Cricket comment Sentiment Analysis on Bangla texts From social media Using Supervised Machine Learning

BY

Hasibul Hasan Tutul
ID: 173-15-10375

Mithun Saha
ID: 173-15-10301

AND

MD. Shaikh Ahemed Shovon
ID: 173-15-10336

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

Supervised By

Ms. Subhenur Latif
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Md. Azizul Hakim
Senior Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2021

APPROVAL

This Project titled "**Cricket comment Sentiment Analysis on Bangla texts From social media Using Supervised Machine Learning** ", submitted by, Hasibul Hasan Tutul ID: 173-15-10375 and Mithun Saha, ID: 173-15-10301 and Md. Shaikh Ahemed Shovon, ID: 173-15-10336 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 *2-1-22*.

BOARD OF EXAMINERS

Chairman

Dr. Md. Ismail Jabiullah

Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

AHossain

Dr. Md. Fokhray Hossain **Internal Examiner**

Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Reduan

Internal Examiner

Md. Reduanul Haque

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

External Examiner

Dr. Mohammad Shorif Uddin

Professor

Department of Computer Science and Engineering

Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Ms. Subhenur Latif, 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:

Ms. Subhenur Latif
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Md. Azizul Hakim
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Hasibul Hasan Tutul
ID: 173-15-10375
Department of CSE
Daffodil International University

Saha

Mithun Saha
ID: 173-15-10301
Department of CSE
Daffodil International University

Ahemed shovon.

MD. Shaikh Ahemed Shovon
ID: 173-15-10336
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 **Ms. Subhenur Latif** ,**Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Field name*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior 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 patients of our parents.

ABSTRACT

People nowadays use various social platforms and video-sharing mediums to communicate their emotions, ideas, Viewpoints, and Proposals. On Twitter, Facebook, and other social media platforms, there are numerous discussions about sports, particularly cricket as well as football. The viewpoint may communicate detraction in various ways, using notation that may include numerous polarities such as positive, negative, or neutral, and understanding the sentiment of each opinion is a difficult and time-consuming effort even for humans. This challenge can be solved by using natural language processing to analyze sentiment in relevant comments (NLP).[8] In NLP tasks such as sentiment analysis, supervised machine learning classifiers are commonly utilized. We created a dataset of real people's attitudes about cricket in Bangla text in three divisions: positive, neutral, and negative. then processed by removing superfluous terms from the dataset.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Introduction

We live in an era of the internet when we Produce Data in excess of 2.5 quintillion bytes every day, and sentiment analysis has turned into one of the most important tools for making sense of this user-generated data.[10] Sentiment analysis is a widely used tool in NLP. opinion mining. The goal of opinion mining or sentiment analysis is to assess a speaker's attitude toward a topic or the overall contextual polarity of a document. Bangla), one of the more prominent Indo-Iranian languages, is the world's sixth most well-liked language, with a population of more than 250 million speakers. In Bangladesh, it is the primary language, and in India, it is the second (Das & Bandyopadhyay, 2010). Sentiment analysis has been studied extensively in a variety of languages, including English, Urdu, Chinese, and others. While sentiment analysis in Bangla is still in the early stages of development. Because of its scarcity of resources and complexity, there are just a few research works on sentiment analysis in Bangla. As a result, the purpose of this research article is to detect sentiment in Bangla text. People nowadays communicate their feelings through social media sites, newspapers, blogs, and other means.[9] There are also forum discussions, comments, and feelings, as well as opinions on certain topics. Cricket has recently garnered enormous popularity in Bangladesh. As a result, people's reactions to this sport are varied. They like to express their thoughts and feelings about the sport on social media in Bangla. It is possible to understand people's feelings towards cricket by analyzing these reviews. However, only a few attempts at Sentiment analysis have been made due to a lack of highly Organized Bangla Language Processing. As a result, cricket comments sentiment analysis on Bangla text based on people's cricket sentiments has turned into an attractive field of interest for us. Today, evaluate text sentiment, Supervised machine learning technique is extensively utilized. and in terms of Accuracy, it has shown to be a reliable tool, since it takes into account past and future words in relation to the word that will be Used to classify the text is the target term. As a result, we are heavily impressed with classifying cricket sentiment from Bangla text data. We focused on polarity classification at the sentence level here, created an automated method that can extract opinions from the Bangla dataset.

1.2 Motivation

The technique of sentiment analysis and the extraction of views from Bangla language is both fascinating and difficult. People express their thoughts and sentiments about cricket most often through social media in Bangla language, as a result of the growing diversity of sentiment for the sport.[8] Understanding people's feelings about cricket have been a fascinating field for us. Cricket sentiment analysis from Bangla text data from actual people's feelings about Cricket has evolved into a fantastic sport for us. As a result, we decided to do a study on the subject, and we believe that our findings will aid us in better understanding people's attitudes toward cricket in the future, by implementing supervised Machine Learning and increasing accuracy.

1.3 Rationale of the Study

Many studies on sentiment analysis have been conducted on various languages. However, sentiment analysis remains an unsolved research challenge in Bangla, and such research work is extremely rare due to a scarcity of resources and the language's perplexity. as a result of the rising diversity of opinions on cricket Analysis of cricket sentiment in Bangla language from various social media platforms The reactions of real people to cricket has become a fascinating site for us. We can classify the polarity of the sentence level in this project. Opinions were gathered and categorized as positive, negative, or neutral.

1.4 Research Questions

There are various phenomena that assist in bringing up some questions about our work. We can improve our search by asking these questions, and everybody can obtain a clear idea.

- Is it possible to analyze sentiment only from Bangla text or from other languages?
- Can we get a result when more than one sentence is provided as input?
- If we give input outside of our dataset, Is it possible to retrieve the input line result or would we get an error?
- Can it possible to always receive the optimal result (positive, negative, or neutral) or is it possible to get something else?

1.5 Expected Outcome

People sentiment analysis in cricket is the expected output of our research-based project. Here, we must create an efficient process or algorithm that will benefit our project, such as when we give a sentence/Text as input from our Dataset, the output will show the result as positive, negative, or neutral, which will aid in the analysis of cricket sentiment; otherwise, it will not always show the correct result. Furthermore, it can investigate more recently discussed, used topics, motivating them to work on new challenges, create new answers, and assist students who are unsure about their subject of interest.

1.6 Report Layout

The report will be described as follows:

The first chapter contains the meat of our research-based endeavor. In this chapter, we first explain ourselves and our idea. Following that, there is content based on what prompted us to develop this project. The following section will explain why this study is beneficial to us, as well as the rationale for the study. In the last section of this chapter 1, some stuff such as research questions and desired output is also written. The second chapter discusses the many types of works that have been done in this sector previously. Furthermore, the second chapter summarizes the connected efforts as well as the issues that arose as a result of the area's constraints. Finally, it describes the obstacles that could not be surmounted. in Chapter 3 elaborates on the strategy for encouraging the use of statistical methods. it discusses data collection procedures and how to use algorithms. We will outline our technique for this task in this chapter, as well as provide some figures. Chapter 4 explains the results of the experiments and correctly discusses them. We include several experimental tables and photographs in this fourth chapter to show the approach and how we finish our project. The fifth chapter is based on the findings of our investigation. Finally, we will discuss our final work restrictions and obstacles in this chapter. And if somebody wants to do in-depth research in this sector, this research is a great place to start.

