

DETECTION OF HATE SPEECH ON SOCIAL MEDIA USING MACHINE LEARNING

BY

RUBEL BISWAS

ID: 161-15-7668

AND

APURBO DATTA

ID: 161-15-7610

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

Supervised By

Raja Tariqul Hasan Tusher

Senior Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Sheikh Abujar

Lecturer

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2019

APPROVAL

This Project titled “**DETECTION OF HATE SPEECH ON SOCIAL MEDIA USING MACHINE LEARNING**”, submitted by Rubel Biswas and Apurbo Datta, ID No: 161-15-7668 and 161-15-7610 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 05-12-19.



BOARD OF EXAMINERS

Dr. Syed Akhter Hossain
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Saiful Islam
Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

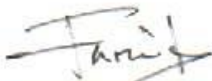
Internal Examiner



Shaon Bhatta Shuvo
Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Dewan Md. Farid
Associate Professor

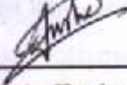
Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

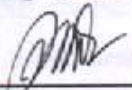
We hereby declare that, this project has been done by us under the supervision of Mr. Raja Tariqul Hasan Tusher, Lecturer, Department of CSE Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Mr. Raja Tariqul Hasan Tusher
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:



Sheikh Abujar
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Rubel Biswas

Rubel Biswas
ID: 161-15-7668
Department of CSE
Daffodil International University

Apurbo Datta

Apurbo Datta
ID: 161-15-7610
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 **Raja Tariqul Hasan Tusher**, Senior Lecturer, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Machine Learning 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. Syed Akhter Hossain** Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

The objective of our research is to detect hate speech on social media. Every day a huge amount of data is generated by the users of different social media. For this research we created a data set collecting data from Twitter. This data set consists of tweets of different kinds of people of different races and religions. In this work we followed a machine learning approach and as we know NB and SVM are the most popular algorithms for sentiment analysis and classifying text, so we used Naïve Bayes and Support Vector Machine algorithms in this work. While using NB we found an accuracy rate of 94.63% and in SVM the accuracy rate was 92.32%. As the actions of particular events on social media are not only bounded to the internet, they affect real-life events as well. Again, anything that spreads faster on social media compared to other media. Many people post hateful things on social media, which hurts others' feelings and then difficulties arise, and people have to face further consequences. By detecting hate speech, we can control these things and avoid such situations. So our work has value in keeping social media free from a few bad things and conflicts between people of different beliefs.

TABLE OF CONTENTS

CONTENS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1-3
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the Study	2
1.4 Research Questions	2
1.5 Expected Outcome	2
1.6 Report Layout	3
CHAPTER 2: BACKGROUND	3-6
2.1 Introduction	3
2.2 Related Work	4
2.3 Research Summary	5
2.4 Scope of Problem	5
2.5 Challenges	6

CHAPTER 3: RESEARCH METHODOLOGY	7-16
3.1 Introduction	7
3.2 Research Subject and Instrumentation	8
3.3 Data Collection Procedure	8
3.4 Statistical Analysis	9
3.5 Implementation Requirements	9
3.5.1 Anaconda	10
3.5.1.1 Installing Anaconda	11
3.5.2 Jupyter Notebook	12
3.5.2.1 Installing Jupyter Notebook	13
3.5.3 Anaconda Navigator	15
3.5.4 Installing Jupyter using Anaconda and conda	16
3.5.5 Natural Language Toolkit	16
CHAPTER 4: Experimental Results and Discussion	16-17
4.1 Introduction	16
4.2 Experimental Results	16
4.3 Descriptive Analysis	17
4.4 Summary	17
CHAPTER 5: Summary, Conclusion, Recommendation and Implication for Future Research	18-19
5.1 Summary of the Study	18

5.2 Conclusions 18

5.3 Recommendations 19

5.4 Implication for Further Study 19

REFERENCES

20-21

LIST OF FIGURES

FIGURES

PAGE NO

Figure 3.1 Proposed Model	7
Figure 3.2 Support Vector Machine	9
Figure 3.3 Twitter api	10
Figure 3.4. Interface of Jupyter Notebook on Chrome Browser	19

LIST OF TABLES

Tables

PAGE NO

Table 3.1 Hardware Requirements	12
Table 4.1 Accuracy	16

CHAPTER 1

INTRODUCTION

1.1 Introduction

At this moment we can clearly visualize that the number of people using social media increasing at great speed. Every day millions of people joining social media like Facebook, Twitter, and Instagram. This means that social media has become a very important communication medium today. By using social media technology people can send the message very quickly, so a particular topic can spread very quickly. Unfortunately, that also same for hate speech, it can also spread very fast. That can be matter of conflicts among different groups of our society. According to Cambridge Dictionary “Hate speech is a public speech that expresses hate or encourages violence towards a person or group based on something such as race, religion, sex, or sexual orientation “ [1]. By those criteria hate speech is defined is varied from one place to another place according to the globe. As we see in America there most of the inconvenient activity occurs on based issues like anti-American African, anti-immigrant, anti-White etc. On the contrary country like India most of the hate speech delivered on the basis of religion, cast, political views.

