting [36]

Firstly, the gradient boosting model has been trained on the labeled dataset. During training, an ensemble of decision trees is created by the Gradient Boosting Classifier. Working together, the ensemble of decision trees is utilized to make predictions. The final prediction is determined based on a combination of the individual tree predictions.

K-nearest neighbors (KNN): Using KNN, a data point is given a classification according to the opinion of its nearest k neighbors [37]. If k=10, a data point's label will be based on the label of its 10 closest neighbors. Euclidean distance is used for determining how far apart two data points are.

3.1.2 Deep Neural Network Models

LSTM (Long Short-Term Memory): LSTMs are made up of memory cells with gating mechanisms to limit the flow of information [38]. The Forget Gate is pivotal because it uses sigmoid activation to decide which bits of data from the previous time step are kept and deleted. Meanwhile, the Input Gate uses a sigmoid layer for updates and a tanh layer for candidate value generation to regulate how new data is integrated into the cell's existing state. Last but not least, the Output Gate uses sigmoid activation to merge the current input with the cell state and decide what should be forwarded to the next time step.

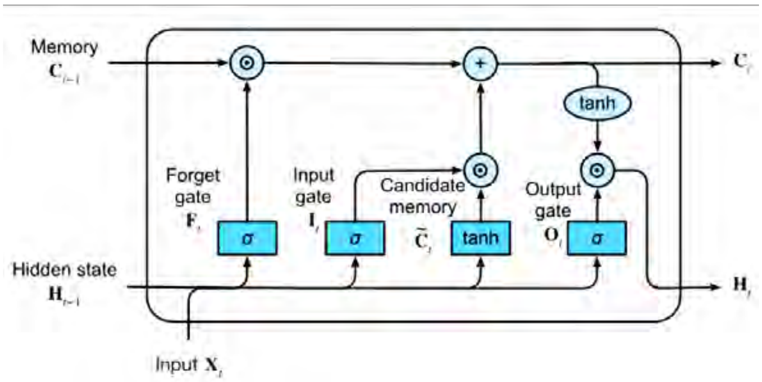

Figure 3.4: LSTM Architecture [39]

Tokenizing the input text into words or subword units and then converting them into numerical embeddings is the first step in using LSTMs for text sentiment analysis. The LSTM network takes these embeddings as input and processes the text token by token while remembering and updating the sequence's context in its memory cells. The output of the LSTM is fed into one or more fully connected layers, where it is then used to generate a sentiment categorization.

BERT (Bidirectional Encoder Representations from Transformers): BERT is a state-of-the-art natural language processing model developed by Google [40]. It utilizes a deep bidirectional Transformer architecture to understand the context and meaning of words in a sentence by considering both the left and right context. BERT has achieved remarkable performance on a wide range of NLP tasks, including text classification, question-answering, and language understanding.

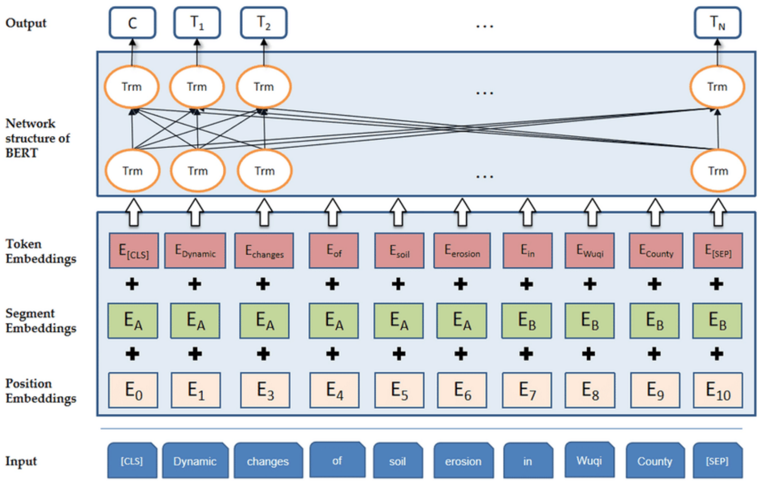

Figure 3.5: BERT Architecture [41]

RoBERTa

RoBERTa is another variant of BERT that focuses on optimizing pretraining tasks and hyperparameters. It has achieved state-of-the-art results on various NLP benchmarks by training on a larger amount of data and fine-tuning with specific techniques. RoBERTa improves upon BERT's performance on tasks like text classification, natural language understanding, and sentiment analysis.
The utilization of BERT for fine-tuning has entailed the training of a pre-trained BERT model on our labeled review dataset to predict sentiment labels. Sentiment-related information has been captured by adjusting the internal representations of the model through this adaptation. Usually, a classification layer has been added atop BERT to map embeddings to sentiment classes, and the combined model has been trained to minimize prediction errors. Once fine-tuned, new reviews can be classified by tokenizing and converting them into numerical embeddings using the same tokenizer. The sentiment label is then predicted by passing these embeddings through the model.

RNN (Recurrent Neural Networks): Recurrent Neural Networks (RNNs) represent a specialized category of artificial neural networks tailored for the analysis of sequential data. Diverging from conventional feedforward networks, RNNs possess intrinsic memory capabilities, facilitating the discernment of temporal dependencies and patterns within sequential input [42]. This architecture proves particularly adept in applications such as language modeling, speech recognition, and time series analysis. Despite their proficiency in capturing temporal dependencies, RNNs encounter challenges in sustaining long-term memory.

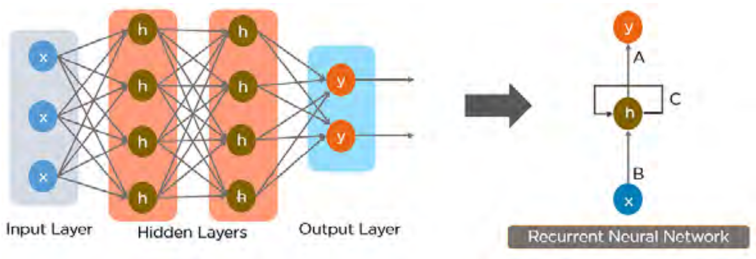

Figure 3.6: RNN Architecture [43]

Recurrent Neural Networks (RNNs) process sequential data by maintaining internal memory. With cyclic connections, each node retains information from previous time steps. The hidden state evolves as it incorporates current input, influencing the output. Training adjusts parameters to improve the network's ability to capture patterns. LSTM and GRU variants address long-term dependency challenges. In sentiment analysis of customer reviews, Recurrent Neural Networks (RNNs) excel by capturing contextual nuances in sequential data. RNNs analyze the ordered nature of language, retaining dependencies between words. This enables more accurate recognition of sentiments, enhancing the model's ability to discern subtle nuances and sentiments expressed in customer feedback.

Chapter 4

Description of the Data

4.1 Data collection method

In the initial stage, customer text-based reviews were collected from online platforms, including the Foodpanda and HungryNaki websites. Raw data from these food portal websites was obtained using web scraping methods, facilitated by the *NoCoding Data Scraper* tool. This data has been compiled in Excel (xlsx) format, encompassing restaurant-by-restaurant review information from various cities in Bangladesh. The collected data includes customer names, reviews, restaurant names, cities, and ratings. Using the Pandas library, all the restaurant customer review data has been merged into a single data frame, and column names such as reviewer_name, ratings, time, review_text, restaurant, and city have been assigned.

The ratings extracted from the Foodpanda website were presented in a string format, such as 4/5 and 5/5. These ratings have been converted into integer form, resulting in values like 4 and 5. In contrast, the ratings extracted from the HungryNaki website have not been provided in numerical or string format but rather in the form of star icon links, featuring two types: white star icons and colored star icons. To calculate the rating count for each row/review, a function has been developed that counts the colored stars based on the icon link. Finally, the Pandas data frame has been saved in an Excel file.

```
df_fp.head()                    #Foodpanda
```

	reviewer_name	ratings	ratings_int	time	review_text	restaurant	city
0	Fatema	1/5	1	1 week ago	The amount of food very very less.	Barcode Cafe_Chittagong	Chittagong
1	Priya	4/5	4	2 weeks ago	Food packaging good. Ordered fried rice set me...	Barcode Cafe_Chittagong	Chittagong
2	Maimuna	2/5	2	2 weeks ago	the rice was horrible, Idk why barcode does it...	Barcode Cafe_Chittagong	Chittagong
3	Samantha	1/5	1	3 weeks ago	I ordered something and received something else	Barcode Cafe_Chittagong	Chittagong
4	Meraj	3/5	3	3 weeks ago	it was ok. half the pizza slid off the bread b...	Barcode Cafe_Chittagong	Chittagong

```
df_hn.head()                    #HungryNaki
```

	reviewer_name	date	review_text	restaurant	area	ratings	scraper
0	Anando Mehran	May 28, 2023	Not satisfied. Oily, undercooked chicken and p...	Alfresco - Tejgaon	Gulshan	2	Priom
1	Warisa Meheraj Chowdhury	Apr 30, 2023	Garlic mashroom was extremely salted and soup ...	Alfresco - Tejgaon	Gulshan	1	Priom
2	NaN	Jan 25, 2023	টেস্ট খুব বাজে, একটু তিতা স্বাদ, বেশি স্পাইসি।...	Alfresco - Tejgaon	Gulshan	2	Priom
3	NaN	Jan 21, 2023	Salad was included in the package. But they di...	Alfresco - Tejgaon	Gulshan	1	Priom
4	Anika Fowzia Jui	Jan 16, 2023	just loved it!	Alfresco - Tejgaon	Gulshan	5	Priom

Figure 4.1: Sample collection of review data

A total of over 47,733 data entries have been collected. Among these, 43,060 reviews from Foodpanda and 4,673 reviews from HungryNaki have been extracted. Initially, two key columns were maintained: one dedicated to the reviews and the other to the ratings. Following this, a sentiment labeling approach has been applied to categorize the ratings into three distinct sentiments: positive, negative, and neutral. Positive reviews have been designated for ratings of 4 and 5, neutral reviews for ratings of 3, and negative reviews for ratings of 1 and 2. Consequently, after processing, every review has been systematically tagged with either a positive, negative, or neutral sentiment based on its associated rating.

	text	ratings	sentiment
0	The amount of food very very less.	1	Negative
1	Food packaging good. Ordered fried rice set me...	4	Positive
2	the rice was horrible, Idk why barcode does it...	2	Negative
3	I ordered something and received something else	1	Negative
4	it was ok. half the pizza slide off the bread ...	3	Neutral
...	...	...	...
47731	food price is increase suddenly.	3	Neutral
47732	I was ordering a chicken breast piece . they s...	1	Negative
47733	So Bad Beef Jhal Fry	1	Negative
47734	awesome	5	Positive
47735	requested to provide chicken breast piece, rec...	2	Negative

47736 rows × 3 columns

Figure 4.2: Sentiment labelled dataset based on rating

Bangla reviews have been translated into English using the Googletrans library. Additionally, some Bangla-romanized reviews have also been translated by this library.

4.2 Data pre-processing method

To begin with, all null or empty columns from the dataset have been removed. Furthermore, the text data has undergone cleaning to eliminate noise. Regular expressions have been employed to execute the following cleaning steps:

- Firstly, any text enclosed within square brackets has been removed, along with the brackets themselves. Square brackets, often used to denote meta-information or citations, have no relevance to the classification task.
- Next, non-word characters have been substituted with spaces. These non-word characters, including hashtags, or punctuation marks, may introduce noise and irrelevance to the classification task.
- Moreover, any URLs have been removed from the text. While URLs are commonly encountered in social media posts, they hold no significance for the classification task.
- Additionally, any HTML tags have been removed from the text. HTML tags are frequently encountered on web pages and may lack relevance for the classification task.
- Furthermore, any punctuation marks have been eradicated from the text. Punctuation marks, such as periods, commas, or exclamation marks, can introduce noise and irrelevance to the classification task.
- Moreover, any newlines have been removed from the text. Newlines, commonly found in text, may lack relevance to the classification task.
- Lastly, any alphanumeric characters containing digits have been eliminated from the text. Alphanumeric characters, such as emojis or emoticons, can introduce noise and irrelevance to the classification task.

	text	ratings	sentiment	p_text
0	The amount of food very very less.	1	Negative	the amount of food very very less
1	Food packaging good. Ordered fried rice set me...	4	Positive	food packaging good ordered fried rice set me...
2	the rice was horrible, Idk why barcode does it...	2	Negative	the rice was horrible idk why barcode does it...
3	I ordered something and received something else	1	Negative	i ordered something and received something else
4	it was ok. half the pizza slide off the bread ...	3	Neutral	it was ok half the pizza slide off the bread ...
...	...	...	...	...
47729	food price is increase suddenly.	3	Neutral	food price is increase suddenly
47730	I was ordering a chicken breast piece . they s...	1	Negative	i was ordering a chicken breast piece they s...
47731	So Bad Beef Jhal Fry	1	Negative	so bad beef jhal fry
47732	awesome	5	Positive	awesome
47733	requested to provide chicken breast piece, rec...	2	Negative	requested to provide chicken breast piece rec...

Figure 4.3: Data after pre-processing

4.3 Data augmentation method

In our dataset, the total number of positive reviews is calculated as 18,233, negative reviews are numbered 23,530, and neutral reviews amount to 5,970. Consequently, there has been an imbalance in the dataset. To address this imbalance, the data augmentation method [44] has been employed. Utilizing the Pegasus Model transformer [45], the number of neutral reviews has been augmented by paraphrasing them. Each neutral review has been paraphrased ten times, and a paraphrased review has been randomly selected for research purposes. In certain cases, 2-3 paraphrased reviews have been chosen for specific neutral reviews. Ultimately, the imbalance in the dataset has been successfully reduced by augmenting the number of neutral reviews, indicating a decreased gap in the review count among the three classes: positive, neutral, and negative.

