

**SENTENCE LEVEL SENTIMENT CLASSIFICATION USING
MACHINE LEARNING APPROCACH IN BENGALI LANGUAGE**

BY

Ahmed Ainun Nahian Kabir

ID: 163-15-8547

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

Supervised By

Mr. ABDUS SATTAR

Assistant Professor

Department of CSE

Daffodil International University



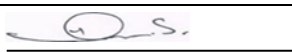
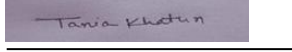

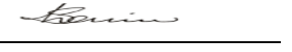
DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2022

APPROVAL

This thesis Project titled — **Sentence level sentiment classification using machine learning approach in Bengali language**, submitted by Name- **Ahmed Ainun Nahian Kabir**, ID No: **163-15-8547**, 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 04-01-2022.

BOARD OF EXAMINERS	
 Dr. S.M Aminul Haque (SMAH)	Chairman
Associate Professor and Associate Head	
Department of Computer Science and Engineering	
Faculty of Science & Information Technology	
Daffodil International University	
 Tania Khatun (TK)	Internal Examiner
Assistant Professor	
Department of Computer Science and Engineering	
Faculty of Science & Information Technology	
Daffodil International University	
 Md. Sazzadur Ahamed (SZ)	Internal Examiner
Senior Lecturer	
Department of Computer Science and Engineering	
Faculty of Science & Information Technology	
Daffodil International University	
 Dr. Shamim H Ripon	External Examiner
Professor	
Department of Computer Science and Engineering	
East West University	

DECLARATION

We hereby declare that; this project has been done by us under the supervision of **Mr. Abdus Sattar, 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:



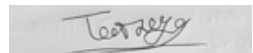
Mr. Abdus Sattar

Assistant Professor

Department of Computer Science and Engineering

Daffodil International University

Submitted by:



Ahmed Ainun Nahian Kabir

ID: 163-15-8547

Department of Computer Science and Engineering

Daffodil International University

ACKNOWLEDGEMENT

First, I express our heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final year project successfully.

I really grateful and wish my profound my indebtedness to **Mr. Abdus Sattar, Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of —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.

I 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.

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

Finally, I must acknowledge with due respect the continuous support and patients of my parents. I appreciate and applaud it with respect and gratitude.

ABSTRACT

In recent times, Emotion extraction from a sentence or word is the most well-known research area in Machine Learning. Huge number of research had already been taken place in natural language processing to detect the gist of a sentence from the text for construct an intelligent system. But enormous number of research has been done on English language. On the other hand, limited number were happened on Bengali dataset through the finite number of Bengali datasets. For my research I used an identical dataset where a lot of sentences were collected from various number of online news portal also different social sites such as Facebook, Twitter, YouTube etc. In my dataset has two columns where ones contain sentences and other contains two types sentiment tags like as positive and negative. My main focus to construct an intellectual system that can detect the emotion of a sentence such as positive or negative.

TABLE OF CONTENTS

Approval Page	i
Board of Examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
List of Figure	viii
Chapter 1: Introduction	01-02
Introduction	01
Motivation	01
Rationale of the Study	02
Research Question	02
Expected Output	02
Project Management and Finance	02
Report Layout	02
Chapter 2: Background	03-04
2.1 Preliminaries/Terminologies	03
2.2 Related Works	03-04
2.3 Comparative Analysis	04
2.4 Scope of the Problem	04
2.5 Challenges	04
Chapter 3: Research Methodology	05-10
3.1 Research Subject and Instrumentation	05
3.2 Data Collection Procedure	05
3.3 Statistical Analysis	06
3.4 Proposed Methodology	07-09
3.5 Implementation Requirements	09-10
Chapter 4: Experimental Results and Discussion	11-16
4.1 Experimental Setup	11-15

4.2 Experimental Results & Analysis	15-16
4.3 Discussion	16
Chapter 5: Impact on Society, Environment and Sustainability	17
5.1 Impact on Society	17
5.2 Impact on Environment	17
5.3 Ethical Aspects	17
5.4 Sustainability Plan	17
Chapter 6: Summary, Conclusion, Recommendation and Implication for Future Research	18-19
6.1 Summary of the Study	18
6.2 Conclusions	18
6.3 Limitation	18
6.4 Implication for Further Study	18-19
References	20
Appendices	20

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.2.1: Snippet of dataset	05
Figure 3.3.1: Quantitative distribution of emotion_2 column.	06
Figure 3.3.2: Dataset objects type	06
Figure 3.4.1: Activity flowchart of proposed methodology	07
Figure 3.4.2: Null value finding	08
Figure 3.4.3: Dataset filtering	08
Figure 3.4.4: Dataset with Label	09
Figure 3.5.1: Model preprocess flow diagram	09
Figure 3.5.2: Base form transformation	10
Figure 3.5.3: Graphical sentiment representation of dataset	10
Figure 4.1.1: Confusion Matrix representation with SVC	11
Figure 4.1.2: Confusion Matrix representation with KNN	12
Figure 4.1.3: Confusion Matrix representation with DT	12
Figure 4.1.4: Confusion Matrix representation with LR	13
Figure 4.1.5: Confusion Matrix representation with NB	14
Figure 4.1.6: Confusion Matrix representation with RF	14
Figure 4.2.1: Accuracy of several machine-learning algorithms	15
Figure 4.2.2: Several algorithms result	16
Figure 4.2.3: Suggest sentiment	16
Figure: Appendices	20

CHAPTER 1

INTRODUCTION

1.1 Introduction

Bangla is the seventh highest number in position all over the world for using this language and fifth position for native speakers [1]. But till now this language has limited resource in Natural Language Processing (NLP) perspective. A lot of number of the research concerning NLP has been taken based on English Language. Due to little number of tools and application were made for Bengali language support such as word level stemmer, chuker, Parts of Speech tagger and word embedder etc. [2]. Through the number of Bengali users in internet are growing day by day, so it inspires us more to research on Bengali language.