CHAPTER 2
Background

2.1 Introduction

Sentiment Analysis has benefited from the expansion of platforms for social media such as Facebook, Twitter, online portal, etc which, generally in the form of writing, give a vast volume of public opinion or sentiment (status, comments, etc). Recently, some study on Bangla sentiment analysis has been published. So We will review the previous study that has previously been done by several researchers . we will highlight the limitations of these works, and we have included the extent of our research as well as the problems in this area.

2.2 Related Works

There is nothing much work using Bangla sentences because it's a complex grammatical structure. Until now, the majority of research has been conducted in English language. However, for this research, various previous works paper on this subject were studied .

In[1]The authors created a deep learning model for training with Bangla language. A tough examination was carried out to differentiate the results with a different DLM (deep learning model) across various word representations.

The fundamental idea is to use a Recurrent Neural Network to represent Bangla sentences as characters and extract information from them (RNN).In[2] The use of contextual valence analysis to assess sentiment from Bangla text .The amount of noun phrases that a verb combines is known as the valence of a verb in linguistics . In terms of overall sense, they compute the entire positive, negative, and neutral of a statement . Using valency analysis, they built their own way for calculating sentiment from Bangla text. in[3] the authors classified a text into six different emotion class using sentiment analysis . They suggested two machine learning strategies to find out emotion from any Bangla sentence: the Nave

Bayes Classification Algorithm and the Topical Approach. in[4] They provide a paradigm for sentiment analyzing in Bangla-language .They develop a classification model using Bangla comments in their proposal . CNN is a type of neural network that generates the model .The classifier model achieves a accuracy of 99.87 %, this is 6.87 % more than the state-of-the art Bangla sentiment classifier.in[5]uses a pre-processing method to extract and save the comments of bloggers on specified issues and identifies the emotion of Bangla language. For 1100 emotional remarks recovered from 20 blog documents, the evaluation yielded precision, recall, and F-Scores of 59.36 %, 64.98 %, and 62.17 %, respectively. In [6] They applied sentiment analysis to a particular domain in this case .They gathered data (comment) from a Facebook page and used two approaches to determine the polarity of each comment .One method is to use Naive Bayes, while another is to use verbal Resources. Following the experiment, they discovered that in certain fields, the lexicon-based method outperforms the others.

in[7] To determine the sentiment of tweets, the authors utilized a semi-supervised technique .They employed a rule-based classifier to annotate the post into positive and negative polarity to provide training data, which they then used to train their sentiment classifier. With emoticons as features, they used the support vector machine (SVM) and the Maximum Entropy (MaxEnt) methods and attained a 93 % accuracy on SVM.

2.3 Research Summary

We looked at a number of publications, research papers, conference papers, and books for our research. Many researchers in computer vision used the same Algorithm on the same dataset .Using the same dataset and Algorithm, you can get a variety of results. Many researcher can not provide enough information which algorithm they used and how processing the dataset. It creates a difficult circumstance in which to implement the algorithm .In addition, multiple implementations are employed in the same procedure, which could result in different results .A small change in landmarks, process, or user input predictability results in a larger change in approach performance .As a result, determining the best ways is a crucial responsibility.

2.4 Scope of the problem

When we were conducting our research, we discovered that for an uninitiated individual, selecting the right research field from the large ocean of knowledge domains is extremely difficult .When a researcher wishes to conduct study, there are numerous fields to choose from . He/she could not choose the majority of the time.

Professors, graduate students, and other researchers in research institutes, colleges, and universities must also find the papers that are most relevant to their study plans. Finding the right papers to read has become an extremely important part of their educational lives .This study document will benefit these individuals by allowing them to serve the best-related papers and also conserving their time .There is a lot of work that is connected to this . No works, on the other hand, have included sufficiently essential information regarding research papers in their recommendations. Furthermore, the majority of the researchers are working with the same dataset.

2.5 Challenges

Dealing with the datasets is the most difficult aspect of this project. Because Bangla sentence is a complex grammatical structure. To address this problem, we'll need some competent ways, but there aren't enough well-known approaches to do it.

CAHPTER 3

Research Methodology

3.1 Introduction

This chapter's main focus is on all of our research's theoretical data. Anyone will get a clear idea of what our thesis is about. To make things easier to comprehend, we quickly gathered some crucial information. In sentiment analysis, it is critical to learn Supervised Machine Learning. So, in order to gain a thorough understanding of our strategy, we will quickly examine the Supervised Machine Learning based methodology. This chapter also covers the research topic and instrumentation.

3.2 Research Subject and Instrumentation

The research is based on Text data in Bangla that was analyzed for the sentiment. using Supervised Machine Learning.

Machine Learning has some advantages-

- rends and patterns are easily discernible.
- There is no need for human intervention.
- Managing data that is multidimensional and diverse
- Higher accuracy
- A wide range of use

Efficiency and scalability of Machine Learning models

They are-

- insanely fast
- remarkably accurate
- easy to tune for resources vs accuracy

Tools or instrument that we use –

Up to now, we have explained the theoretical notions and procedures. Now a list of requirements of instrument are given below

Hardware and software instruments-

- 2.5GHz Intel i5processor
- 8GB memory
- 16000MSzDDR3
- 64-bit operating system

Developing tools-

- Python 3
- NumPy
- Pandas
- Scikit-learn
- Python OpenCV
- matplotlib
- Math

3.3 Steps of Working Process

Workflow refers to how to preprocess data and arrange it step by step so that you can quickly grasp all of the procedures and draw certain limitations. We'll go over what's shown in the diagram step by step. The major purpose of this study is to use a machine learning approach to analyze sentiment from Bangla text. We have broken our work into distinct pieces in order to determine sentiment polarity from raw text.

