Sentiment Classification in Bengali Linguistics Using Directed Machine Learning Techniques

BY

SHAH AL SHIHAB
ID: 183-15-11939

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

Supervised By

Shah Md. Tanvir Siddiquee
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

September 2022

APPROVAL

This Project titled "Sentiment Classification in Bengali Linguistics Using Directed Machine Learning Techniques", submitted by "Shah Al Shihab" to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 13 September 2022.

BOARD OF EXAMINERS

14/9/22

Chairman

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

Tania Khat
13-09-22

Internal Examiner

Tania Khatun (TK)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Mohammad Monirul Islam(MMI)
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Farid

External Examiner

Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

I hereby declare that, this project has been done by me under the supervision of **Shah Md. Tanvir Siddiquee**, Assistant Professor, **Dept. of CSE**, Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Shah Md. Tanvir Siddiquee
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Shihab

Shah Al Shihab
ID: 183-15-11939
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First and foremost, I appreciate God for guiding me down the path to earning an honorable B.Sc. in Computer Science & Engineering from Daffodil International University. Then I'll always be grateful to my parents for their love and support.

I'd like to express my gratitude to my supervisor for providing me with the necessary guidance and advice in order to accomplish this fantastic research project on sentiment classification in Bengali linguistics using directed machine learning techniques. His encouragement and guidance gave me the confidence I needed to accomplish my research assignment correctly. He provided me with all of the necessary materials and knowledge to begin this investigation from beginning. I'd want to express my gratitude to my coworkers for their assistance in shaping the dataset and other relevant duties.

ABSTRACT

Sentimental evaluating is part of NLP research (SA). This dataset was created by scraping information from social media. The contents are also carefully categorized into positive and negative categories. Another name for this is polarity categorization. Emojis are becoming more widely utilized in written communication to express emotions or to repeat statements. Prior AI systems just looked at the arrangement of text, emojis, or pictures, with emojis with text continuously being disregarded, resulting in a variety of feelings being missed. A composite process technique of the "Pipeline" class is used to extract features and train the dataset, which incorporates Count-Vectorizer, transformer, and machine learning classifiers. In terms of accuracy, RF surpassed the other five classifiers. Even though LSVM has the lowest accuracy, it is also nevertheless gives excellent results. However, for current and critical linguistic data, this work has produced superior results, suggesting that adequate feature extraction was used to develop the model. In case of using different classifiers, my dataset was thoroughly concealed and cleansed for the models training, as i can see in some cases that the external symbols may prove to require difficulties while the analysis via models conjugation. The model is mainly based on the method of obtaining the accuracy via the UNIGRAM, BIGRAM and TRIGRAM structure. Which will be later be described through the article. My work was mainly to determine the two binary classes of positive and negative sentiment structure. I am quite satisfied with my achieved result by going through the related research works of such.

TABLE OF CONTENTS

CONTENTS **PAGE**

CHAPTER

LIST OF FIGURES

LIST OF TABLES

Chapter 1
Introduction

1.1 Introduction

Computers are more enticing than Aladdin's magical lamp to humans who have sophisticated technology. The goal of human efforts is to make computers capable of recognizing natural language mechanisms in the same way that people do. One of the most regardful issue in the research sector called natural language processing contains issues like these (NLP). Sentimental analysis (SA) can be labeled as one of the mentionable important and influential topics of research in this era of information artefacts.

In the previous year, many NLP re- searchers have been developed to find attributes from text, sentiment, or subjectivity detection, as well as subject or context categorization Researchers can use Sentiment Analysis to assess if a scenario is favorable, bad, or neutral (SA). Sentiment analysis is the process of acquiring information from public material in order to create people's attitudes, expressions, and thoughts regarding genuine products, information, issues, or forum debates. (i.e., political, cricket) [1]. Manually combing through this enormous stack of documents and uncovering articulated concepts through meticulous data categorization can be laborious and time-consuming [2]. This is a massive problem area. When it comes to ambiguous statements, where positive words might indicate negative meaning or vice versa, a significant case develops. Firms used to spend a lot of time analyzing market demand or current trends, which might be frustrating at times. Bangla content is generated in great amounts in this modern age, with intellectuals', common people, and worldwide happenings blasting our thoughts and cognitive processes. Such occurrences characterize people's perceptions and ideas. In many sectors of human life, the dole of SA has now surpassed the dole of SA, including marketing and customer service in enterprises, social media monitoring, political opinion analysis, and many others. The most advanced sentiment analysis research has been done on text collected from social media platforms using machine learning (ML) algorithms. Regardless, SA on both text and emoji has been mostly ignored due to a paucity of assets and the intricacy of emoticons. Text categorization is one of the most fascinating fields to research since it eliminates sentiment utilizing a variety of

ML and DL techniques, as well as later advances. According to studies, deep learning was infrequently applied in sentiment analysis on both emoji and text data. As a result, the material and emojis were combined in this inquiry to find the notions. Similarly, the study created an emoji vocabulary and assessed the results using a deep learning system that combined emoji vocabularies with content highlights including phrase frequency measurement, inverse document analysis, N-gram, and bag of words. This paper uses a unique and effective strategy. Algorithm with rules The Bangla Text Sentiment Score (BTSC) was created to detect sentence polarity and enhance emotion extraction by scoring a piece of Bangla text an enlarged Bangla emotional dictionary with weighted value will be used to assess the automated system, and the automated system will be categorized using a supervised machine learning technique. This is because the author realized that this method was effective for text categorization. The pattern and behavior of the dataset, as well as the underlying justifications for such a reaction, will be revealed through comparative analysis utilizing different classifiers.