Raw Text	**Pre-processed Text**	**Pegasus-Paraphrased Text**
Taste 😋 ..this is good .but nun's would be much better ..Thanks 😊 👍	taste this is good but nun s would be much better thanks	It is good, but nun's would be better.
অর্ডারের সাথে সস দেয়া হয়নি।	order did not come with sauce	The order did not include sauce.
Rice ok But Chicken tamonee hocchilo onek ager	Rice and Boot Chicken used to be like that a long time ago	Rice and Boot Chicken used to be that way a long time ago.

Figure 4.4: Neutral data after paraphrasing

4.4 Data visualization

We have employed various types of visualization methods, utilizing pyplot, seaborn, and wordcloud libraries, to depict relationships within the data and to visualize the dataset. Furthermore, we have utilized the TfidfVectorizer from the sklearn feature_extraction.text module to extract features from preprocessed review texts.

Histogram: This histogram is a graphical representation used to visualize the distribution of ratings in our dataset, helping us understand the spread and frequency of different rating values.

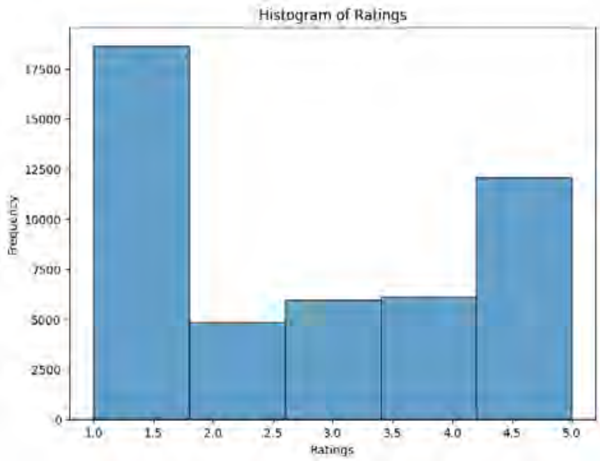

Figure 4.5: Histogram to visualize the distribution of ratings

Bar chart: This bar chart is an effective way to display the distribution of sentiment labels (Negative, Neutral, Positive) within our dataset, offering insights into the overall sentiment makeup.

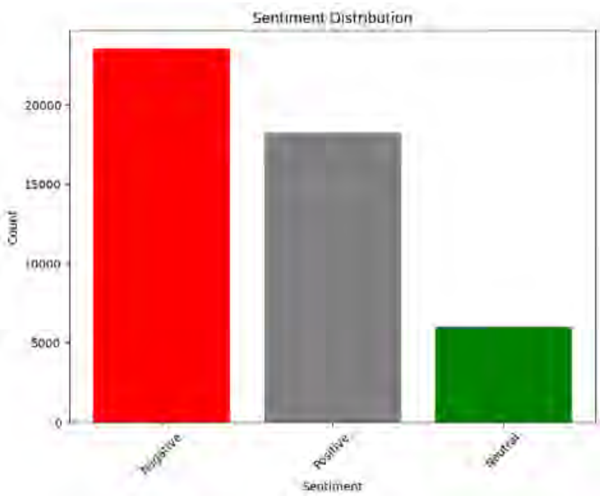

Figure 4.6: Distribution of sentiment labels (Negative, Neutral, Positive)

Pie chart: This pie chart offers a clear and concise representation of the proportion of each sentiment category, making it easy to grasp the relative size of each sentiment group in our dataset.

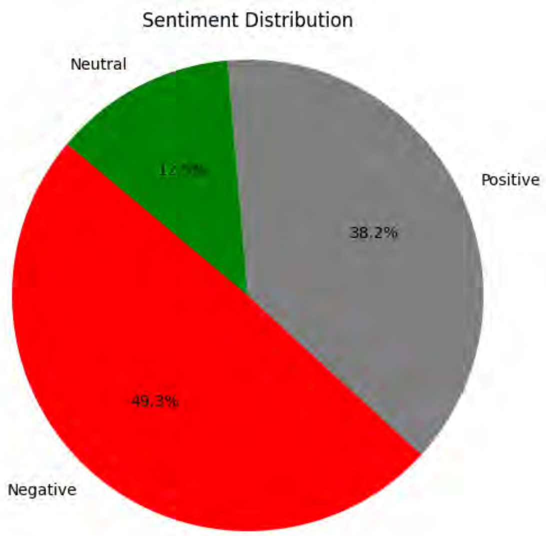

Figure 4.7: Representation of the proportion of each sentiment category

Word cloud: This word cloud is a visual tool that displays frequently occurring words in customer reviews, aiding in identifying common themes or issues discussed in the reviews.

Figure 4.8: Visualizing most frequently occurring words in the customer reviews

Scatter plot: This scatter plot with ratings on the y-axis and review text lengths on the x-axis can help identify any potential correlation between review length and rating, facilitating deeper insights into our data.

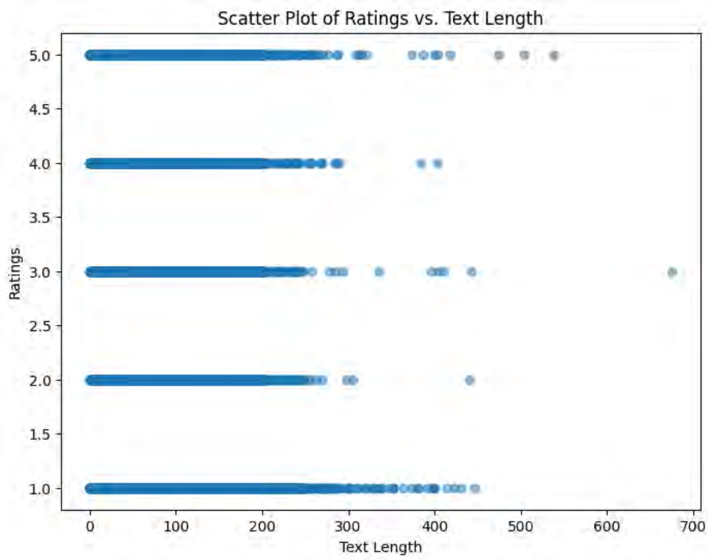

Figure 4.9: Correlation between review length and rating

Heatmap: This heatmap, with ratings on the x-axis and sentiment labels on the y-axis, visualizes the count of reviews with specific combinations of ratings and sentiments, allowing for a comprehensive understanding of the relationship between these two variables.

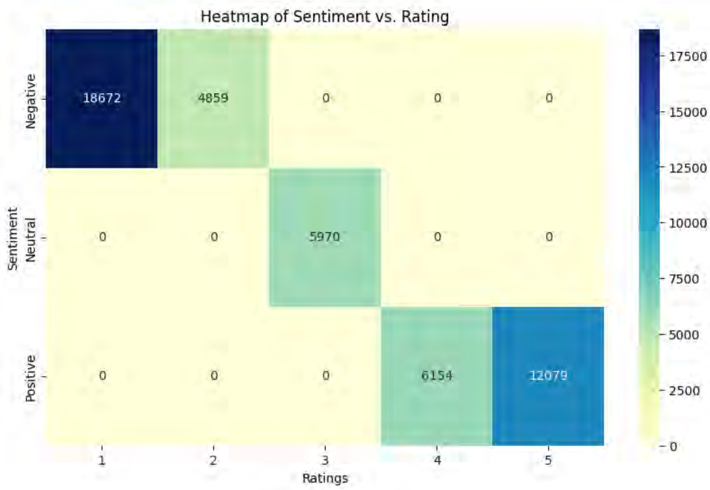

Figure 4.10: Heatmap of Sentiment vs Rating

Sentiment Distribution by text length: Using this bar chart to visualize how sentiment distribution varies with different text lengths can provide insights into how review sentiments are influenced by the length of the text.

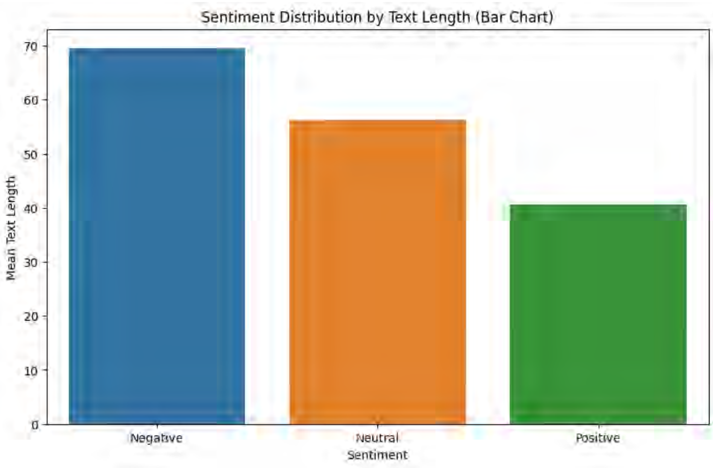

Figure 4.11: Sentiment distribution by text length

Wordcloud for sentiments: These 3 word clouds for positive, negative, and neutral sentiments respectively highlight the most frequently used words in positive, negative, and neutral reviews, enabling the identification of common positive themes.

Figure 4.12: Wordcloud for positive sentiments

Figure 4.13: Wordcloud for negative sentiments

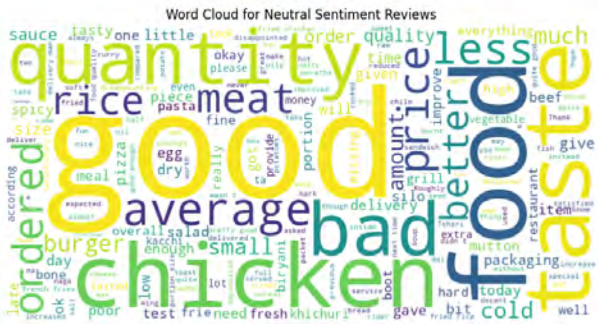

Figure 4.14: Wordcloud for neutral sentiments

Stacked Bar Chart: This stacked bar chart can effectively illustrate how sentiment distribution varies with different review text lengths (e.g., short, medium, long), providing a comprehensive overview of sentiment trends across different text lengths.

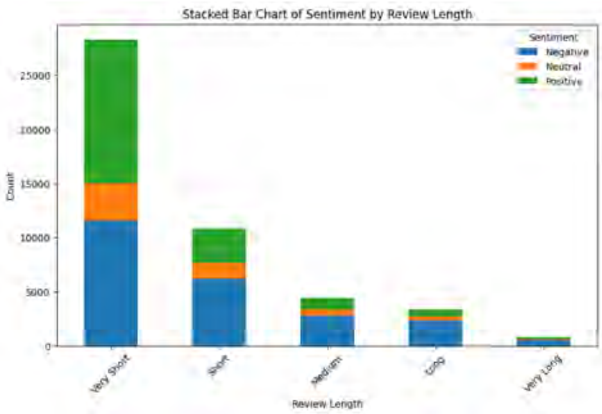

Figure 4.15: Stacked bar chart of sentiment by review length

Review length vs Sentiment (Box plot): This box plot is an effective way to visualize how review text lengths differ among different sentiment categories, helping to identify potential patterns or outliers.

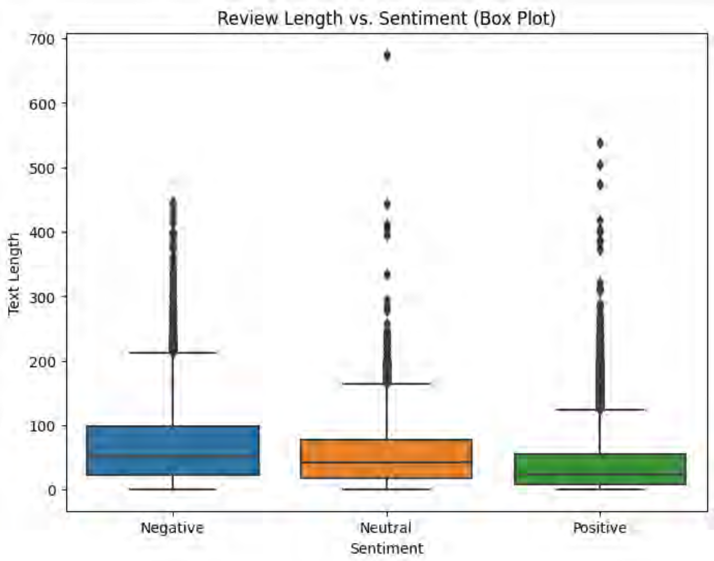

Figure 4.16: Review length vs Sentiment

Positive sentiment words: This bar chart has been used to display the most frequently mentioned words in positive sentiment reviews, providing actionable insights when setting a threshold for the top N words.

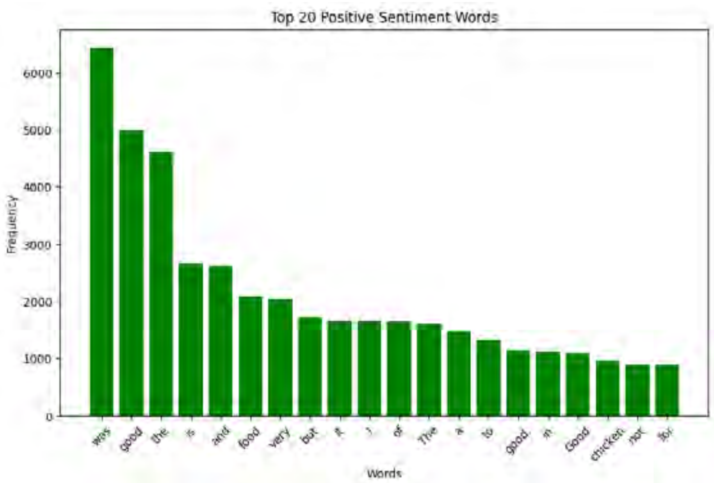

Figure 4.17: Top 20 positive sentiment words

Sentiment transition matrix: This sentiment transition matrix shows how sentiments transition from one review to the next, helping to identify patterns in how positive, negative, and neutral sentiments evolve.