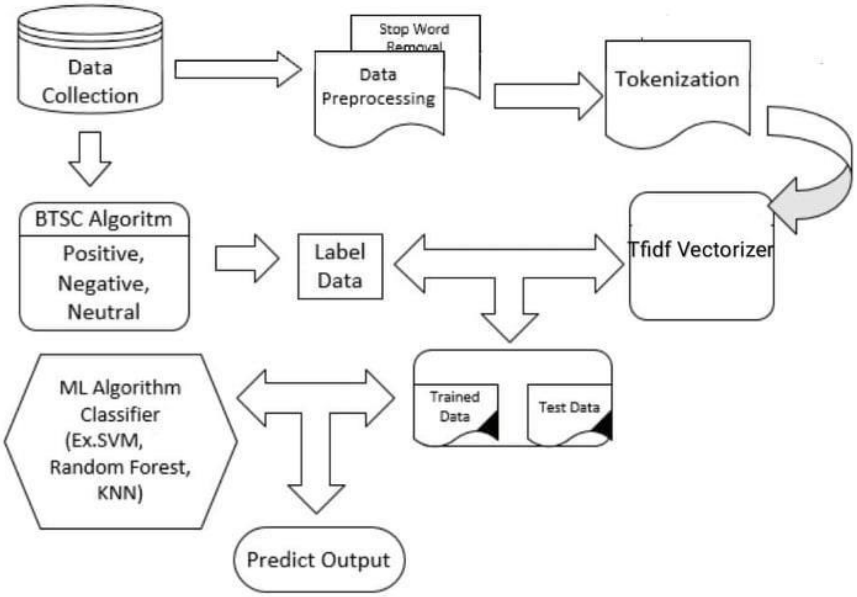


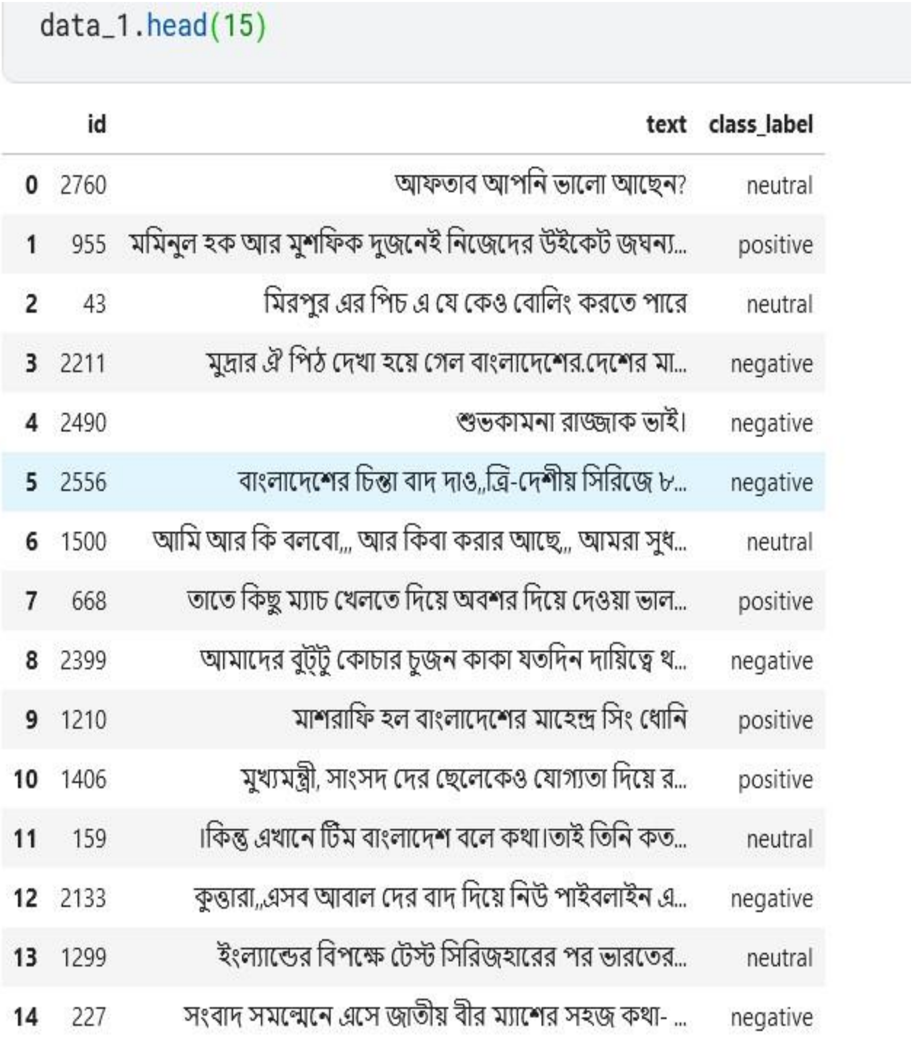
Figure 1 | Visualization of proposed system architecture.

Figure 3.3: Flowchart of the proposed model

3.4 Data Preparation:

3.4.1 Dataset Collection

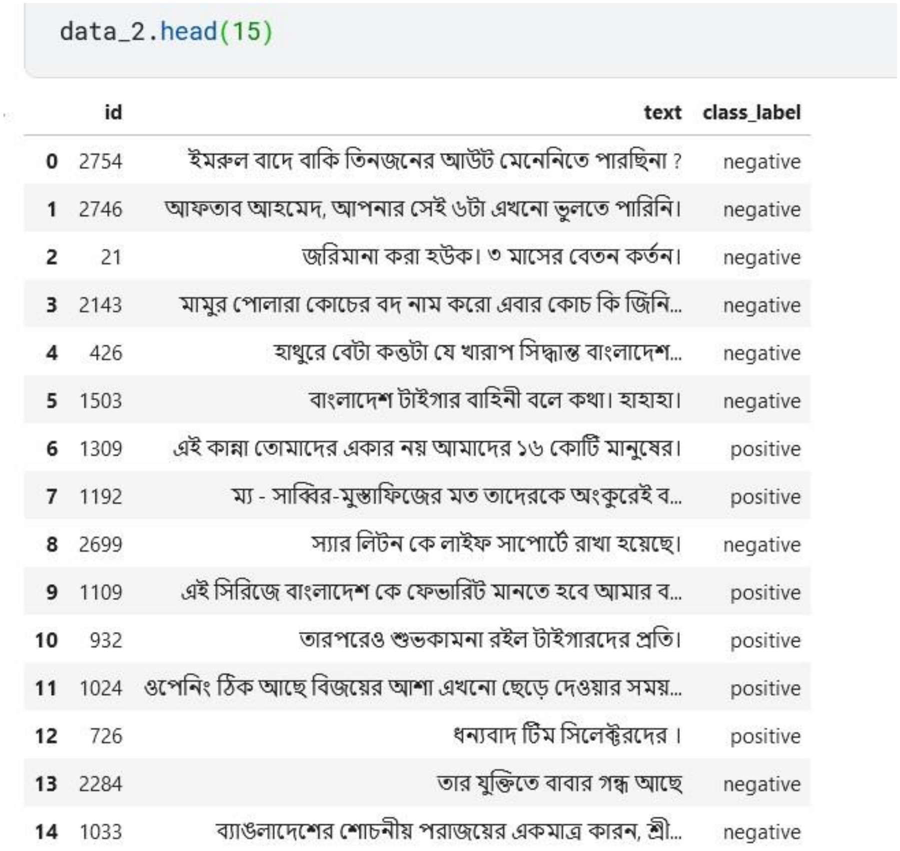
The Cricket dataset contains 2500 unique comments from a variety of web sources, organized into five different aspect groups. The dataset is divided into two parts: train data and test data. As train data, we use 2085 comments, and as test data, we use 380 comments. The majority of the comments were gathered from the BBC Bengali Service and Daily Prothom Alo Facebook pages (https://www.facebook.com/BBCBengaliService/ and https://www.facebook.com/DailyProthomAlo/).Some comments were gathered from BBC Bangla (http://www.bbc.com/ Bengali) and the Daily Prothom Alo (http://www.prothomalo.com), two prominent Bengali websites. The authors of this research produced this dataset. The length of the comments varies, with each review being approximately 3–100 Bangla words. The following are the reasons for using these Websites to collect data .BBC Bangla and Daily Prothom Alo are two of the most popular online news sites for Bengalis all over the world. They are well-known for publishing reliable and accurate information. Bengalis read the news frequently and occasionally write comments to express their views. People submit their thoughts or ideas in both Bangla and English, but they choose Bangla most of the time. We looked at a variety of publications and discovered that people expressed their opinions in Bangla in about 90% of the situations. • Prothom Alo's Facebook page has over 14 million followers, while BBC Bangla's has over 10 million. These two pages have a lot of text posts and a lot of comments on them. • For Bengalis, cricket is one of the most popular sports currently. People are more interested in making comments on cricket-related news than on any other issue, according to our findings. As a result, this was the category we chose for our investigation.



```
data_1.head(15)
```

