Sentimental Analysis of Online Restaurant Reviews: Text Mining

Submitted by

Md. Shohan Bhuiyan

172-35-2165

Department of Software Engineering

Daffodil International University

Supervised By

Ms. Nusrat Jahan

Assistant Professor

Department of Software Engineering

Daffodil International University

This thesis report has been submitted in fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering

APPROVAL

This **Thesis** titled "**Sentimental Analysis of Online Restaurant Reviews: Text Mining**", submitted by **Md. Shohan Bhuiyan**, **ID: 172-35-2165** to the Department of Software Engineering, Daffodil International University has been accepted as pleasing for the partial fulfillment of the requirements for the degree of B.Sc. in software Engineering and approved as to its style and contents.

BOARD OF EXAMINERS

---   Chairman
Dr. Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Daffodil International University

---   Internal Examiner 1
Md Anwar Hossen
Assistant Professor
Department of Software Engineering
Daffodil International University
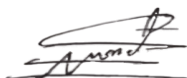
---   Internal Examiner 2
Asif Khan Shakir
Senior Lecturer
Department of Software Engineering
Daffodil International University

---   External Examiner
Professor Dr M Shamim Kaiser
Institute of Information Technology
Jahangirnagar University

---   Internal Examiner 3

Ms. Nusrat Jahan

Assistant Professor

Department of Software Engineering

Daffodil International University

Submitted by

Md. Shohan Bhuiyan

ID: 172-35-2165

Department of Software Engineering

Daffodil International University

THEIS DECLARATION

The thesis entitled "**Sentimental Analysis of Online Restaurant Reviews: Text Mining**" is done under the supervision of Ms. Nusrat Jahan, Assistant Professor, Department of Software Engineering, Daffodil International University.

I assert that the statements made and conclusions drawn are an outcome of the research work. I further declare that to the best of my knowledge and belief that the thesis report does not contain any part of any work which has been submitted for the award of any other degree/diploma/certificate in this University or any other University.

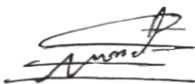
Submitted by

Md. Shohan Bhuiyan

ID: 172-35-2165

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

Certified by,

Ms. Nusrat Jahan

Assistant Professor

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

ACKNOWLEDGEMENT

First of all, I'm grateful to the Almighty Allah who created me and gave me the ability to complete this thesis. My parents have played a great role in bringing me this far. They have supported me through tough times and help me build my career in Software Engineering Sector. I have been able to stand here today with God's undeserved kindness and the blessings of my parents.

I would like to impulse my deepest thanks to my Advisor **Ms. Nusrat Jahan** for the consistent help of my thesis and research work, through his understanding, inspiration, energy, and knowledge sharing. She helped me to solve whatever any problem. Her direction helped me to finding the solutions of research work and reach to my final theory.

I would like to express my extreme sincere gratitude and appreciation to all of my teachers of Software Engineering department for their kind help, generous advice and support during the study.

My special gratitude goes to my friends, seniors, juniors who directly or indirectly have lent their helping hand in this venture.

Md. Shohan Bhuiyan

Table of Contents

List of Figure

List of Table

ABSTRACT

Abstract: Sentimental Analysis is a Natural Language Processing (NLP) procedure used to characterize records for recognizing positive or negative reviews. Customer happiness has recently emerged as one of the most important criteria in the restaurant industry's success. The importance of consumer feedback cannot be overstated. For the sake of social media, people are more motivated to read reviews before coming to a restaurant. Customers who want to choose a restaurant may read a lot of reviews to get a good idea of the restaurant's quality or services. As a result, a nostalgic classification of a large number of audits is required to achieve meaningful experiences, allowing customers to select a restaurant based on their preferences. Sentimental analysis can help with this classification. This research suggests a system for categorizing customer reviews into good and negative categories based on sentimental input. 1000 restaurant evaluations from Tripadvisor, foodpanda, foodbank, and other restaurant review sites were used to test the proposed solutions. In this paper Split Test, 20% Data Testing and 80% Data Training have been used. More specifically, the proposed system has been tested with four supervised algorithms of machine learning: Support Vector Machine (SVM), Multinomial Naïve Bayes, Random Forest and Decision Tree for sentiment classification of comments. The untried result shows that this proposed system can classify restaurant reviews with 71.50% accuracy using SVM, 73% accuracy using Multinomial Naive Bayes, 70.50% accuracy using Random Forest, 65% using Decision Trees.

Keywords— Sentimental Analysis, Natural Language Processing, Restaurant Reviews, Text Mining, Machine Learning.

CHAPTER 1: INTRODUCTION

In this era, the restaurant business is getting more famous and owners feel that it is a profitable business. But the number of restaurants is getting higher due to excessive customer demand (Omar Sharif, 2019). Though there are some restaurants which cannot sustain because of proper strategy to attract their customers. The number of people is getting higher who take food or go to restaurants due to changes in lifestyle, income, and change of food habits (Spoorthi C. B.E., 2018). Due to the increasing use of the internet and the option to choose from so many restaurants, customers like to check the reviews of others (Horrigan, 2008). Thus, many groups and pages have become popular because of the easy way to check the reviews. Hence, reviewing on a restaurant from text, star ratings, remarks become a typical situation (Sungkono, K.R., 2019).

The main focus of this study is to analyze the customer's feedback or review and categorize it into the positive side or negative side. Positive reviews mean customers are satisfied with their food and service and also a symbol of trustworthiness (Veda Waikul, 2019). A higher positive review expresses higher satisfaction of the customer. (Park, at el., 2007). Positive review plays an important role in growing a restaurant. It creates a positive feeling and enforces the other customers to choose that restaurant. On the other hand, a negative review creates an adverse feeling which compels the customers to avoid those restaurants (Ma, Y., 2018). Moreover, the factors associated with reviews regarding customer satisfaction influence business performance. (Jalilvand at el., 2007). As the number of customers is increasing, it is very difficult to understand the attitude of a reviewer. Thus, people admire the experience of the consumer and consent to others and the review on a restaurant is the only way to include others opinion on the restaurant (Wright, L., 2018). In that case, sentiment analysis comes to uncover the motive of customers about certain food.

