

A

Major Project

On

CYBERBULLYING DETECTION ON SOCIAL MEDIA USING MACHINE LEARNING

(Submitted in partial fulfilment of the requirements for award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

BY

Aashish Toshniwal

(177R1A05C3)

Vinay Nagarale

(177R1A05H5)

Harish Gopu

(187R5A0510)

Under the Guidance of

Mr. K. Mahesh

(Asst. professor)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGC Act.1956, Kandlakoya (V), Medchal Road, Hyderabad-501401.

2017-2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



CERTIFICATE

This is to certify that the project entitled “**CYBERBULLYING DETECTION ON SOCIAL MEDIA USING MACHINE LEARNING**” being submitted by **AASHISH TOSHNIWAL (177R1A05C3)**, **VINAY NAGARALE (177R1A05H5)** & **HARISH GOPU (187R5A0510)** in partial fulfilment of the requirements for the award of the degree of B. Tech in Computer Science and Engineering of the Jawaharlal Nehru Technological University Hyderabad, during the year 2019-2020. It is certified that they have completed the project satisfactorily.

INTERNAL GUIDE

Mr. K. Mahesh

(ASST. PROFESSOR)

DIRECTOR

Dr. A. Raji Reddy

HOD

Dr. K. Srujan Raju

EXTERNAL EXAMINER

Submitted for viva voce Examination held on _____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. We take this opportunity to express my profound gratitude and deep regard to my guide

Mr. K. Mahesh, Asst. Professor for his exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to Project Review Committee (PRC) Coordinators: **Mr. J. Narasimha Rao, Mr. B. P. Deepak Kumar, Mr. K. Murali, Dr. Suwarna Gothane and Mr. B. Ramji** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to the Head of the Department **Dr. K. Srujan Raju** for providing excellent infrastructure and a nice atmosphere for completing this project successfully.

We are obliged to our Director **Dr. A. Raji Reddy** for being cooperative throughout the course of this project. We would like to express our sincere gratitude to our Chairman Sri. Ch. Gopal Reddy for his encouragement throughout the course of this project

The guidance and support received from all the members of **CMR TECHNICAL CAMPUS** who contributed and who are contributing to this project, was vital for the success of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity thank our family for their constant encouragement without which this assignment would not be possible. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project.

AASHISH TOSHNIWAL (177R1A05C3)

VINAY NAGARALE (177R1A05H5)

HARISH GOPU (187R5A0510)

ABSTRACT

From the day internet came into existence, the era of social networking sprouted. In the beginning, no one may have thought internet would be a host of numerous amazing services like the social networking. Today we can say that online applications and social networking websites have become a non-separable part of one's life. Many people from diverse age groups spend hours daily on such websites. Despite the fact that people are emotionally connected together through social media, these facilities bring along big threats with them such as cyberattacks, which includes cyberbullying. As social networking sites are increasing, cyber bullying is increasing day by day. To identify word similarities in the tweets made by bullies and make use of machine learning and can develop an ML model automatically detect social media bullying actions. However, many social media bullying detection techniques have been implemented, but many of them were textual based. Under this background and motivation, it can help to prevent the happen of cyberbullying if we can develop relevant techniques to discover cyberbullying in social media. A machine learning model is proposed to detect and prevent bullying on Twitter. Two classifiers i.e. SVM and Naïve Bayes are used for training and testing the social media bullying content.

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO.
Figure 3.1	Project Architecture	8
Figure 3.3	Use case diagram	10
Figure 3.4	Class diagram	11
Figure 3.5	Sequence diagram	12
Figure 3.6	Activity diagram	13

LIST OF SCREENSHOTS

SCREENSHOT NO.	SCREENSHOT NAME	PAGE NO.
Screenshot 5.1	Loading dataset and removing Repeated data	22
Screenshot 5.2	Testing accuracy using Naïve Bayes Algorithm	22
Screenshot 5.3	Testing accuracy using Support Vector Machine	23
Screenshot 5.4	Graph Showing Accuracy of both The Algorithms	23
Screenshot 5.5	Twitter database sample with index	24
Screenshot 5.6	Cyberbullying detected values	24
Screenshot 5.7	Output indicating cyberbullying Detection	25
Screenshot 5.8	Pie chart indicating 0 and 1 values	25

TABLE OF CONTENTS

	PAGE NO'S
ABSTRACT	i .
LIST OF FIGURES	ii.
LIST OF SCREENSHOTS	iii.
1.INTRODUCTION	1
2.SYSTEM ANALYSIS	2
2.1 EXISTING SYSTEM	3
2.2 PROPOSED SYSTEM	3
2.2.1 ADVANTAGES OF PROPOSED SYSTEM	4
2.3 FEASIBILITY STUDY	5
2.3.1 ECONOMIC FEASIBILITY	5
2.3.2 TECHNICAL FEASIBILITY	5
2.3.3 BEHAVIORAL FEASABILITY	5
2.4 HARDWARE AND SOFTWARE REQUIREMENTS	6
2.4.1 HARDWARE REQUIREMENTS	6
2.4.2 SOFTWARE REQUIREMENTS	6
3.ARCHITECTURE	7
3.1 PROJECT ARCHITECTURE	8
3.2 DESCRIPTION	9
3.3 USE CASE DIAGRAM	10
3.4 CLASS DIAGRAM	11
3.5 SEQUENCE DIAGRAM	12

3.6 ACTIVITY DIAGRAM	13
4.IMPLEMENTATION	14
4.1 SAMPLE CODE	15
5.SCREENSHOTS	21
6.TESTING	26
6.1 INTRODUCTION TO TESTING	27
6.2 TYPES OF TESTING	27
6.2.1 UNIT TESTING	27
6.2.2 INTEGRATION TESTING	27
6.2.3 FUNCTIONAL TESTING	28
6.3 TEST CASES	28
6.3.1 UPLOADING DATASET	28
7. CONCLUSION AND FUTURE SCOPE	29
7.1 CONCLUSION	30
7.2 FUTURE SCOPE	30
8.BIBLIOGTAPHY	31
8.1 GITHUB REPOSITORY LINK	32
8.2 REFERENCES	32
8.3 WEBSITES	32

1. INTRODUCTION

1. INTRODUCTION

Social networking sites are being widely used today for multiple purposes like entertainment, networking, etc. Social networking sites are a stop for multiple reasons to billions of people today. All the social media platforms require the consent of all the participating people. Communicating with people is no exception, as technology has changed the way people interact with a broader manner and has given a new dimension to communication. Many people are illegally using these communities. Many youngsters are getting bullied these days. Bullies use various services like Twitter, Facebook, and Email to bully people.

Cyberbullying is one of the most frequently happen Internet abuse and also a very serious social problem especially for teenager. Therefore, more and more researchers are devoting on how to discover and prevent the happen of cyberbullying, especially in social media. Cyberbullying is not just limited to creating a fake identity and publishing/posting some embarrassing photo or video, unpleasant rumours about someone but also giving them threats. The impacts of cyberbullying on social media are horrifying, sometimes leading to the death of some unfortunate victims.

Thus, a complete solution is required for this problem. Cyberbullying needs to stop. The problem can be tackled by detecting and preventing it by using a machine learning approach, this needs to be done using a different perspective.

2. SYSTEM ANALYSIS

2. SYSTEM ANALYSIS

SYSTEM ANALYSIS

System Analysis is the important phase in the system development process. The System is studied to the minute details and analysed. The system analyst plays an important role of an interrogator and dwells deep into the working of the present system. In analysis, a detailed study of these operations performed by the system and their relationships within and outside the system is done. A key question considered here is, “what must be done to solve the problem?” The system is viewed as a whole and the inputs to the system are identified. Once analysis is completed the analyst has a firm understanding of what is to be done.

2.1 EXISTING SYSTEM

Cyberbullying incidents are increasing day by day as technology rolls out. A large number of cyberbullying incidents are reported by companies each year. The existing system doesn't effectively classify and predict the tweets which is presented in the social media. The existing system uses partial automation i.e human intervention is required for the filtering. Moderators are assigned for each particular section of a website or social media page for moderating any bullying which makes the process slow as moderators cannot analyse every post in a world where there are millions of users of Social media. There are a few automated methods too but they are not capable of handling large data sets.

2.1.2 LIMITATIONS

- Doesn't Efficient for handling large volume of data.
- Theoretical Limits
- Incorrect Classification Results.
- Less Prediction Accuracy.

2.2 PROPOSED SYSTEM

The proposed model is introduced to overcome all the disadvantages that arises in the existing system. This system will increase the accuracy of the supervised classification results by classifying the data. An approach is proposed for detecting and preventing Twitter cyberbullying using Supervised Binary classification Machine Learning algorithms. Our model is evaluated on both Support Vector Machine and Naive Bayes. It enhances the performance of the overall classification results.

Support Vector Machine

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot). Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line).

Naïve Bayes

It is a classification technique based on Bayes’ Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as ‘Naive’.

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

2.2.1 ADVANTAGES OF PROPOSED SYSTEM

- High performance.
- Provide accurate prediction results.
- It avoid sparsity problems.
- Reduces the information Loss and the bias of the inference due to the multiple estimates.

2.3 FEASIBILITY STUDY

The feasibility of the project is analysed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. Three key considerations involved in the feasibility analysis are

- Economic Feasibility
- Technical Feasibility
- Social Feasibility

2.3.1 ECONOMIC FEASIBILITY

The developing system must be justified by cost and benefit. Criteria to ensure that effort is concentrated on project, which will give best, return at the earliest. One of the factors, which affect the development of a new system, is the cost it would require. The following are some of the important financial questions asked during preliminary investigation:

- The costs conduct a full system investigation.
- The cost of the hardware and software.
- The benefits in the form of reduced costs or fewer costly errors.

Since the system is developed as part of project work, there is no manual cost to spend for the proposed system. Also, all the resources are already available, it give an indication of the system is economically possible for development.

2.3.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. The developed system must have a modest requirement; as only minimal or null changes are required for implementing this system.

2.3.3 BEHAVIORAL FEASIBILITY

This includes the following questions: • Is there sufficient support for the users? • Will the proposed system cause harm? The project would be beneficial because it satisfies the objectives when developed and installed. All behavioural aspects are considered carefully and conclude that the project is behaviourally feasible.

2.4 HARDWARE & SOFTWARE REQUIREMENTS

2.4.1 HARDWARE REQUIREMENTS

This includes the following questions: • Is there sufficient support for the users? • Will the proposed system cause harm? The project would be beneficial because it satisfies the objectives when developed and installed. All behavioural aspects are considered carefully and conclude that the project is behaviourally feasible.

- Processor: Min. Pentium IV or later
- Hard Disk: 200GB
- RAM: 4GB or Higher

2.4.2 SOFTWARE REQUIREMENTS

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements,

- Operating System: Windows, Linux or Mac
- Language: Python
- IDE: Anaconda Spyder

3. ARCHITECTURE

3. ARCHITECTURE

3.1 PROJECT ARCHITECTURE

This Project architecture shows the procedure followed for cyberbullying detection using machine learning, starting from input to final prediction.