	id	text	class_label
0	2760	আফতাব আপনি ভালো আছেন?	neutral
1	955	মমিনুল হক আর মুশফিক দুজনেই নিজেদের উইকেট জঘন্য...	positive
2	43	মিরপুর এর পিচ এ যে কেও বোলিং করতে পারে	neutral
3	2211	মুদ্রার ঐ পিঠ দেখা হয়ে গেল বাংলাদেশের.দেশের মা...	negative
4	2490	শুভকামনা রাজ্জাক ভাই।	negative
5	2556	বাংলাদেশের চিন্তা বাদ দাও,,ত্রি-দেশীয় সিরিজে ৮...	negative
6	1500	আমি আর কি বলবো,,, আর কিবা করার আছে,, আমরা সুধ...	neutral
7	668	তাতে কিছু ম্যাচ খেলতে দিয়ে অবশর দিয়ে দেওয়া ভাল...	positive
8	2399	আমাদের বুট্টু কোচার চুজন কাকা যতদিন দায়িত্বে থ...	negative
9	1210	মাশরাফি হল বাংলাদেশের মাহেন্দ্র সিং ধোনি	positive
10	1406	মুখ্যমন্ত্রী, সাংসদ দের ছেলেকেও যোগ্যতা দিয়ে র...	positive
11	159	।কিন্তু এখানে টিম বাংলাদেশ বলে কথা।তাই তিনি কত...	neutral
12	2133	কুত্তারা,,এসব আবাল দের বাদ দিয়ে নিউ পাইবলাইন এ...	negative
13	1299	ইংল্যান্ডের বিপক্ষে টেস্ট সিরিজহারের পর ভারতের...	neutral
14	227	সংবাদ সমম্মেনে এসে জাতীয় বীর ম্যাশের সহজ কথা- ...	negative

Figure 3.4.1: Train Dataset



```
data_2.head(15)
```

	id	text	class_label
0	2754	ইমরুল বাদে বাকি তিনজনের আউট মেনেনিতে পারছিনা ?	negative
1	2746	আফতাব আহমেদ, আপনার সেই ৬টা এখনো ভুলতে পারিনি।	negative
2	21	জরিমানা করা হউক। ৩ মাসের বেতন কর্তন।	negative
3	2143	মামুর পোলারা কোচের বদ নাম করো এবার কোচ কি জিনি...	negative
4	426	হাথুরে বেটা কত্তটা যে খারাপ সিদ্ধান্ত বাংলাদেশ...	negative
5	1503	বাংলাদেশ টাইগার বাহিনী বলে কথা। হাহাহা।	negative
6	1309	এই কান্না তোমাদের একার নয় আমাদের ১৬ কোটি মানুষের।	positive
7	1192	ম্য - সাব্বির-মুস্তাফিজের মত তাদেরকে অংকুরেই ব...	positive
8	2699	স্যার লিটন কে লাইফ সাপোর্টে রাখা হয়েছে।	negative
9	1109	এই সিরিজে বাংলাদেশ কে ফেভারিট মানতে হবে আমার ব...	positive
10	932	তারপরেও শুভকামনা রইল টাইগারদের প্রতি।	positive
11	1024	ওপেনিং ঠিক আছে বিজয়ের আশা এখনো ছেড়ে দেওয়ার সময়...	positive
12	726	ধন্যবাদ টিম সিলেক্টরদের ।	positive
13	2284	তার যুক্তিতে বাবার গন্ধ আছে	negative
14	1033	ব্যাঙলাদেশের শোচনীয় পরাজয়ের একমাত্র কারন, শ্রী...	negative

Figure 3.4.1: Test Dataset

3.4.2 Data pre-processing:

Because real-world data is chaotic and often carry errors, extraneous information, and reduplication, data pre-processing is important in natural language processing (NLP). To offer good analytics results, all punctuation and unnecessary words are eliminated, stemmed to their roots, all missing values are replaced with some values, and text cases are combined into a single one, depending on the application's requirements. As a result, we digest our data one step at a time since it isn't particularly significant in the context of the text. Stopwords Removal: Stopwords ar the most prevalent words in a language and they must be removed. However, these terms have no bearing on a sentence's analytical sentiment. the most widely used terms, such as এবং, এবার, এ, এটা, কী, র, পর, ওরা, কক, ককউ, ইত্যাদি applying the proposed approach has little effect on Sentiment Analysis . However, several words, such as না, নাই, কনই, নয় have a significant impact on negative sentiment and some words such as হযা, স্পষ্ট, করর, কাজ, কারজ have a significant impact on positive sentiment.

3.4.2.1 Text process manually

The importance of text processing in NLP tasks cannot be overstated. Sentiment analysis is unaffected by links, URLs, user tags, and mentions from comments, hash-tags, and punctuation marks. As a result, we eliminate these to provide an unbiased text to annotators. Then, from a comment list of roughly 5000 we collect 2500 sentence . After that, we manually categorized the data into three label : Positive, Negative, and Neutral.

3.4.2.2 Text process Automatic

We tokenized all gathered sentences in the automatic part and deleted numerals, digits, and symbols from the tokenized sentence list . Unwanted characters were eliminated such as

['১','২','৩','৪','৫','৬','৭','৮','৯','1','2','3','4','5','6','7', '8','9'] o ['⍰','?',',',';',':','.','(','-','–
','/','_','*','%','!','\"','+','<','>','—','о','='] o ['”','"','|','…',')','`','@','#','।','“','&','–', '_','
😗','🍭','👌','😆','😂'] o [A-Z] o [a-z] We also removed duplicate sentences .

