SENTIMENT ANALYSIS OF GENERAL PEOPLES REACTION ABOUT COVID-19 VACCINATION IN BANGLADESH USING MACHINE LEARNING ALGORITHM FROM BENGALI TEXT DATASET

BY

Puja Sarker
ID: 172-15-9715
AND

Dibbendu Kumar Sarkar
ID: 172-15-10217

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

Supervised By

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Md. Tarek Habib
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

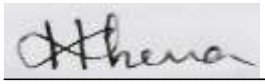
JUNE 2021

APPROVAL

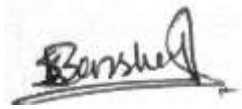
This Project titled "**Sentiment Analysis of General Peoples Reaction About COVID-19 Vaccination in Bangladesh Using Machine Learning Algorithm from Bengali Text Dataset**", submitted by Puja Sarker, ID: 172-15-9715 and Dibbendu Kumar Sarkar ID: 172-15-10217 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 03-June-2021.

BOARD OF EXAMINERS

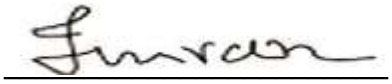
Dr. Touhid Bhuiyan **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Most. Hasna Hena **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Sumit Kumar Banshal **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Shah Md. Imran **External Examiner**
Industry Promotion Expert
LICT Project, ICT Division, Bangladesh

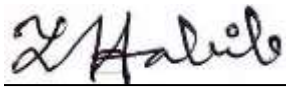
DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Md. Sadekur Rahman, 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:

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

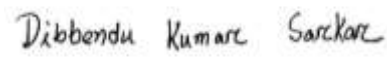
Co-Supervised by:



Md. Tarek Habib
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Puja Sarker
ID: 172-15-9715
Department of CSE
Daffodil International University



Dibbendu Kumar Sarkar
ID: 172-15-10217
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 **Md. Sadekur Rahman**, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of *"Machine Learning" and "Natural Language Processing"* to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts 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 patience of our parents.

ABSTRACT

Nowadays, Sentiment analysis is among the most advance discussable topics in Natural Language Processing field. Sentiment analysis identifies a paragraph's specific pole. Currently, Covid-19 pandemic is one of the most outrageous disease facing by humans. Bangladesh is also suffering by this disease. After detecting the first case of covid. The Government is attempting to coordinate the delivery, vaccination, and distribution of Covid-19 vaccines. Then when the procedure of vaccination starts, the question arises in the minds of the general public whether the vaccine will be good or not? Here, utilizing various classification analysis algorithms based on machine learning, we aim to extract sentiment from the Bengali paragraph, which is ordinary people's reaction to the covid vaccination process. Before, starting the process we have studied various research paper, journal and other online articles to gather knowledge about the process of sentiment analysis. We gathered data for this work from an online survey, multiple social networking sites, and other sources and classify them by Positive, Negative and Neutral class. Preprocessing Bengali text is one of the most difficult aspects of the entire process. We had to overcome several obstacles in order to achieve a satisfactory result. Here, We have implemented six popular classification algorithm which are Naïve Bayes, Random forest, SVM, Decision Tree, K-nearest neighbors, Logistics Regression and two deep learning algorithm, which are LSTM and CNN. Among them CNN provide us the maximum accuracy which is 65.41%.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Coronavirus disease (COVID-19) is an infectious disease caused by a recently discovered virus. It is also known by SARS-Cov2. The majority of patient infected with the COVID-19 virus will have mild to moderate respiratory symptoms and will recover without the need for further treatment. People over the age of 65, as well as those with underlying medical conditions such as cardiovascular disease, diabetes, chronic respiratory disease, and cancer, are at a higher risk of developing serious illness. As per the X. Ou et al [1], The coronavirus is thought to have been transmitted by bats in Wuhan last year. The virus crossed the species barrier as it spread from human to human. After its first patient detection in November 2019 in Wuhan, China, the coronavirus has spread to 220 countries and territories around the world et al [2]. Our Government and WHO have failed to control the outbreak just because it's extremely infectious nature of the virus. As of 30th April,2021 with 150,925,975 total cases, total death is 3,173,003, and newly infected 705,406 confirmed cases have been detected globally [2]. Where, in Bangladesh according to IEDCR there is 756,955 total case, new death 88 and newly effected is 2341 person. [2] We need to follow the safety measurement to stay away from this disease. To resist ordinary people from this atrocious disease our government has started mass vaccination for the general citizen of our country since 10th February, 2021. However, there is still skepticism about the COVID 19 vaccine among some members of the population. Here, in this thesis we have used machine learning algorithm to classify the sentiment of general peoples towards the vaccination process in Bangladesh.

1.2 Motivation

Covid has a very bad impact in our daily life. COVID-19 is a brand-new virus to the human race, and the fundamentals of protective immune responses are poorly understood, so it's unclear which vaccination strategies would be most successful. As a result, it's critical to develop several vaccine platforms and approaches at the same time. Since the outbreak started, scientists around the globe have been striving to develop COVID-19

vaccines, M. Jeyanathan et al [3] stated that at least 166 candidate vaccines now in preclinical and clinical development. People have been subjected to numerous full and partial lockdowns since the year 2020. This has disrupted the usual flow of our schooling, industry, daily lives, and other activities. Many people have lost relatives all over the world. So far, 3,173,003 people have died across the world [2]. Since we live in such a heavily populated country, we are at a higher risk of being affected. Unfortunately, a large portion of our population is unconcerned and unaware of the importance of vaccination, social distancing and follow the safety measurement. This can raise the likelihood of normal people being infected with the virus. Which can be destructive for the society an country. In order to get rid of such a terrible situation. In this thesis work, we're attempting to categorize their attitudes toward vaccination. Because, at present, there is no way to prevent this epidemic without vaccines. As a result, we will identify those who are unaware of the vaccine and covid. Send them appropriate advertising and campaign-related news through their daily used social media to inform them of the effects of covid's influence.

1.3 Rationale of this study

As we know, natural language processing in Bengali is more difficult than other languages. For that reason, less work has been done on Bengali than the other languages. Although various work has already been done on sentiment analysis. But, No work has done directly for the Bengali language, regarding the general public's opinion on the ongoing Covid 19 vaccination process. That’s why we are interested to analysis the sentiment of the people of Bangladesh towards the vaccination process using machine learning.

On the other hand, as we mention earlier that a very large number of people of Bangladesh. Have a negative attitude on vaccine. They don’t understand the value of maintaining social distancing, safety measurement, etc. There is a misconception about vaccines among them. To reduce this, there have no choice but to explain the necessity of vaccination. So, using the text data we can classify their sentiment. After, analysis the data, we can make them aware of a variety of awareness posts, news and campaigns

through social media. Which can be a lifesaving way to aware people about covid. That's why have come up with this idea.

1.4 Research Question

- What is sentiment analysis?
- How do we identify who is aware of Covid and who is unaware?
- What would the state of our original data be?
- Is it necessary to train the machine learning model with our original data?
- What will be the approach to pre-process the text data?
- Does our data and machine learning will be compatible?

