SENTIMENT/CYBER BULLYING DETECTION USING PYTHON & MACHINE LEARNING

Sanjay Patil, Anjeli Sawant, Megha Waghmare, Chaitali Sonawane

Under the guidance of Prof. Vikram Popat Deokate

DEPARTMENT OF COMPUTER ENGINEERING

Al Ameen Educational & Medical Foundation's College of Engineering Koregaon Bhima, Pune-412216

ABSTRACT

In the modern era, the usage of the internet has increased tremendously which in turn has led to the evolution of large amount of data. Cyber world has its own pros and cons. One of the alarming situations in web 4.0 is cyber bullying a type of cyber-crime. When bullying occurs online with the aid of technology it is known as cyber bullying. This research paper has surveyed the work done by 30 different researchers on cyber bullying, and elaborated on different methodologies adopted by them for the detection of bullying.

Three types of features namely textual, behavioral and demographic features are extracted from the dataset as compared to earlier study over the same dataset where only textual features were considered. Textual features include certain bullying words that if exists within the text may lead to a true outcome for cyber bullying. Personality trait features are extracted for the user if it is involved once in bullying may bully in future too. While demographic features extracted from dataset include age, gender and location. The system is evaluated through different performance measures for both classifiers used the performance of the Support Vector Machine classifier is found better than the Bernoulli NB with overall 87.14 Accuracy.

Keywords: *Cyberbully, Support Vector Machine, Textual, classifier.*

CHAPTER 1: INTRODUCTION

1.1 OVERVIEW

Across the globe due to the tremendous increase in the availability of data services, addiction of social media among the society has increased proportionally. Just like other countries, India has also witnessed a drastic rise in the cyber bullying. In this era of web
4.0 where people live in digital and online platforms, it is very difficult to protect the society from the alarming risein cyber-crime. It has been surveyed that the major victims of cyber bullying are adolescents. Different cyber bullying attacks that are performed by attacker are: (1) Sending or posting hateful or abusive comments with an intention to harm the character of an individual (2) Posting an inappropriate image or video. (3) Creation of a false or improper website. (4) Issuing online threats that cause a person to kill themselves or injure another person. (5) Triggering online religious, racial, ethnic or political hatred by posting hate comments or videos.

1.2 MOTIVATION

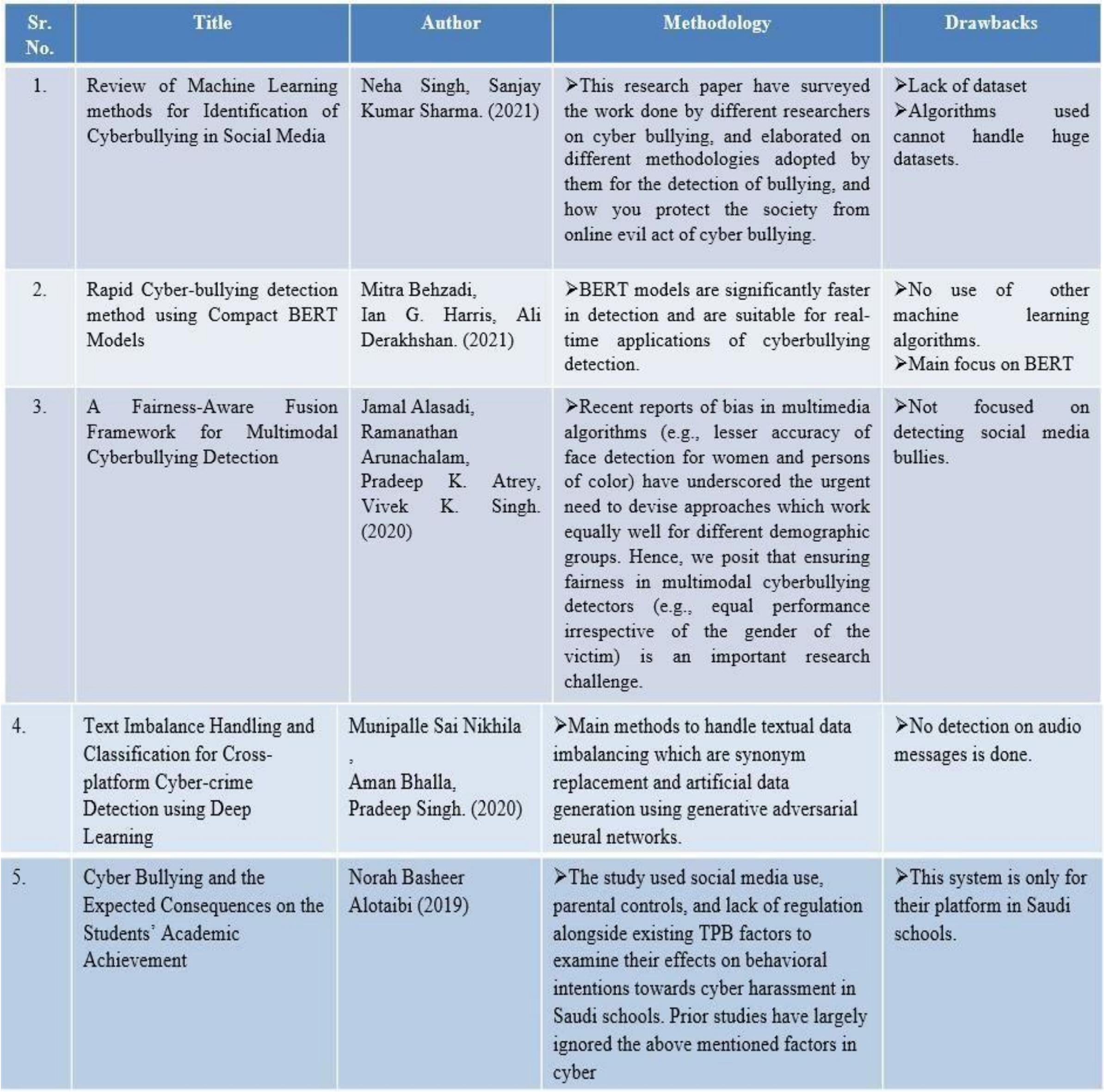
The main Motivation is to Avoid Cyber Bullying and save student or any human life. Althoughsome users indicate they are being sarcastic, most of them do not. Therefore, it might be indispensable to and a way to automatically detect any sarcastic messages Cyber bullying is threatening and destructive act which may result in suicide attempts or negative impact can cause life-long harms to the victims. The detection of Cyber bullying can be considered as a classification problem. An online post can be classified as a bullying post or normal post. We will develop a system by applying different machine learning methods to better detect Cyber bullying and improve performance. For example, the Support Vector Machine (SVM), Forest Classifier

1.3 OBJECTIVE

The objective of the system is to reveal, analyze and stop cyber bullying in social media applications. To identify the occurrence of cyber bullying activity in social media platform which helps the government to yield force before many end-users enhance their Target of cyberbullying. To The system is to give alert message like to warn them, and to identify short hand text and human aggressive behavior on the comment sections. To Also to generate a report which contains the details of bully, and to keep track of count and also by blocking that person along his comment without letting it reaches to victim.

CHAPTER 2: LITERATURE SURVEY

2.1 STUDY OF RESEARCH PAPER



Sr. No.	Title	Author	Methodology	Drawbacks
1.	Review of Machine Learning methods for Identification of Cyberbullying in Social Media	Neha Singh, Sanjay Kumar Sharma. (2021)	➢This research paper have surveyed the work done by different researchers on cyber bullying, and elaborated on different methodologies adopted by them for the detection of bullying, and how you protect the society from online evil act of cyber bullying.	➢Lack of dataset ➢Algorithms used cannot handle huge datasets.
2.	Rapid Cyber-bullying detection method using Compact BERT Models	Mitra Behzadi, Ian G. Harris, Ali Derakhshan. (2021)	➢BERT models are significantly faster in detection and are suitable for real-time applications of cyberbullying detection.	➢No use of other machine learning algorithms. ➢Main focus on BERT
3.	A Fairness-Aware Fusion Framework for Multimodal Cyberbullying Detection	Jamal Alasadi, Ramanathan Arunachalam, Pradeep K. Atrey, Vivek K. Singh. (2020)	➢Recent reports of bias in multimedia algorithms (e.g., lesser accuracy of face detection for women and persons of color) have underscored the urgent need to devise approaches which work equally well for different demographic groups. Hence, we posit that ensuring fairness in multimodal cyberbullying detectors (e.g., equal performance irrespective of the gender of the victim) is an important research challenge.	➢Not focused on detecting social media bullies.
4.	Text Imbalance Handling and Classification for Cross-platform Cyber-crime Detection using Deep Learning	Munipalle Sai Nikhila , Aman Bhalla, Pradeep Singh. (2020)	➢Main methods to handle textual data imbalancing which are synonym replacement and artificial data generation using generative adversarial neural networks.	➢No detection on audio messages is done.
5.	Cyber Bullying and the Expected Consequences on the Students' Academic Achievement	Norah Basheer Alotaibi (2019)	➢The study used social media use, parental controls, and lack of regulation alongside existing TPB factors to examine their effects on behavioral intentions towards cyber harassment in Saudi schools. Prior studies have largely ignored the above mentioned factors in cyber	➢This system is only for their platform in Saudi schools.