Right now it is very important having control over this huge scale of data on social media. There are several works on different methodology done to detect hate speech using data of social media like twitter, Facebook or other sites. There two method popular among one is word bag method, where a data set is created consist of hate word. And another approach is machine learning method. Machine leaning is used in different field like Business data analysis, prediction system, recommendation system, speech or handwriting recognition, bioinformatics, sentiment analysis and more [2]. They uses of machine learning seems very efficient and successful throughout the days on several field including which just mentioned. Using of machine leaning and data mining technique is a great option for deal with enormous collection of data. Where different machine learning algorithms is used to serve those purpose. There are different machine leaning algorithms like SVM, Naïve Bayes, Random Forest, Decision tree etc.

In our research we use machine learning techniques to detect hate speech and we applied SVM and Naïve Bayes algorithms and show the comparison of their performance. We collected data from twitter. For annotate a tweet we consider different factor like hate against other religion, race,

ethnicity, political view, tradition, gender, sexual orientation. In this article we describe the process of our work in different section.

1.2 Motivation

People want a pleasant environment on social media. A small line of word can hurt a lot people. Social media activity's circumstance not bounded only in virtual world Social media is one of the most frequently used communication platform today. People come under the same umbrella regardless their national boarder, socio economical background etc. by the contribution of social media. Social media help any message to be sent very quickly, become far-reaching and even viral when the certain topic takes public attention. As like that hate speech can also spread easily and rapidly and that can be reason of conflicts between different groups in our society. Hate speech especially concerning religion, nationality that we have seen in the recent past.

1.3 Rationale of the Study

Our research aimed to create a new data that consist of random users tweets, here twitter user information will never published. Those set that includes tweets which are not hate speech add those hate speech which including hatred for religion, nationality, ethnicity and gender. In this research we used machine learning techniques as Machine learning is repeatedly used approach for text classification and researcher shows great exertion using ML algorithms to analysis sentiment from text data.

1.4 Expected Outcome

- Maintain a Healthy Environment In Social Media
- Avoid unpleasant and embarrassing situation.
- Get rid of conflicts between groups in society
- Reduce misuse of Social Media

When those kind of violating and disrespectful speech can be detected in real-time on social media. It will be very easy to filter them or taking necessary measures among those person responsible for those kind of activity. And this can be a very efficient way to maintain a good and expected environment in social media like Facebook, Twitter. Thus also help us bypass many embarrassing situations when spending time on social media.

1.5 Report Layout

Report layout describes a summary of all the chapters. A brief summary of all chapters is given below:

- **Chapter 1:** Describes an Introduction of the Motivation, Rationale of the Study, Research Question, Expected Output, Report Layout.
- **Chapter 2:** Describes the background includes Related Work, Research Summary, Scope of the Problems, Challenges.
- **Chapter 3:** Describes the Research Methodology, Research Subject and Instrumentation, Data Collection Procedure, Statistical Analysis. Implementation Requirements
- **Chapter 4:** Describes the Experimental Results and Discussion on Experimental Results, Descriptive Analysis and Summary.
- **Chapter 5:** Describes the Summary, Conclusion, Recommendation and Implication for Future Research.

CHAPTER 2

BACKGROUND

2.1 Introduction

There are few work on detecting hate speech on social media. When ever comes the term in Bangladesh prospective the number of work is very few. Most of the work done in English and many other languages. Which is not related to Bangladesh prospect at any point. We found two work which is related to Bangladesh on of work with English text and the other uses Bangla text or script, as it very complicated to collect Bangla tweets automatically, the manually collect data from many sites and the amount of data set is very low consist only 200-300 words.

2.2 Related Work

Several works have worked on hate speech detection. In this section we will discuss about few previous work and their outcome. In [3] Warner done a brief discussion on hate speech, give a clear concept about hate speech. They proposed a method of detecting hate speech. They collect data from the American Jewish Congress (AJC) and Yahoo. They pick those texts which marked as offensive by those sites readers and found as hate speech according their definition. After analyzing they divided the data by their stereotype, like anti-Hispanic, anti-African American, anti-immigrants, anti-white, anti-sematic language. They build classifier for each stereotype. For classification template-based strategy is used in this work. They used SVM classifier with linear carnal and perform 10 fold cross validation for each classifier. After classification performance and error report is represent in two different tables. The overall accuracy was 96%.

In [4] Waseem collected about 16 thousand tweets and annotates them. They analyses the features for finding the features which can output the best performance and improve hate speech detection. They defines 13 conditions, where a tweet will be consider hate speech if it fulfil one or more pre-defined conditions. They test the impact of various feature on the performance using logistic regression with 10 fold cross validation. For model selection they uses grid search over all possible features. For each tweet n-gram is collected where value of n is 1, 2, 3, 4. Difference between one n-gram to another n-gram is found in their result.

Chen [5] analysis previous methods of text mining. They proposed an approach named Lexical Syntactical Feature (LSF) which may use for detecting offensive text as well as predict the potential of a user use of abusive content. According to their research combining context-specific features, structure features and style features with lexical features may improve existing machine learning methods and prediction of offensiveness.