1.2 Motivation

Whenever it concerns to Narrative research, scientists have long been fascinated by the ability to develop and work on languages in order to understand and predict human behavior. This requirement can only be understood through language. That is why I wanted to focus on this study genre for future advancement. In terms of sentiment detection, sentimental determination towards diverse phrases is a very natural and ongoing strategy nowadays. I can forecast a liberal approach for those phrases by using machine learning to discriminate between bad and positive reviews. Which is really my major incentive for doing this work.

1.3 Research Questions

- What precisely is virtual communities??
- What is Sentiment classification?
- What is NLP?
- Why NLP is important?

1.4 Expected Output

The predicted result was that, by applying the model, which consists of many different algorithms that are to be trained and then evaluated by the dataset that comprises linguistic Bengali text data obtained from multiple social platforms, the model should be able to determine which are negative attitudes and which are positive. When working with text data, the algorithms are typically employed. As we all know, text data in Bengali may be difficult to deal with for true machine learning models; in this instance, the dataset was adequately preprocessed so that the output does not vary depending on the circumstances, which are the trash values.

1.5 Report Layout

The report has six chapters. Every chapter describes the different aspects of the "**Sentiment Classification**". Every chapter has different parts described in detail.

Chapter 1: Introduction

The inspiration is clarified and the proposition objective and introduction are presented.

Chapter 2: Background Studies

The applicable work is talked about and significant popular techniques are introduced corresponding related work.

Chapter 3: Research Methodology

Presents the information assortment, information pre-handling, and the element determination methodology.

Chapter 4: EXPERIMENTAL RESULTS AND DISCUSSION

the philosophies for assessment grouping are clarified and the result discussed.

Chapter 5: IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

The 3-assessment plan, the precision assessment, and the investigation are introduced.

Chapter 6: Conclusion and Future Scope

The end is drawn and my commitments are portrayed.

Chapter 2
Background Studies

2.1 Introduction

In this section, we use taxonomy to lay out the available techniques, which investigates the impacts on various DL designs and discusses how such methods improve their capacity to function in South Africa in the multidisciplinary realm of NLP, sentiment classification was depicted as [3], and a relationship of in the detection of subjectivity and classification of polarity was described. The authors of [4] created a probabilistic model for learning a sequential representation of words while also applying a probabilistic function to learn the sequences of words. [4] describes a simple one-layer-based CNN indication for undertaking text sensitivity analysis. Initially, a classification system was established to track occurrences based on the content of a posting. After then, an effective clustering-based technique was employed to detect and monitor the events, and a memory component was used to retain them. Aldhaheri et al. [5] suggested a novel neural network-based technique for event detection. This work has been broken into five sections that have been arranged in a logical order. Here is a literature review pertaining to our research subject, and it begins by presenting a summary of previous work on SA in other languages, as well as research conducted in Bangla. The methodology utilized in this work is discussed, which includes a discussion of the dataset, data pre-processing techniques, training and testing datasets, and feature extraction algorithms. The experimental data, as well as performance measurements and graphical and tabular representations of the results, are presented. It contains the conclusion and future work, which highlights the contributions of the intended study and suggests some potential next steps.

2.2 Related Work

Among the most frequently used machine learning applications in recent years has been outcome forecasting. These studies concentrated on particular challenges and made use of a range of machine learning approaches ways to deal with them. This paragraph highlights the Many of the professionals in the earlier area took effectively concluded. Sentiment Analysis (SA) includes several a collection of ideas, feelings, and literary subjectivities. In SA research, there are two types of techniques: (a) lexicon- or corpus-based approaches, and (b) ML-based approaches. A dictionary is constructed manually or automatically using a lexicon or corpus-based technique that takes document orientation, words, and phrases into account [6]. This procedure is commonly used in the explained models, and other