CHAPTER 3:

PROBLEM DEFINITION

3.1 PROBLEM STATEMENT

Cyber bullying is a problem in current Situation of the world because All the Student or humans uses social media. To avoid Cyber bullying through project. The purpose is to find an efficient way to detect sarcastic tweets, and study how to use this information (i.e., whether the tweet is sarcastic or not) to enhance the accuracy of Cyber bullying.

3.2 PROPOSED SYSTEM AND METHODOLOGY

This section briefly discusses the research methodology that we used for detecting the severity of cyberbullying with the dataset described in section 3. All steps of our proposed framework are presented in figure 3.1 and discussed in the following sections.

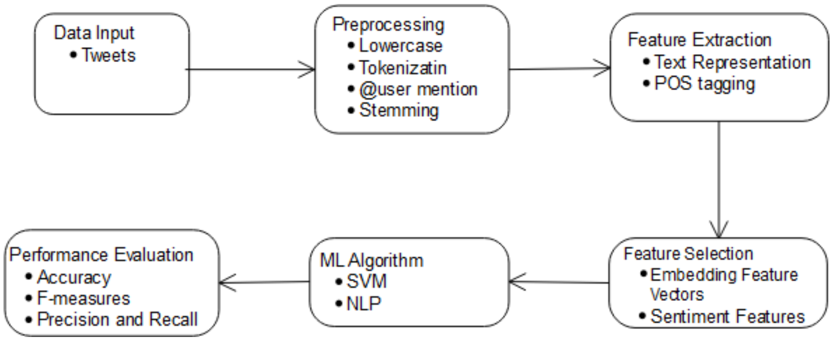


Figure 3.1: Proposed System

3.2.1 Data Collection

For pursuing the task of sentiment analysis, dataset availability is important. The performance of a classifier depends on the quality and size of dataset. Several datasets are available online. However, to perform real time analysis, real time data is required. Such data can be collected using the Twitter API. Once the dataset has been prepared, it has to be split into training and test datasets. The experiments were carried out on three different datasets, two unlabeled and one labilities dataset consisted of 37,411 tweets that were fetched using popular keywords associated with cyber-bullying. It consisted of recent tweets (a period of two weeks prior to the current date). It was assumed that such keywords will capture tweets covering a) Racial Bias b) Religious Bias c) Gender Bias.

3.2.2 Pre-processing

The collected data was pre-processed before assigning severity levels. Tweets were converted to lower case to avoid any sparsity issue, reduced repeated letters, standardized URLs and @usermention to remove noise in the tweets. Tokenization was applied with Twitter-specific tokenizer based on the CMU Tweet NLP library and only words with minimum frequency of 10 were kept. Tokenization is the process of breaking a text corpus up into most commonly words, phrases, or other meaningful elements, which are then called tokens. Finally, stop-words and stemming procedures were performed before feature extraction. Stop words are defined as insignificant words that appear in document which are not specific or discriminatory to the different classes. Stemming refers to the process of reducing words to their stems or roots. For instance, singular, plural and different tenses are consolidated into a single word. We applied stemming with an iterated version of the Lovin's stemmer, it stems the word until it no further changes prior to extracting topic model features.

3.2.3 Feature Extraction Step

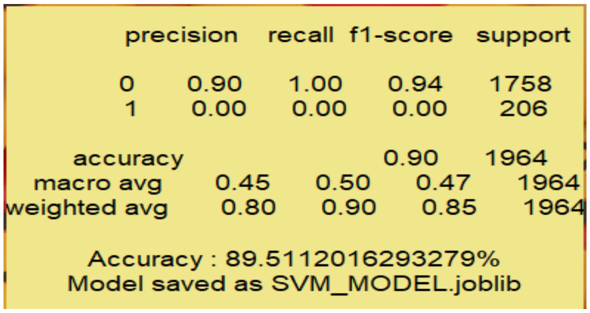
All tweets were represented with bag-of-words which is one of the most appropriate and quickest approaches. In this approach, text is represented by set of words and each word is treated as an independent feature. We applied part-of-speech (POS) tagging with Twitter-specific tagger based on the CMU TweetNLP library for word sense disambiguation. The POS tagger assigns part-of-speech tag to each word of the given text in the form of tuples (*word*, *tag*), for instance, noun, verb, adjectives, etc.

3.2.4 SVM

Choosing the best classifier is the most significant phase of the text classification pipeline. We cannot efficiently determine the most effective model for a text classification implementation without a full conceptual comprehension of each algorithm. The features obtained from the tweets have been used to build a model to detect cyberbullying behaviors and its severity. In order to select the best classifier, we tested several machine learning algorithms namely: Naïve Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, and K-Nearest Neighbors (KNN).

Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features.

SVM Kernel: The SVM kernel is a function that takes low dimensional input space and transforms it into higher-dimensional space, i.e., it converts not separable problem to separable problem. It is mostly useful in non-linear separation problems. Simply put the kernel, it does some extremely complex data transformations then finds out the process to separate the data based on the labels or outputs defined SVM classifier with linear kernel gave an accuracy of 55- 57% over multiple iterations. Multinomial Naive Bayes with ternary classification gave an accuracy of 57 %. But an overall accuracy of 80 % was achieved in the task of binary classification (bullying/non-bullying). This accuracy can be attributed to the lack of clear distinction between neutral/ non-abusive tweets.
Supervised Machine Learning on labelled dataset. The task of ternary classification was performed on dataset using SVM classifier with a training to test split of 7:3, achieving an overall accuracy of 87.51%. The results are shown in below figure.



```
              precision    recall  f1-score   support

           0       0.90      1.00      0.94      1758
           1       0.00      0.00      0.00       206

    accuracy                           0.90      1964
   macro avg       0.45      0.50      0.47      1964
weighted avg       0.80      0.90      0.85      1964

      Accuracy : 89.5112016293279%
    Model saved as SVM_MODEL.joblib
```

Figure 3.1 : Result

3.2.5 Natural Language Processing

NLP enables computers to understand natural language as humans do. Whether the language is spoken or written, natural language processing uses artificial intelligence to take real-world input, process it, and make sense of it in a way a computer can understand. Just as humans have different sensors -- such as ears to hear and eyes to see computers have programs to read and microphones to collect audio. And just as humans have a brain to process that input, computers have a program to process their respective inputs. At some point in processing, the input is converted to code that the computer can understand.

There are two main phases to natural language processing: data preprocessing and algorithm development

- **Tokenization.** This is when text is broken down into smaller units to work with.

- **Stop word removal.** This is when common words are removed from text so unique words that offer the most information about the text remain.

- **Lemmatization and stemming.** This is when words are reduced to their root forms to process.

- **Part-of-speech tagging.** This is when words are marked based on the part-of speech they are -- such as nouns, verbs and adjectives.

Once the data has been preprocessed, an algorithm is developed to process it. There are many different natural language processing algorithms, but two main types are commonly used:

- Rules-based system. This system uses carefully designed linguistic rules. This approach was used early on in the development of natural language processing, and is still used.

- Machine learning-based system. Machine learning algorithms use statistical methods. They learn to perform tasks based on training data they are fed, and adjust their methods as more data is processed. Using a combination of machine learning, deep learning and neural networks, natural language processing algorithms hone their own rules through repeated processing and learning.

3.2.6 Performance Evaluation & Analysis

In order to evaluate the performance of a classifier, we observe the accuracy of the classifier. However, if the data is asymmetric, i.e., if the number of False Positives and False negatives are not same, then other parameters are considered. The main parameters are discussed here in detail with reference to our cyberbullying detection work.