Sentiment analysis is the process of interpretation and classification of emotions where it can be positive, negative, or neutral. It is used to analyze, review and categorize based on scores of sentiments (Gherbi, A., 2017). This paper presents a mechanized system for sentiment analysis for restaurant reviews to classify an opinion into two classes: positive or negative sentiment (Ma, Y., 2018). Four machine learning algorithms such as multinomial naive bayes decision tree, random forest, and support vector machine (svm) classifier are implemented for classifying the reviews (Xu, X., 2019). The classified reviews are helpful for the restaurant to investigate their faults in different areas and improve the quality of food and service in the restaurant (Jaiyen, S., 2015). This paper presents a mechanized system for sentiment analysis for restaurant reviews to classify an opinion into two classes: positive or negative sentiment.

1.1 Background

Day by day the quantity of restaurant surveys are expanding ceaselessly. Such surveys can be discovered on sites and the networks as well as on nearby movement data destinations and SNS locales like TripAdvisor, foodpanda, FoodBank, Twitter and

Facebook. Besides, the expansion of advanced cells has made it conceivable to compose a survey on the spot in any case of the area; consequently the measure of audits proceeds to increment and the impulse for their incorporated pursuit is expanding also (Sungkono, K.R., 2019). Many audit destinations give both a star score and a survey input with text. Be that as it may, there are a few situations when the star score and the survey don't coordinate (Veda Waikul1, 2019). For instance, it's anything but a situation when the survey is acceptable yet a star score is low or the audit isn't acceptable yet a star score is high (Martin, S., 2019). Subsequently, it ought to be useful to give the client a star score by characterizing the assumption of the survey's substance as a checked worth through investigation (Xu, X., 2019).

1.2 Motivation of the Research

A remarkable chance for restaurant business has been set up by online media content. Right now, café business grade to put a solid conspicuousness on client input to improve their business strategy (Omar Sharif1, 2019). Progressively, café business are searching for vital arrangements that will help them impact that data to be sprightlier to advertise requests (Spoorthi C. B.E., 2018). Hence, clients are frequently mentioned to round out a study introduced on the lower part of a receipt or inside an application or email (Sungkono, K.R., 2019). This strategy permits organizations to gather information on their clients' socioeconomics and fulfillment just as the capability of their business. Gathering client criticism through reviews is monetarily and measurably troublesome, substitute techniques for consumer loyalty information assortment as of now exist in web-based media (Veda Waikul1, 2019). Regular online media and cafés locales produces overpowering measures of audits. Via web-based media stages audit like information is energetically accessible and effectively available, looking through great many online remarks is costly in both time and cash (Wright, L., 2018). Huge café organizations might be talked about in huge number of online posts or audits each day, and offering work force time to comprehend, react, or gather the data is strategically bumbling. Besides, innovation is equipped for putting away and preparing text yet can't decipher the importance or assessment. Accordingly to extricate the readily accessible mindfulness from web-based media, organizations need a proficient and financially savvy technique (Xu, X., 2019).

1.3 Problem Statement

Among the creating quantities of restaurant survey sites in Bangladesh, there are numerous famous audit sites like TripAdviso, foodpanda, FoodBank and so on What's more, TripAdvisor is as of now including 1275 café postings of Bangladesh and more than 21785 surveys on the site, which is greater than some other destinations in bd. To have a sensational effect on clients to begin scrutinizing the unwavering quality of audits on notable eatery survey sites. Be that as it may, for what reason do clients actually depend on these eatery audit sites? For what reason are these café survey sites more fruitful than others eatery destinations? Notwithstanding the acknowledgment of online café audits, there is restricted writing reporting the causal cycle in which eatery survey site ascribes motivation clients' reactions. At the point when the positive exactness and negative precision are determined as a normal worth there is an issue with the lessening in the normal grouping exactness. By expanding the exactness of the tolerably lower region between the positive and the negative estimation, there is a need to improve the normal arrangement capability.

1.4 Research Questions

Q1. How to improve restaurant business by classifying the reviews?

Q2. How to classify restaurant reviews into positive and negative sentiment by customers' feedbacks?

1.5 Research Objectives

The primary objective of this paper is to construct a managed learning model to arrange the client's inputs as far as sure and negative suppositions. The arranged surveys are useful for the eatery to explore their shortcomings in various regions and improve the nature of food and administration in the café. Another objective is to track down the best strategy for investigating eatery client survey information by looking at some AI calculation for slant examination. Through the stage given by the eatery audit site clients' can undoubtedly give their input to the café proprietors this would help eatery proprietors in settling issues which occurred during the assistance, refining the café administration execution, or in any event, constructing a superior relationship with clients.Thusly, this examination is intended to give both hypothetical and administrative ramifications for scholastic analysts in cordiality fields and partners in the eatery business.

1.6 Research Scope

To complete this exploration the proposed framework has been utilized four Machine Learning Algorithm like Multinomial Naïve Bayes, Support Vector Machine (SVM), Random Forest and Decision Tree. Examination on opinion investigation has additionally been finished with the Natural Language Processing Techniques. By utilizing these calculation the proposed framework will get the solid worth of precision.

1.7 Thesis Organization

This paper includes five chapter. Those are: Introduction, Literature Review, Research Methodology, Result and discussion, and Conclusion.

- The Chapter 1 Introduction section will be discussing the Background, Research objective, Problem Statement, Research Question, and research scope.
- The Chapter 2 Literature Review section will discuss the related work to this paper and will try to find the research gaps.
- The Chapter 3 Research Methodology section will show the proposed framework for the research and will discuss the research methodology.
- The Chapter 4 Result and Discussion section will show the result of the methodology with proper visualization and discuss it.
- Finally, The last Chapter 5 On Conclusion section will discuss the final output of the result and future recommendations.

CHAPTER 2: LITERATURE REVIEW

Customer review for restaurants is one of the promising research areas for researchers. Schrauwen uses Naive Bayes algorithm, Maximum Entropy and Decision Tree classifier for performance evaluation. (Omar Sharif, 2019) use those classification algorithms by measuring Accuracy, Precision, F1 score and Recall.

