

**SENTIMENT ANALYSIS OF CYBER-VICTIMIZED TEXT DETECTION ALONG
WITH THE LOCATION ON SOCIAL MEDIA**

BY

Md. Jahidul Hasan

ID: 123-15-2034

AND

Sushmita Barai

ID: 171-15-9070

AND

Afsana Akter

ID: 171-15-858

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

Supervised By

Zerin Nasrin Tumpa

Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Ms. Rubaiya Hafiz

Sr. Lecture

Department of CSE

Daffodil International University

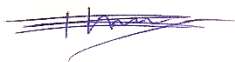


**DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JANUARY 2021**

APPROVAL

This Project/internship titled “**Sentiment Analysis of Cyber-Victimized Text Detection along with The Location on Social Media**”, submitted by **Md. Jahidul Hasan, ID No: 123-15-2034, Sushmita Barai, ID No: 171-15-9070** and **Afsana Akter, ID No: 171-15-858** 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 31 January, 2021.

BOARD OF EXAMINERS



Dr. Touhid Bhuiyan
Professor and Head

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

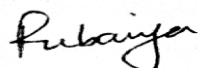
Chairman



Md. Sadekur Rahman
Assistant Professor

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

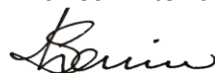
Internal Examiner



Rubaiya Hafiz
Senior Lecturer

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

Internal Examiner



Dr. Shamim H Ripon
Professor

Department of Computer Science and Engineering
East West University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Zerin Nasrin Tumpa, 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.

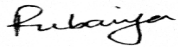
Supervised by:



Zerin Nasrin Tumpa
Lecturer

Department of CSE
Daffodil International University

Co-Supervised by:



Ms. Rubaiya Hafiz
Sr. Lecturer

Department of CSE
Daffodil International University

Submitted by:



Md. Jahidul Hasan

ID: 123-15-2034
Department of CSE
Daffodil International University



Sushmita Barai

ID: 171-15-9070
Department of CSE
Daffodil International University



Afsana Akter

ID: 171-15-858
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 **Zerin Nasrin Tumpa**, 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 draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Zerin Nasrin Tumpa**, Lecturer, **Ms. Rubaiya Hafiz**, Sr. Lecturer, and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

Nowadays the World is getting more and more dependent on the internet. In this Web 2.0 field the Social media getting popularity for its limitless usages with different features. It has become everyone's important part of life as they start to spend more time there. So, cyber-victimization on social media has also awakened due to this. People are now being cheated, harassed, or bullied over the internet. The aim of this research work came up with the idea of monitoring or analyzing the data from a specific location or even from a specific keyword or from a specific user to find out the sentiments behind his/her posts, comments.

The applications of sentiment analysis are broad and powerful in many ways. The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world.

Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. Being able to quickly see the sentiment behind everything from forum posts to news articles means being better able to strategize and plan for the future.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgment	iv
Abstract	v
Table of Contents	vi
List of figures	viii

CHAPTERS	PAGE
CHAPTER 1: Introduction	1-5
1.1 Introduction	1
1.2 Motivation	3
1.3 Rationale of the Study	3
1.4 Research Questions	4
1.5 Expected Output	4
1.6 Report Layout	5
CHAPTER 2: Literature Review	6-8
2.1 Introduction	6
2.2 Related Works	6
2.3 Research Summary	7
2.4 Scope of the Problem	8
2.5 Challenges	8

CHAPTER 3: Research Methodology	9-17
3.1 Introduction	9
3.2 Proposed Model	9
3.3 Research Subject and Instrumentation	10
3.4 Data Collection Procedure	10
3.5 Pre-Processing	10
3.6 Folding Up	11
3.7 Classifier Automated Training Collection	11
3.8 Extraction of Function	12
3.9 Research and Classification	12
3.10 Evaluation	12
3.11 Implementation	13
 CHAPTER 4: Experimental Results and Discussion	 18-24
4.1 Introduction	18
4.2 Experimental Results	19
4.3 Logistic Regression	19
4.4 XGBoost	20
4.5 Decision Tree	21
4.6 Model Comparison	22
4.7 F1 Score Compare	22
4.8 Summary	23

CHAPTER 5: Summary, Conclusion, Recommendation and Implication for Future Research **25-26**

5.1 Summary of the Study	25
5.2 Conclusions	25
5.3 Recommendations	26
5.4 Implication for Future Study	26

REFERENCES **27-28**

PLAGIARISM REPORT **29**

LISTS OF FIGURES

FIGURES	PAGE
Figure 1.5.1 Expected Output	4
Figure 3.2.1 Block Diagram of methodology	9
Figure 3.11.1 Python Modules	13
Figure 3.11.2 Raw Tweets from specific location (Kansas)	14
Figure 3.11.3 Tweets being polished using regex	14
Figure 3.11.4 Tweets with polarity and subjectivity	15

Figure 3.11. 5 Tweets being analyzed	15
Figure 3.11. 6 WordCloud	16
Figure 3.11.7 Expected Output	16
Figure 3.11.8 Expected Output	17
Figure 4.2.1 Experimental Result for Using Bag-of-Words	19
Figure 4.2.2 Experimental Result for Using TF-IDF Features	19
Figure 4.3.1 Output of Bag-of-Words Features for Logistic Regression	20
Figure 4.3.2 Output of TF-IDF Features for Logistic Regression	20
Figure 4.4.1 Output of Bag-of-Words Features for XGBoost	20
Figure 4.4.2 Output of TF-IDF Features for XGBoost	21
Figure 4.5.1 Output of Bag-of-Words Features for Decision Tree	21
Figure 4.5.2 Output of TF-IDF Features for Decision Tree	21
Figure 4.6.1 Model Comparison-1	22
Figure 4.6.2 Model Comparison-2	22
Figure 4.7.1 F1 Score Compare	22
Figure 4.8.1 Precision & Recall	23

CHAPTER 1

INTRODUCTION

1.1 Introduction

Present web 2.0 world, the usage of social media is increasing rapidly. Almost every capable person tries to connect online to know the world's present situation. Various social media such as blogs, social networking sites (e.g. Facebook), and message platforms (e.g. WhatsApp) creates an environment where it becomes possible to communicate with anyone at any time. It creates a possible way for the people who are involved in social media that helps build new friendships and maintain the existing one [2,3]. But everything has a fair side and an unfair side. While social media is a medium of connecting people via virtual, nowadays people are attacking others by using harmful sentences that are a very shameful matter. Most of the attacking sentences contain words, which are related to harassing, bullying, racism and also hates speech. Recent studies show that using offensive constitutes a growing problem among people of all ages. Every work must have its proper clear vision of its reason for creation. We want to work on cyber-victimized work because many of us confused it with cyber-bullying. It's a really pathetic recent study showing that an average of 20% to 40% of all teenagers have been abused online [5]. But cyber-victimized works consist of cyber-bullying, cyber-harassment, cyber-sexual-harassment, offensive words, even hate speech. Though most of them contain the same meaning that it harms our mental health and our daily life work cycle and so on.

