Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2022, 9(7):7-14

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Passive Chat Monitoring and Suspicious Chat Detection using Online Data Mining

Onyemenam G.U.*, Okengwu U.A

Center for Information and Telecommunications Engineering (CITE), University of Port Harcourt, Nigeria
georgeugochukwu92@gmail.com and ugochi.okengwu@uniport.edu.ng

Abstract In this research, passive chat monitoring and suspicious chat detection system using online data mining is achieved. To reduce cyberbullying and save users from being victims, our proposed system autonomously monitors and flags any suspicious chat in plain text by notifying victims to take action and warning perpetrators. This provides real-time monitoring and detection technology for the users without the need for them initiating it themselves, the model was developed and deployed using FASTAPI so that it can be integrated on any chat platform and works autonomously which reduces the vulnerabilities inherent with existing systems. We used agile software development methodology developing the system and utilized python programming language to train and build the model, while we built the chat application for testing the performance of the proposed system using flutter (dart). After building and testing the model with different algorithms, the Support Vector Classifier algorithm produced over 90 percent suspicious chat detection on the various parameter matrix used and as such was used to build the system.

Keywords Cyberbullying, child pornography, cybercrime, Cybersecurity

1. Introduction

The Chartered Trading Standards Institute (CTSI) has been warning people about the fraud that involves a scammer posing as someone you know. It showed “A member of the multitude named Alison accepted a communication on the known messaging platform WhatsApp: Hi mum, I have fell my phone down the latrine, this is my new digit.” Alison responded to the text and questioned if it was child, Will, to which the scammer answered in the affirmative. The veritably coming morning, Alison’s child contacted her asking for over two thousand pounds and clarified that he had gotten involved with loan harpies and was demanded to pay up. Alison did not misdoubt the text for an instant, fortunately, Alison ascertained that this was a fiddle before it was too late.

A lot of people have been victims of cybercrime over the 11 years ago, and these prey cut across all the genders, teenagers and grown-ups.

Detest oration online has been associated to a global rise in vehemence toward nonages, including mass blowups, lynching and ethnical sanctification. Polices capitalized to check detestation declamation threat demarcating autonomous speech and are inconsistently applied. Nations like the United States empower the social media enterprise extensive dominions in managing their news and administering hate declamation ground rules, others, including Germany, can coerce establishments to put off posts within unidentified moments.

On five June 2021, the Nigerian administrator formally located a fathomless embargo on Twitter, confining it from acting it sports in Nigeria after the social media platform deleted tweets placed up with the aid of using the Nigerian President Muhammadu Buhari, cautioning the southeastern human beings of Nigeria of a capacity implicit reprise of the 1967 Biafran Civil War because of a perennial insurgency in Southeastern Nigeria (Wikipedia).

The Nigerian federal government has apportioned just over US$11 million (4.8 billion Nigerian Naira) to the nation's National Intelligence Agency (NIA) to spy on social messaging service whatsapp, as President Muhammadu Buhari's government continues its rush to acquire grip of social forums in the nation (Williams 2021). The allotment is allocated for the WhatsApp intercept Result and for Thuraya Interception Solution a dispatches design used for capturing audio calls of call-related information, SMS, data traffic, among others (Williams 2021).

At the moment, much problem is connected to an increase in several concerns, including cyberbullying, cyber fraud, terrorism and security. Therefore, it is imperative to provide a system that is capable of monitoring, detecting and preventing suspicious online conversations that may lead to cybercrime through social media chatting applications.

2. Review of Related Literature

Murugesan et al. (2016) have used statistical corpus based totally definitely statistics mining approach for the detection of suspicious sports activities on on-line forums. The writer used the concept of save you terms removal and stemming process just so any suspicious text will become clearer and easy to understand. The approach grow to be to healthful the important thing terms with suspicious terms thru the use of matching algorithm. Following this way the suspicious key terms can be identified. Lastly authors have used the important thing-phrase spotting approach, mastering based totally definitely method and hybrid of defined strategies for the overall identification of suspicious human occasion.