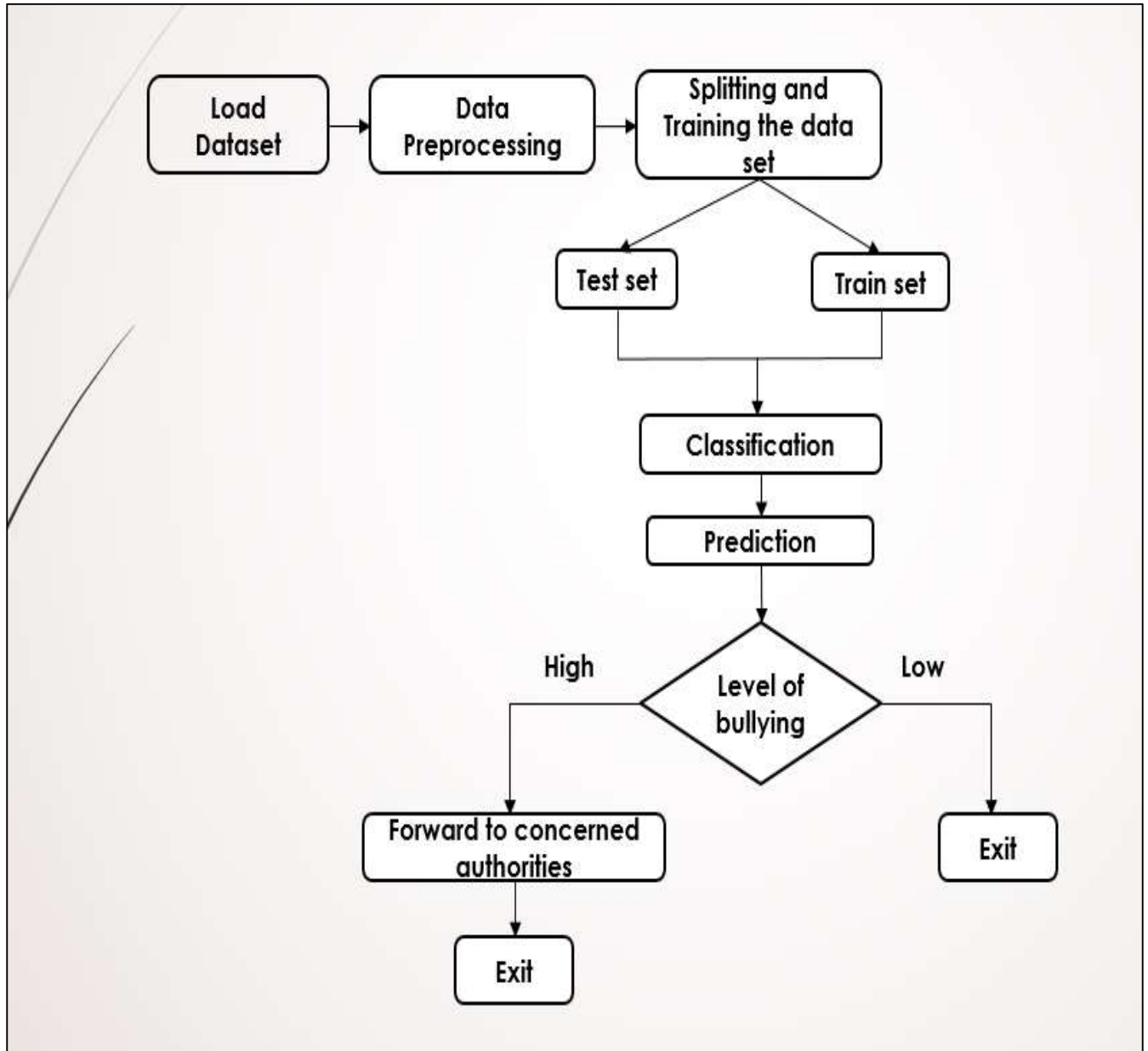


Fig 3.1. System Architecture

3.2 DESCRIPTION

Input Data: Input data is generally given in .csv format where the data is fetched and mapped in the data framed from the source columns.

Reading Data: library files are used to read the data into the data frame.

Separating Features: In this following step we are going to separate the features which we take to train the model by giving the target value i.e. 1/0 for the particular of features.

Normalization: Normalization is a very important step while we are dealing with the large values in the features as the higher bit integers will cost high computational power and time. To achieve the efficiency in computation we are going to normalize the data values.

Training and test data: Training data is passed to the Decision tree classifier to train the model. Test data is used to test the trained model whether it is making correct predictions or not.

Decision Tree Classifier: the purpose of choosing the decision tree classifier for this project the efficiency and accuracy that we have observed when compared to other classifiers.

3.3 USE CASE DIAGRAM

In the use case diagram, we have basically two actors who are the user and the administrator. The user has the rights to login, access to resources and to view the details. Whereas the administrator has the login, access to resources of the users and also the right to update and remove the crime details, and he can also view the user files.

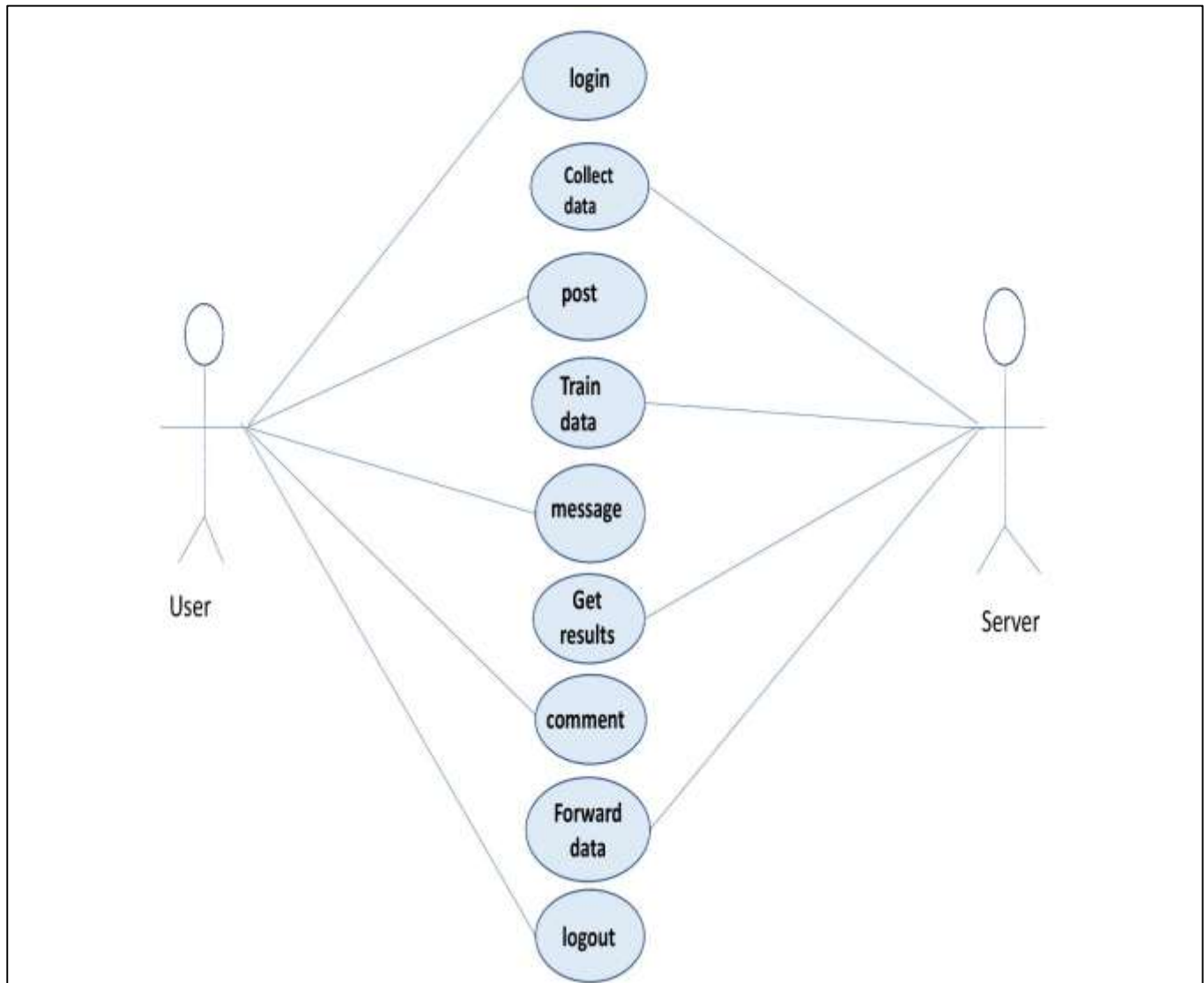


Fig 3.3 Use Case Diagram

3.4 CLASS DIAGRAM

Class Diagram is a collection of classes and objects.