Sentiment analysis is a process where a system can detect the emotion of a word or sentence for improving the sectors of business or a technological situation also very useful for internet platform. The developed of machine learning algorithms various no of way are used to extract the emotion of a sentence such as ruled based machine learning or Deep learning processes.

In my work I use a unique dataset which contains around 4177 Bengali sentences that are used to detect the emotion in supervised learning approach and also compare the accuracy of different supervised learning algorithm.

1.2 Motivation

Now the availability of huge number of data sources around us has concerning the research on Data Science and in the area of Artificial Intelligence. Internet makes our life easier to get any types of information at any time. For that, it is easier to work with much amount of data using several machine learning approaches. Sometimes it is necessary to realize the context of a word or sentences instantly in business or situation perspective, but it is quite impossible for human to process. That's why, my main focus to construct an intelligent system that deal with Bengali language for extract the emotion of a sentence. Here I applied Random Forest (RF), Decision Tree (DT), K-nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM) and Naïve Bayes classifier algorithm.

1.3 Rationale of the Study

Due to the scarcity of quality dataset on Bengali language, I made a dataset which holds around 4177 sentences those were collect from several online news portals, social sites and at last I done it successfully. Now the first steps of my research work are to analyze the accuracy of several applied algorithm and then elect the algorithm which generates the best result were applied for our intelligent system. For uniqueness of my dataset, I think we get unique result.

1.4 Research Question

In below here I mention our research question-

- Build an intelligent system depends on best accuracy among those algorithms which deal with Bengali Language and gives the sentiment of the sentence.

1.5 Expected Outcome

My prime goal is to form an intelligent system which can detect the emotion of a Bengali sentence using different types of supervised machine learning algorithm compering the best one considering the highest accuracy. Another target is to generate the most precise output.

1.6 Project Management and Finance

Project Management and Finance is a flow of a cycle who works following systematic way and take a step to provide all types of materials and tools that needs for the project and also deal with project finance matters. But I done this research works without any finance contribution of any organization and persons.

1.7 Report Layer

Here I made report based on my research works. In Chapter 2, I will introduce the overview of machine learning background and my built project. In Chapter 3, I will represent the work flow of different machine learning project which will be easily constructable following the machine learning algorithm. In Chapter 4, I will discuss about all of my experimental setup and the result which my system generates. In Chapter 5, I show how important of this project for our society and the world. In Chapter 6, I represent the conclusion, abridgement and scope for the further work.

CHAPTER 2

BACKGROUND

2.1 Preliminaries/Terminologies

Machine learning, Bengali dataset, Opinion mining, Sentiment analysis

2.2 Related Works

Now, sentiment extraction has become more significant in the research area of Artificial Intelligence. The earlier research on this area were easy compared to current works. Sentiment lexicon and syntactic parser were used on a benchmark data source in [3]. Authors of [4] already been performed Sequential Minimal Optimization (SMO) algorithm [5] depends on the adjectival appraisal group to detect sentiment. Extraction process for sentiment enhance on a mountainous scale after that. Also, Valence Aware Dictionary for Sentiment contention [6] is updated process that based on sentiment lexicon that act on a par with human raters of sentiment intensity.

In recent times, researcher try to use so much tangled procedure for generate best accuracy compared to prior works. For that Character to Sentence Convolutional Neural Network (Chars CNN) which is known as a novel deep convolutional neural network architecture called [7] was invented to detach emotion from words or a sentence that generates a precise output. In [8], a several procedures were gained where the researchers try to introduce Universal Language Model Fine tuning (ULMFiT) model that is capable to passing learning in NLP works like as sentiment analysis. Recurrent Neural Network (RNN) also be a most popular technique in this works such as text and speech-based prediction and opinion mining. The main reason for this famousness is that RNN performs sequence modeling which has been essential for NLP works. Authors of [9] used Gated Recurrent Neural Network (GRNN) model to detect the emotion on English Language. Using the same dataset of IMDB [10] Yelp [11] they were performed their task.

A few numbers of prominent and remarkable research works based on Natural Language Processing on Bengali language has done compare to English language. Previous work in this area is [12] where classification of positive and negative polarity from Bengali sentences has been done based on SVM [13]. A semi supervised process to detect sentiment was obtained from Bengali microblog posts [14]. Moreover, is a

parallel work where the authors have performed an analysis of contextual valence for each word in a Bengali sentence. An attractive process for sentiment analysis is where researchers have constructed a dataset containing of texts from Bengali horoscopes and performed classification using several Machine Learning algorithms [15].

2.3 Comparative Analysis

In earlier most of the work has already been done on social media tweets for English Language. On the other hand, a limited number of research is done on Bengali language. Here my main target is to sentiment analysis on the level of a Bangla sentence, that's why we can see a large number of the works has been completed on different social medias such as Instagram, Facebook post, tweets comments and also online news portals. Now here I construct a dataset which holds Bengali comments of various online news portal and using word embedding technique on this dataset [16].

In my research works, I used several machine learning algorithms for detect sentiment easily from a Bengali sentence and try to build an intelligent system based on highest accuracy among those algorithms.

2.4 Scope of the Problem

Since this is running a modern era, but sometimes we cannot completely enjoy the all of facilities of modern science due to the inability of expressing correct sentiment by the native language. The reviling of Bengali text in different social sites those cannot deal our information technology machine. That's why machine learning technique is so much appropriate to regain this problem. My main focus is to construct an intelligent system that is capable to find the emotion from a word or sentence by using several machine learning techniques.

2.5 Challenges

Opinion mining is an intention to detect the prime gist for a man on specific topic. An intelligent system is developed by me is susceptible to inform us the sentiment of this text. For building an automated machine here I use several machine learning algorithms which can detect the emotion of a Bengali text instantly which is my prime goal.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research subject and Instrumentation

My research problem is engaged with Natural Language Processing in the area Machine Learning. I used Bengali dataset which contains a large number of Bangla sentences with two types of emotion such as positive and negative. For that model I apply several machine learning algorithms such as K-nearest neighbors (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), Decision Tree (DT) and Logistic Regression (LR) to build my model.