1.5 Expected Outcome

We hope that our research will able to inform people about the effects of this deadly disease, based on their perceptions. People can easily and rapidly determine their risk of being affected by covid using this technique. Machine learning can also help people understand more about prediction. Existing or new machine learning algorithms for sentiment analysis have been successfully deployed. Coronavirus will affect a large number of people, and it will be detrimental to our environment, so we should try to avoid it. We occasionally travel to various locations and environments in search of jobs. We don't know whether someone has been infected by the coronavirus, or if anyone has been affected, to what degree I am susceptible. This study will aid us in analyzing the feelings of uninformed citizens and assisting the government in taking the requisite measures to educate them about the effects of being afflicted by the coronavirus. It will safeguard us and our environment against the harmful effects of the coronavirus. Furthermore, the government, the ministry of health, and law enforcement agencies will work together with our model to recognize unaware and whimsical individuals. Who do not adhere to the safety rules and maintain social distance! Even, to teach them about social distance and personal hygiene. They have the ability to take necessary action. In addition, the development of a broad data set for evaluating public opinion on the vaccination process in Bangladesh. We want to publish one or more papers are in international conference proceedings or journals.

1.6 Report Layout

Chapter 1: Introduction

In this chapter we have discussed the introduction, motivation of the work, Rationale of the study and expected outcome of the research and the report layout.

Chapter 2: Literature Review

In this chapter, we have discussed the background of our research. We also provide the information of some related work, background, research summary, and scope of the problem and the challenges of this research.

Chapter 3: Research Methodology

In this chapter, we have discussed our working procedure. What's are in our proposed solution, how our proposed solution works. Our research subject and what was we used in our research. We also discussed sample data.

Chapter 4: Experimental Results and Discussion

In this chapter, we have discussed our experimental results and discussion about our results.

Chapter 5: Experimental Results and Discussion

In this chapter, we cover this research impact on society

Chapter 6: Summary, Conclusion, and Implication for Future Research

In this chapter, we have discussed Summary of our whole research and some recommendations. We also include what needs for future research.

CHAPTER 2
BACKGROUND

2.1 Terminologies

Natural Language Processing (NLP):

Natural language processing (NLP) is a branch of linguistics, computer science, and artificial intelligence that studies how computers engage with human language, particularly how to design computers to process and analyze massive amounts of natural language data. As a result, a computer can "understand" the contents of documents, such as the intricacies of the language used within them. The system can then extract accurate information and insights from the papers, as well as categorize and organize them. [19]

Sentiment Analysis:

The process of detecting positive or negative sentiment in text is known as sentiment analysis. Businesses frequently utilize it to detect sentiment in social data, assess brand reputation, and gain a better understanding of their customers.

Sentiment analysis is becoming a crucial tool for monitoring and understanding client sentiment as they share their opinions and opinions more openly than ever before. Brands can learn what makes customers happy or frustrated by automatically evaluating consumer feedback, such as comments in survey replies and social media dialogues. This allows them to customize products and services to match their consumers' demands. [20]

Machine Learning:

The term "machine learning" is abbreviated as ML. Machine learning is a field that combines aspects of the computer and statistical sciences. It's a model that learns from data that's been observed. After then, unobserved data is used to predict the results. In most circumstances, supplying more data will help it learn faster and perform better. In a nutshell, it's a set of algorithms that help in statistical data analysis. [21]

K-Nearest Neighbors (KNN):

The abbreviation KNN stands for K Nearest Neighbor. This approach is utilized in classification and regression problems as a machine learning algorithm. When it came to pattern recognition and statistical analysis, KNN excelled. It's a KNN method based on similarity of features. It's a basic algorithm with a high degree of accuracy. As a result, this approach is particularly beneficial for analysis (classes, real value). [22]

Naïve Bias (NB):

The term "naive bias" refers to a new term for simple bias. In probabilistic analysis, it's employed. It is used to anticipate the outcome of a scenario. As a probabilistic analysis, this algorithm is employed in Machine Learning. It performs exceptionally well in classification and multiple-class issues. It assigns the highest likelihood to the prediction's outcome. As a result, this approach is particularly beneficial for analyzing distinct classes with varying attributes. [22]

Decision Tree (DT):

A decision tree is referred to as a decision assistance tool. It produces a tree. It has the ability to make judgments and predict future outcomes. We made use of it (explicitly, visually). This can be used to solve decision-making challenges. It's a decision tree that traces each path to give all possible outcomes. As a result, a decision tree is utilized to solve analytical issues including regression and classification. [22]

Random Forest (RF):

Random decision forests are a technique that can be used to solve classification and regression problems, among other things. It is a versatile machine learning algorithm. The majority of the time, it aids in the discovery of a good result. It's employed for simplicity as well as variety. It aids in the making of many judgments as well as accurate predictions. It produces a large number of trees. It solves the problem of overfitting. Its accuracy improves as a result of this. [22]

Support Vector Machine (SVM):

Support Vector Machine (SVM) is an acronym for Support Vector Machine. Both classification and regression issue analysis are aided by it. It aids in the division of classes. SVM creates the line boundary that splits n-dimensional space into classes. [22]

Logistic Regression (LR):

It's a classification algorithm, not a regression one. It's used to calculate discrete values (like 0/1, yes/no, true/false) from a group of independent variables (s). In simple terms, it fits data to a logit function to forecast the probability of an event occurring. As a result, it's also called logit regression. Its output values are between 0 and 1 because it forecasts probability (as expected). [22]

Long short-term memory (LSTM):

It's a unique type of recurrent neural network that can learn long-term data relationships. This is possible because the model's recurring module is made up of four layers that interact with one another. [22]

Convolutional Neural Network (CNN):

A convolutional neural network (CNN, or ConvNet) is a type of deep neural network used to interpret visual imagery in deep learning. Based on the shared-weight architecture of the convolution kernels or filters that slide along input features and give translation equivariant responses known as feature maps, they are also known as shift invariant or space invariant artificial neural networks (SIANN). [22]

2.2 Related Works

This research paper's literature review section will present recent relevant works on sentiment analysis done by some researchers. We have observed and researched their work in order to gain a better understanding of the processes and approaches which they used-

Varghese. et al [4] separate sentiment analysis into three phases. The levels of sentiment analysis are: document level, sentence level, and expression level. Document Level forecasts a single outcome for the entire document based on a group of sentences. Sentence-level sentiment analysis is likely close to document-level sentiment analysis in that it analyses the sentiment of each one sentence. A single sentence expresses a single idea about the subject's nature. To measure particular sentiment, Phrase Level is the best method. There are several drawbacks to the term Level. Often phrases convey literal meaning, and other times they represent abstract meaning. That is to say, negative structured phrases have positive meaning, whereas positive structured phrases have negative meaning [4].

We use sentence-level sentiment analysis to detect sentiment in vaccination-related posts or comments written in Bengali language. To train for sentence-level sentiment analysis, several popular machine learning classification algorithms were used. The document-level has the drawback of being unable to predict the various forms of sentiment found on a single document [4].

Alberto Holts et al. [5] suggested five text document representations. Frequency representation, Binary representation, tf.rf representation, tf.idf representation, tf representation. The input texts are already part of a dataset, but they don't help with model training or algorithm learning. For algorithm learning and text classification, the input texts must be in a specific format. Each pair of documents is given a Boolean value and a set of predefined categories. They used two different languages in the dataset and found that the SVM and Nave Bayes algorithms produced results that were quite similar.