Online Harassment is an online procedure that indicates psychological harm to any person via sending messages in Social Networks. So from the definition, we can understand they show a different image. The repetitions of any kind of hurting sentence are like bullying though the bullying was only held on educational yield or working place or where there is a business of humans to do any kind of work. But the bullying now comes towards our houses and makes an awkward situation while using online. We ignore it, which makes people do it more and more. Also, some people play a cold after seeing all those on social media.

On the other hand, Offensive sentences used to mock or insult someone or a group of people even by hurting in such a way being aggression against a certain culture, a subgroup of the society, race,

or ideology in a tirade [15]. There is no specific person who faces the problem sometimes it indicates a full nation that is connected with hate –speech.

When anyone harassed someone by using sexual-based words. Sometimes effect very badly in society because even till now many countries in our 3rd world, where talking on the topic of the sexual-based topic even it's for justification is considered taboo.

There is a study that shows that about 73 percent of women in Bangladesh, who have reported online harassment [21]. Though we are a small developing country with a huge population, we face that then what about the other countries that can't imagine.

Though the Internet is a restriction free gateway it has less restrictions on content. For that reason, in terms of the number of followers, contacts, and hub ratings, abusive users have the right to act similarly to spammers. And maximum time we neglect the fact of social victimization and pretend that it's a normal thing and advise others to avoid it and act that there is nothing happening. But the villain, who does the work, seems it's a fun game because of the cold behavior of the people who are victimized and who try to fix the situation by avoiding or trying to advise others to pretend there is nothing to do. Though there is law in almost every country for that behavior, the rate increasing and the mental health of the people also getting down even it creating a situation of suicide or self-harm.

It's a very relatively new and evolving phenomenon in a society that the role and influence of social media impact on suicide-related behavior. That is only beginning to understand by the people of the society [16]

From all that points, there is a difference but the main fact is that whatever it is, it causes a huge effect on our mental health. That is hiding now but day by day we will see the site.

To prevent any kind of accident or incident we have to find the root of the problem. Who facing it, from where the sources are coming toward us also the effect of the bad sides that we are ignoring. In this work we try to build a system for tracking its rate and try to identify them from social media. Mainly we use the English language because of its international language. We aim to develop a model to detect cyber-victimized text through social media along with its location and increasing the awareness of the people of our country and worldwide.

1.2 Motivation

Nowadays the use of social media is increasing rapidly around every age's people around the world. Though when there is no limitation or restriction of expressing one's emotions thought there must be a conflict of a good and bad side. Moral values seemed vanished sometimes on the content or comment or blog or even people judge others opinions. So in our work, we want to identify the places and detecting the victimized content by our little effort. In such a case, our rationale in the postulation was to –

1. Make Participants mindful of Cyber-victimizations side effects.
2. Identify the location for better prevention.
3. Create an image of the present social media environment.
4. Reduce Community-wide resistance and consolidate individuals to battle together against cyber-victimization.
5. Ensuring individual activities just as to guarantee steps taken by the Government and world-wide.

1.3 Rationale of the Study

The effect of our exploration of cyber-victimization and social has a long haul impact on the. Our exploration territories are-

1. Develop an extensive early cautioning and observation framework that has prescient capacity for finding convict location and create awareness around the points.
2. Understand the effect of worldwide mental health due to cyber-victimized inside the setting of the parameters of victoria limit, worldwide portability, and environmental change by finding the location.
3. Enable individuals to comprehend the indications and make strides as right on time as conceivable against cyber-victimized.

1.4 Research Questions

While researching the detection of the cyber-victimized word for its location some questions arose in mind. And the questions are-

1. How to collect correct content for a better result?
2. How to identify which words are cyber-victimized?
3. Which model structure give the appropriate result?
3. What are the effects of those in our daily life?
4. Is identify the location is a good idea?
5. Did we obtain the algorithm with the best accuracy?
6. Should we try more algorithms for better accuracy?

1.5 Expected Output

In this research developing a model for detecting cyber-victimization text along with the location on social media for sentiment analysis.

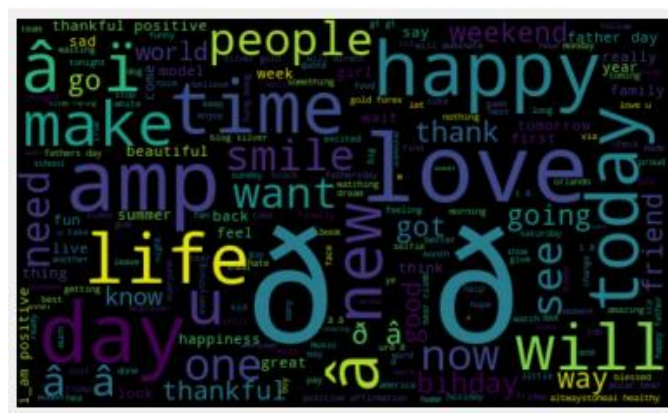


Figure 1.5.1 Expected Output

1. 6 Report Layout

Chapter 1: Introduction

In this part, we've talked about the motivation of ours for doing this research and also the rationale of the study, research questions, and the expected outcome of the research.

Chapter 2: Background

In this chapter, we will describe the related work of our research and comparative studies between our developed model and existing related model as well as the problems and challenges that we've faced.

Chapter 3: Research Methodology

In the chapter, we will describe the research methodology that means we will talk about how we did our research and about the terminology.

Chapter 4: Experimental results and Discussion

In this chapter, we will describe our research experimental results and outputs. We will talk about which results we have found by implementing the algorithms of machine learning over our collected data.

Chapter 5: Summary, Conclusion, Recommendation and Implication for Future Research

In this chapter, we will describe summary, conclusion, recommendation and scope for further development of our research.

CHAPTER 2

Literature Review

2.1 Introduction

This chapter introduces the key concepts and definition of Sentiment Analysis. Limitless usages of Social Media that impacts on users along with their mental health, overview of worldwide mental health present status, dark sides of Social Media are also discussed in this section.