3.2 Data Collection Procedure

All of data are manually collected from different online sources such as Facebook, YouTube and various online news portal. Here my collected data are mainly two types one is positive and another is negative emotion based on customer comments. In figure 3.2.1 we represent of our dataset.

Sentence	Emotion
অত্যন্ত ভালো সংবাদ	1
শেরপুরসহ দেশের বিভিন্ন স্থানে সরকারি দলের গুল্মা বাহিনী হামলা চালিয়েছে	0
যে সব অফিসাররা ভাল দায়িত্ব পালন করবে তাদের পুরস্কার দেওয়া দরকার	1
আমাদের প্রশাসনের সব সেক্টরই দুর্নীতির সঙ্গে জড়িত	0
আনেক ভাল লাগলো আমাদের দেশের এসব ইতিবাচক দিক দেখে	1
ফাস্ট ফুড আমাদের স্বাস্থ্যের জন্য বিপজ্জনক	0
দোয়া করি সার আপনি অনেক দিন বেচে থাকুন	1
খারাপ অভ্যাস কোন সুখ বয়ে আনে না	0
কোন হত্যা কান্ডের সুষ্ঠু বিচার বাংলাদেশে হচ্ছে না	0

Figure 3.2.1: Snippet of dataset

3.3 Statistical Analysis

In my Bengali dataset, there are total 4177 Bengali sentences. Among them around 2300 are holds positive (1) emotion and 1800 holds negative (0) emotion. In bellows figure: 3.3.1 which I represent.

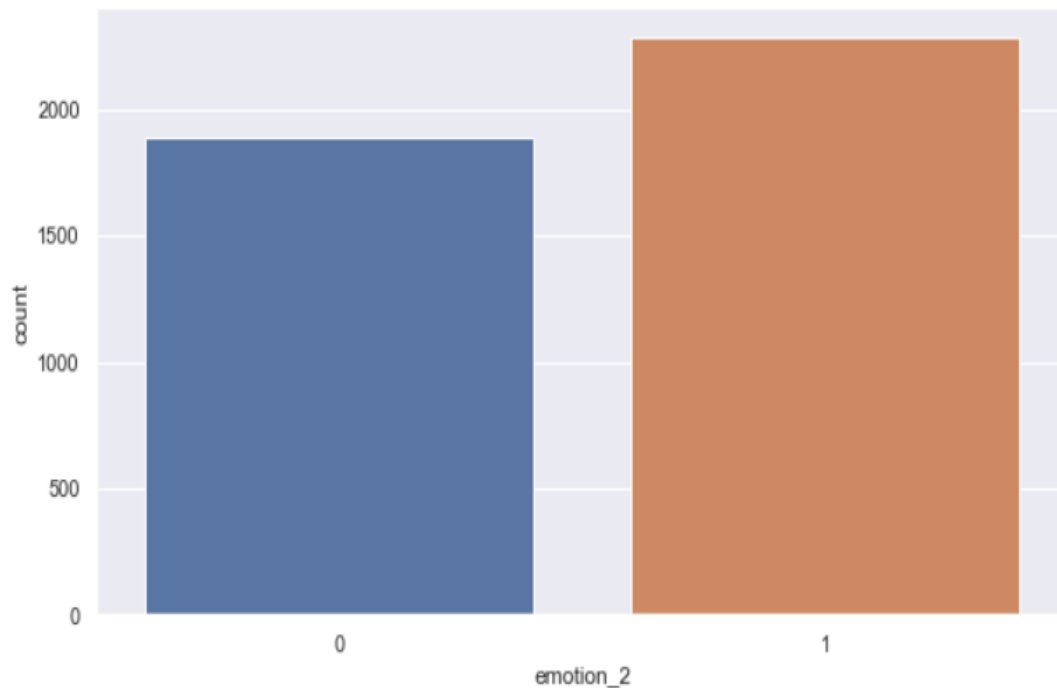


Figure 3.3.1: Quantitative distribution of emotion_2 column.

The sentence column is object type and the emotion_2 column is integer type which are the current state of my used dataset.

```
df.columns
```

```
Index(['sentence', 'emotion_2'], dtype='object')
```

```
df.dtypes
```

```
sentence    object  
emotion_2   int64  
dtype: object
```

Figure 3.3.2: Dataset object types.

3.4 Proposed Methodology

Several supervised machine learning algorithms are applied in my research work. First, I compare the accuracy which they are generate. Among those accuracy, I took highest one for to build my model.

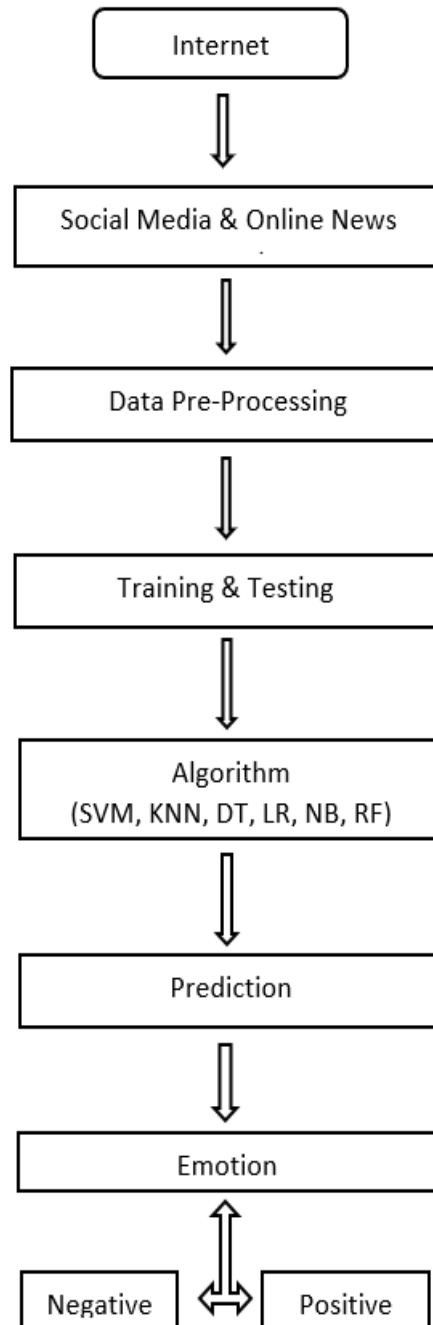


Figure 3.4.1: Activity flowchart of proposed methodology

I collect data from various social sites such as Facebook, YouTube, Twitter and online news portal such as Prothom Alo, Kalerkontho, Bangladesh Protidin etc.