A.H [6] used total 1525 messages from few group messaging services. Among them they annotated 68% of data as “OK” means not offensive and 32 percent as offensive. While they consider criteria like Slurs, Racism, Homophobia, Extremism, Crude language, Provocative language, Taboos, Unrefined language. Where 10 percent of data used for test and rest for training set. After processing data they gone three level classification. They uses Weke software for classification task. They first run few classifier and consider different criteria best classifier with high performance is chosen for classification. In three level they used Complement Naïve Bayes classifier, Multinomial Updatable Naïve Bayes classifier and DTNB (Decision Table/Naive Bayes hybrid classifier respectively. In those classifier level they percentage of correctly classified is varied from 84.26% to 96.72%.

2.3 Research Summary

In this report many related works on sentiment analysis of social media data has described. From the above content. It is clear that research of this topic is seems very important to many researcher across the world and the research on Bangladesh perspective very few. Although more and more research and project are being done in this area, However, People are trying to find out more effective and easy process of virtual for doing this work. For this research we done a study on this topic. Where we find few work related to this. Where many researcher uses different technique to detecting hate and offensive speech of internet and other social media. They use few different technique and different algorithms to this task. Which is discussed in the previous section. Our work is pretty different with respect of those work.

2.4 Scope of the Problem

In present days social media become a very popular and most uses communication medium. By using social media technology people can send the message very quickly, so a particular message or news spread very quickly. Unfortunately, that also same for hate speech, it can also spread very

quickly. That can be matter of conflicts among different community, groups of our society. As this formal definition of hate speech “Hate speech is a public speech that expresses hate or encourages violence towards a person or group based on something such as race, religion, sex, or sexual orientation “ [1]. Not on any particular place but also hate speech is common concern among different part of the world. Cyberbullying is a very harmful thing the pleasant community of internet and it also spread in real world form the internet. For keep pleasant environment in the internet as well as the social media it is very crucial to identify those hate and offensive speech.

2.5 Challenges

We found several challenge in the way of our research work. Collecting data. Firstly for the collection of data we have to gone through several process. For collecting data we have to get twitter api key and password. For this we have to apply for a developer account. Once we send them the information and the purpose of the work. They reply with they need additional information to approve the application. After providing those information the request approve by the twitter authority. Then we generate the api key and password for collecting tweets

The last, but not least which can be even called our biggest challenge in this work has been the factor of time as all of us are working full time and studying beside it. To agree on appointments and meet up was one of the most challenging parts of the job, as the project itself was a very interesting subject and it was fun to get the job done, but we learnt quite a lot from this process to be well structured and well planned.

CHAPTER 3

METHODOLOGY

3.1 Introduction

In the research we done few task sequentially, first it start with study precious work on this related subject then we collect data and create data sets. Then we process the data and label the data in certain classes with their merit. Then we choose few algorithms and with the classifier we train them and test the accuracy In our research we follow this methodology as shown in bellow.

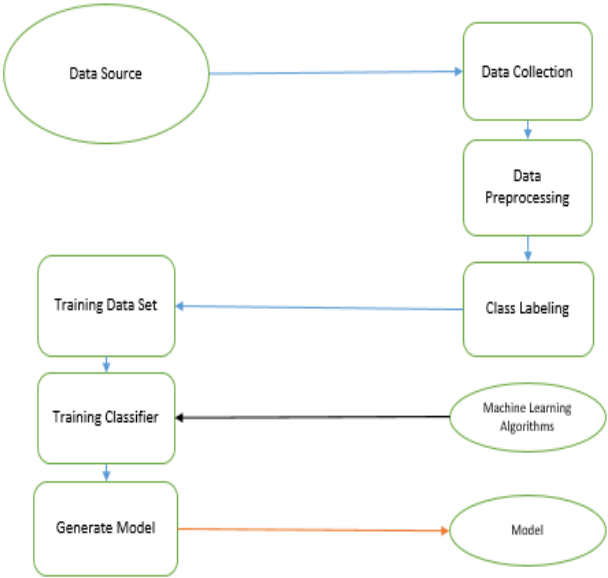


Figure 3.1: Proposed Model

3.2 Research Subject and Instrumentation

We choose machine learning methodology for this work. As machine learning is a great way for analysis dat. We use few machine learning algorithms to perform the task. Those are describe bellow.

Classifier and Algorithm's

Naïve Bayes :

Naïve Bayes is a probabilistic ML algorithm [14], in task like text classification, recommendation system, sentiment analysis, spam detection naïve bays algorithms is frequently used. Inspire of being a less complex algorism it perform many sophisticated algorithms. This algorithms based on Bayes theorem by Thomas Bayes. Since 1960's the popularity of this theorem grown rapidly. The theorem is about the equation given bellow [15].

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)} \dots\dots\dots (1)$$

Where P(A) and P(B) is the probability of A and B's independent observation of each other. Again P(A/B), P(B/A) are both conditional probability, here P(A/B) is the probability of A at the presence of B and P(B/A) is the probability of B at the presence of A .

