ISSN: 2366-1313

Detection of Malevolent statements using machine learning algorithms

[1]P. Swetha, [2]Akshita Sangamkar, [3]D. Vinod Kumar, [4] Ch. Lakshana

[1]Assistant Professor, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

pswetha90@gmail.com

[2]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,
akshitasangamkar@gmail.com

[3]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,
vinoddharavath09@gmail.com

[4]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,
chalvadilakshana@gmail.com

Abstract: *Prior to the innovation of information communication technologies (ICT), social interactions evolved within small cultural boundaries such as geo spatial locations. The recent developments of communication technologies have considerably transcended the temporal and spatial limitations of traditional communications. These social technologies have created a revolution in user generated information, online human networks, and rich human behaviour-related data. However, the misuse of social technologies such as social media (SM) platforms, has introduced a new form of aggression and violence that occurs exclusively online. A new means of demonstrating aggressive behaviour in SM websites are highlighted in this paper. The motivations for the construction of prediction models to fight aggressive behavior in SM are also outlined. We comprehensively review cyberbullying prediction models and identify the main issues related to the construction of cyberbullying prediction models in SM. This paper provides insights on the overall process for cyberbullying detection and most importantly overviews the methodology. Though data collection and feature engineering process has been elaborated, yet most of the emphasis is on feature selection algorithms and then using various machine learning algorithms for prediction of cyberbullying behaviours. Finally, the issues and challenges have been highlighted as well, which present new research directions for researchers to explore.*

Keywords: *social media, cyber bulling, machine learning, information communication technologies*.

I. INTRODUCTION

ISSN: 2366-1313

Machine or deep learning algorithms help researchers understand big data. Abundant information on humans and their societies can be obtained in this big data era, but this acquisition was previously impossible. One of the main sources of human-related data is social media (SM). By applying machine learning algorithms to SM data, we can exploit historical data to predict the future of a wide range of applications. Machine learning algorithms provide an opportunity to effectively predict and detect negative forms of human behaviour, such as cyberbullying. Big data analysis can uncover hidden knowledge through deep learning from raw data. Big data analytics has improved several applications, and forecasting the future has even become possible through the combination of big data and machine learning algorithms.

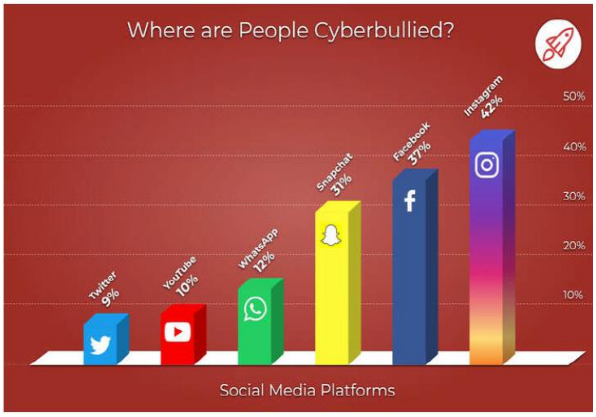

Fig.1 Social media platforms

An insightful analysis of data on human behaviour and interaction to detect and restrain aggressive behaviour involves multifaceted angles and aspects and the merging of theorems and techniques from multidisciplinary and interdisciplinary fields. The accessibility of large-scale data produces new research questions, novel computational methods, interdisciplinary approaches, and outstanding opportunities to discover several vital inquiries quantitatively. However, using traditional methods (statistical methods) in this context is challenging in terms of scale and accuracy. These methods are commonly based on organized data on human behaviour and small-scale human networks (traditional social networks).

Applying these methods to large online social networks (OSNs) in terms of scale and extent causes several issues. On the one hand, the explosive growth of OSNs enhances and disseminates aggressive forms of behaviour by providing platforms and networks to commit and propagate such behaviour. On the other hand, OSNs offer important data for exploring human behaviour and interaction at a large scale, and these data can be used by researchers to develop effective methods of detecting and restraining misbehaviour and/or aggressive behaviour. OSNs provide criminals with tools to perform aggressive actions and networks to commit misconduct. Therefore, methods that address both aspects (content and network)

ISSN: 2366-1313

should be optimized to detect and restrain aggressive behavior in complex systems.

