VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Project Report on

Cyberbullying And Fake Account Detection In Social Media

In partial fulfillment of the Fourth Year (Semester–VII), Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2020-2021.

Project Mentor

Prof. Richard Joseph

Submitted by

Jayesh Samtani D17A-57
Sagar Sidhwa D17A-62
Somesh Tiwari D17A-71
Riya Wadhwani D17A-74

(2020-21)

VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

CERTIFICATE of Approval

This is to certify that ***Jayesh Samtani, Sagar Sidhwa, Somesh Tiwari, Riya Wadhwani*** of Fourth Year Computer Engineering studying under the University of Mumbai has satisfactorily presented the project on "***Cyberbullying and Fake Account Detection in Social Media***" as a part of the course work of PROJECT-I for Semester-VII under the guidance of ***Prof. Richard Joseph*** in the year 2020-2021.

_____

Date

_____ _____

Internal Examiner External Examiner

Project Mentor	Head of the Department Dr. Mrs. Nupur Giri	Principal Dr. J. M. Nair

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mrs. Priya R.L** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to:-

Course Outcome	Description of the Course Outcome
CO 1	Do literature survey/industrial visit and identify the problem of the selected project topic.
CO2	Apply basic engineering fundamental in the domain of practical applications FORproblem identification, formulation and solution
CO 3	Attempt & Design a problem solution in a right approach to complex problems
CO 4	Cultivate the habit of working in a team
CO 5	Correlate the theoretical and experimental/simulations results and draw the proper inferences
CO 6	Demonstrate the knowledge, skills and attitudes of a professional engineer & Prepare report as per the standard guidelines.

ABSTRACT

Enhancement in the Technology trend of using Social Networking is increasing day by day as of now there are more than 50 Crores active users are using different different social media platforms for the interaction which had affected their life just like a coin has two face in a similar way misuse of these platforms is going which cause the tremendous growth of the cyber crime and bullying for e.g Bullying Someone by sending the harmful messages , spreading of the harassment messages by using the fake accounts, using the abusive words on the social media etc.In this new era insulting a person by physically or emotionally is done by cyberbullying and by using fake accounts,So as a preventive measure to ensure the above things should not happen there is a need of detecting the cyber bullying and the fake accounts. In our study to stop this we'll use different Machine Learning algorithms for detecting the Cybercrime and fake accounts so as to report these issues to the system immediately and to stop the crimes to increase in future and develop a secure online environment

INDEX

CHAPTER – 1

INTRODUCTION

1.1 Introduction:

Social networking sites have connected us to different parts of the world However, people are finding illegal and unethical ways to use these communities. We see that people, especially teens and young adults, are finding new ways to bully one another over the Internet. Close to 25% of parents in a study conducted by Symantec reported that, to their knowledge, their child has been involved in a cyberbullying incident.

Other than cyberbullying, Spreading False information is increasingly at a rapid pace. The number of users in social media is increasing exponentially. Instagram has recently gained immense popularity among social media users.

The major source of the fake news are the fake accounts.Business organizations that invest a huge Sum of money on social media influencers must know whether the following gained by that account is organic or not. Hence there is a huge need for the detection of these fake accounts.

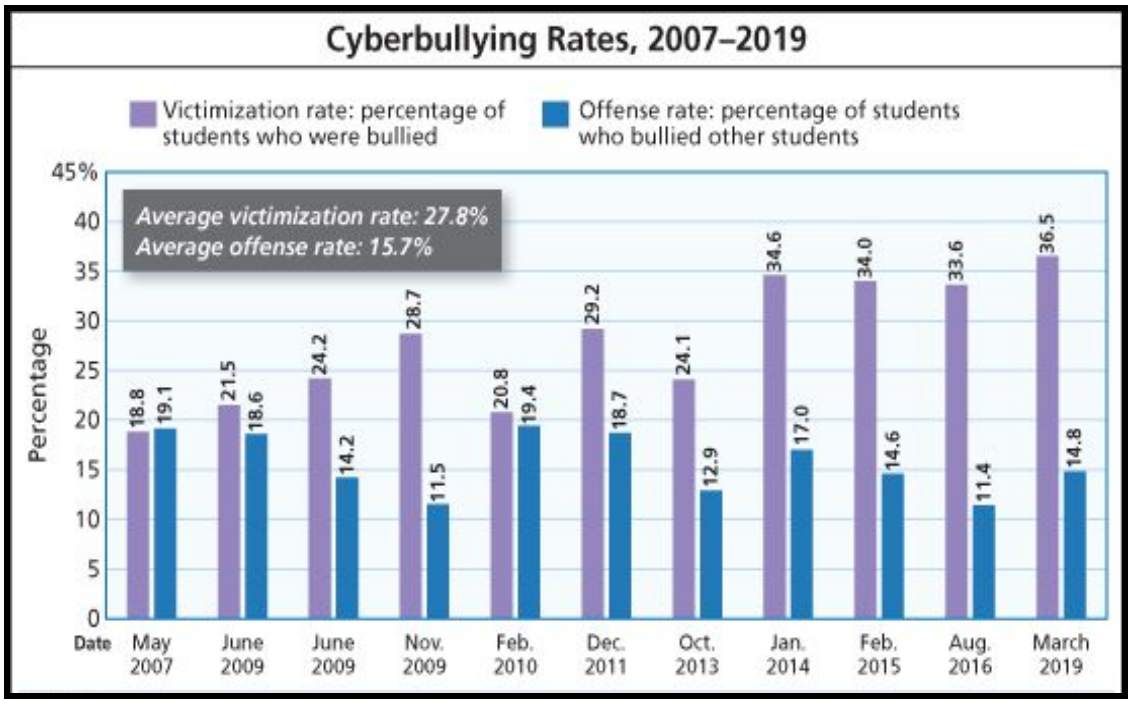


Fig 1.1.1-Graph showing the increase in the rate of CyberBullying in the recent years

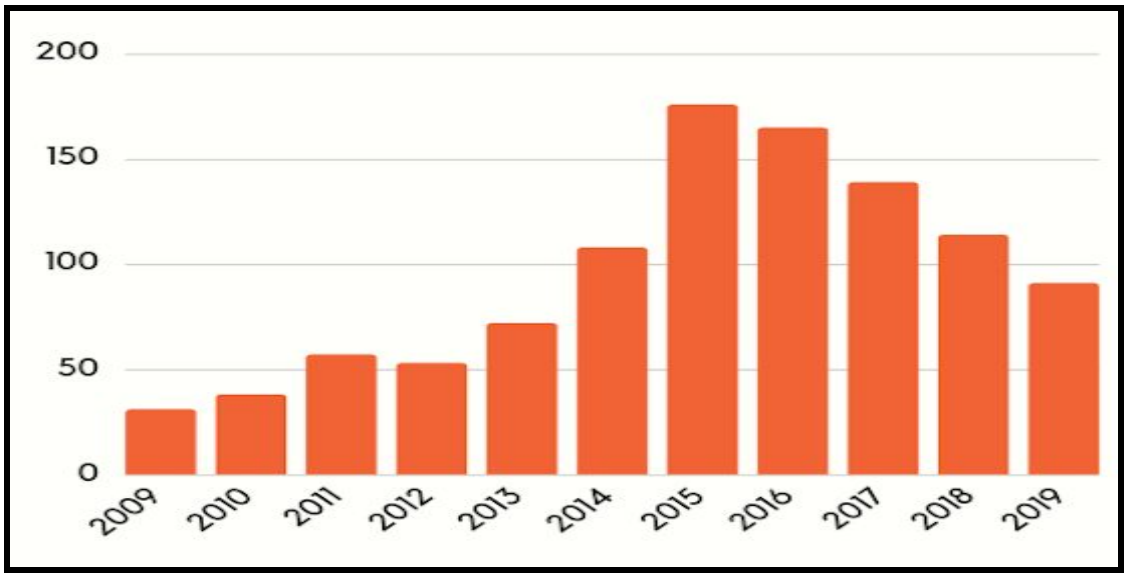


Fig 1.1.2-Graph showing the increase in the number of fake accounts

1.2 Motivation:

The use of social networks such as Facebook, Twitter, Google+, Instagram, and LinkedIn is on the rise. Individuals and organizations use social networks to express their views, advertise their products, and express future policies of their companies and organizations. By expanding the use of social networks, malicious users seek to violate the privacy of other users and abuse their names and credentials by creating fake accounts, which has become a concern for users. Hence, social networks providers are trying to detect malicious users and fake accounts in order to eliminate them from social networking environments. Creating fake accounts in social networks causes more damage than any other cybercrime .

Removing fake accounts has attracted the attention of many researches thus, extensive researches have been carried out on the identification of fake accounts in social networks. Different approaches are proposed to find fake accounts based on attribute similarity hence this generates an awareness in the field of social technology era to remove cyberbullying and eliminate the fake accounts from social media.

1.3. Drawback of the existing system :

1.3.1 Lack of Security :

Existing technology is vulnerable to the dark underbelly of the Internet where an anonymous person sitting on computer miles away can permanently scar their self-esteem by the power of a click, making cyberspace a dangerous and largely unmonitored playground

1.3.2 No Transparency :

There is no Transparency provided to the users about their data or details that users provide on their sites for their protection and theft against the current system .

1.3.3 Only one algorithm is Implemented :

Only one algorithm or technology is implemented based on the previous summary since only one technology will not give the solution to the whole system because there different parameters need to be considered on the current trends hence the system will be a proper working environment with full proof.

1.3.4 Costly to produce reports :

As it is costly to produce reports and provide it to the cyber department but this system will give the reports and the analysis at very less cost.

1.4 Problem definition:

One of the common issues everyone is facing and it is impacting the people, in which some are long period of sadness, anger, irritability, loss of interest in activities, being restless, anxious and worried, even in some cases they go into depression and take steps to scarify their life.It is unfortunate that there are no special Anti-Cyberbullying Laws in India yet.There are some common types of cyberbullying that is Flaming, Harassment, Denigration, Impersonation, Trickery. So to detect cyberbullying we have to make some software that will detect it and then report it to www.cybercrime.gov.in. Similarly, we will detect fake accounts.

1.5 Relevance of the Project:

Most people who are bullied online are also bullied in person. However, while offline bullying allows one the chance to avoid areas and situations that will put them in direct contact with a bully, cyberbullying offers no such reprieve.

Cyberbullying can follow victims wherever they go, whether they are in a crowd or alone. Cyberbullies can reach their victims, 24 hours a day, 7 days a week, 365 days a year. They often post hurtful content online, anonymously, so that they cannot be traced or stopped.

Given the nature of social media, such content is quick to go viral, and reaches a large audience in the blink of an eye, making it difficult, even impossible, for authorities to delete the harmful content before it wrecks damage.

The all-pervasive nature of cyberbullying, as well as the amount of time it takes to trace cyberbullies, makes the growth of cyberbullying an alarming trend across the globe.

Because cyberbullying is difficult to track, many victims feel helpless and unable to cope with it, especially if the bullying is personal and long-drawn. It is no surprise, therefore, that this form of bullying has been known to trigger depression and anxiety in its victims. In many instances, it has also resulted in victims developing suicidal tendencies.

Hence, social networks providers are trying to detect malicious users and fake accounts in order to eliminate them from social networking environments. Creating fake accounts in social networks causes more damage than any other cybercrime.

1.6 Methodology Used:

In this Project our aim is to detect Cyberbullying along with Fake account Detection with respect to the marginal number of attributes methodology that we had adopted to detect cyberbullying is by using computer technology by examining the incoming messages from the dataset of attached label of bullying to the suspected messages.

For each message, users have positively or negatively on a message on the basis of a 0 or 1 0- Non Offensive and 1-Offensive and then we have filtered the mechanism to classify the messages as abusive or non-abusive by using different algorithms and We will use classifiers, namely, SVM (Support Vector Machine), Naive Bayes,Random Forest, Decision Tree , Logistic Regression and Neural Network . A filter can classify all the messages and insults and similar abusive messages

Our next step is to find whether the account that causes bullying to someone is Fake or Not so the attributes of the account is given as input to the models like SVM (Support Vector Machine), Naive Bayes,Random Forest, Decision Tree , Logistic Regression and the model that will give us the best accuracy is used as the best fit to our system and the results that we will get from combined of cyberbullying and the fake account will be send and reported to the system that if the cyberbullying is done by the person or not is yes then we can report this to the cyber branch and this will help them to take an action by them.

CHAPTER – 2

LITERATURE SURVEY

2.1 Journal Papers referred

1. **Fake Twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach[1]:**
    a) **Abstract :** In this paper analysis 62 millions of publicly available Twitter user profiles was conducted and a strategy to retroactively identify automatically generated fake profiles was established. As recently as 2010, when the maximum number of Twitter user IDs was estimated to be less than 800 million , it was feasible to crawl over the entire Twitter user ID space. Spam accounts constitute as much as 93% of all new accounts, 68% of which are detected and automatically suspended or deleted by Twitter. To overcome these issues, they performed a Breadth First Search (BFS) over a given set of seed users. As the social graph is crawled, and eventually the user profiles for these IDs are acquired. This ensures that all user profile requests we make to Twitter include only valid Twitter user IDs.
    b) **Inference :** From this paper we got the information about the fake account dataset quantity, because due to the low number of false positives of fake account data accuracy of model decreased even if the twitter profile database was approx 60 million.

2. **Detection of Behavior Patterns through Social Networks likeTwitter, using Data Mining techniques as a method to detect Cyberbullying[2] :**
    a) Abstract : Social networks such as Twitter or Facebook have revolutionized the communication mechanism between human beings, but have also generated a negative impact due to inappropriate use, this fact is perpetuated by cybercriminals to hurt other people psychologically, these bad practices are called cyberbullying. This research focuses on the detection and analysis of cyberbullying on pages and with pejorative terms in Spanish, taking advantage of

the power of classification of feelings through specialized tools. For the detection of cyberbullying, first the efficiency of classification of each tool is measured, through a set of pejorative terms commonly used to hurt other people.

b) **Inference :** In the analysis stage we use data mining techniques to generate a dictionary of pejorative terms that are related to cyberbullying and thus be able to generate behavior patterns of these terms. And in this way provide better tools so that psychology specialists can optimize their work. The results show which platform is more flexible, and also shows which is best suited to the search of incidences of cyberbullying on Twitter.

3. **Classification of Cyberbullying in Facebook Using Selenium and SVM[3]:**

   a) **Abstract :** Cyberbullying is one of the emerging problems over the past few years especially to teenagers. Approximately 24% of teens go online constantly, facilitated by the widespread availability of smartphones. Almost 21% of teens said the main reason they checked social media always was to make sure nobody was saying mean or bad things to them. Cyberbullying related Facebook posts were harvested by a customized web scraper tool.

   b) **Inference :** In this paper facebook data were used for classification using Support Vector Machines (SVM) models. A total of 2263 data was used for training data, Facebook posts. Based on these posts, the study achieved the precision of 88% and the recall is 87%. So it is quite a good algorithm for cyberbully detection.

4. **Identifying Fake News from the Variables that Governs the Spread of Fake News[4]:**

   a) **Abstract :** Several researchers have attempted to investigate the processes that govern and support the spread of fake news. This paper collates and identifies these variables. This paper then categorised these variables based on three key players that are involved in the process: Users, Content, and Social Networks. The authors conducted an extensive review of the literature and a reflection on the key variables that are involved in the process. The paper has identified a total of twenty-seven variables. Then the paper presents a series of tasks to mitigate or

eliminate these variables in a holistic process that could be automated to reduce or eliminate fake news propagation. Finally, the paper suggests further research into testing the method in lab conditions.

b) **Inference:** In paper has reviewed a variety of variables identified by researchers in the field of understanding the factors that influence fake news. The variables show a significant overlap in views but also concentration on different players. As such, the paper collated all these variables and redistributed them based on the key players. This has helped build a holistic and bigger picture of the environment in which Fake News thrives. The variables identified pinpointed some areas where one can see how different social media platforms have attempted to combat fake news and failed.

5. Detecting Rumors from Microblogs with Recurrent Neural Networks[5]:

a) **Abstract** : Microblogging platforms are an ideal place for spreading rumors and automatically debunking rumors is a crucial problem. To detect rumors, existing approaches have relied on hand-crafted features for employing machine learning algorithms that require daunting manual effort. Upon facing a dubious claim, people dispute its truthfulness by posting various cues over time, which generates long-distance dependencies of evidence. This paper presents a novel method that learns continuous representations of microblog events for identifying rumors. The proposed model is based on recurrent neural networks (RNN) for learning the hidden representations that capture the variation of contextual information of relevant posts over time. Experimental results on datasets from two real-world microblog platforms demonstrate that (1) the RNN method outperforms state-of-the-art rumor detection models that use hand-crafted features; (2) performance of the RNN-based algorithm is further improved via sophisticated recurrent units and extra hidden layers; (3) RNN-based method detects rumors more quickly and accurately than existing techniques, including the leading online rumor debunking services.

b) **Inference :** In this research, we propose a deep learning framework for rumor debunking. Our method learns RNN models by utilizing the variation of

aggregated information across different time intervals related to each event. We empirically evaluate our RNN-based method with three widely used recurrent units, tanh, LSTM and GRU, which perform significantly better than the state-of-the-art.