Figure 4.18: Sentiment transition matrix

Emotion analysis pie chart using TextBlob: Using this TextBlob for sentiment analysis, we have created a pie chart to visualize the distribution of emotions (e.g., happy, sad, neutral) in reviews, offering a more nuanced view of sentiment.

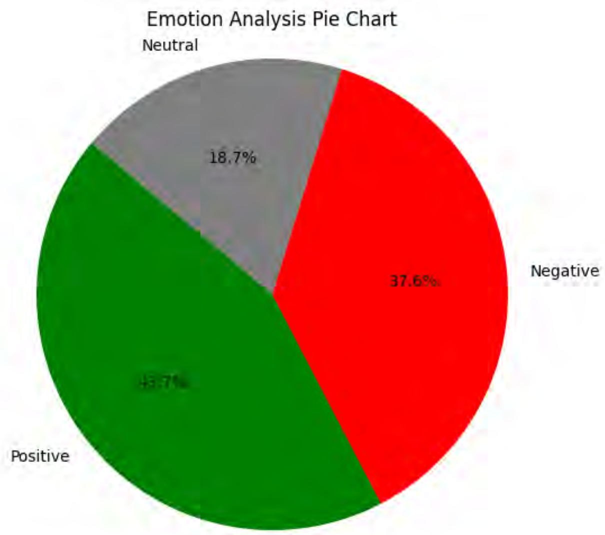

Figure 4.19: Emotion analysis pie chart

Chapter 5

Performance Analysis

5.1 Sentiment Analysis Approach: NLTK

5.1.1 Basic NLP Operations

We have conducted basic NLTK operations to acquire a fundamental understanding of our dataset. The steps have been explained below in order:

1. Tokenization: We have tokenized the customer reviews using the NLTK word tokenization tool.
2. Tagging: The tokens have been passed to the NLTK pos tag tool for tagging.
3. Entity: The tagged words have then been passed to the NLTK tagging tool to identify entities.

Sentence	The packaging was so good. But the amount of rice so.
After tokenization:	['The', 'packaging', 'was', 'so', 'good.But', 'the', 'amount', 'of', 'rice', 'so']
After POS tagging	[('The', 'DT'), ('packaging', 'NN'), ('was', 'VBD'), ('so', 'RB'), ('good.But', 'JJ'), ('the', 'DT'), ('amount', 'NN'), ('of', 'IN'), ('rice', 'NN'), ('so', 'RB')]
Entity recognition	The/DT packaging/NN was/VBD so/RB good.But/JJ the/DT amount/NN of/IN rice/NN

Figure 5.1: Tokenization, Tagging, and Entity Recognition on dataset

5.1.2 VADER: Valence Aware Dictionary and Sentiment Reasoner

This is a bag-of-word approach to do sentiment analysis. It uses a lexicon-based approach, assigning polarity scores to individual words and then combining them to determine the overall sentiment of a text.

In this approach, we have utilized the SentimentIntensityAnalyzer tool from the NLTK Sentiment. We have employed the polarity_scores function from the Natural Language Toolkit (NLTK) library in Python to analyze the sentiment of a given text. In this particular case, customer reviews have been passed as input, and the function has returned a dictionary containing four values: 'neg' (negative sentiment score), 'neu' (neutral sentiment score), 'pos' (positive sentiment score), and 'compound' (a compound score that combines the three aforementioned scores).

VADER Compound Score Visualization:

5.1.2.1 Compound Score Graph: The bar chart reveals that VADER performed admirably in detecting sentiments ranging from 1 to 5 stars. The provided compound scores from VADER encompass a combination of negative, neutral, and positive scores.

- The 1-star sentiment is highly negative, as evidenced by the significantly low bar.
- Similarly, the 2-star sentiment is negative, with the bar positioned below 0.
- The 3-star sentiment represents a neutral point, with the bar neither notably high nor low.
- The 4-star sentiment constitutes a positive review, indicated by the high bar.
- The 5-star sentiment reflects an extremely positive review, with the score prominently high.

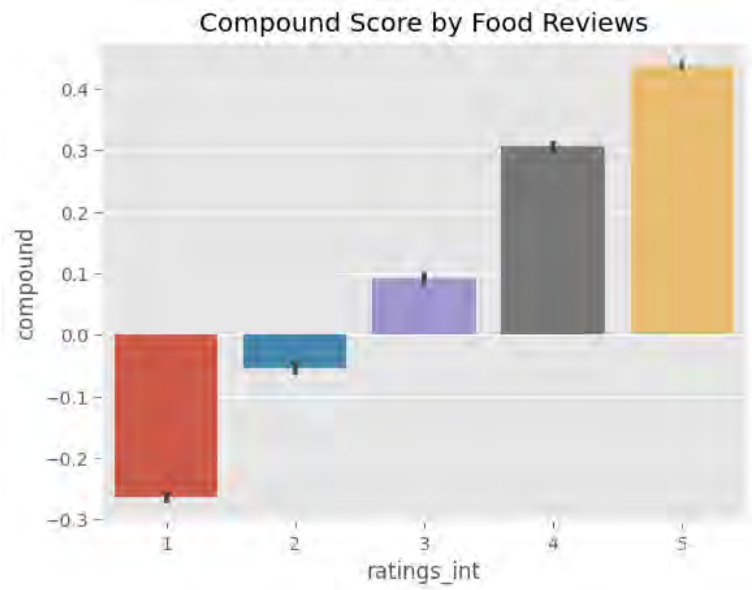

Figure 5.2: Sentiment analysis with VADER: unveiling varied emotions across star ratings

5.1.2.2 Positive, Neutral, and Negative Individual Graph:

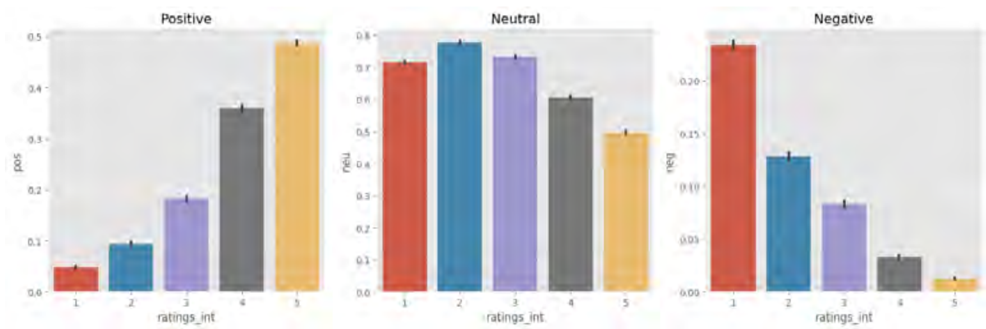

Figure 5.3: Exploring Positive, Neutral, and Negative trends in star ratings

From the graphs, we can observe the individual components.

- Positive: 5 and 4 stars exhibit a highly positive trend, evident from the first graph where the bars corresponding to 5 and 4 stars are significantly higher compared to other star ratings.
- Neutral: This represents the midpoint of sentiment. In this graph, the scores across 1-5 stars are closely clustered, indicating a well-balanced distribution. It essentially depicts a good bar chart, characteristic of neutral scores.
- Negative: In this graph, the bars for 1 and 2 stars should be higher, while others should be lower. This pattern is discernible from the graph, affirming that VADER demonstrates effective negative detection capabilities.

Note: These operations involve basic Natural Language Processing (NLP) techniques for sentiment analysis and do not employ a machine learning approach.

5.1.3 RoBERTa

We have utilized the pre-trained RoBERTa Transformer model for sentiment analysis. We have not trained the model's weights; instead, we have retrieved the model along with its weights and passed our text data to it to obtain sentiment scores. The following steps have been followed to employ this pre-trained RoBERTa model:

- The model and tokenizer have been downloaded from the pre-trained model.
- Text has been passed to the model to obtain sentiment scores for negative, neutral, and positive sentiments.

Pairplot of VADER and RoBERTa Sentiment Analysis for Comparison:

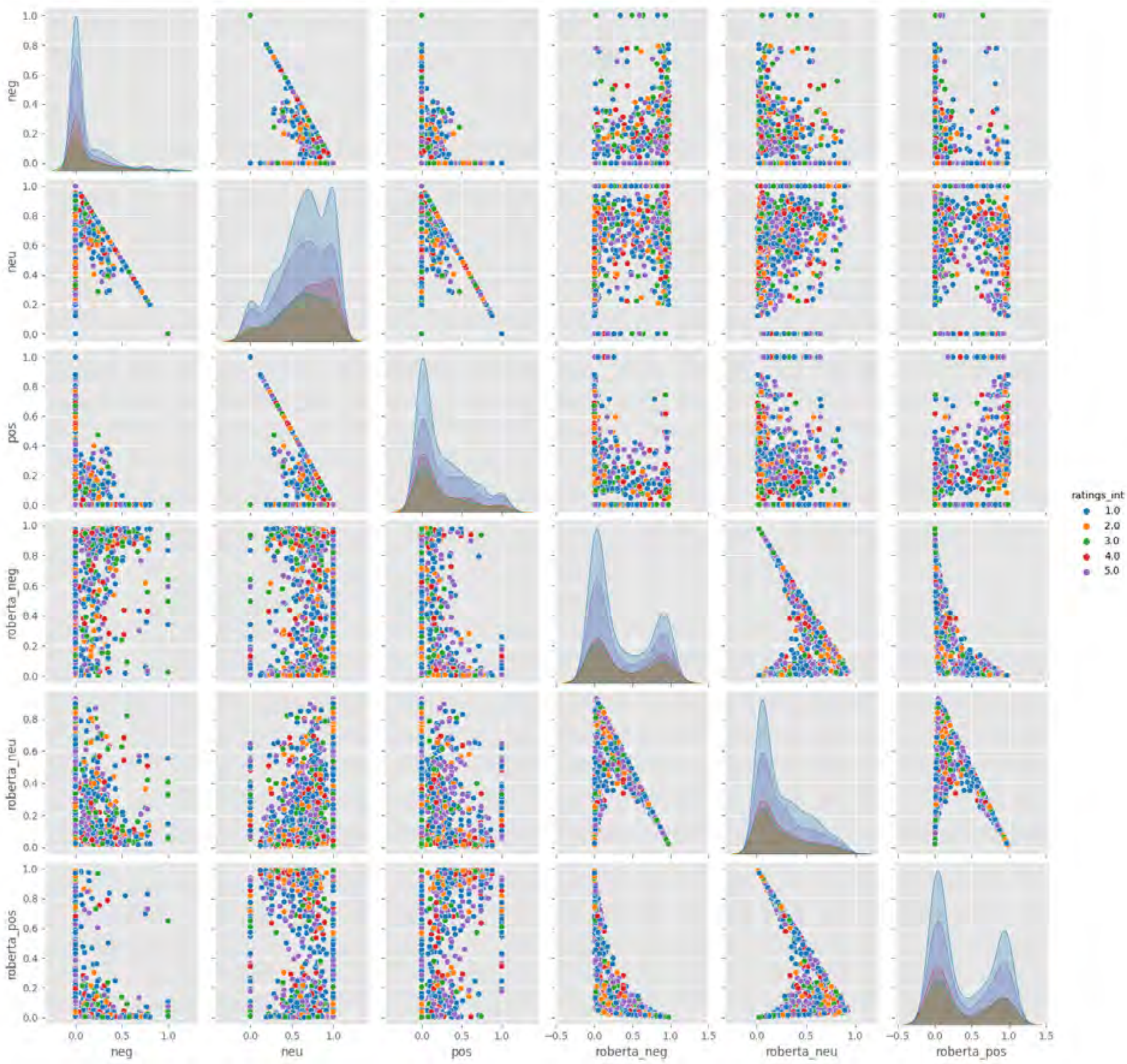

Figure 5.4: Comparative sentiment analysis distribution: VADER vs. RoBERTa models

We have observed the correlation between these two sentiment analyses. The pink dots represent 5-star reviews, while the blue dots represent 1-star reviews. Each cell of the diagonal has displayed a scatter plot comparing two variables: vader_neg, vader_neu, and vader_pos from VADER, and roberta_neg, roberta_neu, and roberta_pos from RoBERTa.

- Both models seem to have similar distributions for the neutral variable, with most data points concentrated around the center of the plot.
- Concerning the positive variable, the RoBERTa model appears to have exhibited a higher concentration of data points in the upper-right quadrant of the plot, indicating more positive sentiment scores.
- Regarding the negative variable, the VADER model seems to have displayed a higher concentration of data points in the lower-left quadrant of the plot, suggesting more negative sentiment scores.

5.1.4 Worst Case Example (RoBERTa vs. VADER)

5.1.4.1 Example of a review where actual rating was 1 but RoBERTa and VADER detected that as positive:

Steps:

1. Filtered all 1-star ratings.
2. Retrieved all 1-star ratings identified as positive.
3. Identified the highest positive value from RoBERTa and VADER.

Example we found:

1. *There was lettuce used in the burgers, pretty sure I filled out the allergies part with lettuce.*
2. *Awful means awry, special paratha or scam, and what about the time, an order was delivered in about 1 hour and 10 minutes.*

5.1.4.2 Example of a review where actual rating was 5 but RoBERTa and VADER detected that as negative:

Steps:

1. Filtered all 5-star ratings.
2. Retrieved all 5-star ratings identified as negative.
3. Identified the highest negative value from RoBERTa and VADER.

Example we found:

1. *Nasi goreng and Mee goreng had authentic flavor and taste with perfect heat. Highly recommended.*

5.2 Performance Metrics

5.2.1 Accuracy

This metric assesses a classifier's overall performance by measuring the ratio of correctly predicted instances to the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

5.2.2 Precision

This metric is defined as the ratio of true positive predictions to the total number of positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

5.2.3 Recall

Recall It measures the proportion of actual positive instances that were correctly predicted as positive by the model in a dataset.