Cyberattacks are currently the most pressing concern in the realm of modern technology. The word implies exploiting a system's flaws for malicious purposes, such as stealing from it, changing it, or destroying it. Malware is an example of a cyberattack. Malware is any program or set of instructions that is designed to harm a computer, user, business, or computer system [1]. The term "malware" encompasses a wide range of threats, including viruses, Trojan horses, ransomware, spyware, adware, rogue software, wipers, scareware, and so on. Malicious software, by definition, is any piece of code that is run without the user's knowledge or consent [2].

In particular, this study demonstrated that detecting harmful traffic on computer systems, and thereby improving the security of computer networks, was possible employing the findings of malware analysis and detection with machine learning algorithms to compute the difference in correlation symmetry (Naive Byes, SVM, J48, RF, and with the proposed approach) integrals.

Malware detection modules are responsible for analysing data they have collected and been trained with to determine whether or not a specific piece of software or network connection constitutes a security concern [3,4]. As an illustration, consider a machine learning system that can explicitly express the principles that underlie the patterns it has observed [5]. Algorithms that have been trained by machine learning systems can improve their ability to predict using feedback regarding how well they performed on previous tasks and using that information to make changes.

II. LITERATURE SURVEY

A survey on Twitter using the Naïve Bayes Classifier method and Support Vector Machine Model was undertaken by Vandana Nandakumar et al. He might determine the probabilities of the feature using the classifier method. He went on to compile the two algorithms' graphs. He could comparison and calculate the data independently of the Twitter data set from his precision factor. He compared the output variables of both algorithms by predicting them. His investigation concluded that a greater accuracy result than a vector holder was supplied by the Naïve Bayes rating. Only the algorithm Naïve Bayes Classifier was available for us to choose and to work on this finding. Researchers at the University of San Carlos have been utilizing the Web scraper tool to collect Facebook-related cyberbullying posts. The Support Vector

ISSN: 2366-1313

Machinery Model was used to classify the collected data model once the data was collected. They attained an accuracy of 88% from their analysis and recorded 87%. One of the issues facing this technique was that this technology could harvest only 24 posts. They were also able to produce incorrect findings, which made SVM's failure to characterize a threat easier. It allowed us to identify the best strategy we can utilize as a classification method based on this research. It is also very crucial in cyberbullying research that we determine the roles of cyberbullying participants. With this in mind, the first scholars to decide the role in the harassment environment were Salmillivali C et al. They conducted

various polls among the young people in which the actual intimidation situation was involved. Six participants/victims were used who were victims of repeated harassment among the bullies. The acquired data were used in the Notation portion where the bullying commentaries were analysed and used to automatically detect.

Livingstone S et al. investigates the problems to be met during the identification of cyberbullying. In his research, he may observe that only an appropriate data set is available is one of the significant problems. Nevertheless, much has been done to improve the data set collecting for the design and building of the model used to detect these remarks. We have platforms like the Kaggle, the Spring Form, which plays a major part in identifying this data.

They used the Fromspring.me a dataset and assessed papers with poor words to identify the harmful words, according to Reynolds et al. He employed numerous technologies to train and classify data during his investigation. The support vector machine and decision trees are common techniques he may implement. The algorithm of the Decision Tree was better and could reach a 78.5% level. On that basis, there were other obstacles and one of them was the

gender information challenge. This involves the usage of diverse vocabulary by men and women. Moreover, on the website, there were several different curse words. (Redmond,2020) Cyberbullying is a rapidly increasing problem that has a harmful influence on society, in particular on the students' numbers, including text, email, mobile phones, chat rooms and websites. The issues of cyberbullying can influence a person exposed to cyberbullying's cognitive capacities. The author asserted that there is no clear definition of cyberbullying because it fluctuates depending on the user's perception. The unpleasant behaviours of the assailant may impact on the

ISSN: 2366-1313

individual's behaviour in the face of society and repeated harassment through the use of digital technologies. The willingness to influence each individual's behaviour and cyberbullying might damage the individual's performance. The sense of stress or mental discomfort might impact the individual's behaviour and violent behaviour among persons exposed to cyberbullying. Unknowledge of the use of technology can influence their behaviour, and ITCs play a significant part in addressing cyberbullying's disadvantages. The person who utilizes technology every day or social media is more susceptible to such problems.