2.2 Related Works

Sentiment analysis on the other hand is called opinion mining, review mining or appraisal extraction, attitude analysis that indicates the process of detecting, extracting, and classifying opinions, sentiments, and attitudes concerning different topics, as expressed in the textual pattern [7]. Nowadays sentiment analysis is using different types of algorithms, processing systems, even machine learning. It is a very powerful model for detecting any outcome over the given data or information. That's why we are using this.

Successful online data detection depends on the effective monitoring of content. Though the size of information on the Web makes it infeasible for moderators to examine manually all user-generated content. Though to categorize the social media sentiment there are many works that contain knowledgeable information. In [1] by using the unoptimised word n-gram-based classifier and keyword-matching system techniques to detect cyberbullying from the social networking site ASKfm by modelling posts written by bullies, victims, and bystanders of online bullying. According to them, Social media sites increase the possibility for the children being faced by threatening situations including grooming or sexually transgressive behavior, signs of depression, suicidal thoughts, and cyberbullying. Internet users are reachable almost 24 hours and are often able to stay nameless if they desire: this makes social media a convenient way for bullies to target their victims outside the school yard. Aggressive users show a similar attitude like spammers in terms of number of friends, and followers. For that to detect them in online platforms Twitter Machine learning algorithms has been also used [9].

In the meantime the term online harassment indicates the process of sending messages over electronic media to harm a victim psychologically. When this kind of message is sent many times

by the same person to the same victim in that situation it is called Cyber Bullying [20]. To prioritize the fact, to detect online harassment in one place binary classification [10] is used another place used libSVM with the linear kernel as classification tool [11] on the data from Social media.

The term harassment and sexual-harassment have a little bit of a difference. Sometimes to feel a person powerless by demonstrating a superior power, people often feel the urge to take the path of harassment by choosing sexual talk or harassment. In social media it also existed and even Linear SVM, LSF Framework [8] techniques are used to detect it from the data of Anonymous social media (ASMs). From the definition of Offensive sentences used to mock or insult somebody or a group of people even by hurting in such a way being aggression against some culture, a subgroup of the society, race, or ideology in a tirade. Another term related with offensive is hate speech that is a difficult matter to define. From Several online sources, even twitter to detect offensive languages using LDA [13] and three-level classifications on training data by using the IAL Dictionary [15]. In [14] the classifications of hate speech tend to reflect their own subjective biases by using logistic regression, random forests Naive Bayes, decision trees, and linear SVMs. People categorized racist and homophobic slurs as hateful, on the other hand just declare sexist language as merely offensive. Cyber-victimized words are becoming a threat to our mental health. It creates mental pressure, even those who can't accept it attempt suicide. Worldwide suicide is considered as a public health problem. Every year more than 30,000 suicide deaths occurs in the United States and nearly 1 million suicide deaths in worldwide [17,18]. Nowadays a phenomenon termed is cyber bullycide, the suicide indirectly or directly influenced by horrible experiences aggressive side from online [19]. To prevent those, finding out the places where the cyber-victimized word is used most is necessary.

2.3 Research Summary

Developing a model for doing the detection the cyber-victimization is being tried which can be seen in future years using some different and effective algorithms of machine learning. Some related work was done by using different types of algorithms over their different types of collected data and information. Some detect cyber-bullying [1, 9], some detect cyber-harassment [10, 11], some sexual harassment [8], some detect offensive words [13, 15], even hate speech [14] also tried to be classified. In this research, the parameters that we use in our dataset to detect cyber-

victimized words, are using some machine learning techniques. After that, cross-validation can be used to differentiate the accuracy rate of different algorithms.

2.4 Scope of the Problem

In every research, there can be a number of problems. Ours is also not exceptional. We collected data from various sites. Then the term pre-processing collected data to smooth our working flow. Even finding a suitable dictionary as the training data set that contains the words that can help us. Choosing a model that can give the desired result that we need at the end of the simulation. If we can't do this properly, our output won't be appropriate and won't be similar to the real-life scenario. If we want to reduce the obstacles we have to work carefully.

2.5 Challenges

No one is perfect in this world. So to build up an appropriate project also tough work with a large amount of data. Successful online data detection depends on the effective monitoring of content. Though the size of information on the Web makes it infeasible for moderators to examine manually all user-generated content. When a sentence is related to emotion, sometimes it seems difficult to classify any sentence according to our literal expression.

For example, sometimes we tweet, or comment, or create a blog to make fun with our close friends with a sarcastic word, or to express the wrongdoing of our society head's, or our present situation, that may also contain words that are offensive or harassed or bullying type. But it makes it difficult to classify the sentences because of those bearing opposite meanings than we classification by the cyber-victimized model that we proposed.

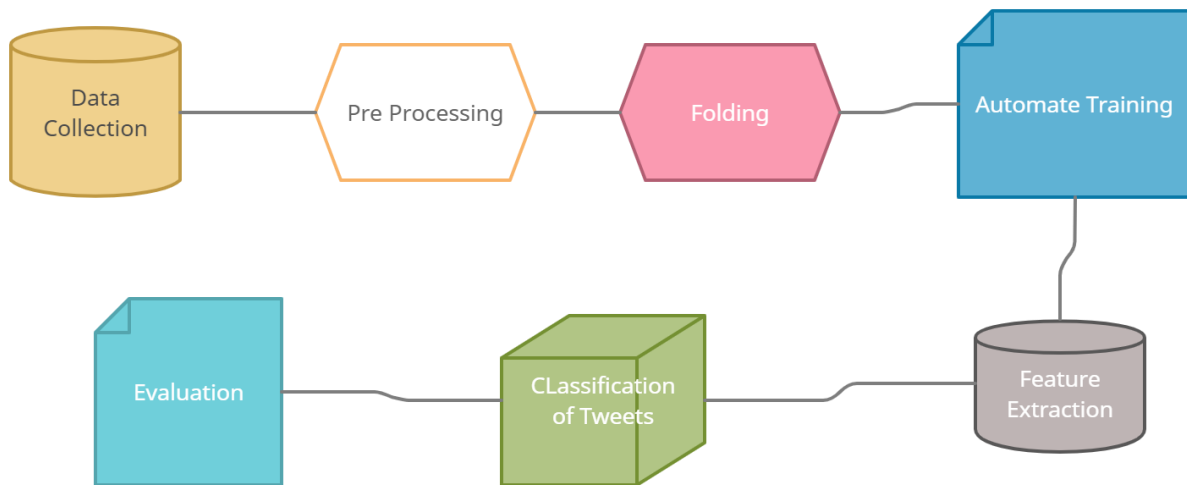
CHAPTER 3

Research Methodology

3.1 Introduction

In this section we will be describing the path or methods we have been following for the investigation. We have used sentiment analysis for our detection. We have used WordCloud for this detection.