Nowadays customers go for online reviews to get feedback on the restaurants they want to visit. So, those reviews are very important for the customers who want to know immaterial attributes of goods in advance (Sungkono, K.R., 2019). Restaurant reviews indicate a restaurant's popularity and it is one of key points for restaurant buiseness (Veda Waikul, 2019). Online reviews allow people to compare restaurants based on some intangible elements.

Research using the Probabilistic Latent Semantic Analysis method is done by (Ma, Y., 2018) for sentiment analysis. They collect data based on the tiltle review, not the whole comment. The results of this research showed 73% accuracy on the data which is also proven by (Wrigh, L. 2018).

(Jin, L., 2017) used restaurant review comments as a dataset and the dataset is in English whereas some researchers work with native language. (Martin, S., 2019) use native language and use classification algorithms like Support Vector Machine, Random Forest, Decision Tree Algorithm, Naïve Bayes method and confusion matrix. Among the classification models, naïve bayes perform better than other algorithms.

2.1 Text Mining

In this way, literary information must first be organized previously it tends to be broke down (Sungkono, K.R., 2019). Text mining is a critical technique for removing information and key thoughts from varieties of scholarly data. Text mining traverses the opening from gigantic unstructured printed data to coordinated data that licenses for cognizance of associations and subjects inside the data. After data has procured a development, we may execute judicious showing, arrangement estimations, clustering examination, and various procedures to isolate information (Spoorthi C. B.E., 2018).

Text mining has in like manner been used to assemble proficiency in business settings by enhancing email strings. College of British Columbia teacher Giuseppe Carenini and accomplices utilized substance mining and snippet of data words to summarize email conversations. They initially fabricated a piece reference outline (an organized diagram) to address conversation strings. By then Carenini and accomplices created Algorithm Clue Word Summarizer which first content mines the email string by then apportions a quantitative worth to each term and sentence. The sentence as denoted the most important score is returned as the email conversation once-over. This grants customers to save time by scrutinizing traces instead of examining the entire string searching for fundamental information (Veda Waikul1, 2019).

Additionally, associations may do message mining methodologies to utilize the brain predominant piece of web substance for their expected advantage. Previously, business information has been amassed from coordinated data figuratively speaking. Parks

underlines, through separating the reports on market designs, reports, and site pages in Internet, monetary experts can get huge business information for instance, new competitors or genuine things turning out in market or customer demand plans advancing". Joining both coordinated data and unstructured content based data licenses associations to see a more unmistakable for the most part image of their customers needs and satisfaction and readies the association to advance with current market demands. For acquiring business knowledge from online media text mining is the initial phase in accomplishing the ideal objective (Wright, L., 2018).

CHAPTER 3: RESEARCH METHODOLOGY

In this paper, the reviews or comments were collected from the web page of TripAdvisor, foodpanda, FoodBank and many other restaurant sites. Then, the preprocessing process was carried out by the Python programming language with the NLTK (Natural Language Tool Kit) library (Veda Waikul, 2019).

The preprocessing stage contains of the symbol removal, case folding step, data cleaning, stop word removal, (Ma, Y., 2018). There exist several steps followed by this research method. Figure 1. depicts the entire methodology of this paper.The following paper is showing the step of this research paper.
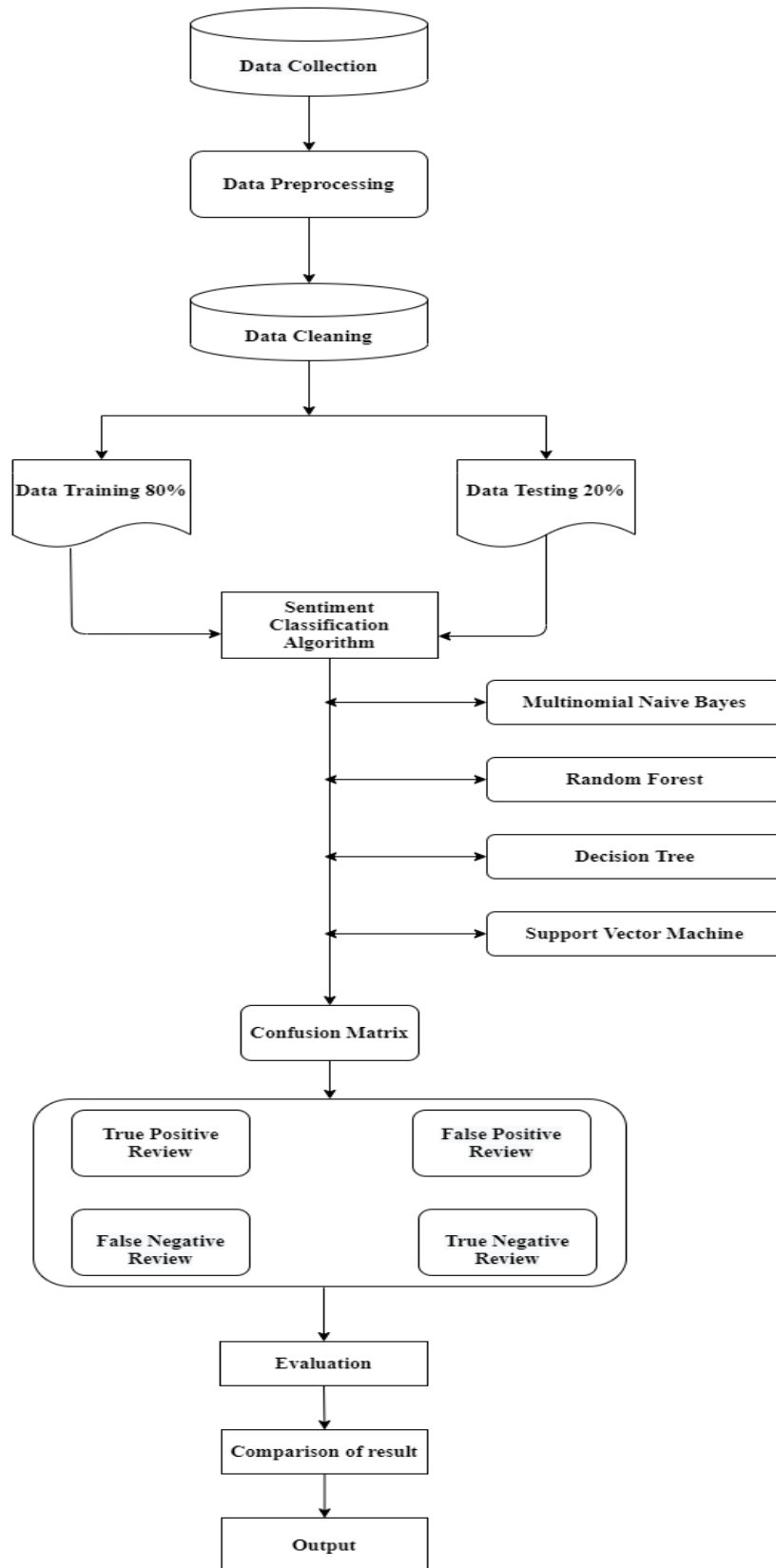
Fig 3.1.1: Proposed Model Framework