classification techniques. On the other hand, the ML methodology is one of the most widely used and advanced methods for categorization. Multilayer Perceptron (MLP), Support Vector Machine (SVM), Multinomial Naive Bayes (MNB), Maximum Entropy (ME), and a range of other basic classifiers have all been utilized successfully [7]. [22] analyzes the statement's positively or negatively meaning using a machine learning algorithm to determine mood in Bangla. The SVM classification method's advantage in topic-based categorization is shown in [23]. There are limited contributions to analyzing the feelings of such under-resourced languages. Thapa and Bal [8] observed the use of supervised ML classifiers on a sample of three hundred and eighty four book and movie reviews in a research study on Nepali language. The results of extracting features using the TF-IDF and Bag-of-Words (BOW) approaches demonstrated that MNB beat SVM and LR. In a 2019 study, Tabassum and Khan [9] collected one thousand and fifty Bangla texts from Facebook and Twitter. The dataset was classified into positive and negative attitudes using an RF classifier and techniques such as unigram, pos tagging, and negation, with an accuracy of 87 percent. Machine learning classifiers have shown potential in extracting feelings in other languages, such as Bangla, according to previous research. As a result, our suggested model will use a pipeline technique that incorporates vectorizers, TF-IDF approaches, and ML classifier modeling to analyze attitudes from Bangla news datasets. Another LSTM-based technique for discriminating polarity positive and negative sentiments was tested on 9337 reviews and found to be 78 percent accurate [10]. [11] extracted six categories of emotion from various types of Bangla YouTube video comments using a CNN and LSTM-based approach, with 65.97 percent and 54.24 percent accuracy on three and five labels sentiment, respectively. On other domains from the Bangla dataset, a CNN-based single channel approach [12] is utilized, however it is impossible to maintain optimal layer tuning. For a product evaluation, another project has been completed. Those who built the different classifiers of Decision Tree, Logistic Regression, KNN, SVM, and Random Forest using 1000 feedback mechanisms. SVM performed with the maximum accuracy, which was 88.81% [17]. Hossain et al. presented a machine learning-based method to discriminate between favorable and unfavorable novels. On 2000 reviews of Bengali literature, they used LR, NB, SVM, and SGD ML algorithms. They found an increased accuracy of 84 percent during using multinomial Naive Bayes [18]. A strategy to recognize abusive comments, which are particularly detrimental to young people, was put out by Maliha Jahan et al. [15]. On the dataset, which they built by collecting input from public Facebook pages, they tried machine learning techniques such as Support Vector Random Forest, Machine, and Adaboost, and they acquired the highest accuracy of all of them, 72.14 percent. Hindi language analysis presented by Mittal et al. gives positive and negative validity of 82.89 percent and 76.59 percent, respectively. To broaden the database's coverage and enhance uniformity, they chose to evaluate emotions. This article outlines an educational program that examines the emotions of Roman Urdu speakers using the sports, software, food and recipe, theater, and political genres. It has 10,021 sentences that were taken from 566 online conversations. The site's main objectives were divided into two phases: (1) creating a Roman Urdu corpus with human annotations for emotionally assessment; and (2) assessing methods for emotion analysis was based around Rule-based, N-gram (RCNN) models [24]. Md Gulzar Hussain and colleagues [16] examined Bangla texts to find offensive Bengali remarks that had been gathered from various social media. To get better outcomes, they suggested using unigram string methods and a root level approach to identify offensive remarks. The efficiency of text classifiers to detect child abuse in chat was investigated by Md Waliur Rahman Miah et al. [20]. They used Classification-via-Regression, Decision trees, and Naive Bayes classifier on the data set that was gathered from various websites. To identify potential child abuse in the data set, Chintan Amrit et al. [21] used machine learning and text mining approaches. The machine learning models they are using are Random Forest and Support Vector Machine, and they received the highest score from AUC-metric. Two different methods were proposed by Tuhin et al. for identifying and classifying feelings in numerous aspects from all across Bangladesh. These were filled with a variety of feelings, such as euphoria, rage, sadness, dread, excitement, and sensitivity. In Naive Bayes, there are two strategies: the relevant answer and the grouping technique. The accuracy rating for the 7400 words from Bangladesh utilized in the collection was 90%. After that, they compared their work to two others who both earned SVM grades of 93% and documentation consistency ratings of 83%. Every article has a unique emotional quality [19].

An RNN network is a type of network. The Bi-LSTM approach was used on a manually labeled dataset of 10000 Facebook comments, and the accuracy was 85.67 percent; nonetheless, it has multiple noteworthy flaws in data preparation. [13]. When making an online purchase, it is essential to comprehend the needs of the buyer, but sometimes firms fall short of this need. In order to validate their assessments, C. Chauhan et al. used a machine learning system to distinguish between positively and negatively comments from potential customers. They studied various publications and found that Naive Bayes gave good results, although the results varied according to the circumstance, the strategy, and the goals [14].

2.3 Research Summary

The experiment's dataset was acquired at random from various websites and platforms on the internet. Extraneous data, numeric values, and special characters were removed from the dataset before to detection in order to produce a full and accurate detection result. The dataset had several repeat occurrences of various numbers and words, which were thoroughly evaluated and deleted for performance reasons.

2.4 Scope of the problem

This study offers and discusses exams in a presumption investigation of Twitter presentations pertaining to US carrier organizations. The purpose of this investigation is to determine whether tweets may be classified as good, negative, or neutral. Through informal communication sites, clients may discuss and exchange their facts, opinions, and speculations. Carrier tweets are now well known and are being proclaimed as a dataset assess customer concerns. In this study, the experts built a model by grouping extremity through explanation using classifiers KNN, Logistic Regression, and Random Forest. The experts gathered tweets from neighboring planes about their participation with the administrations. To better understand the trial's outcome, the analysts decided if a tweet was good, neutral, or negative and gave quantitative and subjective assessments, as well as conclusion examination. Individuals would be one step ahead of dynamics and automation in this field of observation.

2.5 Challenges

In this study, I considered the slant assessment based on voyager feedback in regards to carrier organizations. My proposed method indicated that both element determination and over-inspection techniques are equally important in terms of improving our results. The use of highlight selection algorithms has recovered the best subset of highlights and reduced the calculations required to create our classifiers. However, it has reduced the skewed appropriation of classes present in a major part of our smaller datasets without creating overfitting. My findings provide persuasive evidence that the suggested model has excellent grouping precision in forecasting events structure the two groups positive, negative data. It was also a difficult challenge to organize Bengali text and process it for model training.

Chapter 3

Research Methodology

3.1 Introduction

In this section, I will briefly outline the process I used to finish my study job. To disguise the complex machine learning techniques, I donned several Python programming languages. Because the dataset is the most important aspect of the Natural Language Processing method, the algorithms were chosen and placed on my dataset frequency. These are all advanced machine learning approaches that produced the desired outcome. The purpose of my research was to find an uncommon rule placed algorithm. Figure 1 depicts the technique of visualizing the procedure using a flowchart

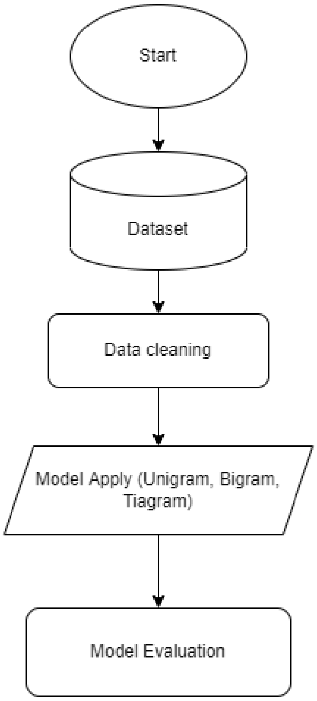


Figure 1: Method Directional Flow Chart

As in the flow chart described at first, I collected a genuine amount of data considering to train and test my model in purpose of our experimentation. And then the dataset was prepared for this propaganda. By preparing I have thoroughly described a small after in this article of how did I cleaned the dataset and got rid of unnecessary elements for my model performance and result's sake. Now in this section I will describe about the model main parts of implementation by describing the Unigram, Bigram and Trigram. And then the model evaluation is described in the Experimental result and Discussion section in afterwards.