3.5 Statistical Analysis

Statistical analysis is a component of data analytics. For classifying various sentences, we gathered around 2500 comments. The dataset we divided by train data and test data. In train total negative data 1515 positive 376 neutral 194. In test total negative data 273 positive 73 neutral 34.

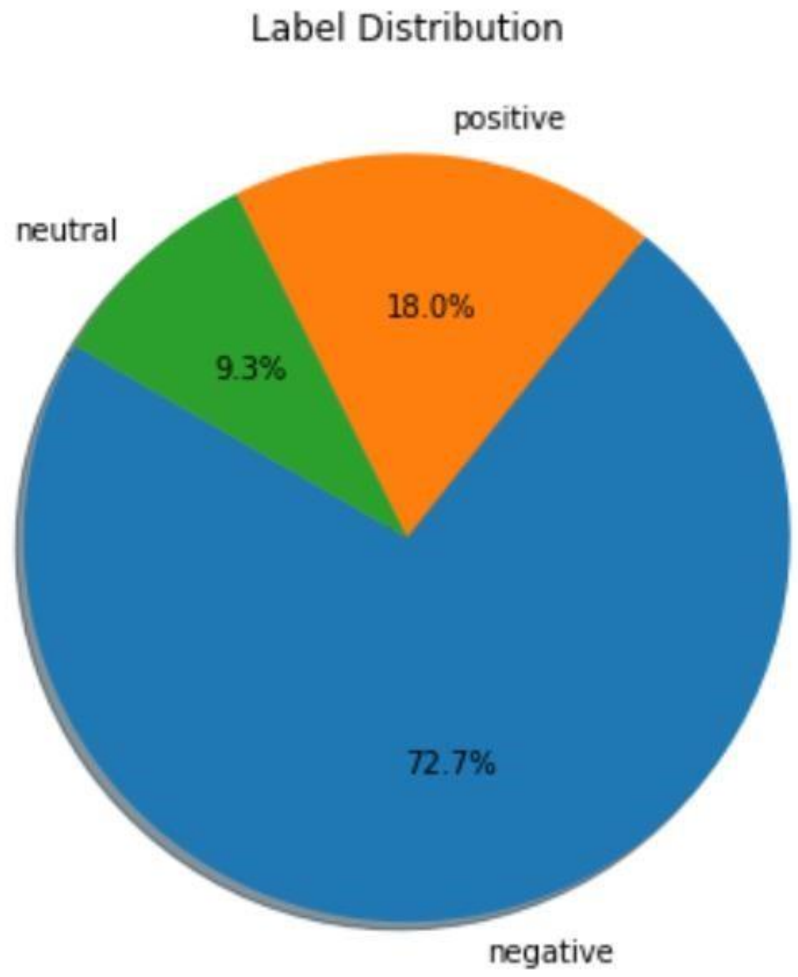


Figure 3.5 : Label Distribution

3.6 SENTIMENT CLASSIFICATION ALGORITHMS

3.6.1 Support vector machine

The Support Vector Machine, or SVM, is a linear model that can be conduct to solve classification and regression issues. It can solve both linear and nonlinear problems and is useful for a large selection of applications. SVM is a basic concept: The method divides the data into classes by drawing a line or hyperplane. The extreme points/vectors that assist create the hyperplane are chosen via SVM. The algorithm is known as a Support Vector Machine, and support vectors are the extreme examples. Consider the diagram below, which illustrates the usage of a decision boundary or hyperplane to classify two distinct categories:

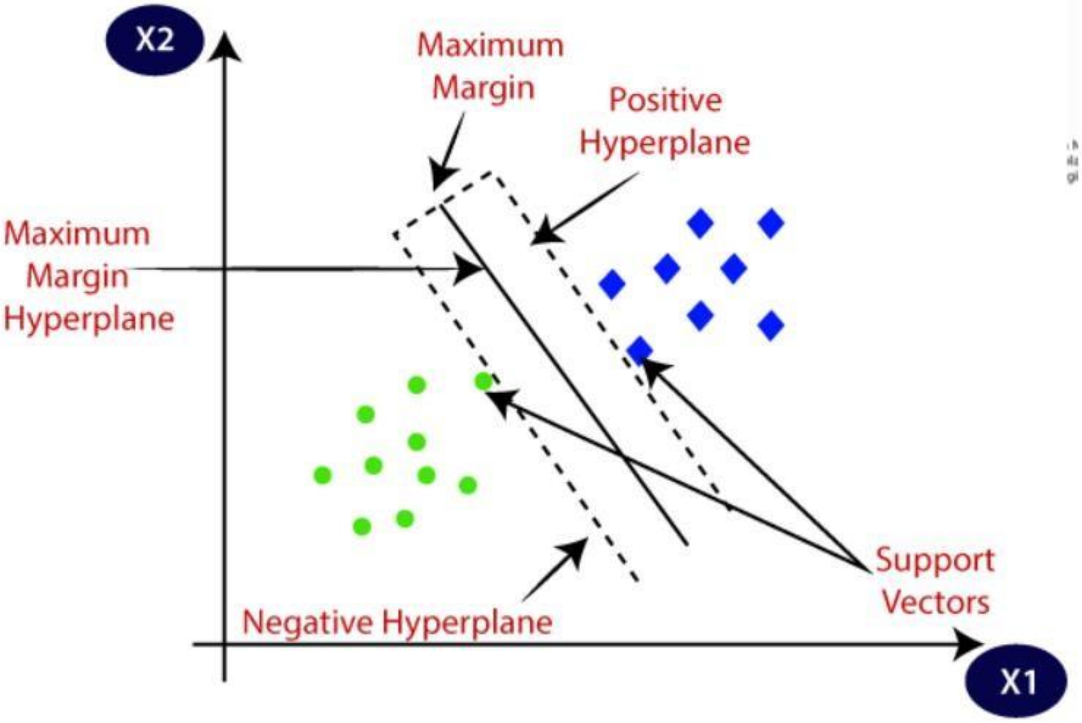


Figure 3.6.1 : Classification of data by support vector machine (SVM)

3.6.2 Random forest

Because it outperformed a single decision tree in terms of accuracy, the random forest classifier was chosen. t's essentially an ensemble method based on bagging. The following is how the classifier works: Given D, the classifier first generates k bootstrap samples of D, each of which is labeled Di. A Di has an equal amount of tuples as D and is sampled using D's replacement. Because sampling with replacement is used, some of the main D tuples may not appear in Di, while others may appear many times. After then, the classifier creates a decision tree based on each Di. As a consequence,