Michael R. et al., (2010) defined approximately the social method to come across malicious content material in internet technology for Facebook with protection heuristics is constrained to pick out malicious internet site links. Detection of suspicious e-mails from desk bound textual content the usage of choice tree induction proposed that is surely depending on the best facts entropy that acknowledges the messages are misleading or non-misleading Appavu, et al. (2019). M. Brindha et al (2015), have proposed a device to show energetic chat and find out suspicious chat over internet. The proposed device analyzes on-line easy text from determined on talk company and classifies the text into awesome company and device will decide whether or not or now no longer the text are regular or suspicious. The device developed is a speak device and it is patron and server based totally definitely. Microsoft 365 consists of verbal exchange compliance that is an insider danger answer that facilitates reduce verbal exchange dangers via way of means of assisting you come across, capture, and act on irrelevant messages to your company in line with Microsoft, this is one of the predominant motives they appoint verbal exchange compliance in Microsoft 365. Designated reviewers can use Pre-described and custom regulations to test inner and outside communications for coverage matches. Scanned email, Microsoft Teams, Yammer, or third-celebration communications to your company may be investigated via way of means of the reviewers in different to take suitable movements to ensure they may be compliant with the requirements of your company message protection. The neighborhood evaluation of internet chat room messaging locations a useful resource venture at the intelligence network due to the time required to take a look at large non-stop chat periods. Chat rooms are used to talk about genuinely any subject, inclusive of pc hacking and bomb making, developing a digital sanctuary for criminals to collaborate Jasonb et al. Given the developing hobby in native land protection troubles, we've advanced a textual content class machine that develops a conceptprimarily based totally profile that concisely represents the subjects mentioned in a talk room via way of means of all members. We then talk this basis of chat profiling machine and illustrates the cap potential to selectively increase the everyday concept of database with new concept of importance Jasonb et al. According to Dem and Tshewang (2018), Active Chat Monitoring and Suspicious

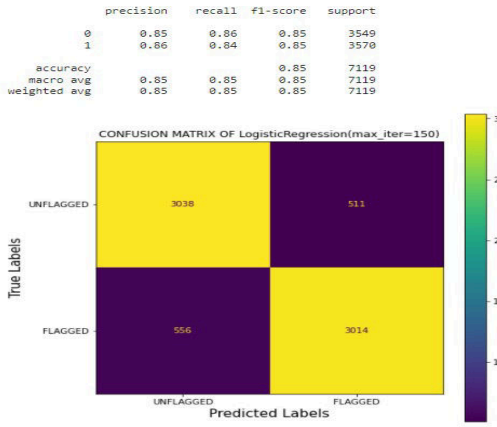
Chat Detection System is used for best the detection of undeniable textual content verbal exchange and does now no longer have the ability to evaluation voice chats or shared documents among users. It is prone to cyber frauds as it's miles not able to come across many suspicious undeniable texts. There can be extensive discount in cybercrime best if the framework proposed at server facet is integrated, . The machine is able to monitoring that phrase even though the announcement isn't cited in that manner. A technique of mechanically tracking textual content-primarily based totally communications of 1 or greater chat room or immediate messaging members to decide if a tracking occasion has occurred. Borden (2004). On-line verbal exchange are constantly monitored and enter to some of sample spotting modules, ideally running in parallel. Using recognize sample-popularity strategies, every sample spotting module can examine an issue of the communications via way of means of imposing positive algorithms and, as equally. A initial literature evaluate suggests that beyond research are mostly difficulty with server patron version that lets in the admin outline how the machine functions, together with specifying the suspicious key phrases, the motion taken as soon as a suspicious suit has been detected, the variety of admin employees offer 24x7 help etc. Limited development has been made on figuring out all of the numerous additives primarily based totally on their functionality, the supply of the key phrases for the machine, consequences for violators, time body for the admin to do so and the variety of admin required to offer 24x7 helps to the machine. The machine desires to be autonomous, it have to be capable of perform records mining pending whilst the admin is to be had to check all actionable occasions and dispose of mistakes which might be susceptible to human sentiment and different constraints that might introduce fraud into the machine.

3. Materials and Methods

It very important to have a clean and well processed data that the model can learn from, “garbage in garbage out”, whatever you feed the model with, is what the model will give out. 80% of this work is dedicated to the data processing and cleaning. After we have been able to achieve preprocessed data it is now time to train and build our model and compare results of our trained model using various probability and classifier algorithms. Our model result will be based on the accuracy, precision, recall, f1 score, and time (sec). The following probability and classifier algorithms were used to evaluate the accuracy, precision, recall, f1 score, and time (sec) in other to identify which algorithm produces the best result to test the model with.

