International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

CYBER BULLYING PREDICTION THROUGH MACHINE LEARNING

M Kanakeshwar Reddy [1],Sevakula Sai Charan[2], Nandikonda Akhila[3],Pallapu Anupriya[4]
[2,3,4] UG Scholars, Department of CSE, **AVN Institute of Engineering and Technology,**Hyderabad, Telangana, India.
[1] Assistant Professor, Department of CSE, **AVN Institute of Engineering and Technology**, Hyderabad, Telangana, India.

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 behavior in SM websites are highlighted in this paper. The motivations for the construction of prediction models to ght 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 behaviors. Finally, the issues and challenges have been highlighted as well, which present new research directions for researchers to explore.

INTRODUCTION

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 behavior, 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.

An insightful analysis of data on human behavior and interaction to detect and restrain aggressive behavior involves multifaceted angles and aspects and the merging of theorems and techniques from multidisciplinary and interdisciplinary_elds. 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 behavior 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 behavior by providing platforms and networks to commit and propagate such

behavior. On the other hand, OSNs offer important data for exploring human behavior and interaction at a large scale, and these data can be used by researchers to develop effective methods of detecting and restraining misbehavior and/or aggressive behavior. OSNs provide criminals with tools to perform aggressive actions and networks to commit misconduct. Therefore, methods that address both aspects (content and network) should be optimized to detect and restrain aggressive behavior in complex systems.

LITERATURAL SURVEY

Title: Predicting human behavior: The next frontiers.

Author: V. Subrahmanian and S. Kumar.

Behavioral choice theories intention to give an explanation for human behavior. Can they assist are expecting it? An open tournament for prediction of human selections in fundamental monetary choice tasks is presented. The results propose that integration of certain behavioral theories as capabilities in system mastering systems gives the excellent predictions. Surprisingly, the most beneficial theories for prediction build on basic homes of human and animal learning and are very distinctive from mainstream selection theories that target deviations from rational choice. Moreover, we find that theoretical capabilities must be based totally not only on qualitative behavioral insights (e.G. "loss aversion"), but additionally on quantitative behavioral foresights generated by practical descriptive models (e.G. Prospect Theory). Our analysis prescribes a recipe for derivation of explainable, useful predictions of human decisions.

Title: Homophily in the digital world: A LiveJournal case study

Author: H. Lauw, J. C. Shafer, R. Agrawal, and A. Ntoulas.

Are two users much more likely to be buddies in the event that they share not unusual interests? Are two users more likely to share commonplace pursuits if they may be buddies? The authors examine the phenomenon of homophily within the digital world by answering those imperati ve questions. Unlike the physical world, the virtual world does not impose any Geographic or organizational constraints on friendships.

So, although online buddies would possibly share not unusual pastimes, a priori there is no cause to consider that two users with not unusual hobbies are more likely to be friends. Using statistics from LiveJournal, the authors show that the answer to each questions is yes.

Title: Cybercrime detection in online communications: The experimental case of cyber bullying detection in the Twitter network.

Author: M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana.

The recognition of online social networks has created huge social conversation among their customers and this results in a big quantity of user-generated conversation data.

In current years, Cyber bullying has grown right into a major hassle with the growth of on-line conversation and social media. Cyber bullying has been recognized these days as a serious country wide health problem among on line social community customers and developi ng an efficient detection version holds high-quality practical significance. In this paper, we've proposed set of

unique capabilities derived from Twitter; community, activity, user, and tweet content, based on those feature, we evolved a supervised machine studying answer for detecting cyber bullying within the Twitter. An evaluation demonstrates that our advanced detection model primarily based on our proposed features, achieved outcomes with an area under the receiver-operating characteristic curve of 0.943 and an f-degree of 0.936. These effects indicate that the proposed model based on these functions presents a viable method to detecting Cyber bullying in online verbal exchange environments. Finally, we compare result obtained the use of our proposed features with the end result acquired from two baseline features. The comparison outcomes display the signific ance of the proposed functions.

Title: Using social media to predict the future: A systematic literature review.

Author: L. Phillips, C. Dowling, K. Shaffer, N. Hodas, and S. Volkova.