3.1 Data Collection

In this study, the dataset was prepared by the comments given towards various restaurants. The data used in this study was prepared from TripAdvisor, foodpanda, FoodBank and many other restaurant sites throughout Bangladesh. The dataset consists of 1000 reviews from TripAdvisor, foodpanda, FoodBank and different restaurant sites which were collected manually. After that dataset labeling was done by using WEKA tool. The dataset contains two columns. To ensure the correct labeling, one attribute was added to the dataset called "liked". This attribute will be the class attribute, where labelling will take place and it will be used to estimate the performance of this system. The reviews are classified in two categories, positive and negative. Thus values of "liked" instances are either "1" or "0". "1" is when the comment is positive and "0" otherwise. There are two columns in the dataset. The first column is consisting of the text reviews which are commented by different users, that are related to the food of the restaurant and the overall review of the restaurant as well. The second column contains the estimation of the sentiment i.e. whether the review is negative or positive.

Dataset has been used splited Test, 80% Data Training and 20% Data Testing. Sentimental Analysis can be considered a classification process (Veda Waikul, 2019). The proposed model should predict if the review is positive or negative (Ma, Y., 2018).

	Review	Liked
0	Food quality and taste is very good.	1
1	Food Prices are very high	0
2	Not tasty and the texture was just nasty.	0
3	Stopped by during the late May bank holiday of...	1
4	The selection on the menu was great and so wer...	1
5	"Friendly staff, good food and homely environm...	1
6	Honeslty it didn't taste THAT fresh.)	0
7	Food prices are very high and several items ar...	0
8	The fries were great too.	1
9	We loved the variety of the well marinated Chi...	1

Fig 3.1.2: Snapshot of the dataset

3.2 Data Preprocessing

The preprocessing steps are an essential advance to achieve spotless and explicit information so that in further, the consequences of order assurance could be more precise. In the preprocessing interaction, there are a few pieces of the cycle inside it (Wright, L., 2018).

- **Case Folding** is the progression where the way toward evolving all words containing capitalized letters into lowercase forms. Like the example below: "Amazing", "Tasty" becomes "amazing", "tasty".

- **Symbol Removal** is the phase where the process of punctuation (period (!), comma (,), question mark (?), exclamation point (!) etc.) is carried out, explicit characters (&, %, $, #, @ and others), and the numbers (0,1,2… to 9).

- **Stopwords** removal is a process of removing words that often seem but do not have meaning in languages. Stop words removal helps to reduce the dataset size then training set takes less time to execute. Moreover, to detect positive and negative reviews, stop words don't play any significant role (Jin, L., 2017). For data clustering, it can be seen that if there are stop words, the focus cannot be given to the main words. Stop words removal also helps to decrease the memory requirements while classifying the reviews. There are some stop words are given below:

  > "a," "the," "of," "to" "you," "it," "and", "a," "they," "or,"

- **Data cleaning** once credits are removed, filling the missing qualities, eliminating conflicting data measuring the central inclination for the characteristic like mean middle, quartile is finished. In information preprocess the information is cleaned and the separated information before examination. Non-text based substance and substance that are superfluous for the examination are recognized furthermore, killed.

'food qualiti tast good', 'food price high', 'tasti textur nasti', 'stop late may bank holiday rick steve recommend love', 'select menu great price'

Fig 3.2.1: Snapshot of the cleaned text

The Preprocessing process is using the Colab a web IDE Python with using the Python Programming Language.

3.3 Sentiment Analysis

Sentiment analysis varies from text mining since it seeks after to selection and arrange assessment instead of intriguing data. Feeling investigation, is that field of study individuals' suppositions, opinions, appraisals, decisions, mentalities, and slants towards elements like items, administrations, the travel industry, motion pictures, associations, policy centered issues, people, issues, occasions, points, and so on (Wright, L., 2018.). Acquiring public and client feelings has for quite some time been

an immense business itself for advertising, advertising, and political mission organizations. The political choice outcomes can moreover be expected from political posts. The districts like online media and little adding to a blog objections are taken an astounding wellspring of information considering the way that various people confer and inspect their experiences about sure and contrary evaluation uninhibitedly (Spoorthi C. B.E., 2018). Assessment Analysis is significant for research locales, for instance, ordinary language planning, data mining and text mining that consistently used to take apart words reliant upon the plans of people recorded as a printed version to find good, contrary, or unprejudiced sentiments. The goal to Sentiment Analysis is to acknowledge how people feel about something from their substance (Veda Waikul1, 2019).

3.4 Classification

The suggested framework's goal is to group the reviews, and then the key piece of the framework is the layout. The suggested model, which can categorize reviews into positive and negative sentiment, is trained using highlights from the surveys that have been removed. The suggested system is primarily built utilizing multinomial nave Bayes, decision trees, random forests, and support vector machines. The reference paper explains why these algorithms are used in the first place.