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

where TP is the number of true positive predictions, TN is the number of true negative predictions, FP is the number of false positive predictions, and FN is the number of false negative predictions.

5.2.4 F1 score

The F1 score is the harmonic mean of precision and recall. It balances these two metrics, providing a single value that considers both false positives and false negatives.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5.4}$$

5.3 Performance Evaluation: Machine Learning Approach

5.3.1 Results: Pre-Data Optimization

In this section, we present a comprehensive comparison of the performance of five different classification models on our datasets before optimization. Table 5.3 illustrates the results obtained from that dataset.

Gradient Boosting Classifier In this case, the Gradient Boosting Classifier model exhibited an accuracy of 76.96% on this dataset, with a precision of 66% and a recall of 61%. This combination resulted in the F1 score of 61%, indicating room for improvement.

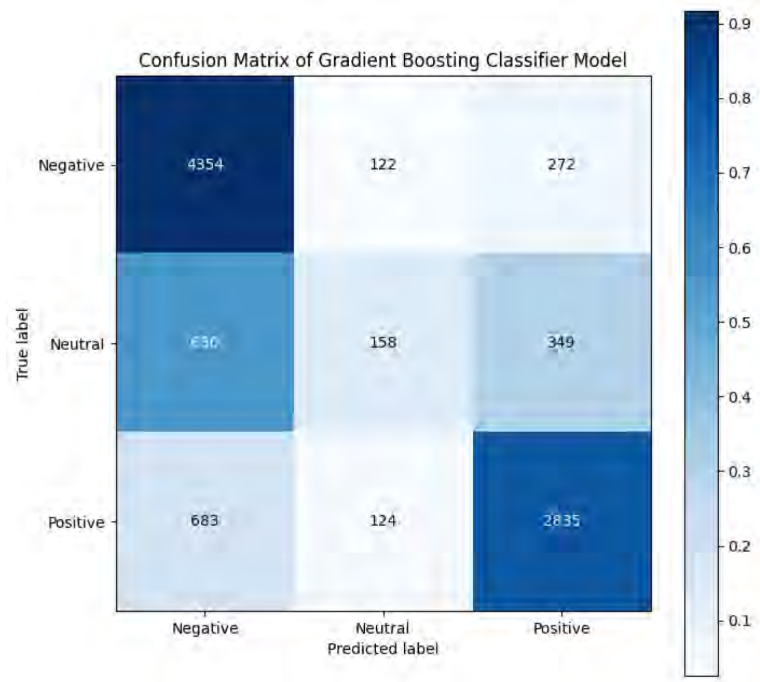

Figure 5.5: Confusion Matrix of Gradient Boosting Classifier

Random Forest Classifier: Then, the Random Forest Classifier achieved an accuracy of 76.74%. Its precision and recall were close at 67% and 60% respectively, resulting in the F1 score of 60%.

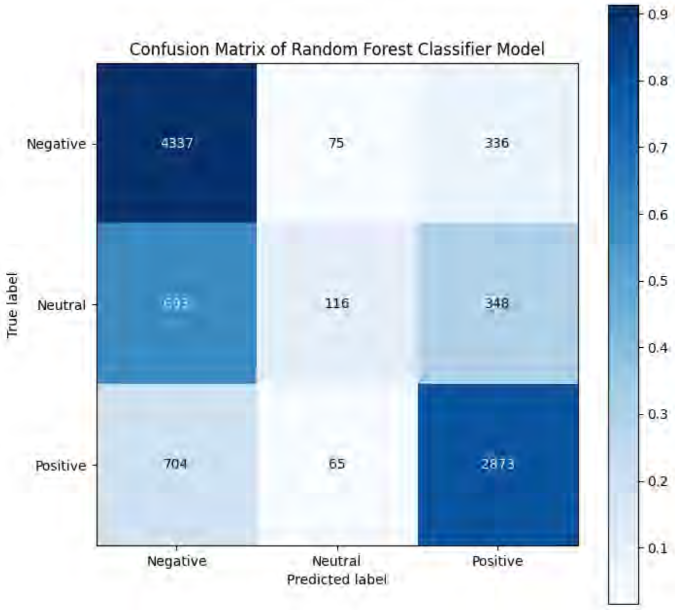

Figure 5.6: Confusion Matrix of Random Forest Classifier

Gaussian Naive Bayes: In addition, the Gaussian Naive Bayes model achieved an accuracy of 74.67%, displaying a precision of 62% and recall of 56%. This imbalanced recall negatively impacted the F1 score, which was 53%.

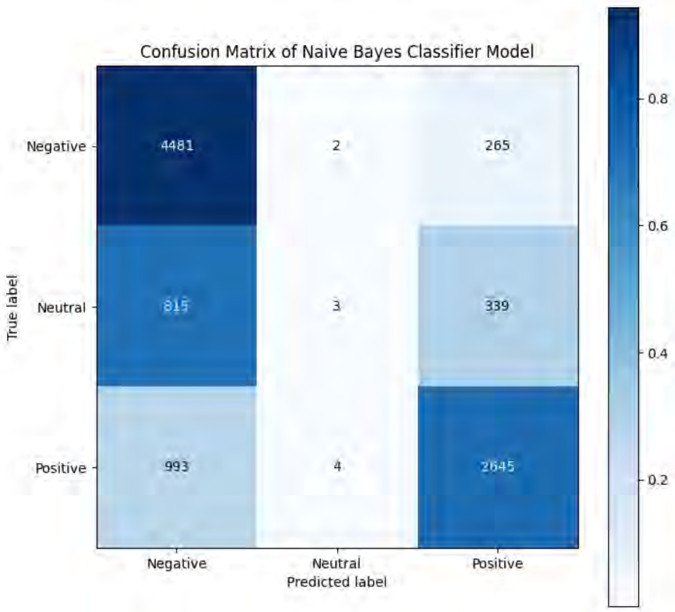

Figure 5.7: Confusion Matrix of Gaussian Naive Bayes Classifier

Logistic Regression: Moreover, the Logistic Regression model achieved an accuracy of 76.37%. While it exhibited precision and recall values at 65% and 62% respectively, the F1 score was relatively lower at 62%.

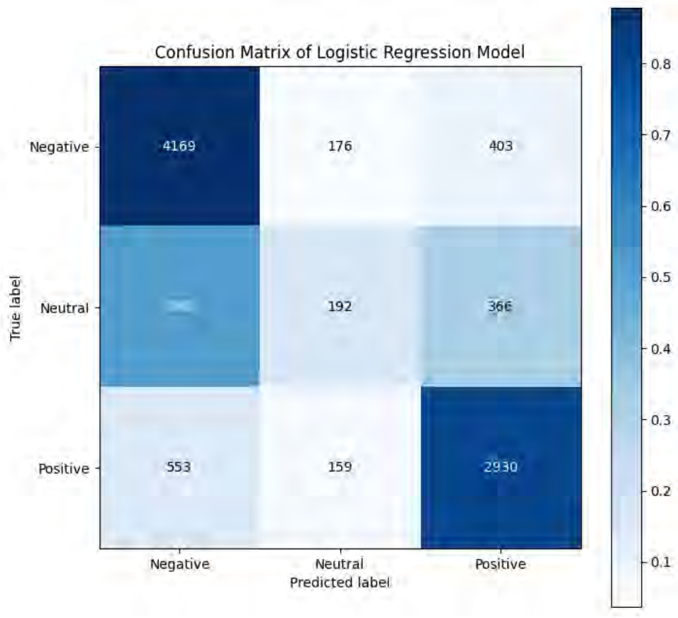

Figure 5.8: Confusion Matrix of Logistic Regression Classifier

K-Nearest Neighbors (KNN): Furthermore, the K-Nearest Neighbors model exhibited an accuracy of 65.09%, with a precision of 55% and a low recall of 56%. This combination resulted in an F1 score of 55%.

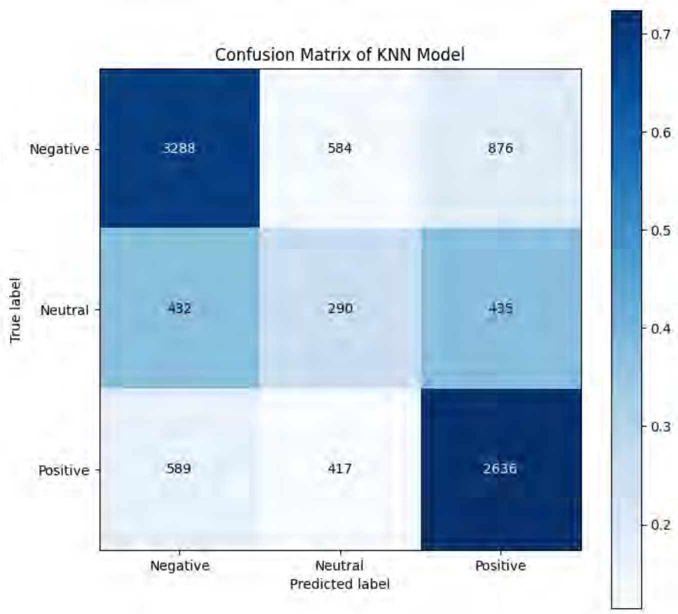

Figure 5.9: Confusion Matrix of K-Nearest Neighbors Classifier

Support Vector Machine (SVM): Besides, the Support Vector Machine performed at an accuracy of 65.05% on this dataset, with a lower precision of 56% and recall 54%, resulting in the F1 score of 54%.

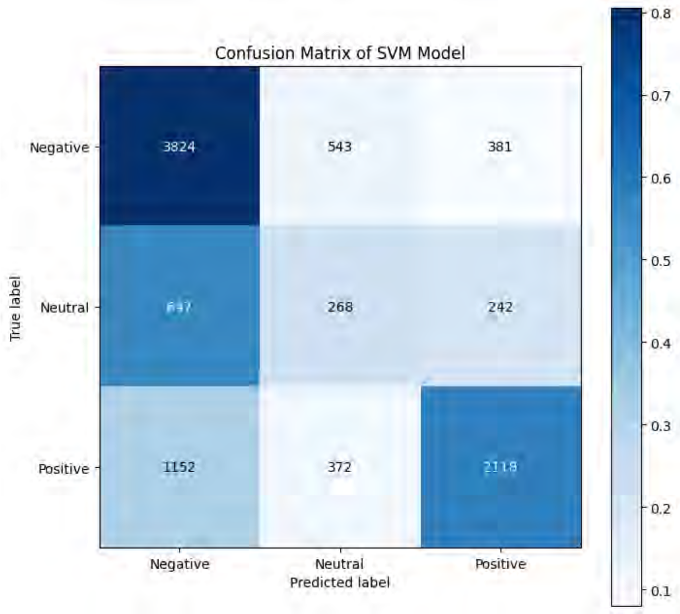

Figure 5.10: Confusion Matrix of Support Vector Machine Classifier

5.3.2 Results: Post-Data Optimization

In this section, we present a comprehensive comparison of the performance of five different classification models on our datasets after optimization. Table 5.4 illustrates the results obtained from that dataset.

Gradient Boosting Classifier Following dataset optimization, the model's accuracy slightly decreased to 75.26%. Both precision and recall improved to 75% each, yielding a more balanced F1 score of 75%. So, we can say that the improvement was notable.

Figure 5.11: Confusion Matrix of Gradient Boosting Classifier

Random Forest Classifier: However, upon optimization of the dataset, the model's performance improved across all metrics. The accuracy improved to 80.31%, with both precision and recall increasing to 80%. Consequently, the F1 score also rose to 80%.

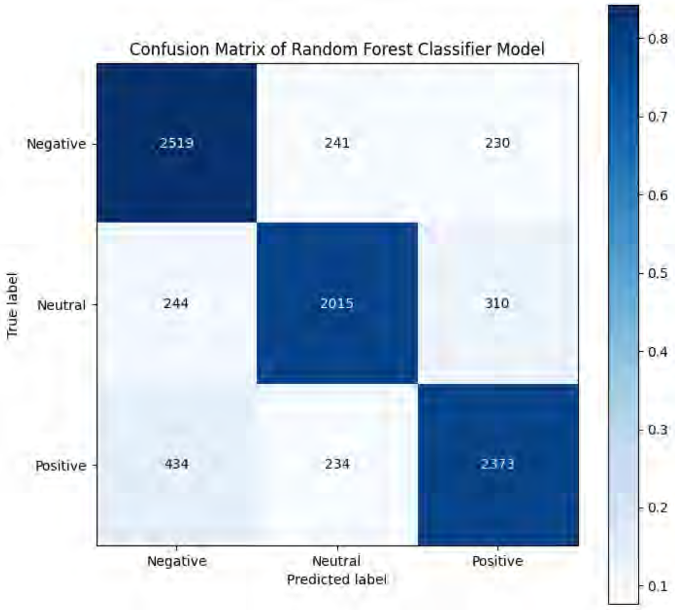

Figure 5.12: Confusion Matrix of Random Forest Classifier

Gaussian Naive Bayes: After dataset optimization, the model's accuracy dropped to 68.52%, showcasing the sensitivity of Naive Bayes class distribution changes. The precision and recall both slightly increased at 68%, leading to the F1 score of 68%.