Social media (SM) records offers a vast record of humanity's everyday thoughts, feelings, and moves at a resolution formerly unimaginable. Because user behavior on SM is a mirrored image of events in the actual world, researchers have found out they can use SM if you want to forecast, making predictions approximately the future. The gain of SM statistics is its relative ease of acquisition, large quantity, and capacity to seize socially relevant informa tion, which may be tough to acquire from other information sources. Promising consequences exist across a wide variety of domains, but one will locate little consensus regarding fine practices in either method or evaluation In this systematic review, we examine applicable literature over the past decade, tabulate mixed results across some of medical disciplines, and identify common pitfalls and best practices. We locate that SM forecasting is limited by using records biases, noisy records, loss of generalizable outcomes, a loss of domain-specific theory, and underlying complexity in lots of prediction tasks. But in spite of these shortcomings, recurring findings and promising effects preserve to impress research ers and call for persevered investigation. Based on the prevailing literature, we identify studies practices which result in success, citing precise examples in each case and making guidelines for great practices. These tips will assist researchers take benefit of the exciting opportunities offered via SM platforms.

Title: Online social **networks & social network services: A technical survey.**

Author: H. Quan, J. Wu, and Y. Shi,

Social capabilities are natural consequences o f human societies. Before communication technologies, social capabilities generally tend to evolve inside cultural boundaries, such as vicinity and families. Communication technologies, from mountain top signaling to Voiceover-IP, have damaged the ones boundaries extra or much less and enabled multi-tradition social features. Empowered with the aid of low-cost, high-energy private computing Devices, the combined computing and networking talents have created a fertile ground for innovative types of social activities. Online social network (OSN) serves as a means of social pastime and has grow to

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

be a mainstream facts media in the industry and in the public. Both authorities and entrepreneurs recognized the fee of OSNs and have positioned forth.

Title: Is social media a gang? Toward a selection, facilitation, or enhancement explanation of cyber violence.

Author: J. K. Peterson and J. Densley.

This paper reviews the prevailing literature on the relationship between social media and violence, such as occurrence rates, typologies, and the overlap between cyber and in-individual violence. This assessment explores the individual-degree correlates and risk elements associated with cyber violence, the group strategies worried in cyber violence, and the macro-stage context of on line aggression. The paper concludes with a framework for reconciling conflicting degrees of clarification and gives a n schedule for destiny studies that adopts a selection, facilitation, or enhancement framework for considering the causal or contingent position of social media in violent offending. Remaining empirical questions and new directions for future research are discussed.

Title: Detecting illicit drugs on social media using automated social media intelligence analysis.

Author: P. A.Watters and N. Phair.

While social media is a new and thrilling technology, it has the capacity to be misused by prepared crime organizations and people in volved within the illicit tablets trade. In particular, social media gives a means to create new advertising and distribution opportunities to a worldwide marketplace, often exploiting jurisdictional gaps between consumer and seller. The sheer quantity of postings affords investigational barriers, however the platform is amenable to the partial Automation of open source intelligence. This paper affords a brand new technique for automating social media data, and provides two pilot research into its use for detecting advertising and distribution of illicit pills focused at Australians. Key technical challenges are identified, and the policy implications of the ease of get right of entry to to illicit drugs are discussed.

Title: Online social networks:Threats and solutions.

Author: M. Fire, R. Goldschmidt, and Y. Elovici.

Many on line social network (OSN) customers are blind to the numerous security risks that exist in those networks, together with privateness violations, identification thef t, and sexual harassment, just to name a few. According to recent studies, OSN users readily expose personal and private details about themselves, such as relationship status, date of birth, college name, e mail address, cellphone number, and even home address. This information, if positioned into the wrong hands, can be used to damage customers both inside the virtual world and within the actual world. These risks come to be even more extreme while the users are children. In this paper, we gift a thorough overview of the different security and privateness risks, which threaten the wellness of OSN users in general, and youngsters in particular. In addition, we gift a top level view of current solutions which could provide higher protection, safety, and privacy for OSN users. We also offer simple-to-implement tips for OSN customers, that could improve their safety and privateness wh

en the usage of these platforms. Furthermore, we suggest destiny studies directions.

Title: Security against sybil attack in social network

Author: N. M. Shekokar and K. B. Kansara.

Sybil (fake) accounts penetrate the OSN security by using hosting more than one threats. Multiple social-graph-based defence schemes have been proposed till date which can effectively hit upon and isolate the sybils based totally on the ground truth of limited social connections between non-sybil (honest) and sybil users. In actual global scenario, sybils can also elude these defenses by imploring many social connections to real users. This situation may also degrade the overall performance of social graph based sybil detection schemes.