All data are level with positive and negative sentiment where positive is denoted by one(1) and negative are denoted by (0).

	sentence	emotion_2
0	চলতি হিস্ট্রি পাতায় অবশ্য সকল তথ্য দৃশ্যমান থ...	0
1	সত্যি অসাধারণ একটি রিলেশন	1
2	এভাবে বারবার ক্লিক করে তথ্য মোছার কাজটা বিরক্...	0
3	খুব সুন্দর হয়েছে	1
4	অনেক হাসির একটা নাটক এবং রুম্যান্টিক	1
5	নিলামে তরঙ্গ বিক্রি না হলে কমপক্ষে পাঁচ হাজার ...	0
6	ভাললোই লাগল	1
7	উল্লিখিত সমস্যাগুলোর একটিও এখন পর্যন্ত সমাধান ...	0
8	বিউটিফুল	1
9	সিম প্রতিস্থাপন কর-সংক্রান্ত বিষয়টি সমাধান না...	0

Figure 3.4.4: Dataset with label

3.5 Implementation Requirement

For implement of my research work some steps are followed which are mention below. Raw data which collects from various online news portal are insert into a excel file with real human sentiment. Then check the null value in dataset and remove noise from dataset.

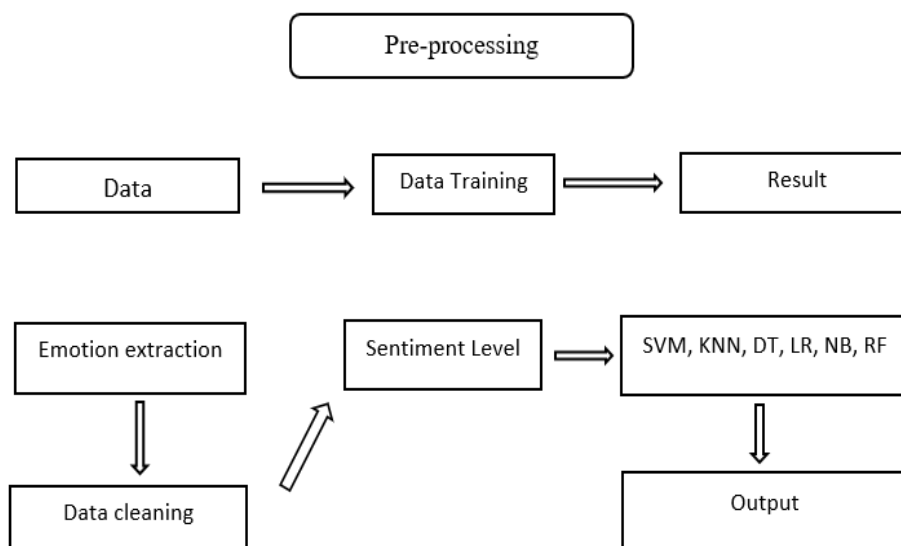


Figure 3.5.1: Model preprocess flow diagram

Sometimes human use the short form of a word. But machine don't understand its proper meaning. That time machine count these are noise. So here I try to remove short form of a word into full form using Bengali phrase tagger.

```

contractions = {
"বি.দ্র ": "বিশেষ দ্রষ্টব্য",
"ড. ": "ডক্টর",
"ডা. ": "ডাক্তার",
"ইঞ্জি: ": "ইঞ্জিনিয়ার",
"রেজি: ": "রেজিস্ট্রেশন",
"মি. ": "মিস্টার",
"মু. ": "মুহাম্মদ",
"মো. ": "মোহাম্মদ",
}

```

Figure 3.5.2: Base form transformation

Total number of data are level with only two sentiments positive (1) and negative (0). Among total number of data eighty five percent are used for training our model and remaining are testing. Positive and negative data are represented in the bellows picture.

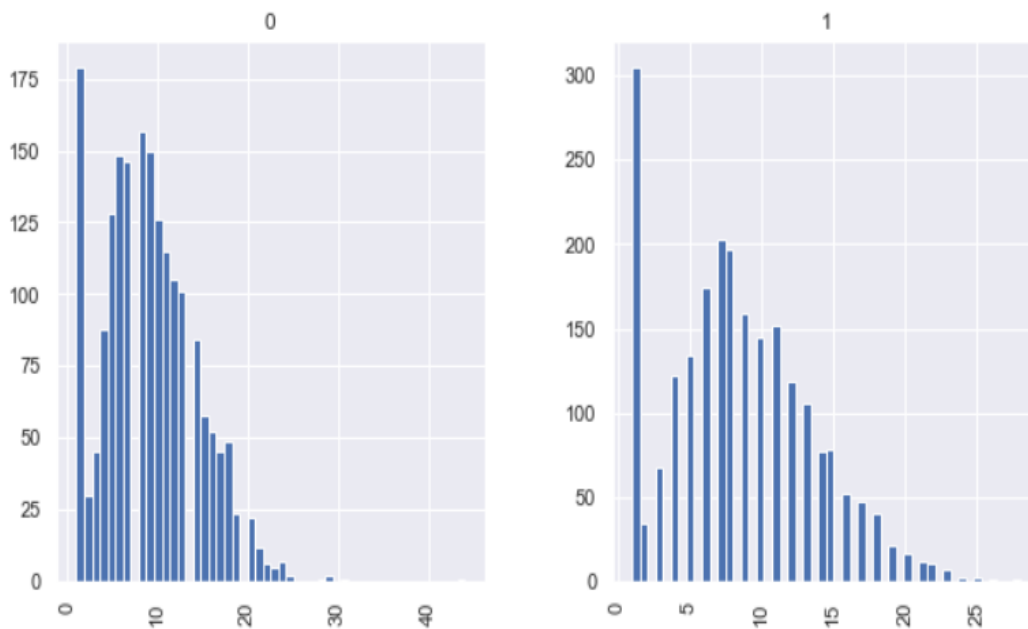


Figure 3.5.3: Graphical sentiment representation of dataset

For my model considering hardware a laptop or pc minimum 4gb ram and considering software Jupyter Notebook and also a lots of packages like NumPy ,CountVectorizer, Scikit-Learn, Seaborn, Matplotlib, and Pandas etc are required.