Precision: It is the ratio of instances that actually lies in that class to the net total of instances that are classified in that same class. This denotes how many tweets are actually instances of cyber-bullying out of all tweets labelled as cyber-bullying.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

Recall: It is the measure of sensitivity which recalls out of all the true cyber-bullying tweets how many of them were labelled as cyber-bullying by us. It is defined as the ratio of instances that are classified to lie in a given class to that of its true number of instances lying in that class.

$$Recall = \frac{TruePositives}{TruePositives + FalsePositives}$$

F-measure: It is entirely dependent on precision and recall which gives the weighted harmonic mean of both. It ranges from 0 to 1 where 0 denotes the worst value and 1 denotes the best value.

$$\mathrm{F-measure} = \frac{2 * \mathrm{Precision} * \mathrm{Recall}}{Precision + Recall}$$

CHAPTER 4 : SYSTEM ANALYSIS

4.1 SYSTEM ARCHITECTURE

The Software Architecture Diagram is a crucial step for software and application developers to describe the basic software structure by separating functional areas into layers. It depicts how a typical software system might interact with its users, external systems, data sources, and services.

Figure 4.1 shows the system architecture in which first we have taken dataset for twitter.

Pre-Processing

Pre-processing is done on dataset where dataset gets transformed into machine understandable language using Tokenization method. It helps to clean, format, and organize the raw data, thereby making it ready-to-go for machine learning models.

Feature Extraction

Machine learning-based automatic cyberbullying detection involves two steps: Representation Learning for Tweets and Classification. Each tweet is converted into a fixed length vector. This constitutes the feature vector space. So higher the features, higher is the dimension of the feature vector space and this accounts to more processing and storage requirements.

SVM

SVM classifier with linear kernel gave an accuracy of 55- 57% over multiple iterations. Multinomial Naive Bayes with ternary classification gave an accuracy of 57 %. But an overall accuracy of 80 % was achieved in the task of binary classification (bullying/non- bullying). This accuracy can be attributed to the lack of clear distinction between neutral/ non-abusive tweets.

After applying algorithm tweets are detected in six types of bully i.e., are normal tweet, religion bully, age bully, other type of bully sentences, gender bully sentence and ethnicity type bully sentence. Once bully sentences are predicted user can take proper action on that user who is posting bully tweets and messages.

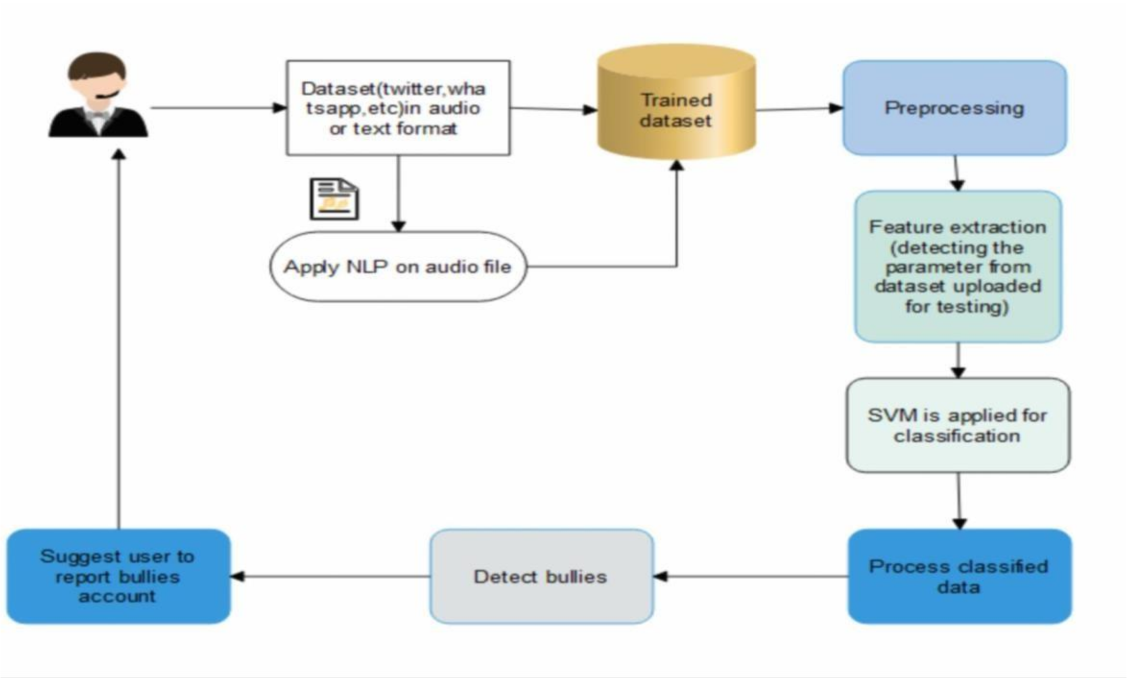

Dataset(twitter,whatsapp,etc)in audio or text format
Trained dataset
Preprocessing
Apply NLP on audio file
Feature extraction (detecting the parameter from dataset uploaded for testing)
SVM is applied for classification
Suggest user to report bullies account
Detect bullies
Process classified data

Figure 4.1: System Architecture

4.1 Data Flow Diagram

In Data Flow Diagram, we Show that flow of data in our system in DVD we show that base DFD(figure 6.2) in which rectangle present input as well as output and circle show our system, In DFD1(figure 6.3) show actual input and actual output of system input of our system is text or image and output is rumor detected likewise in DFD 2(figure 6.3) present operation of user as well as admin.

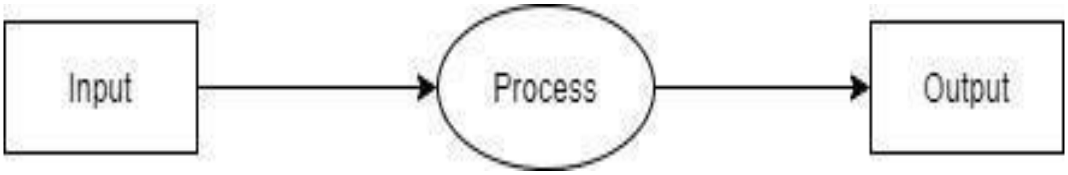


Figure 4.2: Data Flow (0) diagram

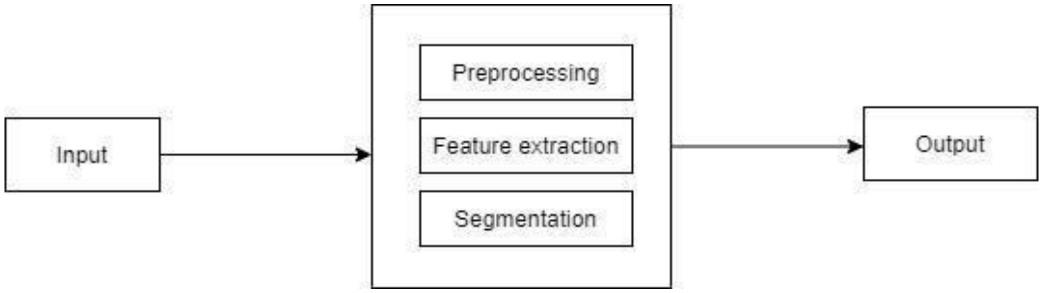


Figure 4.3: Data Flow (1) diagram

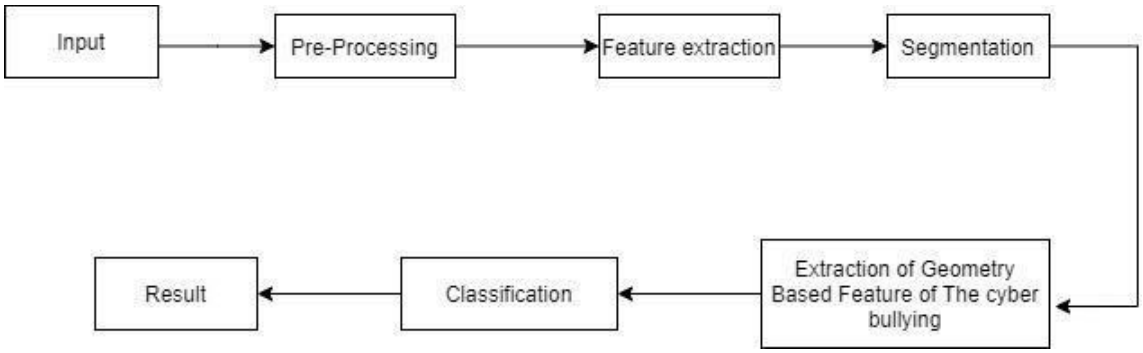


Figure 4.4: Data Flow (2) diagram

4.3 UML Diagrams

Unified Modeling Language is a standard language for writing software blueprints. The UML may be used to visualize, specify, construct and document the artifacts of a soft- ware intensive system. UML is process independent, although optimally it should be used in process that is use case driven architecture centric, iterative, and increment. The Number of UML Diagram is available.
Use case Diagram. (Figure 4.5) Class Diagram. (Figure 4.6) Activity Diagram. (Figure 4.7) Sequence Diagram. (Figure 4.8)