3.2 Proposed Model



Block Diagram of Methodology

Figure 3.2.1 Block Diagram of methodology

3.3 Research Subject and Instrumentation

Sentiment classification was performed on the basis of syntactic and stylistic features such as word-length distributions, vocabulary richness measures, character- and word-level lexical features, and special-character frequencies [4], in particular, to determine if the attitude of the writer towards a particular subject is positive, negative, or neutral. It is a mixture of the encoding of natural languages, text interpretation and linguistic computation.

A sentence is considered positive in this method if it has positive keywords, and if it has negative keywords, it is considered negative. The relation between the sum of each form of content dictates the overall content's positivity and negativity. This research tends to provide an algorithm that can assist in the analysis of words that can lead to the identification of crime, especially on social sites. We are using the sentiment analysis technique of machine learning for our study. Naive Bayes, Support Vector Machine and Neural Network are the three algorithms that we're using. An initial training data set is required for Naive Bayes and Support Vector Machine, which must be labeled with positive, negative and neutral sentiment accordingly. For that, instead of manually classifying 11,500 tweets on our own, we use the lexicon-based textblob library solution.

3.4 Data Collection Procedure

Twitter is one of the largest public and freely available sources of data. It allows one's voice to reach the globe instantly, especially in places where there are barley channels for venting and expressing opinions [12]. Tweets were tracked using specific keywords that were likely to indicate harassment. We've retrieved text tweets from Twitter to compile a corpus of objective posts.

3.5 Pre-Processing

First of all, the spelling errors for tweets were corrected, as many individuals prefer to correct. We later deleted from the tweets all mentions, links, and unused blank spaces. Stop terms which are usually deemed useless. Using some regular expressions, we've created a text-cleaning feature.

3.6 Folding up

From the collected dataset, we have used eighty percent for training and 20 percent for testing purposes. It is best practice to use: sixty percent for training, twenty percent for cross-validation, twenty percent for testing. Out of 15,000 tweets, 11,500 were used to train the classifier, and 3,500 were used.

3.7 Classifier Automated Training Collection

Analysis of machine-learning sentiment involves a collection of tweets labeled positive, negative or objective, which can be produced by hand by manually labeling tweets. Although this will produce a highly accurate lexicon, hundreds of hours of work are needed. By collecting hundreds of thousands of tweets and running an algorithm through those tweets.

1. *Initialize WordList*

2. *Initialize get Analysis function with parameter score*

- 2.1. *If score < 0:*

- 2.1.1. *Return 'Negative'*

- 2.2. *If score == 0:*

- 2.2.1. *Return 'Neutral'*

- 2.3. *Else:*

- 2.3.1. *Return 'Positive'*

After labelling each tweets with their polarities, we tried something different for storing the data. The schema of the twitter data comprised of tweet, sentiment, geolocation, username, full name.

3.8 Extraction of Function

I. Ignore phrases - like auxiliary verbs and articles or any halper words. These terms are sentence helpers and do not imply any emotion. That's why these kinds of terms have been withdrawn. We create a text file with all possible words and ignore them by removing them from the function.

II. Replicating letters - Sometimes people use letters repeatedly to stress their emotion. E.g. very, verrrrrrryyyy for 'very'. We have looked for these kinds of repeated letters and fixed them with a custom built function.

III. Punctuation - We have removed punctuation from sentences such as comma, single/double quote, question marks at the start and end of each word.

IV. Remove Decoration – From each tweet, we have removed mentions of decorations. If a function word is present, we have classed it as 1 for each tweet, otherwise we have marked it as 0. We have some 1s and 0s after that, and a tweet on the basis of this pattern will be labeled as positive, neutral or negative. To check for new tweets, the function words will be extracted as above, and we have one more pattern of 0s and 1s and the classifiers will be able to predict the feeling of the tweet based on the learned model. We have used bigram function terminology for Naive Bayes, as bigram is recognized to be more descriptive than unigrams.

3.9 Research and Classification

The test tweets have been preprocessed and the features were extracted after the classification was trained, and have been run through the classification to detect the sentiments by polarity. We have used it to compare the accuracy of the classifiers after the detection of polarity.

3.10 Evaluation

The accuracy of the classifier, accuracy, recall and f-score of the positive, negative and neutral tweets from the dataset have been calculated after the implementation of the three machine learning approaches to evaluate the most accurate results and is preferable over the other approaches.

3.11 Implementation

We have tried to get the tweets from Twitter. For the implementation purpose, we have used latitude = 37.7186409, longitude = -97.2963301, radius = 5 km, and then request the Twitter API to get the live data from Wichita State University, Kansas, USA that has been posted just the time we requested the API to fetch the tweets. After finding the recent tweets from the location we have implemented the analysis algorithm to find the positivity and negativity from the tweets. Our implementation results are given below:

```
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
```

Figure 3.11.1 Python Modules

	tweet	Location
0	@RyanHernandez A dangerous find	Kansas
1	RT @DeniseH66111046: @tonyschwartz #ImpeachTru...	Kansas
2	2021 is finally here! This is a great piece t...	Kansas
3	RT @GoShockersMBB: A big matchup with Houston ...	Kansas
4	hmmm...https://t.co/l7Rnk9mbTZ	Kansas
5	As far as we have come and as much as we discu...	Kansas
6	Like... can you not? @UniWatch https://t.co/Xh...	Kansas
7	@ICTBiz_dmccoy with news of a @KochIndustries...	Kansas
8	RT @NAIMartens: The new @HuttonBuilds Headquar...	Kansas
9	@kayla_tharp No birth control is a trap	Kansas
10	RT @Greater_Wichita: Partnership investor @Hut...	Kansas
11	Resolution revolution: Less is more https://t....	Kansas
12	@kyrapractor Life do be happening	Kansas
13	Research shows folic acid supplements can help...	Kansas
14	Congrats to Interior Designer Angi Womeldorff ...	Kansas