Impact a person's mental skills, physical behaviours, and personality in the case of young people or 1online students. This form of risk negativizes job development or advancement and also exposes the academic success of the students to such technological hazards. These risks must .be dealt with by the authority of the school and university to improve the education of the student supplied by educators. (Redmond,2020), by giving more literature in this field, created a cyber bullying framework for the number of educators. The author established a conceptual framework. The proposed framework can be implemented in order to implement proactive programming for the number of persons exposed to cyber bullying. The author's framework includes identification, prevention and risk management. Cyber bullying can be identified through consideration of crucial factors such as online disinhibition, constant access, permanent records and power imbalances.

(Bai 2020) described the individual's emotional well-being as a result of cyber bullying and the influence on student learning of cyber-bullying. The educator's personal experience plays a key part in encouraging the kids exposed to cyber bullying difficulties. (Bai,2020) described the pessimism of the individual who is exposed to cyber bullying, another element which may be assisted. Family perspective and understanding contribute to managing the negative consequences of cyber bullying.

Fig.2 Key mediating aspects

The key mediating aspect of this relationship is the family incomparability that is associated positively to cyber

ISSN: 2366-1313

bullying and mental health. An individual's emotional intelligence has a particular effect on hopelessness and family relationships and young people are more prone to such problems. There is also a sensation of dissatisfaction and aggression among those who are exposed to various cyber bullying consequences. The author employed this strategy to address the cyber bullying issues through the theoretical analysis. The total number of participants that the author has assessed in this study is 3030. A focus group consisting of adolescent men and women is used by the author. In this approach, a quantitative analysis was used to evaluate the data. Cyber bullying also has a negative influence on the behaviour of the individual towards their relationship between the student and the family.

(Kim,2020) stressed that cyber bullying is a danger factor that could jeopardize immigrants' academical progress because it has a detrimental impact on kids or students' performance. According to the author, internet bullying is bad to sense of belonging in connection with schools, and through cyber bullying on the conduct of kids, adverse effects have been imposed. The understanding of ICT use in order to avoid cyber bullying problems plays an important part in boosting school pupils' performance or academic achievement. In this method, the author attempted to uncover the proactive approaches utilized to reduce cyber bullying risk among immigrant young people. Indirect use and awareness of the ICT help manage this risk. However, knowledge and skills relating to the efficient use of technology are required for this purpose.

Dinakar et al.2011 identified ways you can improve your performance. First, the bullying terms were labeled as categories and the binary classificator was used for each category. He might employ the following categories in their research: sexuality, race/culture and IQ. He then implemented the tree algorithm via JRip, which is an application of the RIPPER proposalal rule learning algorithm. Use the same technique in his research he has been able to produce better results than utilizing the SVM algorithm using the SMO. Data from comments on the You Tube Video was utilized in his research.

III. PROPOSED METHODOLOGY

The proposed system is constructing cyberbullying prediction models is to use a text classification approach that involves the construction of machine learning classifiers from labelled text instances. Another means is to use a lexicon-based model that involves computing orientation for a document from the semantic orientation of words or phrases in the

ISSN: 2366-1313

document. Generally, the lexicon in lexicon-based models can be constructed manually or automatically by using seed words to expand the list of words. However, cyberbullying prediction using the lexicon-based approach is rare in literature.

The primary reason is that the texts on SM websites are written in an unstructured manner, thus making it difficult for the lexicon-based approach to detect cyberbullying based only on lexicons. However, lexicons are used to extract features, which are often utilized as inputs to machine learning algorithms. For example, lexicon-based approaches, such as using a profane-based dictionary to detect the number of profane words in a post, are adopted as profane features to machine learning models. The key to effective cyberbullying prediction is to have a set of features that are extracted and engineered.