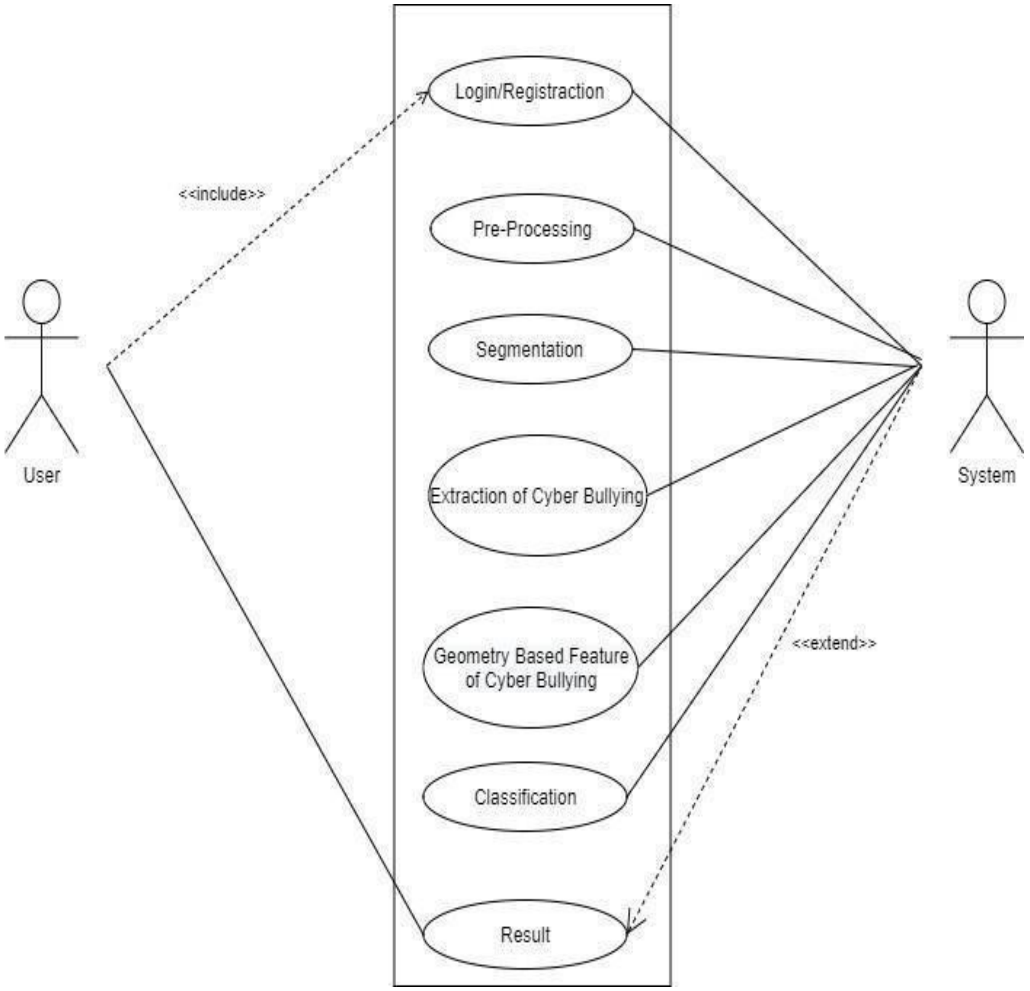


Figure 4.5: Use case Diagram

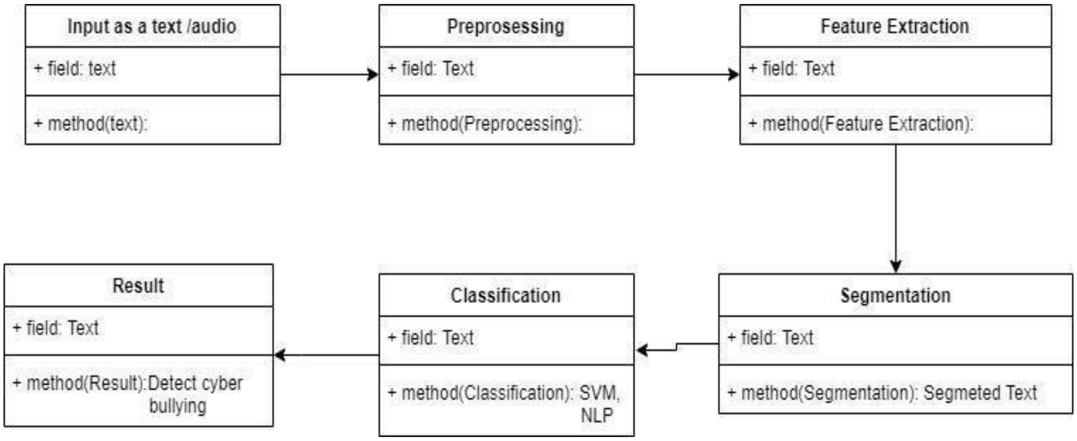


Figure 4.6: Class Diagram Diagram

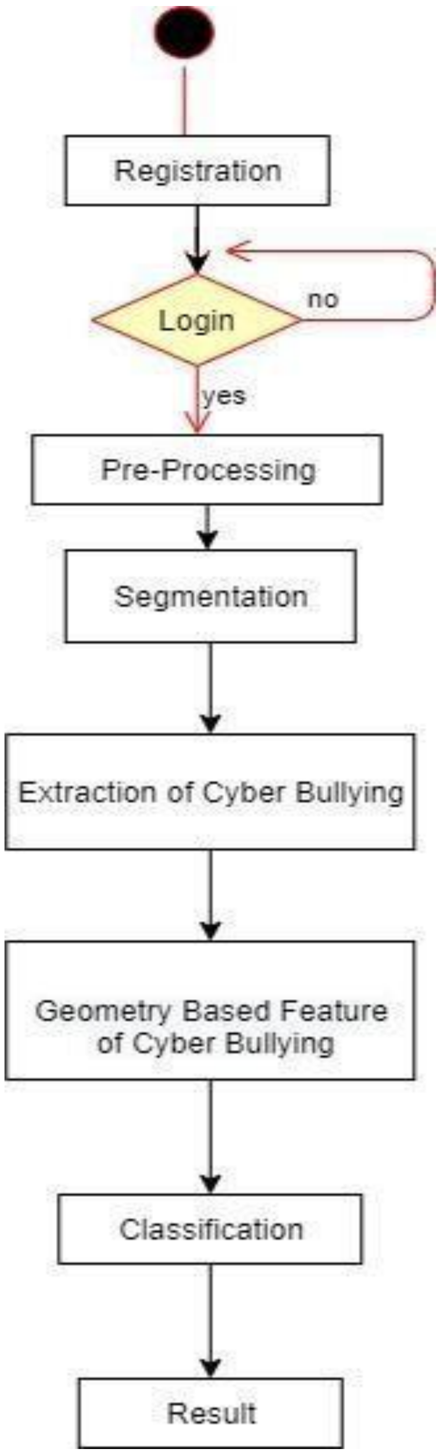

Registration
Login
no
yes
Pre-Processing
Segmentation
Extraction of Cyber Bullying
Geometry Based Feature of Cyber Bullying
Classification
Result

