Sentiment Analysis of Bengali comments on social media using Natural Language Processing and Machine Learning.

BY

Md Mazharul Islam Ullash
ID: 191-15-13022
AND

AMIR HAMZA RONOBI
ID: 191-15-13026
AND

Md. Tahmid Hasan
ID: 191-15-12920

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

Supervised By

Lamia Rukhsara
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
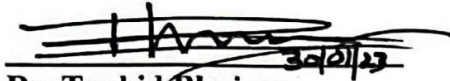Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Thesis titled "**Sentiment Analysis of Bengali comments on social media using Natural Language Processing, Machine Learning**", submitted by Md Mazharul Islam Ullash, ID: 191-15-13022, and AMIR HAMZA RONOBI, ID: 191-15-13026 and, Md. Tahmid Hasan ID: 191-15-12920 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 28 January, 2023.
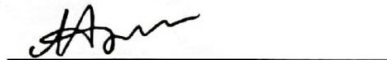
BOARD OF EXAMINERS

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

Nazmun Nessa Moon   Internal Examiner
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Zakia Sultana   Internal Examiner
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Shamim H Ripon   External Examiner
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

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

Supervised by:

Lamia Rukhsara 26.01.2023

Lamia Rukhsara
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

_____

Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Mazharul Islam 29.01.2023

Md Mazharul Islam Ullash
ID: 191-15-13022
Department of CSE
Daffodil International University

Amir Hamza 25.01.2023

AMIR HAMZA RONOBI
ID: 191-15-13026
Department of CSE
Daffodil International University

Tahmid hasan 29-01-2023

Md. Tahmid Hasan
ID: 191-15-12920
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We really grateful and wish our profound our indebtedness to **Supervisor Lamia Rukhsara, Senior Lecturer**, and **Co-supervisor Md. Sazzadur Ahamed, Assistant Professor** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Research-Based Project*" to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

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

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

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

ABSTRACT

Social media is a great way to share your thoughts and opinions with others, but some people on social media use language that is dangerous and disrespectful. Researchers are interested in understanding the emotions and sentiment behind this data, and Bangla Natural Language Processing has become a popular research field because of its diverse applications. Every day, a lot of user-generated Bangla data is created on social media, online news portals, and other websites. Sentiment analysis is a research area that is used to identify sentiments in texts. This paper also provides a comparative evaluation by examining implementations of different types of machine learning algorithms. Instead of previous efforts, we use NLP to process the data. By collecting more comments from social media and processing them through natural language processing, we have tried to extract the sentiments of those comments through machine learning algorithms. Here we use Random Forest, Logistic Regression, Naive Bayes, KNN, SVM, Linear SVC, and AdaBoost algorithms. Based on the results of all proposed classifier implementations, Support Vector Machine (SVM) techniques acquire the highest precision of 80% when compared to the other classifications.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction:

Common people can express their thoughts, emotions, and responses on a variety of topics through social media. For young people to older adults who spend a lot of time networking with their peers, using this public platform becomes a daily habit. However, this internet environment frequently gives rise to debatable subjects like political propaganda, irrational religious beliefs, and random scams. Divided parties in such circumstances exchange hate speech, including threats and profane language, to personally attack one another. Such derogatory language and toxic comments must be taken out in order to maintain a safe user experience on the platform [1]. Bengali users of social media are increasingly uploading text, which makes it possible for others to respond to their remarks fast. This frequently results in negative remarks, which must be filtered out. It is difficult to manually filter out negative comments with harsh word lists and complex rules, but automated technologies have made this process simpler. Additionally, the unstructured nature of social media text makes it difficult to filter out comments with a good acceptability score [2]. Previously, it was challenging to create handwritten rules using manual linguistic features, but today automated systems can be utilized to do this thanks to more affordable computing resources. Machine learning methods have been used by researchers to try and find offensive or harmful phrases. But these models require special tools to help them understand [3] how often different phrases are used on social media, and they can't work well with comments that are not well-formed or predictable.

1.2 Motivation:

Sentiment analysis is a highly automated method for understanding a person's point of view based on spoken words, written content, or even video comments from social media. Bangladeshis have been found to be more likely to engage in online activities such as posting their thoughts and opinions on popular social networking and microblogging platforms such as Facebook, Twitter, Instagram, and YouTube. Bangla is spoken as a first

language by nearly 200 million people worldwide, with 160 million of them being Bangladeshis. Despite the obvious fact that only a human being is capable of comprehending another human being's feelings through written text, doing so manually is difficult given that social media platforms can host millions of users and millions of comments. Given the availability of data and the potential benefits of enhancing social media applications, machine learning can thus play a significant role in this. Use social media to gather information while avoiding offensive statements such as (political propaganda, religious craziness, and random hoax). Toxic language and comments like these must be filtered out to ensure that users can use the platform safely. An increasing number of Bengali social media users post a variety of status updates, comments, graphics, and other content that is instantly available for others to respond to. This results in text with toxic comments that must be filtered out using a machine-learning algorithm.

1.3 Rationale of the Study:

There is a lot of work with data mining and machine learning going on in the social field. Everyone is working with a Twitter comment in their work. But we have done our work with many types of social site data. As a result, our work will be a fresh take on social media's comments sentiment. Using our best efforts, we have improved the existing data mining and machine learning development process. Machine learning might be a fascinating area in and of itself. Combining it with another fascinating discipline of data mining could result in a truly amazing work of energy.

1.4 Research Questions:

Our work will help data mining approaches to solve difficult problems. A professional and useful classifier will be the main goal to get a higher result. To guide a research project, we need to first develop a collection of expert questions, which will all be answered after our job. We've compiled a list of questions that we'll address through our work. They are

1. Is it possible to use different data mining approaches for testing and several performances to measure?

2. Can data mining algorithms be combined with machine learning?

3. How can we choose the best classifier to get the best result?

4. Will the evolutionary methods produce better results than the traditional methods?

1.5 Expected Output:

AI is present in every aspect of our digital lives. This digital world has a vision

Extensive research on machine learning and data mining is another area of AI.

Today, in our medical field, so much is surrounded by machine learning and data mining.

So after our work, we can create a comfort zone for all ages.

people in their medical services. Now we can explain the result as follows.

• Various data mining approaches are used to improve speed and accuracy.

• Dramatically improved execution with optimized data mining results.

• Multiple tests of data mining approaches to select the best classifier output.

• Improved classifier accuracy.

1.6 Project Management and Finance:

Sentiment analysis is a way of understanding the public's opinion on a topic. This is accomplished by looking at the way people are expressing their feelings online. This is useful for social media monitoring tools, which can then give you an idea of how people are feeling about a topic.

Social media is a great way to communicate with people, but it can be difficult to know what people are really feeling. Sentiment analysis is a way to understand the emotions behind social media posts, and this information can be very useful in making decisions. Social media sentiment analysis is a way to measure how people feel about your posts.

1.7 Report Layout:

We used data mining to explore our database and figure out which patterns and trends exist. Then, we used that information to create a query that will allow us to start investigating the data in more depth. Finally, we used the results of that query to organize the data at the beginning of this chapter. Data mining is a process used to analyze data to find patterns and trends. It is becoming more important as Machine Learning becomes more sophisticated. This paper is about a platform that is used to study the effects of a particular medical condition. In the past, people have used this platform to learn more about the

condition. This paper will explain what the platform is and how it works. There is a lot of detail in this paper, so it might be helpful to read the first chapter before reading this one. After reading this paper, you will know more about how this platform works and what kinds of research it can be used for.

CHAPTER 2

BACKGROUND

2.1 Preliminaries:

There has been a recent increase in the level of conflict in this field. This is because two different types of AI work fields are becoming more popular- machine learning and data mining. In the background area, we will look at past work done in this field. In the following part, some background information is offered to help understand the research effort. Finally, some challenges are discussed.

