International Journal on Applications in Engineering and Technology
Volume 8 : Issue 2 : May 2022, pp 62 – 67 www.aetsjournal.com ISSN (Online) : 2455 - 0523

A MACHINE LEARNING BASED METHOD FOR DETECTING AND PROHIBITING OF CYBERBULLYING ON SOCIAL MEDIA PLATFORMS

BRINDA B.M[1] , FARITHA BEGUM N[2], SWARNA N[3], SWATHI S[4]

[1] *Assistant Professor, Department of Computer Science and Engineering, Paavai College of Engineering*

[2,3,4] *Undergraduate student, Department of Computer Science and Engineering, Paavai College of Engineering*

Abstract: *- –* Cyberbullying (CB) is becoming more common on social media platforms. With the popularity and broad use of social media by people of all ages, it is critical to make these platforms more secure from cyberbullying. Increased social networking, on the other hand, frequently has negative societal implications, leading to issues such as online abuse, harassment, cyberbullying, cybercrime, and trolling. SVM and Naive Bayes were used to find CB in a social media network. Its purpose is to construct and develop an effective technique for identifying online abusive and bullying statements using natural language processing and machine learning.

Key words: Cyberbullying, Social Media, Machine Learning, SVM, Naïve bayes, Online abuse, Harassment

1. INTRODUCTION

The wide emergence of social media leads to a considerable growth in the usage of the social networking sites. This promotes an increase in communication between millions of people without any barrier of distances. Nearly every people in the world have profile on any of the social networking sites like Twitter, Facebook, Linked ln, Google etc. The psychology behind this is that it is difficult to connect with people geographically but it is easier to connect digitally. The risk in social networking sites is entirely depending on the amount of information in which people share. It is clear that if the users share more information without considering the privacy and security then it may leads to a great vulnerability. The users of online social networks can create small virtual group inside the network. These virtual structures are called social network communities. Members of a social network community may not be known each other. They come under one cluster or community because of similar interest, opinions, views etc. In this scenario the trust is the major problem. Because there can be chance for spammers or anomalous people within the members of the group. Initially they may act as trusted users and make this trust as an opportunity to perform unacceptable activities which will affect risk factor of using social networks. So it is also important to detect spamming activities inside the social network communities. This study is all about different anomaly detection methods and spam detection in social networking sites. Social networks can be represented as a graph consists of vertices and edges where vertices or nodes are the users and edge shows the relationship between them. Among these nodes some may possess unusual behavior when compared to other nodes. These nodes are called anomalies or anomalous nodes. That is something that deviates from the standard behavior or normal expectation is the anomalies. Anomalous users refer to the people who are deviating from the normal user behavior. Initially the anomalous nodes behaves like a normal legitimate user but after gaining the trust and acceptability it starts performing unlawful activities which leads to serious security threats. Detection of anomalous activities is one of the key are as in the research of social network analysis. Irrelevant or uninvited messages sent over the Internet which aims to reach typically a large number of users, for the purposes of advertising are known as spam messages. Nowadays there is a considerable increase in

the growth of usage of SNS. The high click rate and the effective message propagation make social media as an attractive platform for spammers. Increase in spamming activities affects the people who are using social media adversely. So detection of anomalies and spam messages have equal importance in social network analysis. Most of the existing methods deal with the detection of anomalous users or spam nodes. But it is not a efficient method. Because the attackers can create multiple account and continue performing malicious activities.