Facebook, Twitter, Blog pages, and other social networking websites, according to M. A. I. Talukder et al [6,] contain various formats of data. He explained how Bengali data is preprocessed in this section. To preprocess our Bengali data, we use this method. It is complex to preprocess the data before working with Bengali NLP, but it is essential to preprocess Bengali text while working with it.

Mita K. Dalal et al. [7] categorized unstructured text; the texts are typically not in a standard format, so classification is required. They train a series of text documents as part

of their job. On the dataset, It used preprocessing. Preprocessing entails deleting stemming, HTML tags, and stop-words, among other things. Then they used LSI, Multiword, TF-IDF, and other techniques to delete functionality. They then selected a classification model for Machine Learning, such as Decision Tree, Support Vector Machine, Nave Bayes, Hybrid method, Neural Network, and so on. They learned and tested the classifier with the help of the trained model. Their model predicted the outcome as a result of this.

Wen Zhang et al. [8] suggested a technique based on the supposition that multiple words can't appear more than once in the text of their own type, so the method excludes the multi words from the text. The data collection was used to classify texts. Procedures for data preprocessing were listed. The preprocessing procedures have the goal of removing multiword features from the main dataset. In that paper, a multi word features for a text classification task was defined, along with how the multi word features were implemented and the results. Furthermore, the nonlinear kernel of SVM and the linear kernel of SVM, two techniques for feature construction, were assimilated. They then wrapped up their comments and production.

Tuhin, Rashedul Amin, et al. [9] make use of a number of machine learning algorithms to assess sentiment based on sentence form and article type analysis, also known as document-level sentiment analysis. In Naive Bayes, they achieve approximately 50% accuracy, and in a topical approach, they achieve approximately 70% accuracy. This accuracy is very good for article-type Bengali data, but it does have some limitations.

To evaluate sentiment, Chowdhury, Shaika, and colleagues [10] use support vector machine and Maximum Entropy algorithm. Bengali Micro blog dataset were used to do their work. Using the Support Vector Machine algorithm, they discovered a 93 percent accuracy rate.

The Naive Bayes algorithm is the most efficient and simple machine learning solution for obtaining reliable, accurate, and quick results, as well as predicting and measuring the likelihood of given data. W. Medhat et al [11] finds that Gaussian and Multinomial Naive Bayes are two types of Naive Bayes that function in different ways. Support Vector

Machine (SVM) is a supervised machine learning algorithm that works for classification or regression. It is frequently used for classification problems to plot data that shows n-dimension (n is the number of shapes) with the value of form. Since it trains in a short amount of time, this algorithm performs well for small datasets [12][11]. Random Forest's most common application is classification analysis. This algorithm uses data to construct a decision tree. It then gathers all of the predictions from the Decision Tree and votes on the best solution. Since it eliminates over-fitting on the average of the answer, the Random Forest is more accurate than the Decision Tree [13]. In the process of storing any possible case and categorizing the new data, the K-Nearest Neighbor (KNN) algorithm tests similarity. KNN is commonly used to categorize data based on the classification of nearby elements [14]. The divides and conquers formula are maintained by Decision Tree, a machine learning algorithm. It divides the dataset into subsets and builds trees based on attribute values. It's referred to as a pure subset if the subset only has one answer. It uses a mathematical formula to predict the outcomes [15]. This are some of the previous works related with sentiment analysis, algorithm and approaches.

2.3 Comparative Analysis and Summary

There are some works has already done about sentiment analysis by Bengali Language with the machine learning algorithm and data mining process. Nowadays, the use of machine learning technology has increased with the creation of new section or research field with sentiment analysis. The comparison between these related works has shown in this part. Here, the comparison of different research works with their subject, methodology, and the outcome are given below in Table 2.1.

TABLE 2.1: Summary Of Related Research Work

SL	Author Name	Algorithm	Description	Outcome
1	M. R. H. Khan, et al. [16]	MNB. RF, DT, SVM, KNN, XG Boost	Depression analysis	Multinomial Naïve Bayes provide highest 86.67% accuracy.
2	N. J. Ria, et al. [17]	NB, XGB, RF, SVC, DT, KNN	Classify Saint and common form.	Naïve Bayes provide highest 76.76% accuracy.
3	M. M. Islam, et al [18]	NB, KNN, SVM, DT, RF	Newspaper headline sentiment	SVM provide highest 75% accuracy
4	R. A. Tuhin, et al. [9]	Naive Bayes Classification Algorithm, Topical Approach	Sentiment Analysis from Bangla Text	Topical Approach provide highest accuracy above 90%
5	S. Chowdhury et al. [10]	SVM, Maximum Entropy	Sentiment Analysis from Bangla Microblog Posts	Best accuracy attained by SVM (93%)

2.4 Scope of the problem

Our research focuses on developing a model through data analysis and machine-learning algorithms. Our proposed model will assess public opinion and determine the likelihood of being infected with the coronavirus. This research would have a major effect on society.

People should be made aware of the harmful effects of the coronavirus. The ministry of health and the law enforcement team will be able to use this model for a variety of activities, gaining an understanding of people's psychology and taking the requisite steps

to combat covid. If anyone becomes infected with the coronavirus, it is unsafe for society and families because it is easy for someone to infect others.

As a result, this model would be useful for average people who do not have a thorough understanding of covid and who want to avoid it. The effects of using machine learning and artificial intelligence for various object detection and disease prediction have recently been quite satisfactory. As a result, we made the decision to using machine learning to develop a model that can predict people's ignorance and apathy toward covid and provide them with the knowledge they need to be aware.

2.5 Challenges

Several problems were faced and resolved, including a history analysis for the factor that causes people to be affected by covid, as well as gathering data from the people. To begin, we sorted through all of the information given by Bangladeshis, the majority of whom are students. Then we used a google form to collect data from the survey on various social media sites. We had a tough time collecting data because people are reluctant to fill out forms. We were requesting person to person for giving their point of view about the vaccination and fill up our survey form. Then there was the problem of coping with the missing meaning, which made pre-processing data more difficult. Even, for the initial level, function and algorithm selection is extremely difficult. Despite these obstacles, using machine learning, it was possible to derive visual knowledge from high-risk data for general people in the sample.

We didn't know what anaconda navigator, jupyter notebook, or any new machine learning algorithms were. It took us some time to understand and learn about it at first, but with the guidance of our supervisor and more practice, we are now able to grasp it quickly. Then we carry on with our work, doing it well and enthusiastically.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introductions

This chapter focuses on data collection, pre-processing, and algorithm selection to determine a trend from the data. Here's a quick rundown of what's going on. How we gathered our data, then pre-processed it by removing punctuation, stopping words, adding contraction, cleaning the dataset, and running some algorithms to find the most reliable model for classifying sentiment from the data.

Following the data collection, we first pre-processed our dataset by converting the text information into a vector format using countvectorizer, and then we applied different algorithm to those data for classifications.

As supervised machine learning algorithms, classification algorithms have been used. To determine whether a post/comment is positive, negative, or neutral, we used multiclass classification in our proposed model.