3.4.1   Multinomial Naïve Bayes

This is the most widely used classification method in text mining. It performs admirably and is frequently used in Natural Language Processing (NLP). The Bayes theorem underpins the algorithm. Using naïve Bayes rules for a text review and (M) and class (N). The Bayes theorem estimates the probability P(M|N), where M is the class of possible outcomes and N is the supplied instance to be identified, which represents some specific characteristics.
The formula is given below:

$$P(M|N) = P(M) * P(N|M)/P(N) \qquad (1)$$

Where,
$\qquad$ P(N) = prior probability of N
$\qquad$ P(M) = prior probability of class M
$\qquad$ P(N|M) = occurrence of predictor N given class M probability

With a Testing Accuracy of 92.375 percent and a Testing Accuracy of 73 percent, the suggested system can classify restaurant reviews.

3.4.2   Random Forest

Random Forest is a classifier that stores a number of decision trees on different subsets of a dataset and uses them to improve the dataset's predictive accuracy. A Random Forest is a collection of diverse decision trees with a single initial contrast. Maybe than choosing the best divisor from the whole rundown of components, the calculation chooses an irregular subset of the variables.

With a Testing Accuracy of 99.625 percent and a Testing Accuracy of 70.50 percent, the suggested system can classify restaurant reviews.

3.4.3 Decision Tree

The supervised learning technique includes decision trees, which are a form of classification algorithm. External and internal nodes are utilized to make decisions in a decision tree. A decision tree is a foresight model that constructs a series of valid/false articulations with the goal of characterizing an item. Mathematically Entropy for multiple attributes is symbolized as:

$$E(M,X) = \sum_{c \in X} P(c) \, E(c) \qquad\qquad (2)$$

With a Testing Accuracy of 95.625 percent and a Testing Accuracy of 65 percent, the suggested system can classify restaurant reviews.

3.4.4 Support Vector Machine

Support Vector Machine is a supervised machine learning algorithm which can be used for classification or regression problems. There are specific types of SVMs which can use for individual machine learning problems, like support vector regression (SVR) which is an extension of support vector classification (SVC).

With a Testing Accuracy of 96.875 percent and a Testing Accuracy of 71.50 percent, the suggested system can classify restaurant reviews.

3.5 Confusion Matrix

A confusion matrix is a strategy that has the ability to determine the display of a grouping model as shown by testing data estimate, where the data that is the outcome of expectations is divided into two groups, positive and negative (Omar Sharif, 2019). The system's confusion matrix must have two columns and two rows. The number of true positives, false positives, true negatives, and false negatives is specified in this matrix (Sungkono, K.R., 2019).

– **True Positive (TP):** Reviews with a favorable tone are also categorized as positive.

– **True Negative (TN):** Reviews with a negative tone are likewise categorized negative.

– **False Negative (FN):** Reviews with a favorable tone but are categorized as negative.

– **False Positive (FP):** Reviews with a negative tone but are categorized as positive

	Positive Review (1)	Negative Review (0)
Positive Review (1)	TP	FP
Negative Review (0)	FN	TN

Table 3.5: Confusion Matrix

3.6 Performance Measure Evaluation

From the following equation, we can calculate precision, recall, f1-support of performance measure evaluation for those algorithm.

3.6.1 Precision

Precision is a term that relates to a positive predictive value. The following equation can be used to calculate precision.

$$Precision = TP/ (FP + TP) \qquad (3)$$

If precision is high then the performance of the algorithm is doing well.

3.6.2 Recall

Recall is the percentage of positive reviews that are correctly categorized compared to the total number of positive reviews. The following equation can be used to calculate recall.

$$Recall = TP / (TP + FN) \qquad (4)$$

3.6.3 f_1-score

To choose a particular learning algorithm among numerous algorithms, the **f1-score** of each method must be calculated. The following equation can be used to calculate the f1-score.

$$f_1\text{-score} = (2 * precision * recall) / (precision + recall) \qquad (5)$$

CHAPTER 4: RESULTS & DISCUSSION

This paper has proposed four models and that model has been tested with Decision Tree, Random Forest, Multinomial Naïve Bayes and Support Vector Machine.

4.1 Performance Evaluation

There are two type of performance evaluation.
  i.   Positive Performance Evaluation
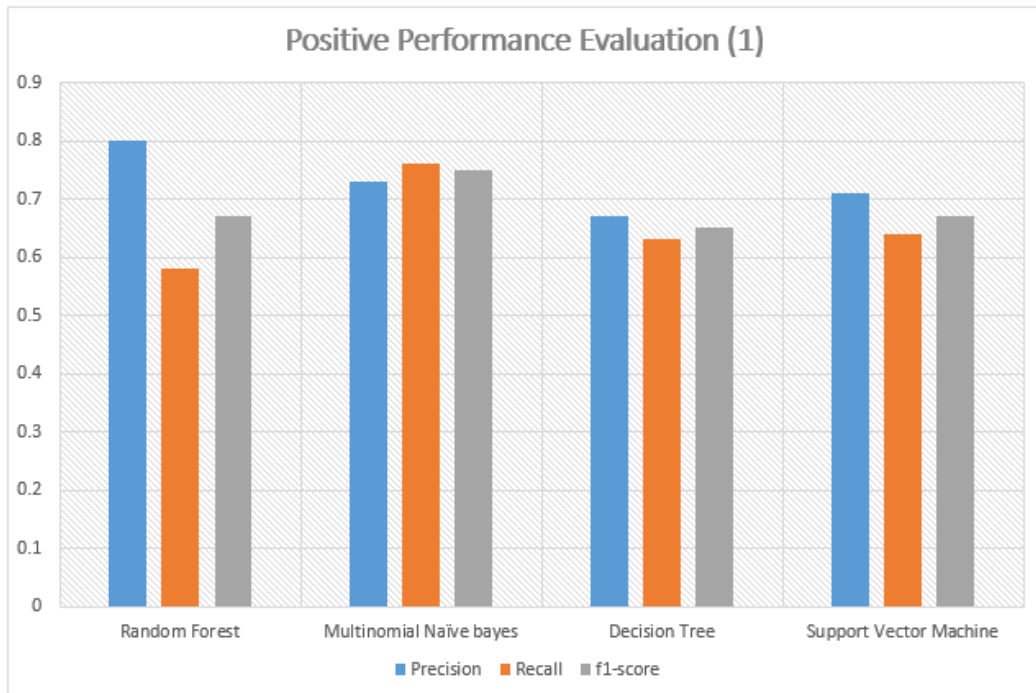 ii. Negative Performance Evaluation

Table 4.1.1, Table 4.1.2 and Figure 4.1.3 Figure 4.1.4 represents the evaluation of positive and negative performance of this algorithms.