CHAPTER 4

Experimental Results and Discussion

4.1 Experimental Setup

4.1.1 Support Vector Machine

Support Vector Machine is used for classification problem. It is a supervised machine learning algorithm. It generates a hyperplane that partitioned n-dimension space into class. That's why the new sample is coming than its place it into most appropriate group.

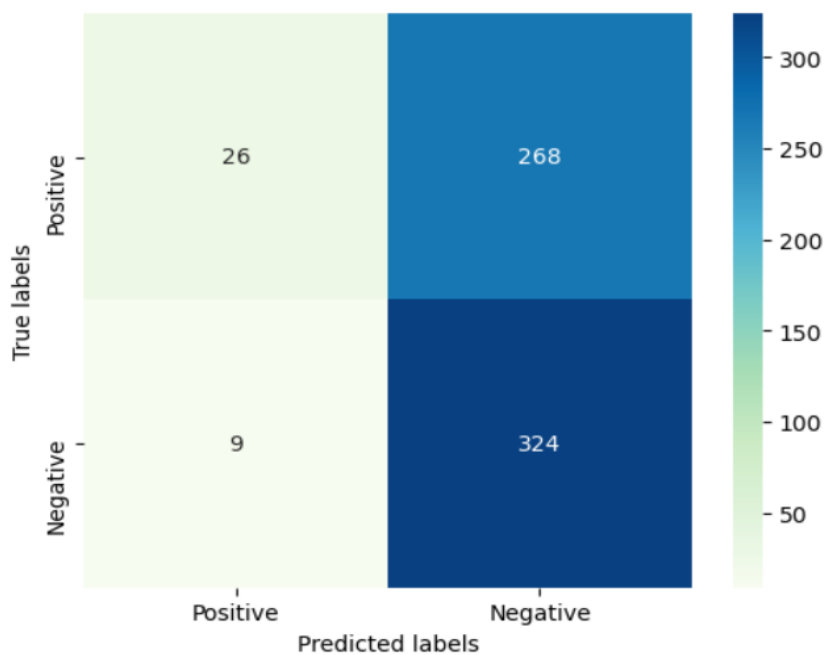


Figure 4.1.1: Confusion Matrix representation with SVC

In above confusion matrix we can see, $(26+324) = 350$ it generates true prediction and $(268+9) = 277$ generates false prediction out of total 627. Also I got accuracy using Support Vector Machine algorithm is 62.36 %.

4.1.2 K-Nearest Neighbor

K-Nearest Neighbor is one of the simplest supervised machine learning algorithms which is mostly used as a classification technique and sometimes for regression. It can category the sample data based on similar feature. If new data come then it put it most suitable group.

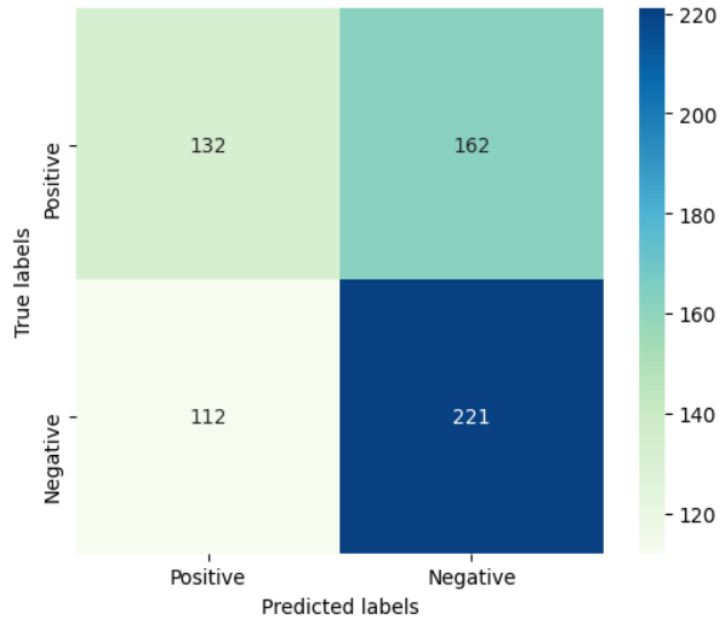


Figure 4.1.2: Confusion Matrix representation with KNN

Above confusion matrix we able to see that among total support 627, $(112+162) = 274$ produce false prediction and remaining gives us true prediction and the accuracy obtained from this algorithm is 56.29 %.

4.1.3 Decision Tree

Decision Tree is a supervised algorithm in the are of machine learning. It generates tree according to the sample and then classify the feature of an object.