Figure 4.7: Activity Diagram

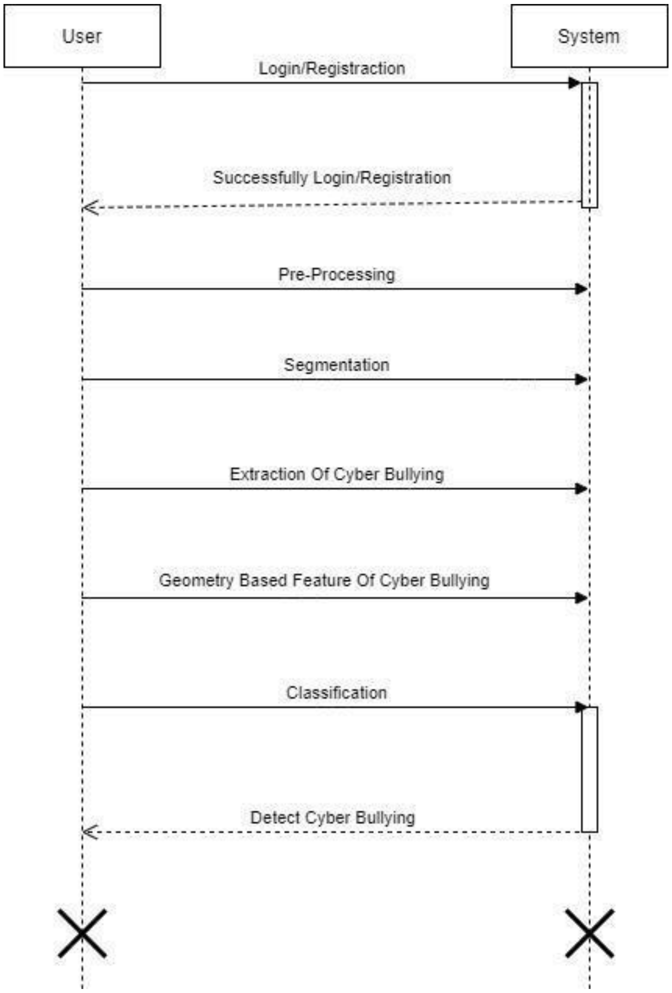

User
System
Login/Registraction
Successfully Login/Registration
Pre-Processing
Segmentation
Extraction Of Cyber Bullying
Geometry Based Feature Of Cyber Bullying
Classification
Detect Cyber Bullying

Figure 4.8: Sequence Diagram

CHAPTER 5: APPLICATIONS

5.1 APPLICATIONS:

- Using purposed system, everyone will feel safe to use social networking sites(SNS).

- If anyone did suicide or some crimes then purposed system is very useful for cyber police. They can easily find out culprits.

- If anyone trying to bully on social media platforms then purposed system quickly identifies that user and help us to report and block that account.

CHAPTER 6: TESTING

6.1 TEST PLANS

A system should always be tested thoroughly before implementing it, as regards its individual programs. This is because implementing a new system is a major job which a lot of man hours and a lot of other resources, so an error not detected be- fore implementation may cost a lot. Effective testing early in the process translates directly into long term cost saving from reduced number of errors. This is also necessary because in some cases, a small error is not detected and corrected before installation, which may explode into much larger problem. Programming and testing are followed by the stage of installing the new computer-based system. Actual implementation of the system can begin at this point using either a parallel or a direct changeover plan, or some blend of two. Testing and implementation of fire fighting robot controlled using android application is carried out as below. Software testingis a critical element of Software Quality Assurance(SQA) and represents the ultimate review of specification, design and coding. The purpose of product testing is to verify and validate the various work products viz. units, integrate unit, final product to ensure that they meet their respective requirements.

6.2 TESTING PROCEDURE

Software Testing is the critical element of the Software Quality Assurance and represents the ultimate review of specification, design and coding. Testing is the process of checking whether software works according to the specification. Testing will beperformed by running the program using the test data. Testing is vital to the success of the system. It will also test whether the system identifies the problem correctly.
System is tested by following steps:

- Unit Testing: Each program is tested individually using dummy records to see whether that program produce satisfactory reports.
- Sequential Testing: The program, whose output will affect the processing done by another program, will be tested using dummy records. Testing: The system is corrected in such a way that it does not affect the forced system failure. This testing is done with low volumes of data.

6.3 TEST STRATEGY

The test strategy consists of a series of different tests that will fully exercise the system. The primary purpose of the test is to uncover the system limitations. Following are the several tests that will be conducted:

6.3.1 Unit Testing:

Testing conducted to verify the implementation of the design for one software element (e.g., unit, module) is called unit testing. The purpose of unit testing is to ensure that the program logic is complete and correct and ensuring that the com- ponent works as designed. In this module, each unit will go through Unit testing after the completion of the module. The bugs in module testing will be reported in Test Log document and will be reported to the developers. After fixing the bug successfully, one more iteration of module testing (Regression Testing) is done. This process is repeated till all critical test cases pass.

6.3.2 Integration Testing:

Testing conducted in which software elements, hardware elements, or both are combined and tested until the entire system has been integrated. The purpose of integration testing is to ensure that design objectives are met and ensures that the software, as a complete entity, complies with operational requirements. This type of testing will be done after all module test cases are passed through module testing, securitytesting, performance testing, user interface testing and regression testing.

6.3.3 Performance Testing:

In developing the system, we are going to use Java which will reduce the response time. In Performance Testing, We are going to test Response time for each Screen. It is a type of non-functional testing. Performance testing is testing that is performed; to determine how fast some aspect of a system performs under a particular work- load. It can serve different purposes like it can demonstrate that the system meets performance criteria. It can compare two systems to find which performs better. Or it can measure what part of the system or workload causes the system to perform badly. This process can involve quantitative tests done in a lab, such as measuring the response time or the number of MIPS (millions of instructions per second) at which a system function.

6.3.4 Regression Testing:

Testing done to ensure that, the changes to the application have not adversely affected previously tested functionality. Here testing will take care of the test cases passed during the first module testing will not be affected in the subsequent rounds of module testing.

6.4 TEST CASES

The listed tests were conducted in the software at the various development stages. Unit testing was conducted. The errors were debugged and regression testing was performed. The integration testing will be performed once the system is integrated with other related systems like Inventory, Budget etc. Once the design stage was over the Black Box and White Box Testing was performed on the entire application. The results were analyzed and the appropriate alterations were made. The test results proved to be positive and henceforth the application is feasible and test approved.

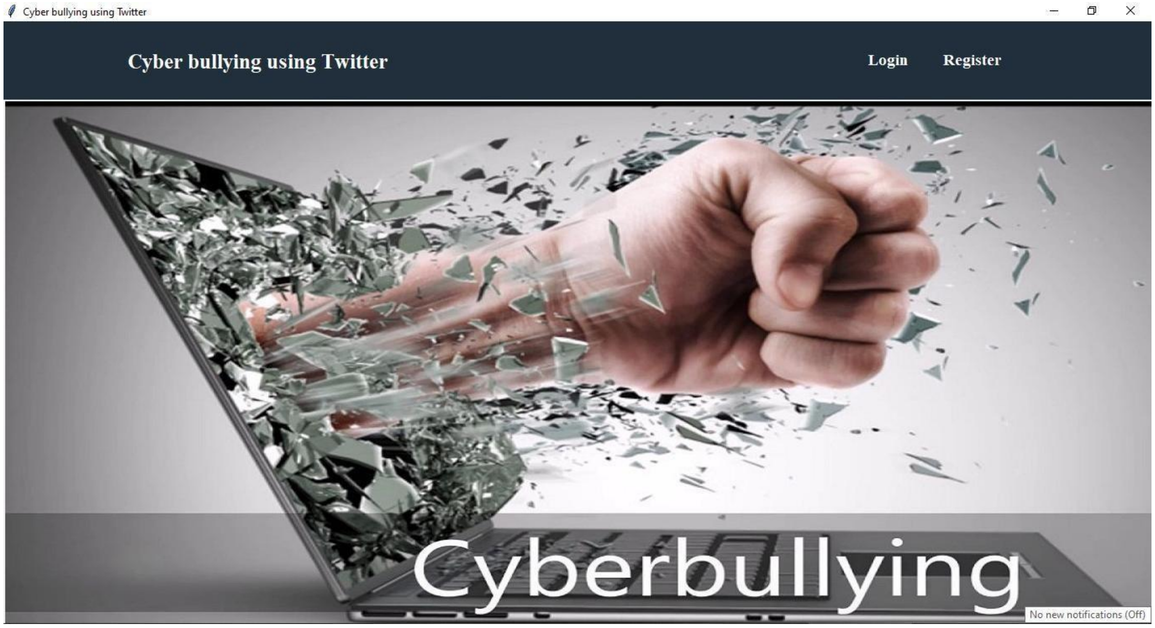
6.5 TEST CASES RESULT

Sr.No	Description	Test Case I/P	Actual Result	Expected	Test Criteria (P/F)
1	Install Python	Python Exe	Should get install properly	Proper Installed	P
2	Installing Libraries	Library command for install	Should Get installed	Library Installed Success-fully	P
3	Training Dataset	Dataset Training	Error in Training Model	Trained Model	F
4	Training Dataset	Dataset Training	Trained Model	Trained Model	P
5	Login Cre- details	User Name and Pass- word	Login Un- successful	Unsuccessful Login	F
6	Login Cre- details	User Name and Pass- word	Login Successful	SuccessfulLogin	P
7	Password	Current and New Pass- word	Password Updated	Update Password	P
8	Prediction	Video as in- put	Should Predict the result	Result Predicted	P