So, current social graph based totally sybil identity further calls for new solutions. A new technique could be extra convincing if it enriches the shape of a social graph with more records about the behavioural analysis of its users. In this paper, we have proposed a sybil node identity (SNI) method based on classical social graph based totally sybil detection state of affairs and SNI-B that is the extension to SNI by means of incorporating person behavioural aspects. Our Simulation results show that social graph based totally SNI outperform if combined with behavioural aspects (SNI-B), of customers even in case of implored sybil connections over the time.

Title: Detecting and tracking political abuse in social media.

Author: J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer.

We have a look at astroturf political campaigns on microblogging platforms: politically-motivated individuals and organiza-tions that use a couple of centrally-managed bills to create the appearance of tremendous help for a candidate or opin-ion. We describe a machine studying framework that com-bines topological, content-based and crowdsourced capabilities of information diff usion networks on Twitter to stumble on the early levels of viral spreading of political misinformation.We present promising preliminary consequences with higher than 96 Acuracy inside the detection of astroturf content inside the run-as much as the 2010 U.S. Midterm elections.

Title: PhishAri: Automatic realtime phishing detection on Twitter.

Author: A. Aggarwal, A. Rajadesingan, and P. Kumaraguru.

With the arrival of on line social media, phishers have commenced using social networks like Twitter, Facebook, and Foursquare to unfold phishing scams. Twitter is an immensely famous micro-blogging network where humans post short m essages of one hundred forty characters known as tweets. It has over one hundred million active users who post about 200 million tweets everyday. Phishers have started the usage of Twitter as a medium to spread phishing due to this vast statistics dissemination. Further, it is difficult to stumble on phishing on Twitter not like emails because of the short spread of phishing links in

the network, brief size of the content, and use of URL obfuscation to shorten the URL. Our technique, PhishAri, detects phishing on Twitter in realtime. We use Twitter specific functions along side URL functions to stumble on whether or not a tweet posted with a URL is phishing or not. Some of the Twitter specific features we use are tweet content material and its characteristics like length, hashtags, and mentions. Other Twitter functions used are the characteristics of the Twitter user posting the tweet together with age of the account, quantity of tweets, and the follower-followee ratio. These twitter specific features coupled with URL based capabilities prove to be a strong mechanism to discover phishing tweets. We use device getting to know classification strategies and hit upon phishing tweets with an accuracy of 92.52%. We have deployed our system for end-customers by supplying an clean to use Chrome browser extension. The extension works in realtime and classifies a tweet as phishing or safe. In this research, we display that we are capable of come across phishing tweets at zero hour with excessive accuracy that's much quicker t han public blacklists and as well as Twitter's very own defense mechanism to discover malicious content. We also executed a short user assessment of PhishAri in a laboratory take a look at to evaluate the usability and effectiveness of PhishAri and confirmed that customers like and find it convenient to apply PhishAri in real-world. To the satisfactory of o.

Title: Detecting spam in a Twitter network.
Author: S. Yardi *et al.*

Spam turns into a hassle as quickly as an online verbal exchange medium becomes popular. Twitter's behavioral and structural properties make it a fertile breeding ground for spammers to proliferate. In this text we examine junk mail round a one–time Twitter meme — "robot pickuplines". We display the life of structural community variations between spa m debts and legitimate users. We conclude by means of highlighting demanding situations in disambiguating spammers from legitimate users.

Title: Analyzing Spammer's Social Networks for Fun and Profit.
Author: C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu.

In this paper, we carry out an empirical analysis of the cyber crook atmosphere on Twitter. Essentially, throughreading inner social relationships inside the criminal account community, we discover that crook debts generally tend tobe socially connected, forming a small-international network.
Wealso find that criminal hubs, sitting within the center of the social graph, are more inclined to follow criminal bills.Throug h analyzing outer social relationships between criminal debts and their social friends out of doors the criminal account community, we reveal three categories of money owed that have close friendships with criminal accounts.Through those analyse s, we provide a singular and effective crook account inference algorithm by exploiting criminalmoney owed' social relationships and semantic coordinations.

SYSTEM ANALYSIS

Existing System:

- State-of-the-art research has developed features to improve the performance of cyberbullying

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

prediction. For example, a lexical syntactic feature has been proposed to deal with the prediction of offensive language; this method is better than traditional learning-based approaches in terms of precision. Dadvar *et al.* examined gender information from profile information and developed a gender-based approach for cyberbullying prediction by using datasets from Myspace as a basis. The gender feature was selected to improve the discrimination capability of a classifier. Age and gender were included as features in other studies, but these features are limited to the information provided by users in their online profiles.

- Several studies focused on cyberbullying prediction based on profane words as a feature. Similarly, a lexicon of profane words was constructed to indicate bullying, and these words were used as features for input to machine learning algorithms. Using profane words as features demonstrates a significant improvement in model performance. For example, the number of ``bad" words and the density of ``bad" words were proposed as features for input to machine learning in a previous work. The study concluded that the percentage of ``bad" words in a text is indicative of cyberbullying. Another research expanded a list of pre-defined profane words and allocated different weights to create bullying features. These features were concatenated with bag-of-words and latent semantic features and used as a feature input for a machine learning algorithm.

Disadvantages

- The System is not much affective due to Semi supervised machine learning techniques.
- The system doesn't have sentiment classification for cyberbullying.

Proposed System:

- The proposed system is constructing cyberbullying prediction models is to use a text classification approach that involves the construction of machine learning classifiers from labeled 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 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

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

set of features that are extracted and engineered.

Advantages

- 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.

IMPLEMENTATION

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 friends request and responses, Add and View Filters, View all posts, Detect Cyber Bullying Users, Find Cyber Bullying Reviews Chart.

Viewing and Authorizing Users

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.

Viewing all Friends Request and Response

In this module, the admin can see all the friends' requests and response history. Details such as Requested User Name and Image, and Requested to User Name and Image, status and date.

Add and View Filters

In this module, the admin can add filters (like Violence, Vulgar, Offensive, Hate, and Sexual) as Categories with the words those related to corresponding filters.

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 corresponding post like Violence (no. of words belongs to Violence Filter used in comments by the users), Vulgar (no. of words belongs to Vulgar Filter used in comments by the users), Offensive (no. of words belongs to Offensive Filter used in comments by the users), Hate (no. of words belongs to Hate Filter used in comments by the users), Sexual (no. of words belongs to Sexual Filter used in comments by the users).

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 viewing their profile details, searching for friends and sending friend requests, Posting Your Messages as Posts by giving details, View and Comment on Friend Posts, viewing all friends posts and comment, view all your cyber bullying comments on your friend posts.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Viewing Profile Details, Search and Request Friends

In this module, the user can see their own profile details, such as their address, email, mobile number, profile Image.

The user can search for friends and can send friend requests or can accept friend requests.

Add Posts

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

View and Comment on Your Friends Post

In this, the user can see his entire friend's post details (post title, description, uses, creator and image of post) and can comment on posts.

View all Friends Posts and Comment (Cyber bullying Related)

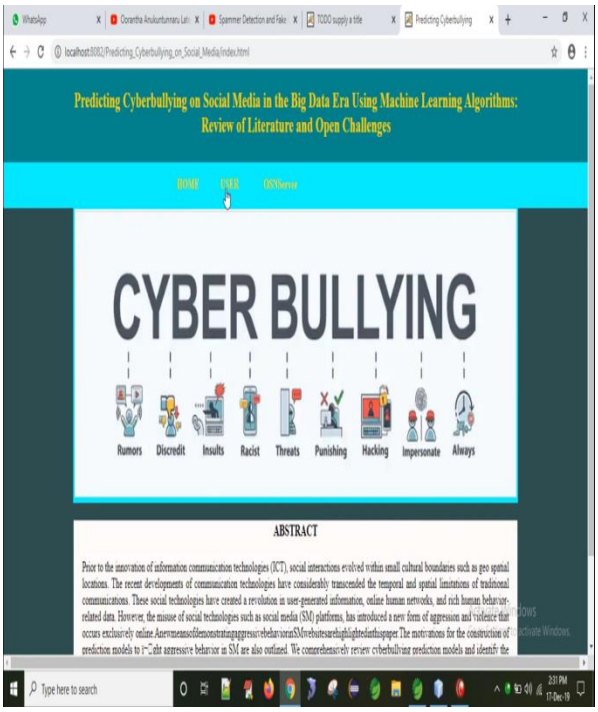
In this, the user can see his all friend's post details (post title, description, uses, creator and image of post) and can comment on posts. Don't Post If the comment consists of Cyber bullying words and Shows the reason why comment is not posted by indicating Detected Cyber Bullying Words like Numbers 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,

View all Your Cyber bullying comments on your friend posts

The user can see all his posted cyber bullying comments on their friend created posts.

SCREEN SHOTS

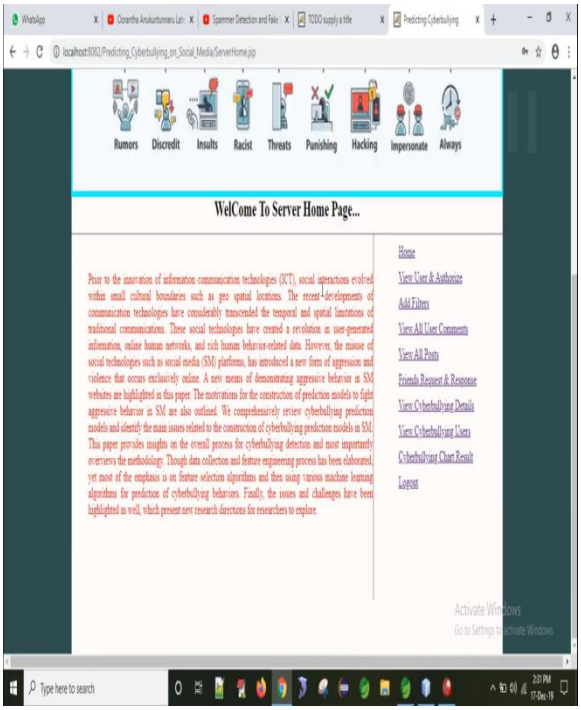
Home screen



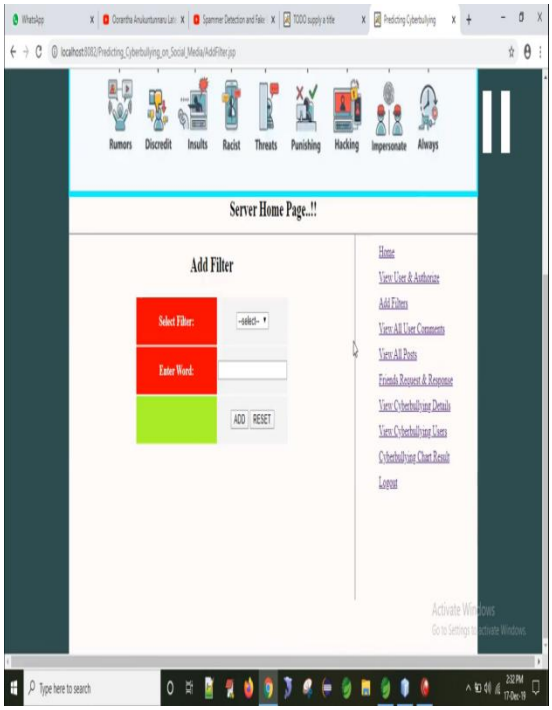
User login screen

Server login screen

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

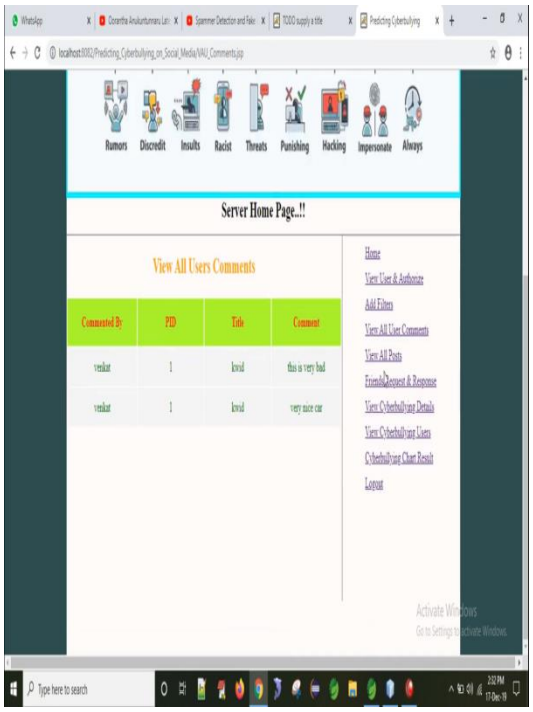


View users and authorize



View all users comments

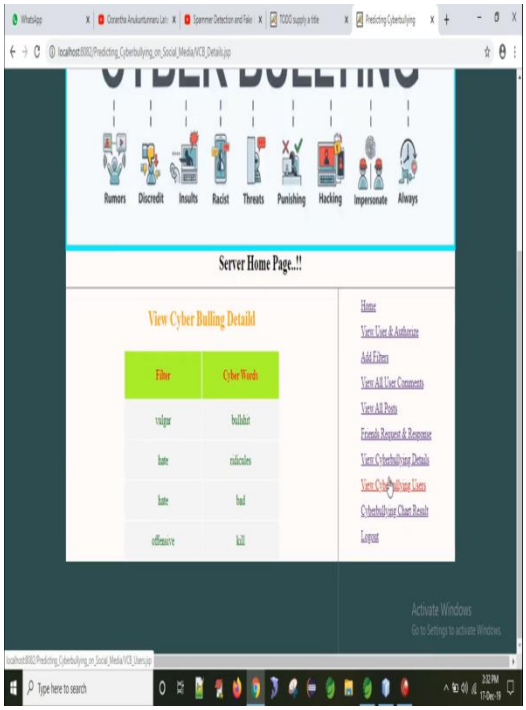
Add filter



View all posts

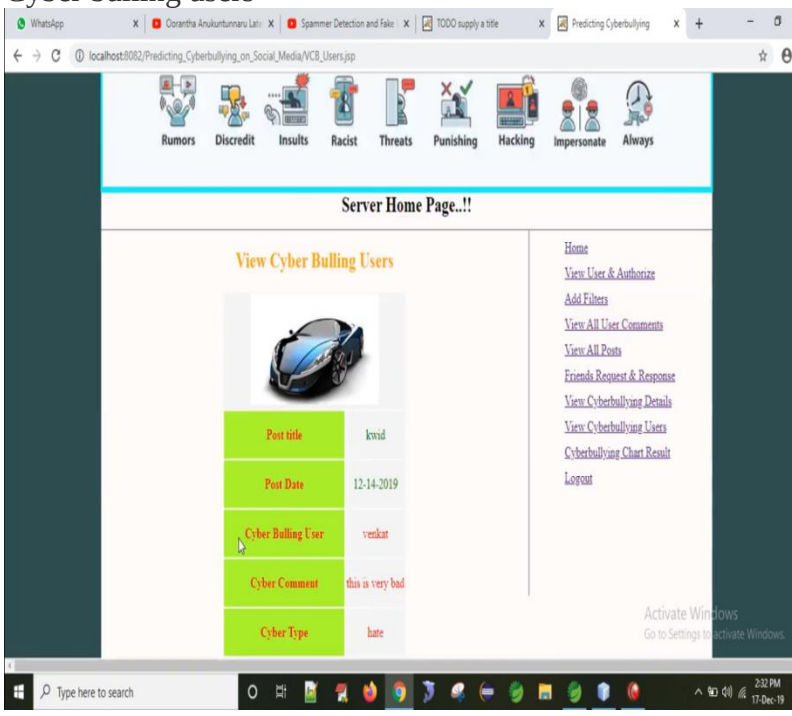
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