3.2 Data collection process

The most critical and difficult aspect of our research is data collection. The most important aspect of machine learning is data. It's difficult to get the right result without correct and reliable data. The machine primarily learns from the data provided to it, gathers information from the data, and then provides the pattern. Our study focuses on one of the most unique aspects of the Covid-19 pandemic. There are several online databases that are scraped from different social media sites, but none of them are entirely based on Bangladesh. As a result, we gathered all of the information using an online survey form that our team set up. It is a particularly difficult aspect of our job because persuading people to engage in the survey is extremely difficult. We also gather information from Facebook, Instagram, Twitter, and famous Youtube channels comment sections. By conducting an online survey and using a variety of social media sites, we were able to collect 2313 real-time data from Bangladeshi people.

3.3 Research subject and instruments

We are working on the basis of the general public's feeling and opinion about the ongoing Covid-19 vaccination campaign in Bangladesh. For this, we gathered opinion data from a survey conducted by our team, as well as some random comments from different social media sites. We have data in both Bengali and English from the survey. We manually translated all of the English sentences into Bengali and added them to our final dataset.

From that point on, we used NLTK (Natural language toolkit) to remove stop words, punctuation, and tokenize the data. Countvectorizer was also used to convert text data to vector format.

The Sci-kit Learn library is a well-known framework for working with machine learning and is designed in the Python programming language. It is currently becoming an immensely powerful and effective tool for analyzing textual data. Sci-kit learning seems to have the advantage of allowing us to integrate a variety of libraries that include explicit algorithm visualization tools. In this study, the sci-kit learning library was used to implement all AI methods.

3.4 Statistical analysis

We were able to gather data from 2313 people in total. We gathered information from people of various ages, occupations, and districts. Figure 3.1 shows that we have received 1034 Positive data, 977 Negative data, and 302 Neutral text data from all of our sources of data. We are only evaluating the class because we are dealing with text data and it is class only. On the basis of this information, we developed our model.

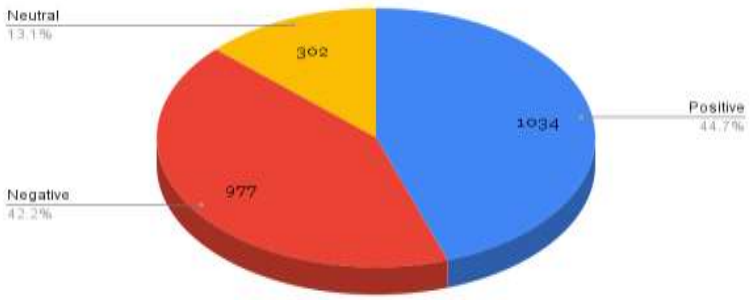


Figure 3.1: Number of Positive, Negative and Neutral Text

3.5 Proposed Methodology

We must go through a series of steps in order to achieve our study objective. They are interconnected to one another. We must pre-process all of the data from our dataset before fitting it into our supervised algorithms and deep learning models since we are working with text data. The entire procedure of our thesis work is depicted in Figure 3.2.

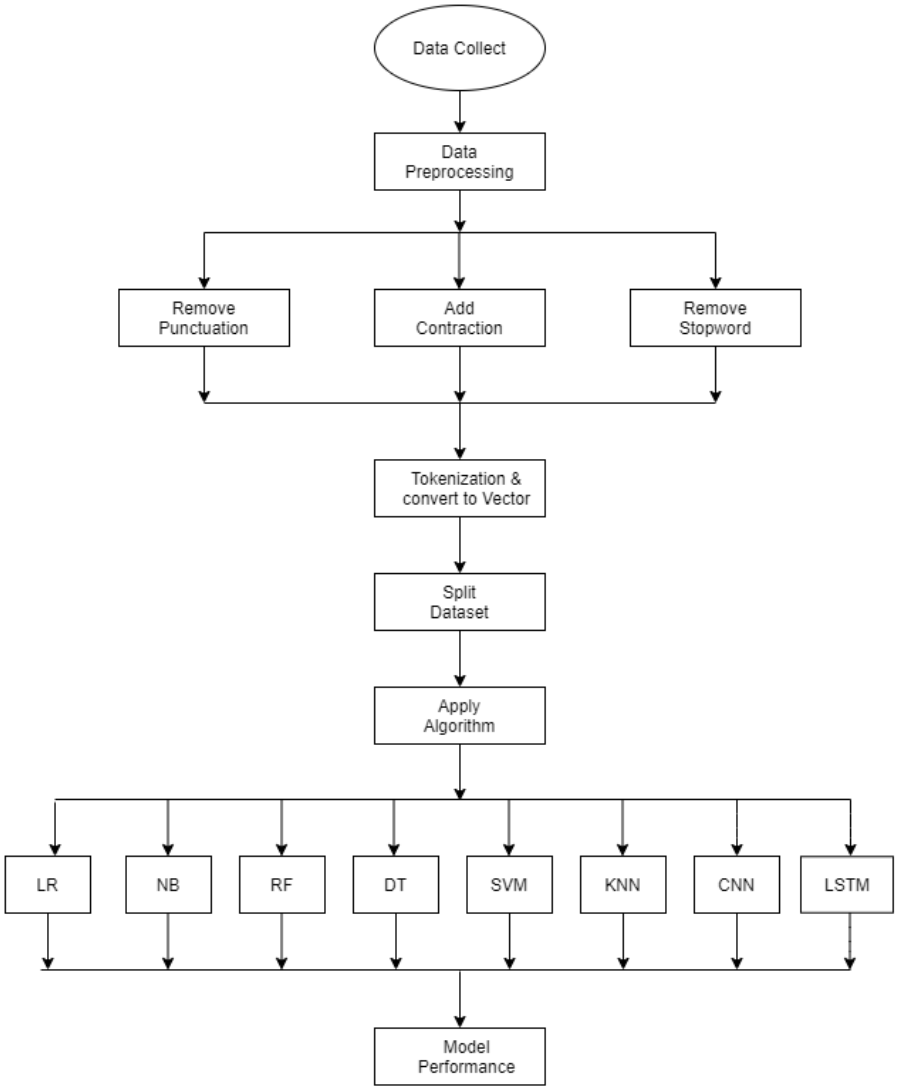


Figure 3.2: Research Methodology

3.5.1 Data Collection

As previously mentioned, data collection was the most difficult aspect of our research project out of all the steps we took. Based on the comments or posts related to Covid-19

vaccination, there was no open-source dataset accessible. We must gather data in solitude. We gathered all of the information from an online survey and manually from the social media website's comments section. In addition, we translate English sentences into Bengali by ourself. Where we needed to put in a little more time and effort to collect data.

```
[ ]  df.head(14)
```

	Data	Class
0	ভ্যাকসিনে ভেজাল করার সম্ভবনা অনেক বেশি তাই আমি...	0
1	আমি মনে করি, ভ্যাকসিন নেয়া টা এই মূহুর্তেই ঠিক...	0
2	যেহেতু বাংলাদেশে দূর্নীতি অনেক বেশি আমার মনে হ...	0
3	বাংলাদেশের দূর্নীতিবাজ লোকেরা ভ্যাকসিন নিয়ে কি...	0
4	ভ্যাকসিন না থাকার চেয়ে থাকা ভালো।	1
5	এটি কাজ করবে না।	0
6	ভ্যাকসিন দেবার সময় যতটা সম্ভব গুজব মুক্ত রাখতে...	1
7	আমরা খুবই গুজব প্রেমি জাতি, তাই দেখা যাবে ভ্যা...	0
8	আমাদের দেশে যত ভালো ভ্যাকসিন আসুক না কেন সেটা ...	0
9	ভালো	1
10	আমি ইতিবাচক এই ব্যাপারে	1
11	নেতিবাচক	0
12	নিরপেক্ষ	2
13	আমি আগ্রহী	1

Fig 3.3: Sample of Dataset

3.5.2 Data Pre-processing

Following the data collection, our dataset contained two types of information: raw text data and class information. This classification was manually assigned by us, and it is based on a voting method. Our honorable supervisor sir oversaw the process, which was double-checked by two other random people. There were two forms of input in our data section: one was English and the other was Bengali. Since we work with Bengali data. As a result, we have translated those data with extreme caution. This was a difficult

challenge for us as well. Some people compose their comments in a phonetic manner; we've taken special care with this exception. For example, “Vaccine ki free te deya hobe?”.