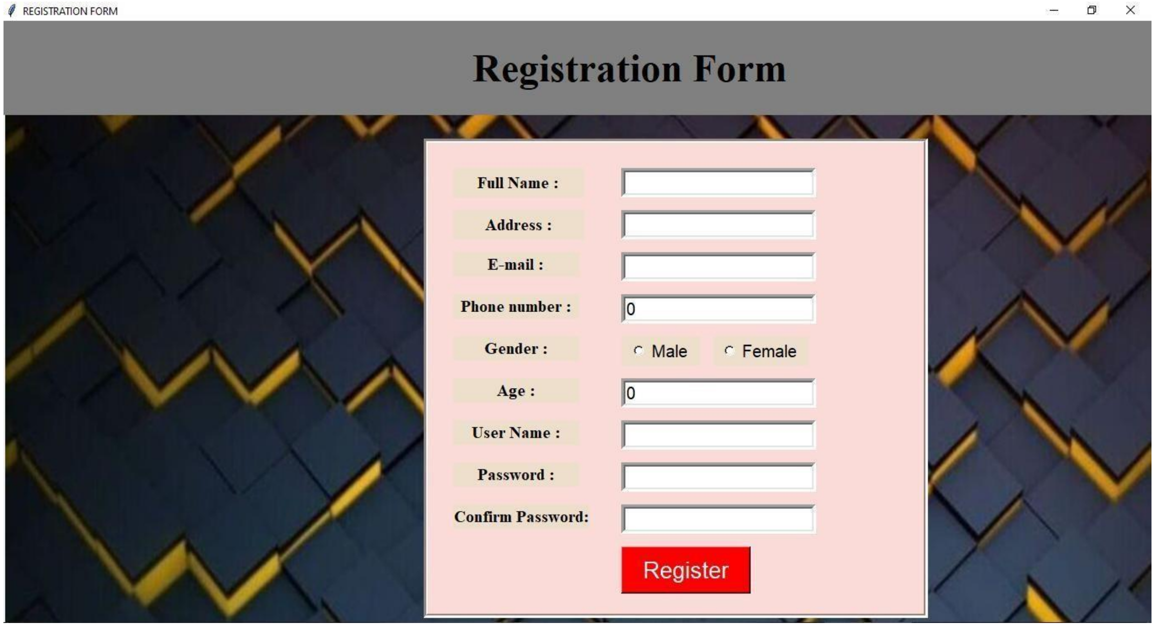
Table 6.1: Test Cases

CHAPTER 7: RESULTS

Home Page


Cyber bullying using Twitter
Cyber bullying using Twitter
Login
Register
Cyberbullying
No new notifications (Off)

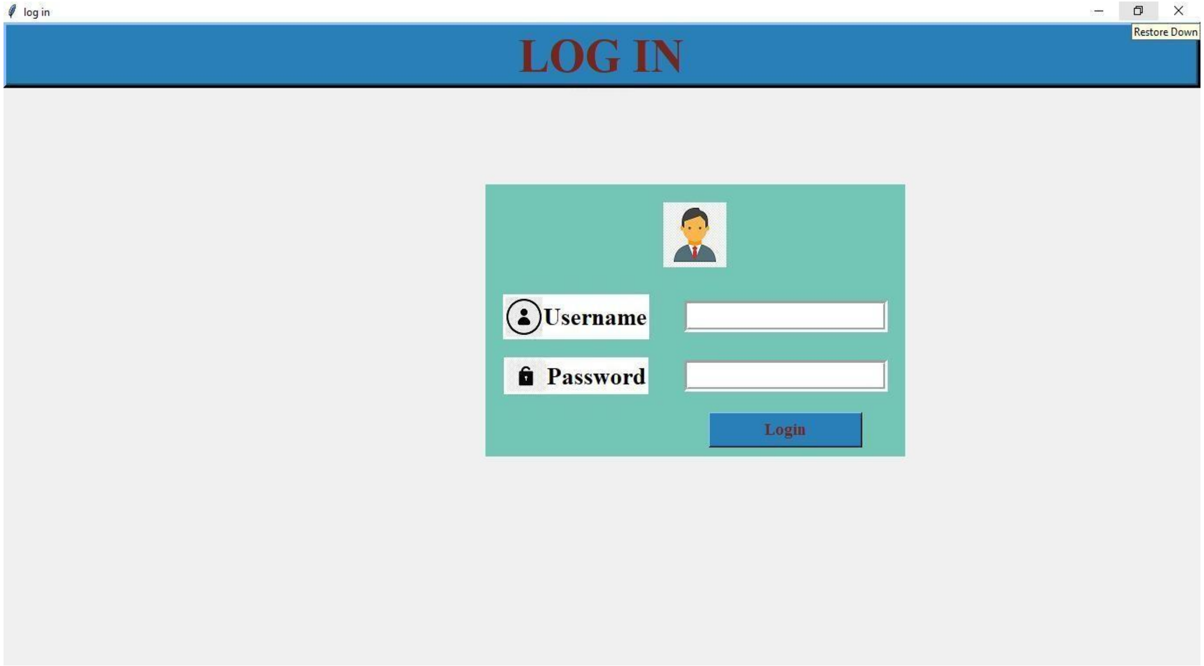
Registration Window



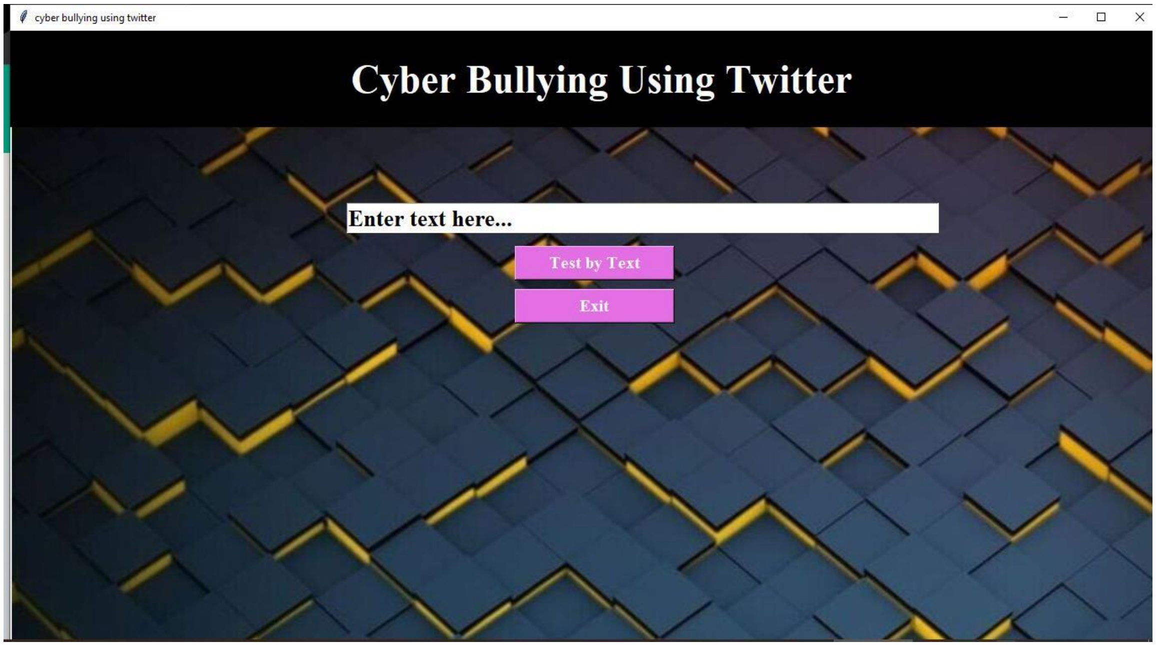
Registration Form

Full Name :	
Address :	
E-mail :	
Phone number :	0
Gender :	○ Male ○ Female
Age :	0
User Name :	
Password :	
Confirm Password:	