The system is more effective due to logistic regression Classification and unsupervised machine learning. An effective cyberbullying prediction models is to use a text classification approach that involves the construction of machine learning classifiers from labeled text instance and also is to use a lexicon-based model that involves computing orientation for a document from the semantic orientation of words or phrases in the document.

IV. IMPLEMENTATION

Development Tools Used

PYTHON

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. An interpreted language, Python has a design philosophy that emphasizes code readability, and a syntax that allows programmers to express concepts in fewer lines of codes than might be used in languages such as C++ or Java

DJANGO

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

MACHINE LEARNING ALGORITHMS:

K-nearest neighbour

The k-nearest neighbour algorithm is a pattern recognition model that can be used for classification as well as regression. Often abbreviated as k-NN, the k in k-nearest neighbour is a positive integer, which is typically small. In either classification or regression, the input will

ISSN: 2366-1313

consist of the k closest training examples within a space

Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

SVM Algorithm

SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. 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.

Naïve Bayes

Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

Decision tree

The goal of decision tree learning is to create a model that will predict the value of a target based on input variables.

Extreme Machine Learning

Extreme learning machine (ELM) is a training algorithm for single hidden layer feedforward neural network (SLFN), which converges much faster than traditional methods and yields promising performance

MODULES

Admin:

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as view and authorize users, View all posts, Detect Cyber Bullying Users, Find Cyber Bullying Reviews Chart.

Viewing and Authorizing User : In this module, the admin views all users details and authorize them for login permission. User Details such as User Name, Address, Email Id, Mobile Number.

View all posts: In this module, the admin can see all the posts added by the users with post details like post name, description and post image.

Detect Cyber Bullying Users: In this module, the admin can see all the Cyber Bullying Users (The users who had posted a comment on posts using cyber bullying words which are all listed by the admin to detect and filter). In this, the results shown as, Number of items found for a

ISSN: 2366-1313

corresponding post like Violence, Vulgar, Offensive, Hate, Sexual.

Find Cyber Bullying Reviews Chart: In this module, the admin can see all the posts with number of cyber bullying comments posted by users for particular post.

USER:

In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like Posting Your Messages as Posts by giving details, view all your cyber bullying comments on your friend posts.

Add Post : In this, the user can add their own posts by giving post details such as, post title, description, uses, and image of post.

View all Friends Posts and Comment (Cyber bullying Related) In this, the user can see his all-friend's post details (post title, description, uses, creator and image of post) and can comment on posts. Don't Post If the comment consists of Cyber bullying words and Shows the reason why comment is not posted by indicating Detected Cyber Bullying Words like Numbers of Cyber Bullying words Related to Filter Violence found in comment, Numbers of Cyber Bullying words Related to Filter Vulgar found in comment, Numbers of Cyber Bullying words Related to Offensive found in comment, Numbers of Cyber Bullying words Related to Hate found in comment, Numbers of Cyber Bullying words Related to Sexual found in comment.

V. RESULTS

SYSTEM ARCHITECTURE

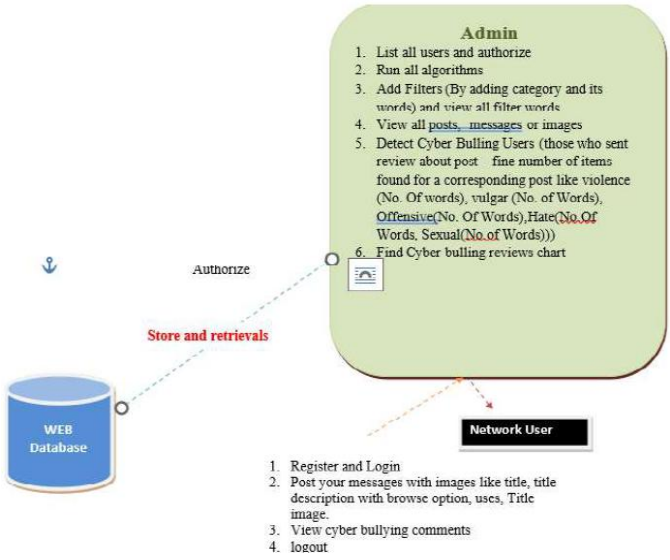

Fig.3 System architecture.