2.2 Related Works:

We looked at research to see if there is a gap in previous work on this type of problem. We found that there are a lot of complications that change over time, so we will look at different publications to see if their findings match up.

The authors of [4] develop a system that employs Natural Language Processing to assess customer reviews from online purchasing and offer a ratio of favorable and unfavorable reviews written in Bangla.

On their preprocessed dataset, the authors of [4] utilized five conventional ML algorithms to assess the accuracy of their findings. Five different training and testing sets—30%, 40%, 50%, 60%, and 70%—are utilized to gauge efficiency. Support Vector Machines (SVM), which are used in pattern recognition and machine learning, have improved performance. SVM generated an accuracy of 70.81% out of all the algorithms using the 30% test data. Using 70% training data and 30% test data, logistic regression provided the second-highest accuracy at 70.09% efficiency. By using 70% training data. The ratio of genuine positive and negative reviews for a certain product, where positive reviews make up 62.5% and negative reviews make up 37.5%. The ratio of favorable to negative reviews in the expected output from their model is fairly close to the actual ratio. The author of the study [5] wants to mechanically extract the sentiments or opinions expressed by users from Bangla microblog postings and then identify the overall polarity of texts as either negative or

positive. With the use of emoticons and the training corpus, SVM, MaxEnt, and Naive Bayes can be used to get exceptional accuracy. The authors [5] create a sentiment lexicon by hand that includes terms that express both positive and negative emotions. In their implementation of the support vector machine (SVM), they applied linear SVC. The author of the paper [6] presents a thorough collection of methods for extracting sentiment and emotions from Bangla text comments on youtube and for classifying sentences with a three-class (positive, negative, neutral) and a five-class (strongly positive, positive, neutral, negative, strongly negative) sentiment label and the accuracy for three and five labels using the suggested approach is 65.97% and 54.24%, respectively. At [7] the author has introduced "BanglaSenti," a lexicon-based corpus or dataset created exclusively to detect sentiment analysis from social media comments Bangla textual data.[8] provided feature analysis and sentiment analysis techniques using a dataset they built themselves that contains 1500 brief comments. independently created a dataset of 1500 Facebook comments. They used 920 training data, the remaining part of testing data, and 52.98% accuracy was the greatest they could achieve.

The author provides two publicly available datasets to perform the ABSA task in Bangla [20]. One dataset contains human-annotated user comments on cricket, while the other contains customer reviews of restaurants. SVM provides the highest accuracy of 77% when compared to cricket data. In [21], a sentimental analysis of Bengali song reviews from a particular YouTube channel was presented in order to assess the public opinion of a rising young star. In order to identify actual public sentiment, the research employed a backtracking algorithm that had a 71.23% accuracy rate. [22] This article provides a thorough analysis of several current techniques for text classification-based sentiment analysis on a Bangla dataset. Through the effective application of 7 machine learning techniques, this research aims to close some of those existing research gaps. Among all algorithms, applying 5-fold SVM yields the best accuracy score of 76.48%, while not using 5-fold Gradient Boosting yields the best accuracy score of 76.95%. Using word2vec and sentiment extraction of words, a method of classifying the sentiment of Bengali comments is presented. The accuracy reached is 75.5 5% respectively [23].

2.3 Research Summary:

Machine Learning:

Machine learning is a way for computers to learn from data and get better at doing things over time. This involves using statistical models and algorithms to look at patterns in data, and then make decisions based on that information.

Machine learning algorithms can learn from large data sets to make predictions or decisions about new data. This helps them do things that would be difficult or impossible for humans to do, such as identifying objects in images or transcribing speech to text.

A machine learning algorithm is a set of instructions that can learn to predict things like labels (i.e. "This is a dog" or "This is a cat") from sets of input data (features, like length, color, etc.). The goal of the algorithm is to figure out how the input data relates to the label so that it can make predictions on new data.

Machine learning is a way of making predictions about things using data. It's being used in many different areas of life, like healthcare, finance, and retail. It's been helpful in solving a lot of problems, and it's getting better and better every day!

Machine learning is a way of using computers to learn from data, and it's been growing in popularity in recent years because of the availability of lots of data, the development of powerful computing resources, and improvements in algorithms and techniques. This has made it possible to create more accurate and efficient models, which is why machine learning is so widely used these days.

The fundamental paradigm of machine learning is supervised learning. Machine learning algorithms in supervised machine learning must be taught using labeled data; otherwise, the approach will not perform effectively. When applied correctly, supervised learning may be tremendously powerful. Supervised machine learning algorithms will always strive to improve, discovering new forms and forming relationships as they train themselves on fresh data.

Machine learning is a hot issue in the research world. All of the latest machine learning approaches are constantly seeking ways to improve. The pace and intricacy of this sector,

along with new approaches, can make it tough even for professionals. So, tell us about alternative ways for demystifying machine learning and providing a learning route for folks who are unfamiliar with the original concept. The following strategies offer an outline of what you can develop using your machine-learning knowledge and skills: Pattern Recognition, Classification, Association, Outlier Detection, Clustering, Regression, and Prediction.

Data mining specialists use strategies and technology from the intersection of database management statistics and machine learning to better understand their carriers. Machine learning works by understanding how patterns emerge from data. Machine learning is a way to figure out what happens in the world by using computers. The first part is figuring out what variables and properties are involved, and the second part is using the best information to make the system easier to understand and to help it learn.

Recently, advances in machine learning and predictive analytics technologies have simplified matters. Machine learning produced tangible results in the prediction of human diseases. Machine learning often incorporates certain implementations by utilizing data and training their algorithms to make predictions, exposing solution vision inside data mining operations. Machine learning has already demonstrated several beneficial uses of data and algorithms in the data mining industry in the prediction section. We also employed a variety of machine learning methods to assess our work. After labeling the data, we run the classifiers through it to see which classifier produces the best results.

We can find out which classifier functioned best and had the best accuracy by measuring. Machine learning can quickly find trends and patterns in their field's most recent work. There is no need for human interference in their job. Machine learning is gaining popularity in the eyes of the globe as it continues to develop. It can manage data that is multidimensional and diverse. Machine learning generates a large number of diverse applications. Machine learning is gaining popularity, and it is shown promise in predictive analytics. The key reason for this is that it uses data flow to train its algorithms before integrating with data mining initiatives. As a result, it stands to outperform other approaches.

Data Mining:

Data mining is the use of discrepancies in a collection to forecast results and determine correlations. Data mining may be a source of information and is also known as database knowledge discovery. It also includes a novel method for detecting relevant patterns and forming associations with massive amounts of data. This discipline involves combining technologies from statistics, artificial intelligence, and machine learning with database control to analyze massive amounts of data. The majority of data mining applications are found in industry, scientific research, and government security.

Data mining strategies and technologies are capable of anticipating future movement and acquiring additional information about the company sector. Using a variety of data mining approaches, you may utilize data to increase revenues, lower prices, reduce risks, and improve customer interactions, among other things. It is a method used by businesses to convert hard facts into useful info. The data mining technique and its efforts are well-known in the sector for producing satisfactory outcomes. Data mining will seem more sophisticated in the next generation's purpose, in addition to other variables and their linkages developed for every model, and further refinement is feasible via research that will build new ways for choosing the most interesting aspects.

Data mining is the exploration and analysis of large data blocks to obtain significant forms and patterns. Data mining may be used for a variety of purposes, including spam email filtering, fraud detection, illness prediction, database marketing, and so on. Its works are becoming increasingly popular. It may be combined with other AI subfields by producing strong outcomes in their prediction job. So, given that we know about their subsequent works, we can understand how it works. The data mining process is separated into five stages, which are as follows: The system begins by collecting data and then packing it into storage. The obtained data was then handled and secured. Then, access the data and figure out how to organize it. These tools are used to find the best airfare and phone number of a long-lost classmate, as well as the greatest price on lawnmowers. Assume intelligent agents have lost control over medical research data and atomic particle physics. Computers can bring new insights into the nature of the cosmos as well as revolutionary illness treatments.