Although before I explain the detailed structural and equations concealed of the n grams, we would like to explain the reason I chose this n-gram based model then typical machine learning existing models, The main reason in such case is that of its flexibility and adaptability. As we know that Bengali NLP can be proven to be quite difficult to execute than certain other NLP languages. I chose this method to be quite adaptative towards my dataset performative concertation.

N-Gram

For n = 1, the n-gram model is referred to as a unigram model. Bi-gram is defined similarly for n = 2, and in the case of tri-gram is n = 3. This straightforward statistical tagging technique is called the unigram (n-gram, n = 1) tagger. It provides the label which most closely matches the text within each token. Because the term often is more frequently employed as an adjective (e.g., a common word) than like a verb, it will, for instance, label all instances of a word with the letter. A unigram pos tagging should be taught on a training data before use to tag data. The much more popular tags for every phrase are identified using the database. Any token not found in the training examples will be given the preset tag Nothing by the unigram tagger.

Let S(wi,wj) provide the proportion of prefix similarity. After a comparison of the prefixes of two distinct nouns, the total compensation yields this proportion.

S (wi,,wj) = $length(matchPrefix(\boldsymbol{wi},\boldsymbol{wj}))$ / $Min(length(\boldsymbol{wi}),length(\boldsymbol{wj}))$ × 100

We initially compute the overall amount of words shared between such a set of phrases' listings for their preceding and succeeding phrases in order to determine how comparable their contexts are.

$matchPrv = count(matchPreviousList(\boldsymbol{wi},\ \boldsymbol{wj}))$

[N-gram Figure]

Applying N-gram Feature (সবাই কে অসংখ্য ধন্যবাদ)

Table 1: Differentiate between n-gram feature on a Bengali text

Unigram	Bigram	Trigram
সবাই কে অসংখ্য ধন্যবাদ	সবাই কে অসংখ্য ধন্যবাদ	সবাই কে অসংখ্য কে অসংখ্য ধন্যবাদ

3.2 Research Subject and Instrumentation

My selected title is " Sentiment Classification in Bengali Linguistics Using Directed Machine Learning Techniques ". This is a key research sense in Natural Language Processing. To date, I have inspected the way toward doing estimating investigations in Bangla using the defined and theoretical approach. A remarkable learning model necessitates a high structure computer and many instruments. The example of an idea analyzer is given beneath the primary instrument for this model.

Hardware and Software:

- 4GB RAM and Intel core i7 9th generation.
- 1 TB Hard Disk.

Tools:

- Windows 10
- Python 3.8
- Jupyter Notebook
- NLTK

- Pandas

- Numpy

3.3 Data Collection

My gathered amount of a certain dataset comprises 16211 data's that make up sentences. The lines were culled from a variety of social media networks, including Facebook, Twitter, and Instagram. The dataset was separated into two categories: Positive and Negative. The dataset distribution is depicted here based on class differences.

Table 2: Class information

Class Name	Amount
"0" or Positive Data	7939
"1" or Negative Data	8272

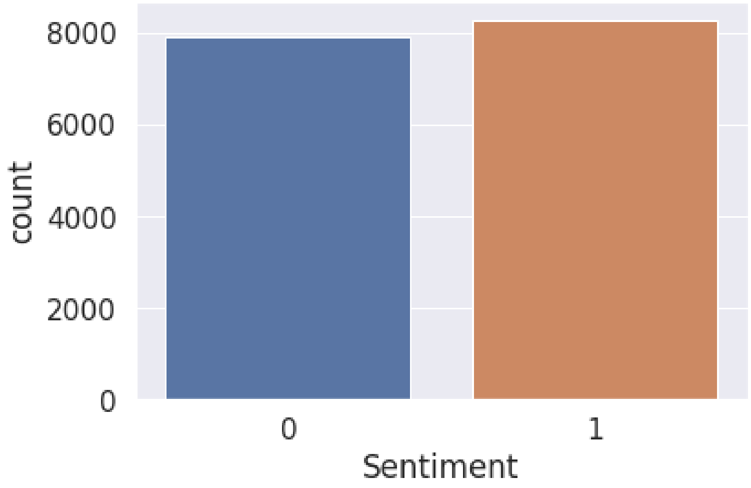


Figure 2: Data statistics Diagram

3.4 Dataset Distribution

After all of the cleaning and clearing of untamed data, the dataset was finally ready to be divided into classes for train and test. The whole amount of data was divided in an 90/10 proportion. The train data size was 14589, and the test data size was 1622.

