JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ACTIVE CHAT MONITORING AND SUSPICIOUS CHAT DETECTION

[1] Devendra Kumar Singh, [2] Binit Kumar Yadav, [3] Dr. Pankaj Kumar
[1] Student , [2] Student , [3] Associate Professor
[1,2,3]CSE Department,
[1,2,3]SRMCEM, Lucknow, India
[1]devendrasingh71022@gmail.com, [2]bkyadav2806@gmail.com, [3]pk79jan@gmail.com

Abstract: In the research of online social networks, detection of anonymous user behavior, detection of offensive data, etc. are traditional and important research works. This project is focused on the detection of offensive data, and bully statements in shared data of the social networks. For this system, using Machine Learning algorithms with Text Mining concepts predicted the offensive data to get more accurate results. This project proposed a system of "Chat Detection (CD) in Social Networking" for predicting bully data. This project is used two datasets, namely, 'Hate Speech and Offensive Language Dataset' and 'Harassment-Corpus Dataset'. This project used three Machine Learning classifiers such as Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), and Neural Network (NN) Algorithms and calculated performance results for comparing the performance for both datasets. To demonstrate this system designed and developed a Python-based Django web application and showed the results.

Keywords— Suspicious Chat detection; Chat Monitoring ; Machine Learning(ML); Algorithms; Naïve Bayes; Neural Network

I. INTRODUCTION

Most chat applications in the internet like WhatsApp, Messenger or other social networking apps offering chat communication tools for text messages, media data sharing, web data sharing etc. In current trend many social networking sites created and providing services of communications, multi-media services, e-commerce etc immensely. For example twitter social media provide major services of micro-blogging massively, it has more than 700 million users and 400 million micro-blogs produces per day. According to research survey many more than 30% of dummy or duplicate or fake accounts are present in all social media services like twitter, facebook, etc [1]. But in the current social sites not focus on services like tracking the user behavior of anonymous behavior. In current system, social network sites need to focus the user microblogs and need to capture the user behavior whether his/she anonymous user or not. Few surveys' providing concepts to tracking the attackers like using profile matching techniques and network based techniques etc. But in real-time to apply those concepts in social network is less practical. Crawling the user information from the user micro blogs is also less practical. Anonymous Users can easily manipulate the public profile information. In social networks user may share their messages by using the chat applications. For every social networking sites has their own chat applications, for this facebook is main example. And another way is sharing the multimedia data like images or videos. For this Facebook and Instagram best examples. For communication between users chat applications will most useful for share their information, thoughts, views etc. But in the same way it may also cause the security loophole of user's security which is cyber bullying. Such text based content may security threat to the users because of the people can share cyber bullying words to the users with their fake accounts. [2], [3]. Based on these disadvantages detection malicious users is active topic in the study of social media

II. LITERATURE SURVEY

Except many blessings and convenient of superior technology, there arises many crimes Like cyber bullying, fraud, and many extra. It is vital to save you criminals from the Beginning and is vital to invent machine that might music suspects at any cost. Murugesan et Al. [1] (2016) have used statistical corpus based facts mining approach for the detection of Suspicious sports on online boards. The writer used the concept of forestall phrases elimination And stemming process so that any suspicious textual content will become clearer and easy to understand .The technique changed into to suit the key phrases with suspicious phrases through the use of matching Set of rules. On this manner the suspicious key words can be identified. In the end authors have used