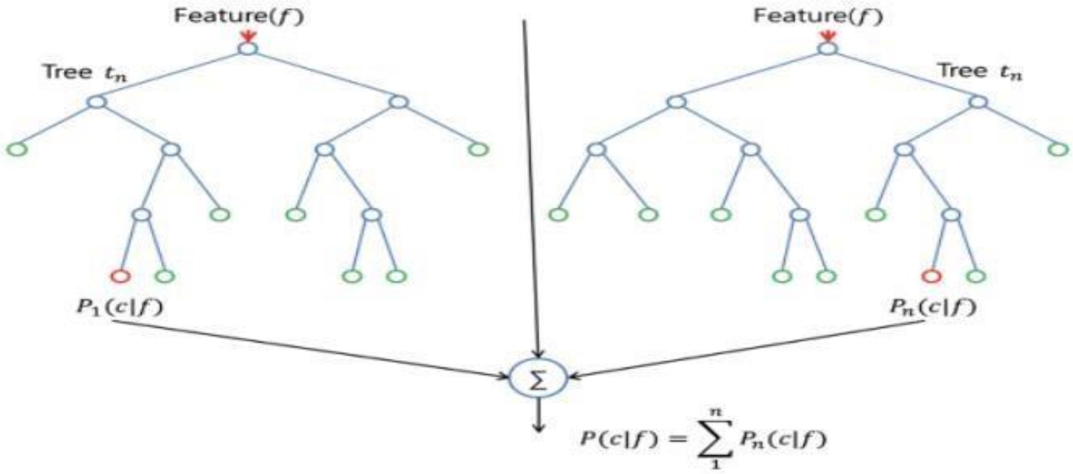


Figure 3.6.2 : Classification of data by Random forest

A "forest" made up of k decision trees is created. Each tree gives its class prediction as one vote to classify an unknown tuple, X. The one with the most votes gets to make the final decision in X's class. Random forest is a supervised machine learning technique based on group learning. Ensemble learning is a type of machine learning in which multiple versions of the same algorithm are combined to create a more accurate prediction model. The random forest technique combines various similar methods, such as multiple decision trees, to create a forest of trees, therefore the name "Random Forest.". Both regression and classification jobs can benefit from the random forest approach. An RF classifier is a collection of tree-structured classifiers. It's a more complex variant of Bagging with the addition of randomization rather than utilizing the best split among all variables, RF divided each node using the best split among a group of predictors randomly chosen at that node.

3.6.3 KNN Algorithm

The K-Nearest Neighbor method is one of the most important Machine Learning algorithms. It is based on the Supervised Learning technique. The KNN algorithm classifies data by finding the K closest matches in training data and then predicting using the label of the closest matches. The K-NN approach saves all available data and categorizes new data points depending on how similar they are to the current data. This means that utilizing the K-NN approach, fresh data can be swiftly sorted into a well-defined category.

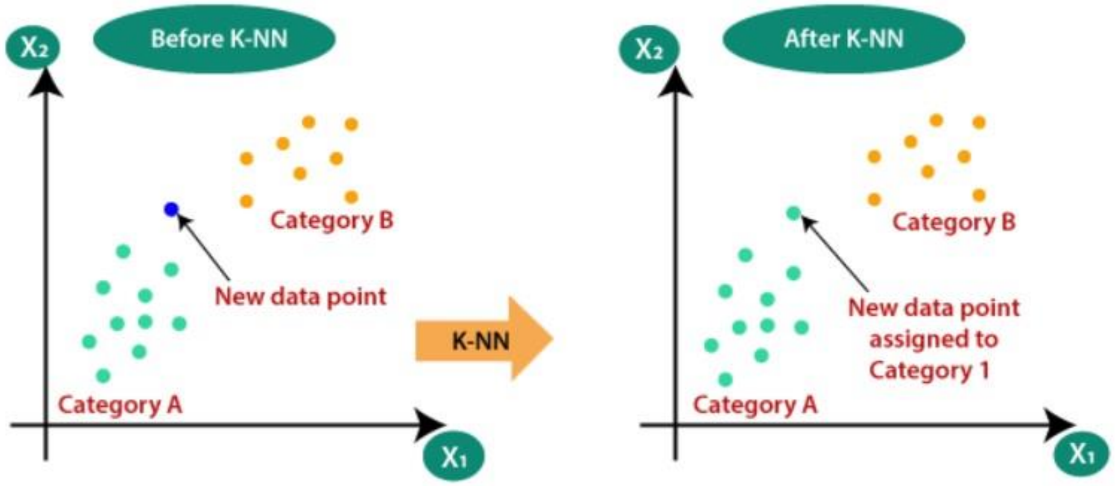


Figure 3.6.3 : Classification of data by KNN

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

We made use of an effective algorithm for sentiment analysis on Bangla text and achieved better precision with a short training period**.** At the end of this chapter, we will understand and find the best algorithm with most accurate outcome by using different type of algorithm. I hope that future academics will follow suit and do fresh study using new data.

4.2 Experimental Result

We utilized NumPy Scikit-learn OpenCV media pipe matplotlib Math and other libraries in this project. We first collect all text comments from various social media sites, then process and read the data, and last apply the Support Vector Machine, Random Forest, and KNN Algorithms.

Support Vector Machine: SVM Algorithm for Sentiment Analysis from Bangla Comments that we Proposed We got 71.58 percent accuracy, 65.55 percent precision, and 71.58 percent recall for F1 66.70 percent.

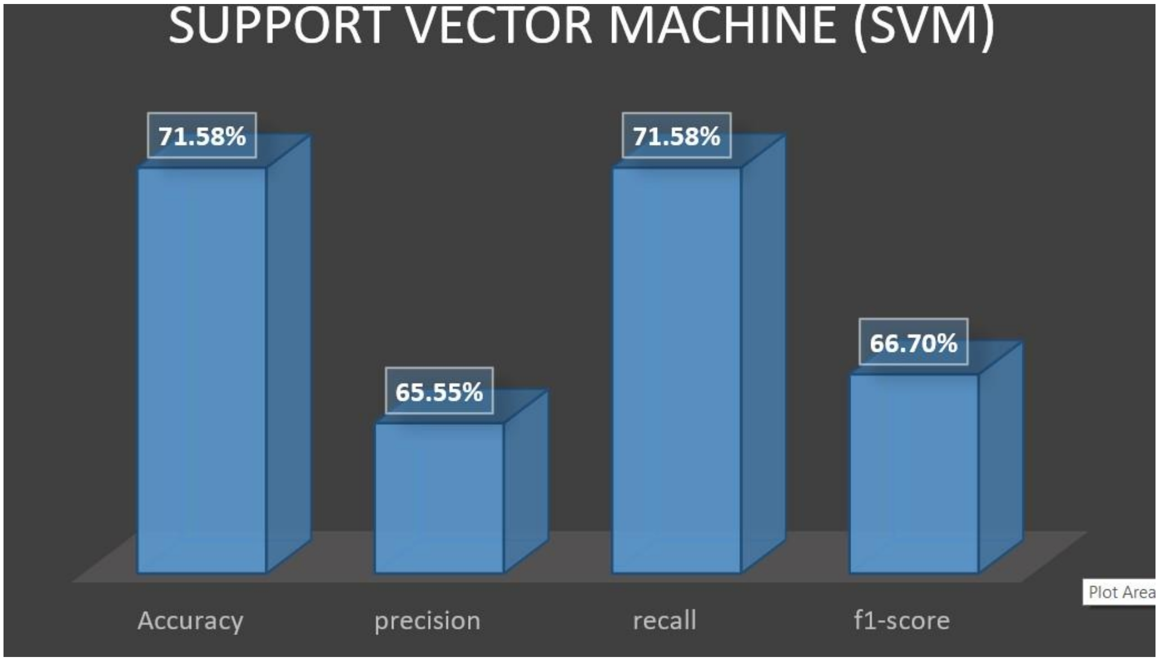


Figure 4.2 : SVM Algorithm performance score

The description of the figures is shown in the following table.