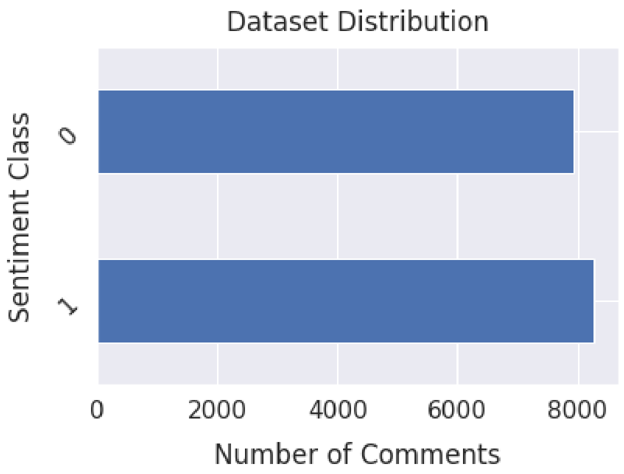


Figure 3: Dataset Distribution based on sentiment class

3.5 Preprocessing

To prepare the dataset for the train case of the algorithms, I required to refresh the entire dataset by eliminating the junk character. What are special characters ("!", "@", "#", "$", "percentage ", "", "*") number character (1, 2, 3, 4, 5, 6, 7, 8, 9, 0) English alphabets (A to Z both capital and small letter style), white space, and duplicate character? The duplicate character might possibly be kept since the character a certain was reconciled repeatedly when accepting the data set. In order for the computer to recognize the difference between classes, the dataset must be raw.

3.6 Stop word remove

The stop word can be a commonly used term, in case of example, ".", ",", "", "|", etc., that a web index has been configured to ignore, both while sorting parts for viewing and when recovering them as the result of a pursue enquiry. I wouldn't need these terms to take up space in our database or take up substantial handling time. I can effectively evacuate them for this by securing a list of terms that we believe to be stop words. Python's NLTK (Natural Language Toolkit) provides a list of stop words stored in 16 different dialects. We can find them in the NLTK data index.

3.7 Tokenization

Tokenization is the process of breaking down the argument in relation to the sentence, and these single arguments are referred to as tokens. Tokenization is essential in such procedures for training the input to the algorithm. In the instance of labeling, the dataset was divided into two classes: positive and negative.

3.8 Statistical Analysis

1. In the dataset total 16211 data is presented.
2. The dataset is divided into 2 classes.
3. 14589 data is used for the train.
4. 1622 data is used for the test.
5. Highest accuracy achieved by Unigram 67%, Bigram 68%, Trigram 68%.

3.9 Implementation Requirements

Python was the programming language I used to create the machine learning model. Panda's library is used for loading the dataset, while the NLTK library is utilized for preprocessing. The entire implementation is built in Google Colab.

Chapter 4

Experimental Results and Discussion

4.1 Introduction

In the instance of my identification of negative and positive outcomes, the algorithms were able to achieve a very volatile and appreciating level of result. The graph below depicts the Unigram, Bigram, and Trigram of sentiment detection by our chosen methods. The methods were chosen and placed on our dataset frequency because, in Natural Language Processing, the dataset is the most important aspect of the entire operation. The algorithms we used were Linear Regression, Decision Tree, Random Forest, Multinomial Naive Bayes, KNN, Linear SVM, and RBF SVM. These are all advanced machine learning approaches that produced the desired outcome.

4.2 Model Performance

In terms of the unigram feature, MNB earned the maximum accuracy of 67.45 percent. Linear SVM earned the greatest F1-score of 69.77. The LR obtained the greatest precision score of 68.27, while the KNN achieved the highest recall score of 94.84. The MNB earned the greatest accuracy for Bigram performance characteristics at 68 percent. RBF SVM also earned the highest F1 score of 72.46. The MNB had the greatest accuracy score of 69, while the KNN had the best recall score of 98.56. In terms of the final Trigram feature for my study, the MNB earned the maximum accuracy at 68.55 percent. The Linear SVM obtained the greatest accuracy score of 87, as well as the highest F1 score of 69.22. The MNB also had the highest recall score of 91.

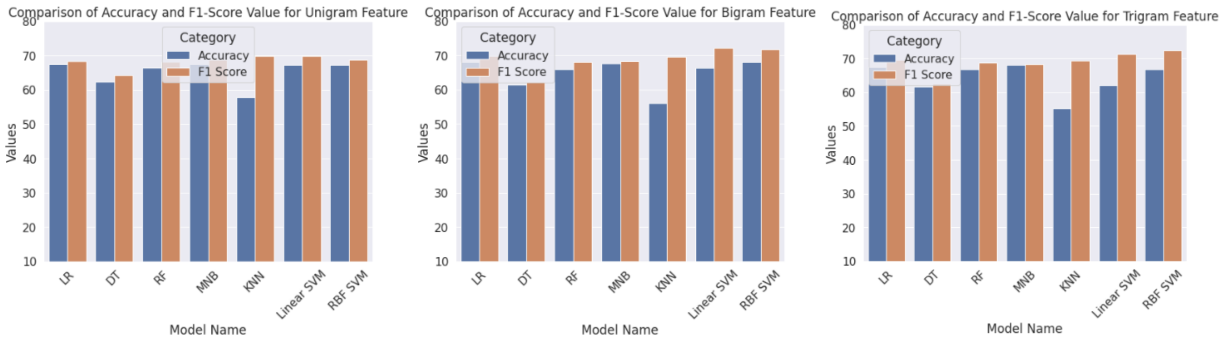
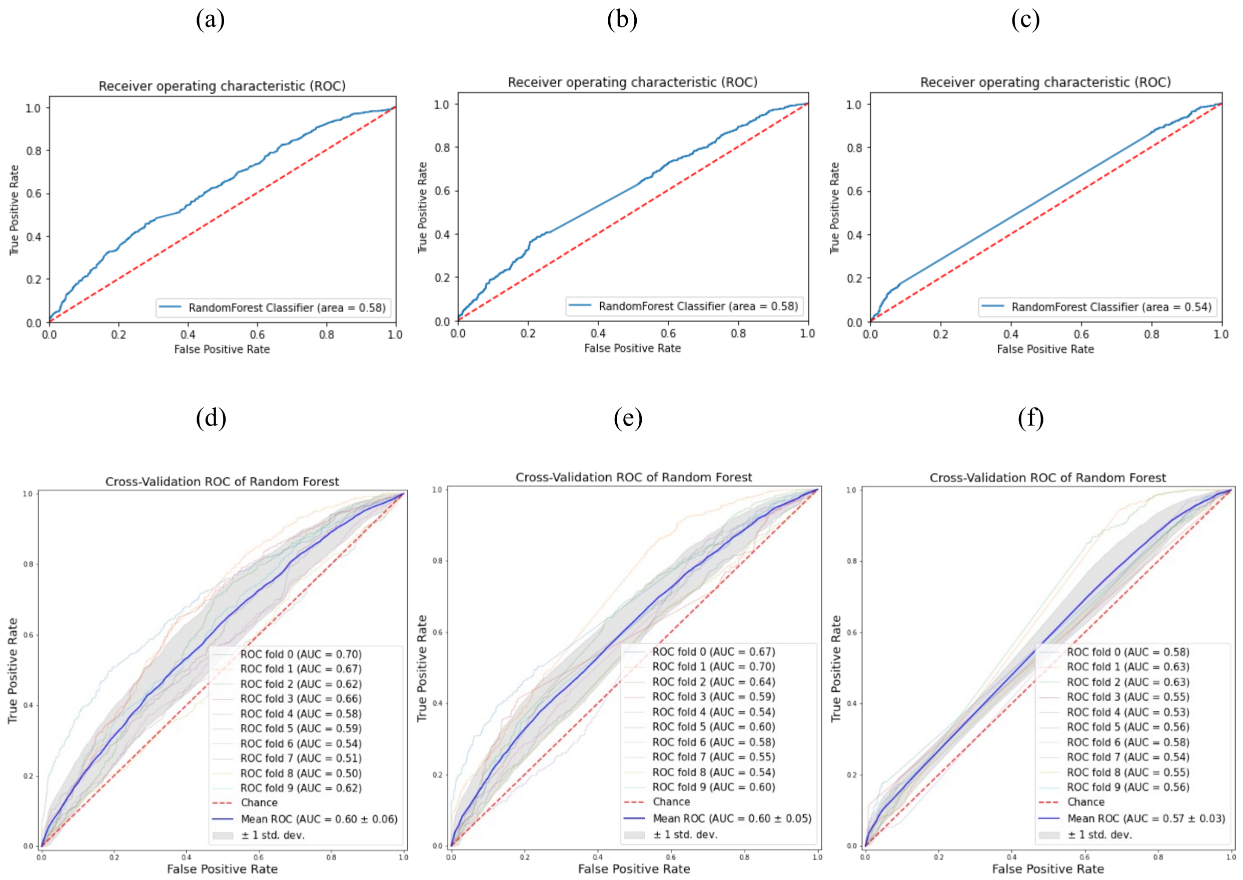


Figure 4: Model performance

The ROC (Receiver Operating Characteristic) curve given shows another describing graphical illustration where by determining the graphs performance, one can evaluate the binary classifier allocation.


(a)

(b)

(c)

(d)

(e)

(f)

Figure 5: (a) ROC curve of RF in unigram, (b) ROC curve of RF in bigram, (c) ROC curve of RF in trigram, (d) Cross-Validation ROC curve of RF in unigram, (e) Cross-Validation ROC curve of RF in bigram, (f) Cross-Validation ROC curve of RF in trigram

4.4 Summary

Many words were rendered incomprehensible once special characters were removed, as these were also responsible for assessment in the event of sentiment expression. In some circumstances, I needed to take a different approach by re-processing them into a more positive perspective.

Chapter 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Every sentiment of human being nowadays may be allocated to the words we see on numerous web platforms on a daily basis. In this case, it is vital for these platforms to have a system in place to distinguish between genuine sentiments and planted aggression. This is why I've decided to concentrate my efforts on one of the most exciting genres of all time: comedy. I hope to do this by ushering in a more definite and diversified digital era.

5.2 Impact on Environment

The main impact on environment happens when there is massiveness of social sentimental corrosion. Say for an example, if a products rating is good or bad is discriminated via the social media comments, in such case the positive and negative emotion detection can come is very handy in case of preventing the product or persay any kind of social pool to uphold the contradiction to remain stated. In many cases when the negative multitude of work reaches to a certain depth of immoral views, it can be highly possible that there could occur an outrage of events. To help in dictate in such cases the work of mine could be proven quite efficient in stablishing a promising order. And let's say by measuring such concept of positive and negative sentiments i can also help to remove of those which could be labeled harmful by the genuine people from their own experiences. The work of mine could be able to fasten the progression in such case.

5.3 Ethical Aspects

Because internet networking allows you to hunt for organizations that are interested in your benefits and pastimes, web-based living is one of the finest methods to meet and connect with new people who share your interests. This is great for meeting new people, but it's also great for love interests and web dating, which has grown more popular than traditional face-to-face meetings due to online life and any similarities to Tinder.

The internet is a wonderful tool for fast spreading news throughout the world, with "breaking news" tweets gaining a huge number of retweets in minutes. This might be highly valuable for informing people about crucial information, such as weather updates and missing children. As previously said, internet networking has had a positive influence on society in a variety of ways, all at no extra cost because all essential web-based life stages are free. Evaluate another object or administration that has ever transformed your life in terms of the internet, and then consider its cost.

5.4 Sustainability

● There are around two and three billion total dynamic social media-based life customers.

● 90% of major corporate companies have at least two social media and other platform based life cycles.

● in cases they are unable to get through their profile, 65 percent of individuals feel uneasy and uncomfortable.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

I feel that this study may make a new addition to the work's notion of the continuous developing age of BNLP in a vast globe where emotional intelligence is growing more connected with all the turbulence between our daily life structure and habits. Emotion detection has long been a source of contention among academics in this subject, and I hope that my work in this contribution will inspire others. So that i might envision a future in which technology augments one's intelligence in the case of a tragedy. I outline a method for classifying opinions of various sentiments in Bengali literature. I gathered information from a variety of reliable internet resources to achieve my purpose. I compiled data in a novel way, presenting just the words that are absolutely necessary to convey sentiment, leading to excellent accuracy.

6.2 Recommendations

The most recent trend in understanding the demands of the general public is experimental analysis; it's a simpler and more intelligent technique to study how people feel about a specific issue and the brand effect of smaller scale blogging. In this case, I analyzed people's sentiments regarding the aviation business, as well as United Airlines' ongoing problems and how the general public saw them. The investigation supported my beliefs about how successful a Twitter assumption investigation method is. The Logistic Regression and Random Forest classifiers employed in the computation, together with two programming for better results, clearly demonstrate the mass group assumption and, as a consequence, The airplane might readily examine the data and profit from it by attempting to enhance the features that look unpleasant or hated by the led audience. There are various options for this assignment, including:

- Take out any bias from the dataset.
- The percentage of data categories is evenly distributed.
- More categorization algorithms based on machine learning should be utilized.
- Classification necessitates parameter tweaking

6.3 Implication for Further Study

Because of the sensitive development of information on the web and internet-based live locations, businesses may use conclusion examination to gain insight into clients' perspectives about their items or administrations. In modern literature, sentiment based comment inquiry is frequently conducted utilizing a few days' worth of social media data. This barrier prevents factually significant and significant consequences from being realized unless social media material is frequently viewed. A comprehensive analysis of tweets must meet a few characteristics in order to provide a factually massive client evaluation.

REFERENCES

1.Liu Bing. Sentiment analysis and opinion mining. Synth Lect Hum Lang Technol 2012;5(1):1–167.

2. Kumar and A. Jaiswal. (2020). Systematic Literature Review of Sentiment Analysis on Twitter using Soft Computing Techniques. Concurrency and Computation: Practice and Experience, 32(1), e5107.

3. Pang Bo, Lee Lillian. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. 2004, arXiv preprint Cs/0409058.

4. Bengio Yoshua, Ducharme Réjean, Vincent Pascal, Janvin Christian. A neural probabilistic language model. J Mach Learn Res 2003;3:1137–55.

5. Aldhaheri, A.; Lee, J. Event detection on large social media using temporal analysis. In Proceedings of the 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 9−11 January 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6. [CrossRef]

6. P. Yang and Y. Chen. (2017). A Survey on Sentiment Analysis by using Machine Learning Methods. 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 117-121: IEEE.

7. Rani and P. Kumar. (2019). A Journey of Indian Languages over Sentiment Analysis: A Systematic Review. Artificial Intelligence Review, 52(2), 1415-1462.

8. L. B. R. Thapa and B. K. Bal. (2016). Classifying Sentiments in Nepali Subjective Texts. 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA), 1-6: IEEE.

9. N. Tabassum and M. I. Khan. (2019). Design an Empirical Framework for Sentiment Analysis from Bangla Text using Machine Learning. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 1-5: IEEE.

10. Hassan Asif, Amin Mohammad Rashedul, Al Azad Abul Kalam, Mohammed Nabeel. Sentiment analysis on bangla and romanized bangla text using deep recurrent models. In: 2016 International workshop on computational intelligence. IEEE; 2016, p. 51–6.

11. Tripto Nafis Irtiza, Ali Mohammed Eunus. Detecting multilabel sentiment and emotions from bangla youtube comments. In: 2018 International conference on bangla speech and language processing. IEEE; 2018, p. 1–6.

12. Alam Md Habibul, Rahoman Md-Mizanur, Azad Md Abul Kalam. Sentiment analysis for bangla sentences using convolutional neural network. In: 2017 20th International conference of computer and information technology. IEEE; 2017,p. 1–6.

13. Sharfuddin Abdullah Aziz, Tihami Md Nafis, Islam Md Saiful. A deep recurrent neural network with bilstm model for sentiment classification. In: 2018 International conference on bangla speech and language processing. IEEE; 2018, p.1–4.

14. Chauhan, Chhaya and Smriti Sehgal. “Sentiment analysis on product reviews.” 2017 International Conference on Computing, Communication and Automation (ICCCA) (2017): 26-31.

15. Maliha Jahan, Istiak Ahamed, Md. Rayanuzzaman Bishwas and Swakkhar Shatabda, “Abusive Comments Detection in Bangla-English Code-mixed and Transliterated Text”, IEEE

16. Md Gulzar Hussain∗ , Tamim Al Mahmud, and Waheda Akthar, “An Approach to Detect Abusive Bangla Text”, International Conference on Innovation in Engineering and Technology (ICIET)

17. Shafin, Minhajul Abedin, et al. "Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language." 2020 23rd International Conference on Computer and Information Technology (ICCIT). IEEE, 2020.

18. Hossain, Eftekhar, Omar Sharif, and Mohammed Moshiul Hoque. "Sentiment polarity detection on bengali book reviews using multinomial naive bayes." Progress in Advanced Computing and Intelligent Engineering. Springer, Singapore, 2021. 281-292

19. 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.

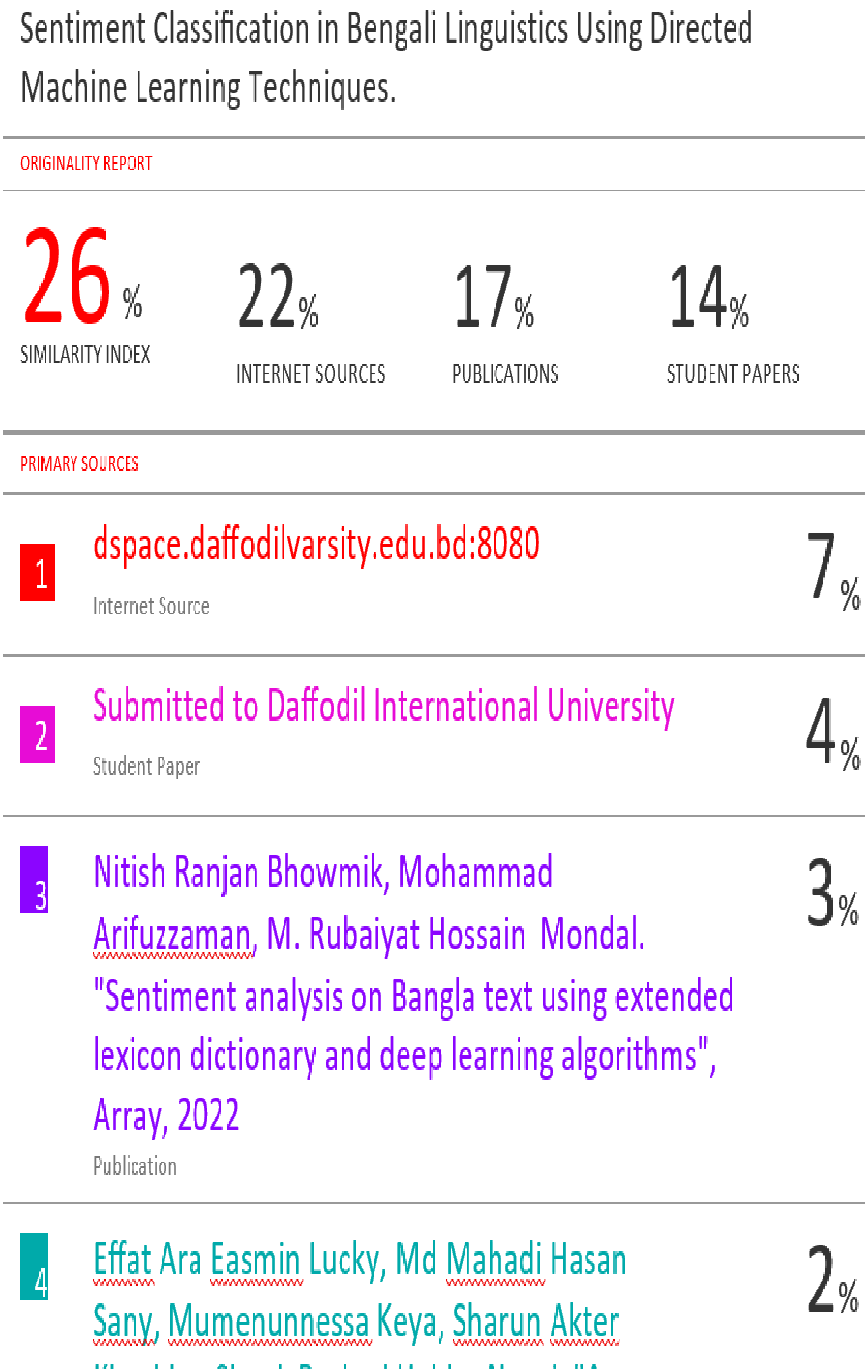
20. Md Waliur RahmanMiah,John Yearwood, and Sid Kulkarni, "Detection of child exploiting chats from a mixed chat dataset as a text classification task", Proceedings of the Australasian

21. Chintan Amrit∗ Tim Paauw† Robin Aly‡ . Miha Lavric §, "Using text mining and machine learning for detection of child abuse", arXiv:1611.03660v2

22. KM Azharul Hasan, Md Sajidul Islam, GM Mashrur-E-Elahi, Mohammad Navid Izhar, "Sentiment recognition from bangla text" Technical Challenges and Design Issues in Bangla Language Processing, IGI Global, 2013.

23. P. Turne., "Thumbs Up or Thumbs Down ? Semantic Orientation ." Proceeding of Applied to Unsupervised Classification of Reviews the 40th Annual Meeting of the Association for Computational (ACL) ,Philadelphia ,pp.417-424 ,2002.

24. Mittal, Namita, et al. "Sentiment analysis of hindi reviews based on negation and discourse relation." Proceedings of the 11th Workshop on Asian Language Resources. 2013



Sentiment Classification in Bengali Linguistics Using Directed Machine Learning Techniques.

ORIGINALITY REPORT

26 %	22%	17%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	7 %
2	Submitted to Daffodil International University Student Paper	4 %
3	Nitish Ranjan Bhowmik, Mohammad Arifuzzaman, M. Rubaiyat Hossain Mondal. "Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms", Array, 2022 Publication	3 %
4	Effat Ara Easmin Lucky, Md Mahadi Hasan Sany, Mumenunnessa Keya, Sharun Akter	2 %