Naïve Bayes classifier assume value of one feature without concerning about another feature. Each feature contribute independently to the overall probability. So that any change of value in one feature doesn't bother another features values. It's also a drawback of Naïve Bayes. Naive Bayes models are also known as independence Bayes and simple Bayes. There are different NB classifier model Gaussian naive Bayes, Multinomial naive Bayes, Bernoulli naive Bayes. The advantage is the algorithms is it can be coded really easy with programming language like Python or R and this classifier can be trained very efficiently. Another advantage is it can work very quickly, it can work fine with very large amount of data. So this algorithms is often use in many application.

Support Vector Machine:

In machine learning, support-vector machines (SVM), also known as support-vector networks are supervised learning models. In Supervised learning we have known the class label [16]. When the class label are unknown then supervised learning is not possible, then used unsupervised learning. There is also svm clustering algorithms, which is created by Hava Siegelmann and Vladimir Vapnik, In this work we used Supervised learning as we know the target classes. SVM is available in many tools like weka, sci-kit learn, MATLAB, kernlab and OPEN CV [16].

SVM helps to draw a line between 2 categories of data exist in a data set. The margin between data know as Support Vector. There are two kind of margin one is hard margin and other is soft margin

[17]. When data of two class can be divided linearly into two categories then used hard margin on the other hand when data is not linearly separable then it's uses soft margin.

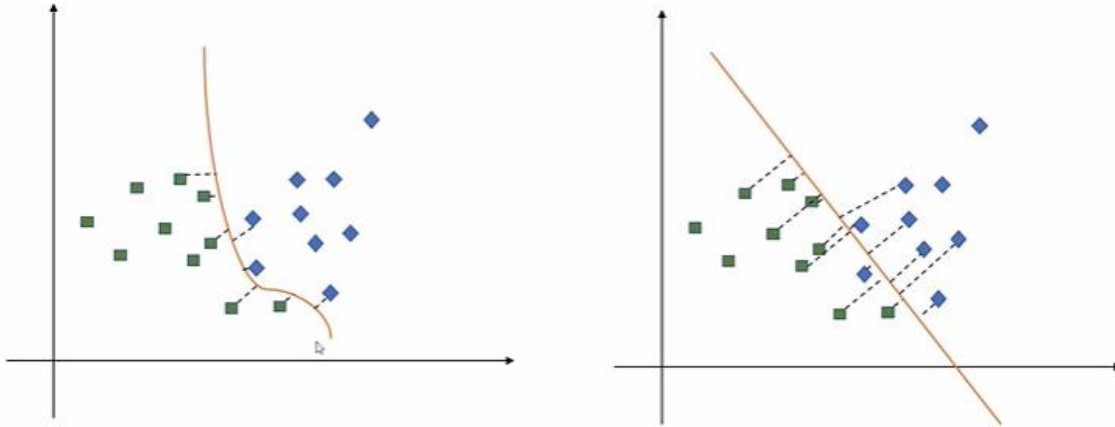


Figure 3.2: Support Vector Machine

Task like Classification, Detection and Regression can be perform with support vector machine. Over the year SVM is a very popular in use in industrial work. SVM frequently uses fields like face detection, categorize of image, text classification, bioinformatics, hand writing recognition and more.

3.3 Data Collection Procedure

Creating the dataset consists of few steps like selecting data source, collecting and labeling the data. Those process are discussed below.

3.3.1 Data Collection

For collecting data for this work we chose twitter as data source. From twitter data can be collected form twitter API. For this you need to follow few steps. First of all you need to create a twitter developer account, for this you need to apply with required information asked by twitter. For this you have to acknowledge them what exactly you are going to do with this account and what is the purpose of collecting data form twitter. They also asked how you going to use the data and do your keep user identification secret or you going to share them. If twitter team is not satisfied with the information you provided they will reply to you asking additional information. You have to provide them those information so they can clearly understand your purpose. Then when your

application approved, you will receive an email from twitter team then you can get access token secret and with this you can collect tweets from particular topic as your requirement. By using particular keywords. We export the data to a csv file.