ISSN: 2366-1313

Fig.4 In above screen, now click on 'Register' button to add details

Fig.5 In above screen login as 'admin' by giving username as 'admin' and password as admin

ISSN: 2366-1313

Fig.6 In above screen admin will enter username and then select ‘Accept’ or ‘Reject’ option to give permission.

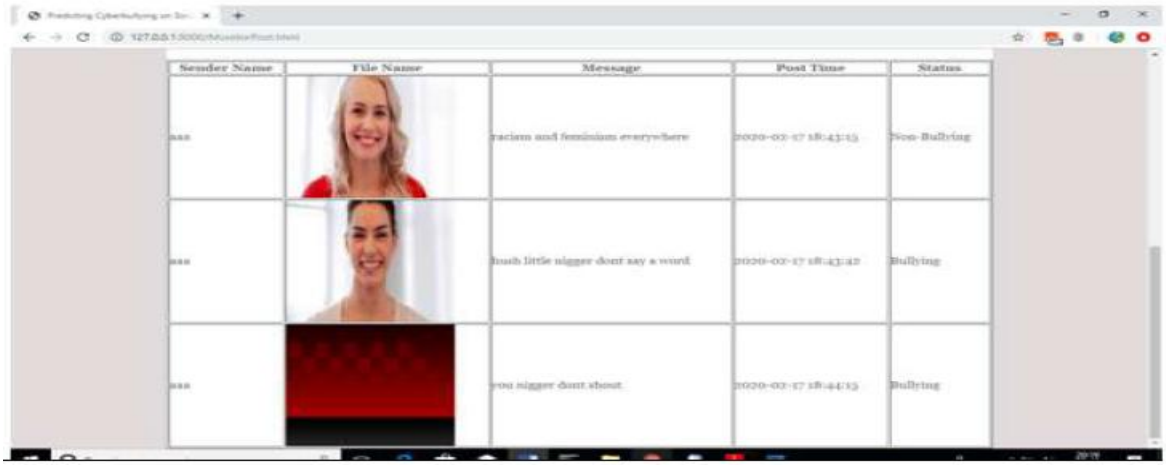

Fig.7 In above screen applications will automatically detect whether message is non bullying or bullying from machine learning algorithms.

Fig.8 Here the admin has to select each algorithm and click on submit button to train model and we will get accuracy also for each algorithm

ISSN: 2366-1313

Fig.9 In the above screen advance “Extreme machine learning” gave 98% accuracy .

Fig.10 In above screen I added some messages and uploaded a photo also.