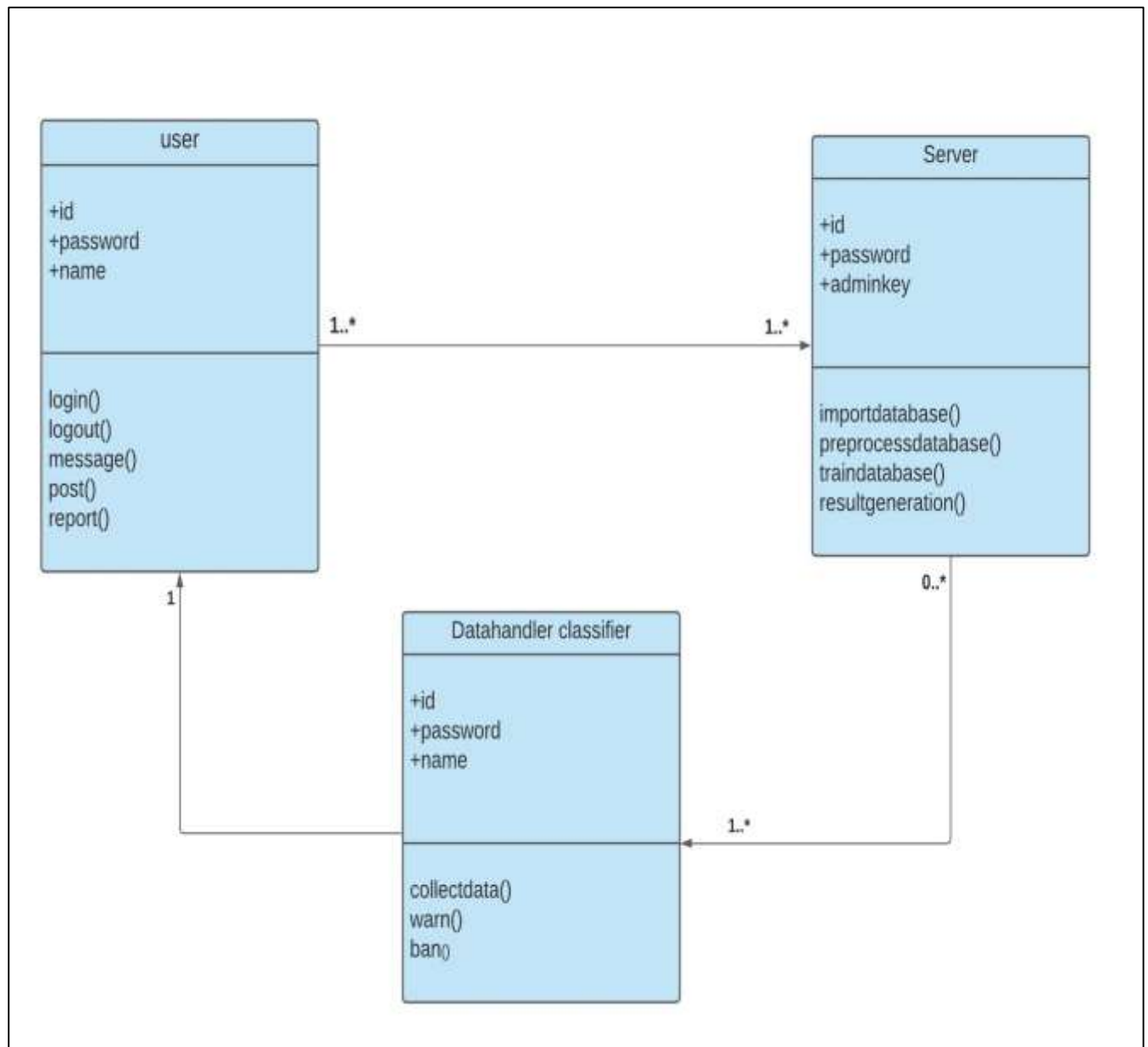


Fig 3.4 Class Diagram

3.5 SEQUENCE DIAGRAM

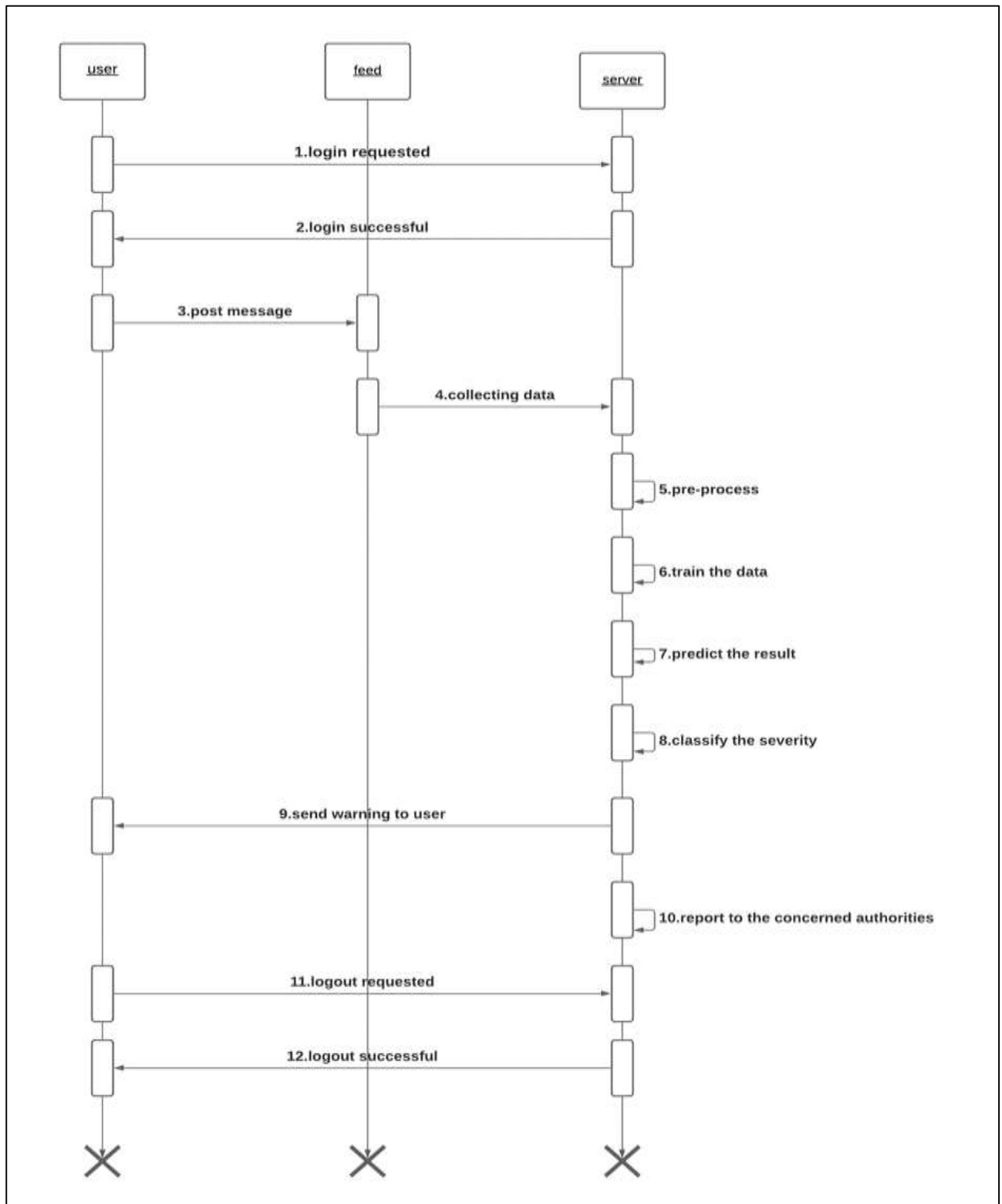


Fig 3.5 Sequence Diagram

3.6 ACTIVITY DIAGRAM

It describes about the flow of activity states.

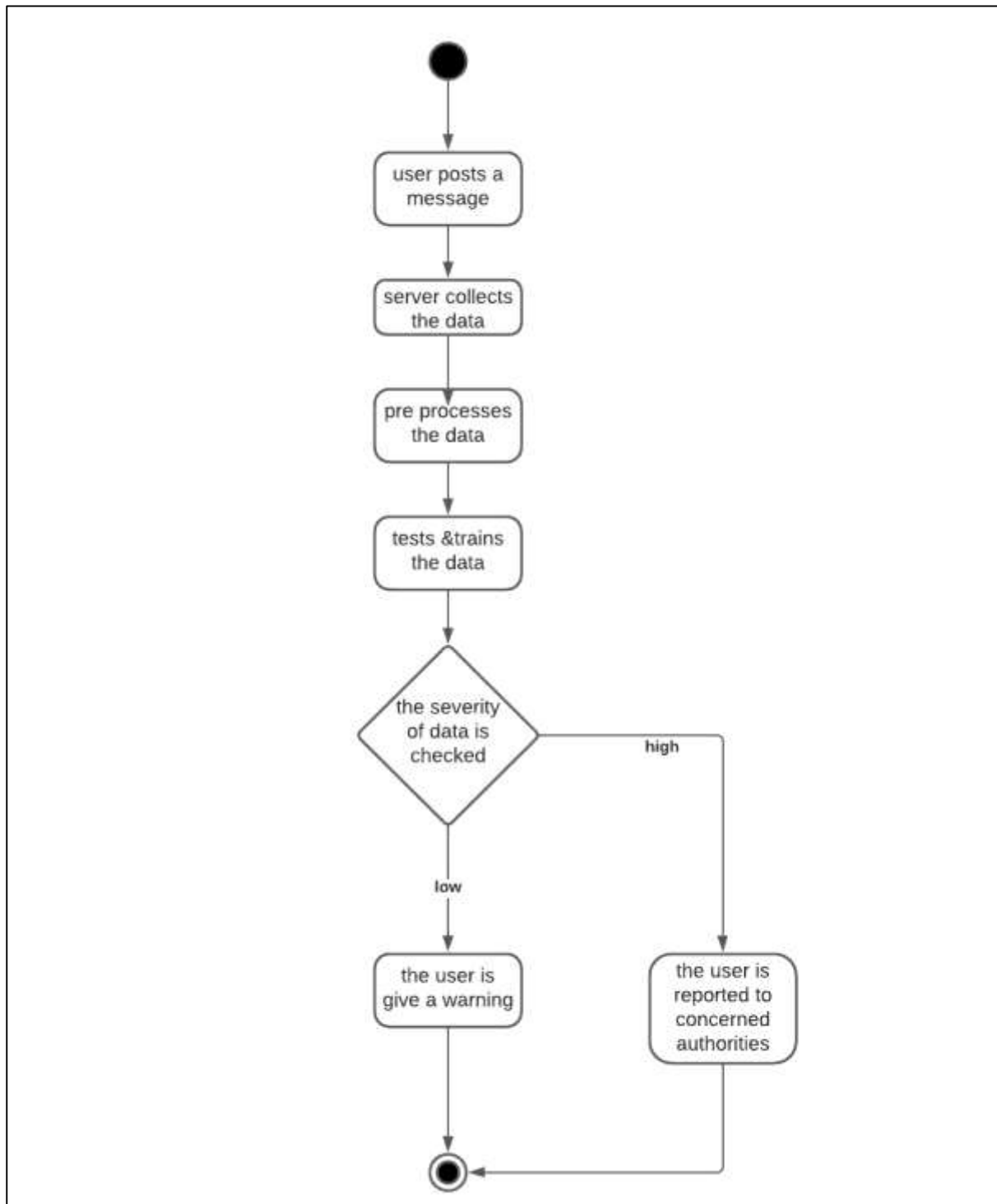


Fig 3.6 Activity Diagram

4. IMPLEMENTATION

4. IMPLEMENTATION

4.1 SAMPLE CODE

```
#Importing libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn import naive_bayes, svm
from sklearn.metrics import classification_report, accuracy_score
import re

from sklearn.feature_extraction.text import TfidfTransformer
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

#Importing data set
data=pd.read_csv('Cyberbullying.csv')

#remove repeated(duplicates) tweets
data.drop_duplicates(inplace = True)
print(data.head(5))
print(data.shape)

#drop table
data_1=data.drop(['annotation'], axis = 1)

#Corpus bag of words
corpus = []
```



```

for i in range (0, len(data)):
    review = re.sub('[A-Z^a-z]', ' ', data['content'][i])
    review = review.lower()
    review = review.split()
    review = ' '.join(review)
    corpus.append(review)

#corpus

bow_transformer = CountVectorizer()
bow_transformer = bow_transformer.fit(corpus)
print(len(bow_transformer.vocabulary_))
messages_bow = bow_transformer.transform(corpus)
print(messages_bow.shape)

tfidf_transformer = TfidfTransformer().fit(messages_bow)

#sentiment analysis
analyzer = SentimentIntensityAnalyzer()
data_1['compound'] = [analyzer.polarity_scores(x)['compound'] for x in data_1['content']]
data_1['neg'] = [analyzer.polarity_scores(x)['neg'] for x in data_1['content']]
data_1['neu'] = [analyzer.polarity_scores(x)['neu'] for x in data_1['content']]
data_1['pos'] = [analyzer.polarity_scores(x)['pos'] for x in data_1['content']]

#Labelling
data_1['comp_score'] = data_1['compound'].apply(lambda c: 0 if c >=0 else 1)

#Splitting dataset into Train and Test set
X_train, X_test, y_train, y_test = train_test_split(data_1['content'], data_1['comp_score'],
random_state=40)

print('Number of rows in the total set: {}'.format(data.shape[0]))
print('Number of rows in the training set: {}'.format(X_train.shape[0]))