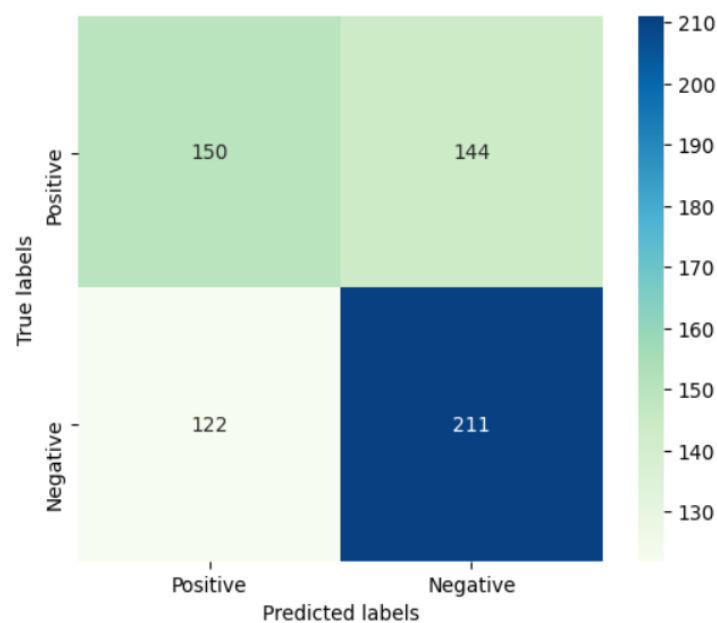


Figure 4.1.3: Confusion Matrix representation with DT

True prediction is $(150+211) = 362$ and false prediction is 266 compare with total 627. Accuracy is 58.05 %.

4.1.4 Logistic Regression

It is a supervised machine learning algorithm which deals at time both discrete and continuous variable. Both classification and regression problem it can be used.

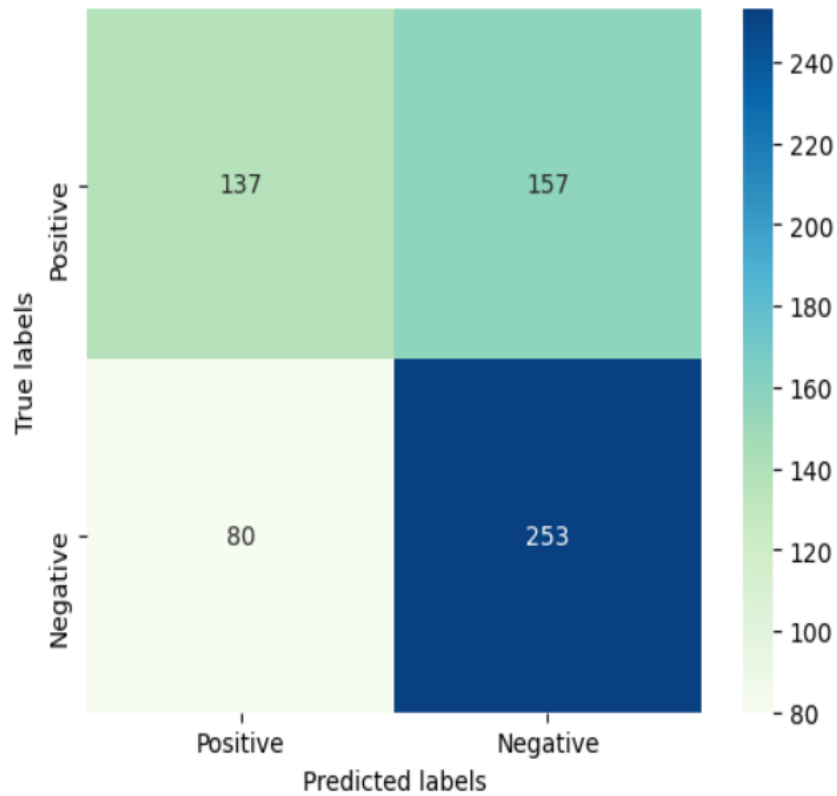


Figure 4.1.4: Confusion Matrix representation with LR

The amount of total false prediction from support 627 is $(80+157) = 237$ and remaining are true prediction that generates by our model and also the accuracy that this algorithm produce is 62.20 %.

4.1.5 Naïve Bayes

Naïve Bayes algorithm is constructed based on the theory of Bayes Theorem. Basically, it works depends on the probability concepts. It is also a supervised algorithm in machine learning technique which is capable to classify of an object depends on probability.

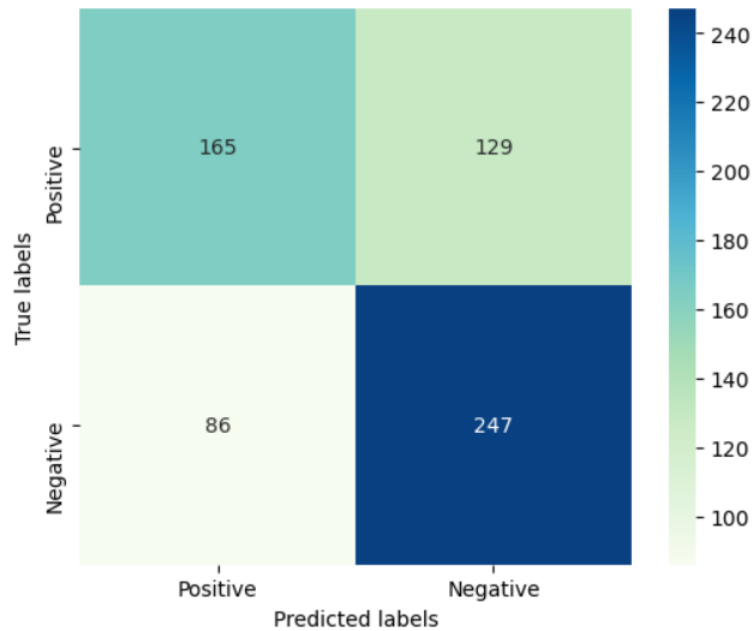


Figure 4.1.5: Confusion Matrix representation with NB

Here we can see true prediction is 412 and false prediction is 215 from 627 and also the accuracy obtained 65.70 %.

4.1.6 Random Forest

The main working procedure of Random Forest is to generate a number of decision trees and takes the average accuracy of the mentioned dataset. Though a much larger number of trees it produces, it combines them and predicts us a comparatively most accurate result. Another advantage of it is it consumes less training time so generates results first.