It is obvious from the explanation that data mining is becoming increasingly effective in its new task by employing new strategies.

2.4 Scope of the Problem:

The need for a model that can estimate the quality of comments from text reviews is highly enticing in today's society. Many of us will benefit from the deployment of a scheme like this. Almost everything around us is now review-dependent. It takes a long time to read a thorough review. People only care about the quality of reviews when they leave comments. As a result, the ideal answer to this problem may be a model that can anticipate the quality of the comments. Almost every user will profit from the model since they will no longer need to read the complete comment review to determine the quality of the remark. People may readily determine the quality of the remark using this methodology. This will save consumers time while also increasing their efficiency and productivity. My research will have a working scope that includes comments on quality prediction difficulties. The future of emotion or sentiment analysis is bright. Nowadays, emotion analysis in social media has a long influence on deciding user behavior. Consumers will increasingly seek tailored and directly connected evaluations as a consequence of a deeper and better comprehension of the feelings, emotions, and sentiments expressed in comments. In my research, I employed a dataset including user comments. My dataset is in Bengali, and practically all of the work has been completed. It is an excellent chance to delve further into the study, discover intriguing patterns in the dataset, and solve difficulties. My study has a rather broad working area.

2.5 Challenges:

There are so many variables involved, the problem is highly difficult. There are several subproblems that must be addressed before arriving at a stunning solution. Natural language is inherently difficult since we cannot compel people to obey any set of norms. The complexities of natural language must be treated with extreme caution. Much research has been conducted to address this issue, and new models are being produced on a daily basis. Because the same purpose may be obtained in several forms of words, properly parsing natural language is more crucial than anything else in this situation. In natural

language, one word can take several forms, and dealing with all of them is neither straightforward nor practicable. I required comments, as well as confirmation that the information in the reviews was correct. The most significant tasks for this study are feature selection, word tokenization, and word selection. Converting the comments into Negative, Positive, and Neutral categories was a vital assignment for me, as was selecting the best algorithm and software libraries to work with. My dataset had a large number of related terms that needed to be recognized. The difficulty of cleaning data for numerous datasets is comparable. Displaying the mathematical depictions of the traits and contrasting them with the attributes was another crucial task. The accuracy % was the key problem of this study. A strong accuracy level can portray the true feeling, thus calculating an exact accuracy was a must-do assignment. It took a long time to figure out the term and model choices. I needed a good dataset, and this one fell short. For this investigation, feature selection became even more critical. Of course, algorithm application was a significant problem. It was also difficult to determine which set of regulations would be most suited. As novice researchers, we frequently have to deal with novel situations. Thus, it was challenging to retain patience when working or studying.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation:

Many groups who have learned a lot about experimentation are beginning to develop machine learning and data mining models that are quite good at what they're meant to do. Then they reduce them in their research efforts as they improve. Many individuals have recently used machine learning and data mining to categorize objects. Data mining is a form of Evolutionary Algorithm that is used to create data-driven judgments.

3.2 Data Collection Procedure/Dataset Utilized:

The issue of this research is a medium that was considered for sentiments. So, in our research, we used machine learning and data mining. We knew about how our work impacts both our everyday operations and our ability to develop models and collect data. We have collected a lot of data from various social sites and various websites which have been used for our research work. The other important part of the research was using high-tech tools and methods. We operated the Windows Operating System, Python with many libraries like seaborn, Pandas, NumPy, etc. In the operating system where we used Google Collab, many applications for all the training and testing processes. This is a list of the most common commands used for machine learning in Python.

Workflow:

Primary 4 Stages:

- Collection Data
- Preprocessing Data
- Implementing algorithm
- Evaluate Performance

3.3 Statistical Analysis:

After prepping the data for this project, we split it into training and testing sets. In this instance, training purposes utilized 80% of the total data set. For testing, the remaining 20% of the total data set was employed.

In an effort to determine which classifier performs the best, we used the data that is most pertinent to our research as well as data mining approaches. We utilized the classifier to generate predictions about the remaining data after training it.

Implementation Procedure:

This work is about figuring out what people feel about things. First, we have collected a lot of data from different types of social media. Which will help us to predict in the survey. We have pre-processed that data in different ways. We used some pre-processing steps to input this data into the classifier. Stop words are words that are often eliminated prior to natural language processing. One of the most common words in any language is this one, provide little information to the text (along with articles, prepositions, pronouns, conjunctions, etc.).Examples of a few stop words {'এ', 'হয়', 'কি', 'কী', 'এর',........,'সম্পর্কে', 'সর্বাধিক', 'বিরুদ্ধে', 'অন্যান্য'}.We removed all these stop_words. Tokenization is the division process of a piece of text into smaller components. Then CountVectorizer Breaking Down A Sentence Or Any Text Into Words By Performing Preprocessing Tasks, Thus Removing Special Characters.

We labeled all the data. After preprocessing, our data was divided into two sets: training data and testing data. 80% of the total data set was used for training purposes in this example. The remaining 20% of the data set was used for testing purposes only. This is a random division of the data. The training data was then used to train the classifiers. We used test data to predict the specific comments that would be made after training our classifiers. Some of the performance evaluation measures have been collected here.

The working procedure of Sentiment analysis of Bengali comments in social media using machine learning is given below.
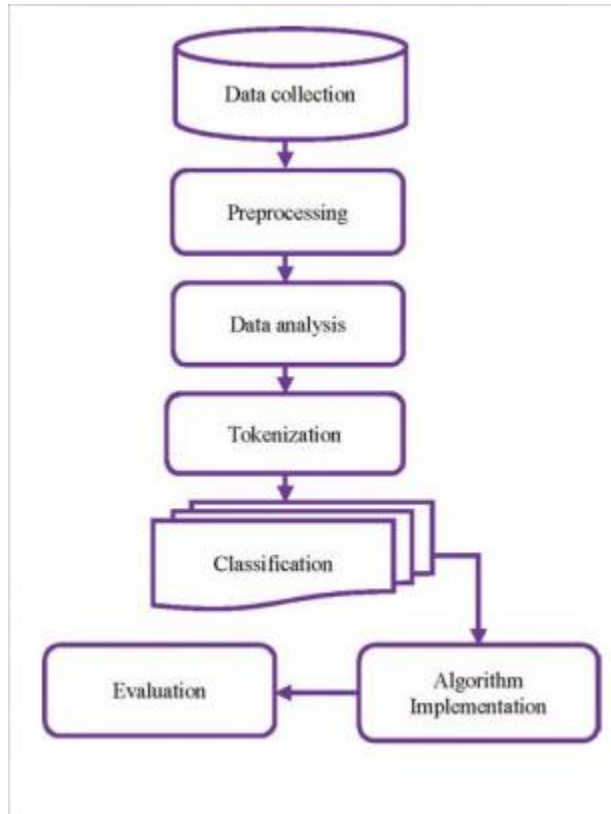
Figure 3.1: Implementation Procedure.