```

```
print('Number of rows in the test set: {}'.format(X_test.shape[0]))

#CountVectorizer method
vector = CountVectorizer(stop_words = 'english', lowercase = True)

#Fitting the training data
training_data = vector.fit_transform(X_train)

#Transform testing data
testing_data = vector.transform(X_test)

#Classification
#Naive Bayes
print()
print("-----")
print("Naive Bayes")

Naive = naive_bayes.MultinomialNB()
Naive.fit(training_data, y_train)
nb_pred = Naive.predict(testing_data)

#Analysis Report
print()
print("-----Classification Report-----")
print(classification_report(nb_pred,y_test))

print("-----Accuracy-----")
print(f"The Accuracy Score :{round(accuracy_score(nb_pred,y_test)*100)}")
print()
nb=round(accuracy_score(nb_pred,y_test)*100)
```

```
#Support vector Machine
```

```
print()
```

```
print("-----")
```

```
print("Support vector Machine")
```

```
sv = svm.SVC(kernel='linear') # Linear Kernel
```

```
sv.fit(training_data, y_train)
```

```
sv_pred = sv.predict(testing_data)
```

```
#Analysis Report
```

```
print()
```

```
print("-----Classification Report-----")
```

```
print(classification_report(sv_pred,y_test))
```

```
print()
```

```
print("-----Accuracy-----")
```

```
print(f"The Accuracy Score :{round(accuracy_score(sv_pred,y_test)*100)}")
```

```
svm=round(accuracy_score(sv_pred,y_test)*100)
```

```
#Data Visualization
```

```
# comparison Graph
```

```
objects = ('Naive Bayes', 'SVM')
```

```
y_pos = np.arange(len(objects))
```

```
performance = [nb,svm]
```

```
plt.bar(y_pos, performance, align='center', alpha=0.5)
```

```
plt.xticks(y_pos, objects)
```

```
plt.ylabel('Accuracy')
```

```
plt.title('Cyberbullying')
```

```
plt.show()
```

```
#pie graph
plt.figure(figsize = (7,7))
counts = data_1['comp_score'].value_counts()

plt.pie(counts, labels = counts.index, startangle = 90, counterclock = False, wedgeprops = {'width' :
0.6},autopct='%1.1f%%', pctdistance = 0.55, textprops = {'color': 'black', 'fontsize' : 15}, shadow =
True,colors = sns.color_palette("Paired")[3:])

plt.text(x = -0.35, y = 0, s = 'Total Tweets: {}'.format(data.shape[0]))
plt.title('Distribution of Tweets', fontsize = 14);

#Heatmap
plt.figure(figsize=(15, 15))
sns.heatmap(data_1.corr(), linewidths=.5)

#Histogram
fig, axis = plt.subplots(2,3,figsize=(8, 8))
data_1.hist(ax=axis)
```

4.2 Result Generation

The Final Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures like,

- Accuracy

Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

$$AC = \frac{TP+TN}{TP+TN+FP+FN}$$

- Precision

Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives.

$$\text{Precision} = \frac{TP}{TP+FP}$$

- Recall

Recall is the number of correct results divided by the number of results that should have been returned. In binary classification, recall is called sensitivity. It can be viewed as the probability that a relevant document is retrieved by the query.

$$\text{Recall} = \frac{TP}{TP+FN}$$

- F-Measure

F measure (F1 score or F score) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

$$\text{F-measure} = \frac{2TP}{2TP+FP+FN}$$

5. SCREENSHOTS

5. SCREENSHOTS

5.1 LOADING DATASET AND REMOVING REPEATED DATA

```
Python 3.7.3 (default, Mar 27 2019, 17:13:21) [MSC v.1915 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 7.4.0 -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/aashi/Downloads/Source code_Cyberbullying/main.py
code_Cyberbullying')
   Unnamed: 0  ...      user
0           0  ...  scotthamilton
1           1  ...    mattycus
2           2  ...    ElleCTF
3           3  ...    Karoli
4           4  ...    joy_wolf

[5 rows x 4 columns]
(20001, 4)
281
(20001, 281)
Number of rows in the total set: 20001
Number of rows in the training set: 15000
Number of rows in the test set: 5001
```

Screenshot 5.1: Loading Dataset and removing repeated data and split the data into train-test

5.2 TEST ACCURACY USING NAÏVE BAYES ALGORITHM

```
-----
Naive Bayes

-----Classification Report-----
              precision    recall  f1-score   support

         0           0.79       0.85       0.82       2350
         1           0.85       0.80       0.83       2651

   micro avg       0.82       0.82       0.82       5001
   macro avg       0.82       0.82       0.82       5001
weighted avg       0.83       0.82       0.82       5001

-----Accuracy-----
The Accuracy Score :82.0
```

Screenshot 5.2: Test Accuracy using Naïve Bayes Algorithm

5.3 TEST ACCURACY USING SUPPORT VECTOR MACHINE

```

-----
Support vector Machine
-----Classification Report-----
              precision    recall  f1-score   support

     0           0.89       0.88       0.89       2538
     1           0.88       0.89       0.88       2463

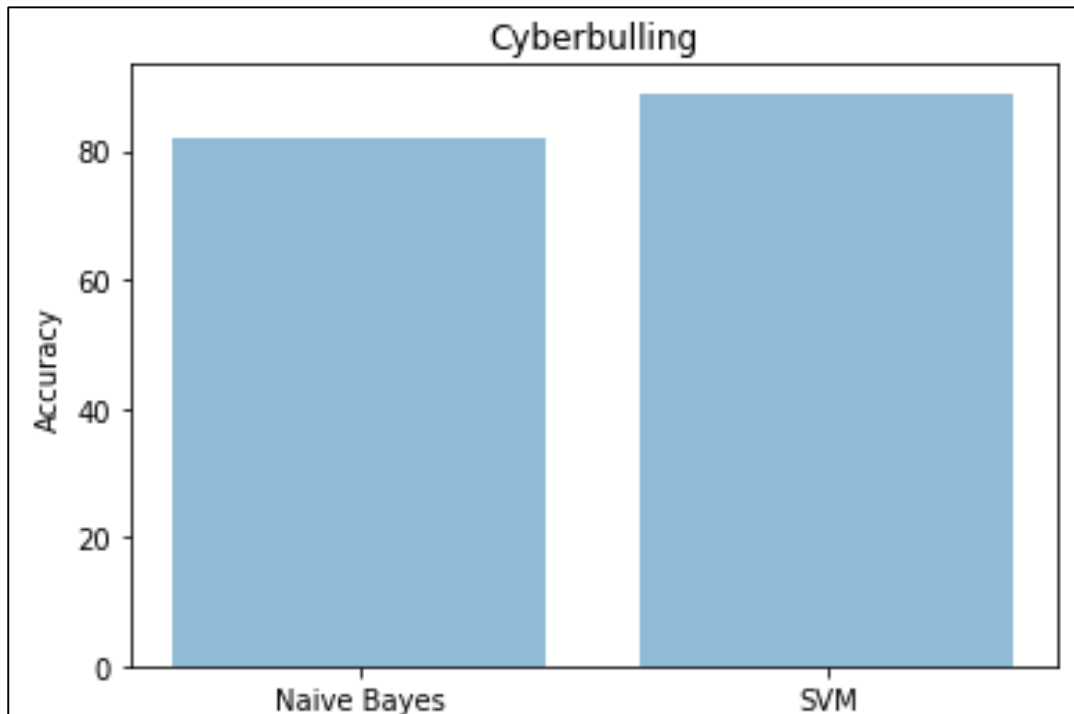
  micro avg       0.89       0.89       0.89       5001
  macro avg       0.89       0.89       0.89       5001
 weighted avg       0.89       0.89       0.89       5001

-----Accuracy-----
The Accuracy Score :89.0

```

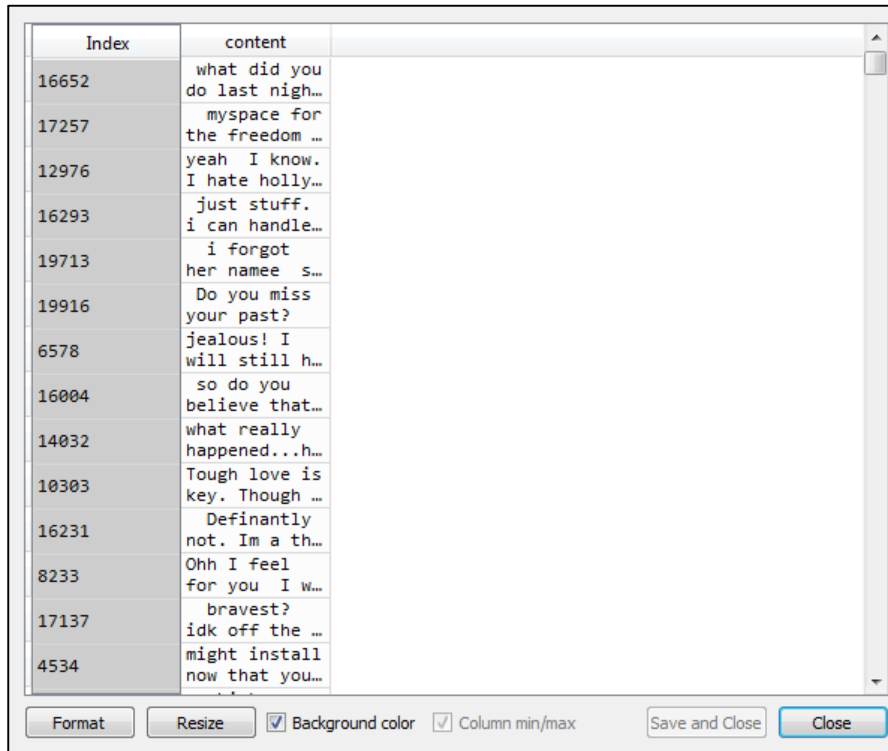
Screenshot 5.3: Test Accuracy using Support Vector Machine (SVM)

5.4 GRAPH SHOWING ACCURACY OF BOTH THE ALGORITHMS



Screenshot 5.4: Graph showing accuracy of both algorithms

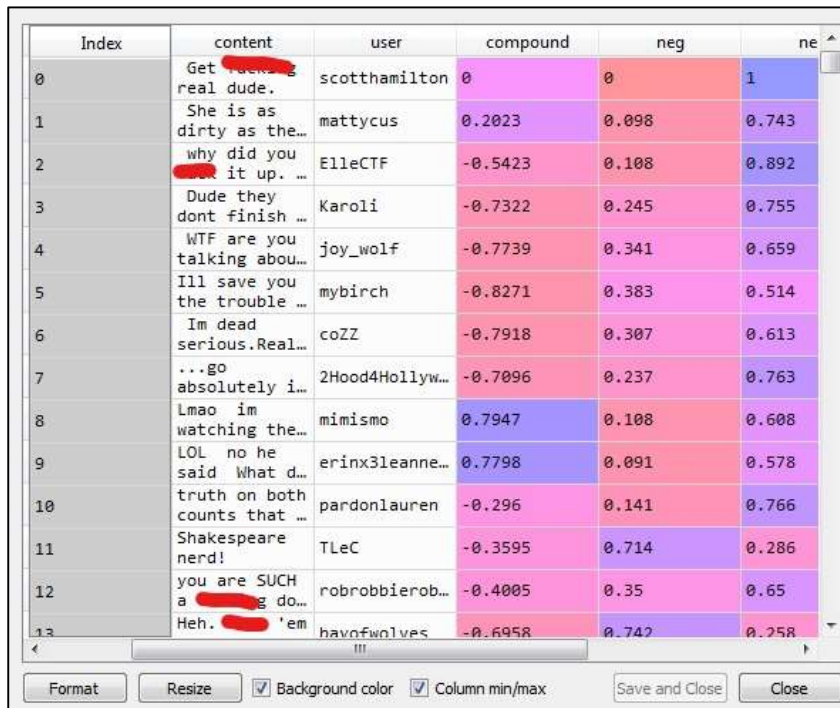
5.5 TWITTER DATABASE SAMPLE WITH INDEX



Index	content	
16652	what did you do last nigh...	
17257	myspace for the freedom ...	
12976	yeah I know. I hate holly...	
16293	just stuff. i can handle...	
19713	i forgot her namee s...	
19916	Do you miss your past?	
6578	jealous! I will still h...	
16004	so do you believe that...	
14032	what really happened...h...	
10303	Tough love is key. Though ...	
16231	Definantly not. Im a th...	
8233	Ohh I feel for you I W...	
17137	bravest? idk off the ...	
4534	might install now that you...	