TABLE 4.2 : SVM Algorithm performance score

SVM	Negative	Neutral	Positive
Precision	.75	.25	.49
Recall	.92	.03	.29
F1 score	.83	.05	.36

Random forest: Algorithm for Sentiment Analysis from Bangla Comments that we Proposed. We got 70.26 percent accuracy, 60.29 percent precision, and 70.26 percent recall for F1 63.62 percent.

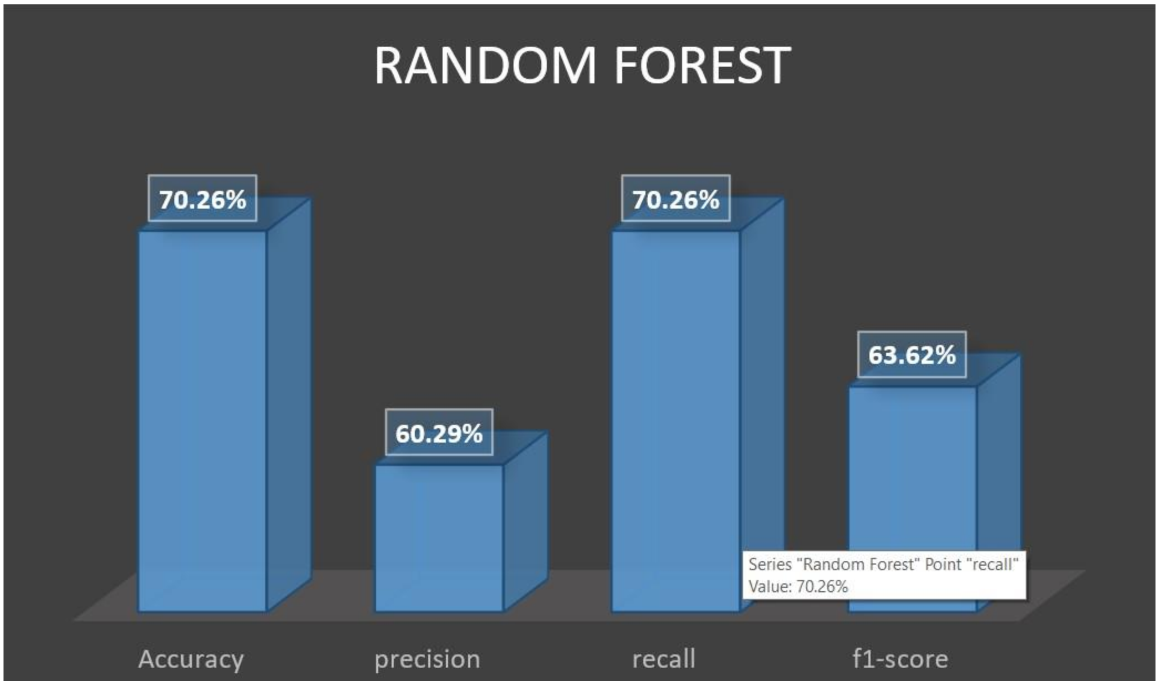


Figure 4.2 : Random forest Algorithm performance score

The description of the figures is shown in the following table.

TABLE 4.2: Random forset Algorithm performance score

Random forest	Negative	Neutral	Positive
Precision	.73	0.00	.41
Recall	.93	0.00	.18
F1 score	.82	0.00	.25

KNN Algorithm: Algorithm for Sentiment Analysis from Bangla Comments that we Proposed. We got 69.74percent accuracy, 60.20 percent precision, and 76.74 percent recall for F1 63.58 percent.

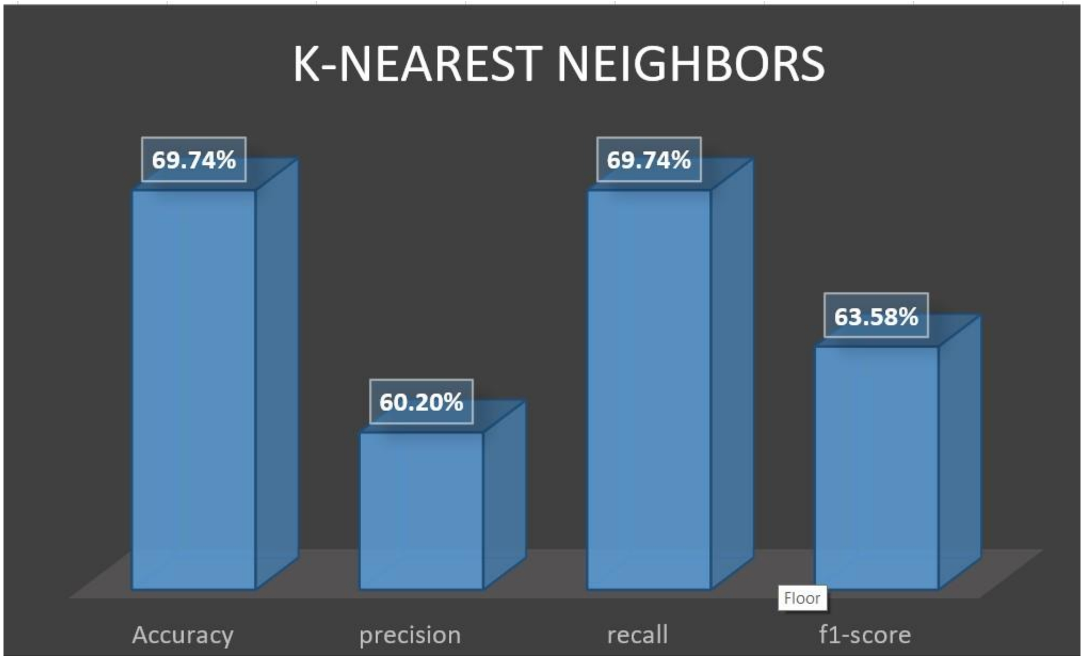


Figure 4.2 : KNN Algorithm performance score

The description of the figures is shown in the following table.