3.5.3 Add Contraction

In our daily conversation we have often we used some shortened word for our convenient while conversation. This types of shortened from can be confusing for our machine to understand the text. So, to avoid this kind of event we can add contraction. Which makes a comfortable and understanding event for our system to understand the appropriate meaning of the certain shortened form.

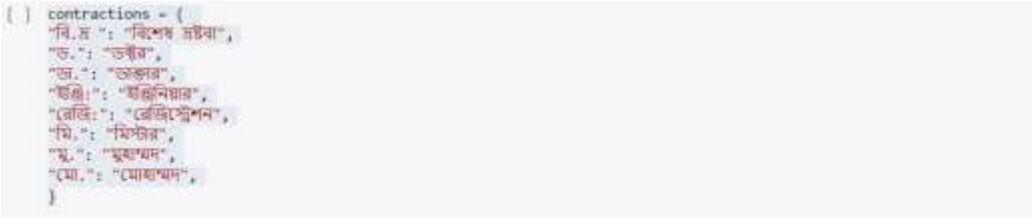


Fig 3.4: Adding Contraction

3.5.4 Remove Punctuation and Stop Words

In text data there are such a large number of punctuations. Also, punctuation doesn't add meaning to a sentence that which likewise confuses the system. So, we removed punctuation.

Before Remove punctuation: Here we can see two sentence where there is ‘Coma’ and ‘Dari’ which are known as Bengali punctuation.

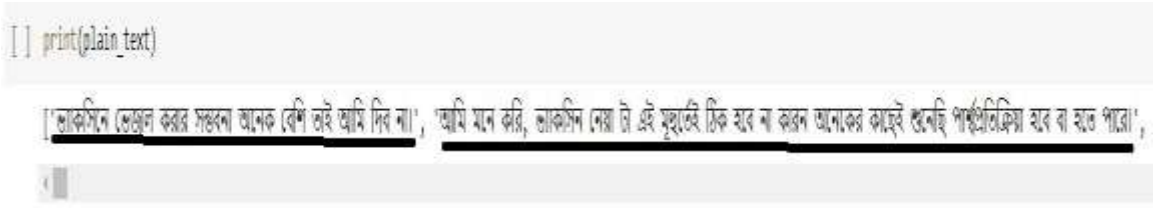


Fig 3.5: Data before remove punctuation

After Remove Punctuation: From the given figure below we can see, the Bengali punctuation have removed from the data after using the python codes print(processed_plain_text).

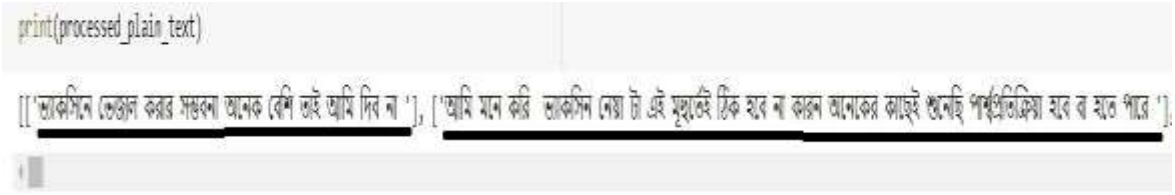


Fig 3.6: Data after remove punctuation

Remove Stop words: In Bengali language, there are some stop words. Which is used with the sentences. This type of words indicates the end of a sentence or complete the meaning of a sentence. These words don't have their own meaning. The express their meaning with the context of those sentence, with whom they used. They don't have any significant effect on the sentences and they can make redundancy. That's why it should be removed. Removing stop word can improve the performance of a model while solving the text classification problem. We can remove stop word using various libraries, But, NLTK is the most well-known among them. In our research work we have used NLTK library to remove stop words.

Figure 3.7: List of stop words.

In the given figure, we can see there is a list of stop words in Bengali language. Which will be automatically removed from the sentences in our dataset. When we will run this list with the NLTK library to remove this stop words.

3.5.5 Tokenization

Sentence tokenizer was used to determine the meaning of each sentence in our dataset. Before we can vectorize the terms, we must first separate the words from the sentence. As a result, the tokenizer will break the sentence down into terms. We can then vectorize the word and apply a machine learning algorithm to the vectorized value. We used the Natural Language Tool-word Kit's tokenizer to differentiate the total number of word amounts. These words become presently accessible for use on our training dataset.

Fig 3.8: After applying Tokenization

3.5.6 Implement Machine Learning Algorithm

Our study is based on the Supervised Learning algorithm and also is a Classification Model problem. There are a variety of tools for implementing algorithms and techniques, each of which is distinct. We used the Sci-Kit Learn library, which is very unique and well-known.

Similar algorithms also produce different outcomes. As a result, choosing the right algorithm is indeed a must-do job. We used the Multinomial Naive Bayes Algorithm, Random Forest Classifier, Decision Tree, Support Vector Machine, K-nearest neighbour, Logistic Regression, Convolution Neural Network (CNN) and Long Short-Term Memory to solve this problem (LSTM). We built a classifier model using these Supervised algorithms.

3.5.7 Accuracy

For our study, we used a variety of supervised algorithms, as discussed in the previous sub-section, and we tried to find the best accuracy for each algorithm.

3.6 Implementation Requirement

- Operating System - Windows
- Environment -Google Colab with Python version 3.7.1
- Data Preprocessing- Pandas, NumPy
- Data Visualization -Seaborn Matplotlib for graph plotting
- Score Showing- classification report, accuracy score
- Sci-kit learn framework
- Algorithm
  - ✓ Logistic Regression
  - ✓ Naïve Bias
  - ✓ Decision Tree
  - ✓ Random Forest Tree
  - ✓ Support Vector Machine
  - ✓ K-nearest Neighbors
  - ✓ Long short-term Memory
  - ✓ Convolution Neural Network

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Introduction

Our research project's test findings are dependent on the discovery of suitable data and the use of the best machine learning models. The language pre-processing part is extremely important for building a Model using a textual information dataset and obtaining the best possible result.

We split our dataset into two (2) sections in order to apply the model on it. There are two datasets: one for training and one for testing. A model training algorithm was applied after the dataset was divided.