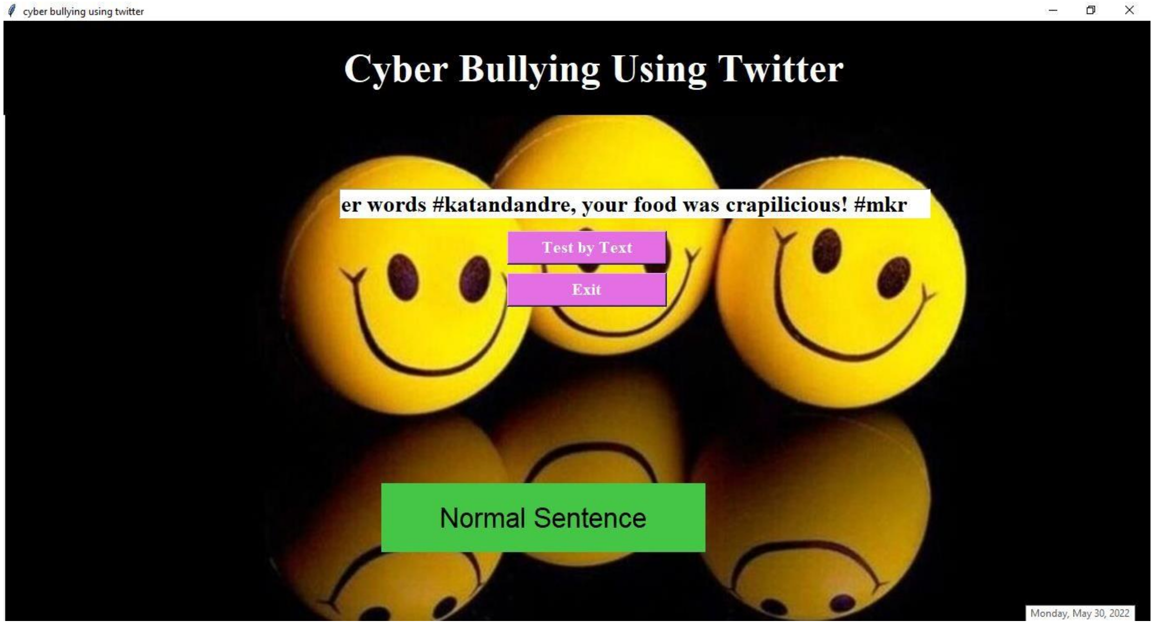
Login Window



Main Input Window



First type of Bully Detection



cyber bullying using twitter

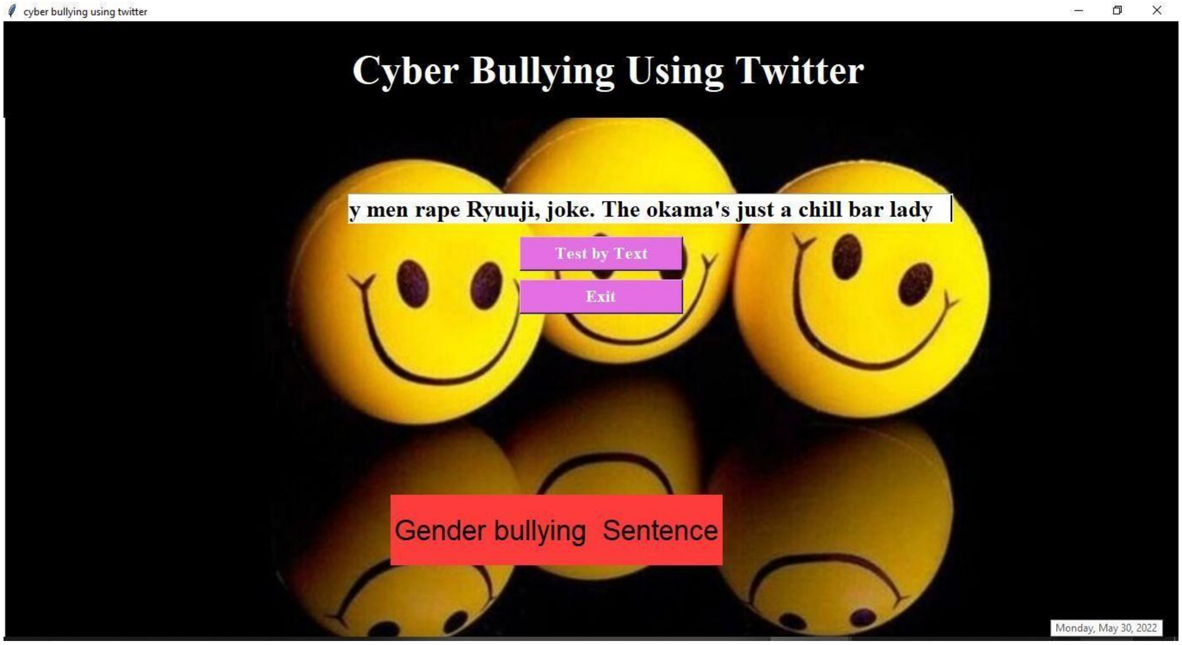
Cyber Bullying Using Twitter

er words #katandandre, your food was crapilicious! #mkr

Normal Sentence

Monday, May 30, 2022

Second Type of Bully Detection



cyber bullying using twitter

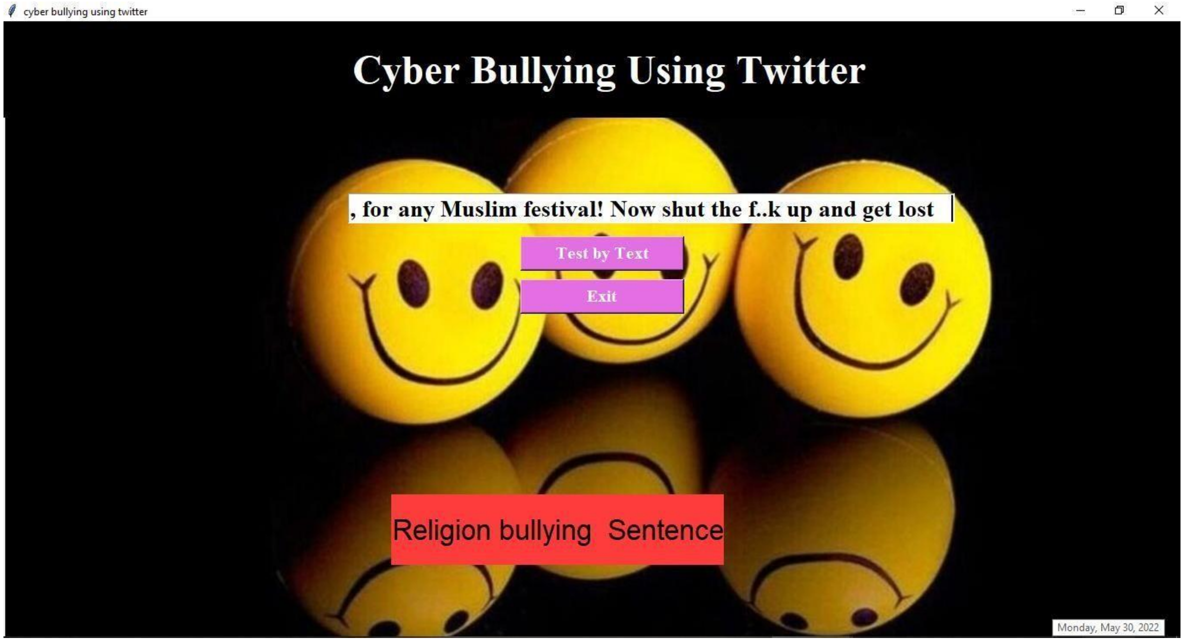
Cyber Bullying Using Twitter

y men rape Ryuuji, joke. The okama's just a chill bar lady

Gender bullying Sentence

Monday, May 30, 2022

Third type of Bully Detection

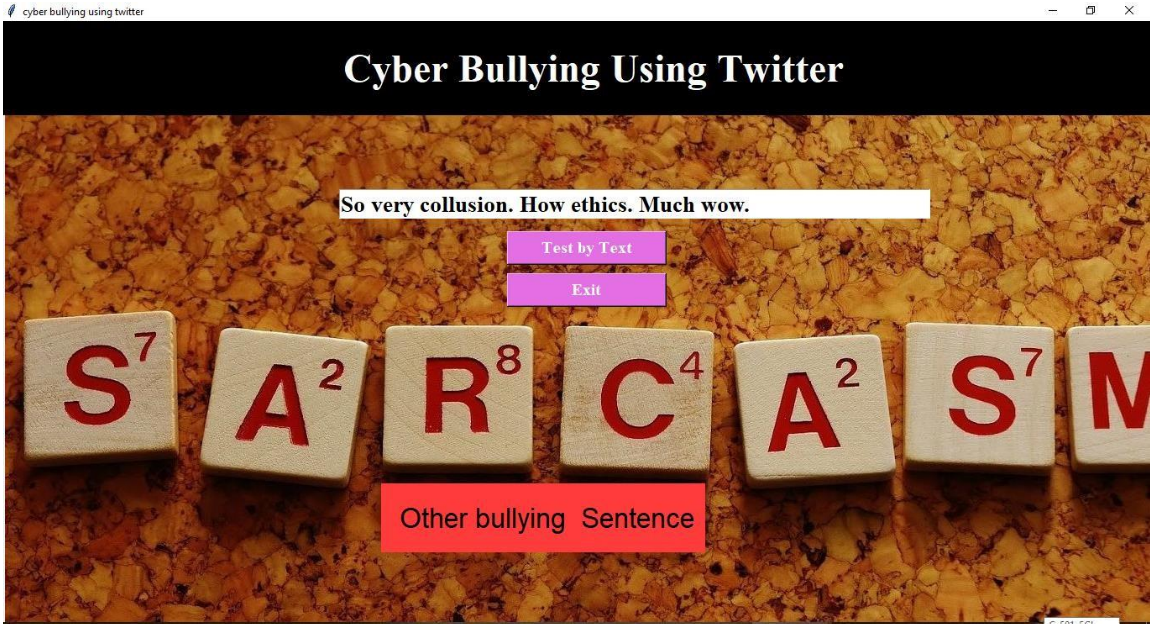


Cyber Bullying Using Twitter

, for any Muslim festival! Now shut the f..k up and get lost

Religion bullying Sentence

Forth type of Bully Detection

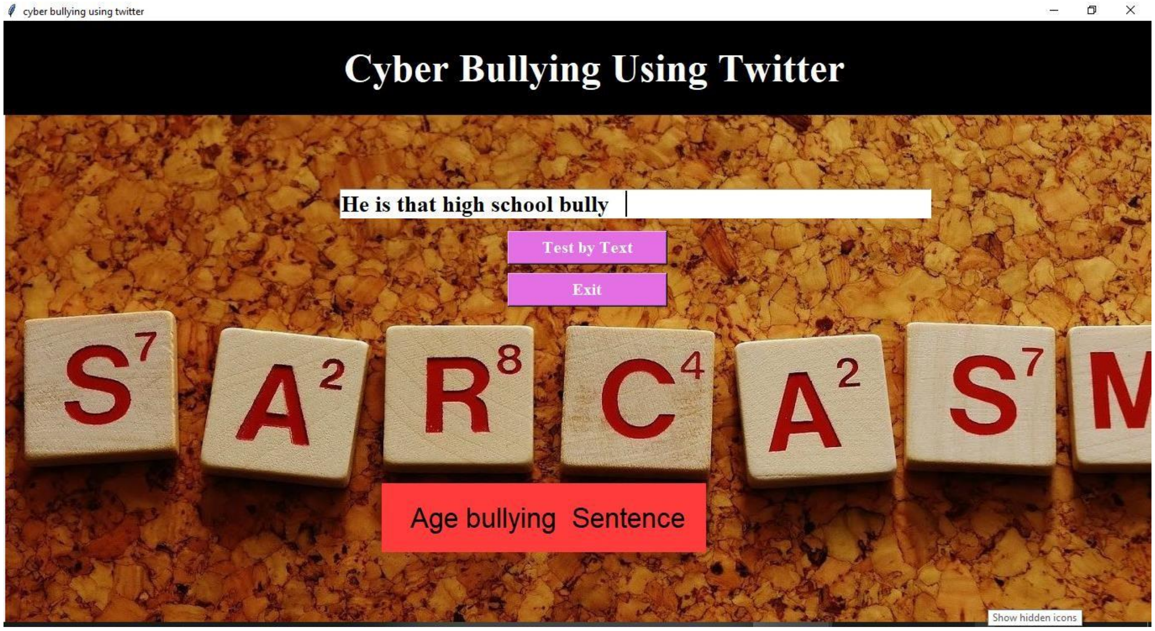


Cyber Bullying Using Twitter

So very collusion. How ethics. Much wow.

Other bullying Sentence

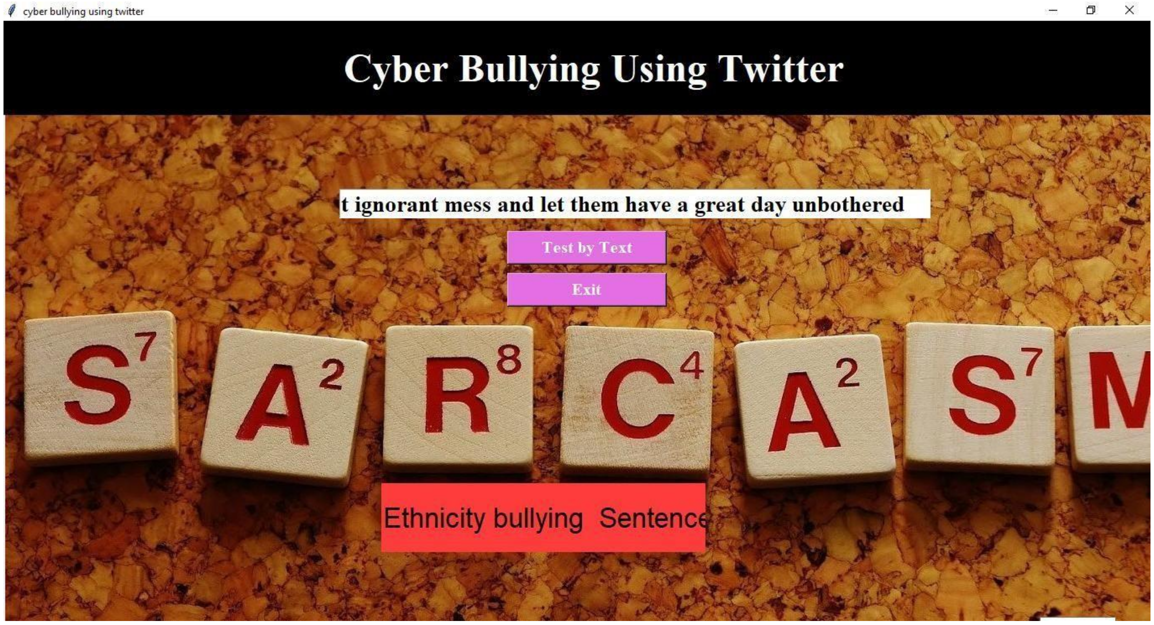
Fifth type of Bully Detection



He is that high school bully

Age bullying Sentence

Sixth type of Bully Detection



t ignorant mess and let them have a great day unbothered

Ethnicity bullying Sentence

CHAPTER 8: CONCLUSION

CONCLUSION

In this work, a system is proposed which detects on English as well as on Hindi tweets in Twitter. Cyber bullying is very dependent and highly contextual; therefore, sentiment and othercontextual clues to help detect the Cyber bullying . For future work, data from multiple social media platforms can be considered , apart from text. Image , video must be taken into account for experimentation.