Figure 3.11.2 Raw Tweets from specific location (Kansas)

	tweet	Location
0	A dangerous find	Kansas
1	: ImpeachTrumpNow TonySchwartzWarnedUs \n ...	Kansas
2	2021 is finally here! This is a great piece t...	Kansas
3	: A big matchup with Houston on deck. Catch _l...	Kansas
4	hmmm...	Kansas
5	As far as we have come and as much as we discu...	Kansas
6	Like... can you not?	Kansas
7	_dmccoy with news of a investment.	Kansas
8	: The new Headquarters is finished. Congrats ...	Kansas
9	_tharp No birth control is a trap	Kansas

Figure 3.11.3 Tweets being polished using regex

	tweet	Location	Subjectivity	Polarity
0	A dangerous find	Kansas	0.900000	-0.600000
1	: ImpeachTrumpNow TonySchwartzWarnedUs \n ...	Kansas	0.000000	0.000000
2	2021 is finally here! This is a great piece t...	Kansas	0.551136	0.234091
3	: A big matchup with Houston on deck. Catch _l...	Kansas	0.100000	0.000000
4	hmmm...	Kansas	0.000000	0.000000
5	As far as we have come and as much as we discu...	Kansas	0.618750	0.193750
6	Like... can you not?	Kansas	0.000000	0.000000
7	_dmccoy with news of a investment.	Kansas	0.000000	0.000000
8	: The new Headquarters is finished. Congrats ...	Kansas	0.553030	0.357576
9	_tharp No birth control is a trap	Kansas	0.000000	0.000000
10	_Wichita: Partnership investor has completed ...	Kansas	0.454545	0.136364
11	Resolution revolution: Less is more	Kansas	0.283333	0.166667
12	Life do be happening	Kansas	0.000000	0.000000
13	Research shows folic acid supplements can help...	Kansas	0.000000	0.000000
14	Congrats to Interior Designer Angi Womeldorff ...	Kansas	0.000000	0.000000

Figure 3.11.4 Tweets with polarity and subjectivity

	tweet	Location	Subjectivity	Polarity	Analysis
0	A dangerous find	Kansas	0.900000	-0.600000	Negative
1	: ImpeachTrumpNow TonySchwartzWarnedUs \n ...	Kansas	0.000000	0.000000	Neutral
2	2021 is finally here! This is a great piece t...	Kansas	0.551136	0.234091	Positive
3	: A big matchup with Houston on deck. Catch _l...	Kansas	0.100000	0.000000	Neutral
4	hmmm...	Kansas	0.000000	0.000000	Neutral
5	As far as we have come and as much as we discu...	Kansas	0.618750	0.193750	Positive
6	Like... can you not?	Kansas	0.000000	0.000000	Neutral
7	_dmccoy with news of a investment.	Kansas	0.000000	0.000000	Neutral
8	: The new Headquarters is finished. Congrats ...	Kansas	0.553030	0.357576	Positive
9	_tharp No birth control is a trap	Kansas	0.000000	0.000000	Neutral
10	_Wichita: Partnership investor has completed ...	Kansas	0.454545	0.136364	Positive
11	Resolution revolution: Less is more	Kansas	0.283333	0.166667	Positive
12	Life do be happening	Kansas	0.000000	0.000000	Neutral
13	Research shows folic acid supplements can help...	Kansas	0.000000	0.000000	Neutral
14	Congrats to Interior Designer Angi Womeldorff ...	Kansas	0.000000	0.000000	Neutral