4.2 Experimental result

To achieve the expected result, we have used 8 different classification algorithms. Among them Long Short-term Memory (LSTM) gives us the best accuracy which is around 65.41% and the second best accuracy was given by Convolution Neural Network (CNN) which is 63.78%.

Here, in Table 4.1 all the accuracy score of those algorithm which we have used in this thesis work is given below-

Table 4.1: Accuracy Score Measurement Of The Used Algorithm For Dataset

Serial No.	Algorithm	Accuracy Score
01	Long Short-term memory (LSTM)	65.41%
02	Convolution Neural Network (CNN)	63.78%
03	Logistics Regression (LR)	56.15%
04	Naïve Bayes (NB)	52.26%
05	Random Forest (RF)	54.85%
06	Decision Tree (DT)	43.41%
07	Support Vector Machine (SVM)	51.82%
08	K-nearest Neighbour (KNN)	48.81%

4.3 Analysis

Taking a look at all the 8 algorithm we finally got the ideal algorithm which has given us the clear idea that which algorithm is more accurate to understand the sentiment behind a sentence. Here in Table 4.2 and 4.3, we will see the prediction for several algorithm to detect positive and negative sentence.

Table 4.2: Predict Positive Post

Original Text	আমাদের দেশে প্রাপ্ত ভ্যাকসিন করোনা ভাইরাস প্রতিরোধ করতে পারে।
Original Prediction	Positive Comment
Input Text	আমাদের দেশে প্রাপ্ত ভ্যাকসিন করোনা ভাইরাস প্রতিরোধ করতে পারে।
Prediction of the used algorithm	
Long Short-term memory (LSTM)	Positive Comment
Convolution Neural Network (CNN)	Positive Comment
Logistics Regression (LR)	Positive Comment
Naïve Bayes (NB)	Positive Comment
Random Forest (RF)	Positive Comment
Decision Tree (DT)	Positive Comment
Support Vector Machine (SVM)	Positive Comment
K-nearest Neighbour (KNN)	Positive Comment

Table 4.3: Predict Negative Post

Original Text	যে দেশ আমাদের ভ্যাকসিন দিয়েছে ঐ দেশের চিকিৎসকরাই ভ্যাকসিন নিতে অনীহা প্রকাশ করেছেন।
Original Prediction	Negative Comment
Input Text	যে দেশ আমাদের ভ্যাকসিন দিয়েছে ঐ দেশের চিকিৎসকরাই ভ্যাকসিন নিতে অনীহা প্রকাশ করেছেন।
Prediction of the used algorithm	
Long Short-term memory (LSTM)	Negative Comment
Convolution Neural Network (CNN)	Negative Comment
Logistics Regression (LR)	Negative Comment
Naïve Bayes (NB)	Negative Comment
Random Forest (RF)	Negative Comment
Decision Tree (DT)	Negative Comment
Support Vector Machine (SVM)	Negative Comment
K-nearest Neighbor (KNN)	Negative Comment

4.4 Discussion

We are happy with this precision because the majority of the increased findings come from LSTM and CNN; but, if we want to increase the accuracy level, we must properly set up the dataset and pre-process it. Since, based on the literature review, we also know that, despite the challenges, we must pre-process the data as accurately as possible in order to achieve higher precision. Data cleaning is the only way to increase accuracy at that stage. The more data that has been preprocessed, the more accurate the assumptions that this classifier can make.

Based on the study, I believe our dataset's data is quite complex. Since many people's thought processes are likely to vary from one another, our precision was very limited. In comparison of other people. As a result, if we can fine-tune our data further in the future, we may see improved accuracy.

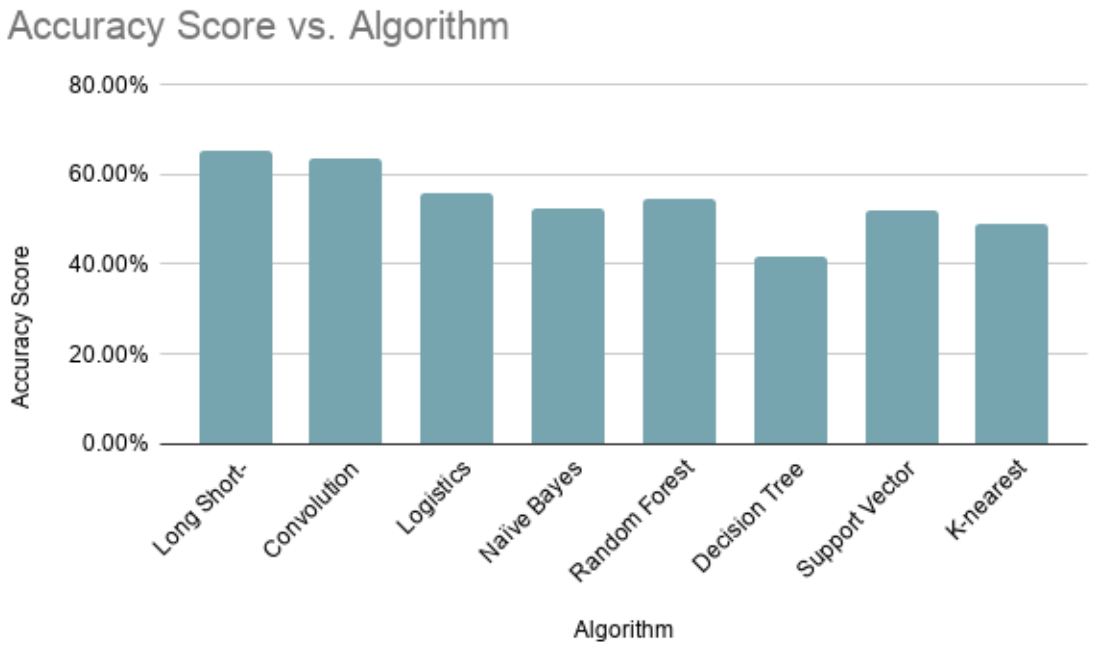


Figure 4.1: Accuracy Score vs. Algorithm

The results we obtained, as well as the existing constraints, are shown in the bar chart above. We can get a good understanding of the algorithm's accuracy based on our dataset by looking at this chart.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMEN AND SUSTAINABILITY

5.1 Impact on Society

Our study would benefit society in a positive way. People of all faiths and castes must work together to survive in a community because humans are social beings. We are all conscious that Covid-19 is currently a major issue in our country's existence. Covid-19 has erected the greatest barrier to living a safe and peaceful life in society. The whole world is working to eradicate this devastating disease. Scientists are working nonstop to save the planet from this deadly coronavirus, and vaccines are being developed to prevent it, as well as attempts to improve its efficacy so that the human body can combat it. However, the vaccine's efficacy, durability, and consequences, as well as how the vaccination process is carried out, are all being questioned, as is the vaccine's impact on the country's and nation's lives. Our research model would benefit the general public in this regard. Who has a negative or neutral attitude toward vaccination? Our model will assist them by sending posts that are relevant to raising awareness about the benefits of vaccinations.

People are thinking about this vaccine and the vaccination process one by one. People are being affected in one way or another by the perspective of others. Their ideas are affecting one another. Furthermore, these viewpoints play a significant role in the study of seasoned experts.