Using these characteristics, we derived the optimal classifier to predict in this scenario. Several performance indicators in % have been determined using the equations below based on the confusion matrix generated by the classifier.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\%$$

$$Sensitivity \text{ or } Recall \text{ or } True \text{ } Positive \text{ } Rate \text{ } (TPR) = \frac{TP}{TP + FN} \times 100\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$F1 \text{ } Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

3.4 Proposed Methodology/Applied Mechanism:

a) Random Forest:

The Random Forest classifier is a supervised learning algorithm that helps you understand by batching lots of different trees together. It's good for classifying and predicting things, like what a person's favorite color is. Random sampling is a way of solving classification problems. It takes a random sampling from a dataset to help you figure out which groups it belongs to. By using data samples, a decision tree can be created that will help you predict what might happen in the future [9]. Random Forest is a machine learning algorithm that helps you learn about patterns in data. It can be used for both classification (sorting things into different groups) and regression (figuring out how changes in one thing affect other things). Random Forest works by using a lot of different classifiers (like teachers) to come up with a better solution than any one of them could on its own. Random Forest is a type of computer algorithm that uses a number of decision trees to make predictions about a given dataset. The algorithm tries to improve its predictions by averaging the results of those trees. Random forest is a kind of machine learning algorithm that takes the predictions from a group of decision trees and uses the combined information from those trees to make a prediction about the final outcome. This kind of algorithm is more accurate than decision trees because it uses more data to make its predictions [10]. In splitting a node, from a particular distribution, the computer searches for the best attribute. This makes the network more varied, which leads to a better model. Because it averages the outcomes, Random Forest is superior to a single decision tree in that it prevents overfitting.
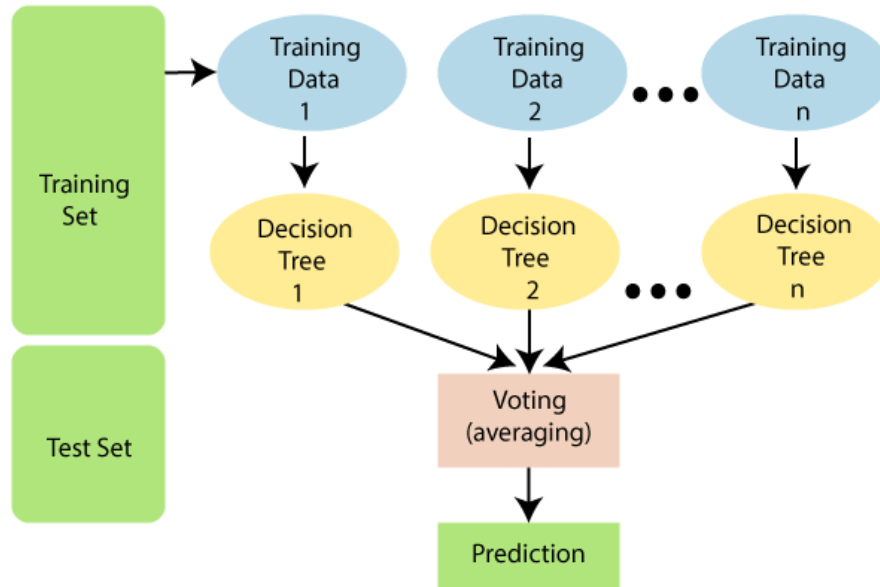
Figure 3.2: This diagram presents the working of the Random Forest algorithm.

b) Logistic Regression:

Based on a set of independent variables, a machine learning technique called logistic regression can be used to predict the category of a given dependent variable. Logistic regression predicts the likelihood of an event, such as whether a person is a smoker or not. The probability of an event occurring can be measured in terms of odds (0, 1), or as a percentage (between 0 and 100%). Logistic regression is a type of regression analysis used to solve problems related to regression, such as predicting the probability that an event will happen [11]. A technique used to predict the categorical outcome variable using merely a collection of individual factors is the supervised learning approach of logistic regression [12]. This is a strong technique since it can provide fresh knowledge about both continuous and discrete data. A mathematical formula called the logistic model exists that can predict what will happen based on a range of possible outcomes. It uses a number between 0 and 1 to represent how likely it is for an event to happen.
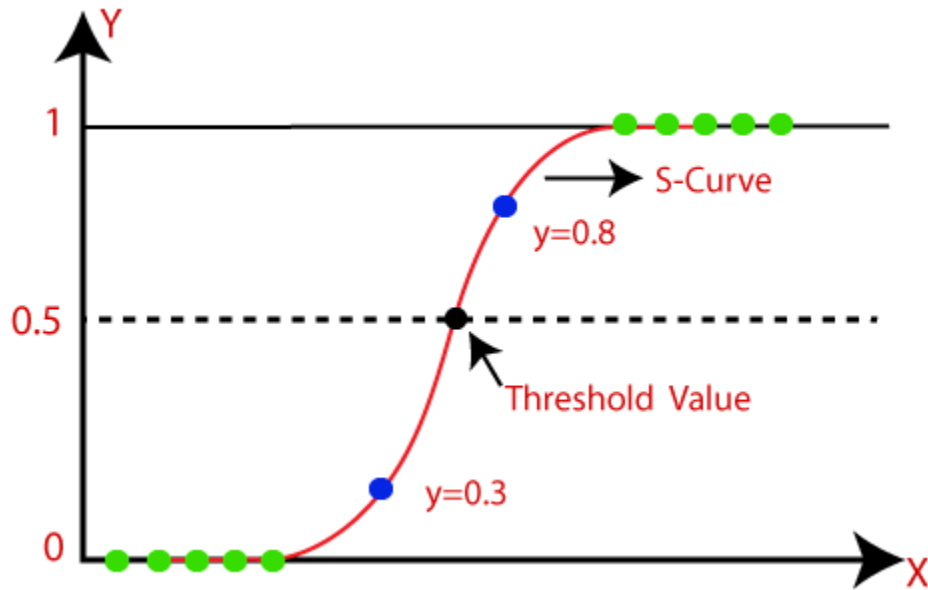
Figure 3.3: This image is showing the logistic function.

c) Naive Bayes:

Based on Bayes' theorem, the Naive Bayes algorithm is a supervised learning algorithm. It is employed to address classification issues with large training datasets. Naïve Bayes Classifier is one of the simplest and most effective classification algorithms, and it can help you build fast machine learning models that can make quick predictions. The Nave Bayes Algorithm is a type of classifier that makes predictions based on the likelihood of an event. It can be used, for instance, to categorize articles, assess text sentiment, and filter spam. [13]. Naive Bayes classifiers are a type of machine learning that can be used to predict outcomes based on data sets. They work by preprocessing the data set so that it is in a form that can be easily tabulated, after which determining the categorization that best fits the new data using a mathematical method. The categorization determined to be best-fit is the one having the highest fitness value. An example of a probabilistic classification system is the Naive Bayes Classifier, which is based on Thomas Bayes's posthumous Bayesian Theorem. Finding an effective approach to link fresh data to categories in a particular field is the primary objective of classification. [14] [15]. This assumption might lead to inaccurate predictions because it is based on a limited number of data points [16].This classifier is versatile

and can be used for a variety of classification purposes. To ensure the accuracy of any probabilistic computations, a sample data set is used to test it.
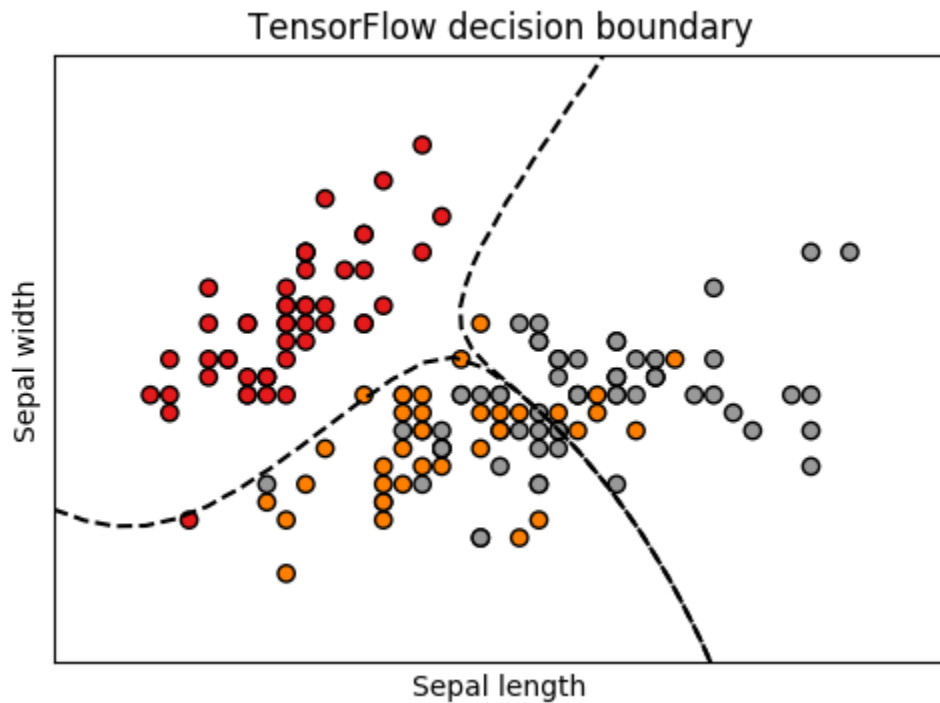
Figure 3.4: This diagram presents the working of the Naive Bayes algorithm.