Screenshot 5.5: Twitter Dataset sample with index

5.6 CYBERBULLYING DETECTED DATAVALUES



Index	content	user	compound	neg	ne
0	Get [redacted] real dude.	scotthamilton	0	0	1
1	She is as dirty as the...	mattycus	0.2023	0.098	0.743
2	why did you [redacted] it up. ...	ElleCTF	-0.5423	0.108	0.892
3	Dude they dont finish ...	Karoli	-0.7322	0.245	0.755
4	WTF are you talking abou...	joy_wolf	-0.7739	0.341	0.659
5	Ill save you the trouble ...	mybitch	-0.8271	0.383	0.514
6	Im dead serious.Real...	coZZ	-0.7918	0.307	0.613
7	...go absolutely i...	2Hood4Hollyw...	-0.7096	0.237	0.763
8	Lmao im watching the...	mimismo	0.7947	0.108	0.608
9	LOL no he said What d...	erinx3leanne...	0.7798	0.091	0.578
10	truth on both counts that ...	pardonlauren	-0.296	0.141	0.766
11	Shakespeare nerd!	TLeC	-0.3595	0.714	0.286
12	you are SUCH a [redacted] do...	robrobberob...	-0.4005	0.35	0.65
13	Heh. [redacted] 'em havnfwolves	hvnfwolves	-0.6958	0.742	0.258

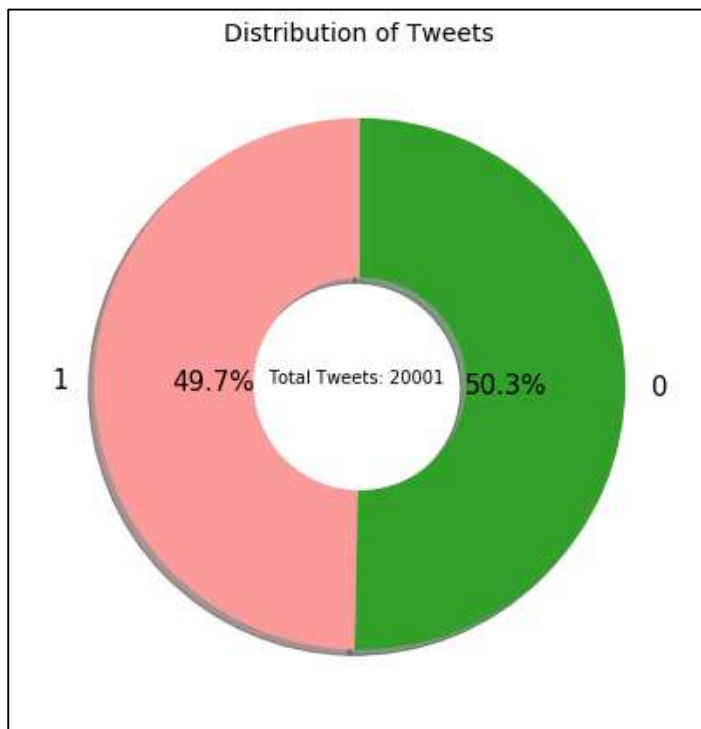
Screenshot 5.6: Cyberbullying Detected Data values with username

5.7 OUTPUT INDICATING CYBERBULLYING DETECTION

Index	comp_score
11131	1
13062	0
17693	0
14857	0
17105	0
4389	1
9827	1
1737	1
19793	1
6640	1
8162	1
13470	1
2411	1
17869	0

Screenshot 5.7: Output Indicating Cyber bullying detection

5.8 PIE CHART INDICATING 0 and 1 Values



Screenshot 5.8 Pie chart indicating 0 and 1 values i.e. Cyber bullying or not

6. TESTING

6. TESTING

6.1 INTRODUCTION TO TESTING

System testing is the stage of implementation, which aimed at ensuring that system works accurately and efficiently before the live operation commence. Testing is the process of executing a program with the intent of finding an error. A good test case is one that has a high probability of finding an error. A successful test is one that answers a yet undiscovered error. The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product.

6.2 TYPES OF TESTING

6.2.1 UNIT TESTING

Unit testing is the testing of each module and the integration of the overall system is done. Unit testing becomes verification efforts on the smallest unit of software design in the module. This is also known as ‘module testing’. The modules of the system are tested separately. This testing is carried out during the programming itself. In this testing step, each model is found to be working satisfactorily as regard to the expected output from the module. There are some validation checks for the fields. For example, the validation check is done for verifying the data given by the user where both format and validity of the data entered is included. It is very easy to find error and debug the system.

6.2.2 INTEGRATION TESTING

Data can be lost across an interface, one module can have an adverse effect on the other sub function, when combined, may not produce the desired major function. Integrated testing is systematic testing that can be done with sample data. The need for the integrated test is to find the overall system performance. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

6.2.3 FUNCTIONAL TESTING

Functional testing is a formal type of testing performed by testers. Functional testing focuses on testing software against design document, Use cases and requirements document. Functional testing is a black box type of testing and does not require internal working of the software unlike white box testing. Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes.

6.3 TEST CASES

6.3.1 UPLOADING AND LOADING DATASET

Test case ID	Test case name	Purpose	Input	Output
1	User posts a tweet	Use it to check and predict if it is cyber bullying or not	The user uploads a tweet	Uploaded /Loaded Successfully
2	User posts another tweet	Use it to check and predict if it is cyber bullying or not	The user uploads a tweet	Uploaded /Loaded Successfully

6.3.2 PREDICTION

Test case ID	Test case name	Purpose	Input	Output
1	Prediction Test 1	To check if the predictor performs its task.	Tweets data is loaded and given	Cyberbullying is detected by value of 1 or 0
2	Prediction Test 2	To check if the predictor performs its task.	Tweets data is loaded and given	Cyberbullying is detected by value of 1 or 0
3	Prediction Test 3	To check if the predictor performs its task.	Tweets data is loaded and given	Cyberbullying is detected by value of 1 or 0

7. CONCLUSION & FUTURE SCOPE

7. CONCLUSION & FUTURE SCOPE

7.1 PROJECT CONCLUSION

We have developed an approach towards the detection of cyberbullying behaviour. If we are able to successfully detect such posts which are not suitable for adolescents or teenagers, we can very effectively deal with the crimes that are committed using these platforms. An approach is proposed for detecting and preventing Twitter cyberbullying using Supervised Binary classification Machine Learning algorithms. Our model is evaluated on both Support Vector Machine and Naive Bayes, also for feature extraction, we used the TFIDF vectorise. As the results show us that the accuracy for detecting cyberbullying content has also been great for Support Vector Machine which is better than Naive Bayes. Our model will help people from the attacks of social media bullies.

7.2 FUTURE SCOPE

In future, it is possible to provide extensions or modifications to the proposed clustering and classification algorithms to achieve further increased performance. Apart from the experimented combination of data mining techniques, further combinations and other clustering algorithms can be used to improve the detection accuracy and to reduce the rate offensive tweets. Finally, the cyberbullying detection system can be extended as a prevention system to enhance the performance of the system.

8. BIBLIOGRAPHY

8. BIBLIOGRAPHY

8.1 GITHUB REPOSITORY LINK

https://github.com/Aashish15/major_project_cyberbullyingdetection

8.2 REFERENCES

1. Amanpreet Singh, Maninder Kaur, "Content-based Cybercrime Detection: A Concise Review", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8, pages 1193-1207, 2019.
2. Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. "Detecting offensive language in social media to protect adolescent online safety". In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom), pages 71– 80. IEEE, 2012.
3. NaliniPriya. G and Asswini. M (2015) "A Dynamic Cognitive System For Automatic Detection And Prevention Of Cyber-Bullying Attacks", ARPN Journal of Engineering and Applied Sciences ©2006-2015 Asian Research Publishing Network (ARPN). VOL. 10, NO. 10, JUNE 2015.
4. "Protective shield shield for social networks to defend cyberbullying and online grooming attacks" in Proceedings of 40 th IRF International Conference, Pune, India, ISBN: 978-93-85832-16-1, 2015.
5. Gupta, P. Kumaraguru, and A. Sureka, "Characterizing pedophile conversations on the internet using online grooming," CoRR, vol. abs/1208.4324, 2012.
6. K. Jedrzejewski and M. Morzy, "Opinion Mining and Social Networks: A Promising Match," 2011 Int. Conf. Adv. Soc. Networks Anal. Min., pp. 599–604, Jul. 2011.
7. H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network."
8. Kelly Reynolds, April Kontostathis, Lynne Edwards, "Using Machine Learning to Detect Cyberbullying", 2011 10th International Conference on Machine Learning and Applications volume 2, pages 241–244. IEEE, 2011.

8.3 WEBSITES

1. <https://developers.google.com/machine-learning/crash-course>
2. <https://ieeexplore.ieee.org/document/8626869>
3. <https://www.sciencedirect.com/science/article/abs/pii/S0747563218306071>

Cyber Bullying Detection in Social Media using Supervised ML Techniques

K. Mahesh, Suwarna Gothane, Aashish Toshniwal, Vinay Nagarale, Harish Gopu

Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad (JNTUH), India

ABSTRACT

Article Info

Volume 7, Issue 3

Page Number: 410-416

Publication Issue :

May-June-2021

Article History

Accepted : 25 May 2021

Published : 31 May 2021

From the day internet came into existence, the era of social networking sprouted. In the beginning, no one may have thought internet would be a host of numerous amazing services like the social networking. Today we can say that online applications and social networking websites have become a non-separable part of one's life. Many people from diverse age groups spend hours daily on such websites. Despite the fact that people are emotionally connected together through social media, these facilities bring along big threats with them such as cyber-attacks, which includes cyberbullying.

Keywords: Cyberbullying, social media, Support Vector Machine, Naïve Bayes, Test-Train Split, Classification, Detection

I. INTRODUCTION

Social networking sites are being widely used today for multiple purposes like entertainment, networking, etc. Social networking sites are a stop for multiple reasons to billions of people today. All the social media platforms require the consent of all the participating people. Communicating with people is no exception, as technology has changed the way people interact with a broader manner and has given a new dimension to communication. Many people are illegally using these communities. Many youngsters are getting bullied these days. Bullies use various services like Twitter, Facebook, and Email to bully people.

Cyberbullying is one of the most frequently happen Internet abuse and also a very serious social problem

especially for teenager. Therefore, more and more researchers are devoting on how to discover and prevent the happen of cyberbullying, especially in social media. Cyberbullying is not just limited to creating a fake identity and publishing/posting some embarrassing photo or video, unpleasant rumours about someone but also giving them threats. The impacts of cyberbullying on social media are horrifying, sometimes leading to the death of some unfortunate victims.

Thus, a complete solution is required for this problem. Cyberbullying needs to stop. The problem can be tackled by detecting and preventing it by using a machine learning approach, this needs to be done using a different perspective.