Many people in our country are or have been deceived about vaccinations and the vaccination process, which is currently unsuitable. If the myths persist, people would be unable to be informed, limiting Corona's ability to advance in society. And it has the potential to disrupt society's normal dynamics. The Sentiment Analysis of COVID-19 Vaccination Process model would work in that situation. This model will be able to detect their misconceptions so that we can provide them with the right information later.

There are a lot of positive and poor stories floating around about the corona vaccine and how it works. As a consequence, we are moving away from true news, which can also place human life at risk. Furthermore, how people view the corona vaccine from a

political, economic, social, and religious standpoint, how they view the problem, how it will affect society, what the negative aspects are, and how to solve it. It'll be able to fix the issue by looking at the procedure and the issues that people are having with vaccination, as well as the adverse effects that people are having. It'll be able to make the process simpler and take the proper approach by looking at the process and the problems that people are having with vaccination.

It'll discuss the vaccine or its operation, as well as other topics that are unclear to most people. People are prone to becoming fascinated with superstitions and having a wide range of feelings.

It'll be able to spot them and react appropriately. It'll be able to raise awareness about vaccines and the vaccination process, and It'll be able to address any gaps in the system or deficiencies. Based on the views and feelings of the consumer of the vaccine. We'll be able to respond based on what news is hitting the public and what they're thinking about it.

Aside from that, it will be able to react appropriately to any issues that people have during or after getting the vaccine. If we discover any side effects after the vaccine, we will be able to take appropriate measures and find a solution by consulting the experts.

As the world reacts to the COVID-19 pandemic, we are confronted with an overabundance of virus-related knowledge. Some of this knowledge may be incorrect and dangerous.

Rumors travels quickly, making it more difficult for the public to recognize validated evidence and recommendations from reliable sources like their local health authority or the World Health Organization.

Why are those who are taking the vaccine receiving it, and why some of them are not taking it, why are they not doing it, why are they indecisive? We will be able to interpret their views about how they will determine whether or not to be vaccinated based on any facts, and we will be able to take the appropriate measures to relay the right information to them by reviewing the information gap they are experiencing. COVID-19 Vaccination

Process Sentiment Analysis This model will include the information and data required here so that he or she can know the true news and what to do, and it will also assist our scientists in determining the vaccine reaction among people so that they can take the appropriate measures. In this way, we can provide relevant information to our community, and scientists can study and take measures to avoid the spread of COVID-19. We believe that the Sentiment Analysis of COVID-19 Vaccination Process model would be beneficial to all members of society.

5.2 Impact on environment

Our model is not harmful to the environment in any way. This model does not require any chemicals, combustibles, or organic acids to work. As a result, this model would have no negative consequences for the climate or biodiversity. There is no way to submit any hazardous substances, such as carbon emissions or anything else that affects the environment. People will be able to get the right information and make the right decisions as a result of this research, increasing their knowledge of the vaccine and therefore reducing the spread of the virus, which will benefit the environment and scientists alike. Will attempt to improve human immunity and minimize the occurrence of covid, resulting in less virus spread in the environment, which is a very positive feature for our environment.

This project would also assist citizens in analyzing the sentiment of the COVID-19 vaccination process while causing no harm to the community. As a result, we can conclude that this model would undoubtedly benefit our climate.

5.3 Ethical Aspects

The Sentiment Analysis of COVID-19 Vaccination Process model is not immoral or infringes on human rights. This model does not collect any personal data, such as a person's name or identity. Though collecting data from the citizens of Bangladesh, we did not collect any email addresses or personal details. As a result, there would be no issue with privacy. We haven't shared the datasets with anyone outside the research team until now. This model does not infringe on a person's right to enjoy or use, but it does play a role in raising awareness. The COVID-19 Vaccination Process Sentiment Analysis model

was developed with all forms of laws in mind, as well as privacy and confidentiality concerns. As a result, the model of Sentiment Analysis of COVID-19 Vaccination Process can be controlled without difficulty using machine-learning technology.

5.4 Sustainability Plan

The three components of the sustainability strategy are community, environmental, and operational.

The Sustainability Strategy gives us a fair portrayal of how a project can work and what the project's future aspirations are. Our model cohort's goal is to categorize public opinion on vaccination and to educate them about the risks and benefits of vaccination. This model must be tailored to make it simple for people to adapt, and it is important to remember that using this model does not imply that people are inferior. This model can be used by scientists, police, law enforcement, and ministry of health departments to step up their coronavirus control efforts.

CHAPTER 6
SUMMARY, CONCLUSION, RECCOMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

Our main aim was to create a model that would study people's reactions to the COVID-19 vaccination process in Bangladesh. Data collection, data preprocessing, methodology implementation, and experimental assessment are all aspects of our project. We gathered the necessary text data using a survey form and social media platforms such as Facebook, YouTube, Instagram, and Twitter. Positive, negative, and neutral data were collected. Voting was used to categorize data that was not listed as positive, negative, or neutral. That is, if a data set receives more than 80% of the vote from a panel of three person to identify it as positive, negative, or neutral, then set is approved. We do data preprocessing and implementation using Anaconda Navigator and Jupyter Notebook after collecting the data. We perform five machine-learning algorithms after preprocessing: Random forest, Decision tree, Naïve bayes, Support vector machine, and K-nearest neighbor, and evaluate their output based on accuracy, precision, recall, F1 score, support, and confusion matrix. It is clear that the Long short-term memory algorithm produces the best results. As a result, the Sentiment of the COVID-19 Vaccination Process in Bangladesh was modeled using the Random Forest algorithm for Sentiment analysis.

6.2 Limitations and Conclusions

Sentiment Analysis of COVID-19 Vaccination Process in Bangladesh Using Machine Learning Algorithm from Bengali Text Dataset is the subject of our research. In our work and model, we have some limitations and deficiencies. The data set we used was not particularly large; in order to achieve good accuracy, it would have been preferable to use a larger and more diverse data set. We were unable to collect data from the people of various occupations, counties, and social classes due to some limitations. For data processing, a variety of advanced methods could be used, and the model could be presented in a fascinating way using various variations of algorithm implementation.

It is possible to study the sentiment of ordinary people about the COVID-19 vaccination process using our proposed model. We hope that once this model is fully developed, the general public will be able to understand the significance of this model in raising consciousness about the coronavirus. To stop the coronavirus and to be aware of the negative effects of the coronavirus in our lives, we must always be careful. Since the coronavirus is a highly infectious disease, if people are not aware of how to maintain social distance and protective measures, they will eventually be interacted with another covid affected patient and become infected. If they do not take appropriate precautions at the outset, it would be difficult to prevent the virus from infecting them. We hope that by using this model, people will be more aware of the negative aspects of coronavirus and will be able to monitor their situation.

6.3 Implication for Further Study

Technology and modern science have made our lives simpler and faster in recent years. In the future, we hope to use our model in a program or the backend of a social media web application or an Android application, as information technology and the internet become more widely used in our country. We would be able to improve the consistency of our model in the future by using a larger database. Furthermore, the model's applications can be made accessible to the public by developing user-friendly GUIs. The model can be made more successful in the future by introducing new algorithms, adding different parameters, and adding more functionality. In addition, we have chosen more significant reasons associated with wretchedness in order to increase the information collection metric, model accuracy, and the ability to discover more important and superfluous reasons. A comprehensive database can be built in the future by collecting data from various categories of citizens according to the district. In addition, the model can be expanded and advanced with the aid of a major analysis on human activity in social media.