CHAPTER 9: REFERENCES

[1] M. Kaplan and M. Heinlein, "Users of the world, unite! The challenges and opportunities of social media," Business horizons, vol. 53, no. 1, pp. 59–68, 2019.

[2] B. K. Biggs, J. M. Nelson, and M. L. Sampilo, "Peer relations in the anxiety– depression link: Test of a mediation model," Anxiety, Stress, Coping, vol. 23, no. 4,
pp. 431–447, 2022.

[3] K. Dinakar, B. Jones, C.Havasi, H. Lieberman, and R. Picard. "Common sense reasoning for detect ion, prevent ion, and mitigation of cyber bullying." ACM Transact ions on Interactive Intelligent Systems (TiiS) 2, no. 3, 2012, p. 18.

[4] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R.Lattanner, "Bullying in the digital age: A critical review and misanalysis of cyber bullying research among youth." 2018

[5] V. Nahar, S. Unankard, X. Li, and C. Pang. "Sentiment analysis for effective detect ion of cyber bullying." In Asia-Pacific Web Conference, Springer, Berlin, Heidelberg, 2019, pp. 767-774.

[6] V. Nahar, X. Li, C. Pang, and Y. Zhang. "Cyber bullying detect ion based on text stream classification." In The 11th Australasian Data Mining Conference (AusDM 2013), 2013.

[7] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A re- view and new perspectives," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 35,no. 8, pp. 1798–1828, 2013