1.1 Objectives of the project

- Support intended for educating: social network systems are improved to keep social gatherings and also improves causal astuteness for the new users so that this helps in expanding for education in social groups.
- Support given for individuals in a community: social network systems utilize every individual from community not only who are engaged with work but also informal organizers to enhance the network on training.
- Engaging through informal: inactive usage of new users also gives significant utilization inside the group and also criticizes the institutional administrations that gives assessment for moral concerns.
- Ease to access the information and presentations: The ease of use of various long ranges interpersonal communication administrations can offer points of interest to customers.
- Systematic boundary: A potential benefit of new organizations is usual crossing point pass through work/social boundaries. So such directions are frequently used in close to home boundary Interface by the manner in which the direction naturally points the line, along these lines preventing the making and boosting predictable are misused by the directors in expert committee. In the same manner the boundary for every user who wish to have same limits in work and social networks are illustrated as

1.2 Project Description

Irrelevant or unsolicited messages which send over the internet in order to destroy the normal communication are referred to as spamming messages. Usually spam messages are sent as bulk messages which target a large number of users. Spamming messages need not be sent by human individuals it can be generated from the third party tools such as machines. Because it is difficult for an intruder to manage multiple account and send these kinds of message. This is done for avoiding early detection of spammers. A large number or variety devices such as mobiles, laptops, tablet computers desktop computers comes into popular there by use of social networking is also increased. Spamming can be spread to new technologies rapidly. Micro-blogging services like twitter became a prominent platform for many activities like campaigning. Spam promoting campaigns are need to be detected.

II. SYSTEM ANALYSIS

2.1 Existing System

The concept of Digital Media Marketing has become very popular in the recent times mainly because of the increasing use of social media by more and more people day by day. With the growing usage of social media platforms, the digital media marketing importance has increased over the time. Hence, there are a number of marketing tools that helps the marketing agencies to target users and sell their products and services. There exists a number of applications that provides analysis of the social media usage. A good example of such a system is Google Analytics, Face book Insight and Audience Insights by Twitter. Google Analytics tracks down the activities of a website where as Face book or Twitter Analytic Tool use social science and computer science together to show the valuable insights gathered from stakeholders and use the same for business development decisions. Spam is a problem throughout the Internet, and Twitter is not immune. In addition, Twitter spam is much more successful compared to email spam. Various methods have been proposed by researchers to deal with Twitter spam, such as identifying spammers based on tweeting history or social attributes, detecting abnormal behavior, and classifying tweet-embedded URLs.

2.2 Proposed System

The proposed system aims at utilizing the data collected from the three of the most popular social media platforms that are Face book. The users would be classified into different Trending Keyword based categories. Using Deep Neural Network Algorithm Trend keyword classification occurs, then the user will targeted for Marketing based on Trend Result. The system aims to investigate the utility of linguistic features for detecting the spam twitter accounts and tweets. We take a supervised approach to the problem, but leverage existing hash tags in the Twitter data for building training data.

III. METHOD

3.1 SVM Algorithm

SVM (Support Vector Machine) is a common Supervised Learning technique for Classification and Regression. However, it is most commonly employed in Machine Learning for Classification issues.

The SVM algorithm's purpose is to find the optimal line or decision boundary that can divide n-dimensional space into classes so that fresh data points may be readily placed in the proper category in the future. A hyperplane denotes the optimal choice boundary.

The hyperplane is created using SVM, which selects the extreme points/vectors. Support vectors are the extreme situations, and the Support Vector Machine technique is named after them. Consider the figure below, which shows two different types of classifications.

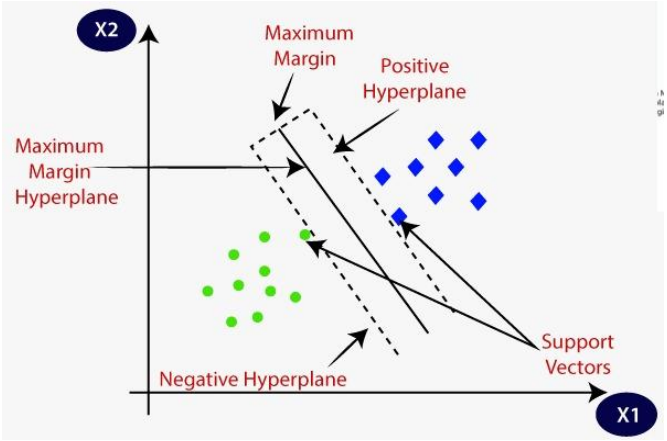

Figure 1: Support Vector Machine

3.2 NAIVE BAYES ALGORITHM

The Naive Bayes method is a supervised learning technique for addressing classification issues that is based on the Bayes theorem.