REFERENCES

[1] X. Ou et al., “Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV,” Nat. Commun., vol. 11, no. 1, p. 1620, Dec. 2020,doi: 10.1038/s41467-020-15562-9..

[2] Worldometer.info, available at <<https://www.worldometers.info/coronavirus/#countries>> last accessed on 30-04-2021 at 1:55 AM

[3] M. Jeyanathan, S. Afkhami, F. Smaill, M. S. Miller, B. D. Lichty, and Z. Xing, “Immunological considerations for COVID-19 vaccine strategies,” Nat. Rev. Immunol., vol. 20, no. 10, pp. 615–632, Oct. 2020, doi: 10.1038/s41577-020-00434-6.

[4] Varghese, R., Jayasree, M.: A survey on sentiment analysis and opinion mining. In: International Journal of Research in Engineering and Technology, eISSN. 2319- 1163, pISSN. 2321-7308, vol. 2 (2013)

[5] A. Holts, C. Riquelme and R. Alfaro, "Automated Text Binary Classification Using Machine Learning Approach," 2010 XXIX International Conference of the Chilean Computer Science Society, 2010, pp. 212-217, doi: 10.1109/SCCC.2010.30.

[6] M. A. I. Talukder, S. Abujar, A. K. M. Masum, F. Faisal and S. A. Hossain, "Bengali abstractive text summarization using sequence to sequence RNNs," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-5.

[7] M. K. Dalal and M. A. Zaveri, "Automatic Text Classification: A Technical Review", International Journal of Computer Applications, vol. 28, no. 2, pp. 37-40, 2011. Available: 10.5120/3358-4633 [Accessed 31 May 2021]..

[8] Wen Zhang, Taketoshi Yoshida, & Xijin Tang. (2007), “Text classification using multi-word features”, 2007 IEEE International Conference on Systems, Man and Cybernetics. doi:10.1109/icsmc.2007.4414208

[9] R. A. Tuhin, B. K. Paul, F. Nawrine, M. Akter and A. K. Das, "An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), 2019, pp. 360-364, doi: 10.1109/CCOMS.2019.8821658.

[10] S. Chowdhury and W. Chowdhury, "Performing sentiment analysis in Bangla microblog posts," 2014 International Conference on Informatics, Electronics & Vision (ICIEV), 2014, pp. 1-6, doi: 10.1109/ICIEV.2014.6850712.

[11] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093-1113, 2014. Available: 10.1016/j.asej.2014.04.011 [Accessed 31 May 2021].

[12] W. Noble, "What is a support vector machine?", Nature Biotechnology, vol. 24, no. 12, pp. 1565-1567, 2006. Available: 10.1038/nbt1206-1565 [Accessed 31 May 2021].

[13] B. Xu, Y. Ye and L. Nie, "An improved random forest classifier for image classification", 2012 IEEE International Conference on Information and Automation, 2012. Available: 10.1109/icinfa.2012.6246927 [Accessed 31 May 2021].

[14] T. Cover and P. Hart, "Nearest neighbor pattern classification," in IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, January 1967, doi: 10.1109/TIT.1967.1053964.
[15] S. Sharma, J. Agrawal and S. Sharma, "Classification Through Machine Learning Technique: C4. 5 Algorithm based on Various Entropies", International Journal of Computer Applications, vol. 82, no. 16, pp. 28-32, 2013. Available: 10.5120/14249-2444 [Accessed 31 May 2021].
[16] M. R. H. Khan, U. S. Afroz, A. K. M. Masum, S. Abujar and S. A. Hossain, "Sentiment Analysis from Bengali Depression Dataset using Machine Learning," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-5, doi:10.1109/ICCCNT49239.2020.9225511.
[17] N. J. Ria, S. A. Khushbu, M. A. Yousuf, A. K. M. Masum, S. Abujar and S. A. Hossain, "Toward an Enhanced Bengali Text Classification Using Saint and Common Form," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-5, doi: 10.1109/ICCCNT49239.2020.9225358.
[18] M. M. Islam, A. K. M. Masum, M. G. Rabbani, R. Zannat and M. Rahman, "Performance Measurement of Multiple Supervised Learning Algorithms for Bengali News Headline Sentiment Classification," 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), 2019, pp. 235-239, doi: 10.1109/SMART46866.2019.9117477.
[19] En.wikipedia.org. 2021. Natural language processing - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Natural_language_processing#:~:text=Natural%20language%20processing%20(NLP)%20is,amounts%20of%20natural%20language%20data.> [Accessed 27 May 2021].
[20] Monkeylearn.com. 2021. [online] Available at: <https://monkeylearn.com/sentiment-analysis/> [Accessed 27 May 2021].
[21] En.wikipedia.org. 2021. *Machine learning - Wikipedia*. [online] Available at: <https://en.wikipedia.org/wiki/Machine_learning> [Accessed 27 May 2021].
[22] Codes), C., 2021. *Commonly Used Machine Learning Algorithms | Data Science*. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/> [Accessed 27 May 2021].

APPENDICES

Abbreviation:

LSTM = Long Short-term memory

CNN = Convolution Neural Network

LR = Logistics Regression (LR)

NB = Naïve Bayes

RF = Random Forest

DT = Decision Tree

SVM = Support Vector Machine

KNN = K-nearest Neighbor

Appendix A: Research Reflection

We knew very little about machine learning and artificial intelligence detection and prediction when we started this research project. Our supervisor was very helpful and sincere. He provided us with invaluable advice and was very helpful. We learned several new strategies, new knowledge, how to use algorithms, and how to work with various approaches during the course of our study. The Anaconda-navigator and Jupyter notebook, as well as the Python programming language, were both new to me.

There were some difficulties at first, but we eventually got more comfortable with Anaconda-navigator, Jupyter notebook, and Python.

Finally, we gathered confidence and were encouraged to do more in the future as a result of our study.

Appendix B: Web Based Interface

After the outbreak, our daily lives have become more reliant on social networking sites in a variety of ways. For instance, education, information sharing, business, entertainment, and income are all examples of sources of income. People spend a significant amount of time on social networking sites. Where they articulate themselves and focus on this website to learn about the world. So, what happens if we apply our templates to such social networking sites? This site can be used as a tool to raise awareness of the corona virus and its negative consequences.

PLAGIARISM REPORT

Sentiment Analysis of General People's Reaction about Covid-19 Vaccination in Bangladesh using Machine Learning Algorithm from Bengali Text Dataset

ORIGINALITY REPORT

13%	8%	6%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Md. Rafidul Hasan Khan, Umme Sunzida Afroz, Abu Kaisar Mohammad Masum, Sheikh Abujar, Syed Akhter Hossain. "Sentiment Analysis from Bengali Depression Dataset using Machine Learning", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020 Publication	2%
2	Submitted to Daffodil International University Student Paper	2%
3	Nushrat Jahan Ria, Sharun Akter Khushbu, Mohammad Abu Yousuf, Abu Kaisar Mohammad Masum, Sheikh Abujar, Syed Akhter Hossain. "Toward an Enhanced Bengali Text Classification Using Saint and Common Form", 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020 Publication	1%