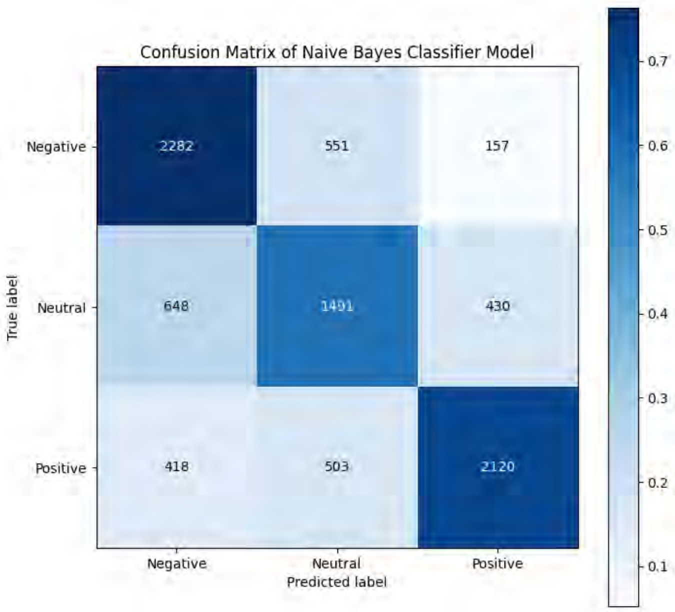

Figure 5.13: Confusion Matrix of Gaussian Naive Bayes Classifier

Logistic Regression: In contrast, the Logistic Regression model accuracy significantly decreased to 70.86%. While it exhibited precision, recall, and F1 score values at 70%.

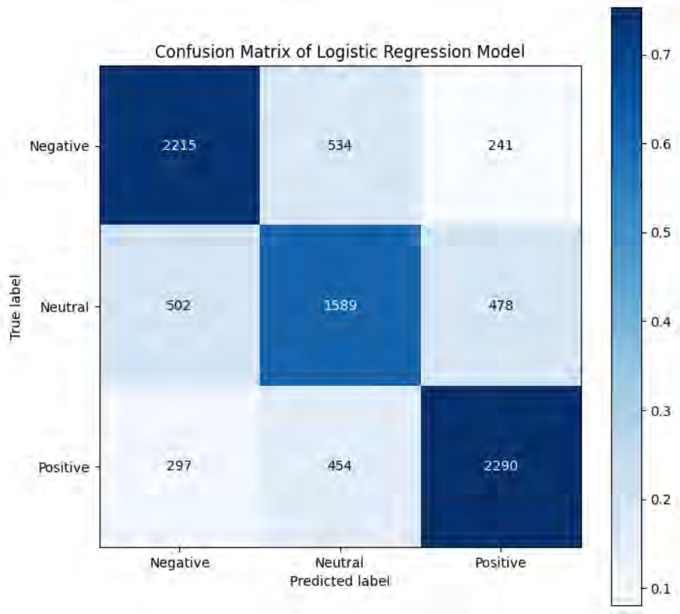

Figure 5.14: Confusion Matrix of Logistic Regression Classifier

Support Vector Machine (SVM): Lastly, the Support Vector Machine performed at an accuracy of 48.37% on this dataset which is relatively the worst, with a lower precision of 50% and recall 48%, resulting in an F1 score of 46%.

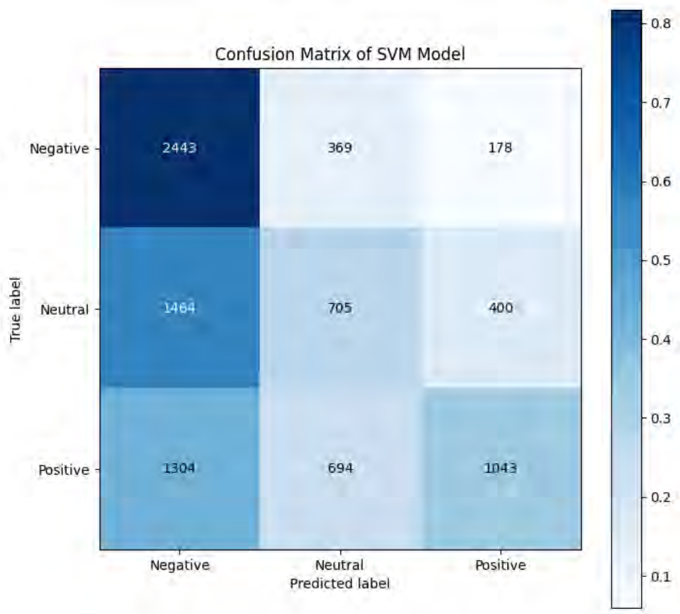

Figure 5.15: Confusion Matrix of Support Vector Machine Classifier

K-Nearest Neighbors (KNN): After dataset balancing, the KNN's accuracy substantially improved to 70.21%. The model exhibited both precision and recall at 71%, leading to the F1 score of 70%.

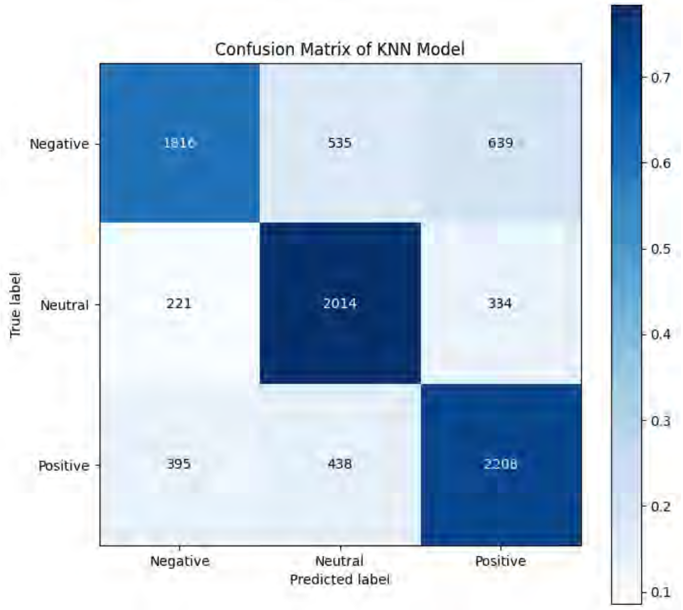

Figure 5.16: Confusion Matrix of K-Nearest Neighbors Classifier

Table 5.1: **Accuracy, Precision, Recall, F1 Score with pre-optimized dataset**

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Gradient Boosting Classifier	76.96%	66%	61%	61%
Random Forest Classifier	76.74%	67%	60%	60%
Gaussian Naive Bayes Classifier	74.67%	62%	56%	53%
Logistic Regression Classifier	76.37%	65%	62%	62%
K-Nearest Neighbors Classifier	65.09%	55%	56%	55%
SupportVectorMachine Classifier	65.05%	56%	54%	54%

Table 5.2: **Accuracy, Precision, Recall, F1 Score with post-optimized dataset**

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Random Forest Classifier	80.31%	80%	80%	80%
Gradient Boosting Classifier	75.26%	75%	75%	75%
K-Nearest Neighbors Classifier	70.21%	71%	71%	70%
Logistic Regression Classifier	70.86%	70%	70%	70%
Gaussian Naive Bayes Classifier	68.52%	68%	68%	68%
SupportVectorMachine Classifier	48.73%	50%	48%	46%

In summary, before optimizing the dataset, models like the Random Forest Classifier, Logistic Regression, Gaussian Naive Bayes Classifier, and Gradient Boosting Classifier showcased substantial enhancements in precision, recall, and F1 score. However, certain models like SupportVectorMachine and K-Nearest Neighbors Classifier exhibited a decrease in performance, underscoring the need for careful consideration when selecting algorithms for imbalanced datasets.

On the contrary, after the dataset optimization, models including the Random Forest Classifier, Logistic Regression, K-Nearest Neighbors Classifier, and Gradient Boosting Classifier showcased substantial enhancements in the precision, recall, and F1 score. However, certain models like SupportVectorMachine and Gaussian Naive Bayes Classifier exhibited a decrease in performance, underscoring the need for careful consideration when selecting algorithms for imbalanced datasets.

5.4 Performance Evaluation: Deep Learning Approach

5.4.1 Results: Pre-Data Optimization

Long Short-Term Memory (LSTM): In this pre-optimized dataset, the LSTM performed at an accuracy of 74%, with a precision of 66%, recall of 67%, and F-1 score of 67%, suggesting room for improvement.

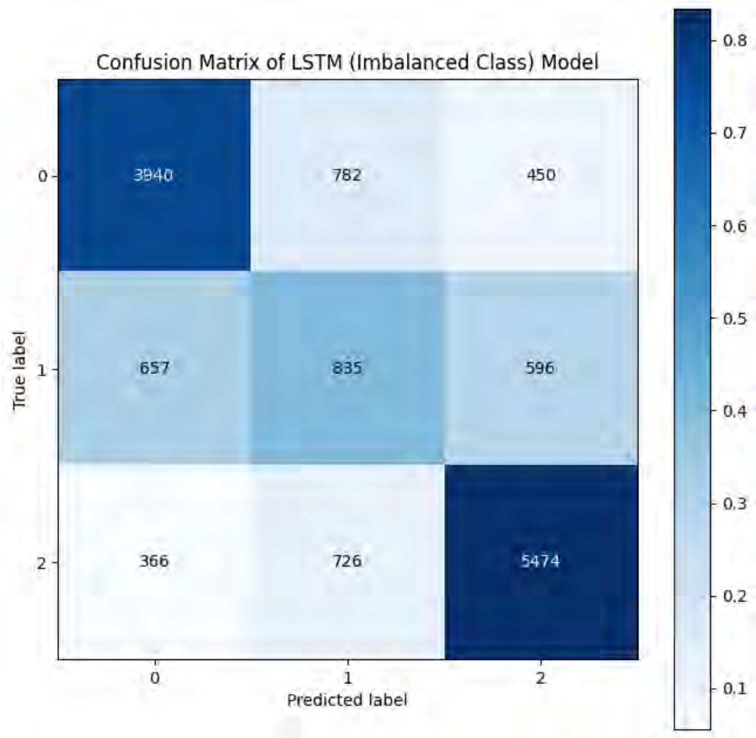

Figure 5.17: Confusion Matrix on imbalanced dataset of LSTM Classifier

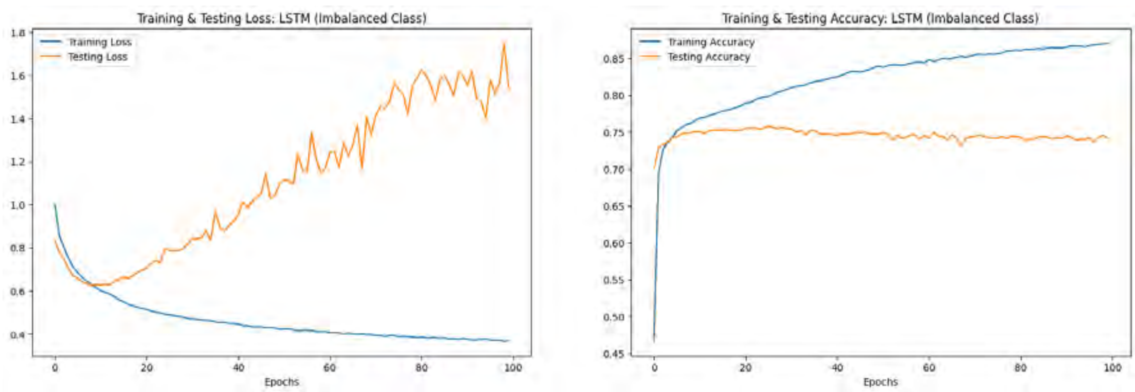

Figure 5.18: History plot on imbalanced dataset of LSTM Classifier

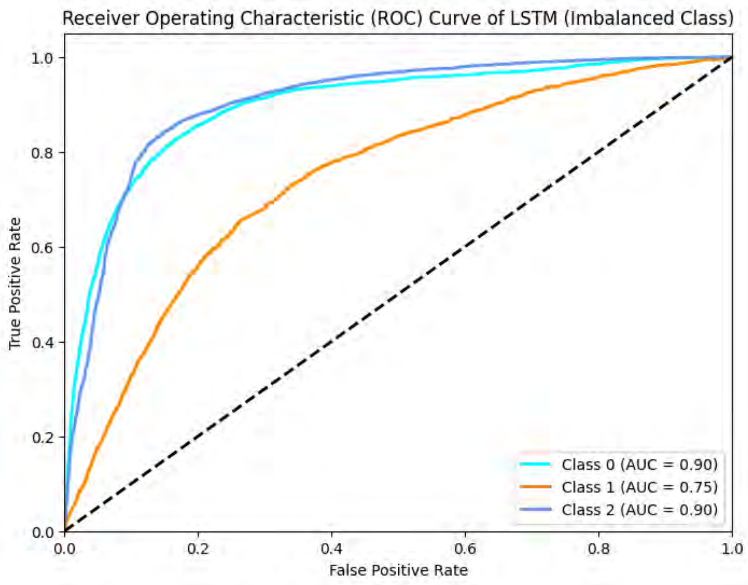

Figure 5.19: ROC Curve on imbalanced dataset of LSTM Classifier

Recurrent Neural Network (RNN): In this pre-optimized dataset, the RNN performed at an accuracy of 75%, with a precision of 67%, recall of 66%, and F-1 score of 66%.