View all friends request and response



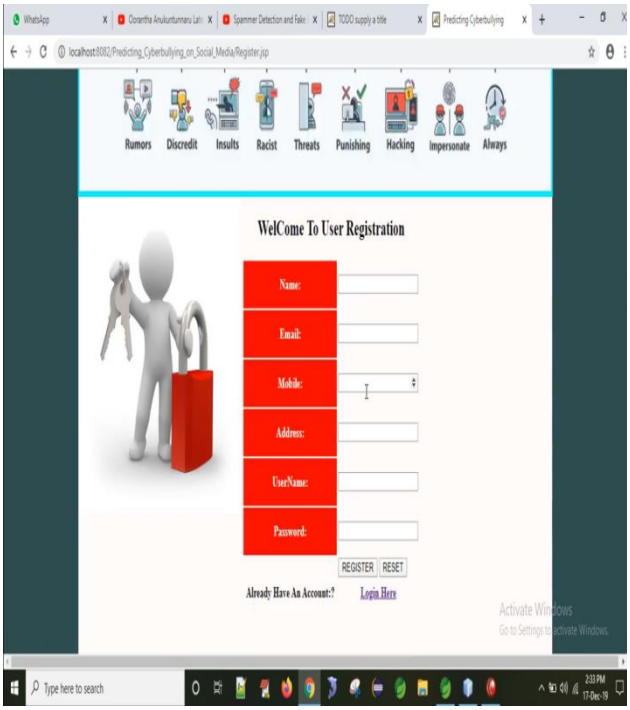
Cyber bulling users

Cyber bulling details



User Registration screen

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362



User home screen

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 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 modeling the behaviors 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

[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.