d) KNN:

The KNN method compares the current specific instance to old instances to see if there is a similarity between them. If there is, it assigns the new case to the category that is closest to the classifications.KNN uses the number of neighbors to decide what to do. This is important because it determines how well the system will work. K-Nearest Neighbour is a machine learning algorithm that uses supervised learning to learn about similarities between new cases and other cases. This algorithm looks at how similar new cases are to other cases that are already in the database and then classify new cases accordingly. This makes it easier for the algorithm to quickly find a category for new cases and makes it easier for the algorithm to correctly classify new cases [17]. The K-NN algorithm is used for regression (measuring how one variable relates to another) and classification (determining whether something is a certain kind of object), but it's mostly used for classification problems. K-NN is a non-parametric algorithm, which means it doesn't rely on assumptions about

the data. The KNN algorithm is designed to be lazy, meaning it doesn't immediately learn from the training set. Instead, it stores the data set and uses it to make predictions at the training phase. Once it has new data, it uses that data to figure out how similar it is to the training data.
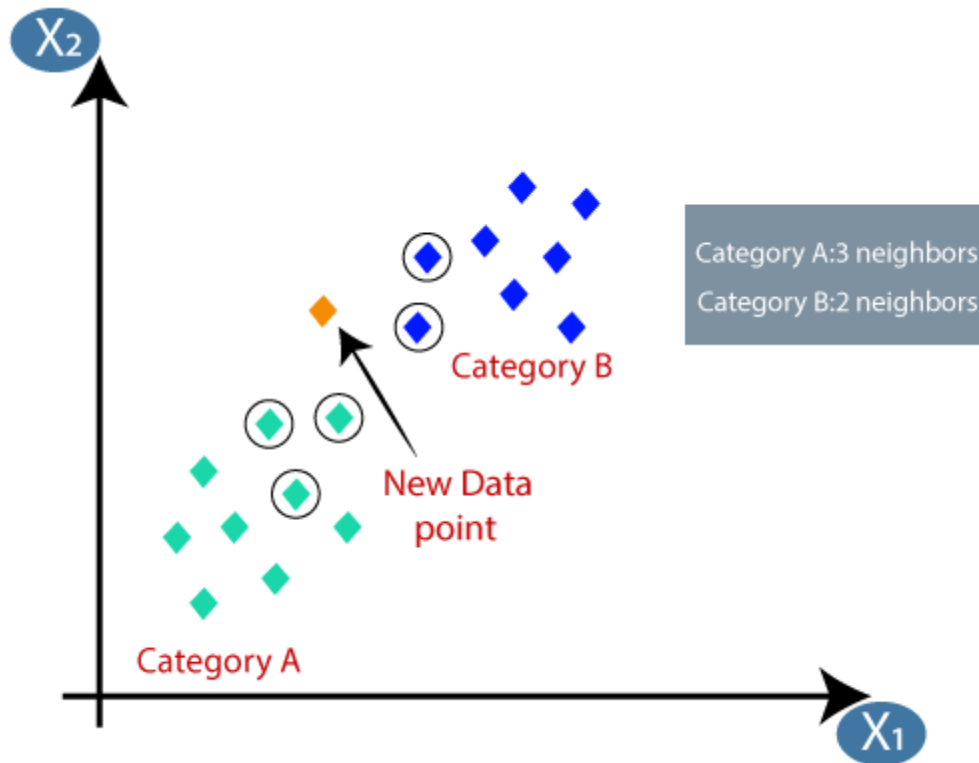
Figure 3.5: This diagram presents the working of the KNN algorithm.

e) SVM:

The supervised machine learning algorithm Support Vector Machine (SVM) is used to categorize and forecast data. SVM sorts the data points accurately by locating a hyperplane in an N-dimensional space amount of features in the data determines the hyperplane's dimension. The hyperplane is a line if there are just two features. The hyperplane turns into a 2-D plane if there are three features. However, once there are more than three aspects, it is more difficult to visualize. [18]. A computer method called the support vector machine (SVM) is capable of learning how to name objects by using examples. By studying several reports of both fraudulent and legitimate credit card activity, for instance, the SVM can learn to identify fraudulent

credit card activity. A vast collection of scanned photographs of handwritten zeros, ones, and other numbers can also be used to train the SVM to recognize handwritten numerals [19].
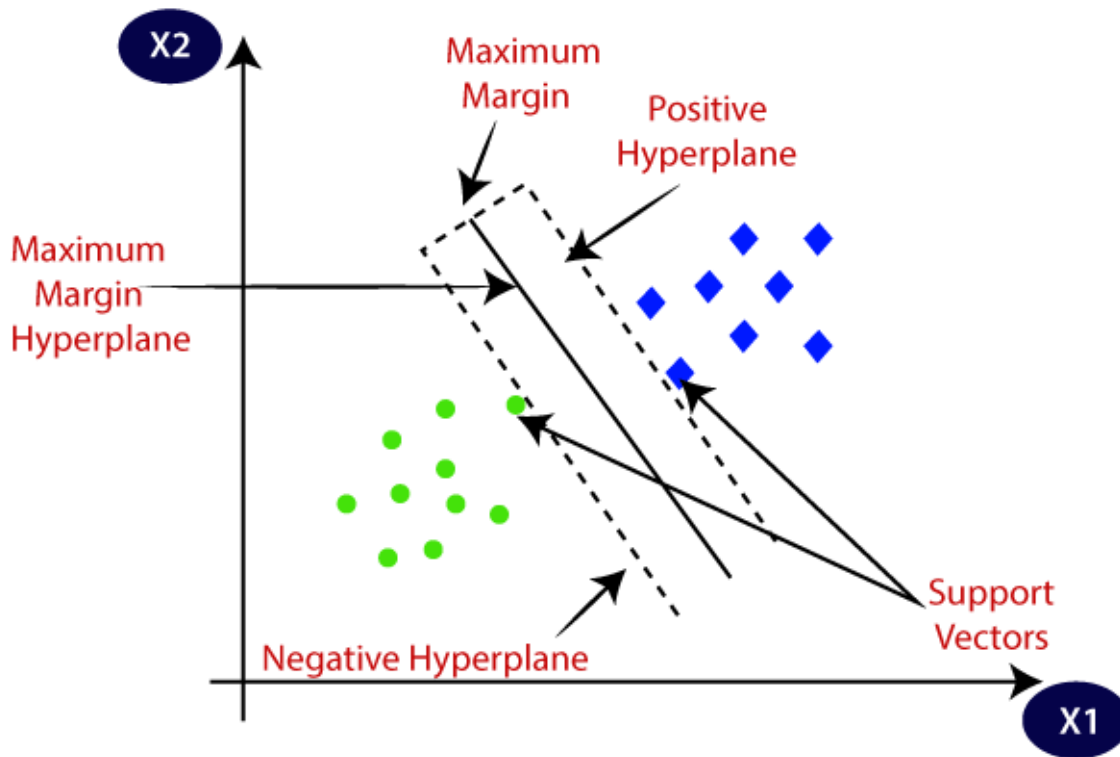
Figure 3.6: This diagram presents the working of the SVM algorithm

f) Linear SVC:

The Linear Support Vector Classifier (SVC) approach conducts classification using a linear kernel function and works effectively with a large number of data. When compared to the SVC model, the Linear SVC contains extra parameters such as penalty normalization ('L1' or 'L2') and loss function. Because linear SVC is based on the kernel linear technique, the kernel method cannot be modified.
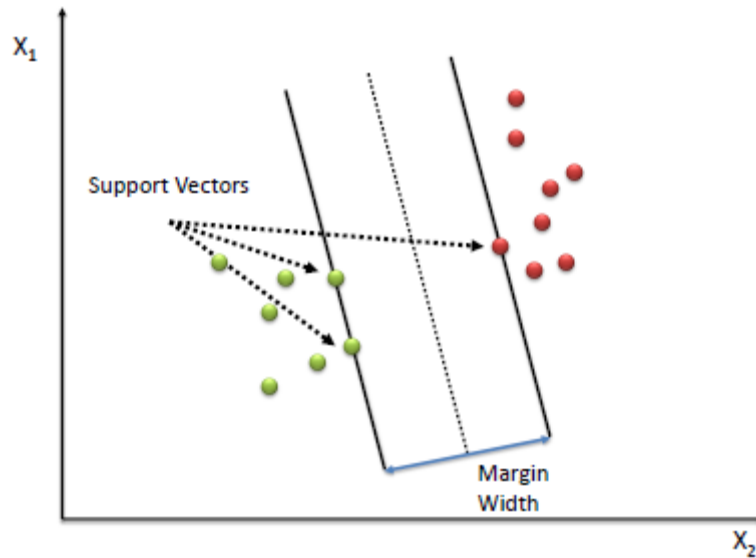
Figure 3.7: This diagram presents the working of the SVC algorithm

g) AdaBoost:

AdaBoost is an ensemble method for sequential training and deploying trees. In order to improve the classification of data that has been incorrectly classified by a previous weak classifier, a series of weak classifiers are connected together using the boosting technique, which is implemented by AdaBoost. Boosting accomplishes this by stringing together weak classifiers to create a strong classifier. Since decision trees used in boosting methods are shallow models that do not overfit but may be biased, they are known as "stump" methods. A specific tree is trained to concentrate only on the shortcoming of the tree that came before it. A previously misclassified sample will have its weight boosted so that the future tree can focus on correctly classifying the misclassified sample. Classification accuracy increases as more weak classifiers are successively added to the model; nevertheless, this may lead to significant overfitting and a loss of generalization abilities. AdaBoost performs well with unbalanced datasets but poorly when there is noise present. AdaBoost requires more time to train. Optimization of AdaBoost hyperparameters are far more difficult than the optimization of RF classifiers.

Figure 3.8: This diagram presents the working of the AdaBoost algorithm.

3.5 Implementation Requirements

• Hardware:

       ▪ Processor: AMD Ryzen 7

       ▪ Ram: 8GB

• Software:

       ▪ Jupyter Notebook

       ▪ Language: Python

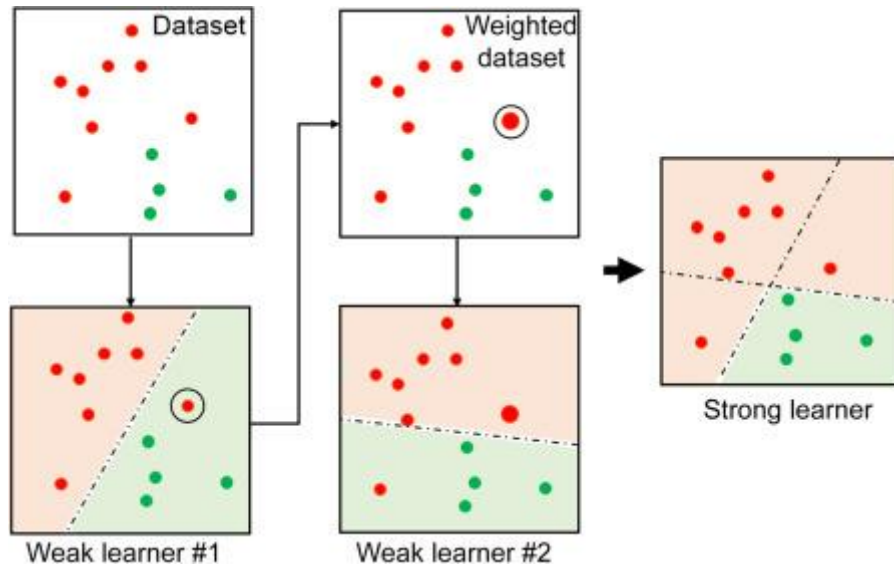   ▪ Libraries: Numpy, pandas, sklearn,nltk etc.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

Introduction:

We will be discussing the results of the general test in this section. We will look at how to measure execution and see what results in this information provides at the end.

4.1 Experimental Setup:

We were aware of how our work affected both our daily operations and our capacity to construct models and gather data. We gathered a large amount of information from numerous social media platforms and websites for our investigation. We used Python on Windows with several libraries such as Seaborn, Pandas, NumPy, etc. Many programs for all training and testing activities were available on the operating system where we utilized Jupyter Notebook. This is a collection of the most often used Python machine-learning commands.

4.2 Results & Analysis:

As a result of the fact that this is essentially a multiclass problem, the classifier produced a confusion matrix. The outcome matrix for each of the classifiers is displayed in Table 3. The confusion matrix mentioned above is used to determine the accuracy, F1, precision, and recall scores for this study. Table 1 demonstrates that the Part classifier outperforms the other four classifiers when looking at the findings as a whole. The Part classifier has the highest accuracy of all the classifiers for all classes 80,79,78,78,78 and 74. We discovered SVM to be the best classification method after applying Linear SVC, AdaBoost, Logistic Regression, Naive Bayes, Random Forest, KNN, and SVM algorithms. The result is shown below:

Table 4.1: shows the resulting matrix for each of the classifiers.

Classifier	Index	F1 (%)	Precision (%)	Recall (%)
Logistic	0	84	77	92
	1	77	82	72
	2	24	67	14

Naive Bayes	0	81	77	86
	1	69	76	63
	2	30	61	30
Random Forest	0	84	79	78
	1	76	80	78
	2	27	44	13
KNN	0	84	79	90
	1	76	70	73
	2	27	44	19
SVM	0	84	80	90
	1	78	80	76
	2	35	61	25

The train test split is only a method for assessing how well machine learning works. It can be applied to supervised learning algorithms and is utilized for classification or regression problems. A dataset split is necessary for a fair assessment of prediction accuracy. The process entails splitting the dataset into two subsets. As a result, I trained my model using the training set before testing it using the test set. This is how I assessed the effectiveness of my model. I used train test split from sklearn to divide the data. Additionally, the training to testing ratio was 8:2. I repeatedly trained my models with various parameters to increase their accuracy. Finally, a decent score was obtained for the SVM model. A comparison of the five experimental models is promised in Table 2.

Table 4.2: Experimental models Accuracy.

Model Name	Accuracy
Logistic	79%
Naive Bayes	74%
Random Forest	78%

KNN	78%
SVM	80%
Linear_SVC	78%

4.3 Result Discussion

The classifier generated a confusion matrix because this is primarily a multiclass problem. Accuracy, F1, Precision, and Recall scores are calculated from the above confusion matrix to evaluate this work. In the overall examination of the results, Table 1 shows that the Part classifier outperforms the other five classifiers. The Part classifier has the highest accuracy of all the classifiers for all classes 80,79,78,78,78 and 74. We examined numerous quality assessment criteria to evaluate the effective classification algorithm. After applying Linear SVC, AdaBoost, Logistic Regression, Naive Bayes, Random Forest, KNN, and SVM algorithm we found the best classifier algorithm as SVM.

