VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belgaum – 590 018

A PROJECT REPORT ON

"Cyber-bullying detection using machine learning"

Submitted in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

INFORMATION SCIENCE AND ENGINEERING

BY

Saurav Singh (1CR16IS095)

Shakambhari (1CR16IS097)

Sumeet kumar (1CR16IS112)

Under the guidance of

Dr R Joshua Samuel Raj

(Professor, Dept. of ISE, CMRIT)

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

CMR INSTITUTE OF TECHNOLOGY,

BANGALORE-37

CMR INSTITUTE OF TECHNOLOGY
BANGALORE-37

Department of Information Science & Engineering

Certificate

This is to certify that the project entitled **"Cyber-bullying detection using machine learning"** has been successfully completed by **Saurav singh, USN 1CR16IS095, Shakambhari, USN 1CR16IS097** and **Sumeet kumar, USN 1CR16IS112,** bonafide students of CMR Institute of Technology in partial fulfillment of the requirements for the award of degree of Bachelor of Engineering in **Information Science and Engineering** of the Visvesvaraya Technological University, Belgaum during the academic year 2019-2020. It is certified that all the corrections/suggestions indicated for Internal Assessment have been incorporated in the project report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said degree.

Name & Signature of Guide (Dr R Joshua Samuel Raj)	Name & Signature of HOD (Mrs. Farida Begum)	Signature of Principal (Dr.Sanjay Jain)

External Viva

Name of the Examiners	Signature with date
1.	
2.	

CMR INSTITUTE OF TECHNOLOGY
BANGALORE-37

Department of Information Science & Engineering

DECLARATION

We, **Saurav Singh, USN:1CR16IS095, Shakambhari, USN:1CR16IS097** and **Sumeet Kumar, USN:1CR16IS112,** bonafide students of CMR Institute of Technology, Bangalore, hereby declare that the dissertation entitled, **"Cyber-bullying detection using machine learning"** has been carried out by us under the guidance of **Dr R Joshua Samuel Raj, Professor**, CMRIT, Bangalore, in partial fulfillment of the requirements for the award of the degree of Bachelor of Engineering in **Information Science and Engineering**, of the Visvesvaraya Technological University, Belgaum during the academic year 2019-2020. The work done in this dissertation report is original and it has not been submitted for any other degree in any university.

Saurav Singh (USN:1CR16IS095)

Shakambhari (USN:1CR16IS097)

Sumeet Kumar(USN:1CR16IS112)

ABSTRACT

With the exponential increase of social media users, cyber bullying has been emerged as a form of bullying through electronic messages. Social networks provide a rich environment for bullies to uses these networks as vulnerable to attacks against victims. Given the consequences of cyber bullying on victims, it is necessary to find suitable actions to detect and prevent it. Machine learning can be helpful to detect language patterns of the bullies and hence can generate a model to automatically detect cyber-bullying actions. This project proposes a supervised machine learning approach for detecting and preventing cyber-bullying. Several classifiers are used to train and recognize bullying actions.

Keywords: cyber bullying, misbehaving, social media, disturb.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany a successful completion of any task would be incomplete without the mention of people who made it possible, success is the epitome ofhard work and perseverance, but steadfast of all is encouraging guidance.

So, it is with gratitude that we acknowledge all those whose guidance and encouragement served as beacon of light and crowned our effort with success.

We would like to thank **Dr. Sanjay Jain,** Principal, CMRIT, Bangalore, for providing an excellent academic environment in the college and his never-ending support for the B.E program.

We would like to express our gratitude towards **Dr. Farida Begum,** Assoc. Professor and HOD, Department of Information Science and Engineering CMRIT, Bangalore, who provided guidance and gave valuable suggestions regarding the project.

We consider it a privilege and honour to express our sincere gratitude to our internal guide **Dr R Joshua Samuel Raj,** Professor, Department of Information Science & Engineering for their valuable guidance throughout the tenure of this project work.

We would also like to thank all the faculty members who have always been very Co-operative and generous. Conclusively, we also thank all the non-teaching staff and all others who have done immense help directly or indirectly during our project.

SAURAV SINGH

SHAKAMBHARI

SUMEET KUMAR

Table of Contents

LIST OF FIGURES

LIST OF TABLE

Chapter 1

PREAMBLE

1.1 Introduction

Cyber bullying is the use of technology as a medium to bully someone. Social networking sites provide a fertile medium for bullies, and teens and young adults who use these sites are vulnerable to attack. Through machine learning, we can detect language, patterns used by bullies and their victims and develop rules to automatically detect cyber bullying content. Cyberbullying is harassing, threatening, embarrassing someone or making targeting sharings about that person through technology. Cyberbullying actions, which are more common among young children and young people, can also be observed in adults. In such cases, severe legal sanctions are imposed on adults, such as prison sentences.

In contrast to the typical "bullying" actions, there is no need for physical force or face-to-face communication for cyberbullying. Anyone using any device with an Internet connection can perform a cyberbullying action. Bullies can be anonymous, as well as from close friends of children and young people.