1. **Logistic Regression**: With logistic regression we are to predict the probability of the model in terms of its accuracy, precision, recall, and the f1 scores of our model.
2. **Multinomial Naïve Bayes:** This is also a probalistic learning algorithm mostly employ in NLP. It is based on Bayes theorem and prediction.
3. **Support Vector Classifier SVC:** The aim of SVC is to create the best decision region that isolates or segregate the unflagged and flagged labels.
4. **Decision Trees Classifier:** These are built through an algorithm method that identifies ways to split a dataset based on variety of conditions.
5. **Random Forest Classifier:** Random Forest Classifier produces decision trees on various samples and takes their majority count for classification and finds the average results.

Figure 1 to 5 shows the result for probability and classifier algorithm.

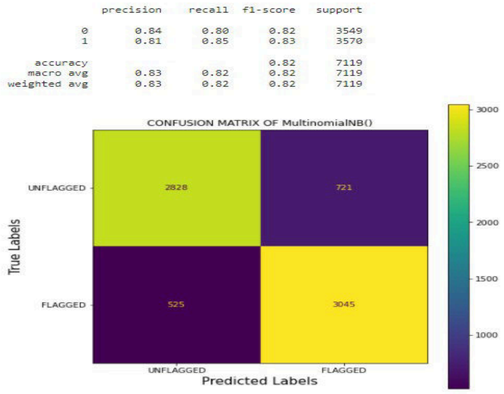


```
              precision    recall  f1-score   support

           0       0.85      0.86      0.85      3549
           1       0.86      0.84      0.85      3570

    accuracy                           0.85      7119
   macro avg       0.85      0.85      0.85      7119
weighted avg       0.85      0.85      0.85      7119
```

Figure 1: Logistic Regression

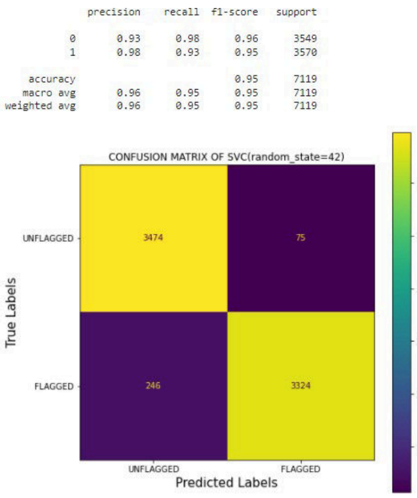


```
              precision    recall  f1-score   support

           0       0.84      0.80      0.82      3549
           1       0.81      0.85      0.83      3570

    accuracy                           0.82      7119
   macro avg       0.83      0.82      0.82      7119
weighted avg       0.83      0.82      0.82      7119
```

Figure 2: Multinomial Naïve Bayes

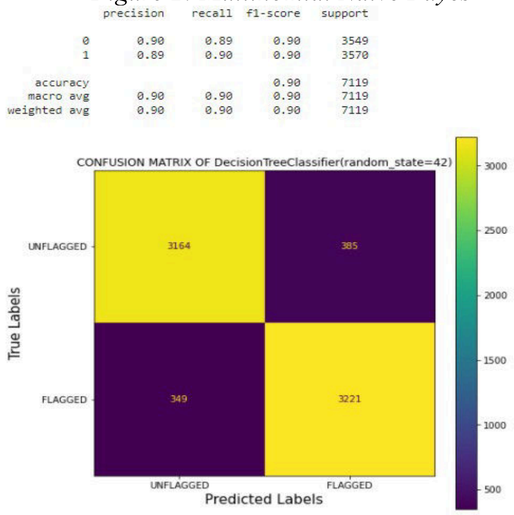


```
              precision    recall  f1-score   support

           0       0.93      0.98      0.96      3549
           1       0.98      0.93      0.95      3570

    accuracy                           0.95      7119
   macro avg       0.96      0.95      0.95      7119
weighted avg       0.96      0.95      0.95      7119
```