It is mostly utilised in text classification tasks that need a large training dataset.

The Nave Bayes Classifier is a simple and effective classification method that aids in the development of rapid machine learning models capable of making quick predictions.

It's a probabilistic classifier, which means it makes predictions based on an object's likelihood.

Spam filtration, sentiment analysis, and article classification are all common uses of the Nave Bayes Algorithm.

IV. TESTING TECHNIQUES

Software testing is a critical element of software quality assurance and represents the ultimate review of specification, design and code generation. At the time of developing the code and executing the code, he list of errors are identified (both syntax and semantic) and corrected. After the system designed, code is written, there is usually a procedure in place for testing the system for bugs, performance and reliability. System Testing is an important phase. Testing represents an interesting anomaly for the software. Thus a series of testing are performed for the proposed system before the system is ready for user acceptance testing. A good test case is one that has a high probability of finding an as undiscovered error. A successful test is one that uncovers an as undiscovered error.

4.1 BLACK BOX TESTING

Black box testing is defined as a testing technique in which functionality of the Application under Test (AUT) is tested without looking at the internal code structure, implementation details and knowledge of internal paths of the software. This type of testing is based entirely on software requirements and specifications. In Black Box Testing we just focus on inputs and output of the software system without bothering about internal knowledge of the software program.

4.2 WHITE BOX TESTING

White box testing is a software testing method in which the internal structure/design/implementation of the item being tested is known to the tester. The tester chooses inputs to exercise paths through the code and determines the appropriate outputs. Programming know-how and the implementation knowledge is essential. White box testing is testing beyond the user interface and into the nitty-gritty of a system.

4.3 UNIT TESTING

The goal of unit testing to separate each part of the program and test that the individual parts are working correctly and as intended. While performing unit tests, application code functions are executed in a test environment with sample input. The output obtained is then compared with the expected output for that input. If they match the test passes. If not it is a failure. Unit tests are great for confirming the correctness of the code. Let’s take a look at a sample algorithm that illustrates the concept.

V. SYSTEM IMPLEMENTATION

Implementation is the final stage of the project where the theoretical design is turned in to working design. It is the key stage in achieving successful system, since it involves much upheaval in the employee of the company. Implementation is carefully planned. The executive are trained fully about the calculation part of the system before use. The system test in implementation confirms that all is correct and shows the user that the system works. This involves training the end user in the office, system testing by the user and implementation. The term implementation has different meaning, ranging from the conversion of a basic application to a complete replacement of a computer system. Implementation is used here to mean the process of converting a new or a revised system design into an operational one.

5.1 PHYSICAL DESIGN

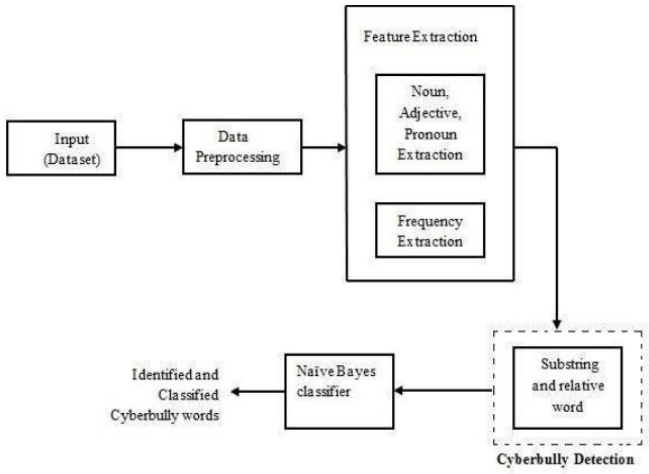

Figure -2: PHYSICAL DESIGN

VI. CONCLUSION

In this System, Classifier based approach is given to solve the detection of spam messages. A classification model is mainly based on machine learning algorithm which gives the output in the form of binary value. Here the feature extraction is important phase of project to add more benefits to the system. A performance evaluation is carried out on a large dataset which includes around 600 tweets to identify the spammer also system helps to categories the spam and non spam message.