Cyberbullying is the most common in these platforms:

- Popular social networks like Facebook, Instagram, Twitter and Snapchat
- Text messages sent via direct devices (SMS)
- Instant messaging features offered by e-mail providers, applications or social networks
- Chat rooms and e-mails.

Types of Cyberbullying:

There are various types of Cyberbullying. In terms of structure, all of the actions that you can experience or encounter may have differences in themselves.

To understand what the problem is, or to protect your children from cyberbullying, you might know the types of issues you may encounter:

1. **Disclosure:** The victim's social media accounts are revealed by the seizure of the actions that will cause him to appear funny or lose his or her dignity. This method can cause permanent damage to the victim's digital reputation as it is virtually impossible to erase and destroy the information shared on the Internet.
2. **Pretending as someone else:** based on the use of electronic and information technologies to further automate production.
3. **Indictment:** Through accusations; someone can share things to humiliate your child and harm his/her digital dignity and friendship relationships. Generally, these attacks are personal and cause anger in the victim.
4. **Trolling:** We are all familiar with what trolls are, but yes; trolling is also a cyberbullying act when the pressure is set for the person to blame and respond.
5. **Cheat and Blackmail:** The trickers learn the secrets of the children by gaining their trust. Then they share these secrets openly with everyone on the internet. Sometimes they use the information they get to blackmail. These people may be from your child's close circle, or someone you don't even know.

Although **symptoms of cyberbullying** vary, children and adolescents who are victims of bullying usually have the following symptoms:

- Emotional anger after using the Internet or mobile devices
- Overprotective behaviour about digital life
- Getting away from family members, friends and general routine activities
- Avoiding school and group meetings
- Performance decrease in class and academic success
- To exhibit angry and irritated behaviour at home
- Continuous changes in mood, behaviour, sleep and appetite
- Extraordinary; stop using devices such as computers and telephones
- Tense and hasty when an instant message or email arrives
- Avoiding discussions about the use of computers and telephones

Chapter 2

LITERATURE SURVEY

In order to get required knowledge about various concepts related to the present application, existing literature were studied. Some of the important conclusions were made through those are listed below.

Language Features Dadvar et al. proposed gender-specific language features to classify users into male and female groups to improve the discrimination capacity of a classifier for cyberbullying detection. Chen et al. study the detection of offensive language in social media, applying the lexical syntactic feature (LSF) approach that successfully detects offensive content in social media and users who send offensive messages. Dinakar et al. focus on detecting of textual cyberbullying in YouTube comments. They collected videos involving sensitive topics related to race and culture, sexuality, and intelligence. By manually labeling 4,500 YouTube comments and applying binary and multi-class classifiers, they showed that binary classifiers outperform multi-class classifiers.

Social-Structure Features Some researchers consider social-structure features in cyberbullying analysis. Fore example, Huang et al. investigate whether analyzing social network features can improve the accuracy of cyberbullying detection. They consider the social network structure between users and derived features such as number of friends, network embeddedness, and relationship centrality. Their experimental results showed that detection of cyberbullying can be significantly improved by integrating the textual features with social network features. Tahmasbi et al. investigate the importance of considering user's role and their network structure in detecting cyberbullying. Chatzakou et al. extract features related to language, user, and network; then, they study which features explain the behavior of bullies and aggressors the best.

Linguistic and Statistical Analysis Hosseinmardi et al. conducted several studies analyzing cyberbullying on Ask.fm and Instagram, with findings that highlight cultural differences among the platforms. They studied negative user behavior in the Ask.fm social network, finding that properties of the interaction graph—such as in-degree and outdegree—are strongly related to the

negative or positive user behaviors . They studied the detection of cyberbullying incidents over images in Instagram, providing a distinction between cyberbullying and cyber-aggression . They also compared users across two popular online social networks, Instagram and Ask.fm, to see how negative user behavior varies across different venues. Based on their experiments, Ask.fm users show more negativity than Instagram users, and anonymity tends to result in more negativity (because on Ask.fm, users can ask questions anonymously) . Rafiq et al. propose a highly scalable multi-stage cyberbullying detection, which is highly responsive in raising alerts.

Deep Learning Pitsilis et al. applied recurrent neural networks (RNN) by incorporating features associated with users tendency towards racism or sexism with word frequency features on a labeled Twitter dataset. Al-Ajlan et al. applied convolutional neural network (CNN) and incorporates semantics through the use of word embeddings. Zhao et a. extended stacked denoising autoencoder to use the hidden feature structure of bullying data and produce a rich representation for the text. Kalyuzhnaya et al. classify a tweet as racist, sexist, or neither using deep learning methods by learning semantic word embeddings. Dadvar et al. investigate the performance of several models introduced for cyberbullying detection on Wikipedia, Twitter, and Formspring as well as a new YouTube dataset. They found out that using deep learning methodologies, the performance on YouTube dataset increased.