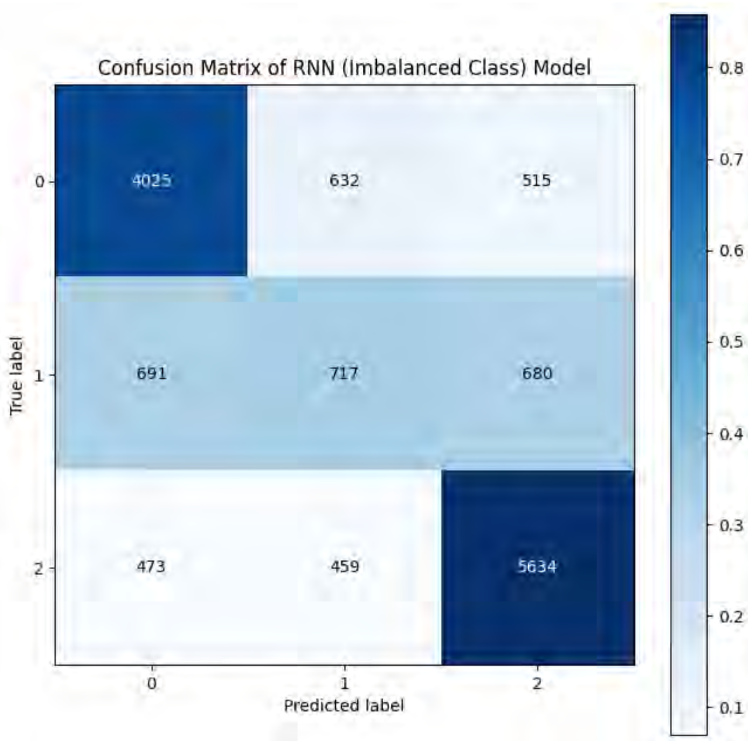

Figure 5.20: Confusion Matrix on imbalanced dataset of RNN Classifier

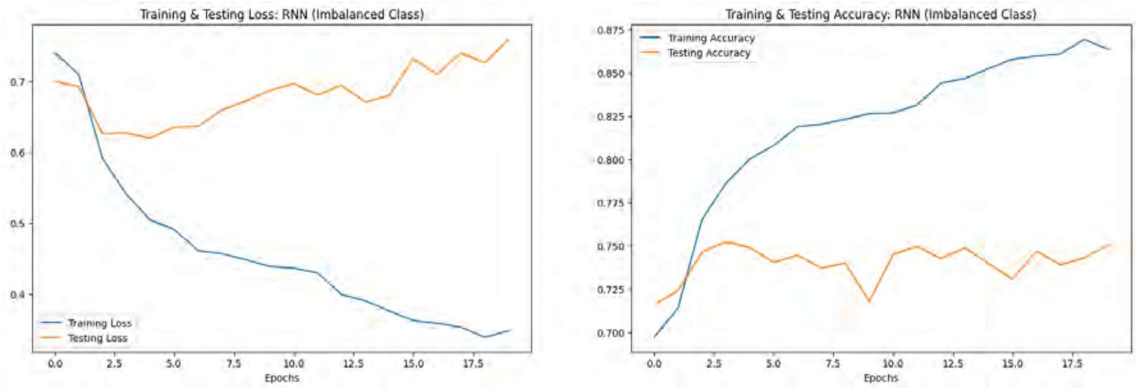

Figure 5.21: History plot on imbalanced dataset of RNN Classifier

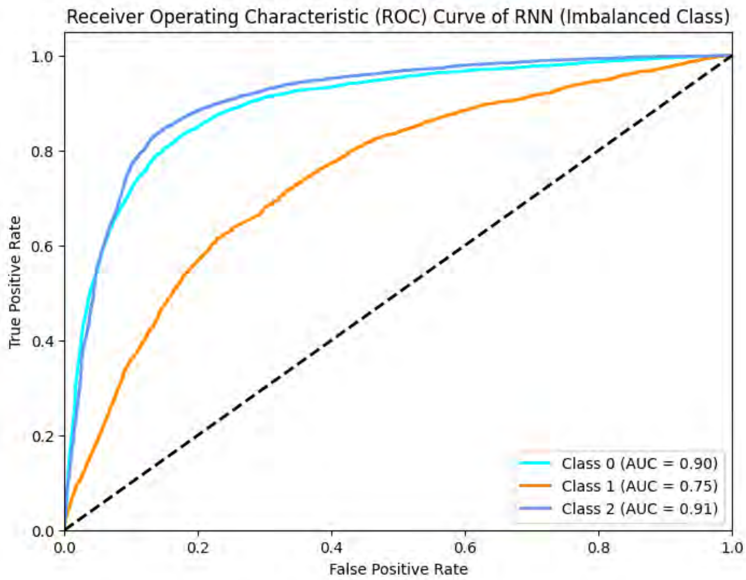

Figure 5.22: ROC Curve on imbalanced dataset of RNN Classifier

5.4.2 Results: Post-Data Optimization

Long Short-Term Memory (LSTM): In this post-optimized dataset, the LSTM performed at an accuracy of 79%, with a precision of 80% and a recall of 79%. This combination resulted in the F1 score of 79%, indicating room for improvement.

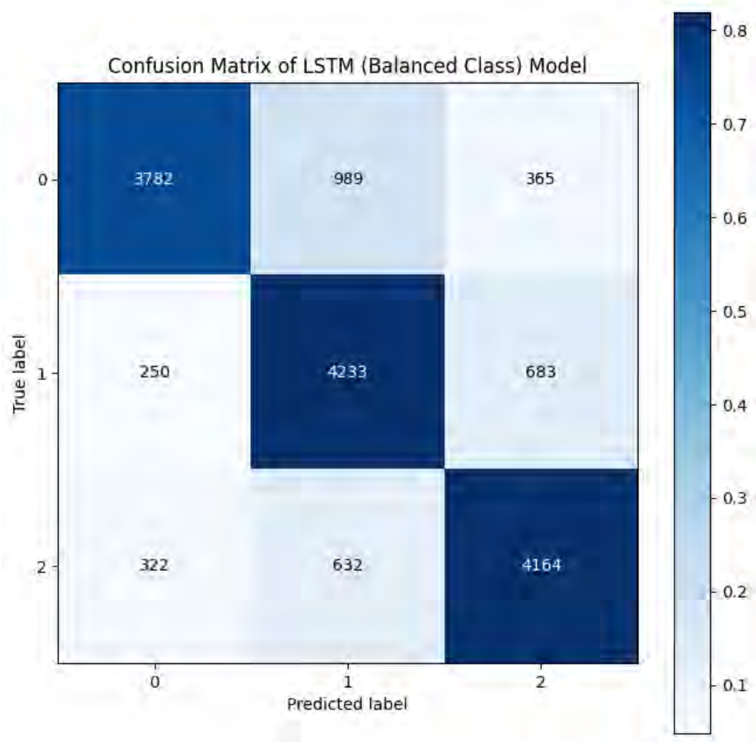

Figure 5.23: Confusion Matrix on balanced dataset of LSTM Classifier

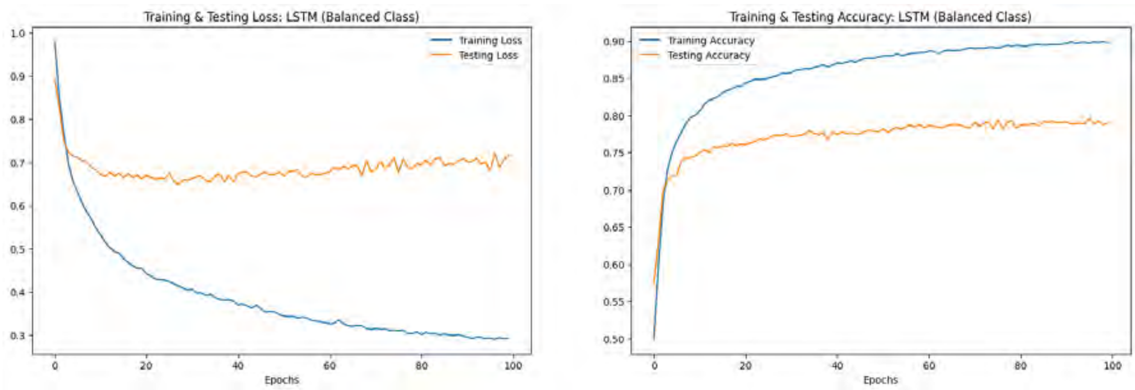

Figure 5.24: History plot on balanced dataset of LSTM Classifier

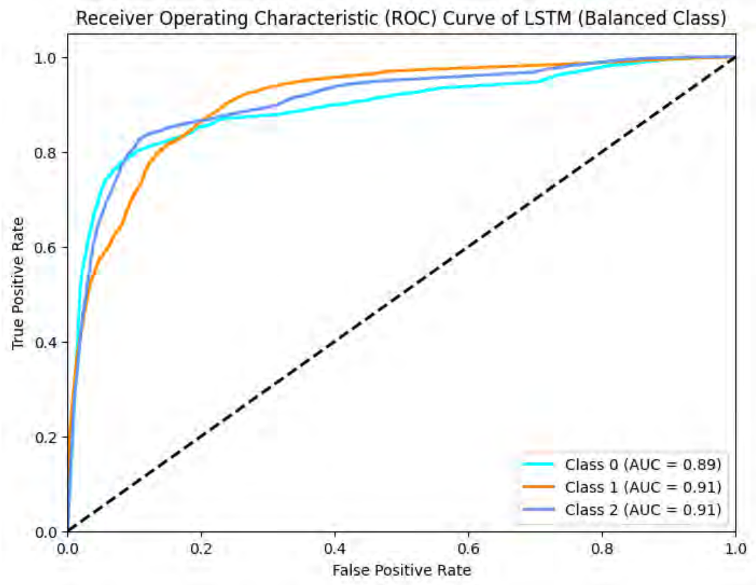

Figure 5.25: ROC Curve on balanced dataset of LSTM Classifier

Recurrent Neural Network (RNN): In this post-optimized dataset, the RNN performed at an accuracy of 77%, with precision, recall, and F1-score each of 77%.

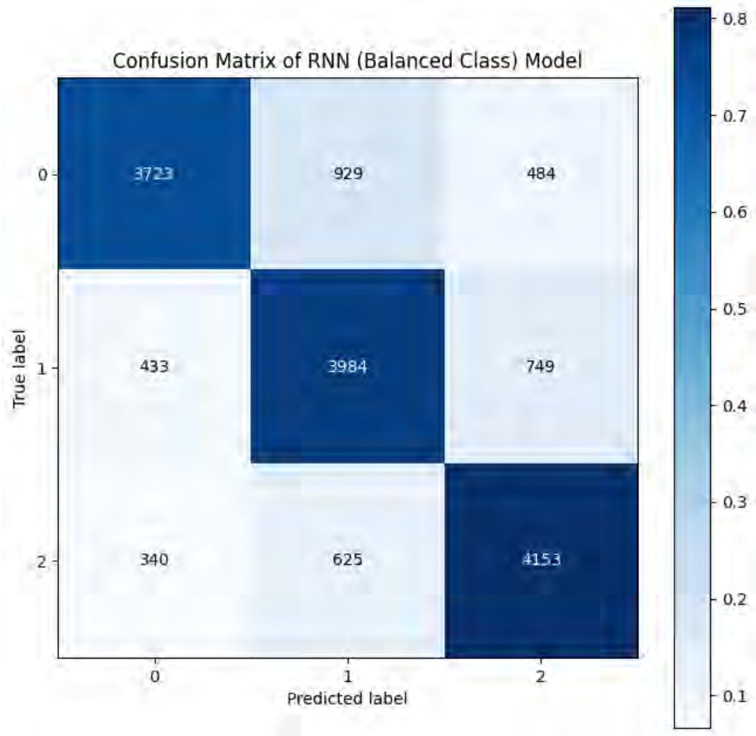

Figure 5.26: Confusion Matrix on balanced dataset of RNN Classifier

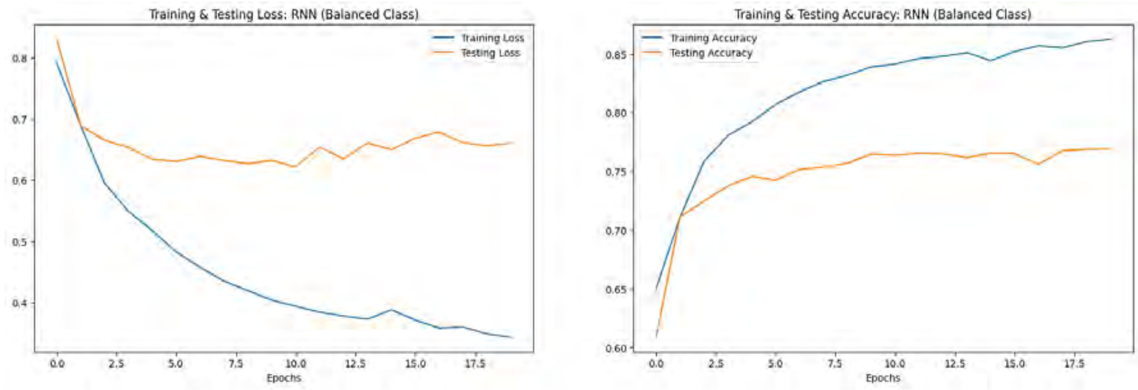

Figure 5.27: History plot on balanced dataset of RNN Classifier

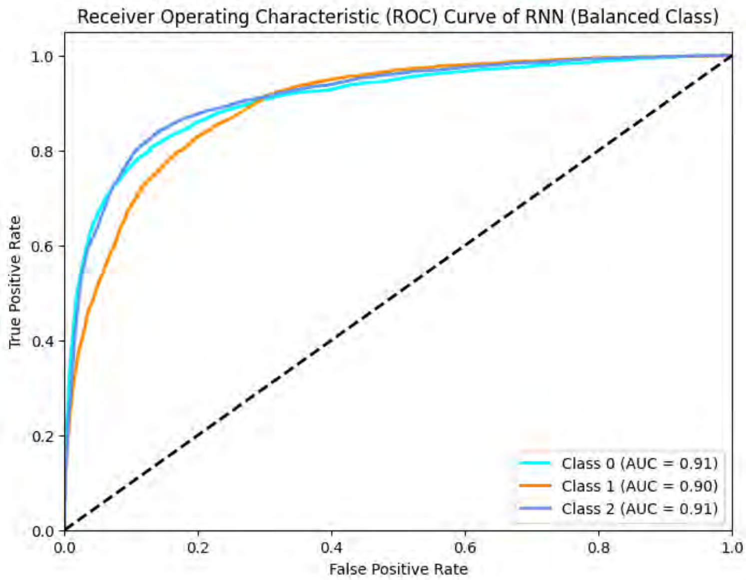

Figure 5.28: ROC Curve on balanced dataset of RNN Classifier

Bidirectional Encoder Representations from Transformers (BERT): In this post-optimized dataset, the BERT performed at an accuracy of 89%, with precision, recall, and F1-score each of 89%, which outperformed other models.