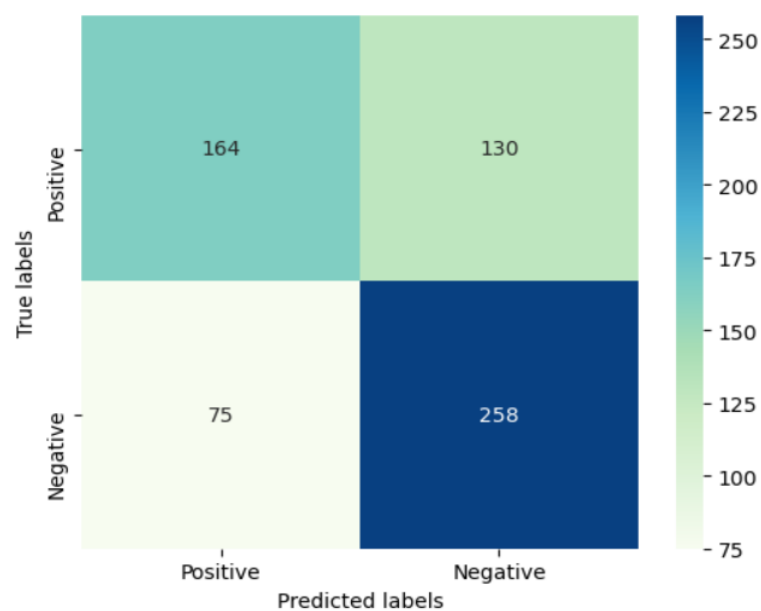


Figure 4.1.6: Confusion Matrix representation with RF

From the above confusion matrix, there are $164+258= 422$ correct predictions, and $130+75 = 205$ generate incorrect predictions out of total support 627. Maximum accuracy which I got using this algorithm is 67.30%.

4.2 Experimental Result & Analysis

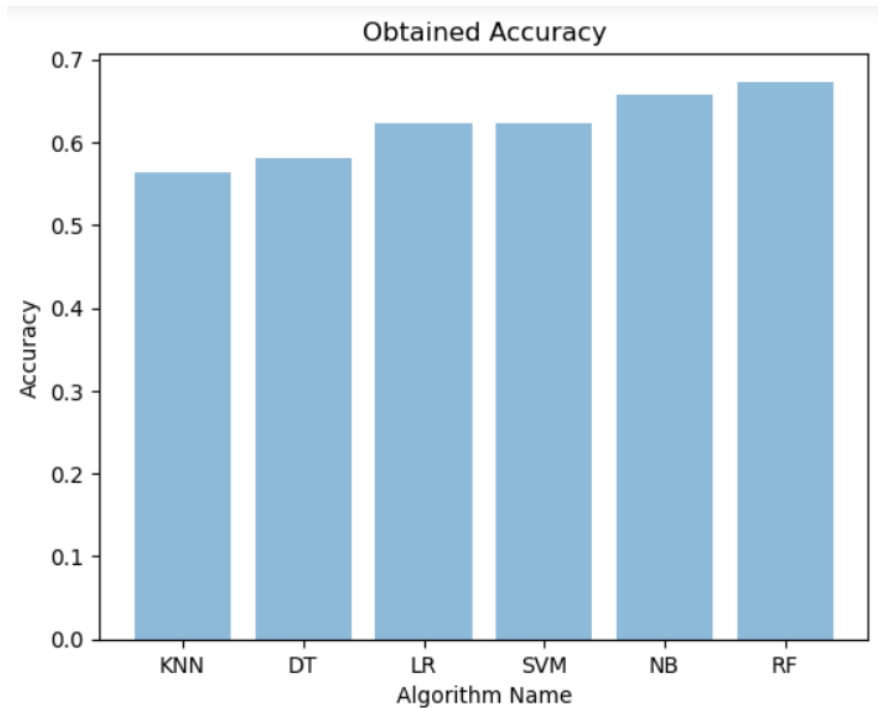


Figure 4.2.1: Accuracy of several machine-learning algorithms

Algorithm Name	KNN	DT	LR	SVC	NB	RF
Accuracy	56.29%	58.05%	62.20%	62.36%	65.70%	67.30%

Figure 4.2.2: Several algorithm result

All accuracy which we obtain using several machine learning algorithms are mention in Figure: 4.2.2.

```
In [97]: p=input()
art=vocab.transform([p])
s1=rmfr.predict(art)
if s1==0:
    print("Negative Sectence")
else:
    print("Positive Sectence")
```

অসাধারণ একটা গল্প অনেক ভালো লাগলো
Positive Sectence

```
In [106]: p=input()
art=vocab.transform([p])
s1=rmfr.predict(art)
if s1==0:
    print("Negative Secture")
else:
    print("Positive Secture")
```

ততক্ষণে ভয়ে থরথারয়ে কাঁপতে শুরু করেছেন সোহেল, গা কাঁপিয়ে এসে গিয়েছে জ্বর
Negative Secture

Figure 4.2.3: Suggest sentiment

4.3 Discussion

The experimental output already shows with result table and graph. Six supervised machine learning algorithm I used. They are K-Nearest Neighbor, Decision Tree, Logistic Regression, Support Vector Machine, Naïve Bayes Classifier and also Random Forest algorithm. Among them Random Forest generates highest accuracy. That's why I select this algorithm for my model to get best accuracy. After testing my model by random input my model is capable to classify the sentences with correct sentiment.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Day by day a large amount of population in our society are influenced rapidly by several social media and online news portal due to the cheaper rate of internet or cope up the modern world. Now data is enormous but it is not suitable to process by human brain. For that machine learning is so much suitable to process the large amount of data at instantly that helps our lifestyle so much easy.

5.2 Impact on Environment

For knowing the future upcoming pattern machine learning technique is so favorable. Because it can analyze the several previous year data and can produce a pattern which will be happen in future such as business strategy, culture types, human capability, weather condition which makes our life more comfortable. So, it is not excess to say that without machine learning life will may be difficult.

5.3 Ethical Aspects

We are all confined to state law and should follow this. Without any permission we should not try to collect another person's data of any types and not process this data to determine the gist of his daily activities. Sometimes it would be a reason to get unhappy of a person due to social harassment. So, we must ensure the public privacy and at a time to work more and more to makes our future more comfortable by using machine learning technique.