(a) Twitter dataset

Method	Class	Accuracy	Precision	Recall	F_1
DT-Rank	R	0.644	0.638	0.675	0.656
	N		0.652	0.613	0.632
SVM-RBF	R	0.722	0.856	0.526	0.651
	N		0.663	**0.914**	0.769
DTC	R	0.731	0.724	0.757	0.740
	N		0.739	0.704	0.721
RFC	R	0.772	0.717	0.908	0.801
	N		0.870	0.634	0.734
SVM-TS	R	0.808	0.735	**0.963**	0.834
	N		**0.947**	0.652	0.772
tanh-RNN	R	0.827	0.847	0.833	0.840
	N		0.804	0.820	0.812
LSTM-1	R	0.855	0.855	0.883	0.869
	N		0.854	0.820	0.837
GRU-1	R	0.864	**0.857**	0.900	0.878
	N		0.872	0.820	0.845
GRU-2	R	**0.881**	0.851	0.950	**0.898**
	N		0.930	0.800	**0.860**

Fig 2.1-Rumor detection Results(R: Rumor, N: Non-Rumor)

6. **Automatic detection of cyberbullying in social media text[6]:**

a) **Abstract** : The focus of this paper is on automatic cyberbullying detection in social media text by modelling posts written by bullies, victims, and bystanders of online bullying. We describe the collection and fine grained annotation of a cyberbullying corpus for English and Dutch and perform a series of binary classification experiments to determine the feasibility of automatic cyberbullying detection. We make use of linear support vector machines exploiting a rich feature set and investigate which information sources contribute the most for the task. Experiments on a hold-out test set reveal promising results for the detection of cyberbullying-related posts. After optimisation of the hyperparameters, the

classifier yields an F1 score of 64% and 61% for English and Dutch respectively, and considerably outperforms baseline systems .

b) **Inference :**A set of binary classification experiments were conducted to explore the feasibility of automatic cyberbullying detection on social media. In addition, we sought to determine which information sources contribute most to the task. Two classifiers were trained on an English and Dutch ASKfm corpus and evaluated on a hold-out test of the same genre. Overview of the most related cyberbullying detection approaches.After feature and hyperparameter optimisation of our models, a maximum F1 score of 64.32% and 58.72% was obtained for English and Dutch

CHAPTER – 3

REQUIREMENTS

3.1 Proposed model

The proposed approach contains three main steps namely Preprocessing, features extraction and classification step.

In the preprocessing step we clean the data by removing the noise and unnecessary text.
The preprocessing step is done in the following: -

Tokenization: In this part we take the text as sentences or whole paragraphs and then output the entered text as separated words in a list. -

Lowering text: This takes the list of words that got out of the tokenization and then lower all the letters Like: 'THIS IS AWESOME' is going to be 'this is awesome'.

Stop words and encoding cleaning: This is an essential part of the preprocessing where we clean the text from those stop words and encoding characters like \n or \t which do not provide meaningful information to the classifiers.

The second step of the proposed Model is the features extraction step. In this step the textual data is transformed into a suitable format applicable to feed into machine learning algorithms.

The last step in the proposed approach is the classification step where the extracted features are fed into a classification algorithm to train, and test the classifier and hence use it in the prediction phase. We will use classifiers, namely, SVM (Support Vector Machine), Naive Bayes,Random Forest,Decision Tree , Logistic Regression and Neural Network.

The neural network will contain three layers: Input, hidden, output layer. The output layer is a Boolean output.

Accuracy of different algorithms will be Compared to get the best possible result.
For the fake profile detection this paper proposes the detection process starts with the selection of the profile that needs to be tested. After selection of the profile the suitable attributes ie., features are selected on which the classification algorithm is being implemented ,the attributes extracted are passed to the trained classifier .

Different Classifier algorithms such as Gradient Booster, random forest Decision trees ,Support Vector Machine and Neural Networks such as RNN and CNN can be used.
The model generated by the learning algorithm should both fit the input data correctly and correctly predict the class labels of the learning algorithm is to build the model with good generality capability.

Data set of both fake and genuine profiles with various attributes like number of friends ,followers, status count. Dataset is divided into training and testing data. Classification algorithms are trained using training dataset and testing data set is used to determine the efficiency of the algorithm .From the dataset used 80% of both (real and fake) are used to prepare a training data set and 20% of both profiles are used to prepare a testing dataset.

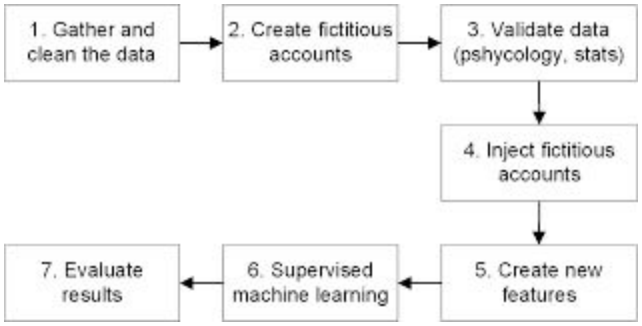


Fig 3.1 - Flow of the Proposed Model

3.2 Functional Requirements

- **Social Media-**

  The government uses social media technologies to extract opinion from people regarding a specific issue , these social media technologies have now been seen as the most extensively used platforms to conduct electronic participation activities. According to smallbiztrends.com, Facebook, Twitter and Instagram are one of the famous social networking sites that a lot of people are using including establishments and organizations .With high online activity using these sites, teens can be a victim or a perpetrator of cyberbullying
- User authentication module should detect and reject malicious authentication attempts.
- The system should decide whether a suspicious content should be checked for malicious behaviour or for fake identity or activity.
- Each time a malicious behaviour is detected, an account should be added in the appropriate malicious behaviour reputation list.
- **Admin-** It can view and block all the malicious accounts

3.3 Non-Functional Requirements

- **Availability -** The system is available all the time, no time constraint

3.4 Hardware & Software Requirements

- Intel Pentium Processor
- RAM > 4GB
- Django
- Machine Learning Algorithms
- Anaconda

3.5 Constraints of working

- Continuous Internet Connection
- Risk Management
- System Failure

CHAPTER – 4

PROPOSED DESIGN

4.1 Block Diagram of the proposed system (with explanation)

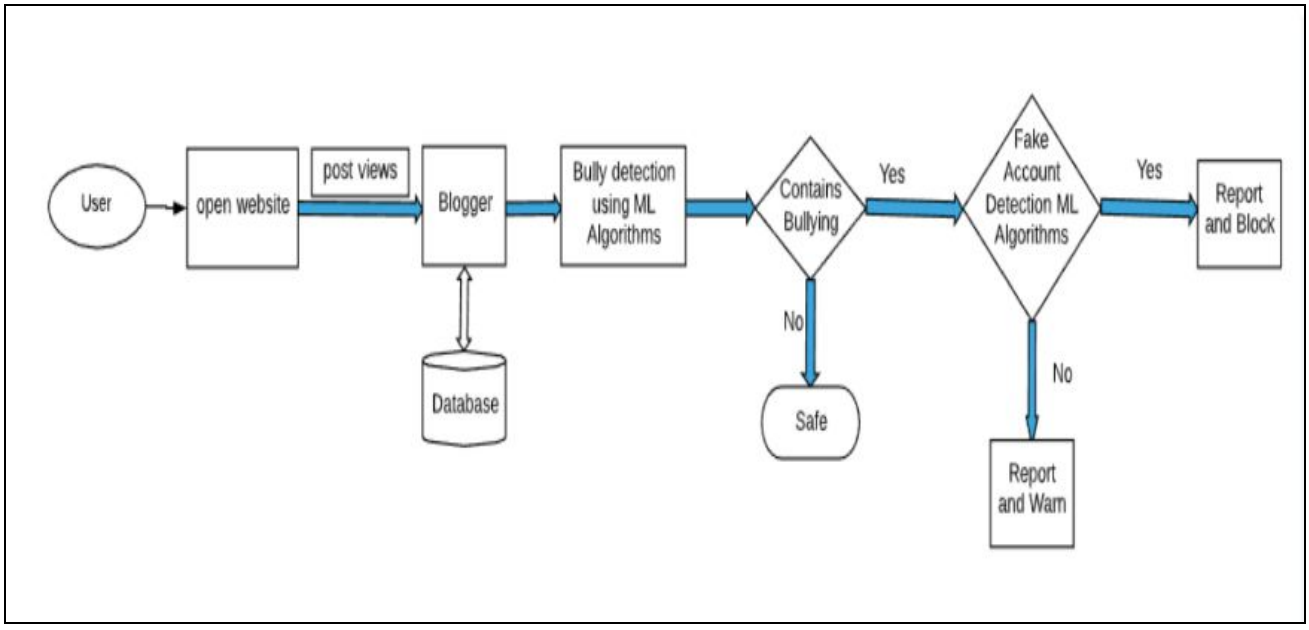


Fig 4.1 Block Diagram

The user will first login in to our portal with his/her credentials or can sign up on our portal for new registration. The portal can work as a simple blog where users can post their views and can read other people's views.on posting the views it will undergo certain machine learning algorithm processing where it will determine whether the following post is cyberbullying or not . If found that the post is vulnerable the system will further check that if the account is fake or not.if found fake it will block the account and if the account is not fake the system will report the tweet.

4.2 Modular diagram (with explanation)

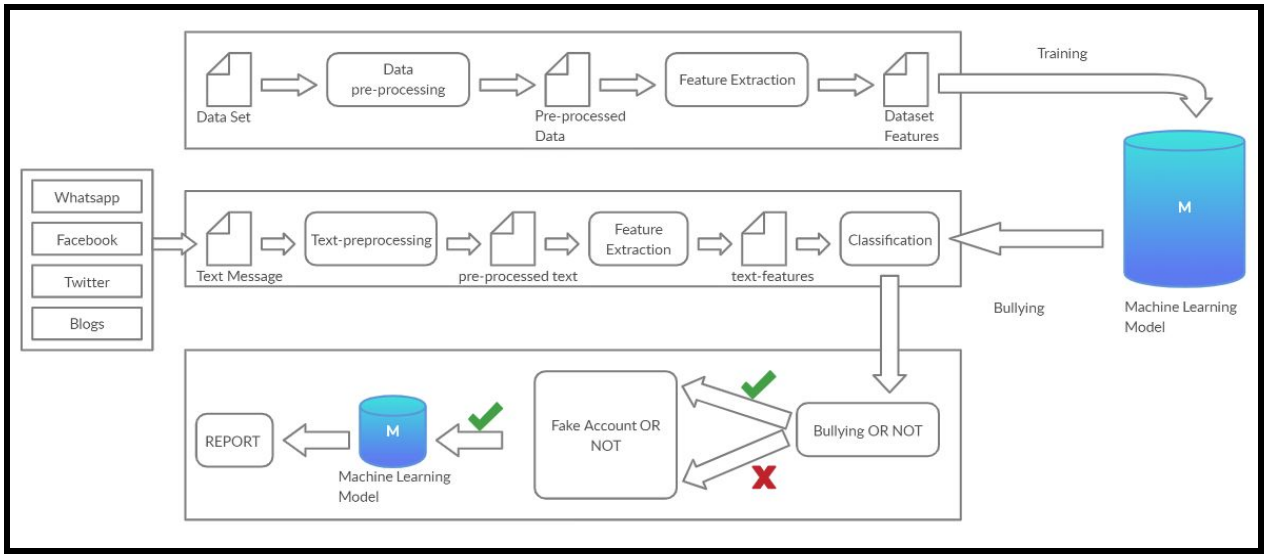


Fig 4.2 Modular diagram

In this Project our aim is to detect Cyberbullying along with Fake account Detection with respect to the marginal number of attributes the proposed methodology consists of different steps.

First step is the preprocessing of the datasets and find the minimum set of attributes from two datasets i.e Cyberbullying and the second one is the fake account so in this step we will separate number of the attributes that will actually used to identify these bullying, contains abusive words etc and save them the dataset for further processing in the second dataset we will preprocess and extract the attributes e.g name, contact, email, etc and will be stored for further processing and to identify that weather the account is fake or not.

The Second step is to create different Machine Learning Models like Support Vector Machine Classifier, Random Forest algorithm,Naïve Bayes, Logistic Regression,K-mean clustering,ADT and BFT tree and Neural Networks etc and applying the following algorithms on the datasets of Cyberbullying and the fake account by splitting the datasets in to training and test approximately in the ratio of 80:20 and to find the algorithm which best suits or fit for our system to achieve the highest accuracy.

Third step is to test the messages that are extracted from the chats or the tweets or the blog which is posted which can cause bullying or use of abusive words and then sending the words to different Machine Learning models that are created for testing or checking to give the input to these models and running these models to test that which one of them gives the best solution and fits best in to our system.

In the Fourth step is to find whether the account that causes bullying to someone is Fake or Not so the attributes of the account is given as input to the models and the model that will give us the best accuracy is used as the best fit to our system and the results that we will get from the third step and fourth step will be combined and will be send and reported to the system that if the cyberbullying is done by the person or not is yes then we can report this to the cyber branch and this will help them to take an action by them.

4.3 Detailed Design (DFD-level 0,1,2, Flowchart, ER Diagram, etc...)

a. ER DIAGRAM

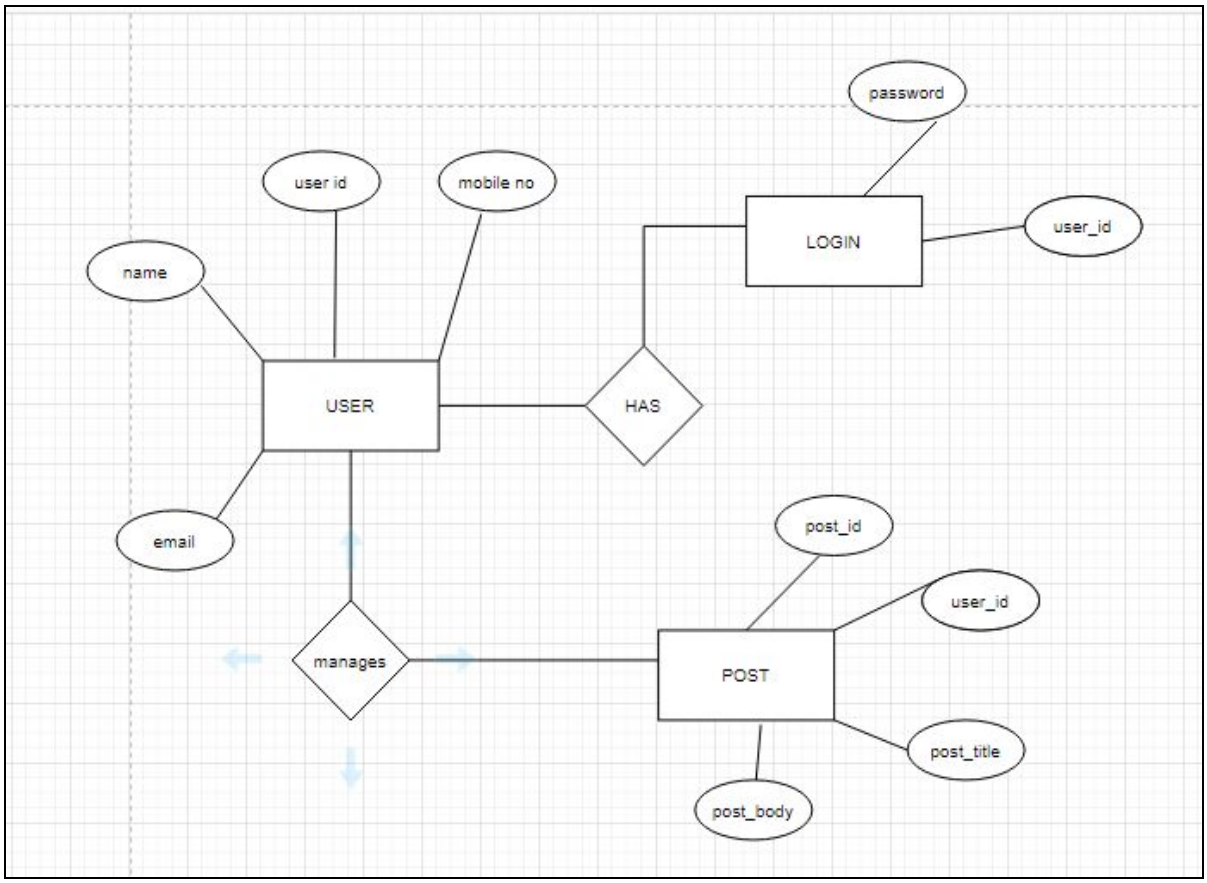


Fig 4.3.1 ER Diagram

As shown in above E-R Diagram which contains the data storage process where first the user needs to register on the website by giving the details like name ,number etc also if the user further posts something on the site then its post will be saved on the database based on the user id which creates the post and then the user needs to login to the page after logging out from the website.

b. USE CASE

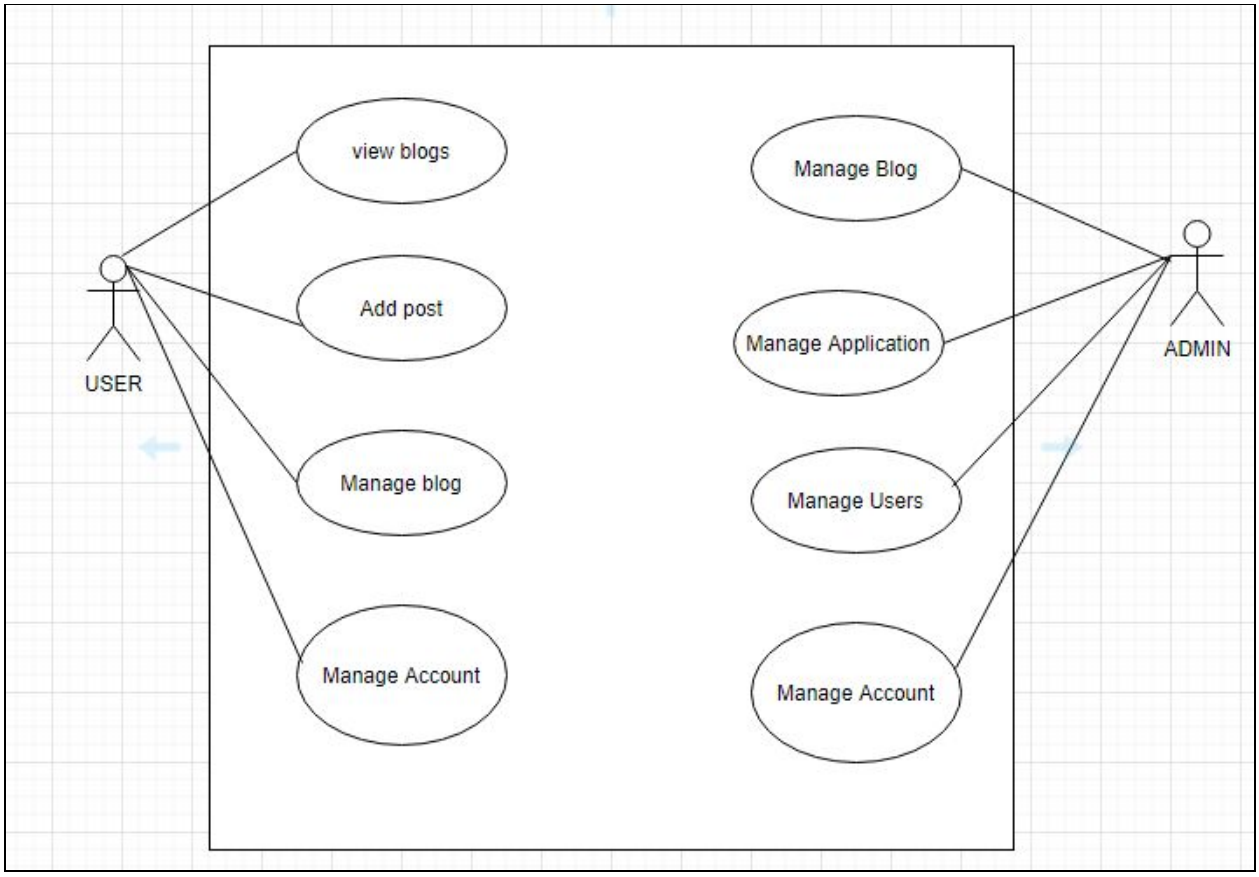


Fig 4.3.2 Use Case Diagram

This is the sample Use Case Diagram representing the Cyberbullying and fake Account detection model ,If we look into diagram we can see the Use Cases as Follows:

Manage Account, Manage Blog, Add Post, View Blogs,Manage application,Manage Users,Account and Blog. We have two Actors as we can see in the diagram User and Admin.

c. DFD Diagram:

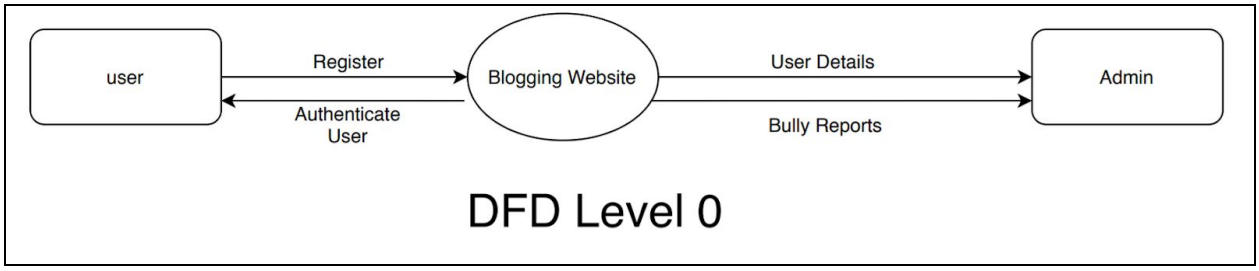


Fig 4.3.3 DFD Level-0 Diagram

As shown above the DFD Level-0 diagram which shows that the user first needs to register to the portal then the user details will be forwarded to the admin where if the user posts something abusive at the website then the bully reports will be sended to the admin and the admin can authenticate the user**.**

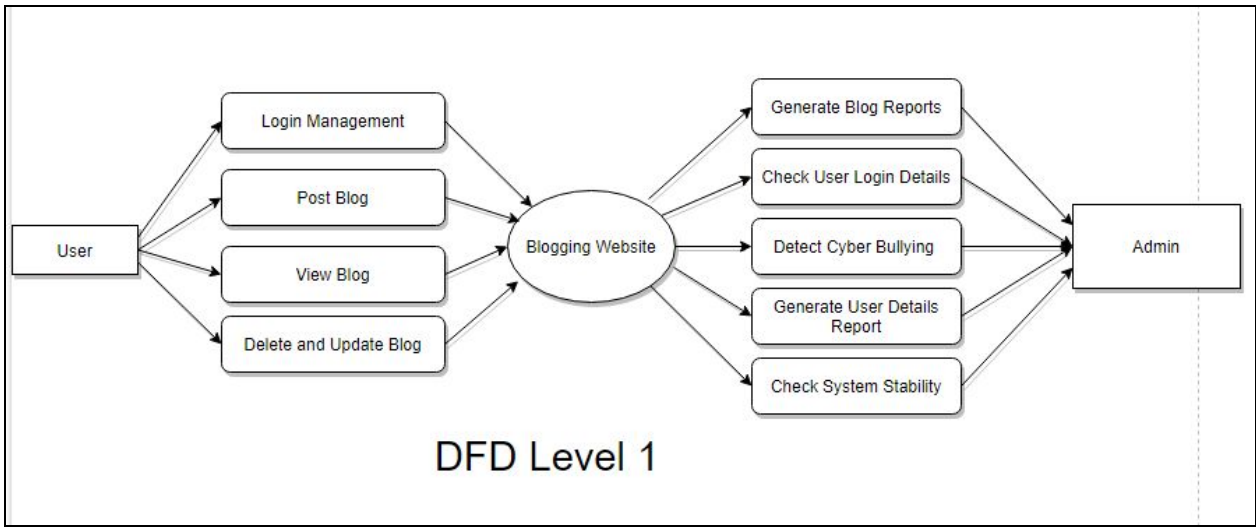


Fig 4.3.4 DFD Level-1 Diagram

In the DFD Level-1 Figure the user can perform the functionality like it can do login,post blog,view its blog and delete or update its blog on the blogging website and from the admin side It can generate blog reports ,check user login details ,detect cyberbullying, generate user details reports,check system stability.

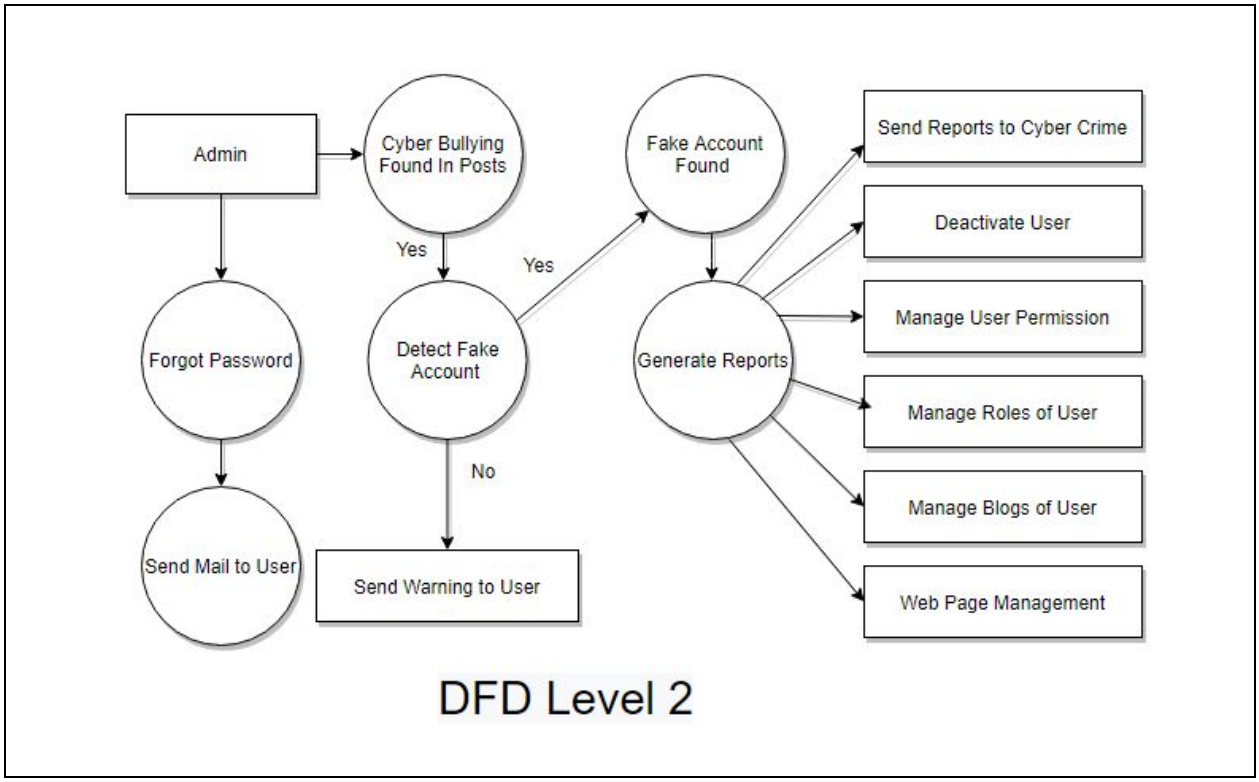


Fig 4.3.5 DFD Level-2 Diagram

In DFD Level 2 Diagram, as shown above the flow is like User will Register and then he/she will log into the blogging Website where he/she will post some blogs and can view other users blogs also.If he/she has posted some offensive or bully content then that will be detected and then Account verification will be also done if the account is fake or not valid then a report will be generated and that will be sent to the Cyber Crime and Users Account will be blocked.