TABLE 4.2: KNN Algorithm performance score

KNN	Negative	Neutral	Positive
Precision	.73	0.00	.39
Recall	.92	0.00	.19
F1 score	.82	0.00	.26

4.3 Summery

Accuracy table for different algorithm

TABLE 4.3 : All Algorithm performance score

	Accuracy	**Precision**	**Recall**	**F1 Score**
SVM	**71.58%**	**65.55%**	**71.58%**	**66.70%**
Random forest	**70.26%**	**60.29%**	**70.26%**	**63.62%**
KNN	**69.74%**	**60.20%**	**69.74%**	**63.58%**

TABLE 4.3: Label wise performance score

	SVM		**Random forest**		**KNN**	
	F1 Score	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**	**Accuracy**
Negative	**.83**	**71.58%**	**.82**	**70.26%**	**.82**	**69.74%**
Positive	**.36**		**.25**		**.26**	
Neutral	**.05**		**0.00**		**0.00**	

At the end of the discussion, we saw that from all the algorithms Support Vector Machine performs better. And hopefully, this research paper will help the student and the researcher who wants to research more on this topic.

CHAPTER 5

Summary, Conclusion, Recommendation, and Implication for Future Research

5.1 Summary of the study

We provide a method for analyzing the sentiment of cricket remarks in Bangla language in this study. The Supervised Algorithm of Machine Learning utilized in Sentiment Analysis was the subject of our research. For sentiment analysis from text data, different Machine Learning algorithms have been used. We also employ the Matplotlib toolkit and the Scikit-learn framework, both of which are useful in analysis.

This part ,we talk about the research's result, recommendations, and future enhancement ideas .

5.2 Recommendations

It is recommended:

- More data set will give excellent output on this research work
- that a rise of knowledge diversity can facilitate to predict a lot of accurately;

5.3 Conclusions

Sentiment analysis is a type of research where the researcher explores people's feelings, attitudes, and emotions about a particular system. Sentiment analysis is a relatively new topic in text mining and computational linguistics that has gotten a lot of press lately. Sentiment polarity categorization from cricket remarks in Bangla text is the subject of this research, which addresses a fundamental topic in sentiment analysis. The end outcome is both satisfying and inspiring. This work, we feel, will also encourage researchers.

5.4 Implication for Further Study

A new dataset will be added to our method in the future, as well as preprocessing processes, making our dataset better formation. We also intend to expand the selected class in order to create a perfect NLP model for this problem**.**

References

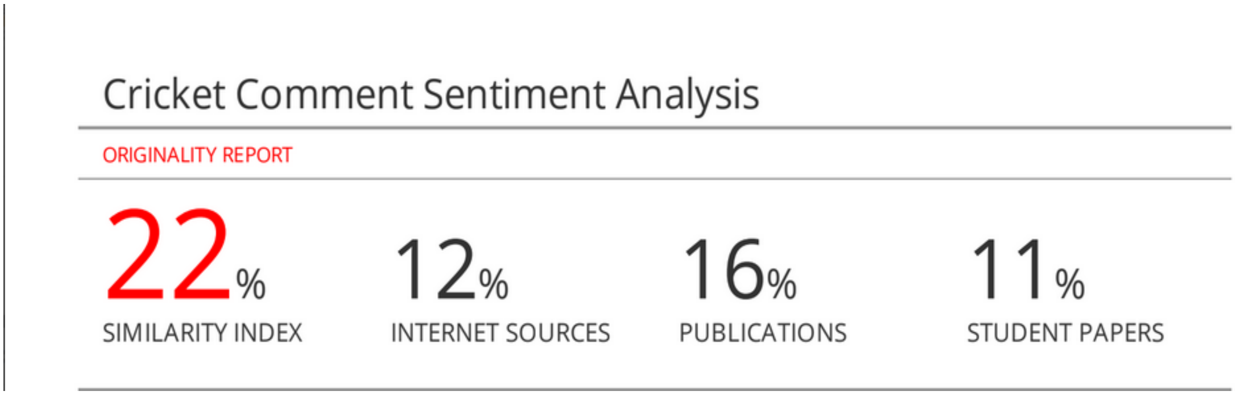
1. Haydar, M.S., Al Helal, M. and Hossain, S.A., 2018, February. Sentiment extraction from bangla text: A character level supervised recurrent neural network approach. In 2018 international conference on computer, communication, chemical
2. 2.Hasan, K.A. and Rahman, M., 2014, December. Sentiment detection from bangla text using contextual valency analysis. In 2014 17th International Conference on Computer and Information Technology (ICCIT) (pp. 292-295). IEEE.
3. 3.Tuhin, R.A., Paul, B.K., Nawrine, F., Akter, M. and Das, A.K., 2019, February. An automated system of sentiment analysis from bangla text using supervised learning techniques. In 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS) (pp. 360364). IEEE.
4. 4.Alam, M.H., Rahoman, M.M. and Azad, M.A.K., 2017, December. Sentiment analysis for Bangla sentences using convolutional neural network. In 2017 20th International Conference of Computer and Information Technology (ICCIT) (pp. 1-6). IEEE.
5. 5.Das, D., Roy, S. and Bandyopadhyay, S., 2012, June. Emotion tracking on blogs-a case study for bengali. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 447-456). Springer, Berlin, Heidelberg.
6. 6.Akter, S. and Aziz, M.T., 2016, September. Sentiment analysis on facebook group using lexicon based approach. In 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT) (pp. 1-4). IEEE.
7. 7.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.
8. Wahid, M.F., Hasan, M.J. and Alom, M.S., 2019, September. Cricket sentiment analysis from bangla text using recurrent neural network with long short term memory model. In *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)* (pp. 1-4). IEEE.
9. Bhowmik, N.R., Arifuzzaman, M., Mondal, M.R.H. and Islam, M.S., 2021. Bangla Text Sentiment Analysis Using Supervised Machine Learning with Extended Lexicon Dictionary. *Natural Language Processing Research*, *1*(3-4), pp.34-45.
10. Haque, S., Rahman, T., Shakir, A.K., Arman, M.S., Biplob, K.B.B., Himu, F.A., Das, D. and Islam, M.S., 2020, February. Aspect based sentiment analysis in Bangla dataset based on aspect term extraction. In *International Conference on Cyber Security and Computer Science* (pp. 403-413). Springer, Cham

.

Appendix

Appendix A: Research Reflection

when we conduct our research, we discover that solving all troublesome features is not simple. We had to first figure out which methodological approach would be best for our project. Furthermore, there was little connected study in this field based on our project, which is a limitation for our effort. Another issue was that we didn't fully comprehend Supervised Machine Learning. But we never give up and devote a lot of time and effort to this area, employing supervised classifiers such as KNN, SVM, and Random Forest, among others, to improve accuracy. Whatever the issue, gathering bangla text data from various social media was exhausting for us, as we had to continue collecting data manually. We eventually completed our entire research-based project following a lengthy term and a great deal of effort



Cricket Comment Sentiment Analysis

ORIGINALITY REPORT

22%	12%	16%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS