

# Combating Online Abuse and Online Reporting Functions on Social Networking Sites

Mr.M.Mohan Periyasamy[AP/IT], M.E.,  
Department of Information Technology,  
V.S.B Engineering College, Karur,  
Tamil Nadu, India.

Ms.P.Praveena, B.Tech., Student,  
Department of Information Technology,  
V.S.B Engineering College, Karur,  
Tamil Nadu, India.

Ms.R. Muthu Prabha, B.Tech., Student,  
Department of Information Technology,  
V.S.B Engineering College, Karur,  
Tamil Nadu, India.

Ms.K.Somamugi, B.Tech., Student,  
Department of Information Technology,  
V.S.B Engineering College, Karur,  
Tamil Nadu, India.

**Abstract**-Online abuse has emerged as a huge problem across the Internet and especially on social networking sites (SNS). Online social network is key for spread the message and abuse videos to the people. The main objective of this study is to identify the factors driving people to use online reporting functions on Social network sites we need to analysis and block that website based on features of machine learning algorithm.

**Keywords**-Tweets messages, Abuse detection, Social crime, Preprocessing, Tokenization and Stamping.

## I.INTRODUCTION

Social networking site has become the world's most popular online social venue where many people get connected with others easily. Many social networking sites (i.e., Facebook, Twitter, Instagram and Snapchat) have responded to the call for prevention and intervention of online abuses and launched built-in functions on their platforms. The rapid growth of digital technology has revolutionized our lives, transforming the way we connect and communicate.

## II.PROJECT DESCRIPTION

The severity of abuse and the damage to victims has increased. A more coordinated attempt aiming to the perception of a widely accepted definition was proposed in the Convention on Cybercrime. Cybercrime involves a mixture of diverse typical crimes with new illegal acts. Individual cybercrime incidents are occurrences of particular criminal offences and, as multiple national crime statistics and surveys demonstrate are steadily increasing.

## III.FEASIBILITY STUDY

The feasibility study deals with all the analysis that takes up in developing the project. Each structure has to be thought of in the developing of the project, as it has to serve the end user in a user- friendly manner. One must know the type of information to be gathered

and the system analysis consist of collecting. The main objective of the system analysis is to study the existing operation and to learn and accomplish the processing activities. Calculating cloud area status at a given refresh period through windows application needs to be analysed well. Cloud areas must be grouped based on their processing ability. According to their processing and storage power, the partial job needs to assign to them. The details are processed through coding themselves. It will be controlled by the programs alone.

### A) ECONOMIC FEASIBILITY

The organization has to buy a personal computer with a keyboard and a mouse. The user can be given responses on asking questions, justification of any capital out lays that it will reduce expenditure or improve the quality of service or goods, which in turn may be expected to provide the increased profits.

### B) OPERATIONAL FEASIBILITY

The Proposed system accessing process to solves problems what occurred in existing system. The current day-to-day operations of the organization can be fit into this system.

### C) TECHNICAL FEASIBILITY

The cost and benefit analysis may be concluded that computerized system is favourable in today's fast moving world. The assessment of technical feasibility must be based on an outline design of the system requirements in terms of input, output, files, programs and procedure. The project aims to assign multiple nodes after the job is split according to the nodes capability from the given application. The current system aims to overcome the problems of the existing system. The current system is to reduce the technical skill requirements so that more number of users can access the application

IV. PROPOSED SYSTEM

The proposed system predict and block the abuse videos and messages on online social network Abuse words and website features are extracted from the data set and trained into the ANN network Its machine learning algorithm its analyze those features trained into the layers This project aims to contribute toward better understanding cybercrime by proposing a schema-based cybercrime incident description that:

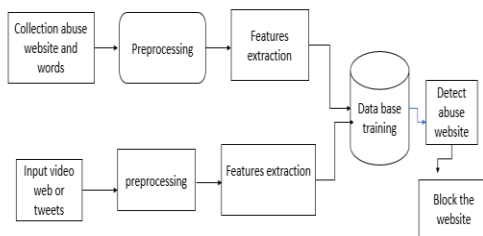
- 1) Identifies the features of a cybercrime incident and their potential elements and
- 2) Provides a two-level offence classification system based on specific criteria.

The proposed schema can be extended with a list of recommended actions, corresponding measures and effective policies that counteract the offence type and subsequently the particular incident.

V. ADVANTAGES OF PROPOSED SYSTEM

- High accuracy
- It provides co-ordination and cooperation among all countries of the world for security of cyberspace.
- More accurate for detection of cyber system.

VI. ARCHITECTURE FOR PROPOSED SYSTEM



VII. EXISTING SYSTEM

Gordon and Ford proposed a typology consisted of two categories. Type I offences characterize singular or discrete events facilitated by the introduction of malware programs such as keystroke loggers, viruses, and rootkits. Type II offences are facilitated by programs that are not classified as crime ware, and they are generally repeated contacts or events from the perspective of the user. A much broader classification,

VIII. DISADVANTAGES OF EXISTING SYSTEM

- Predicting rate is low
- Easy to unblock those service
- Less secure
- High data leakage
- Easier to spread the message

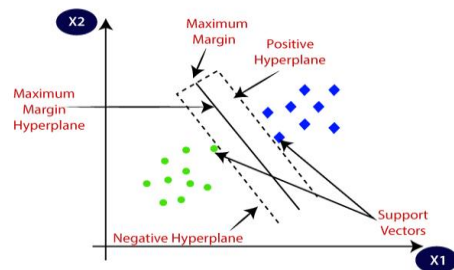
IX. ALGORITHM

a) SUPPORT VECTOR MACHINE ALGORITHM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyper plane.

SVM chooses the extreme points/vectors that help in creating the hyper plane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyper plane:



b) Types of SVM

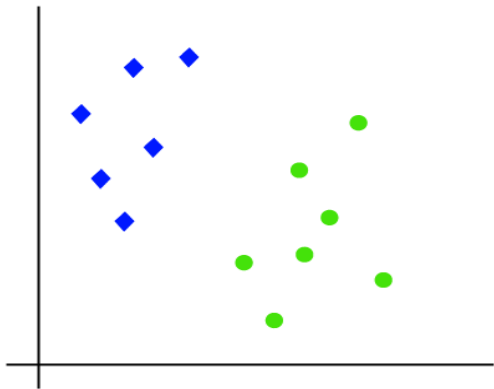
SVM can be of two types:

- Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

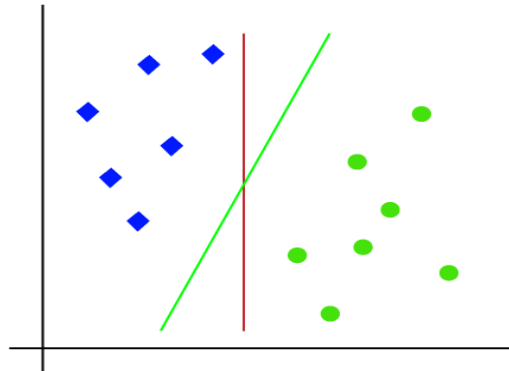
c) How does SVM works?

Linear SVM:

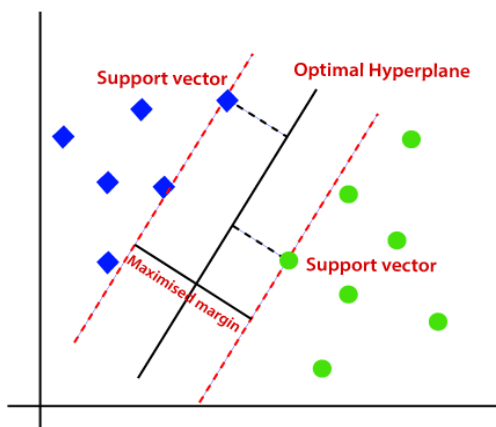
The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features  $x_1$  and  $x_2$ . We want a classifier that can classify the pair  $(x_1, x_2)$  of coordinates in either green or blue. Consider the below image:



So as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the below image:

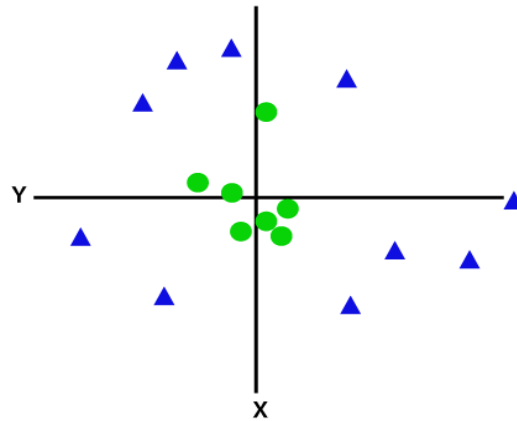


Hence, the SVM algorithm helps to find the best line or decision boundary; this best boundary or region is called as a hyper plane. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyper plane is called as margin. And the goal of SVM is to maximize this margin. The hyper plane with maximum margin is called the optimal hyper plane



Non-Linear SVM:

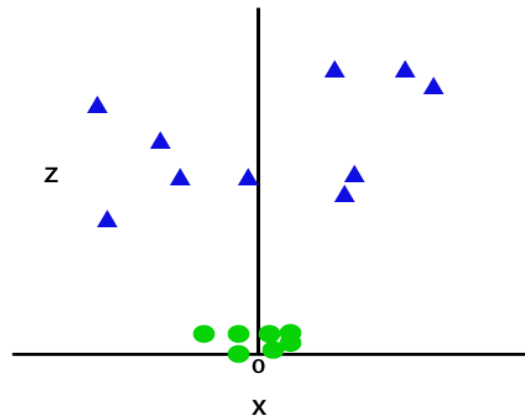
If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the below image:



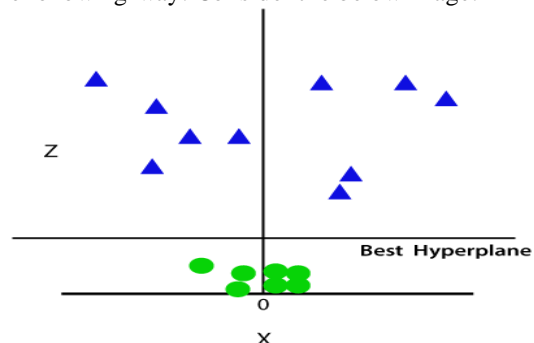
So to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as:

$$z = \sqrt{x^2 + y^2}$$

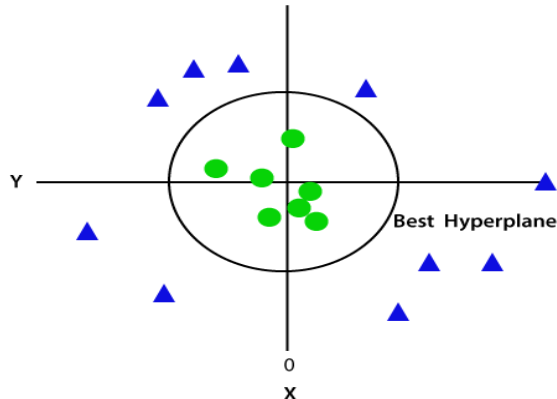
By adding the third dimension, the sample space will become as below image:



So now, SVM will divide the datasets into classes in the following way. Consider the below image:



Since we are in 3-d Space, hence it is looking like a plane parallel to the x-axis. If we convert it in 2d space with  $z=1$ , then it will become as:



Hence we get a circumference of radius 1 in case of non-linear data.

*X.MODULE*

1. Data set collection
2. Pre-processing
3. Features extraction
4. Training
5. performance evaluation

**DATA SET COLLECTION**

Data collection system (DCS) is a computer application that facilitates the process of data collection, allowing specific, structured information to be gathered in a systematic fashion, Subsequently enabling data analysis to be performed on the information.

**PREPROCESSING**

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

*FEATURE EXTRACTION*

When the input data to an algorithm is too large to be processed and it is suspected to be redundant. Then it can be transformed into a reduced set of features Determining a subset of the initial features is called feature selection

*TRAINING*

Training data set, you include both the input data together along with the corresponding expected output. This provides your model with “ground truth” data that’s usually prepared by humans or in a semi-automated way. During the testing phase, you present this data to train your model by pairing the input with

Its respective desired output.

*PERFORMANCE EVALUATION*

Performance Analysis is a specialist discipline involving systematic observations to enhance performance and improve decision making, primarily delivered through the provision of objective statistical (Data Analysis) and visual feedback (Video Analysis).

*XI.CONCLUSION*

To formalize the problem, a social relationship model is proposed to demonstrate the follower and friend relationships. A Web crawler using Twitter API methods is also mentioned to extract real data set from publicly available in formation based on religious keywords. The extracted different activities are then flrtd out and detected various frauds acts. Finally, the performance of the framework is measured with three classifiers. The results showed much closer accuracy of each classifier. So, we claim that the proposed system shows a feasible detecting and tracking mechanism of any religious abuses in social media.

*XII.REFERENCES*

[1] P. Analytics, Twitter study, 2009. [Online]. Available: [http://www. Pearanalytics.com/wp-content/uploads/2009/08/](http://www.Pearanalytics.com/wp-content/uploads/2009/08/). [Accessed: 27/08/2018].

[2] A. H. Wang, “Don’t follow me: Spam detection in twitter,” in Security and cryptography (SECURITY), proceedings of the 2010 international conference on. IEEE, 2010, pp. 1–10.

[3] J. Ratkiewicz, M. Conover, M. R. Meiss, B. Goncalves, A. Flammini, and F. Menczer, “Detecting and tracking political abuse in social media.” ICWSM, vol. 11, pp. 297–304, 2011.

[4] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, “Who is tweeting on twitter: human, bot, or cyborg?” in Proceedings of the 26th annual computer security applications conference. ACM, 2010, pp. 21–30.

[5] C. Honey and S. C. Herring, “Beyond microblogging: Conversation and collaboration via twitter,” in System Sciences, 2009. HICSS’09. 42nd Hawaii International Conference on. IEEE, 2009, pp. 1–10.

[6] T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake shakes twitter users: real-time event detection by social sensors,” in Proceedings of the 19th international conference on World Wide Web. ACM, 2010, pp. 851–860

[7] M. Jiang, P. Cui, and C. Faloutsos, “Suspicious behavior detection: Current trends and future directions,” IEEE Intelligent Systems, vol. 31, no. 1, pp. 31–39, 2016.

[8] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[9] Religion Spirituality Keywords. [Online]. Available: <https://kdp.amazon.com/>. [Accessed: 27.08.2018].