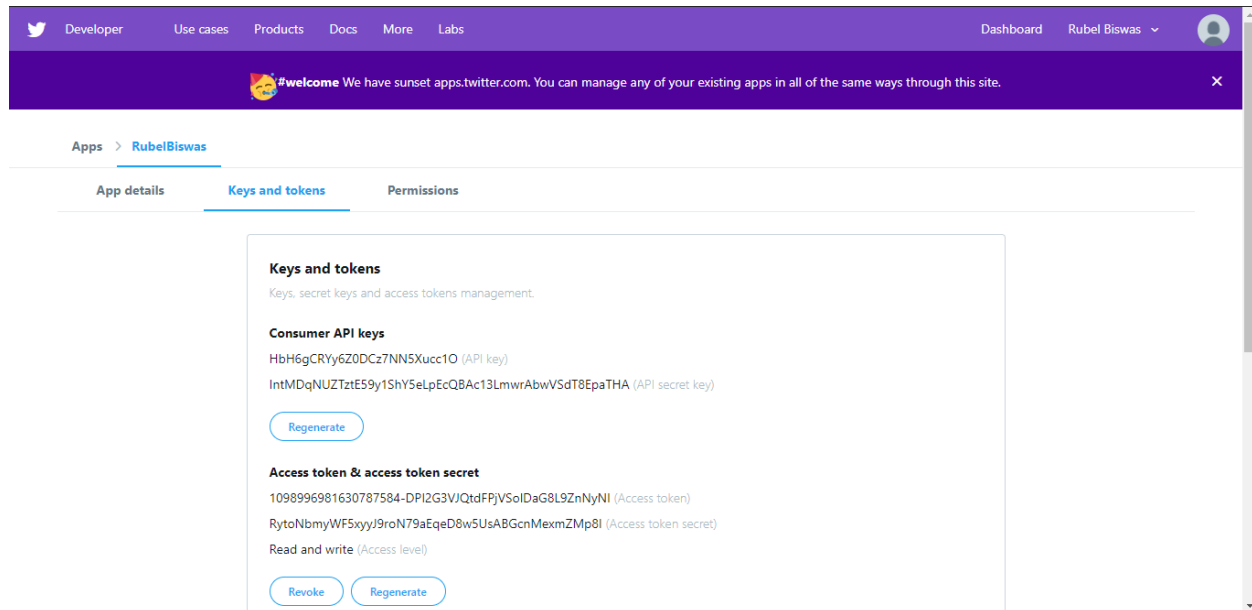


Figure 3.3: Twitter developer account API Keys and Token

3.3.2 Labeling Data

After collecting tweets they should be identified is it a hate speech or not. For this data labeled manually. Those data labeled abjectly two category those tweets which seems hateful labeled as “hate speech” and other tweets consider as “non-hate tweets”. Which tweet is hate which given 1 as class label and which in non-hate is given 0 as class label.

3.3.3 Preprocessing

The preprocessing method used by this work have those steps.

- 1) Retweet removal
- 2) Cleaning text
- 3) Lowercasing

- 4) Correcting Spell
- 5) Negation handling
- 6) Stop word removal.

For preprocessing tweets we used Natural Language Processing Toolkits. Now we will see an overview of some NLP tools which are commonly used in NLP NLTK is a python based platform to work with NLP. It was developed by Steve Bird and Edward Loper. NLTK was made available to uses from year 2001 under Apache 2.0 license. It provides various tools for classification, POS tagging, parsing, tokenization, stemming and semantic reasoning alone provides 50 corpora and lexical resources. NLTK support research in NLP and some other closely related research areas such as machine learning, artificial intelligence, linguistics and information retrieval.

2. Features Extraction: After gathering large tweet corpus, we have built and train classifier for tweet sentiment analysis. We examine mainly two classifiers: Naïve Bayes and Support Vector Machine. For each classifier we extract the same features from the tweets to classify on it. To build feature set, we process each tweet and extract meaningful feature and create feature matrix by unigram technique. For example, if positive tweet contains word “sorrow”, a feature for classification would be whether or not a tweet contains the word “sorrow”.

3. Training Module: The generated data is used as training dataset to train the model for sentiment analysis. On inspecting the model on test dataset, we receive the tweet sentiment labels as an output. We will use this dataset for detecting hate speech.

4. Classification and Evaluation: We used supervised learning approach in classifying tweets. We would compare the performance of NB, SVM algorithm with our dataset.

3.4.1 Anaconda

Anaconda is a open-source software package of the Python and R programming languages for scientific computing (data science, data mining and machine learning applications, predictive analytics, large-scale data processing, etc.) [7], that motivate to simplify package management and deployment. Package management system of conda manages package versions. Over 15 million users using Anaconda distribution which includes exceeding 1500 popular data-science packages. Anaconda support different operating system like MacOS, Linux and Windows.

The requirements for Anaconda Installation is discussed below.

Table 3.1: Hardware requirements [12].

Category	Description
CPU:	2 x 64-bit 2.8 GHz 8.00 GT/s CPUs
RAM:	32 GB (or 16 GB of 1600 MHz DDR3 RAM)
Storage:	300 GB. (600 GB for air-gapped deployments.)*
Communication:	Internet access to download the files from Anaconda Cloud or a USB drive containing all of the files you need with alternate instructions for air gapped installations

*Additional space recommended if the repository will be used to store packages built by the customer. With an empty repository, a base install requires 2 GB.

3.4.1.1 Installing Anaconda

This is a open source software and anyone can use it for free. For using it you have to download and install. For downloading you have to go there official website where they provide anaconda for different operating system. For windows operating system we choose Windows 10s version. One and chose either Python 3.6 version or Python 2.7 version as they uses. Then they ask for mail address you may provide your email address or just skip this option, then download will be started. After completing download, click and install the file, it will prompt with anaconda minimal interface then follow the instruction and read the license agreement and click ok. Then the installing process will start and after while the package will be installed.