Cyberbullying is a relatively new medium through which bullying occurs (e.g., chat rooms, text messages). Cyberbullying has been defined as an individual or a group wilfully using information and communication involving electronic technologies to facilitate deliberate and repeated harassment or threat to another individual or group by sending or posting cruel text and/or graphics using technological means. Many of the methods used in traditional bullying are used in cyberbullying. Direct cyberbullying can occur when one person calls another a name through an electronic message. Relational bullying can also occur online. For example, with the numerous social networking sites now online (e.g., Facebook, Myspace), 'hate groups' have become a popular approach to bullying. In a hate group, a student creates an online social group against a schoolmate, allows others to join, and collectively the group posts negative comments about the student. Fortunately, social networking sites have begun taking action against the creation of hate groups. When creating a group on Facebook, for instance, a warning is placed near the bottom of the page that reads, "Note: groups that attack a specific person or group of people (e.g., racist, sexist, or other hate groups) will not be tolerated."

II. RELATED WORK

The results of the proposed model demonstrate significant improvement in the performance of classification on all the datasets in comparison to recent existing models. The success rate of the SVM classifier with the excellent recall is 0.971 via tenfold cross-validation, which demonstrates the high efficiency and effectiveness of the proposed model. [2] Author of the Work in references Detecting Offensive Language in Social media is *Ying chen, Yilu Zhou, Sencun Zhu* and *Heng Xu* who came up with a methodology of user-level offensiveness detection seems a more feasible approach. so, the

Lexical Syntactic Feature (LSF) architecture to detect offensive content and identify potential offensive users in social media. We distinguish the contribution of pejoratives/profanities and obscenities in determining offensive content, and introduce hand-authoring syntactic rules in identifying name-calling harassments. [3] Another Author *K. Jedrzejewski* and *M. Morzy* had a different methodology where The role and importance of social networks in preferred environments for opinion mining and sentiment analysis. Selected properties of social networks that are relevant with respect to opinion mining are described and general relationships between the two disciplines are outlined. The related work and basic definitions used in opinion mining is given. [4] *H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra* , Cyberbullying is a growing problem affecting more than half teens. The main goal is to study cyberbullying incidents in the social network. In this work, we have collected a sample data and their associated comments. We then designed a study and employed human contributors at the crowd-sourced Crowd Flower website to label these media sessions for cyberbullying. [5] *Kelly Reynolds, April Kontostathis, Lynne Edwards* The results of the proposed model demonstrate significant improvement in the performance of classification on all the datasets in comparison to recent existing models. The success rate of the SVM classifier with the excellent recall is 0.971 via tenfold cross-validation, which demonstrates the high efficiency and effectiveness of the proposed model.

III. PROPOSED WORK

The proposed model is introduced to overcome all the disadvantages that arises in the existing system. This system will increase the accuracy of the supervised classification results by classifying the data. An approach is proposed for detecting and preventing

Twitter cyberbullying using Supervised Binary Classification Machine Learning algorithms. Our model is evaluated on both Support Vector Machine and Naive Bayes. It enhances the performance of the overall classification results. This proposed method is supposed to have high performance while providing accurate prediction results. It also avoids sparsity problems. It is less prone to information loss. Below figure depicts the architecture of the proposed system.

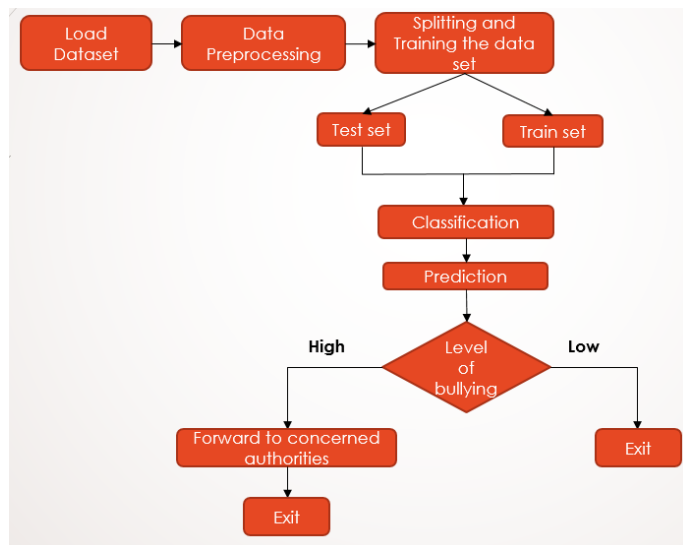


Fig 1. Proposed System Architecture

This is the project architecture where the dataset is loaded and pre-processing is done where the unwanted data is removed and then the data is split trained into test and train sets which is then classified using algorithms and then a prediction is obtained which shows us the level of bullying as high or low and then the execution of the program exits for low severity and forwards to concerned authorities and exits if High severity is present.

IV. IMPLEMENTSTION OF PROPOSED SYSTEM

The proposed system consists of six modules. The data selection is the process of selecting the data for detecting the attacks. In this project, the cyberbullying tweets dataset is used for detecting offensive and non-offensive tweets. The dataset

which contains the information about the user name and tweets label.

Data pre-processing is the process of removing the unwanted data from the dataset. Missing data removal, Encoding Categorical data.

Missing data removal: In this process, the null values such as missing values are removed using imputer library.

Encoding Categorical data: That categorical data is defined as variables with a finite set of label values. That most machine learning algorithms require numerical input and output variables. That an integer and one hot encoding is used to convert categorical data to integer data.

Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes. One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance. Separating data into training and testing sets is an important part of evaluating data mining models. Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

The Supervised classification algorithm such as Naïve Bayes and Support vector machine is used in Data Mining.

Support vector machine:

For implementing this we have used Support vector machine (SVM) model is basically a representation of different classes in a hyper plane in multidimensional space. The hyper plane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyper plane.

Naive Bayes classifier:

it is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'.

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Predictive analytics algorithms try to achieve the lowest error possible by either using "boosting" or "bagging".

Accuracy – Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

Speed – Refers to the computational cost in generating and using the classifier or predictor.

Robustness – It refers to the ability of classifier or predictor to make correct predictions from given noisy data.

Scalability – Scalability refers to the ability to construct the classifier or predictor efficiently; given large amount of data.

Interpretability – It refers to what extent the classifier or predictor understands. It's a process of predicting the offensive and non-offensive tweets

from the dataset. This project will effectively predict the data from dataset by enhancing the performance of the overall prediction results.

V. RESULT GENERATION

The Final Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures like,

Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

$$AC = (TP+TN)/(TP+TN+FP+FN)$$

Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives.

$$\text{Precision} = TP/(TP+FP)$$

Recall is the number of correct results divided by the number of results that should have been returned. In binary classification, recall is called sensitivity. It can be viewed as the probability that a relevant document is retrieved by the query.

$$\text{Recall} = TP/(TP+FN)$$

F measure (F1 score or F score) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

$$F\text{-measure} = 2TP/(2TP+FP+FN)$$

VI. OUTPUT AND SCREENSHOTS

Below are the screenshots of the output obtained from the proposed project.

Index	content	user	compound	neg	ne
0	Get real dude.	scotthamilton	0	0	1
1	She is as dirty as the...	mattycus	0.2023	0.098	0.743
2	why did you it up...	ElleCTF	-0.5423	0.108	0.892
3	Dude they dont finish...	Karoli	-0.7322	0.245	0.755
4	WTF are you talking about...	joy_wolf	-0.7739	0.341	0.659
5	I'll save you the trouble...	mybirc	-0.8271	0.383	0.514
6	Im dead serious.Real...	coZZ	-0.7918	0.307	0.613
7	...go absolutely i...	2Hood4Hollyw...	-0.7096	0.237	0.763
8	Lmao im watching the...	mimiso	0.7947	0.108	0.608
9	LOL no he said what d...	erinx3leanne...	0.7798	0.091	0.578
10	truth on both counts that ...	pardonlauren	-0.296	0.141	0.766
11	Shakespeare nerd!	TLeC	-0.3595	0.714	0.286
12	you are SUCH a ...	robrobriob...	-0.4005	0.35	0.65
13	Heh. ...em	havofwolves	-0.6958	0.742	0.258

Fig 2. Cyberbullying Detection Output

Here we can see how the variable explorer indicates and shows the data of the bad words out of the whole data set from the labels 1.

Index	content
16652	what did you do last nigh...
17257	myspace for the freedom ...
12976	yeah I know. I hate holly...
16293	just stuff. i can handle...
19713	i forgot her namee s...
19916	Do you miss your past?
6578	jealous! I will still h...
16004	so do you believe that...
14032	what really happened...h...
10303	Tough love is key. Though ...
16231	Definantly not. Im a th...
8233	Ohh I feel for you I w...
17137	bravest? idk off the ...
4534	might install now that you...

Fig 3. Sample Dataset

A sample of the data from the dataset with index in this way is generated after classifying and training the data set to separate the unwanted data and classify the important data to detect cyberbullying in the twitter database which is labelled with 1 or 0 where 1 means Positive and 0 means negative.

Index	comp_score
11131	1
13062	0
17693	0
14857	0
17105	0
4389	1
9827	1
1737	1
19793	1
6640	1
8162	1
13470	1
2411	1
17869	0

Fig 4. Labelled Values of the Dataset

As we can see above the data indexes with 0 as comp_score are not the users bullying and the indexes with 1 as value are the ones bullying. 0 and 1 indicate mainly if the data value is having abusive content which is tagged by labelling using sentiment analysis.

Number of rows in the total set: 20001

Number of rows in the training set: 15000

Number of rows in the test set: 5001

```

-----
Naive Bayes
-----Classification Report-----
              precision    recall  f1-score   support

     0       0.79       0.85       0.82       2350
     1       0.85       0.80       0.83       2651

 accuracy          0.82
 macro avg         0.82
 weighted avg      0.83

-----Accuracy-----
The Accuracy Score :82

-----
Support vector Machine
-----Classification Report-----
              precision    recall  f1-score   support

     0       0.89       0.88       0.89       2538
     1       0.88       0.89       0.88       2463

 accuracy          0.89
 macro avg         0.89
 weighted avg      0.89

-----Accuracy-----
The Accuracy Score :89

```

Fig 5. Accuracy of both the algorithms

As we can see now SVM has more accuracy than Naïve Bayes. Support vector machine has an accuracy of 89 whereas Naïve Bayes has an accuracy of 82.

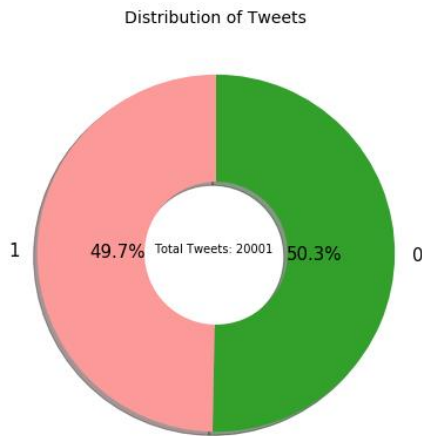


Fig 6. Distribution of Tweets

VII. FUTURE WORK

In future, it is possible to provide extensions or modifications to the proposed clustering and classification algorithms to achieve further increased performance. Apart from the experimented combination of data mining techniques, further combinations and other clustering algorithms can be used to improve the detection accuracy and to reduce the rate offensive tweets. Finally, the cyberbullying detection system can be extended as a prevention system to enhance the performance of the system.