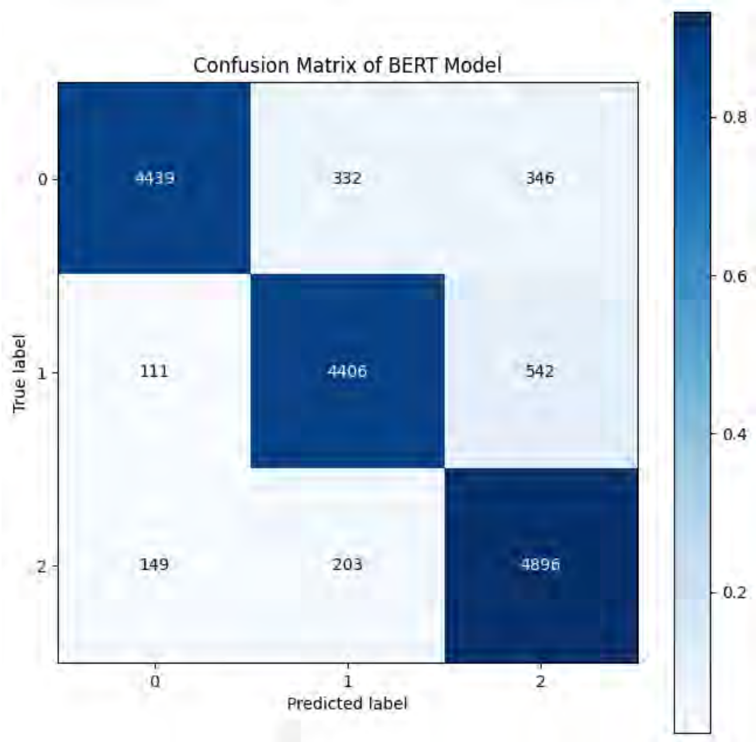

Figure 5.29: Confusion Matrix on balanced dataset of BERT Classifier

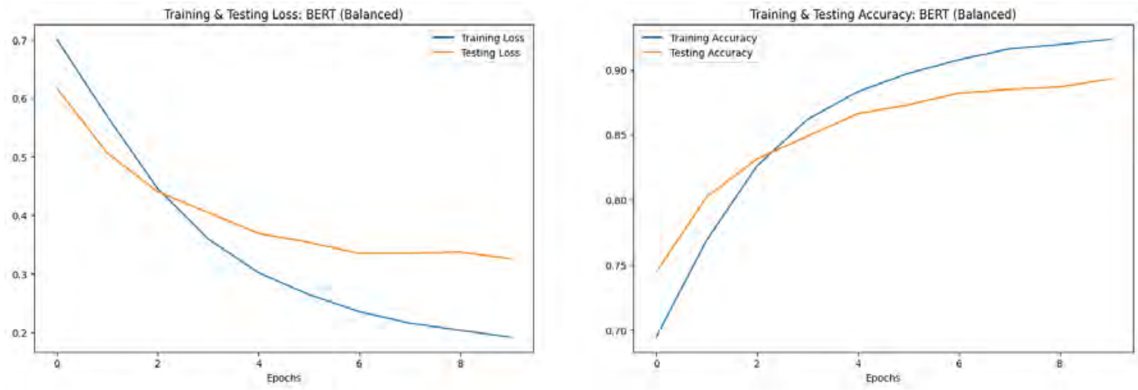

Figure 5.30: History plot on balanced dataset of BERT Classifier

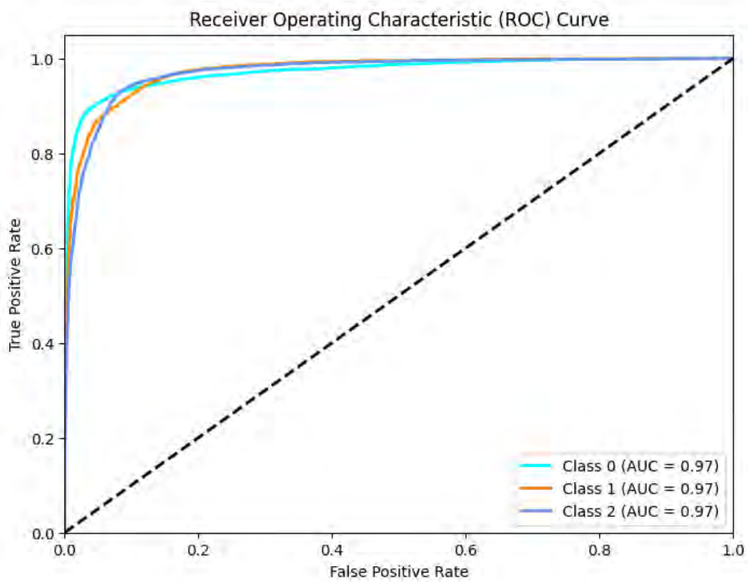

Figure 5.31: ROC Curve on balanced dataset of BERT Classifier

Table 5.3: **Accuracy, Precision, Recall, F1 Score with pre-optimized dataset**

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
LSTM Classifier	74%	66%	67%	67%
RNN Classifier	75%	67%	66%	66%

Note: The utilization of our imbalanced dataset for training the BERT model has been avoided, as an examination of the performance of our previous model indicated that satisfactory results were not achieved with an imbalanced dataset. Consequently, the decision has been made to exclude this aspect from our current approach.

Table 5.4: **Accuracy, Precision, Recall, F1 Score with post-optimized dataset**

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
BERT Classifier	89%	89%	89%	89%
LSTM Classifier	79%	80%	79%	79%
RNN Classifier	77%	77%	77%	77%

In conclusion, the BERT Classifier exhibits superior performance across accuracy, precision, recall, and F1 score, making it the optimal choice for the given task. However, the LSTM Classifier, with slightly lower but competitive metrics, remains a viable alternative, especially when computational resources are limited. The RNN Classifier, while simpler, falls behind in performance and may be chosen only if resource efficiency is a top priority, accepting a trade-off in overall model effectiveness.

Table 5.5: **Sentiment prediction using LSTM**

Review Sentence	Negative	Neutral	Positive	Prediction
I cannot say that the burger was bad. It was really tasty.	0.0070942	0.00863617	0.9842696	Positive
The pasta had very little chicken.	0.56334	0.34022307	0.09643696	Negative
Rice was not boiled perfectly.	0.86752933	0.09210572	0.04036488	Negative
They never disappoint in milkshakes.	1.5348438	1.9002538e-04	9.9981004e-01	Positive
The pizza was tasty.	0.0176545	0.18772379	0.79462165	Positive

Table 5.5 summarizes sentiment analysis results for diverse review sentences, presenting the model's prediction scores for negative, neutral, and positive sentiments. Notably, the model predicts positive sentiments for sentences expressing satisfaction with food, such as a tasty burger and pizza. Conversely, it accurately identifies negative sentiments for reviews indicating shortcomings in dishes like pasta with insufficient chicken and imperfectly boiled rice. The numerical values provide insights into the model's level of confidence in each sentiment prediction, offering a comprehensive overview of its performance.

Table 5.6: **Sentiment prediction using BERT**

Review Sentence	Negative	Neutral	Positive	Prediction
I cannot say that the burger was bad. It was really tasty.	0.07329996	0.82868564	0.09801441	Neutral
The pasta had very less chicken.	0.14138314	0.8010528	0.05756408	Neutral
Rice was not boiled perfectly.	0.3579989	0.5664827	0.07551836	Neutral
They never disappoint in milkshake.	0.04581137	0.01241603	0.94177264	Positive
The pizza was tasty.	0.02359549	0.10217245	0.87423205	Positive

Table 5.6 rates food-related review sentences on a scale of negative, neutral, and positive. Despite varied sentiment probabilities, the overall predictions suggest a neutral outlook for the first three sentences and a positive sentiment for the latter two, highlighting the model's interpretation of the reviews.

Chapter 6

Website Implementation

6.1 A Comprehensive Overview of Key Features

A website has been developed by us, featuring four primary functionalities:

1. **Top n Restaurants Listing:** This feature enables the display of the top n restaurants in a specific city based on their positive sentiment percentage. For instance, upon searching for the top 3 restaurants in Dhaka, the website generates a corresponding list.

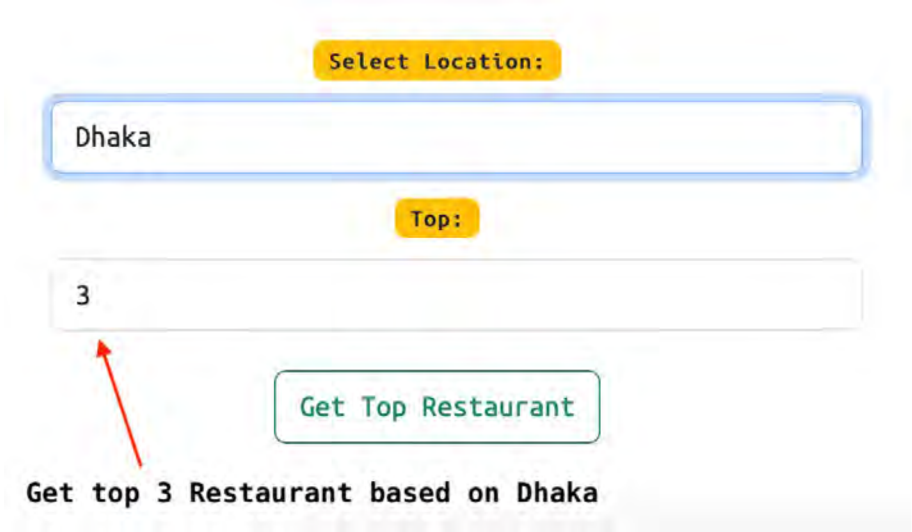

Figure 6.1: Searching top 3 Restaurants in Dhaka

S.No	Top Restaurants	Positive Sentiment Percentage
1	90sDhaka	76.66666666666667
2	CoopersDhanmondi27_Dhaka	73.33333333333333
3	HalalKitchen	72.41379310344827

Top 3 Restaurants in Dhaka

Figure 6.2: Showing top 3 Restaurants in Dhaka

2. **Restaurant Overall Sentiment Analysis:** This functionality involves the determination of a particular restaurant's sentiment through a table displaying the percentage of positive, negative, and neutral sentiments. This analysis aids in predicting the overall service quality of the restaurant. For instance, *Pizza Burg Dhanmondi* is identified as having a negative overall sentiment due to a higher prevalence of negative reviews.

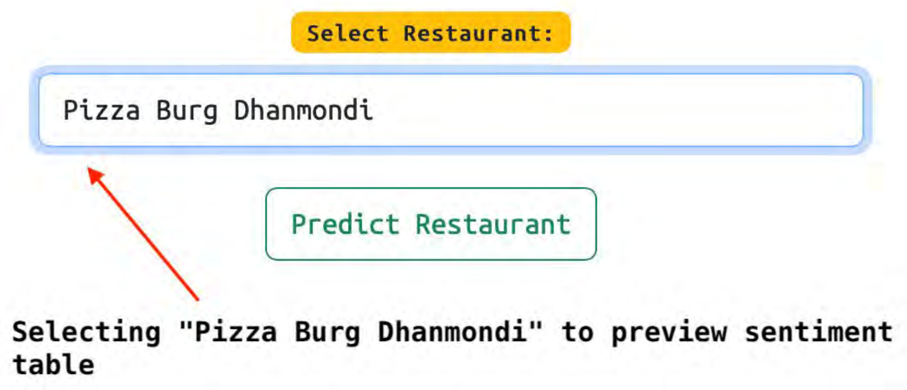

Figure 6.3: Searching sentiment table of Pizza Burg Dhanmondi

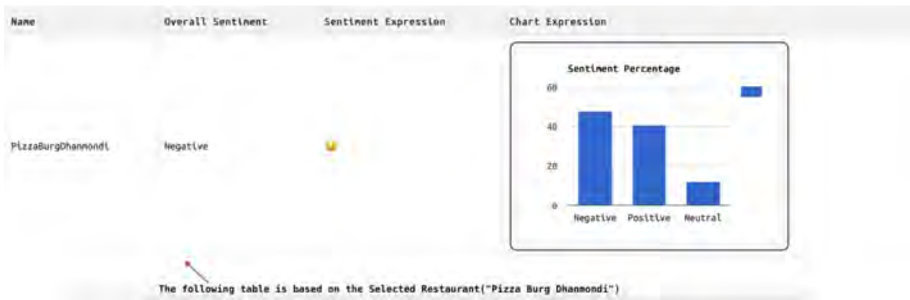

Figure 6.4: Showing sentiment table of Pizza Burg Dhanmondi

3. **Review Sentiment Prediction:** The system analyzes reviews to categorize them as positive, negative, or neutral, providing users with valuable insights into the sentiment conveyed in each review. For example here, *the food was tasty* - this review is predicted as positive in our model.

Figure 6.5: Review sentiment prediction for given review

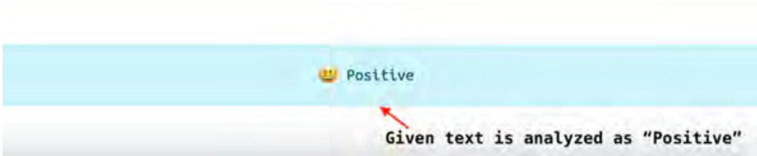

Figure 6.6: Review sentiment predicted for given review

4. **Statistical Insights:** This feature offers comprehensive statistical information, including the total number of enlisted restaurants, their locations, restaurant counts by location, and the average positive score by city within our dataset. This allows users to gain a holistic understanding of restaurant distribution and performance metrics.