Figure 3.11.5 Tweets being analyzed

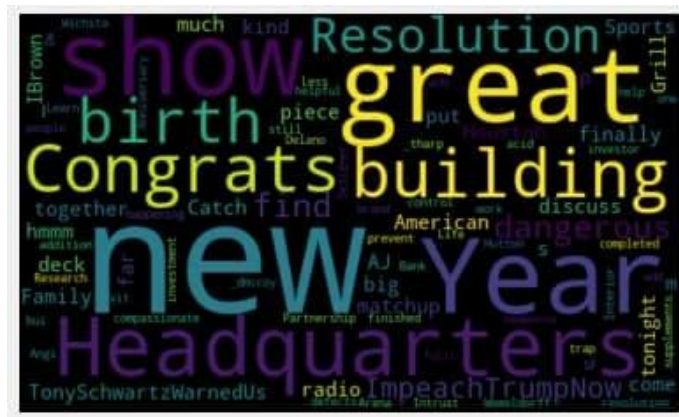


Figure 3.11.6 WordCloud

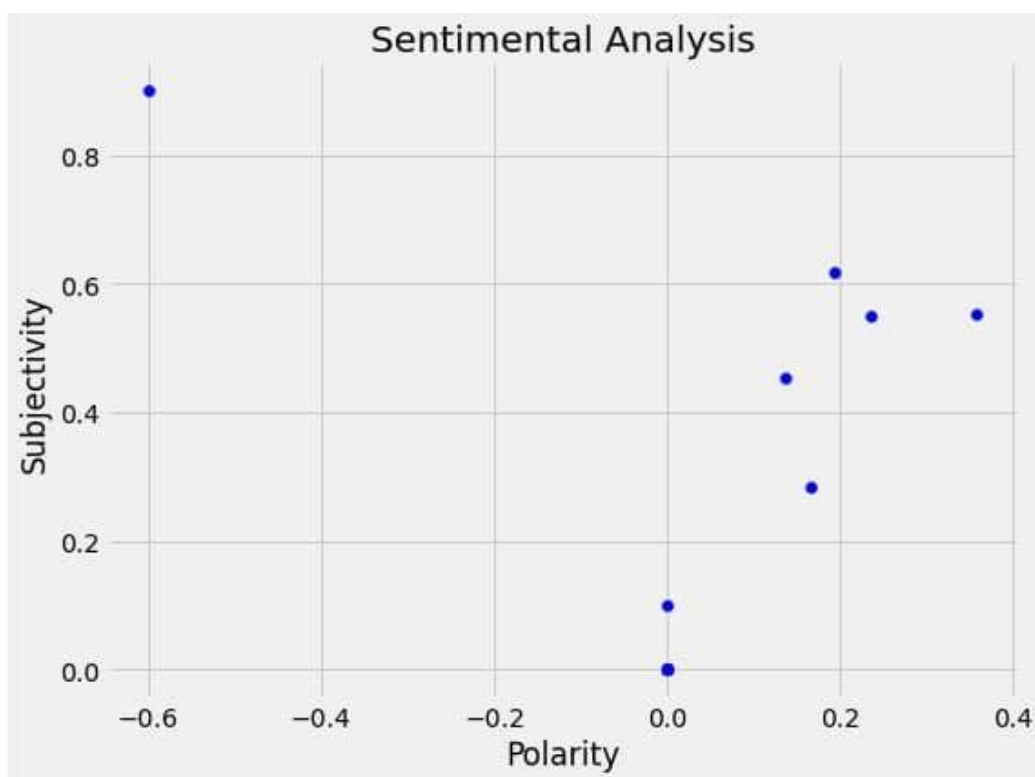


Figure 3.11.7 Expected Output

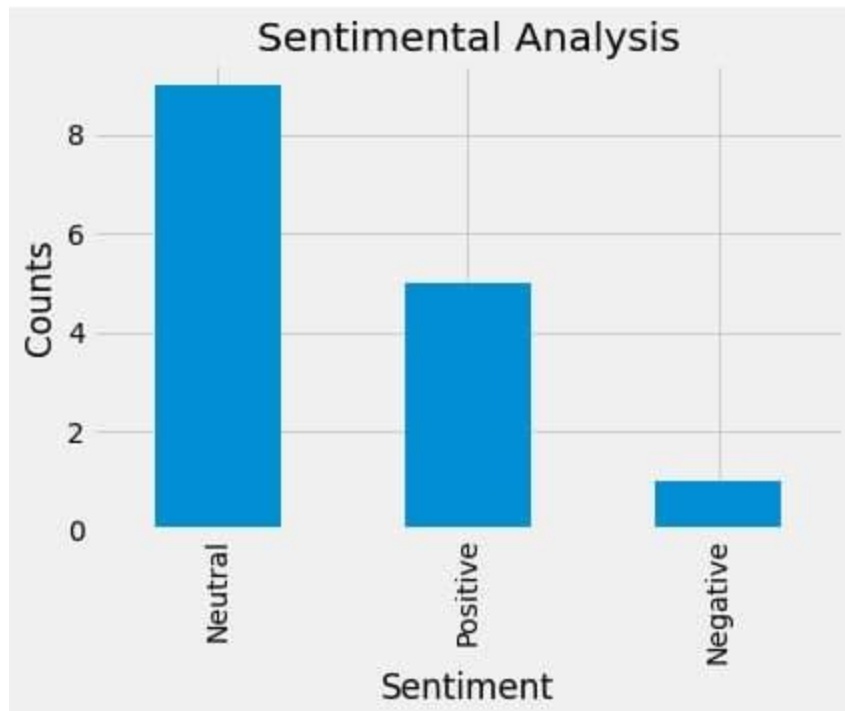


Figure 3.11.8 Expected Output

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

Supervised learning is used for the majority of functional machine learning.

Supervised learning is where you have input variables (x) and an output variable (Y) and learn the mapping function from the input to the output by using an algorithm.

$Y=f(x)$

The aim is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

It is called supervised learning since, as an instructor supervising the learning process, the process of an algorithm learning from the training dataset can be considered. The correct answers are known, the algorithm iteratively makes predictions on the training data and the instructor corrects them. Learning stops when an appropriate level of output is reached by the algorithm.

Regression and classification issues can be further categorized into supervised learning problems.

- **Classification:** A classification issue is when a category is the output element, such as "**red**" or "**blue**" or "**disease**" and "**no illness**" or "**positive**" or "**negative**" in our case.
- **Regression:** A regression problem is when a real value, such as "**dollars**" or "**weight**" is the output variable.

Our conflict falls under the heading of classification because we have to classify our findings into either **Positive** or **Negative** categories.

[6]There use the raw sentiment scores to track two trends over time:

Polarity: Is the sentiment associated with the entity positive or negative?

Subjectivity: How much sentiment (of any polarity) does the entity garner?

Another type of machine learning algorithm is called **Unsupervised Machine Learning**, where you have input data, but no corresponding output variables. In order to learn more about the data, the purpose of unsupervised learning is to model the underlying structure or distribution of the data. But for this issue's argument, that is of no concern to us.

So we see from the above dataset division that we will use features for our **Machine Learning Models** from the **Bag-of-Words** and **TF-IDF**.

We typically use various models to see which suits our dataset best, and then we use that model to predict test data outcomes.

We will be using 3 different versions here.

- **Logistic Regression**
- **XGBoost**
- **Decision Trees**

Then we will compare their performance and choose the best possible model to predict results on our test data with the best possible feature extraction technique.

Importing f1_score from sklearn

We will use F1 Score throughout to assess our model's performance instead of accuracy.

4.2 Experimental Results

Using Bag-of-Words

```
x_train_bow,x_valid_bow,y_train_bow,y_valid_bow = train_test_split(train_bow,train['label'],test_size=0.3,random_state=2)
```

Figure 4.2.1 Experimental Result for Using Bag-of-Words

Using TF-IDF Features

```
x_train_tfidf,x_valid_tfidf,y_train_tfidf,y_valid_tfidf = train_test_split(train_tfidf_matrix,train['label'],test_size=0.3,random
```

Figure 4.2.2 Experimental Result for Using TF-IDF Features

4.3 Logistic Regression

Output of Bag-of-Words Features

```
array([[9.86764207e-01, 1.32357931e-02],
       [9.99593900e-01, 4.06099967e-04],
       [9.13482248e-01, 8.65177523e-02],
       ...,
       [8.95267697e-01, 1.04732303e-01],
       [9.59786822e-01, 4.02131779e-02],
       [9.67596054e-01, 3.24039462e-02]])
```

Figure 4.3.1 Output of Bag-of-Words Features for Logistic Regression

Output of TF-IDF Features

```
array([[0.98505728, 0.01494272],
       [0.97945171, 0.02054829],
       [0.94191583, 0.05808417],
       ...,
       [0.98630046, 0.01369954],
       [0.96766546, 0.03233454],
       [0.99058281, 0.00941719]])
```

Figure 4.3.2 Output of TF-IDF Features for Logistic Regression

4.4 XGBoost

Output of Bag-of-Words Features

```
array([[9.91007686e-01, 8.99230968e-03],
       [9.99111533e-01, 8.88437440e-04],
       [9.27937031e-01, 7.20629618e-02],
       ...,
       [8.93149197e-01, 1.06850825e-01],
       [9.27937031e-01, 7.20629618e-02],
       [9.27937031e-01, 7.20629618e-02]], dtype=float32)
```

Figure 4.4.1 Output of Bag-of-Words Features for XGBoost

Output of TF-IDF Features

```
array([[0.98698187, 0.01301816],
       [0.98737913, 0.01262084],
       [0.95263565, 0.04736437],
       ...,
       [0.98640716, 0.01359282],
       [0.9865412 , 0.01345879],
       [0.99581224, 0.00418777]], dtype=float32)
```