Figure 3: Support Vector Classifier

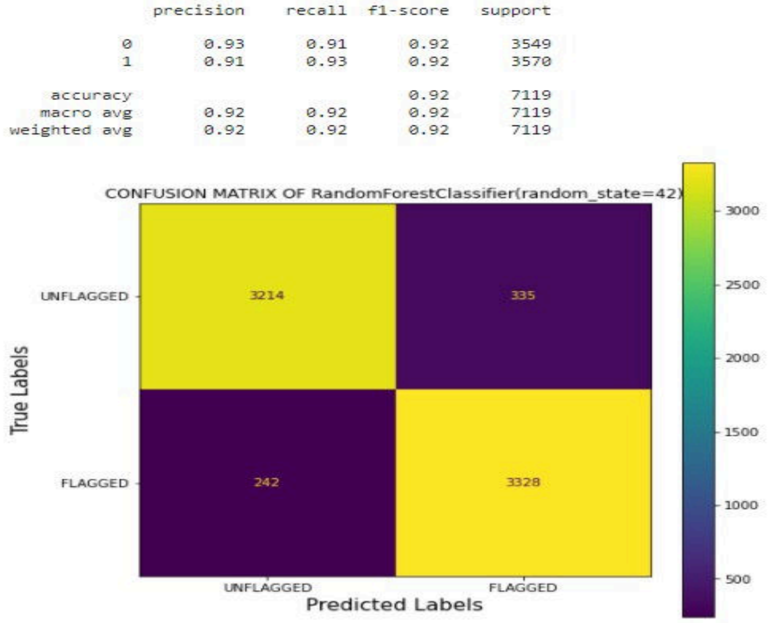


```
              precision    recall  f1-score   support

           0       0.90      0.89      0.90      3549
           1       0.89      0.90      0.90      3570

    accuracy                           0.90      7119
   macro avg       0.90      0.90      0.90      7119
weighted avg       0.90      0.90      0.90      7119
```

Figure 4: Decision Tree Classifier



```
              precision    recall  f1-score   support

           0       0.93      0.91      0.92      3549
           1       0.91      0.93      0.92      3570

    accuracy                           0.92      7119
   macro avg       0.92      0.92      0.92      7119
weighted avg       0.92      0.92      0.92      7119
```

Figure 5: Random Forest Classifier

4. Results and Discussion

It very important to have a clean and well processed data that the model can learn from, “garbage in garbage out”, whatever you feed the model with, is what the model will give out. 80% of this work is dedicated to the data processing and cleaning. After we have been able to achieve preprocessed data it is now time to train and build our model and compare results of our trained model using various probability and classifier algorithms. Our model result will be based on the accuracy, precision, recall, f1 score, and time (sec). The following probability and classifier algorithms were used to evaluate the accuracy, precision, recall, f1 score, and time (sec) in other to identify which algorithm produces the best result to test the model with.

1. **Logistic Regression**: With logistic regression we are to predict the probability of the model in terms of its accuracy, precision, recall, and the f1 scores of our model.
2. **Multinomial Naïve Bayes:** This is also a probalistic learning algorithm mostly employ in NLP. It is based on Bayes theorem and prediction.
3. **Support Vector Classifier SVC:** The aim of SVC is to create the best decision region that isolates or segregate the unflagged and flagged labels.
4. **Decision Trees Classifier:** These are built through an algorithm method that identifies ways to split a dataset based on variety of conditions.
5. **Random Forest Classifier:** Random Forest Classifier produces decision trees on various samples and takes their majority count for classification and finds the average results.