4.4. Proposed algorithms

The main step of the proposed Model is the features extraction step. In this step the textual data is transformed into a suitable format applicable to feed into machine learning algorithms. The last step in the proposed approach is the classification step where the extracted features are fed into a classification algorithm to train, and test the classifier and hence use it in the prediction phase. We will use classifiers, namely, SVM (Support Vector Machine), Naive Bayes,Random Forest, Decision Tree , Logistic Regression

4.4.1 Logistic Regression

Logistic regression is one of the well-known techniques introduced from the field of statistics by machine learning.Logistic regression is an algorithm that constructs a separate hyper-plane between two datasets utilizing the logistic function. The logistic regression algorithm takes features (inputs) and produces a forecast according to the probability of a class suitable for the input. For instance, if the likelihood is ≥0.5, the instance classification will be a positive class; otherwise, the prediction will be for the other class (negative class) , as given in Equation. logistic regression was used in the implementation of predictive cyberbullying models. hθ (x) = 1 1 + e −θTx , (1) if hθ (x) ≥ 0.5, y = 1 (Positive class) and if hθ (x) ≤ 0.5, y = 0 (Negative class)

4.4.2 Random Forest algorithm

The Random Forest Algorithm is composed of different decision trees, each with the same nodes, but using different data that leads to different leaves. It merges the decisions of multiple decision trees in order to find an answer, which represents the average of all these decision trees
The random forest algorithm is a supervised learning model; it uses labeled data to "learn" how to classify unlabeled data. This is the opposite of the K-means Cluster algorithm,which is an unsupervised learning model. The Random Forest Algorithm is used to solve both regression and classification problems, making it a diverse model that is widely used by engineers.
When performing Random Forests based on classification data, you should know that you are often using the Gini index, or the formula used to decide how nodes on a decision tree branch

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2 \quad .$$

This formula uses the class and probability to determine the Gini of each branch on a node, determining which of the branches is more likely to occur. Here, *pi* represents the relative frequency of the class you are observing in the dataset and *c* represents the number of classes.

4.4.3 Naives bayes classifier

Naives bayes classifiers are a group of machine learning algorithms that all use the Bayes' Theorem to classify data points. The Bayes' Theorem is named after Reverend Thomas Bayes, a man who studied probability and binomial distributions in the 18th century. The mathematics behind Naive Bayes The algorithm completely depends upon Bayes theorem since the classifiers simply apply the formula to sets of data. This theorem consists of a formula assessing probabilities of different events occurring. The formula below is the simplest version of it, with only two events — Event A and B.

$$P(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

4.4.4 Adaptive boosting

(AdaBoost) is an ensemble learning method, and it is a prevalent boosting technique that was initially developed to make binary classifiers more efficacious.It uses an iterative approach to learn from weak classifiers' errors, and transform them into strong ones. Therefore, each training observation is initially assigned equal weights. It uses several weak models and attributes higher weights to experimental misclassification observations. As the results of the definitive boundaries obtained during several iterations are combined using several low models, the accuracy of the erroneously classified observations is improved. Thus, the accuracy of the overall iteration is enhanced An example of AdaBoost classifier implantation is shown in Figure 1, where it showed a similar dataset that has two features and two classes in which week learner 2 improve by mistake made by weak learner #1 and the accuracy of the misclassified observations is further improved when two weak classifiers are combined.

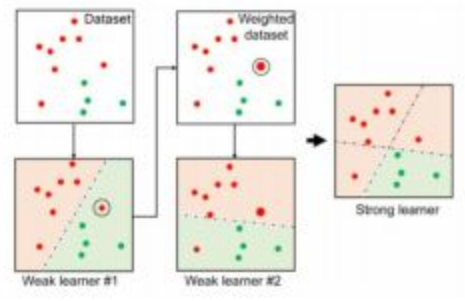


Fig - 4.4.4 Adaptive Boosting Sample Training

4.4.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning classifier widely utilized in text classification . SVM turns the original feature space into a user-defined kernel-based higher-dimensional space and then seeks support vectors for optimizing the distance (margin) between two categories. SVM originally approximates a hyperplane separating the two categories. SVM accordingly selects samples from both categories, which are nearest to the hyperplane, referred to as support vectors.

SVM seeks to efficiently distinguish the two categories (e.g., positive and negative). If the dataset is separable by nonlinear boundaries, specific kernels are implemented in the SVM to turn the function space appropriately. Soft margin is utilized to prevent overfitting by giving less weighting to classification errors along the decision boundaries for a dataset that is not easily separable [101]. In this research, we utilize SVM with a linear kernel for the basis function. Figure 2 shows the SVM classifier implementation for a dataset with two features and two categories where all samples for the training are depicted as circles or stars. Support vectors (referred to as stars) are for each of the two categories from the training samples, meaning that they are nearest to the hyperplane among the other training samples.