VIII. CONCLUSION

We have developed an approach towards the detection of cyberbullying behaviour. If we are able to successfully detect such posts which are not suitable for adolescents or teenagers, we can very effectively deal with the crimes that are committed using these platforms. An approach is proposed for detecting and preventing Twitter cyberbullying using Supervised Binary Classification Machine Learning algorithms. Our model is evaluated on both Support Vector Machine and Naive Bayes, also for feature extraction, we used the TFIDF vector. As the results show us that the accuracy for detecting cyberbullying content has also been great for Support Vector

Machine which is better than Naive Bayes. Our model will help people from the attacks of social media bullies.

IX. ACKNOWLEDGEMENT

We take this opportunity to express our gratitude and respect to all the faculty members who have guided us in this project. We take privilege to extend our profound gratitude and sincere thanks to our guide **Mr. K. Mahesh** and our PRC co-ordinator **Dr. Suwarna Gothane**, Department of computer science and Engineering, CMR Technical Campus, who constantly supported us at every stage of the project and helped us in our difficult times to make the project a success.

We are thankful to HOD Dr. K Srujan Raju. Department of Computer Science and Engineering, CMR Technical Campus, for his immense support and encouragement. We also take this opportunity to thank Dr A Raji Reddy, Director CMR Technical Campus, for providing us an encouraging environment to work with.

X. REFERENCES

- [1]. Amanpreet Singh, Maninder Kaur, "Content-based Cybercrime Detection: A Concise Review", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8, pages 1193-1207, 2019.
- [2]. Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. "Detecting offensive language in social media to protect adolescent online safety". In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom), pages 71– 80. IEEE, 2012.

- [3]. K. Jedrzejewski and M. Morzy, "Opinion Mining and Social Networks: A Promising Match," 2011 Int. Conf. Adv. Soc. Networks Anal. Min., pp. 599–604, Jul. 2011.
- [4]. H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Analysing Labelled Cyberbullying Incidents on the Instagram Social Network."
- [5]. Kelly Reynolds, April Kontostathis, Lynne Edwards, "Using Machine Learning to Detect Cyberbullying", 2011 10th International Conference on Machine Learning and Applications volume 2, pages 241–244. IEEE, 2011.

Cite this article as :

K. Mahesh, Suwarna Gothane, Aashish Toshniwal, Vinay Nagarale, Harish Gopu, "Cyber Bullying Detection in Social Media using Supervised ML Techniques", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 3, pp. 410-416, May-June 2021.
Available at
doi : <https://doi.org/10.32628/CSEIT217381>
Journal URL : <https://ijsrcseit.com/CSEIT217381>

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

CERTIFICATE OF PUBLICATION

Ref : IJSRCSEIT/Certificate/Volume 7/Issue 3/7240

31-May-2021

This is to certify that **Aashish Toshniwal** has published a research paper entitled 'Cyber Bullying Detection in Social Media using Supervised ML Techniques' in the International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 7, Issue 3, May-June 2021.

This Paper can be downloaded from the following IJSRCSEIT website link

<https://ijsrcseit.com/CSEIT217381>

DOI : <https://doi.org/10.32628/CSEIT217381>

IJSRCSEIT Team wishes all the best for bright future

A handwritten signature in blue ink.

Editor in Chief
IJSRCSEIT

website : <http://ijsrcseit.com>

Peer Reviewed and Refereed International Journal

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

CERTIFICATE OF PUBLICATION

Ref : IJSRCSEIT/Certificate/Volume 7/Issue 3/7240

31-May-2021

This is to certify that **Vinay Nagarale** has published a research paper entitled 'Cyber Bullying Detection in Social Media using Supervised ML Techniques' in the International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 7, Issue 3, May-June 2021.

This Paper can be downloaded from the following IJSRCSEIT website link

<https://ijsrcseit.com/CSEIT217381>

DOI : <https://doi.org/10.32628/CSEIT217381>

IJSRCSEIT Team wishes all the best for bright future

A handwritten signature in blue ink, appearing to be 'Anita', located above the Editor in Chief text.

Editor in Chief
IJSRCSEIT

website : <http://ijsrcseit.com>

Peer Reviewed and Refereed International Journal

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

CERTIFICATE OF PUBLICATION

Ref : IJSRCSEIT/Certificate/Volume 7/Issue 3/7240

31-May-2021

This is to certify that **Harish Gopu** has published a research paper entitled 'Cyber Bullying Detection in Social Media using Supervised ML Techniques' in the International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 7, Issue 3, May-June 2021.

This Paper can be downloaded from the following IJSRCSEIT website link

<https://ijsrcseit.com/CSEIT217381>

DOI : <https://doi.org/10.32628/CSEIT217381>

IJSRCSEIT Team wishes all the best for bright future

A handwritten signature in blue ink, appearing to be 'Anita', is located above the Editor in Chief text.

Editor in Chief
IJSRCSEIT

website : <http://ijsrcseit.com>

Peer Reviewed and Refereed International Journal

Cyber Bullying Detection in Social Media using Supervised ML Techniques

K. Mahesh, Suwarna Gothane, Aashish Toshniwal, Vinay Nagarale, Harish Gopu

Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad (JNTUH), India

ABSTRACT

Article Info

Volume 7, Issue 3

Page Number: 410-416

Publication Issue :

May-June-2021

Article History

Accepted : 25 May 2021

Published : 31 May 2021

From the day internet came into existence, the era of social networking sprouted. In the beginning, no one may have thought internet would be a host of numerous amazing services like the social networking. Today we can say that online applications and social networking websites have become a non-separable part of one's life. Many people from diverse age groups spend hours daily on such websites. Despite the fact that people are emotionally connected together through social media, these facilities bring along big threats with them such as cyber-attacks, which includes cyberbullying.

Keywords: Cyberbullying, social media, Support Vector Machine, Naïve Bayes, Test-Train Split, Classification, Detection

I. INTRODUCTION

Social networking sites are being widely used today for multiple purposes like entertainment, networking, etc. Social networking sites are a stop for multiple reasons to billions of people today. All the social media platforms require the consent of all the participating people. Communicating with people is no exception, as technology has changed the way people interact with a broader manner and has given a new dimension to communication. Many people are illegally using these communities. Many youngsters are getting bullied these days. Bullies use various services like Twitter, Facebook, and Email to bully people.

Cyberbullying is one of the most frequently happen Internet abuse and also a very serious social problem

especially for teenager. Therefore, more and more researchers are devoting on how to discover and prevent the happen of cyberbullying, especially in social media. Cyberbullying is not just limited to creating a fake identity and publishing/posting some embarrassing photo or video, unpleasant rumours about someone but also giving them threats. The impacts of cyberbullying on social media are horrifying, sometimes leading to the death of some unfortunate victims.

Thus, a complete solution is required for this problem. Cyberbullying needs to stop. The problem can be tackled by detecting and preventing it by using a machine learning approach, this needs to be done using a different perspective.

Cyberbullying is a relatively new medium through which bullying occurs (e.g., chat rooms, text messages). Cyberbullying has been defined as an individual or a group wilfully using information and communication involving electronic technologies to facilitate deliberate and repeated harassment or threat to another individual or group by sending or posting cruel text and/or graphics using technological means. Many of the methods used in traditional bullying are used in cyberbullying. Direct cyberbullying can occur when one person calls another a name through an electronic message. Relational bullying can also occur online. For example, with the numerous social networking sites now online (e.g., Facebook, Myspace), 'hate groups' have become a popular approach to bullying. In a hate group, a student creates an online social group against a schoolmate, allows others to join, and collectively the group posts negative comments about the student. Fortunately, social networking sites have begun taking action against the creation of hate groups. When creating a group on Facebook, for instance, a warning is placed near the bottom of the page that reads, "Note: groups that attack a specific person or group of people (e.g., racist, sexist, or other hate groups) will not be tolerated."

II. RELATED WORK

The results of the proposed model demonstrate significant improvement in the performance of classification on all the datasets in comparison to recent existing models. The success rate of the SVM classifier with the excellent recall is 0.971 via tenfold cross-validation, which demonstrates the high efficiency and effectiveness of the proposed model. [2] Author of the Work in references Detecting Offensive Language in Social media is *Ying chen, Yilu Zhou, Sencun Zhu and Heng Xu* who came up with a methodology of user-level offensiveness detection seems a more feasible approach. so, the

Lexical Syntactic Feature (LSF) architecture to detect offensive content and identify potential offensive users in social media. We distinguish the contribution of pejoratives/profanities and obscenities in determining offensive content, and introduce hand-authoring syntactic rules in identifying name-calling harassments. [3] Another Author *K. Jedrzejewski and M. Morzy* had a different methodology where The role and importance of social networks in preferred environments for opinion mining and sentiment analysis. Selected properties of social networks that are relevant with respect to opinion mining are described and general relationships between the two disciplines are outlined. The related work and basic definitions used in opinion mining is given. [4] *H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra* , Cyberbullying is a growing problem affecting more than half teens. The main goal is to study cyberbullying incidents in the social network. In this work, we have collected a sample data and their associated comments. We then designed a study and employed human contributors at the crowd-sourced Crowd Flower website to label these media sessions for cyberbullying. [5] *Kelly Reynolds, April Kontostathis, Lynne Edwards* The results of the proposed model demonstrate significant improvement in the performance of classification on all the datasets in comparison to recent existing models. The success rate of the SVM classifier with the excellent recall is 0.971 via tenfold cross-validation, which demonstrates the high efficiency and effectiveness of the proposed model.

III. PROPOSED WORK

The proposed model is introduced to overcome all the disadvantages that arises in the existing system. This system will increase the accuracy of the supervised classification results by classifying the data. An approach is proposed for detecting and preventing

Twitter cyberbullying using Supervised Binary Classification Machine Learning algorithms. Our model is evaluated on both Support Vector Machine and Naive Bayes. It enhances the performance of the overall classification results. This proposed method is supposed to have high performance while providing accurate prediction results. It also avoids sparsity problems. It is less prone to information loss. Below figure depicts the architecture of the proposed system.

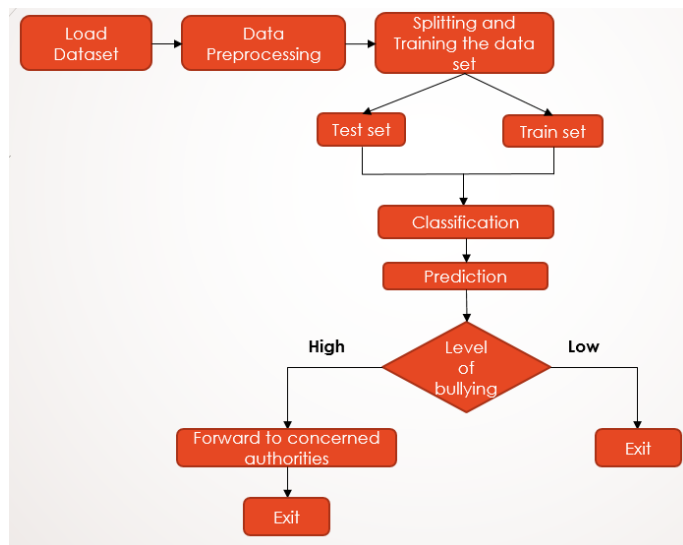


Fig 1. Proposed System Architecture

This is the project architecture where the dataset is loaded and pre-processing is done where the unwanted data is removed and then the data is split trained into test and train sets which is then classified using algorithms and then a prediction is obtained which shows us the level of bullying as high or low and then the execution of the program exits for low severity and forwards to concerned authorities and exits if High severity is present.

IV. IMPLEMENTSTION OF PROPOSED SYSTEM

The proposed system consists of six modules. The data selection is the process of selecting the data for detecting the attacks. In this project, the cyberbullying tweets dataset is used for detecting offensive and non-offensive tweets. The dataset

which contains the information about the user name and tweets label.