3.4.2 Anaconda Navigator

It is available on major operation system like Windows, macOS and Linux [9]. Anaconda navigator is an alternative of command line, this comes with anaconda distribution. It make easy for user to launch application. Conda packages, channels and package can be managed by anaconda navigator without command line. Search operation for package on both Local anaconda repository and Anaconda cloud can be perform with the help of Anaconda Navigator. Installing

packages in an environment, run and update them is less complex in anaconda navigator with comparison to Command line.

3.4.3 Jupyter Notebook

Jupyter Notebook comes with anaconda full package, its previously known as IPython Notebooks. This is a run on computer browser like a web application [10]. The codes are written on input box called “cell”. The advantage of this in jupyter notebook particular block of code can run separately while other interpreter run the whole code at the same time. In the interface of jupyter can contain code, text, mathematical plot or other kind of media. The file of the jupyter can save on any local disk of computer. Those files generally contains “.ipynb” in the end of file as extension.

Using “Download As” option from web interface Jupyter Notebook can be save as different standard formats like HTML, PDF, python file and more format also available. It can be convert from shell using “jupyter nbconvert” or nbconvert library. Nb convert library is provided via NbViewer . Which convert notebook document to HTML or other formats by taking the URL of the document. It simplify the visualization for display to the user.

Python Notebook interface

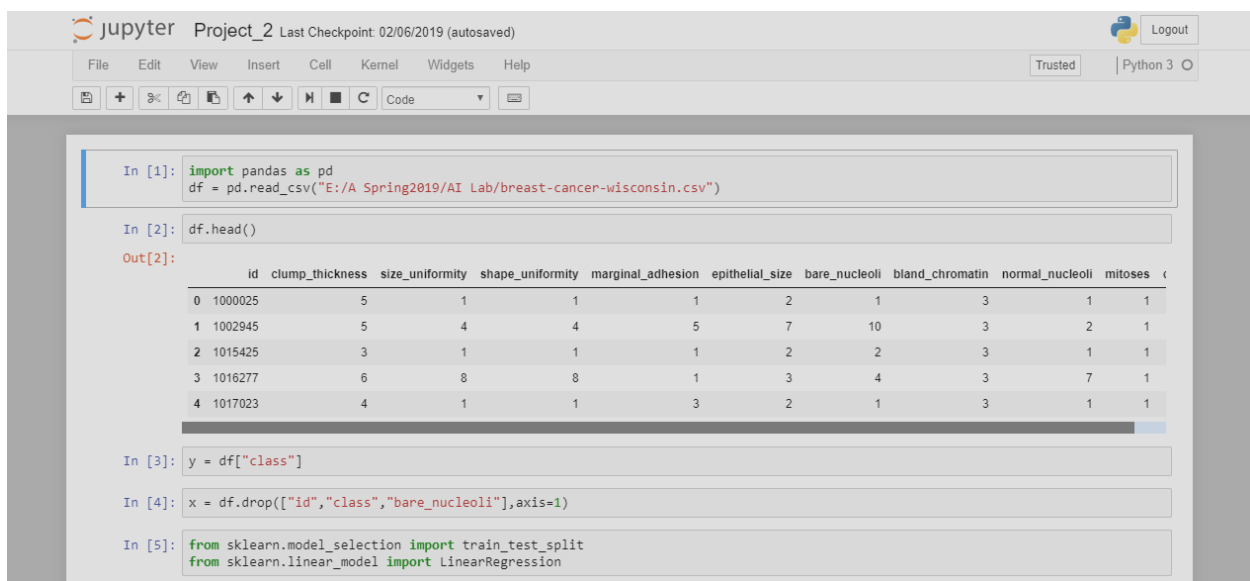


Figure 3.4: Interface of Jupyter Notebook on Chrome Browser

A Browser based REPL is provided by Jupyter notebook which built contains number of open-source libraries. For programming is different languages Jupyter Notebook allow to connect to a lot of carnal. There are already 49 Jupyter kernels available for programming many languages like R, Python, Julia and Haskell. Python kernel comes with Jupyter notebook as default kernel.

In December 2011 the notebook interface added to Ipython on it's 0.12 release. Later on 2015 it renamed as Jupyter Notebook. Though its interface is similar to Maple, Sage Math or Mathematica interface, it gains vast popularity recent years. It overtook the popularity of Mathematic in 2018.

3.4.4 Installing Jupyter Notebook

3.4.4.1 Prerequisite

Python

As previously discussed Jupyter can run code of several languages. There is different requirement for each of programming language. Python is required for running python on Jupyter notebook. We installed it using Anaconda. Which installation process is describe in previous section.

3.2.1.2 Installing Jupyter using Anaconda and conda

For new users, it is best course installing Anaconda [11]. Anaconda conveniently installs the Jupyter Notebook, Python and other popular packages for data science, machine learning and scientific other computing.

3.4.4.3 Installation steps:

- First of all need to download Anaconda. It's better to download Anaconda's latest Python 3 version (current version is Python 3.5).
- Install Anaconda which you downloaded, the instructions are given on the download page.
- When installation complete jupyter notebook will be ready for run.