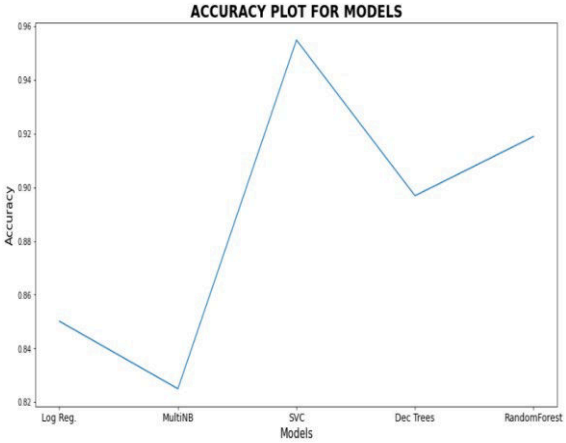


Figure 6: Accuracy Plot for the Model

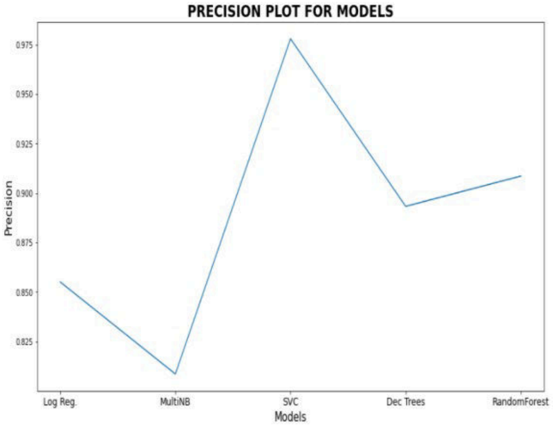


Figure 7: Precision Plot for the Model

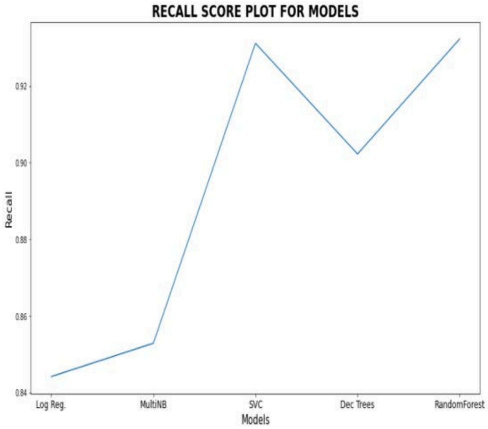


Figure 8: Recall Score Plot for the Model

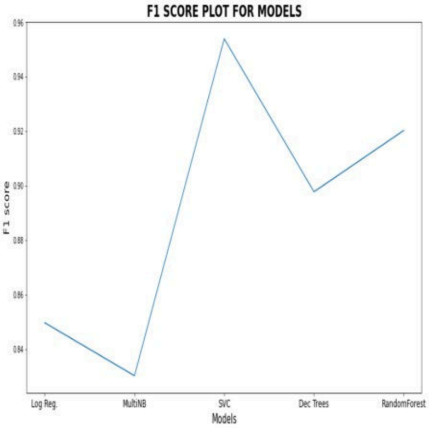


Figure 9: F1 Score for the Model

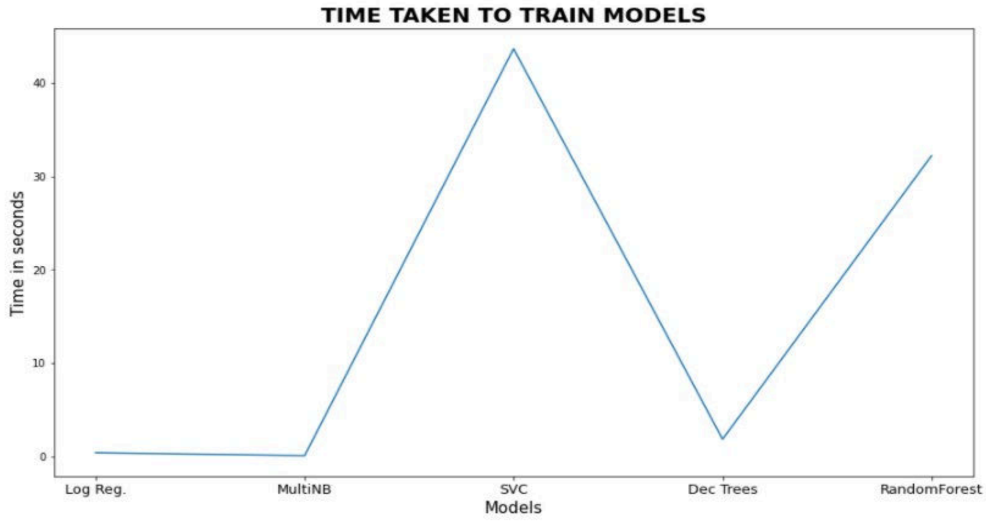


Figure 10: Time (sec) Plot Taken to Train the Model

Figure 6 to Figure 10 show the graph plots of the accuracy, precision, f1 score and time taken to train the model, the results as can been seen shows that Support Vector Classifier SVC produced the best result of over 90% for all the probability and classifier algorithms used. The cost of the result is on the time taken to train the model with SVC which is over 40 seconds unlike other algorithms that were far below 40 seconds.

5. Conclusion

With the increase number of deaths, mental cases and depression recorded on a daily bases worldwide and the amount of cybercrimes committed through online chat and web applications it has become paramount that developers, owners and stakeholders of these apps put in place technologies that is able to monitor and detect suspicious activities on their platform automatically and also notify users on a real-time bases so that they can be guarded and protected, which will in return boost the confidence of users using the application and on the other hand increase and sustain app owners users on their platform.

References

[1]. Alami, Salim, and Omar EL Beqqali (2015). Detecting Suspicious Profiles Using Text Analysis Within_Social Media. Published by JATIT.

[2]. Ali, Mohammed Mahmood, Khaja Moizuddin Mohammed, and Lakshmi Rajamani (2014). Framework for surveillance of instant messages in instant messengers and social networking sites using data mining and ontology," published by IEEE.

[3]. C. Serrano-Cinca (1996). Self organizing neural networks for financial diagnosis," Decision Support Systems, vol. 17, no. 3, pp. 227–238.

[4]. David W. Cheung, and et al. (1996). Maintenance of discovered association rules in largedatabases: an incremental updating technique," published by IEEE.

[5]. Daya C. Wimalasuriya and Dejing Dou (2010). Ontology-Based Information Extraction: An Introduction and a Survey of Approaches. Journal of Information Science, Volume 36, No. 3, pp. 306-323.

[6]. E. Aleskerov, B. Freisleben, & B. Rao (1997). *Cardwatch: A neural network based database mining system for credit card fraud detection.* In Proceedings of the 1997 IEEE/IAFE Conference on Computational Intelligence for Financial Engineering, CIFEr, pp.220–226, IEEE, March.

[7]. E. Duman & M. H. Ozcelik (2011). *Detecting credit card fraud by genetic algorithm and Scatter search*. Expert Systems with Applications, vol. 38, no. 10, pp. 13057–13063.

[8]. E. M. Carneiro, L. A. V. Dias, A. M. D. Cunha, & L. F. S. Mialaret (2015). *Cluster analysis and artificial neural networks: a case study in credit card fraud detection.* In Proceedings of the 12th International Conference on Information Technology: New Generations, ITNG 2015, pp. 122–126, IEEE.

[9]. Harsh Arora & Govind Murari Upadhyay (2015). *A Framework for the Detection of Suspicious Discussion on Online Forums using Integrated approach of Support Vector Machine and Particle Swarm Optimization.* Published by IJARCS.

[10]. Hosseinkhani, Javad, Mohammad Koochakzaei, Solmaz Keikhaee, & Javid Hosseinkhani Naniz (2015). *Detecting suspicion information on the Web using crime data mining techniques*. Published by IJARCS.

[11]. I. H. Witten, E. Frank, M. A. Hall, & C. J. Pal (2016). *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann. J. Friedman, T. Hastie, & R. Tibshirani (2001). *The Elements of Statistical Learning*, vol. 1, Springer, Berlin, Germany. J. Han, J. Pei, & M. Kamber (2011). *Data mining: concepts and techniques*, Elsevier.

[12]. J. Hosseinkhani. (2014). *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.

[13]. John Resig & Ankur Teredesai (2018). *Data Mining Research Group*, Department of Computer Science, Rochester Institute of Technology, {jer5513,amt}@cs.rit.edu.

[14]. J. Rollins & C. Wilson (2006). *Terrorist capabilities for cyberattack*. Overview and policy Issues. J.West & M. Bhattacharya (2016). *Intelligent financial fraud detection: a comprehensive review*. Computers & Security, vol. 57, pp. 47–66.

[15]. K. RamaKalyani & D. UmaDevi (2012). *Fraud detection of credit card payment system by genetic algorithm.* International Journal of Scientific & Engineering Research, vol. 3, no. 7, pp. 1–6.

[16]. Kumar, A.S. & Singh, S. (2013). *Detection of User Cluster with Suspicious Activity in Online Social Networking Sites*. Advanced Computing, Networking and Security (ADCONS), 2nd International Conference. 220,225, 15-17
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6714167&isnumber=6714118

[17]. M. Brindha, V. Vishnupriya, S. Rohini, M. Udhayamoorthi, & K.S. Mohan. (2018). *Active Chat Monitoring and Suspicious Chat Detection over Internet*.

[18]. M. Syeda, Y.-Q. Zhang, & Y. Pan (2002). *Parallel granular neural networks for fast credit card fraud detection.* In Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZIEEE '02), vol. 1, pp. 572–577, 2002.

[19]. M. Zanin, D. Papo, P. A. Sousa et al. (2016). *Combining complex networks and data mining: why and how.* Physics Reports, vol. 635, pp. 1–44.

[20]. Michael Robertson, Yin Pan, & Bo Yuan (2010). *A Social Approach to Security: Using Social Networks to help detect malicious web content*," published by IEEE.

[21]. Ms. Pooja S. Kade1, & Prof. N.M. Dhande (2017). *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.

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

[23]. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun (2011). *The application of data mining techniques in financial fraud detection*. A classification framework and an academic review of literature, Decision Support Systems, vol. 50, no. 3, pp. 559–569.

[24]. Murugesan, M. Suruthi, R. Pavitha Devi, S. Deepthi, V. Sri Lavanya, & Annie Princy. (2016). *Automated Monitoring Suspicious Discussions on Online Forums Using Data Mining Statistical Corpus Based Approach*. International Conference on, pp. 2015-2020. IEEE. Imperial Journal of Interdisciplinary Research.

[25]. Placida Tellis & N. Deepika (2015). *Expert System to Detect Suspicious Words in Online Messages for Intelligence Agency Using FP-growth Algorithm*," published by IJCSMC.

[26]. P. J. Bentley, J. Kim, G.-H. Jung, & J.-U. Choi (2000). *Fuzzy darwinian detection of credit card fraud. I*n Proceedings of the 14th Annual Fall Symposium of the Korean Information Processing Society, vol. 14, pp. 1–4.

[27]. Q. Lu & C. Ju (2011). *Research on credit card fraud detection model based on class weighted support vector machine.* Journal of Convergence Information Technology, vol. 6, no. 1, pp. 62–68.

[28]. R. Albert & A. Barab´asi (2002). *Statistical mechanics of complex networks*. Reviews of Modern Physics, vol. 74, no. 1, pp. 47–97.

[29]. R. Brause, T. Langsdorf, & M. Hepp (1999). *Neural data mining for credit card fraud detection.* In Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence, pp. 103–106, IEEE.

[30]. R. Patidar & L. Sharma (2011). *Credit card fraud detection using neural network*. International Journal of Soft Computing and Engineering (IJSCE), vol. 1.

[31]. Rob Kavet & Gabor Kenzo (2010). *A Perspective on Chat Associated with Suspicious Chat Technology.* Published by IEEE.

[32]. Shakil Ahmed, Frans Coenen, & Paul Leng (2012). *Treebased Partitioning of Data for Association Rule Mining.* Published by IEEE.

[33]. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, & D. W. Hwang (2006) *Complex networks: Structure and dynamics*. Physics Reports, vol. 424, no. 4-5, pp. 175–308.

[34]. S. Ghosh & D. L. Reilly (1994). *A tutorial on hidden markov models and selected applications In speech recognition.* In Proceedings of the 27th Hawaii International Conference on System Sciences: Information Systems: Decision Support and Knowledge-Based Systems, vol. 3, pp. 621–630.

[35]. S. H. Strogatz (2001). *Exploring complex networks.* Nature, vol. 410, no. 6825, pp. 268– 276.

[36]. S. Maes, K. Tuyls, B. Vanschoenwinkel, & B. Manderick (2002) *Credit card fraud detection Using Bayesian and neural networks*, In Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologies, pp. 261–270.

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

[38]. T. P. Bhatla, V. Prabhu, & A. Dua (2003). *Understanding credit card frauds*. Cards Business Review, vol. 1.

[39]. V. Zaslavsky & A. Strizhak (2006). *Credit card fraud detection using self-organizing maps.* The International Journal of Information & Security, vol. 18, pp. 48–63.

[40]. Y. G. Sahin & E. Duman (2011). "*Detecting credit card fraud by decision trees and support Vector machines.* In Proceedings of the International Multi-Conference of Engineering and Computer Statistics, vol. 1.