Figure 4.4.2 Output of TF-IDF Features for XGBoost

4.5 Decision Tree

Output of Big-of-Words

```
array([[1., 0.],
       [1., 0.],
       [1., 0.],
       ...,
       [1., 0.],
       [1., 0.],
       [1., 0.]])
```

Figure 4.5.1 Output of Bag-of-Words Features for Decision Tree

Output of TF-IDF Features

```
array([[1., 0.],
       [1., 0.],
       [1., 0.],
       ...,
       [1., 0.],
       [1., 0.],
       [1., 0.]])
```

Figure 4.5.2 Output of TF-IDF Features for Decision Tree

4.6 Model Comparison

	1	2	3	4	5	6
Model	LogisticRegression(Bag-of-Words)	XGBoost(Bag-of-Words)	DecisionTree(Bag-of-Words)	LogisticRegression(TF-IDF)	XGBoost(TF-IDF)	DecisionTree(TF-IDF)
F1_Score	0.57308	0.581818	0.517134	0.587727	0.571201	0.544462

Figure 4.6.1 Model Comparison-1

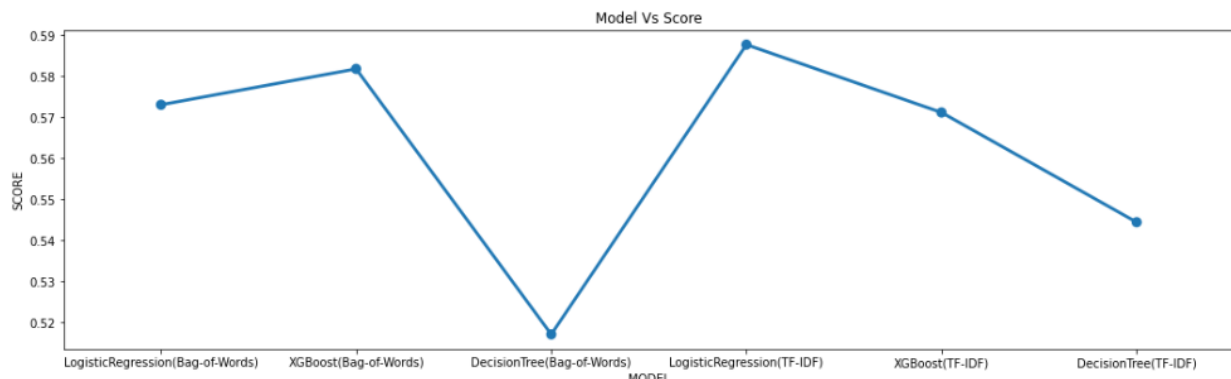


Figure 4.6.2 Model Comparison-2

4.7 F1 Score Compare

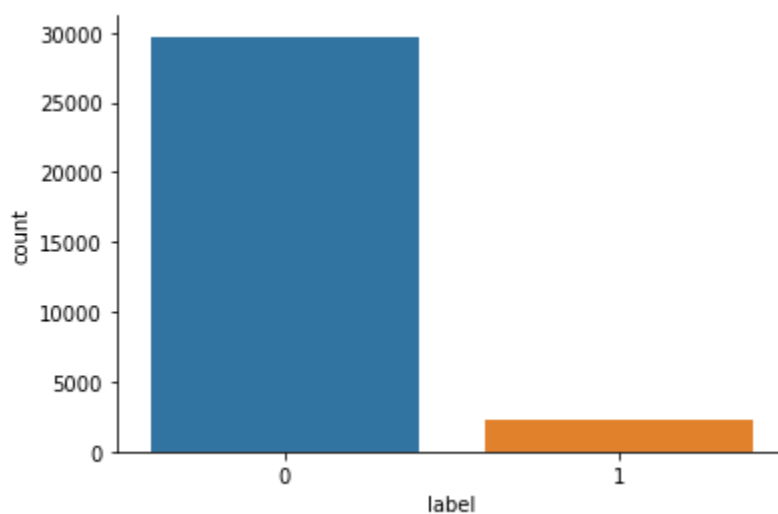


Figure 4.7.1 F1 Score Compare

4.8 Summary

Instead of Accuracy, why use F1-Score?

- We see how imbalanced our dataset is from the counting plot produced above. In contrast to the values with labels: 1 i.e. negative feelings, we can see that the values with label: 0 i.e. positive feelings are very high in number.
- So while we hold accuracy as our measurement criterion there may be situations where we may encounter high number of false positives.

Precision & Recall

- The percentage of your results that are important is precision.
- Recall refers to the percentage correctly classified by your algorithm of the total applicable data.

$$\begin{aligned}\text{Precision} &= \frac{\text{True Positive}}{\text{Actual Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \\ \text{Recall} &= \frac{\text{True Positive}}{\text{Predicted Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \\ \text{Accuracy} &= \frac{\text{True Positive} + \text{True Negative}}{\text{Total}}\end{aligned}$$

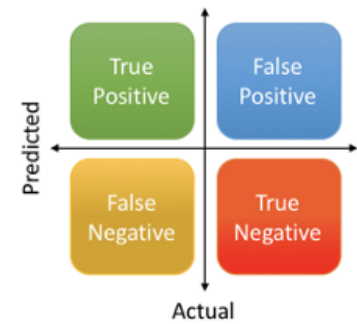


Figure 4.8.1 Precision & Recall

We still face a Precision-Recall trade-off, i.e. Low recall and vice versa are given by High Precision.

- In most of the issues, depending on the issue you are trying to solve, you may either give greater priority to optimizing precision, or recall. There is a simpler metric in general that takes both precision and recall into account, so you can try to optimize this number to make your model better. The F1-score, which is simply the harmonic mean of accuracy and recall, is known as this metric.

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

- So this metric is a lot simpler and easy to deal with, since instead of juggling two different scores, you just have to optimize one score.

CHAPTER 5

Summary, Conclusions, Recommendations and Implication for Future Research

5.1 Summary of the Study

In the 3rd world, almost everyone is connected by web 2.0 and there are many ways to create communication with each other. Social media is gaining popularity for communication, sharing thought, and so on. But some culprit or mentally sick people use different types of harmful words in this platform that create an environment uncomfortable for many people. Those are cyber-victimized words. It can be related to harassing, bullying, sexual-harassing, offensive, or hate speech, or so more. But the root of the problem is the main of all problems. Here we want to aware people of the topic and create awareness. We use machine learning and try to find a result. Finally, we have concluded that we could work with detecting those words and that would be so beneficial for people. So a model for detecting cyber-victimized words using a machine language algorithm has been developed. After collecting the data, preprocessing and finding the best algorithm was the main challenge for our research work.