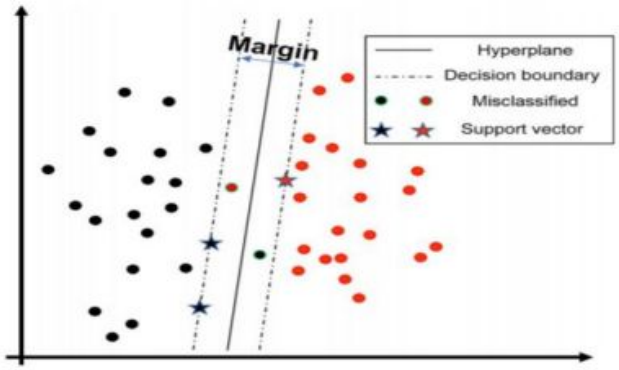


Fig - 4.4.5 SVM Model

4.5 Project Scheduling & Tracking using Time line / Gantt Chart

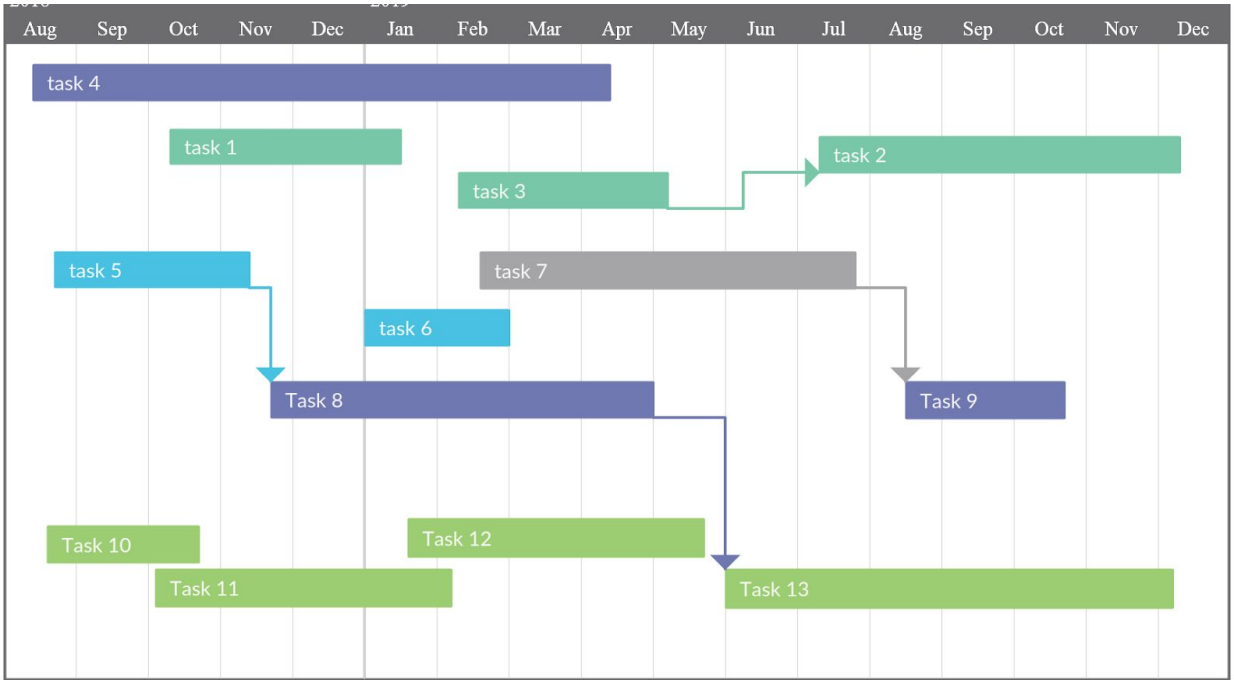


Fig 4.5.1 Gantt Chart

CHAPTER – 5

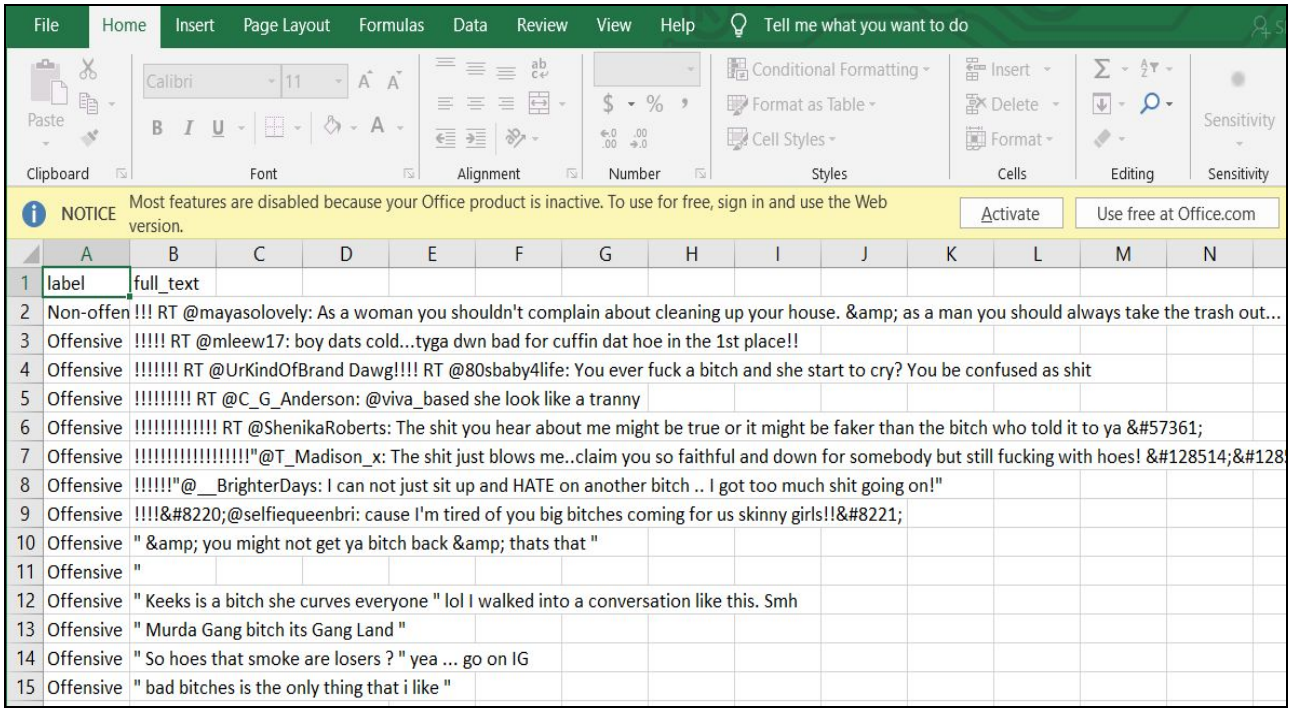
RESULTS & DISCUSSION

5.1 Determination of efficiency

5.1.1 Cyberbullying Dataset

The Cyberbullying Dataset contains two parameters that is 1) Label 2)Full_text

1) Label - This column describes that weather the text associated with this is Offensive or Not we then converted it to Offensive to 1 and Non-Offensive to 0
2) Full-Text - The following parameter contains the bad words or the bully words in each tuple and this column is processed using the Count Vectorizer Method which groups the similar words in to one so that if the similar kind of bully or bad words if used then can be recognized in the post



	A	B
1	label	full_text
2	Non-offen	!!! RT @mayasolovely: As a woman you shouldn't complain about cleaning up your house. & as a man you should always take the trash out...
3	Offensive	!!!!! RT @mleew17: boy dats cold...tyga dwn bad for cuffin dat hoe in the 1st place!!
4	Offensive	!!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby4life: You ever fuck a bitch and she start to cry? You be confused as shit
5	Offensive	!!!!!!!!! RT @C_G_Anderson: @viva_based she look like a tranny
6	Offensive	!!!!!!!!!!!!! RT @ShenikaRoberts: The shit you hear about me might be true or it might be faker than the bitch who told it to ya 
7	Offensive	!!!!!!!!!!!!!!!!!!"@T_Madison_x: The shit just blows me..claim you so faithful and down for somebody but still fucking with hoes! 😂&#128
8	Offensive	!!!!!!"@__BrighterDays: I can not just sit up and HATE on another bitch .. I got too much shit going on!"
9	Offensive	!!!!“@selfiequeenbri: cause I'm tired of you big bitches coming for us skinny girls!!”
10	Offensive	" & you might not get ya bitch back & thats that "
11	Offensive	"
12	Offensive	" Keeks is a bitch she curves everyone " lol I walked into a conversation like this. Smh
13	Offensive	" Murda Gang bitch its Gang Land "
14	Offensive	" So hoes that smoke are losers ? " yea ... go on IG
15	Offensive	" bad bitches is the only thing that i like "

Fig 5.1.1 Cyberbully Dataset

5.1.2 Fake Account Dataset

The Fake Account Detection dataset has the following parameters i.e

1) Name - It has names of the account holders.

2) Status Count - Total Status updated by users.

3) Followers Count - It shows the number of people the account holder follows.

4) Friends Count - It shows the total number of friends count of users.

5) Url - The user profile's URL

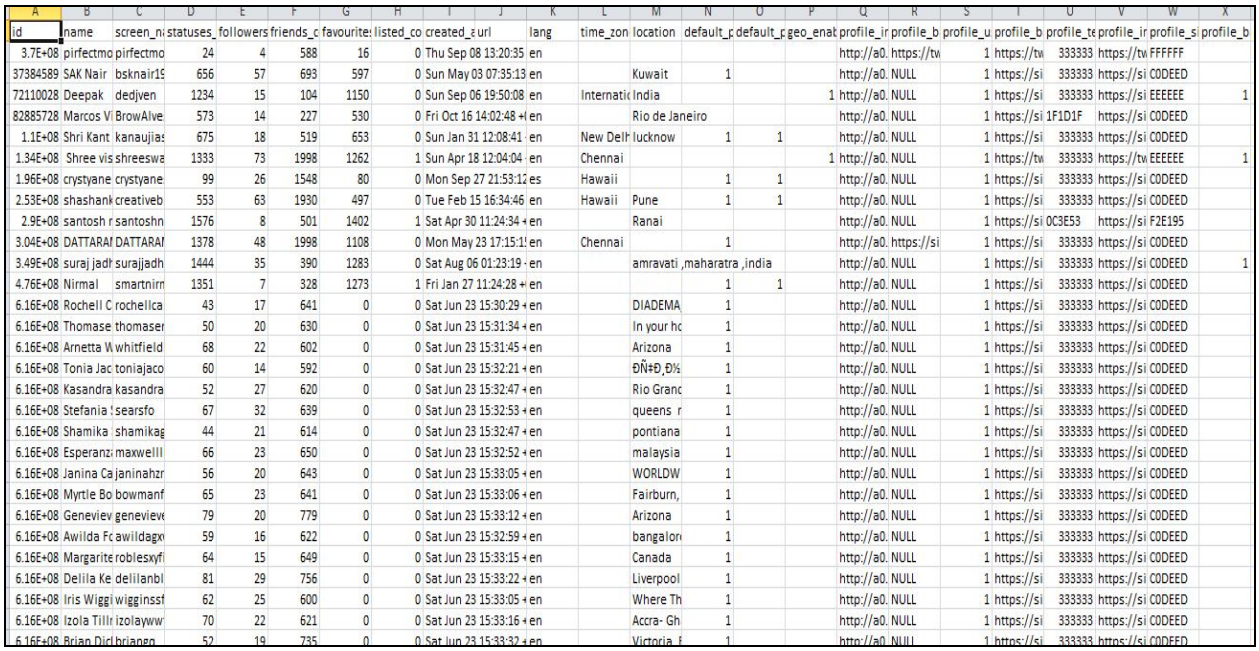
6) Time Zone - It is the time zone.

7) Listed Count - It counts how many websites the user visits.

8)Screen Name - It is the display name of the user on the website.

9) Profile Bio - It describes the the profile of the user on different social media.