Chapter 5

Impact on Society, Environment, and Sustainability

5.1 Impact on Society:

The popularity of sentiment analysis research, which gathers people's opinions, evaluations, attitudes, and emotions from texts in society, has been significantly influenced by Facebook, Twitter, and MySpace. The Impact of Commenting on Social Media The Bengali community uses online marketplaces to shop with friends and relatives and share opinions and ideas. By posting comments on news websites to do so. Bad comments on these posts have a bad effect socially.

5.2 Impact on Environment:

Nowadays social media is used by many people all over the world. And there they express their personal opinions. There are many people who use social media well. And there are many people who abuse social media. Many people express their opinions. After publication, there are naturally many people making many kinds of comments. Those comments are also of different types. Some are negative, some are positive and some are neutral. These social media comments also affect the environment. As a result, it starts all over the world. There are various types of arguments and even fights. The Effects of Making a Comment on Social Media The Bengali community makes use of online marketplaces to shop with friends and family and exchange ideas and opinions. by commenting on news websites to accomplish this. Socially, negative comments on these posts have a negative impact.

5.3 Ethical Aspects:

People may simply make the appropriate judgment and plan if they understand the good and bad parts of a text's mood. Toxic language and statements, such as political propaganda, religious crazy, and random fabrication, are easily detected as unethical. If the majority of good comments about an item on social media outnumber negative remarks, the user can be certain that the object is genuine.

5.4 Sustainability Plan:

Users can quickly form an opinion by knowing the longevity feelings of comments from the majority of favorable and negative public posts on social media. Numerous individuals worldwide use social media today. They also voice their own opinions there. Numerous individuals effectively utilize social media. Additionally, numerous individuals misuse social media. Numerous individuals voice their opinions. After distribution, there are normally many individuals offering numerous sorts of remarks. These remarks also come in a variety of forms. Some are in the negative, some are in the positive, and some are in between. The environment is also impacted by these social media comments. Consequently, it begins worldwide. Arguments of all kinds and even fights occur.

If we can mine the opinions of people on social media and express their sentiments and apply machine learning algorithms to extract the sentiments of those comments, then we can easily understand the sentiment of a person's opinions through those comments.

CHAPTER 6

SUMMARY AND CONCLUSION

6.1 Summary of the Study

Sentiment analysis is the study of judgments, responses, and feelings created by texts. It is widely utilized in domains such as data mining, web mining, and social media analytics since sentiments are crucial features for judging human behavior. This field is making waves in both scientific and industrial communities. Sentiments might be positive, negative, or neutral, or they can include a numerical score expressing the sentiment's efficiency. Sentiments may be conveyed by calculating people's opinions on a certain subject, approach, and sensation toward a unit, where a team might be an occurrence, a theme, or even a character. In many circumstances, sentiment analysis and opinion mining are used interchangeably, yet there are certain instances where they differ somewhat. Sentiment analysis is concerned with identifying views, classifying the attitude they communicate, and finally dividing them divisionally. In the procedure, reviews are first gathered, then their sentiment is detected, characteristics are chosen, sentiments are categorized, and lastly, sentiment polarization is found or computed. Finding the right dataset is a critical consideration when dealing with sentiment analysis. Sentiment analysis is performed mostly because not every review received provides a clear "good" or "poor" impression. Though sentiment analysis is extremely useful, the quantity of training data given into the system determines how effective the analysis is. In general, there are several approaches to categorizing attitudes, one of which is the Machine Learning technique. When considering the Machine Learning technique, may be further divided into supervised and unsupervised learning. While supervised learning is described as the act of learning from previously known data in order to construct an initial model and then predict a target class for the specific data.

6.2 Conclusion:

Various data mining techniques are used to achieve this. The objective of this study is to employ data mining and machine learning to provide high-quality output for social media

comments. To fulfill this assignment, the classifier needs to be trained on 80% and tested on 20% of the total data, respectively. To assess the functional classifier, we took into account a variety of variables. We found that part-based classifiers are better than other data mining methods. SVM is a classifier that is really good at distinguishing between different types of objects. We need help getting good output from the Other 3 or 4 classifiers. We got the least score from the Naive Bayes classifier. So, we believe that our work gives a better output for detecting sentiment to a social media post.

6.3 Implication for Further Study:

In this work, we operate with some specific data, so firstly we collect a very big number of datasets. And apply different types of classifiers algorithm on datasets. In our work, we apply five and more classifiers, and we work with Social media comments. So, in the future, we will work with more classifiers and data mining techniques. After all, we can say that in the future, we will continue our work to make this work as best work.

REFERENCES

[1] T. Cooper, C. Stavros, and A. R. Dobele, "Domains of influence: exploring negative sentiment in social media," Journal of Product & Brand Management, 2019.

[2] A. Goyal, V. Gupta, and M. Kumar, "Recent named entity recognition and classification techniques: A systematic review," Computer Science Review, vol. 29, pp. 21–43, 2018.

[3] N. Banik and M. H. H. Rahman, "Evaluation of na¨ıve bayes and support vector machines on bangla textual movie reviews," in 2018 International Conference on Bangla Speech and Language Processing (ICBSLP). IEEE, 2018, pp. 1–6.

[4] Shafin, M.A., Hasan, M.M., Alam, M.R., Mithu, M.A., Nur, A.U. and Faruk, M.O., 2020, December. Product review sentiment analysis by using NLP and machine learning in Bangla language. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-5). IEEE.

[5] Chowdhury, S. and Chowdhury, W., 2014, May. Performing sentiment analysis in Bangla microblog posts. In *2014 International Conference on Informatics, Electronics & Vision (ICIEV)* (pp. 1-6). IEEE.

[6] Tripto, N.I. and Ali, M.E., 2018, September. Detecting multilabel sentiment and emotions from bangla youtube comments. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)* (pp. 1-6). IEEE.

[7] Ali, H., Hossain, M.F., Shuvo, S.B. and Al Marouf, A., 2020, July. Banglasenti: A dataset of bangla words for sentiment analysis. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-4). IEEE.

[8] Nabi, M.M., Altaf, M.T. and Ismail, S., 2016. Detecting sentiment from Bangla text using machine learning technique and feature analysis. *International Journal of Computer Applications*, *153*(11), pp.28-34.

[9] Random Forest: https://www.tutorialspoint.com/machine-learning-withpython/machine-learning-with-python-classification-algorithms-random-forest.htm. Last accessed 11/01/2023

[10] Random Forest: https://www.javatpoint.com/machine-learning-random-forest-algorithm, last accessed on 27-12-2022 at 11:46 PM.

[11] Logistic Regression: https://www.javatpoint.com/logistic-regression-in-machine-learning, last accessed on 27-12-2022 at 01:57 AM.

[12] Logistic Regression: https://www.javatpoint.com/logistic-regression-inmachinelearning, last accessed 20/09/2021

[13] Naive Bayes:https://www.javatpoint.com/machine-learning-naive-bayes-classifier, last accessed on 28-12-2022 at 09:23 PM.

[14] T. Bayes, "An Essay Towards Solving a Problem in the Doctrine of Chances," Philosophical Transactions of the Royal Society, Volume 53, Issue 1, 1763, pp. 370-418.

[15] J. Tabak, "Probability and Statistics: The Science of Uncertainty," NY: Facts On File, Inc., USA, 2004, pp. 46-50.

[16] F. Yang, "Eliciting an Overlooked Aspect of Bayesian Reasoning," ACM SIGCSE Bulletin, Volume 39, Issue 4, 2007, pp. 45-48.

[17] KNN: https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning, last accessed on 29-12-2022 at 10:55 AM.

[18] SVM: https://www.geeksforgeeks.org/support-vector-machine-algorithm/,last accessed on 30-12-2022 at 07:35 PM.