Classifier	Precession	Recall	f$_1$ score	Accuracy
Random Forest	0.80	0.58	0.67	70.50
Multinomial Naive Bayes	0.73	0.76	0.75	**73.00**
Support Vector Machine	0.71	0.64	0.67	71.50
Decision Tree	0.67	0.63	0.65	65.00

Table 4.1.1: Positive Performance Evaluation

Classifier	Precession	Recall	f$_1$ score	Accuracy
Random Forest	0.65	0.84	0.73	70.50
Multinomial Naive Bayes	0.73	0.70	0.71	**73.00**
Support Vector Machine	0.65	0.71	0.68	71.50
Decision Tree	0.63	0.67	0.65	65.00

Table 4.1.2: Negative Performance Evaluation

Fig: 4.1.3: Positive Performance Evaluation
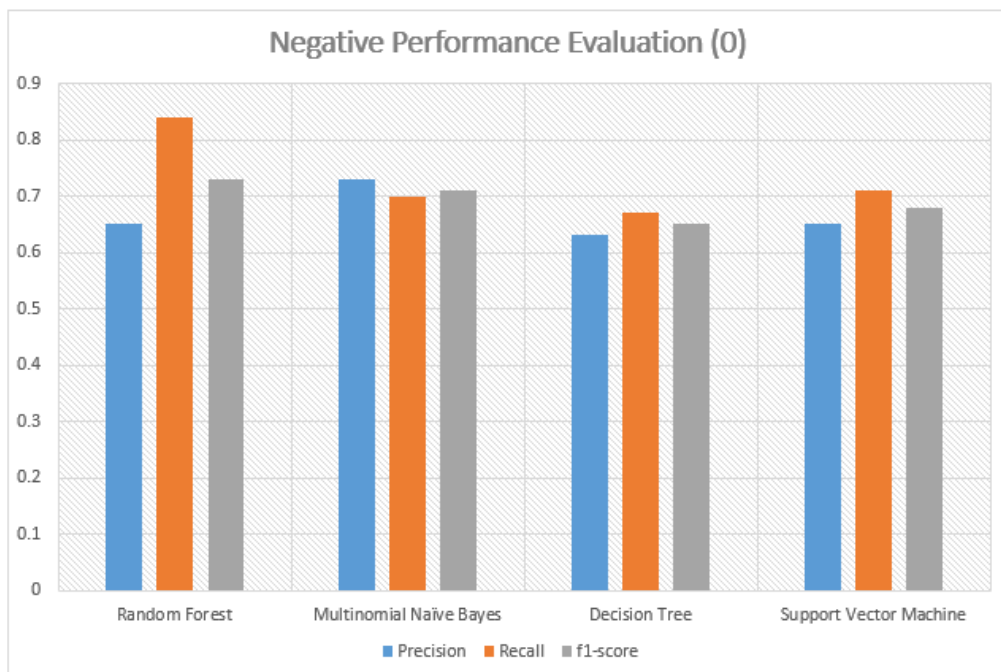
Fig: 4.1.4: Negative Performance Evaluation

4.2 Proposed Methods Accuracy

i.   Multinomial Naïve Bayes

Accuracy = 73%						
Actual (Expert)						
	Positive	**Negative**	**Precision**	**Recall**	**f₁-score**	**Support**
Positive	TP= 67	FP= 29	0.73	0.70	0.71	104
Negative	FN= 25	TN= 79	0.73	0.76	0.75	96

Table 4.2.1: Accuracy of the multinomial naïve bayes

ii.   Random Forest

Accuracy = 70.50%						
Actual (Expert)						
	Positive	**Negative**	**Precision**	**Recall**	**f₁-score**	**Support**
Positive	TP= 81	FP= 15	0.80	0.58	0.67	104
Negative	FN= 44	TN= 60	0.65	0.84	0.73	96

Table 4.2.2: Accuracy of the random forest

iii.   Decision Tree

Accuracy = 65%						
Actual (Expert)						
	Positive	**Negative**	**Precision**	**Recall**	**f₁-score**	**Support**
Positive	TP= 64	FP= 32	0.67	0.63	0.65	104
Negative	FN= 38	TN= 66	0.63	0.67	0.65	96

Table 4.2.3: Accuracy of the decision tree

iv.   Support Vector Machine

Accuracy = 71.50%						
Actual (Expert)						
	Positive	**Negative**	**Precision**	**Recall**	**f₁-score**	**Support**
Positive	TP= 85	FP= 11	0.71	0.64	0.67	104
Negative	FN= 46	TN= 58	0.65	0.71	0.68	96