5.4 Sustainability plan

Due to huge amount of data around us, it is very tough for human to process and find out a pattern. But it is comparatively easy for a machine by using several machine learning techniques in the area of artificial intelligence. Machine can predict the future events which will may be happen. It helps us in business sector such as the owner of the business realize the future demand of a specific product that customer wants. Another advantage is it is capable to predict the weather condition in future also predict the economic condition of a country. So, I think it play's major role for sustainable development of a country.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Summary of the Study

My research main target is to detect the emotion of a Bengali context either it has positive or negative meaning. For that I use a Bengali dataset which is manually crafted by person realizing human level real emotion from several social and online sites. This dataset is trained by six supervised machine learning algorithms. Then compare the accuracy and choose algorithm with highest accuracy to build my intelligent system. After all my system is capable to classify the Bengali content either it positive or negative by Random Forest algorithm.

6.2 Conclusions

Now machine learning is an important part of modern technology. It takes away our life into comfort zone. Instantly human can now take a decision by the help of machine learning which is the part of an artificial intelligence. All types of native speaker are capable to process their language with machine learning technology. Here I build a system which deal with Bengali language and gives us the decision instantly. My intelligent model is trained with total six supervised machine learning algorithm. Among those which generates highest accuracy I used this algorithm to build my model and now it acts a full intelligent system which can classify the Bengali sentences or word based on sentiment.

6.3 Limitation

Sentiment classification by human of their native language has much limitation. Sometimes the real emotion is not realized of a human. Because the sentiment of speeches may vary in several situation. It also varies considering time to time and also political, social and economic perspective. Frequently based on culture of a country same word has different meaning which is sometimes unrealizable for human. Since this dataset is created by human manually so it may have some noise. For noisy data it will be difficult to deal for a system.

6.4 Implication for Further Study

In my research work, I have only considered two types of emotion positive and

negative. But in my feature work I will try to level my data with more accurate emotion such as strongly positive, mildly positive, positive, neutral, strongly negative, mildly negative, negative etc. Sometimes we can see there are many types of emoji already exists in different social media which express a clear emotion. But in my data set now it is considered noise and I remove it. Though it has clear emotion so next time I will be work on it. At this time for sentiment analysis so much updated platform is Deep Learning method. In future my next goal is to work on sentiment classification using Deep Learning.


References

- [1] Bengali Language. [Online] Available at: https://en.wikipedia.org/wiki/Bengali_language
- [2] Awesome Bangla [Online] Available at: <https://github.com/banglakit/awesome-bangla>
- [3] Nasukawa, T. and Yi, J. 2003. Sentiment analysis: capturing favorability using natural language processing. In Proceedings of the 2nd international conference on Knowledge capture (KCAP '03). Association for Computing Machinery, New York, NY, USA, 70–77.
DOI: <https://doi.org/10.1145/945645.945658>.
- [4] Whitelaw, Casey & Garg, Navendu & Argamon, Shlomo. (2005). Using appraisal groups for sentiment analysis. International Conference on Information and Knowledge Management, Proceedings. 10.1145/1099554.1099714. DOI: <https://doi.org/10.1145/1099554.1099714>
- [5] John C. Platt. 1999. Fast training of support vector machines using sequential minimal optimization. Advances in kernel methods: support vector learning. MIT Press, Cambridge, MA, USA, 185–208
- [6] Hutto, C.J. & Gilbert, Eric. (2015). VADER: A Parsimonious Released Model for Sentiment Analysis of Social Media Text. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014
- [7] Dos Santos, Cicero, and Maira Gatti. "Deep convolutional neural networks for sentiment analysis of short texts." Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014
- [8] Howard, Jeremy, and Sebastian Ruder. "Universal language model fine-tuning for text classification." arXiv preprint arXiv:1801.06146 (2018)
- [9] Tang, Duyu, Bing Qin, and Ting Liu. "Document modeling with gated recurrent neural network for sentiment classification." Proceedings of the 2015 conference on empirical methods in natural language processing. 2015
- [10] Yelp Open Dataset [Online] Available at: <https://www.yelp.com/dataset>
- [11] IMDb Data Files [Online] Available at: <https://datasets.imdbws.com/>
- [12] Das, A. and Bandyopadhyay, S., 2010. Phrase-level polarity identification for Bangla. Int. J. Comput. Linguist. Appl. (IJCLA), 1(1-2), pp.169-182
- [13] Weston, J., 2014. Support vector machine. Tutorial, http://www.cs.columbia.edu/kathy/cs4701/documents/jason_svm_tutorial.pdf, accessed, 10(0), pp.0-5
- [14] 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
- [15] 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.
- [16] Ghosal, T., Das, S.K. and Bhattacharjee, S., 2015, December. Sentiment analysis on (Bengali horoscope) corpus. In 2015 Annual IEEE India Conference (INDICON) (pp. 1-6). IEEE
- [17] M. Al-Amin, M. S. Islam and S. D. Uzzal, "Sentiment analysis of Bengali comments with Word2Vec and sentiment information of words," in International Conference on Electrical, Computer and Communication Engineering, Cox's Bazar, Bangladesh, 2017 describes about our proposed work flow for this experiment

PLAGIARISM REPORT

Plagiarism Checked by

Abdus Sattar, Assistant Professor, Department of CSE



26-12-2021

SENTENCE LEVEL SENTIMENT CLASSIFICATION

ORIGINALITY REPORT

7 %

SIMILARITY INDEX

7 %

INTERNET SOURCES

1 %

PUBLICATIONS

4 %

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Daffodil International University

Student Paper

2 %

2

dspace.daffodilvarsity.edu.bd:8080

Internet Source

2 %

3

arxiv.org

Internet Source

<1 %

4

doctorpenguin.com

Internet Source

<1 %

5

mljar.com

Internet Source

<1 %

6

studentsrepo.um.edu.my

Internet Source

<1 %

7

ieomsociety.org

Internet Source

<1 %

8

ebin.pub

Internet Source

<1 %

9

worldwidescience.org

Internet Source

<1 %