[19] Boser, B.E., Guyon, I.M. & Vapnik, V.N. A training algorithm for optimal margin classifiers. in 5th Annual ACM Workshop on COLT (ed. Haussler, D.) 144–152 (ACM Press, Pittsburgh, PA, 1992).

[20] Rahman, M.A. and Kumar Dey, E., 2018. Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation. Data, 3(2), p.15.

[21] Rabeya, T., Chakraborty, N.R., Ferdous, S., Dash, M. and Al Marouf, A., 2019, February. Sentiment analysis of bangla song review-a lexicon based backtracking approach. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-7). IEEE.

[22] Soumik, M.M.J., Farhavi, S.S.M., Eva, F., Sinha, T. and Alam, M.S., 2019, December. Employing machine learning techniques on sentiment analysis of google play store bangla reviews. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-5). IEEE.

[23] Al-Amin, M., Islam, M.S. and Uzzal, S.D., 2017, February. Sentiment analysis of Bengali comments with Word2Vec and sentiment information of words. In *2017 international conference on electrical, computer and communication engineering (ECCE)* (pp. 186-190). IEEE.

APPENDICES

To complete this work, we faced some complicated issues. Our main issue was collecting data from people. Because of this, when we surveyed for data collection many people didn't give their proper data or didn't have any data. So, we have wasted a lot of time on data collection. We have faced another issue, that was based on what we thought we would do in our works of sentiment prediction. Then we decided that take social media comments to detect user sentiment for doing this work. And we also face some issues as we need to number our collected data for applying the algorithm and also use a label encoder for numbering the sentiment.

Sentiment Analysis of Bengali comments on social media using Natural Language Processing and Machine Learning

22%	13%	15%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Nayan Banik, Md. Hasan Hafizur Rahman. "Toxicity Detection on Bengali Social Media Comments using Supervised Models", 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), 2019 Publication	2%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	Jabir Al Nahian, Abu Kaisar Mohammad Masum, Sheikh Abujar, Md. Jueal Mia. "Common human diseases prediction using machine learning based on survey data", Bulletin of Electrical Engineering and Informatics, 2022 Publication	2%
4	Submitted to Daffodil International University Student Paper	1%
5	"Proceedings of International Conference on Trends in Computational and Cognitive	1%

Engineering", Springer Science and Business Media LLC, 2021
Publication

6	www.iaeng.org Internet Source	1%
7	Submitted to University of Salford Student Paper	1%
8	Submitted to University of Derby Student Paper	1%
9	lib.buet.ac.bd:8080 Internet Source	1%
10	Submitted to Ahsanullah University of Science and Technology Student Paper	1%
11	Submitted to Napier University Student Paper	1%
12	M. F. Mridha, Md. Saifur Rahman, Abu Quwsar Ohi. "Human Abnormality Detection Based on Bengali Text", 2020 IEEE Region 10 Symposium (TENSYMP), 2020 Publication	<1%
13	Siddharth Misra, Hao Li. "Noninvasive fracture characterization based on the classification of sonic wave travel times", Elsevier BV, 2020 Publication	<1%

14 Hasmot Ali, Md. Fahad Hossain, Shaon Bhatta Shuvo, Ahmed Al Marouf. "BanglaSenti: A Dataset of Bangla Words for Sentiment Analysis", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020
Publication

<1 %

15 Submitted to Pandit Deendayal Petroleum University
Student Paper

<1 %

16 S. Sumathi, S. N. Sivanandam. "Introduction to Data Mining and its Applications", Springer Science and Business Media LLC, 2006
Publication

<1 %

17 Submitted to Alliance University
Student Paper

<1 %

18 Submitted to Mar Baselios Engineering College
Student Paper

<1 %

19 www.nature.com
Internet Source

<1 %

20 Submitted to New Jersey Institute of Technology
Student Paper

<1 %

21 Nolak Kapali, Tariquzzaman Tuhin, Anik Pramanik, Md. Sadekur Rahman, Sheak Rashed Haider Noori. "Sentiment Analysis of

<1 %

Facebook and YouTube Bengali Comments Using LSTM and Bi-LSTM", 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2022
Publication

22   Shaika Chowdhury, Wasifa Chowdhury. "Performing sentiment analysis in Bangla microblog posts", 2014 International Conference on Informatics, Electronics & Vision (ICIEV), 2014
Publication   <1%

23   Submitted to The University of Law Ltd
Student Paper   <1%

24   ebin.pub
Internet Source   <1%

25   A. Stachowski, M. Szydłowski, K. Urbanowski. "From Quantum Mechanics to running Λ cosmologies", Journal of Physics: Conference Series, 2019
Publication   <1%

26   Minhajul Abedin Shafin, Md. Mehedi Hasan, Md. Rejaul Alam, Mosaddek Ali Mithu, Arafat Ulllah Nur, Md. Omar Faruk. "Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language", 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020
                                                     <1%

Publication

27	www.researchgate.net Internet Source	<1%
28	Submitted to Midlands State University Student Paper	<1%
29	dokumen.pub Internet Source	<1%
30	Submitted to National University of Ireland, Galway Student Paper	<1%
31	Submitted to Liverpool John Moores University Student Paper	<1%
32	Submitted to Fresno City College Student Paper	<1%
33	Submitted to Helsinki Highschool Student Paper	<1%
34	Solairaj A., Sugitha G., Kavitha G.. "Enhanced Elman spike neural network based sentiment analysis of online product recommendation", Applied Soft Computing, 2022 Publication	<1%
35	iarjset.com Internet Source	<1%
36	sci-hub.se Internet Source	<1%

37	Submitted to Universiti Teknikal Malaysia Melaka Student Paper	<1%
38	Submitted to University of Birmingham Student Paper	<1%
39	mdpi-res.com Internet Source	<1%
40	Submitted to HotChalk Inc Student Paper	<1%
41	Submitted to Aston University Student Paper	<1%
42	Amit R. Kulkarni, Makarand S. Ballal. "Synergism of synchrophasor measurements and data analytics for enhancing situational awareness of power grid", Computers & Electrical Engineering, 2021 Publication	<1%
43	Submitted to Jacksonville University Student Paper	<1%
44	sigit-purnomo.github.io Internet Source	<1%
45	ulspace.ul.ac.za Internet Source	<1%
46	www.classcentral.com Internet Source	<1%

47	www.tib.eu Internet Source	<1%
48	www.yumpu.com Internet Source	<1%
49	"Intelligent Computing & Optimization", Springer Science and Business Media LLC, 2022 Publication	<1%
50	dspace.bracu.ac.bd Internet Source	<1%
51	www.jetir.org Internet Source	<1%
52	Nafis Irtiza Tripto, Mohammed Eunus Ali. "Detecting Multilabel Sentiment and Emotions from Bangla YouTube Comments", 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), 2018 Publication	<1%
53	conf.kln.ac.lk Internet Source	<1%
54	cs229.stanford.edu Internet Source	<1%
55	ethesisarchive.library.tu.ac.th Internet Source	<1%
56	irjmst.com Internet Source	

<1 %

| 57 | techieyantechnologies.com
Internet Source | <1 % |

| 58 | www.ijraset.com
Internet Source | <1 % |

| 59 | www.ncbi.nlm.nih.gov
Internet Source | <1 % |

| 60 | zenodo.org
Internet Source | <1 % |

| 61 | "Cyber Security and Computer Science", Springer Science and Business Media LLC, 2020
Publication | <1 % |

| 62 | ahmed sakr, mohamed amin, Tamer Grwany. "Performance Evaluation in Arabic Sentiment Analysis during the Covid-19 Pandemic", The Egyptian Journal of Language Engineering, 2021
Publication | <1 % |

| 63 | Md. Serajus Salekin Khan, Sanjida Reza Rafa, Al Ekram Hossain Abir, Amit Kumar Das. "Sentiment Analysis on Bengali Facebook Comments To Predict Fan's Emotions Towards a Celebrity", Journal of Engineering Advancements, 2021
Publication | <1 % |