10) Location - it identifies the current location from where the user is operating.



id	name	screen_n	statuses_	followers	friends_c	favourite	listed_co	created_a	url	lang	time_zon	location	default_p	default_p	geo_enat	profile_ir	profile_b	profile_u	profile_b	profile_te	profile_ir	profile_s	profile_b
3.7E+08	pirfectmo	pirfectmo	24	4	588	16	0	Thu Sep 08 13:20:35		en						http://a0.	https://tw	1	https://tw	333333	https://tw	FFFFFF	
37384589	SAK Nair	bsknair19	656	57	693	597	0	Sun May 03 07:35:13		en		Kuwait	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
72110028	Deepak	dedjven	1234	15	104	1150	0	Sun Sep 06 19:50:08		en	Internatic	India			1	http://a0.	NULL	1	https://si	333333	https://si	EEEEEE	1
82885728	Marcos Vi	BrowAlve	573	14	227	530	0	Fri Oct 16 14:02:48 +		en		Rio de Janeiro				http://a0.	NULL	1	https://si	1F1D1F	https://si	C0DEED	
1.1E+08	Shri Kant	kanaujias	675	18	519	653	0	Sun Jan 31 12:08:41		en	New Delh	lucknow	1	1		http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
1.34E+08	Shree vis	shreeswa	1333	73	1998	1262	1	Sun Apr 18 12:04:04		en	Chennai				1	http://a0.	NULL	1	https://tw	333333	https://tw	EEEEEE	1
1.96E+08	crystyane	crystyane	99	26	1548	80	0	Mon Sep 27 21:53:12		es	Hawaii		1	1		http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
2.53E+08	shashank	creativeb	553	63	1930	497	0	Tue Feb 15 16:34:46		en	Hawaii	Pune	1	1		http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
2.9E+08	santosh r	santoshn	1576	8	501	1402	1	Sat Apr 30 11:24:34 +		en		Ranai				http://a0.	NULL	1	https://si	0C3E53	https://si	F2E195	
3.04E+08	DATTARAI	DATTARAI	1378	48	1998	1108	0	Mon May 23 17:15:11		en	Chennai		1			http://a0.	https://si	1	https://si	333333	https://si	C0DEED	
3.49E+08	suraj jadh	surajjadh	1444	35	390	1283	0	Sat Aug 06 01:23:19 +		en		amravati ,maharatra ,india				http://a0.	NULL	1	https://si	333333	https://si	C0DEED	1
4.76E+08	Nirmal	smartnirn	1351	7	328	1273	1	Fri Jan 27 11:24:28 +		en			1	1		http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Rochell C	rochellca	43	17	641	0	0	Sat Jun 23 15:30:29 +		en		DIADEMA	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Thomase	thomaser	50	20	630	0	0	Sat Jun 23 15:31:34 +		en		In your hc	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Arnetta W	whitfield	68	22	602	0	0	Sat Jun 23 15:31:45 +		en		Arizona	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Tonia Jac	toniajaco	60	14	592	0	0	Sat Jun 23 15:32:21 +		en		ÐÑ‡Ð¸Ð½	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Kasandra	kasandra	52	27	620	0	0	Sat Jun 23 15:32:47 +		en		Rio Granc	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Stefania !	searsfo	67	32	639	0	0	Sat Jun 23 15:32:53 +		en		queens r	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Shamika	shamikag	44	21	614	0	0	Sat Jun 23 15:32:47 +		en		pontiana	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Esperanz	maxwellI	66	23	650	0	0	Sat Jun 23 15:32:52 +		en		malaysia	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Janina Ca	janinahzr	56	20	643	0	0	Sat Jun 23 15:33:05 +		en		WORLDW	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Myrtle Bo	bowmanf	65	23	641	0	0	Sat Jun 23 15:33:06 +		en		Fairburn,	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Geneviev	geneviev	79	20	779	0	0	Sat Jun 23 15:33:12 +		en		Arizona	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Awilda Fc	awildagx	59	16	622	0	0	Sat Jun 23 15:32:59 +		en		bangalor	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Margarite	roblesxyf	64	15	649	0	0	Sat Jun 23 15:33:15 +		en		Canada	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Delila Ke	delilanbl	81	29	756	0	0	Sat Jun 23 15:33:22 +		en		Liverpool	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Iris Wiggi	wigginssf	62	25	600	0	0	Sat Jun 23 15:33:05 +		en		Where Th	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Izola Tillr	izolayww	70	22	621	0	0	Sat Jun 23 15:33:16 +		en		Accra- Gh	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	
6.16E+08	Brian Dic	brianeg	52	19	735	0	0	Sat Jun 23 15:33:32 +		en		Victoria, E	1			http://a0.	NULL	1	https://si	333333	https://si	C0DEED	

Fig 5.1.2 Fake Account Dataset

5.2 Determination of Accuracy

5.2.1 Percentage of Messages found Offensive

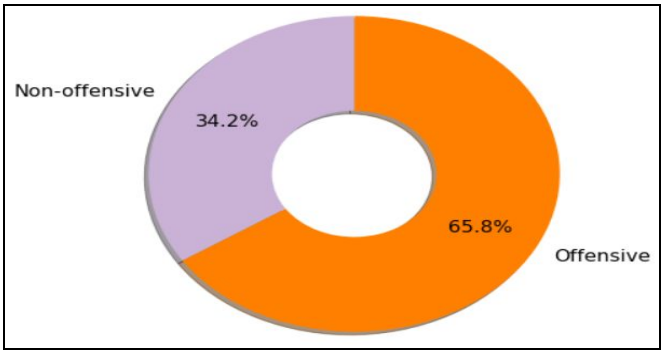


Fig 5.2.1 Offencive Message(%)

5.2.2 Accuracy Comparison For Cyber Bullying

	Algorithm	Accuracy: Test
0	SGDClassifier	0.929474
1	BaggingClassifier	0.926568
2	LogisticRegression	0.926679
3	DecisionTreeClassifier	0.921426
4	LinearSVC	0.920756
5	RandomForestClassifier	0.916620
6	AdaBoostClassifier	0.907902
7	MultinomialNB	0.893596
8	KNeighborsClassifier	0.855594

Fig 5.2.2 Accuracy Comparison For Cyber Bullying

5.2.3 Accuracy Comparison for Fake Account

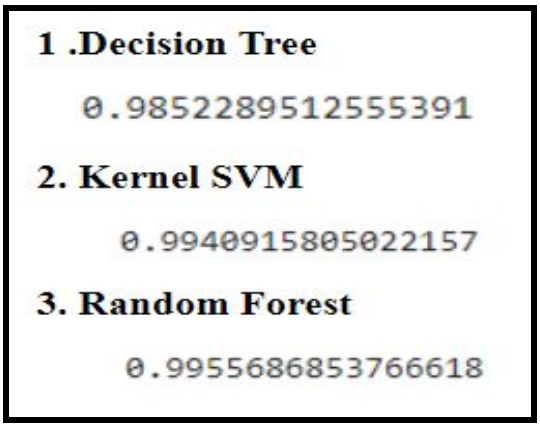


Fig 5.2.3 Accuracy Comparison For Fake Account

CHAPTER – 6

PLAN OF ACTION FOR THE NEXT SEMESTER

6.1.Work done till date

We had collected the dataset for both cyber bullying and fake account, preprocessed it and extracted the necessary parameters and trained it on the different models like Decision Tree, Random forest, SVM , SGD Classifier and we had got the accuracy of >=0.92 for the Cyberbullying model and >=0.98 for the fake account detection.

6.2.Plan of action for project II

We will create the blogging website with GUI where the user will first login to the portal with his/her credentials or can sign up on our portal for new registration and connect to the database and also we will try to use train out dataset on the neural network and we will try to use the NLP for the cyberbullying text where we will try to increase the accuracy and will integrate with our system also we will try to add an additional 24 hours authentication in the system so that user can verify its details as and also as an addition we will try to add an feature of getting the IP address of the user so that it will be helpful to the cyber branch to locate the user.

CHAPTER – 7

CONCLUSIONS

In this project, we proposed an approach to detect cyberbullying using machine learning techniques. We have evaluated our model on First Cyberbully by comparing accuracies of the different algorithms we can conclude that SGDClassifier gives us the best accuracy of 92% and from the model results evaluated on the Fake Account we found that the best accurate algorithm is Decision Tree having the accuracy of By using machine learning algorithms to its full extent, we have eliminated the need for manual prediction of a fake account, which needs a lot of human resources and is also a time-consuming process.

CHAPTER – 8
REFERENCES

Journal Paper referred :

[1] S. Gurajala, J. S. White, B. Hudson, and J. N. Matthews, “Fake Twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach,” in Proceedings of the 2015 International Conference on Social Media & Society (SMSociety'15), Toronto, Ontario, Canada, 2015.View at: Publisher Site | Google Scholar.

[2]Freddy Tapia; Cristina Aguinaga; Roger Luje,”Detection of Behavior Patterns through Social Networks like Twitter, using Data Mining techniques as a method to detect Cyberbullying”2018 7th International Conference On Software Process Improvement (CIMPS),Guadalajara, Jalisco, Mexico, Mexico,28 January 2019.

[3]Kim D. Gorro, M. J. Sabellano, Ken Gorro, C. Maderazo, Kris Capao,”Classification of Cyberbullying in Facebook Using Selenium and SVM ”, Published 2018 Computer Science 2018 3rd International Conference on Computer and Communication Systems (ICCCS)

[4]Milan Dordevic, Pardis Pourghomi, Fadi Safieddine, "Identifying Fake News from the Variables that Governs the Spread of Fake News.", *Semantic and Social Media Adaptation and Personalization (SMAP) 2020 15th International Workshop on*, pp. 1-6, 2020.

[5]Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B.J., Wong, K.F. and Cha, M., 2016. Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (pp. 3818-3824).

[6]Van Hee C, Jacobs G, Emery C, Desmet B, Lefever E, Verhoeven B, et al. (2018) Automatic detection of cyberbullying in social media text. PLoS ONE 13(10): e0203794.

CHAPTER–9

APPENDIX

a. List of Figures