Others Ashktorab et al. provide a number of potential mitigation solutions for cyberbullying among teenagers and the technologies required to implement these solutions. Yao et al. and Zois et al. formulate cyberbullying detection as a sequential hypothesis testing problem to reduce the number of features used in classification. Li et al. propose a method that take advantage of the parent-child relationship between comments to receive the reaction from a third party to detect cyberbullying. Soni et al. [228] use multiple audio and visual features along with textual features for cyberbullying detection. Cheng et al. introduce a framework that creates a heterogeneous network from social media data, and then a node representations.

Gender Shades Joy Buolamwini, an MIT researcher, was using a 3D camera with facial recognition software for her studies on social robots. Because the software was not able to detect her darker skin face, she needed to wear a white mask to complete her research . “Gender

Shades" is a project initiated by Joy to investigate three commercial face recognition systems. They discovered systematic bias in these automated facial analysis algorithms and datasets with

respect to race and gender . They demonstrated that there are significant performance gaps across different populations at classification tasks. The accuracy of all three systems was pretty high in detecting the gender using popular benchmark (around 90%); however, the accuracy of these systems dropped noticeably for females compared to males and for dark-skinned people compared to light-skinned ones. More specifically, the most misclassified group were darker-skinned females with the error rate of roughly 34%.

Wikipedia Wager et al. examined the potential gender inequalities in Wikipedia articles. They found that men and women are covered and featured equally well in many Wikipedia language editions. However, a closer investigation reveals that the way women are portrayed is different from the way men are portrayed. Women on Wikipedia tend to be more linked to men; i.e. the Wikipedia articles about women more often highlight their gender, husbands, romantic relationships, and family-related issues, while this is not true for articles about men. They also discovered that certain words such as "husband" is remarkably more frequent in women-related articles, while "baseball" is more frequent in men-related Wikipedia pages.

Criminal Justice System A recent report by the Electronic Privacy Information Center shows that machine learning algorithms are increasingly used in court to set bail, determine sentences, and even contribute to determinations about guilt or innocence . There are various companies that provide machine learning predictive services such as criminal risk assessment tools to many criminal justice stakeholders. These risk assessment systems take in the details of a defendants profile, and then estimate the likelihood of recidivism for criminals to help judges in their decision-making. Once a suspect is arrested, they are pre-trialed using these risk assessment tools. The results will be shown to the judge for the final decision. The judge then decides to release or to incarcerate that person. A low score would lead to a kinder verdict; A high score does precisely the opposite .

Microsoft Chat Bot Tay was an AI chatbot released by Microsoft via Twitter to mimic and converse with users in real time as an experiment for "conversational understanding." A few hours after the launch of Tay, some Twitter users (trolls) took advantage of Tay's machine learning capabilities and started tweeting the bot with racist and sexist conversations. A few hours later, Tay quickly began to repeat these sentiments back to the users and post inflammatory and offensive tweets . Around 16 hours after its release, Microsoft shut down the Twitter account and deleted Tay's sensitive tweets .

Predicting income Based on a research by Chen et al. , an income-prediction system falsely classified female employees to be low-income much more than wrongly labeling male employees as high-income. They found that this misclassification issue would decrease 40% by increasing the size of training set size ten times.

Beauty.AI The first international online beauty contest judged by artificial intelligence held in 2016 after the launch of Beauty. AI.7 Roughly 6,000 men and women from more than 100 countries submitted their photos to be judged by artificial intelligence, supported by complex algorithms. Out of 44 winners, the majority of them were White, a handful were Asian, and only one had dark skin; while half of the contestants were from India and Africa . Their algorithm was trained using a large datasets of photos; but the main problem was that the data did not include enough minorities; i.e. there were far more images of white women; and many of the dark-skinned images were rejected for poor lighting. This leads to learning the characteristics of lighter skin to be associated with the concept of beauty .

Recommendation Systems Recommender systems have been applied successfully in a wide range of domains, such as entertainment, commerce, and employment . The studies show that in the electronic marketplace, online recommendations can change not only consumers' preference ratings but also their willingness to pay for products . Recommendation systems are seen as tools to accelerate content discover and lead customer engagement. Examples are Netflix's personalized movie recommendations.

Amazon Hiring System In 2014, Amazon created a new recruiting system using machine learning algorithms to review applicants' resumes in order to automatically find the top talent applicants. The system was designed to give job candidates scores ranging from one to five stars . In 2015, they noticed that their system was not rating female applicants similarly to male applicants for technical jobs such as software developer. It was more favorable toward male applicants for technical jobs. The models were trained based on resumes from the past 10 years, which were mostly male applicants. Consequently, Amazon's system prefer men that women for technical jobs. Although they adjusted the system to make it neutral against gender, there was no guarantee that the model wont exhibit discriminatory behaviors in future. Eventually, they stopped the system when they observed such bias in their hiring process using this AI-based system.

Dinakar et al. [12]	2011	Used supervised machine learning approach in which binary & multiclass Classifiers classify bullying sensitive topics.	Label-specific (binary) classifiers are more effective than multiclass classifiers at detecting content-based cybercrime.	Dataset is of stand-alone posts, pragmatics of conversation are not considered, only for supervised learning techniques. Predators and victims were not identified.	Youtube comments from different videos after clustering into Physical appearance, sexuality, race & culture.
Bayzick et al. [13]	2011	Proposed a rule-based system called BullyTracer and also developed a truth-set of MySpace threads to check accuracy of proposed system.	Correctly identify 85.3% as cyberbullying posts and 51.91% as innocent posts of MySpace dataset.	Falsely flag a lot of innocent posts as cyberbullying. Only uses rule-based system, no supervised or unsupervised learning technique was used.	Thread-style forum transcripts crawled from MySpace media. Link: MySpace.com
Reynolds et al. [14]	2011	Supervised machine learning approach in conjunction with labelled data was used to learn the system to identify bullying content.	Model was capable to recognize 78.5% posts in Formspring dataset that have cyberbullying in a small sized sample.	Only for Supervised Learning techniques, pragmatics of conversation are not considered, dataset is of stand-alone posts. Predators and victims were not identified.	18554 user"s data which contain 1 to 1000 posts is used. Link: *www.Formspring.me*
Dinakar et al. [15]	2012	Used common-sense knowledge base with associated reasoning techniques in addition to machine learning classifiers.	In the task of detection of textual cyberbullying, binary classifiers outperform multiclass classifiers.	Other aspects of the problem like pragmatics of conversation and dialogue did not considered by the authors.	Manually Labelled corpus of Youtube and Formspring data. Link: *www.Formspring.me*
Nahar et al. [16]	2012	Proposed a sentimental analysis technique for cyberbullying content detection by using PLSA (a method of feature selection).	Finds the Most Influential persons (predator or Victims) using HITS Algorithm.	Not focused on Indirect Cyberbullying.	Datasets of Kongregate, Slashdot, MySpace websites Link: www.*caw3.barcelonamedia.org*
Nahar et al. [17]	2013	Proposed a session-based framework which incorporated an ensemble of one-class classifier and addressed the real-world scenario where just minimal set of positive instances were given.	Effectively classifies bully instances using session-based one-class ensemble classifier which uses small set of labelled data and huge unlabelled data.	Baseline swear-keywords method can be incorporated along to improve the accuracy.	Datasets of Twitter, MySpace, Kongregate, and Slashdot websites Link: *http://caw2.barcelonamedia.org*

Table 2.1 literature survey

Chapter 3

SYSTEM REQUIREMENT SPECIFICATION

A System Requirement Specification (SRS) is basically an organization's understanding of a customer or potential client's system requirements and dependencies at a particular point prior to any actual design or development work. The information gathered during the analysis is translated into a document that defines a set of requirements. It gives the brief description of the services that the system should provide and also the constraints under which, the system should operate. Generally, SRS is a document that completely describes what the proposed software should do without describing how the software will do it. It's a two-way insurance policy that assures that both the client and the organization understand the other's requirements from that perspective at a given point in time.

SRS document itself states in precise and explicit language those functions and capabilities a software system (i.e., a software application, an ecommerce website and so on) must provide, as well as states any required constraints by which the system must abide. SRS also functions as a blueprint for completing a project with as little cost growth as possible. SRS is often referred to as the "parent" document because all subsequent project management documents, such as design specifications, statements of work, software architecture specifications, testing and validation plans, and documentation plans, are related to it.

Requirement is a condition or capability to which the system must conform. Requirement Management is a systematic approach towards eliciting, organizing and documenting the requirements of the system clearly along with the applicable attributes. The elusive difficulties of requirements are not always obvious and can come from any number of sources.

3.1 Functional Requirement

Functional Requirement defines a function of a software system and how the system must behave when presented with specific inputs or conditions. These may include calculations, data

manipulation and processing and other specific functionality. In this system following are the functional requirements:-

Following are the functional requirements on the system:

1. For detecting bullying content we need a proper data set which is implemented in real time.
2. A proper algorithm for analyzing the data set.
3. The model should give a respective output and it should not deviate from the expected output.

3.2 Non Functional Requirement

Non functional requirements are the requirements which are not directly concerned with the specific function delivered by the system. They specify the criteria that can be used to judge the operation of a system rather than specific behaviors. They may relate to emergent system properties such as reliability, response time and store occupancy. Non functional requirements arise through the user needs, because of budget constraints, organizational policies, the need for interoperability with other software and hardware systems or because of external factors such as:-

- Product Requirements
- Basic Operational Requirements

3.2.1 Product Requirements

Platform Independency: Standalone executables for embedded systems can be created so the algorithm developed using available products could be downloaded on the actual hardware and executed without any dependency to the development and modeling platform.

Correctness: It followed a well-defined set of procedures and rules to compute and also rigorous testing is performed to confirm the correctness of the data.

Ease of Use: Model Coder provides an interface which allows the user to interact in an easy manner.

Modularity: The complete product is broken up into many modules and well-defined interfaces are developed to explore the benefit of flexibility of the product.

Robustness: This software is being developed in such a way that the overall performance is optimized and the user can expect the results within a limited time with utmost relevancy and correctness.

Non functional requirements are also called the qualities of a system. These qualities can be divided into execution quality & evolution quality. Execution qualities are security & usability of the system which are observed during run time, whereas evolution quality involves testability, maintainability, extensibility or scalability.

Chapter 4

SYSTEM DESIGN

Design is a meaningful engineering representation of something that is to be built. It is the most crucial phase in the developments of a system. Software design is a process through which the requirements are translated into a representation of software. Design is a place where design is fostered in software Engineering. Based on the user requirements and the detailed analysis of the existing system, the new system must be designed. This is the phase of system designing. Design is the perfect way to accurately translate a customer's requirement in the finished software product. Design creates a representation or model, provides details about software data structure, architecture, interfaces and components that are necessary to implement a system. The logical system design arrived at as a result of systems analysis is converted into physical system design.

4.1 System development methodology

System development method is a process through which a product will get completed or a product gets rid from any problem. Software development process is described as a number of phases, procedures and steps that gives the complete software. It follows series of steps which is used for product progress.

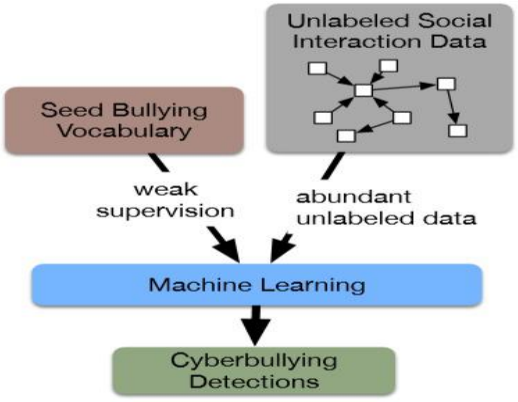

Fig 4.1 methodology

Series Of Steps

PRAW- It is an acronym for python reddit API wrapper, is a python package that allows for simple access to reddit's API.

PRAW can be used for web applications, installed applications and script application.

The reddit class provides convenient access to reddit's API.

Instances of this class are the gateway to interacting with reddit's API through PRAW.

DATA ACQUISITION:

Our dataset is compiled from two datasets online:

• The first dataset contained 25K tweets, each labelled as hate speech, offensive language, or neither

• The second dataset contained 1.6M tweets, each labelled as negative, neutral, or positive

Our data is currently stored as a pickled pandas dataframe.

• Pandas is a library in Python that creates dataframes to store large amounts of data. You can imagine a dataframe to be similar to a spreadsheet, except optimized for Python.

• Pickle is another library in Python that can compress and save Python objects for later use.

ALGORITHMS THAT WE HAVE USED:

1. Bag of words(BoW): It is a representation of text that describes the occurrence of words within a document.

It involves a vocabulary of known words and a measure of the presence of known words and while extracting document vectors we use Boolean value 0 for absent and 1 for present.

The idea is to turn each tweet into a list of word counts. This can be done easily and relatively quickly with CountVectorizer().

As the vocabulary size increases, the vector representation also increases and in such case we are using text cleaning techniques like:

• removing stopwords

• stemming each words

• removing non-alphanumeric characters

• converting everything to lowercase

PROBLEM WITH THIS ALGORITHM: A problem with this algorithm is that with scoring word frequency is that highly frequent words start to dominate in the document, but may not contain as much information content to the model.

One another problem with this approach is that each data gets turned into a very very long list of word counts which can significantly slow down some machine learning algorithms like Support Vector Machines(SVM).

So, to overcome the problem of BoW we have used another algorithm with BoW that is 'Term frequency-Inverse document frequency (TF-IDF)'.

2. Term frequency-Inverse document frequency (TF-IDF):

TF- It is a scoring of the frequency of the word in the current document.

IDF- It is a scoring of how rare the word is across elements.

3. Support vector machine: For the custom feature extraction we have used SVM.

Each additional features causes a significant increase in computational complexity. The word count for each word is its own feature, so BoW and TF-IDF have a very high number of features(we call this high dimensionality) and thus we use Naïve Baiyes in those situations as it suffers greatly from the curse of dimensionality.

Our custom feature extraction is very low dimensionality so SVM is ideal.

ADDING ENSEMBLE LEARNING:

We are using this multiple algorithms at a same time to obtain prediction with an aim to have better prediction than the individual one.

For that we are using the ensemble method 'Voting'.

MODEL TRAINING: We are using Naïve Bayes Classifier when choosing Bow for feature extraction but we use

SVM(Support Vector Machine) when doing a custom feature extraction.

SVM tends to outperform Naïve Bayes in most situations so for custom feature extraction we chose SVM.

MODEL EVALUATION: We are computing three different metrics, that we use to evaluate the performance of our machine learning model.

1. Recall: It deals with the sensitivity of the model that is how many of the cyberbullying

instances does our model successfully identify.

2. Precision: It deals with how reliable our model is when it identifies something as cyberbullying it checks how many positives are false positives.