The keyword spotting strategies, leaning based totally approach and hybrid of described tactics forThe general popularity of suspicious human activity. Tayal et al. [2] (2015) have proposed the Approach of crime detection and crook identification (cdci) using statistics mining method.The writer made modules like records extraction (de) in order to extract all unstructured crime Records from web resources. Statistics processing (dp) a good way to lessen the ones extracted data into Based crime facts. Different modules like clustering, google map, category and waka Implementation are useful for crime detection, crook identification and crime verification Respectively. The writer has used k-method clustering set of rules for reading crime detection And google map improves visualization to okay-manner. Knn category is used to research Crook identity and prediction. The verification of criminals is achieved with the help of Waka. The main motive for creator to expand this kind of device become to lessen crimes And for the betterment of the society. It also helps investigating businesses in crime detection And identification of criminals. So, this system has helped in reducing the crime charge in the With the development of chat packages, ic has grow to be day after day application for all people and It is one of the today's chat in which customers can send messages, chat, documents and so on. Yahoo chat is A unfastened on-line chat room carrier for most effective yahoo customers. It presents customers to create public chat And send messages. In addition, skype is likewise an immediate messaging app that gives online Textual content messaging and video offerings. It allows change of digital files. However, with Convenience of such apps come the unlawful activities inclusive of increasing terrorist via spreading The message and influencing the innocence, threatening via message, fraud and etc. In Order to reduce such activities and enhance the advantages of immediately chat programs, Packages together with mspy that's the parental control app for clever telephones that permits Mother and father to reveal textual content messages, calls, present day gps vicinity, snapchat, whatsapp and much More. It collects statistics from the tool on which it is established and show it inside the manage Panel which we will get admission to using any net browser. Flexispy is software program that we set up Cellular cellphone. It secretly statistics activities that manifest on the smartphone and can provide this Statistics to an internet account, where we will view these reports 24 x 7 from any internet Enabled computer or cell telephone. The truthspy works at once at the goal tool and Gives huge spectrum of statistics to the person. On line social networking web sites like fb And google are the splendid and maximum famous manner to preserve in contact with our loved ones and Associates. However, such social networking may be used for the motive of unlawful Sports which includes spreading hate comments which may placed the victim into despair and Making plans terrorism associated works. So, kumar and singh (2013) [3] have proposed a machine For the detection of suspicious customers based totally on their sentiments over chat conversations and Remarks exchanged on on-line social networking websites and chat messangers. The proposed gadget also identifies the cluster of people replacing comments approximately same Topics which may additionally result in suspicious activities. The main objective of their system is to Examine the chat messages in social networking websites and to perceive institution of suspicious customers Involved in suspicious activities. Data mining is one of the maximum powerful methods of extracting Useful statistics or best approach to come across underlying relationships amongst statistics the usage of System getting to know and artificial intelligence. Now with the development of era, Criminals' charge is increasing as they may be also adapting smarter ways to devote crimes. So, Statistics mining is powerful device to discover the activities of criminals by means of analyzing thir document and Information, subsequently prevents the crimes in destiny. Hosseinkhani et al (2014) [4] have proposed A gadget for the present concepts of crime facts mining techniques for the detection of Suspicious data on the net. In keeping with him, the most important challenge for the detection of Suspicious activities are the daily growth in number of cyber statistics, common online Transactions and busy community visitors which results to massive quantity of illegal activities. He Delivered concept such as internet mining, criminal identities and crime facts mining Strategies. John resig ankur teredesai [5] have proposed a framework for the analysis of Huge scale communique media popularly called im (instant messaging) and diverse Statistics mining troubles and the way those relate to immediately messaging and terrorism activities. They Also focused on person pattern evaluation, restrained message length and anomaly detection. Their Paintings does no longer tend to absolutely detect suspicious messages now not even detection of topics. M. Brindha et al [6] have proposed a machine to reveal energetic chat and detect suspicious chat Over internet. The proposed system analyzes on line undeniable text from decided on discussion organization And classifies the textual content into distinct organization and machine will decide whether or not the text are normal or suspicious. The device evolved is chat system and is based on purchaser server.

III. OBJECTIVES

Active chat monitoring and suspicious chat detection is a chat system where suspicious users are identified by determining the keywords used by him/her. The main purpose of this project is to develop software that can be used to find a system that identifies deception in messages through communication. The system is designed in a way that the users can easily interact with the system with minimum knowledge to browse the internet. It detects what kinds of information are being passed from the users and can detect their whole conversation without their notice. It can be used along with most widely used social sites like face book, twitter, etc. to provide security against cyber crime. By the use of this system, one can be saved from involving in illegal activities and from being a cyber victim.

IV. PROPOSED SYSTEM

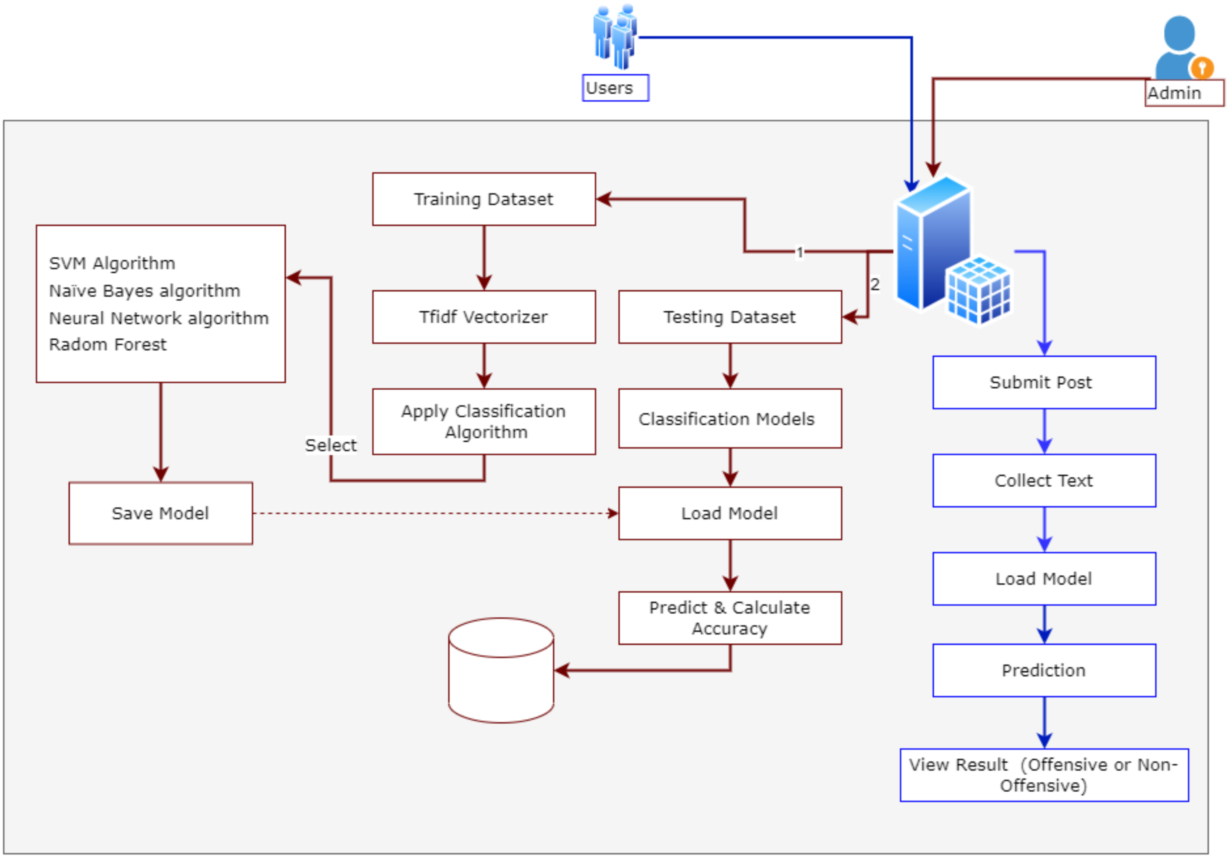
The proposed System will analyze online plain text sources from selected discussion forums and will classify the text into different groups and system will decide which post is legal and illegal. This system will ensures that the admin may not watch all the chat at a time, so in order to stop chatting illegally, the keywords are set by the admin as suspicious words which will be blocked or it cannot be able to view by the other person. So, in order to prevent illegal or suspicious chatting, admin sets the keywords as suspicious words which will be deleted as an alert will be sent to the admin.

V. PREDICTION ALGORITHMS

- We choose to use neural networks algorithm to obtain correct chat detection..
- a neural network is a chain of algorithms that endeavors to recognize underlying relationships in a fixed of statistics via a manner that mimics the manner the human brain operates.
- A neural community will encompass 3 kinds of layers:
  1. The enter layer: it gets all the inputs and the closing layer is the output layer which gives the desired output
  2. The hidden layer: all of the layers in among these layers are called hidden layers. There can be some of hidden layers. The hidden layers and perceptron in every layer will rely upon the use-case you are attempting to resolve.
  3. The output layer: it gives the preferred output. Upon which all the enter is received and it forwarded to the following layer.
- Depending at the vicinity of the weather station, it offers historic dataset as well as modern-day dataset.
- We are able to be imposing neural networks algorithms on the to be had datasets and we will be also analyzing the prediction traits on extraordinary types of datasets. Maximum information units were obtained from kaggle.

VI. ARCHITECTURE

This architecture will describe the online social user process of the chat procedure and detection of anonymous users, in this flow there is a manual procedure and another one is system procedure in the detection process.



VII. CONCLUSION

In current trend many social networking sites created and providing services of communications, multi-media services, e-commerce etc immensely. Lot of anonymous users accounts are creating very rapidly. We need to focus for the tracking the anonymous users. In our project we have calculated the user behavior according to the chat statements of the user which he/she do with others. By the taking the advantages Machine Learning algorithms we classify the anonymous users. Here we are using Naïve Bayes algorithm to perform the classification of the users.

VIII. Acknowledgment

We would like to thank all the working staff of **Computer Science and Engineering Department, SRMCEM, Lucknow** and a great thank to our project guide and Coordinator **Dr. Pankaj Kumar** who gave her best to make us do this great project. Moreover, thanks to all the authors whose papers and books we referred during this project. Without help of all these resources, we would have never completed this project.

IX. References

[1] Murugesan, M. Suruthi, R. Pavitha Devi, S. Deepthi, V. Sri Lavanya, and Annie Princy. "Automated Monitoring Suspicious Discussions on Online Forums Using Data Mining Statistical Corpus Based Approach." Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on, pp. 2015-2020. IEEE, 2016. Imperial Journal of Interdisciplinary Research [8].

[2] Tayal, Devendra Kumar, Arti Jain, Surbhi Arora, Surbhi Agarwal, Tushar Gupta, and Nikhil Tyagi. "Crime detection and criminal identification in India using data mining techniques." 2, no. 5 (2016)

[3]] J. Hosseinkhani, "Detecting suspicion information on the Web using crime data mining techniques," International Journal of Advanced Computer Science and Information Technology, vol. 3, pp. 32-41, 2014.

[4] John Resig Ankur Teredesai, "Data Mining Research Group", Department of Computer Science, Rochester nstitute of Technology, {jer5513,amt}@cs.rit.edu.

[5] M.F. Porter, (1980) "An algorithm for suffix stripping", Program, Vol. 14 Issue: 3, pp.130-137".

[6] M. Brindha1 , V. Vishnupriya2 , S. Rohini3 , M. Udhayamoorthi4 , K.S.Mohan5 , "Active Chat Monitoring and Suspicious Chat Detection over Internet", 1,2,3UG Scholars, Deparment of IT, SNS College of Technology, Coimbatore, Tamilnadu, India. 4,5Assistant Professor, Department of IT, SNS College of Technology.

[7] Ms. Pooja S. Kade1, Prof. N.M. Dhande, " A Paper on Web Data Segmentation for Terrorism Detection using Named Entity Recognition Technique" presented at International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056,Volume: 04 Issue: 01 | Jan -2017.

[8] T.Bhaskar, —"Fast identification of stop words for font learning and keyword spotting".