5.2 Conclusions

As shown in our research, we have tried to show how different machine learning performs to detect the sentiments of cyberbullying from social media. We have successfully detected using three approaches to indicate the negative tweets by bullying and harassment words. Among them, Big-of-Words had the best results, and TF-IDF had the second best. We browsed a dataset of tweets, those containing cyber-victimization, in this paper. In addition, efficacy about that automated detection of cyber bullying has been explored incessantly by running experiments to obtain the best results. Our analysis was finally able to show the reason for each of their performances. We would be able to bring this research to perfection with further work.

5.3 Recommendations

Machine Learning (ML) is the logical investigation of calculations and factual models that PC frameworks use to play out a particular assignment without utilizing unequivocal directions, depending on examples and surmising. It has been viewed as a subset of artificial intelligence. Machine Learning calculations assemble a numerical model dependent on test information, known as training data, to settle on expectations or choices without being expressly modified to play out the work. In our work, we Big-of-Words **to detect the cyber-victimized words from social media because it obtains a much better accuracy rate for our expected goal.** The three algorithms are- **Logistic Regression, XGBoost, and Decision Trees.** While doing our research work over our dataset similar accuracy rate has been found from these different algorithms. Though, it can vary from dataset to dataset, of the accuracy of algorithms for our dataset the accuracy rate is similar for the used algorithms. When we increase or decrease our test numbers the accuracy of our model gets higher than other algorithms. So we have a recommendation for you to use or try Big-of-Words to detect research work.

5. 4 Implication for Further Study

In this research work, a model for detecting Cyber-victimization words has been developed by using different machine learning algorithms. In the future, we want to go far with this plan. We know finding a way always shows the path to the solution. So in our work, we want to identify the location. That will be beneficial for us because we are connected with each other via social media, and removing the trash minded people from there will make a healthy social media environment.

REFERENCES

- [1] Van Hee, C., Jacobs, G., Emmery, C., Desmet, B., Lefever, E., Verhoeven, B., ... & Hoste, V. (2018). Automatic detection of cyberbullying in social media text. *PloS one*, 13(10), e0203794.
- [2] McKenna, K. Y., & Bargh, J. A. (2000). Plan 9 from cyberspace: The implications of the Internet for personality and social psychology. *Personality and social psychology review*, 4(1), 57-75.
- [3] Gross, E. F., Juvonen, J., & Gable, S. L. (2002). Internet use and well-being in adolescence. *Journal of social issues*, 58(1), 75-90.
- [4] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.
- [5] Van Hee, C., Lefever, E., Verhoeven, B., Mennes, J., Desmet, B., De Pauw, G., ... & Hoste, V. (2015). Automatic detection and prevention of cyberbullying. In *International Conference on Human and Social Analytics (HUSO 2015)* (pp. 13-18). IARIA.
- [6] Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *Icwsn*, 7(21), 219-222.
- [7] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, *Decision Support Systems* 53 (2012) 675–679.
- [8] Nova, F. F., Rifat, M. R., Saha, P., Ahmed, S. I., & Guha, S. (2019, January). Online sexual harassment over anonymous social media in Bangladesh. In *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development* (pp. 1-12).
- [9] Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. (2017, June). Mean birds: Detecting aggression and bullying on twitter. In *Proceedings of the 2017 ACM on web science conference* (pp. 13-22).
- [10] Bretschneider, U., Wöhner, T., & Peters, R. (2014). Detecting online harassment in social networks.
- [11] Yin, D., Xue, Z., Hong, L., Davison, B. D., Kontostathis, A., & Edwards, L. (2009). Detection of harassment on web 2.0. *Proceedings of the Content Analysis in the WEB*, 2, 1-7.
- [12] Alhumoud, S., Albuhairi, T., & Alohaideb, W. (2015, November). Hybrid sentiment analyser for Arabic tweets using R. In *2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)* (Vol. 1, pp. 417-424). IEEE.
- [13] Xiang, G., Fan, B., Wang, L., Hong, J., & Rose, C. (2012, October). Detecting offensive tweets via topical feature discovery over a large scale twitter corpus. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 1980-1984).
- [14] Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *arXiv preprint arXiv:1703.04009*.

- [15] Razavi, A. H., Inkpen, D., Uritsky, S., & Matwin, S. (2010, May). Offensive language detection using multi-level classification. In *Canadian Conference on Artificial Intelligence* (pp. 16-27). Springer, Berlin, Heidelberg.
- [16] Luxton, D. D., June, J. D., & Fairall, J. M. (2012). Social media and suicide: a public health perspective. *American journal of public health*, 102(S2), S195-S200.
- [17] Centers for Disease Control and Prevention. Suicide and self-inflicted injury. Available at: <http://www.cdc.gov/nchs/fastats/suicide.htm>. Updated January 27, 2012. Accessed November 9, 2011.
- [18] World Health Organization. Mental health: suicide prevention. Available at: http://www.who.int/mental_health/prevention/en. Accessed November 9, 2011
- [19] Hinduja, S., & Patchin, J. W. (2009). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying*. Thousand Oaks, CA: Sage Publications (Corwin Press).
- [20] Tokunaga, R. S. 2010. "Review: Following You Home from School: A Critical Review and Synthesis of Research on Cyberbullying Victimization," *Computers in Human Behaviour* (26:3), pp. 277-287.
- [21] Saraban Tahura Zaman, Lisa Gansheimer, Sofia Bordin Rolim, and Tahrira Mridha. 2017. *Legal Action on Cyber Violence Against Women*. Dhaka: Bangladesh Legal Aid Services Trust (BLAST). <http://www.blast.org.bd/>

PLAGIARISM REPORT

Defense 2021 final version

ORIGINALITY REPORT

21%

SIMILARITY INDEX

17%

INTERNET SOURCES

11%

PUBLICATIONS

11%

STUDENT PAPERS

PRIMARY SOURCES

1	towardsdatascience.com Internet Source	3%
2	Mifta Sintaha, Moin Mostakim. "An Empirical Study and Analysis of the Machine Learning Algorithms Used in Detecting Cyberbullying in Social Media", 2018 21st International Conference of Computer and Information Technology (ICCIT), 2018 Publication	3%
3	Submitted to Daffodil International University Student Paper	2%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	journals.plos.org Internet Source	1%
6	Submitted to Jacobs University, Bremen Student Paper	1%
7	Submitted to Macquarie University Student Paper	1%