Table 4.2.4: Accuracy of the support vector machine

© Daffodil International University
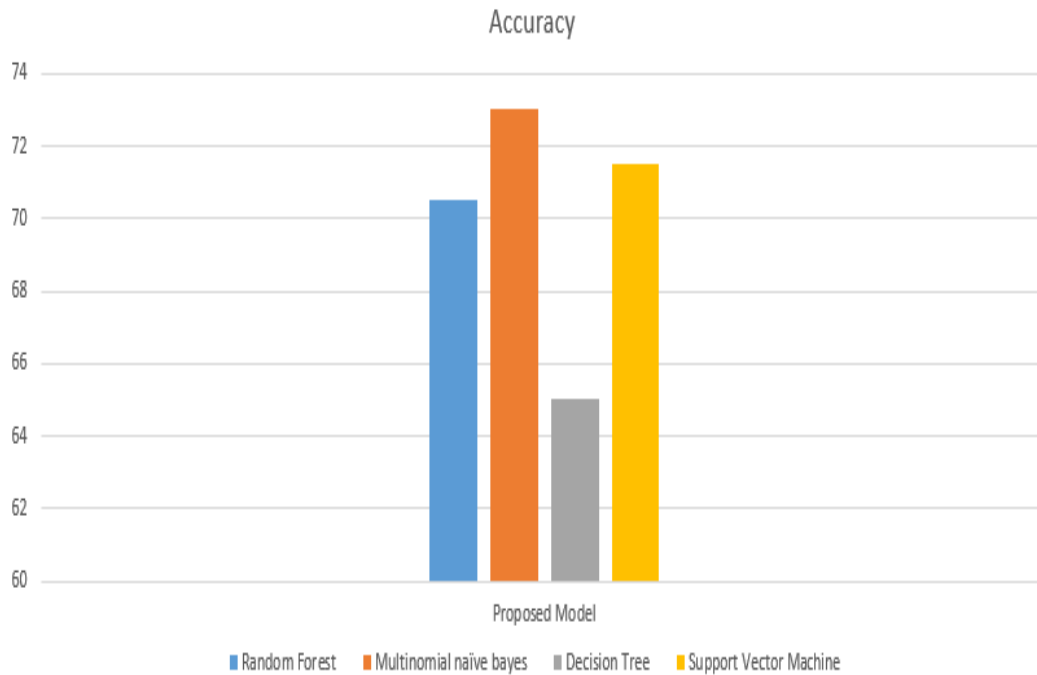
Fig 4.1.5: Accuracy of the proposed models

4.3 Discussion the Outputs

For positive value analysis, Multinomial Naive Bayes has precision as 0.73, recall 0.70 and f1 score 0.71, Random forest has precision as 0.80, recall 0.58 and f1 score 0.67, Decision tree has precision as 0.67, recall 0.63 and f1 score 0.65, and Support machine has precision as 0.71, recall 0.64 and f1 score 0.67.

For negative value analysis, Multinomial Naive Bayes has precision as 0.73, recall 0.76 and f1 score 0.75, Random forest has precision as 0.65, f1 score 0.73 and recall 0.84, Decision tree has recall as 0.67, precision as 0.63, f1 score as 0.65, and Support vector machine has precision as 0.64, recall 0.71 and f1 score 0.68.

From the above figure 3 and figure 4, it can be concluded that the positive and negative performance of multinomial naïve Bayes is higher than other models. And the accuracy is 73% which is greater than others models. Precision, recall and f1-score of both positive and negative each class can be represented by a classification result. This report can help for further analysis of that algorithm. Classification report shows that, the value of f1-score is 0.75 for positive response and 0.71 for the negative response. That means, this proposed system can properly classify 75% of positive and 71% of negative reviews. Naive Bayes is well suited for definite input variables than the numerical variables. And there the reviews comments are not dependent on each other. Thus multinomial naïve Bayes shows better results with the dataset.

CHAPTER 5: CONCLUSIONS & RECOMMENDATIONS

5.1 Findings and Contributions

In this paper, for restaurant reviews, a sentiment analysis structure is presented. The data were collected from social media and different restaurant sites on English language. In this study, various machine learning algorithms have been laboring to classify the original sentiments of the customer's comments. Other than this, a comparison of the accuracy on the applied models has also been shown. Here, the multinomial naïve Bayes acquired the highest 73% accuracy among all other used algorithms such as Support Vector Machine, Random forest and Decision Tree. This sentiment analysis has become very operative for the owners of the restaurants, as it can help the owner by providing a real view of what customers think about his/her restaurant. This study is a tiny effort to pay off for the insufficiency.
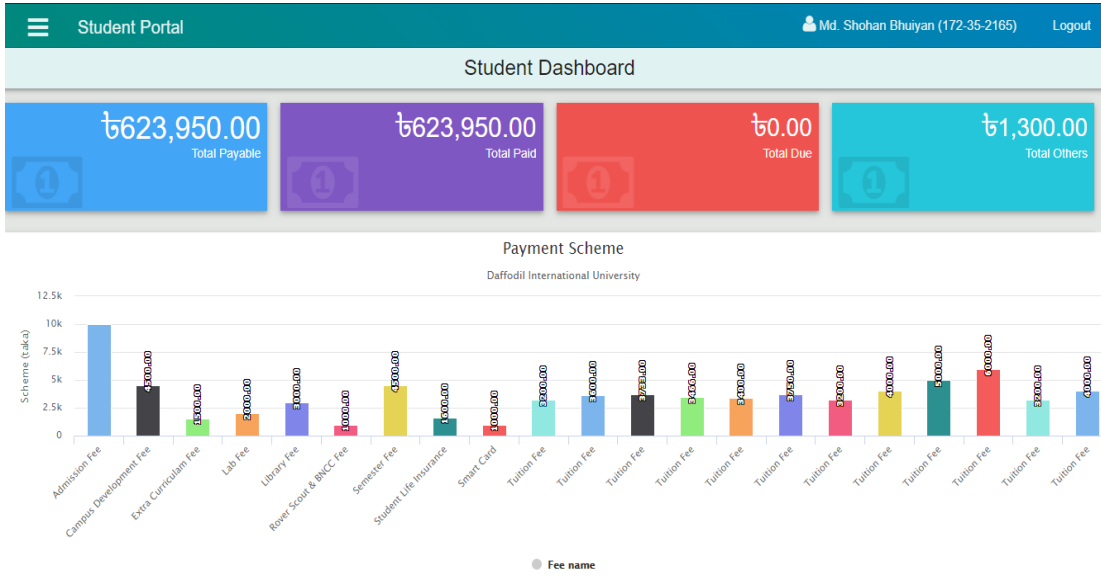
5.2 Recommendations for Future Works

In future, more powerful algorithms and more features will be implemented to discover the semantic connection of various words of a comment that would assist us to detect sentiment accurately. It must be done by using other methods for comparing materials. To accomplish that objective, becoming the dataset size should likewise be possible. Variety of review data can help to increase the value of accuracy.