3.4.5 Natural Language Toolkit

The Natural Language Toolkit, or commonly known as NLTK [12], is a suite of libraries as well as programs for symbolic and statistical natural language processing (NLP) for script written in English in Python language. NLTK was developed by Edward Loper and Steven Bird in the Computer and Information Science department at the University of Pennsylvania [13]. Graphical demonstrations and sample data is comprised in it. Natural Language Toolkit is associated by a

book. The underlying concepts behind the language processing tasks and supported toolkit is explained in this book, also having a cookbook.

NLTK is designed to support research and teaching in NLP or compactly related areas, including cognitive science, empirical linguistics, information retrieval, artificial intelligence and machine learning. NLTK has been applied effectively as a teaching tool, as a distinct study tool, and as a entresol for prototyping and creating research systems. There are as much as 32 universities in the US and at least 25 countries using NLTK in their courses [12]. NLTK supports classification, stemming, tokenization, parsing tagging, and semantic reasoning functionalities.

3.5.6 Library highlights

- Lexical analysis: Word and text tokenizer
- n-gram and collocations
- Part-of-speech tagger
- Word lemmatize.
- Tree model and Text chunker for capturing

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Introduction

Using those Support Vector Machine and Naïve Bayes algorithms we get 91% and 90% accuracy. The accuracy is given below. Rate of Precision, Recall, True Positive, False Positive, True Negative, False Negative represented below with a table. The resulting dataset had a size of 5000, consists of both hate tweets non-hate tweets. Based on the experimental results, The Naïve Bayes shows accuracy of 79.63% and in SVM the accuracy rate was 85.62%. The results also showed that instead of using word bigram or trigram alone, it was better to union word bigram and word trigram. In adding character n-gram and negative sentiment to the feature sets was not needed. While previous study said that RFDT, BLR, and SVM had the same performance in detecting hate speech, we found out that NB performance was better than SVM.

4.2 Experimental Results

Performance of selected model measured by using Confusion Matrix. Using SVM the accuracy rate is 85.62 % and error rate is 14.27% where using Naïve Bayes Accuracy and Error rate is 79.63% and 19.12%. This result is shown in table 4.1.

Table 4.1: Accuracy Rate

Classifier	Accuracy	Error Rate
SVM	85.62	14.27
Naïve Bayes	79.63	19.12

For SVM precision, recall and F1-score is sequentially .725, .743 and .825 on the other hand for Naïve Bayes precision, recall and F1-score is .697, .682 and .774 respectively.

Table 4.2: Precision, Recall and F1 Score

Classifier	Precision	Recall	F1-score
SVM	.725	.743	.826
Naïve Bayes	.697	.682	.774

4.3 Summary

We used two different classifier and compare their performance into prepared data set. The resulting dataset had a size of 5000 tweets, consists of both hate tweets non-hate tweets which is labeled before according to the merit of the tweet. We tested the data several times, based on the average of experimental results, using those algorithms we get 85.62% and 79.63% accuracy, where error rate is 14.27% and 19.12% respectively for SVM and NB. The Naïve Bayes shows less accuracy than SVM the accuracy rate. The results also showed that instead of using word bigram or trigram alone, it was better to union word bigram and word trigram.

CHAPTER 5

SUMMARY CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

5.1 Summary of the Study

In the research we done some task, first it start with study previous work on this related subject then we collect data and create data sets. Then we process the data and label the data in certain classes with their merit. For processing data we used NLTK Then we choose few algorithms and with the classifier we train them and test the accuracy. The accuracy of the classifier present in result section. In this report we describe the steps and all requirements of this work.

5.2 Conclusion

In this paper we discussed about few work which previously done by researcher and wrote about their techniques and methodology. And make complete study, besides our main focus was to represent our work. In our research we collected raw data, processed them and create a dataset consist of 5000 clear tweets which consists of both hate tweets non-hate tweets. The annotation of data done manually after that use them in our classifier. We used applied two different algorithms SVM and NB. Where SVM accuracy is higher than NB accuracy. Using SVM the accuracy rate is 85.62 % and error rate is 14.27% where using Naïve Bayes Accuracy and Error rate is 79.63% and 19.12%. In our point of view this work has social value to make the environment of social media pleasant and user friendly. By keeping the bully's away, social media can be more continent for calm people. For future work we will use motivate do research on detection of hate speech in Bengali language with improve techniques.

5.3 Recommendations

Detection of hate speech in social media like Facebook, twitter as real-time application can prevent this kind of activity by user. Using more advance algorithms the accuracy of this process can be improved. Analysis data with similar technique many crime of virtual world can be predict or detect.

5.4 Implication for Further Study

In this work we work with we work with a moderate number of data, using more data give more competitively better accuracy. Using more advance algorithms the accuracy of this process can be improved. For the future work on hate speech detection in the Bengali language.