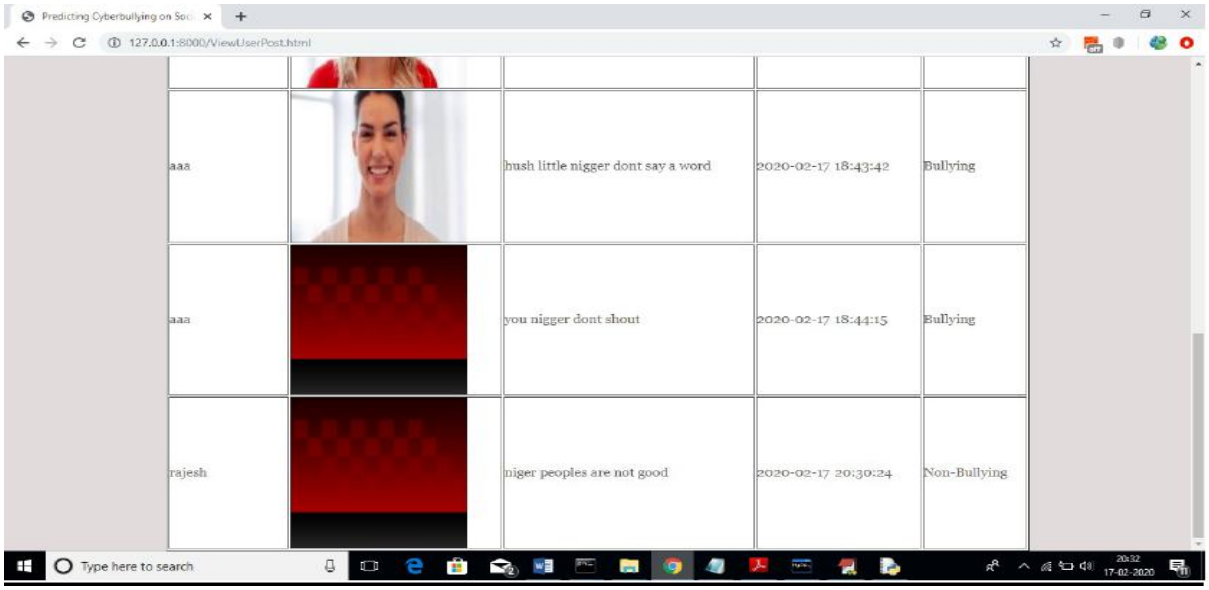

Fig.11 In above screen we are seeing posts from all users and rajesh post predicted as ‘Non - Bullying.

VI. CONCLUSION

This study reviewed existing literature to detect aggressive behavior on SM websites by using machine learning approaches. We specifically reviewed four aspects of detecting cyberbullying messages by using machine learning approaches, namely, data collection, feature engineering, construction of cyberbullying detection model, and evaluation of constructed

ISSN: 2366-1313

cyberbullying detection models. Several types of discriminative features that were used to detect cyberbullying in online social networking sites were also summarized. In addition, the most effective supervised machine learning classifiers for classifying cyberbullying messages in online social networking sites were identified. One of the main contributions of current paper is the definition of evaluation metrics to successfully identify the significant parameter so the various machine learning algorithms can be evaluated against each other. Most importantly we summarized and identified the important factors for detecting cyberbullying through machine learning techniques specially supervised learning. For this purpose, we have used accuracy, precision recall and f-measure which gives us the area under the curve function for modelling the behaviours in cyberbullying. Finally, the main issues and open research challenges were described and discussed.

REFERENCES

[1] V. Subrahmanian and S. Kumar, ‘‘Predicting human behavior: The next frontiers,’’ Science, vol. 355, no. 6324, p. 489, 2017.

[2] H. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas, ‘‘Homophily in the digital world: A LiveJournal case study,’’ IEEE Internet Comput., vol. 14, no. 2, pp. 15–23, Mar./Apr. 2010.

[3] M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana, ‘‘Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network,’’ Comput. Hum. Behav., vol. 63, pp. 433–443, Oct. 2016.

[4] L. Phillips, C. Dowling, K. Shaffer, N. Hodas, and S. Volkova, ‘‘Using social media to predict the future: A systematic literature review,’’ 2017, arXiv:1706.06134. [Online]. Available: https://arxiv.org/abs/1706.06134

[5] H. Quan, J. Wu, and Y. Shi, ‘‘Online social networks & social network services: A technical survey,’’ in Pervasive Communication Handbook. Boca Raton, FL, USA: CRC Press, 2011, p. 4.

[6] J. K. Peterson and J. Densley, ‘‘Is social media a gang? Toward a selection, facilitation, or enhancement explanation of cyber violence,’’ Aggression Violent Behav., 2016.

[7] BBC. (2012). Huge Rise in Social Media. [Online]. Available: http://www.bbc.com/news/uk-20851797

ISSN: 2366-1313

[8] P. A. Watters and N. Phair, ‘‘Detecting illicit drugs on social media using automated social media intelligence analysis (ASMIA),’’ in Cyberspace Safety and Security. Berlin, Germany: Springer, 2012, pp. 66–76.

[9] M. Fire, R. Goldschmidt, and Y. Elovici, ‘‘Online social networks: Threats and solutions,’’ IEEE Commun. Surveys Tuts., vol. 16, no. 4, pp. 2019–2036, 4th Quart., 2014.

[10] N. M. Shekokar and K. B. Kansara, ‘‘Security against sybil attack in social network,’’ in Proc. Int. Conf. Inf. Commun. Embedded Syst. (ICICES), 2016, pp. 1-5

[11] Prasadu Peddi (2019), "Data Pull out and facts unearthing in biological Databases", International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.