Figure 6.7: Statistical information of our dataset

Chapter 7

Conclusion

Customers need to have awareness regarding both the quality of food and the standard of service offered by a restaurant before selecting meals for their convenience. This enables consumers to make well-informed selections regarding both the restaurant and the delivery service. In this study, we have constructed and assessed machine learning models to detect sentiment in text-based reviews on food ordering platforms in Bangladesh. In the scenario with the pre-optimized dataset, all models exhibited reasonable accuracy but demonstrated varying precision, recall, and F1 scores. The Random Forest Classifier, Logistic Regression, and Gradient Boosting Classifier displayed balanced performance, achieving the highest accuracy of around 77%. However, in the manually optimized dataset scenario, models showed overall improvement with the highest accuracy of 89%. The Random Forest Classifier excelled in accuracy, precision, and recall in the machine learning approach with an accuracy of 80.31%, and the BERT Classifier performed well in the deep learning approach with an accuracy of 89%, while SVM and RNN continued to face challenges due to their sensitivity to data imbalance. Random Forest excels in sentiment analysis of customer reviews due to its effective feature selection and robustness to overfitting. On the other hand, BERT excels in sentiment analysis on customer reviews due to its bidirectional context understanding, contextualized word embeddings, and pre-training on extensive datasets. In our research endeavor, we curated a distinctive and extensive dataset encompassing reviews from Bangladeshi food ordering portals, a resource previously unavailable in the academic domain. Our chosen model demonstrated superior performance compared to existing methodologies within the specific domain, as evidenced by robust performance metrics. Lastly, we engineered a website incorporating the aforementioned features, thus presenting a unique contribution to the field. In our future endeavors, we plan to enhance our platform by collecting more comprehensive data. Additionally, we will explore the integration of new machine learning models that we have not used due to their lower performance, optimizing their capabilities through meticulous tuning. Moreover, we will seek assistance from professional annotators who will help us properly annotate the reviews in our dataset. This is particularly crucial as some reviews exhibit improper annotation between the review and rating, leading to confusion. For instance, some reviews may appear neutral but are rated as positive or negative. The ultimate objective is to provide users with refined recommendations, highlighting the best restaurants based on specific culinary preferences on our website.

Bibliography

[1] A. T. Saad, "Factors affecting online food delivery service in bangladesh: An empirical study," *British Food Journal*, vol. 123, pp. 535–550, Sep. 2020. DOI: 10.1108/BFJ-05-2020-0449.

[2] O. Sharif, M. M. Hoque, and E. Hossain, "Sentiment analysis of bengali texts on online restaurant reviews using multinomial naïve bayes," in *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2019, pp. 1–6. DOI: 10.1109/ICASERT.2019.8934655.

[3] R. A. Laksono, K. R. Sungkono, R. Sarno, and C. S. Wahyuni, "Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes," in *2019 12th International Conference on Information and Communication Technology and System (ICTS)*, 2019, pp. 49–54. DOI: 10.1109/ICTS.2019.8850982.

[4] P. Sasikala and M. I. S. Lourdusamy, "An empirical study of sentiment analysis techniques for online reviews," *GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES*, vol. 5, no. 5, pp. 48–54, May 2018. DOI: 10.5281/zenodo.1249390. [Online]. Available: https://doi.org/10.5281/zenodo.1249390.

[5] V. Nama, V. Hegde, and B. Satish Babu, "Sentiment analysis of movie reviews: A comparative study between the naive-bayes classifier and a rule-based approach," in *2021 International Conference on Innovative Trends in Information Technology (ICITIIT)*, 2021, pp. 1–6. DOI: 10.1109/ICITIIT51526.2021.9399610.

[6] S. Vanaja and M. Belwal, "Aspect-level sentiment analysis on e-commerce data," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2018, pp. 1275–1279. DOI: 10.1109/ICIRCA.2018.8597286.

[7] E. R. Rhythm, R. A. Shuvo, M. S. Hossain, M. F. Islam, and A. A. Rasel, "Sentiment analysis of restaurant reviews from bangladeshi food delivery apps," in *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2023, pp. 1–5. DOI: 10.1109/ESCI56872.2023.10100214.

[8] H. M. Ahmed, M. Javed Awan, N. S. Khan, A. Yasin, and H. M. Faisal Shehzad, "Sentiment analysis of online food reviews using big data analytics," *Hafiz Muhammad Ahmed, Mazhar Javed Awan, Nabeel Sabir Khan, Awais Yasin, Hafiz Muhammad Faisal Shehzad (2021) Sentiment Analysis of Online Food Reviews using Big Data Analytics. Elementary Education Online*, vol. 20, no. 2, pp. 827–836, 2021.

[9] A. K. Bitto, M. H. I. Bijoy, M. S. Arman, I. Mahmud, A. Das, and J. Majumder, "Sentiment analysis from bangladeshi food delivery startup based on user reviews using machine learning and deep learning," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 4, pp. 2282–2291, 2023, ISSN: 2302-9285. DOI: 10.11591/eei.v12i4.4135. [Online]. Available: https://beei.org/index.php/EEI/article/view/4135.

[10] Pankaj, P. Pandey, Muskan, and N. Soni, "Sentiment analysis on customer feedback data: Amazon product reviews," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 320–322. DOI: 10.1109/COMITCon.2019.8862258.

[11] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*, C. C. Aggarwal and C. Zhai, Eds. Boston, MA: Springer US, 2012, pp. 415–463, ISBN: 978-1-4614-3223-4. DOI: 10.1007/978-1-4614-3223-4_13. [Online]. Available: https://doi.org/10.1007/978-1-4614-3223-4_13.

[12] Z. Singla, S. Randhawa, and S. Jain, "Statistical and sentiment analysis of consumer product reviews," in *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2017, pp. 1–6. DOI: 10.1109/ICCCNT.2017.8203960.

[13] A. U. Rehman, A. Malik, B. Raza, and W. Ali, "A hybrid cnn-lstm model for improving accuracy of movie reviews sentiment analysis," *Multimedia Tools and Applications*, vol. 78, Sep. 2019. DOI: 10.1007/s11042-019-07788-7.

[14] M. Trupthi, S. Pabboju, and G. Narasimha, "Sentiment analysis on twitter using streaming api," in *2017 IEEE 7th International Advance Computing Conference (IACC)*, 2017, pp. 915–919. DOI: 10.1109/IACC.2017.0186.

[15] Y. Woldemariam, "Sentiment analysis in a cross-media analysis framework," in *2016 IEEE International Conference on Big Data Analysis (ICBDA)*, 2016, pp. 1–5. DOI: 10.1109/ICBDA.2016.7509790.

[16] M. R. Khan, S. Nawsad Rahmatullah, M. F. Islam, A. R. M. Kamal, and M. A. Hossain, "Sentiment analysis of covid-19 vaccination in bangla language with code-mixed text from social media," in *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*, 2022, pp. 76–79. DOI: 10.1109/ICECE57408.2022.10088478.

[17] S. Mukherjee and P. Bhattacharyya, "Feature specific sentiment analysis for product reviews," vol. 7181, Jun. 2013, ISBN: 978-3-642-28603-2. DOI: 10.1007/978-3-642-28604-9_39.

[18] J. Jin and P. Ji, "Mining online product reviews to identify consumers' fine-grained concerns," in *12th International Symposium on Operations Research and its Applications in Engineering, Technology and Management (ISORA 2015)*, 2015, pp. 1–9. DOI: 10.1049/cp.2015.0622.

[19] S. Lee, H. Shin, I. Choi, and J. Kim, "Analyzing the impact of components of yelp.com on recommender system performance: Case of austin," *IEEE Access*, vol. 10, pp. 128 066–128 076, 2022. DOI: 10.1109/ACCESS.2022.3225190.

[20] J. Zheng and L. Zheng, "A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification," *IEEE Access*, vol. 7, pp. 106 673–106 685, 2019. DOI: 10.1109/ACCESS.2019.2932619.

[21] S. Reshma and K. Kalpana, "Analysing the emotions of food products reviewsusing nlp and adaboost algorithm," in *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, 2023, pp. 1–6. DOI: 10.1109/ACCAI58221.2023.10200516.

[22] J. Ara, M. T. Hasan, A. Al Omar, and H. Bhuiyan, "Understanding customer sentiment: Lexical analysis of restaurant reviews," in *2020 IEEE Region 10 Symposium (TENSYMP)*, 2020, pp. 295–299. DOI: 10.1109/TENSYMP50017.2020.9230712.

[23] S. M. Asiful Huda, M. M. Shoikot, M. A. Hossain, and I. J. Ila, "An effective machine learning approach for sentiment analysis on popular restaurant reviews in bangladesh," in *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, 2019, pp. 170–173. DOI: 10.1109/AiDAS47888.2019.8970976.

[24] M. R. Bhuiyan, M. H. Mahedi, N. Hossain, Z. N. Tumpa, and S. A. Hossain, "An attention based approach for sentiment analysis of food review dataset," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1–6. DOI: 10.1109/ICCCNT49239.2020.9225637.

[25] M. A. Shafin, M. M. Hasan, M. R. Alam, M. A. Mithu, A. U. Nur, and M. O. Faruk, "Product review sentiment analysis by using nlp and machine learning in bangla language," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, 2020, pp. 1–5. DOI: 10.1109/ICCIT51783.2020.9392733.

[26] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts," *Mathematical and Computational Applications*, vol. 23, no. 1, 2018, ISSN: 2297-8747. DOI: 10.3390/mca23010011. [Online]. Available: https://www.mdpi.com/2297-8747/23/1/11.

[27] R. Hu, L. Rui, P. Zeng, L. Chen, and X. Fan, "Text sentiment analysis: A review," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018, pp. 2283–2288. DOI: 10.1109/CompComm.2018.8780909.

[28] E. Tyagi and A. K. Sharma, "Sentiment analysis of product reviews using support vector machine learning algorithm," *Indian Journal of Science and Technology*, vol. 10, pp. 1–9, Jun. 2017. DOI: 10.17485/ijst/2017/v10i35/118965.

[29] R. Obiedat, R. Qaddoura, A. Al-Zoubi, *et al.*, "Sentiment analysis of customers' reviews using a hybrid evolutionary svm-based approach in an imbalanced data distribution," *IEEE Access*, vol. 10, pp. 1–1, Jan. 2022. DOI: 10.1109/ACCESS.2022.3149482.

[30] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Information Sciences*, vol. 488, Mar. 2019. DOI: 10.1016/j.ins.2019.02.064.

[31] F.-J. Yang, "An implementation of naive bayes classifier," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2018, pp. 301–306. DOI: 10.1109/CSCI46756.2018.00065.

[32] T. Haifley, "Linear logistic regression: An introduction," in *IEEE International Integrated Reliability Workshop Final Report, 2002.*, 2002, pp. 184–187. DOI: 10.1109/IRWS.2002.1194264.

[33] Y. Xiao, W. Huang, and J. Wang, "A random forest classification algorithm based on dichotomy rule fusion," in *2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, 2020, pp. 182–185. DOI: 10.1109/ICEIEC49280.2020.9152236.

[34] Q. Wang, "Support vector machine algorithm in machine learning," in *2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 2022, pp. 750–756. DOI: 10.1109/ICAICA54878.2022.9844516.

[35] I. tabarez paz, N. Hernández-Gress, and M. Gonzalez-Mendoza, "Classification of database by using parallelization of algorithms third generation in a gpu," vol. 459, Sep. 2014, pp. 25–38, ISBN: 978-3-319-11070-7. DOI: 10.1007/978-3-319-11071-4_3.

[36] H. Deng, Y. Zhou, L. Wang, and C. Zhang, "Ensemble learning for the early prediction of neonatal jaundice with genetic features," *BMC Medical Informatics and Decision Making*, vol. 21, Dec. 2021. DOI: 10.1186/s12911-021-01701-9.

[37] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 2019, pp. 1255–1260. DOI: 10.1109/ICCS45141.2019.9065747.

[38] A. Pulver and S. Lyu, "Lstm with working memory," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 845–851. DOI: 10.1109/IJCNN.2017.7965940.

[39] R. Kang, B. Park, Q. Ouyang, and N. Ren, "Rapid identification of foodborne bacteria with hyperspectral microscopic imaging and artificial intelligent classification algorithms," *Food Control*, vol. 130, p. 108 379, Jun. 2021. DOI: 10.1016/j.foodcont.2021.108379.

[40] A. Marium and G. S. Mamatha, "Bert model for classification of fake news using the cloud processing capacity," in *2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC)*, 2021, pp. 1–6. DOI: 10.1109/R10-HTC53172.2021.9641632.

[41] J. Sun, Y. Liu, J. Cui, and H. He, "Deep learning-based methods for natural hazard named entity recognition," *Scientific Reports*, vol. 12, p. 4598, Mar. 2022. DOI: 10.1038/s41598-022-08667-2.

[42] M. Kaur and A. Mohta, "A review of deep learning with recurrent neural network," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2019, pp. 460–465. DOI: 10.1109/ICSSIT46314.2019.8987837.

[43] W. Feng, N. Guan, Y. Li, X. Zhang, and Z. Luo, "Audio visual speech recognition with multimodal recurrent neural networks," May 2017, pp. 681–688. DOI: 10.1109/IJCNN.2017.7965918.

[44] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, “Understanding data augmentation for classification: When to warp?” In *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2016, pp. 1–6. DOI: 10.1109/DICTA.2016.7797091.

[45] D. Jadeja, A. Khetri, A. Mittal, and D. K. Vishwakarma, “Comparative analysis of transformer models on wikihow dataset,” in *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 2022, pp. 655–658. DOI: 10.1109/ICSCDS53736.2022.9761043.