REFERENCES

- [1] Cambridge Dictionary, "hate speech," Cambridge Dictionary, [Online]. Available: <https://dictionary.cambridge.org/dictionary/english/hate-speech>. [Accessed 23 09 2019].
- [2] Daffodil Software, "9 Applications of Machine Learning from Day-to-Day Life," Medium, 31 07 2017. [Online]. Available: <https://medium.com/app-affairs/9-applications-of-machine-learning-from-day-to-day-life-112a47a429d0>. [Accessed 23 09 2019].
- [3] W. W. a. J. Hirschberg, "Detecting Hate Speech on the World Wide Web," in *Proceedings of the 2012 Workshop on Language in Social Media (LSM 2012)*, Montreal, Canada, 2012.
- [4] Z. W. a. D. Hovy, "Hateful Symbols or Hateful People?," in *Association for Computational Linguistics*, San Diego, California, 2016.
- [5] Y. Chen, Y. Zhou, S. Zhu and H. Xu, "Detecting Offensive Language in Social Media to Protect Adolescent Online Safety," in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, Amsterdam, Netherlands, 2012.
- [6] A. H. Razavi, D. Inkpen, S. Uritsky and S. Matwin, "Offensive Language Detection Using Multi-level Classification," in *Canadian Conference on Artificial Intelligence*, Berlin, Heidelberg, 2010.
- [7] Wikipedia, "Anaconda (Python distribution)," Wikipedia, the free encyclopedia, [Online]. Available: [https://en.wikipedia.org/wiki/Anaconda_\(Python_distribution\)](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)). [Accessed 23 09 2017].
- [8] conda.io, "Installation," conda.io, [Online]. Available: <https://docs.conda.io/projects/conda/en/latest/user-guide/install/>. [Accessed 23 09 2019].
- [9] Anaconda, "Anaconda Navigator," Anaconda, [Online]. Available: <https://docs.anaconda.com/anaconda/navigator/>. [Accessed 23 09 2019].
- [10] Wikipedia, "Project Jupyter," Wikipedia, the free encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/Project_Jupyter. [Accessed 23 09 2019].
- [11] Jupyter, "Installing Jupyter Notebook," Jupyter, [Online]. Available: <https://jupyter.readthedocs.io/en/latest/install.html>. [Accessed 23 09 2019].
- [12] NLTK, "Natural Language Toolkit," NLTK, [Online]. Available: <https://www.nltk.org/>. [Accessed 23 09 19].
- [13] Wikipedia, "Natural Language Toolkit," Wikipedia, the free encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/Natural_Language_Toolkit. [Accessed 23 09 2019].
- [14] Wikipedia, "Naive Bayes classifier," 25 07 2019. [Online]. Available: https://en.wikipedia.org/wiki/Naive_Bayes_classifier.
- [15] C. W. A. S. Andrew Ellinor, "Bayes' Theorem and Conditional Probability," Brilliant, [Online]. Available: <https://brilliant.org/wiki/bayes-theorem/>. [Accessed 23 09 2019].
- [16] Wikipedia, the free encyclopedia, "Support-vector machine," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Support-vector_machine. [Accessed 23 09 2019].

[17] codebasics, "Support Vector Machine (SVM)," Youtube, 19 12 2018. [Online]. Available: <https://www.youtube.com/watch?v=FB5EdxAGxQg>. [Accessed 23 09 2019].

Document Viewer

Turnitin Originality Report

Processed on: 28-Nov-2019 11:43 +06

ID: 1223166088

Word Count: 5081

Submitted: 1

DETECTION OF HATE SPEECH
ON SOCIAL MEDIA Usin.... By
Rubel Biswas 161-15-7668

Similarity Index	Similarity by Source	
28%	Internet Sources:	12%
	Publications:	16%
	Student Papers:	18%

[exclude quoted](#) [exclude bibliography](#) [exclude small matches](#) mode:
 quickview (classic) report [print](#) [refresh](#) [download](#)

4% match (publications)

[Ika Alfina, Rio Mulla, Mohamad Ivan Fanany, Yudo Ekanata, "Hate speech detection in the Indonesian language: A dataset and preliminary study", 2017 International Conference on Advanced Computer Science and Information Systems \(ICACSIS\), 2017](#)

2% match (publications)

[Tejas Mankar, Tushar Hotchandani, Manish Madhwani, Akshay Chidrawar, C.S Ufna, "Stock Market Prediction based on Social Sentiments using Machine Learning", 2018 International Conference on Smart City and Emerging Technology \(ICSCET\), 2018](#)

2% match (publications)

[H. Sankar, V. Subramaniaswamy, "Investigating sentiment analysis using machine learning approach", 2017 International Conference on Intelligent Sustainable Systems \(ICISS\), 2017](#)

2% match (Internet from 21-Nov-2019)

https://en.wikipedia.org/wiki/Natural_Language_Toolkit

2% match (student papers from 29-Sep-2019)

[Submitted to Daffodil International University on 2019-09-29](#)

1% match (student papers from 04-Nov-2019)

[Submitted to Daffodil International University on 2019-11-04](#)

1% match (student papers from 07-Apr-2019)

[Submitted to University of Hull on 2019-04-07](#)

1% match (student papers from 04-Nov-2019)

[Submitted to Daffodil International University on 2019-11-04](#)