3. F1-Score: It combines recall and precision into a single metric.

Chapter 5

IMPLEMENTATION

Praw Login

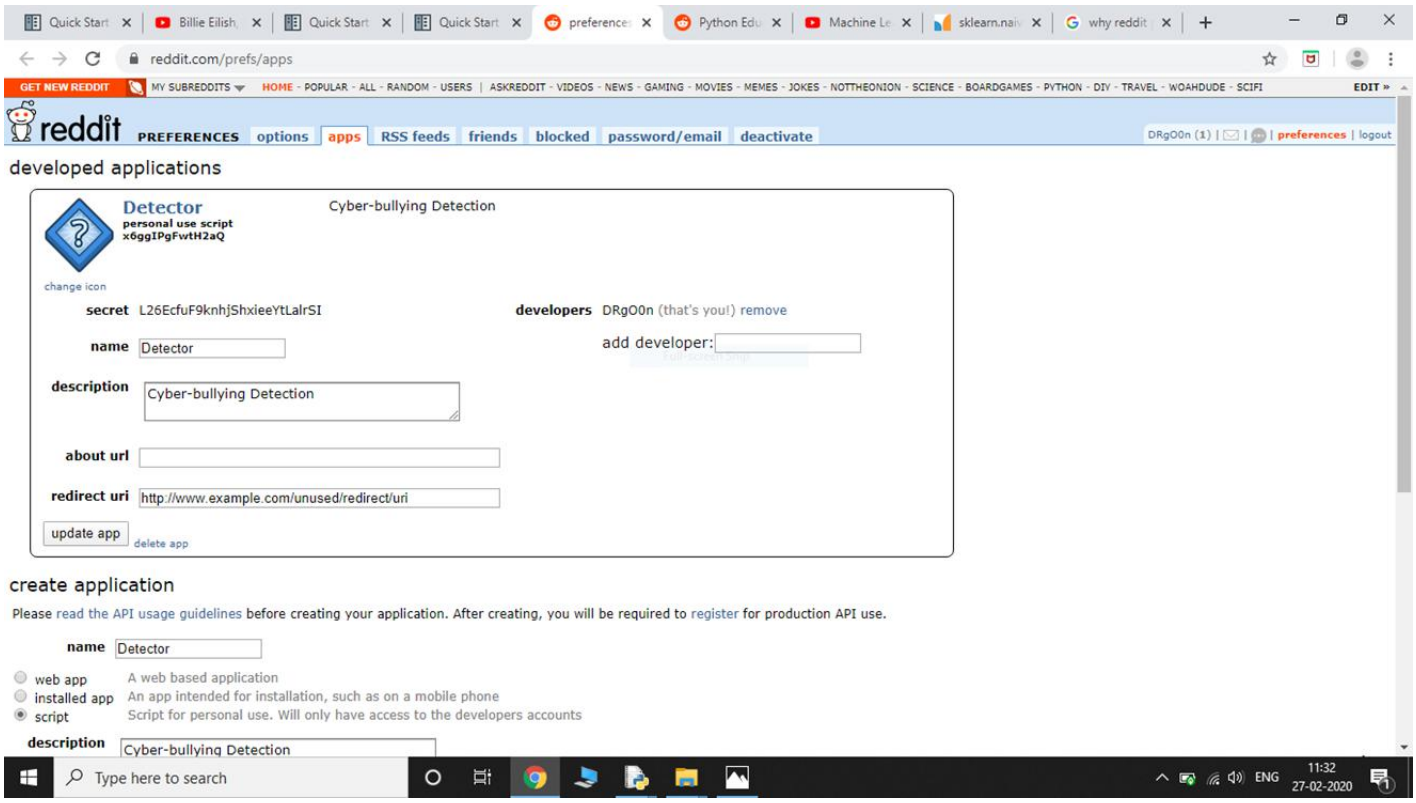

Fig 5.1

Subreddit

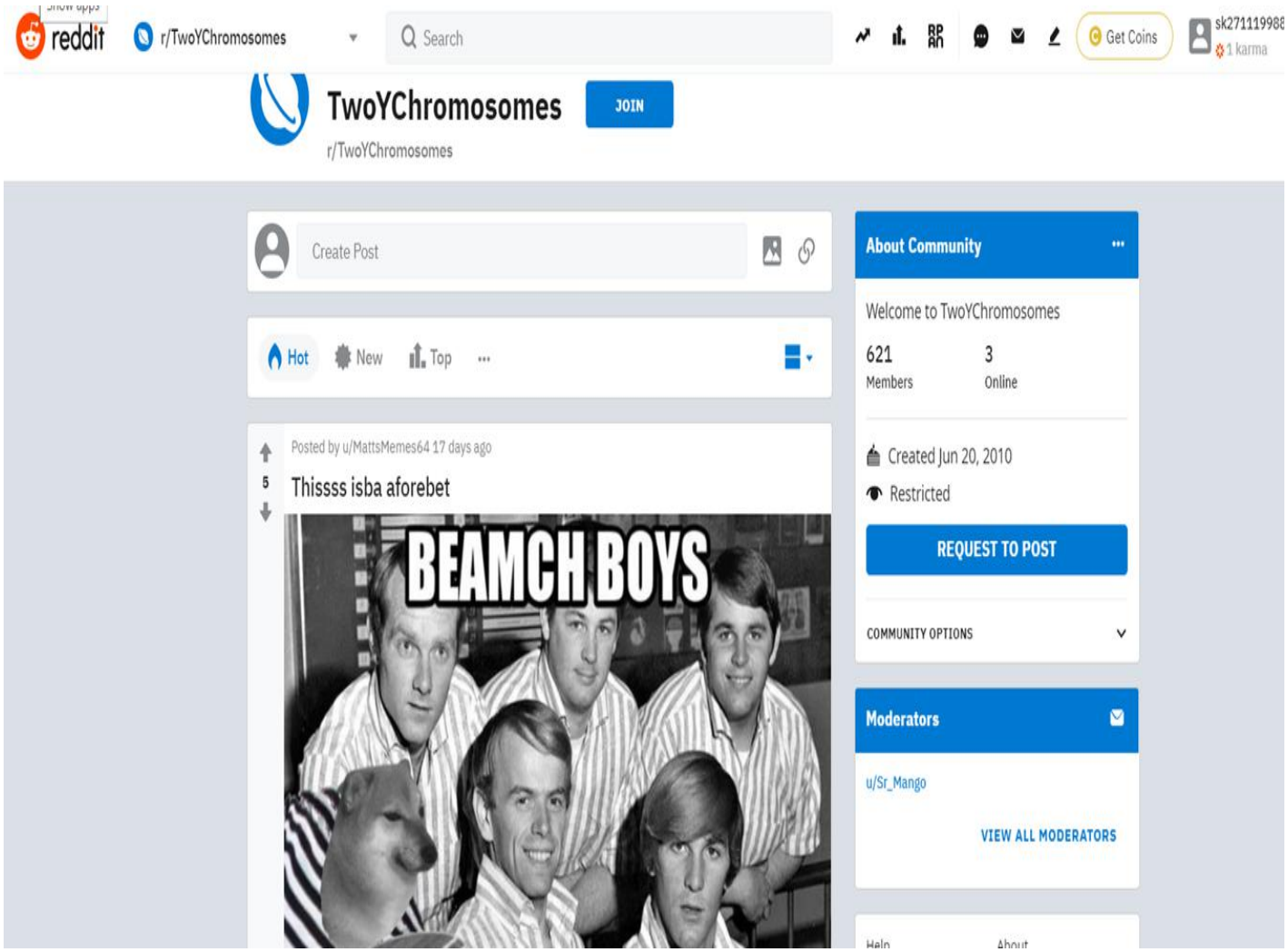

Fig 5.2

Code

```
if __name__ == '__main__':

    reddit = praw.Reddit(

        client_id='x6ggIPgFwtH2aQ',

        client_secret='L26EcfuF9knhjShxieeYtLalrSI',

        user_agent='reddit_script'

    )
    #from sklearn.model_selection import train_test_split
    #xTrain, xTest, yTrain, yTest = train_test_split(x, y, test_size = 0.2, random_state = 0)
    #TrollXChromosomes
    tempp = reddit.subreddit('TwoYChromosomes').hot(limit = 2)


    k=0
    for i in tempp:
        j=0

        print(" POST -> ",k+1," ",i.title)
        print("------------------------------------------------------------------")
    #new_comments = tempp.comments(limit = 1000)
        new_comments = i.comments
    #print(len(new_comments), "LENGTH")

        for comment in new_comments:
            j=j+1
            print(j," ",comment.body,"\n")
        print(j, "COMMENTS")
        print("------------------------------------------------------------------")
       #print(comment.body,"\n")
        k=k+1
    queries = [comment.body for comment in new_comments]



    engine = CyberbullyingDetectionEngine()
    ty = []

    data=engine.load_corpus('./data/final_labelled_data.pkl', 'tweet', 'class')
```

Fig 5.3

```
engine.train_using_bow()
print(engine.evaluate())
print(engine.predict(queries))

engine.train_using_tfidf()

print(engine.evaluate())

print(engine.predict(queries))


engine.load_lexicon('hate-words')

engine.load_lexicon('neg-words')

engine.load_lexicon('pos-words')

engine.load_lexicon('second_person_words')

engine.load_lexicon('third_person_words')

engine.train_using_custom()

print(engine.evaluate())

print(engine.predict(queries))
```