CHAPTER 6: REFERENCES

1. Sharif, O., Hoque, M. M., & Hossain, E. (2019, May). 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)* (pp. 1-6). IEEE.

2. Ravikumar, D. and MJ, M., 2019. Sentiment Analysis of Customer Feedback on Restaurant Reviews.

3. Waikul, V., Ravgan, O. and Pavate, A., Restaurant Review Analysis and Classification Using SVM. *optimization*, *2*, p.7.

4. Wright, L., 2018. Classifying textual fast food restaurant reviews quantitatively using text mining and supervised machine learning algorithms

5. Adnan, M., Sarno, R. and Sungkono, K.R., 2019, September. Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm. In *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 121-126). IEEE.

6. Kang, H., Yoo, S.J. and Han, D., 2012. Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, *39*(5), pp.6000-6010.

7. Gan, Q., Ferns, B.H., Yu, Y. and Jin, L., 2017. A text mining and multidimensional sentiment analysis of online restaurant reviews. *Journal of Quality Assurance in Hospitality & Tourism*, *18*(4), pp.465-492.

8. TAMA, B.A., 2015. DATA MINING FOR PREDICTING CUSTOMER SATISFACTION IN FAST-FOOD RESTAURANT. *Journal of Theoretical & Applied Information Technology*, *75*(1).

9. Ganu, G., Kakodkar, Y. and Marian, A., 2013. Improving the quality of predictions using textual information in online user reviews. *Information Systems*, *38*(1), pp.1-15.

10. Titz, K., Lanza-Abbott, J.A. and Cruz, G.C.Y., 2004. The anatomy of restaurant reviews: An exploratory study. *International Journal of Hospitality & Tourism Administration*, *5*(1), pp.49-65.

11. Çalişkan, C., Yeşilyurt, H., Güler, O. and Kodal, S., CONSUMERS'POSITIVE AND NEGATIVE REVIEWS TOWARDS LOCAL FOOD EXPERIENCES: THE CASE OF RESTAURANTS IN GAZIANTEP CITY ON TRIPADVISOR. COM. *ITHMC*, p.90.

12. Elmurngi, E. and Gherbi, A., 2017, August. An empirical study on detecting

fake reviews using machine learning techniques. In *2017 seventh international conference on innovative computing technology (INTECH)* (pp. 107-114). IEEE.

13. Elmurngi, E. and Gherbi, A., 2017. Detecting fake reviews through sentiment analysis using machine learning techniques. *IARIA/data analytics*, pp.65-72.

14. Jonathan, B., Sihotang, J.I. and Martin, S., 2019, December. Sentiment Analysis of Customer Reviews in Zomato Bangalore Restaurants Using Random Forest Classifier. In *Abstract Proceedings International Scholars Conference* (Vol. 7, No. 1, pp. 1719-1728).

15. Varudharajulu, A.K. and Ma, Y., 2018, September. Feature-based Restaurant Customer Reviews Process Model using Data Mining. In *Proceedings of the 2018 International Conference on Computing and Big Data* (pp. 32-37).

16. Luo, Y. and Xu, X., 2019. Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: A case study of yelp. *Sustainability*, *11*(19), p.5254.

17. Claypo, N. and Jaiyen, S., 2015, January. Opinion mining for thai restaurant reviews using K-Means clustering and MRF feature selection. In *2015 7th International Conference on Knowledge and Smart Technology (KST)* (pp. 105-108). IEEE.

18. Auty, S., 1992. Consumer choice and segmentation in the restaurant industry. *Service Industries Journal*, *12*(3), pp.324-339.

© Daffodil International University

ACCOUNTS CLEARANCE

© Daffodil International University

PLAGIARISM REPORT

5% match (publications)
Omar Sharif, Mohammed Moshiul Hoque, Eftekhar Hossain. "Sentiment Analysis of Bengali Texts on Online Restaurant Reviews Using Multinomial Naïve Bayes", 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019

2% match ()
Wright, Lindsey. "Classifying textual fast food restaurant reviews quantitatively using text mining and supervised machine learning algorithms", Digital Commons @ East Tennessee State University, 2018

2% match (student papers from 04-Apr-2018) Class: Article 2018
Assignment: Journal Article Paper ID: 940850308

2% match (publications)
Masrur Adnan, Riyanarto Sarno, Kelly Rossa Sungkono. "Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm", 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), 2019

1% match (Internet from 17-Apr-2021)
http://dspace.daffodilvarsity.edu.bd:8080/handle/123456789/2803?show=full

> 1% match (Internet from 15-Mar-2020)
> http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/3553/P13659%20%2829%25%29.pdf
> isAllowed=y&sequence=1

> 1% match (Internet from 18-Jan-2020)
> http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/3523/P13624%20%2817%25%29.pdf
> isAllowed=y&sequence=1

1% match (Internet from 10-Nov-2020) http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/3552/P13658%20%287%25%29.pdf?
isAllowed=y&sequence=1

1% match (Internet from 07-Apr-2021)
https://www.coursehero.com/file/63075362/Reportdocx/

1% match (publications)
Bern Jonathan, Jay Idoan Sihotang, Stanley Martin. "Sentiment analysis of customer reviews in zomato bangalore restaurants using random forest classifier", Abstract Proceedings International Scholars Conference, 2019

1% match (Internet from 26-Mar-2019) https://ar.scribd.com/document/347684701/bok-3A978-1-4842-
2388-8

1% match (student papers from 06-Sep-2020)
Submitted to National University of Ireland, Galway on 2020-09-06

< 1% match (Internet from 07-Apr-2021) http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5149/162-
15- 7878%20%2818_%29.pdf?isAllowed=y&sequence=1

> < 1% match (Internet from 24-Feb-2020)
> http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/3555/P13663%20%2821%25%29.pdf
> isAllowed=y&sequence=1