Data pre-processing is the process of removing the unwanted data from the dataset. Missing data removal, Encoding Categorical data.

Missing data removal: In this process, the null values such as missing values are removed using imputer library.

Encoding Categorical data: That categorical data is defined as variables with a finite set of label values. That most machine learning algorithms require numerical input and output variables. That an integer and one hot encoding is used to convert categorical data to integer data.

Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes. One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance. Separating data into training and testing sets is an important part of evaluating data mining models. Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

The Supervised classification algorithm such as Naïve Bayes and Support vector machine is used in Data Mining.

Support vector machine:

For implementing this we have used Support vector machine (SVM) model is basically a representation of different classes in a hyper plane in multidimensional space. The hyper plane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyper plane.

Naive Bayes classifier:

it is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'.

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Predictive analytics algorithms try to achieve the lowest error possible by either using "boosting" or "bagging".

Accuracy – Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

Speed – Refers to the computational cost in generating and using the classifier or predictor.

Robustness – It refers to the ability of classifier or predictor to make correct predictions from given noisy data.

Scalability – Scalability refers to the ability to construct the classifier or predictor efficiently; given large amount of data.

Interpretability – It refers to what extent the classifier or predictor understands. It's a process of predicting the offensive and non-offensive tweets

from the dataset. This project will effectively predict the data from dataset by enhancing the performance of the overall prediction results.

V. RESULT GENERATION

The Final Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures like,

Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

$$AC = (TP+TN)/(TP+TN+FP+FN)$$

Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives.

$$\text{Precision} = TP/(TP+FP)$$

Recall is the number of correct results divided by the number of results that should have been returned. In binary classification, recall is called sensitivity. It can be viewed as the probability that a relevant document is retrieved by the query.

$$\text{Recall} = TP/(TP+FN)$$

F measure (F1 score or F score) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

$$F\text{-measure} = 2TP/(2TP+FP+FN)$$

VI. OUTPUT AND SCREENSHOTS

Below are the screenshots of the output obtained from the proposed project.

Index	content	user	compound	neg	ne
0	Get real dude.	scotthamilton	0	0	1
1	She is as dirty as the...	mattycus	0.2023	0.098	0.743
2	why did you it up...	ElleCTF	-0.5423	0.108	0.892
3	Dude they dont finish...	Karoli	-0.7322	0.245	0.755
4	WTF are you talking about...	joy_wolf	-0.7739	0.341	0.659
5	Ill save you the trouble...	mybirc	-0.8271	0.383	0.514
6	Im dead serious.Real...	coZZ	-0.7918	0.307	0.613
7	...go absolutely i...	2Hood4Hollyw...	-0.7096	0.237	0.763
8	Lmao im watching the...	mimiso	0.7947	0.108	0.608
9	LOL no he said what d...	erinx3leanne...	0.7798	0.091	0.578
10	truth on both counts that ...	pardonlauren	-0.296	0.141	0.766
11	Shakespeare nerd!	TLeC	-0.3595	0.714	0.286
12	you are SUCH a ...	robrobriob...	-0.4005	0.35	0.65
13	Heh. ...em	havofwolves	-0.6958	0.742	0.258

Fig 2. Cyberbullying Detection Output

Here we can see how the variable explorer indicates and shows the data of the bad words out of the whole data set from the labels 1.

Index	content
16652	what did you do last nigh...
17257	myspace for the freedom ...
12976	yeah I know. I hate holly...
16293	just stuff. i can handle...
19713	i forgot her namee s...
19916	Do you miss your past?
6578	jealous! I will still h...
16004	so do you believe that...
14032	what really happened...h...
10303	Tough love is key. Though ...
16231	Definantly not. Im a th...
8233	Ohh I feel for you I w...
17137	bravest? idk off the ...
4534	might install now that you...

Fig 3. Sample Dataset

A sample of the data from the dataset with index in this way is generated after classifying and training the data set to separate the unwanted data and classify the important data to detect cyberbullying in the twitter database which is labelled with 1 or 0 where 1 means Positive and 0 means negative.

Index	comp_score
11131	1
13062	0
17693	0
14857	0
17105	0
4389	1
9827	1
1737	1
19793	1
6640	1
8162	1
13470	1
2411	1
17869	0

Fig 4. Labelled Values of the Dataset

As we can see above the data indexes with 0 as comp_score are not the users bullying and the indexes with 1 as value are the ones bullying. 0 and 1 indicate mainly if the data value is having abusive content which is tagged by labelling using sentiment analysis.

Number of rows in the total set: 20001

Number of rows in the training set: 15000

Number of rows in the test set: 5001

```

-----
Naive Bayes
-----Classification Report-----
              precision    recall  f1-score   support

     0       0.79       0.85       0.82       2350
     1       0.85       0.80       0.83       2651

 accuracy         0.82
 macro avg       0.82       0.82       0.82       5001
weighted avg       0.83       0.82       0.82       5001

-----Accuracy-----
The Accuracy Score :82

-----
Support vector Machine
-----Classification Report-----
              precision    recall  f1-score   support

     0       0.89       0.88       0.89       2538
     1       0.88       0.89       0.88       2463

 accuracy         0.89
 macro avg       0.89       0.89       0.89       5001
weighted avg       0.89       0.89       0.89       5001

-----Accuracy-----
The Accuracy Score :89

```

Fig 5. Accuracy of both the algorithms

As we can see now SVM has more accuracy than Naïve Bayes. Support vector machine has an accuracy of 89 whereas Naïve Bayes has an accuracy of 82.

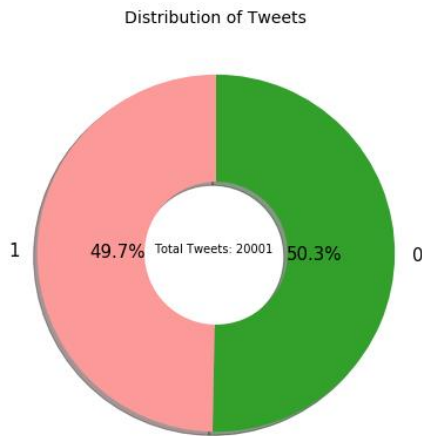


Fig 6. Distribution of Tweets

VII. FUTURE WORK

In future, it is possible to provide extensions or modifications to the proposed clustering and classification algorithms to achieve further increased performance. Apart from the experimented combination of data mining techniques, further combinations and other clustering algorithms can be used to improve the detection accuracy and to reduce the rate offensive tweets. Finally, the cyberbullying detection system can be extended as a prevention system to enhance the performance of the system.

VIII. CONCLUSION

We have developed an approach towards the detection of cyberbullying behaviour. If we are able to successfully detect such posts which are not suitable for adolescents or teenagers, we can very effectively deal with the crimes that are committed using these platforms. An approach is proposed for detecting and preventing Twitter cyberbullying using Supervised Binary Classification Machine Learning algorithms. Our model is evaluated on both Support Vector Machine and Naive Bayes, also for feature extraction, we used the TFIDF vector. As the results show us that the accuracy for detecting cyberbullying content has also been great for Support Vector

Machine which is better than Naive Bayes. Our model will help people from the attacks of social media bullies.

IX. ACKNOWLEDGEMENT

We take this opportunity to express our gratitude and respect to all the faculty members who have guided us in this project. We take privilege to extend our profound gratitude and sincere thanks to our guide **Mr. K. Mahesh** and our PRC co-ordinator **Dr. Suwarna Gothane**, Department of computer science and Engineering, CMR Technical Campus, who constantly supported us at every stage of the project and helped us in our difficult times to make the project a success.

We are thankful to HOD Dr. K Srujan Raju. Department of Computer Science and Engineering, CMR Technical Campus, for his immense support and encouragement. We also take this opportunity to thank Dr A Raji Reddy, Director CMR Technical Campus, for providing us an encouraging environment to work with.

X. REFERENCES

- [1]. Amanpreet Singh, Maninder Kaur, "Content-based Cybercrime Detection: A Concise Review", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8, pages 1193-1207, 2019.
- [2]. Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. "Detecting offensive language in social media to protect adolescent online safety". In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom), pages 71– 80. IEEE, 2012.

- [3]. K. Jedrzejewski and M. Morzy, "Opinion Mining and Social Networks: A Promising Match," 2011 Int. Conf. Adv. Soc. Networks Anal. Min., pp. 599–604, Jul. 2011.
- [4]. H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra, "Analysing Labelled Cyberbullying Incidents on the Instagram Social Network."
- [5]. Kelly Reynolds, April Kontostathis, Lynne Edwards, "Using Machine Learning to Detect Cyberbullying", 2011 10th International Conference on Machine Learning and Applications volume 2, pages 241–244. IEEE, 2011.

Cite this article as :

K. Mahesh, Suwarna Gothane, Aashish Toshniwal, Vinay Nagarale, Harish Gopu, "Cyber Bullying Detection in Social Media using Supervised ML Techniques", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 3, pp. 410-416, May-June 2021.
Available at
doi : <https://doi.org/10.32628/CSEIT217381>
Journal URL : <https://ijsrcseit.com/CSEIT217381>

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

CERTIFICATE OF PUBLICATION

Ref : IJSRCSEIT/Certificate/Volume 7/Issue 3/7240

31-May-2021

This is to certify that **Aashish Toshniwal** has published a research paper entitled 'Cyber Bullying Detection in Social Media using Supervised ML Techniques' in the International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 7, Issue 3, May-June 2021.

This Paper can be downloaded from the following IJSRCSEIT website link

<https://ijsrcseit.com/CSEIT217381>

DOI : <https://doi.org/10.32628/CSEIT217381>

IJSRCSEIT Team wishes all the best for bright future

A handwritten signature in blue ink.

Editor in Chief
IJSRCSEIT

website : <http://ijsrcseit.com>

Peer Reviewed and Refereed International Journal

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

CERTIFICATE OF PUBLICATION

Ref : IJSRCSEIT/Certificate/Volume 7/Issue 3/7240

31-May-2021

This is to certify that **Vinay Nagarale** has published a research paper entitled 'Cyber Bullying Detection in Social Media using Supervised ML Techniques' in the International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 7, Issue 3, May-June 2021.

This Paper can be downloaded from the following IJSRCSEIT website link

<https://ijsrcseit.com/CSEIT217381>

DOI : <https://doi.org/10.32628/CSEIT217381>

IJSRCSEIT Team wishes all the best for bright future

A handwritten signature in blue ink, likely of the Editor in Chief.

Editor in Chief
IJSRCSEIT

website : <http://ijsrcseit.com>

Peer Reviewed and Refereed International Journal

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

CERTIFICATE OF PUBLICATION

Ref : IJSRCSEIT/Certificate/Volume 7/Issue 3/7240

31-May-2021

This is to certify that **Harish Gopu** has published a research paper entitled 'Cyber Bullying Detection in Social Media using Supervised ML Techniques' in the International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), Volume 7, Issue 3, May-June 2021.

This Paper can be downloaded from the following IJSRCSEIT website link

<https://ijsrcseit.com/CSEIT217381>

DOI : <https://doi.org/10.32628/CSEIT217381>

IJSRCSEIT Team wishes all the best for bright future

A handwritten signature in blue ink, appearing to be 'Anita', is located above the Editor in Chief text.

Editor in Chief
IJSRCSEIT

website : <http://ijsrcseit.com>

Peer Reviewed and Refereed International Journal