Fig 5.4

Output

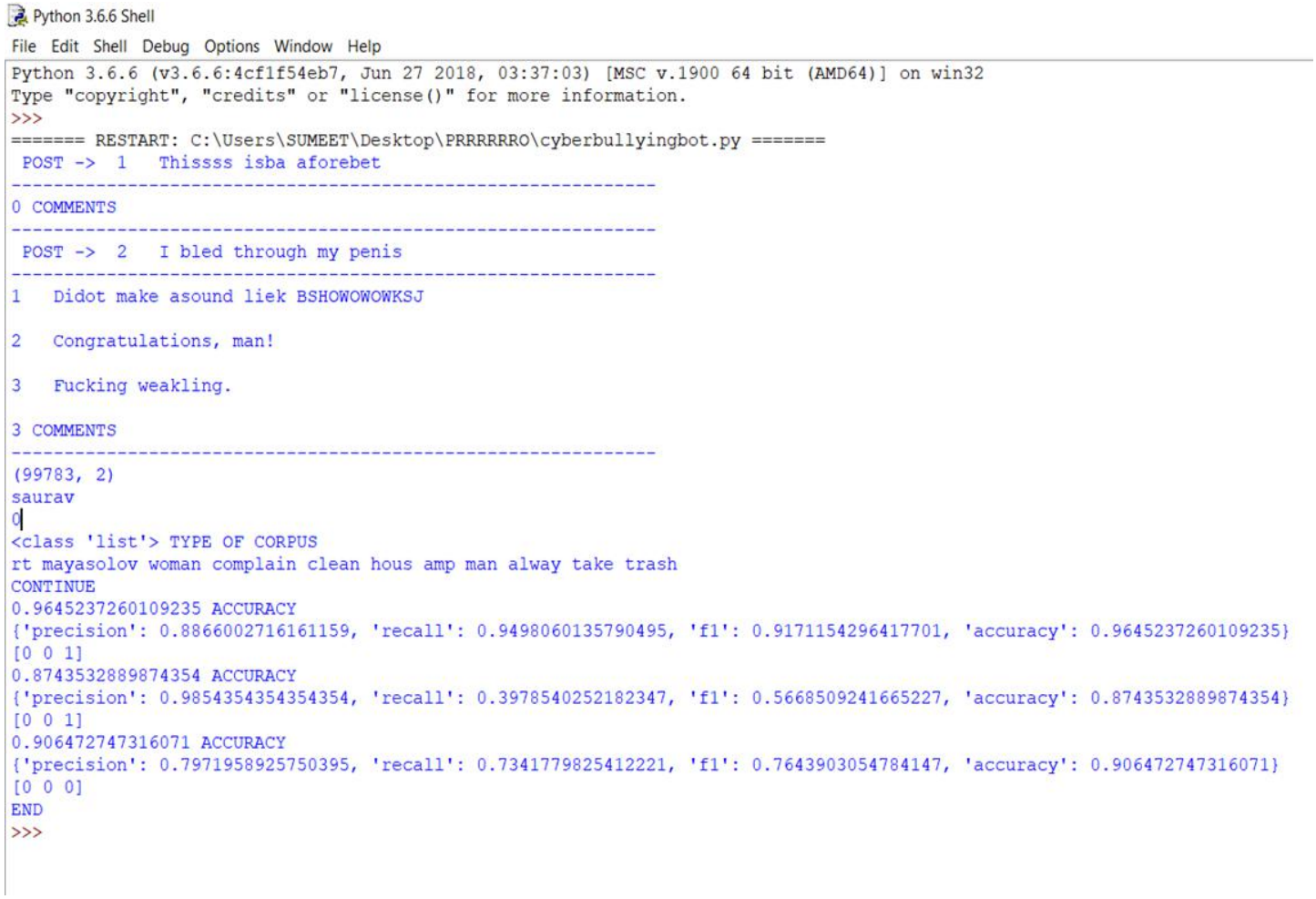

Fig 5.5

Chapter 6

FUTURE SCOPE AND CONCLUSION

Using ensemble learning to combine different machine learning models (such as Naive Bayes classifier with the TF-IDF word vectors and the Support Vector Machine using our custom word vectors) to improve the model evaluation metrics.

Creating more detailed custom word vectors for the SVM to train creating more detailed custom word vectors for the SVM to train.

As we know, social media became a common platform for most of the people where they share their views but few teenagers are there who perform bullying on these types of platforms and that bullying may disturb someone mentally and emotionally. So, In order to stop this type of teasing or bullying, we developed our project to detect cyberbullying using machine learning.

Early detection of harmful social media behaviors such as cyberbullying is necessary for identifying threatening online abnormalities and preventing them from increasing. So, In this project we successfully fetched the comments from the subreddit using praw, and I also able to identify the vulgar comments by using three machine learning algorithm bag of words, term frequency inverse document frequency, support vector machine.

REFERENCES

1. V. Balakrishnan. Cyberbullying among young adults in malaysia. Comput. Hum. Behav., 46(C):149–157, May 2015.

2. J. Bayzick, A. Edwards, and L. Edwards. Detecting the presence of cyberbullying using computer software. 02 2019

3. https://scholar.google.com//

4. https://www.coursera.org/

5. https://ieeexplore.ieee.org/search/searchresult.jsp?newsearch=true&contentType=courses&queryText=ml

6. http://www.w3schools.com/

7. https://www.dosomething.org/us/facts/11-facts-about-cyber-bullying