VII. FUTURE ENHANCEMENT

Spammer Detection has strong commercial interest because companies or individuals want to improve the security on social media. In future the picture message and location for detecting spammer. Enhancing the detecting model by considering other features and applying network analyzing to improve accuracy in the model.

REFERENCES

[1]. C. Fuchs, Social media: A critical introduction. Sage, 2017.

[2]. N. Selwyn, “Social media in higher education,” The Europa world of learning, vol. 1, no. 3, pp. 1–10, 2012.

[3]. H. Karjaluoto, P. Ulkuniemi, H. Keinanen, and O. Kuivalainen, “An- ¨ tecedents of social media b2b use in industrial marketing context: customers’ view,” Journal of Business & Industrial Marketing, 2015.

[4]. W. Akram and R. Kumar, “A study on positive and negative effects of social media on society,” International Journal of Computer Sciences and Engineering, vol. 5, no. 10, pp. 351–354, 2017.

[5]. D. Tapscott et al., The digital economy. McGraw-Hill Education,, 2015.

[6]. “Datasets,”https://www.kaggle.com/datasets, accessed: June 2020.

[7]. D. L. Hoff and S. N. Mitchell, “Cyberbullying: Causes, effects, and remedies,” Journal of Educational Administration, 2009.

[8]. S.Hinduja and J. W. Patchin, “Bullying, cyberbullying, and suicide,” Archives of suicide research, vol. 14, no. 3, pp. 206–221, 2010.

[9]. D.Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, “Detection of harassment on web 2.0,” Proceedings of the Content Analysis in the WEB, vol. 2, pp. 1–7, 2009.

[10]. K.Dinakar, R.Reichart, and H.Lieberman, “Modeling the detection of textual cyberbullying,” in In Proceedings of the Social Mobile Web. Citeseer, 2011’’

[11]. K. Kumari, J. P. Singh, Y. K. Dwivedi, and N. P. Rana, ‘‘Bilingual cyberaggression detection on social media using LSTM autoencoder,’’ Soft Comput., vol. 25, no. 14, pp. 8999–9012, Jul. 2021”

[12]. P. Nand, R. Perera, and A. Kasture, ‘‘‘How bullying is this message’: A psychometric thermometer for bullying,’’ in Proc. 26th Int. Conf. Comput. Linguistics, Tech. Papers (COLING), 2016, pp. 695–706.

[13]. B. A. H. Murshed, S. Mallappa, O. A. M. Ghaleb, and H. D. E. Al-ariki, ‘‘Efficient Twitter data cleansing model for data

analysis of the pandemic tweets,'' in Studies in Systems, Decision and Control, vol. 348. Springer, 2021.

[14]. T. Anuprathibha and C. S. Kanimozhiselvi, ''Penguin search optimization based feature selection for automated opinion mining,'' Int. J. Recent Technol. Eng., vol. 8, no. 3, pp. 648–653, 2019.

[15]. H. M. Abdulwahab, S. Ajitha, and M. A. N. Saif, ''Feature selection techniques in the context of big data: Taxonomy and analysis,'' Appl. Intell., Jan. 2022.

[16]. T. Anuprathibha and C. S. Kanimozhiselvi, ''Enhanced medical tweet opinion mining using improved dolphin echolocation algorithm based feature selection,'' Int. J. Innov. Technol. Explor. Eng., vol. 8, no. 10, pp. 2049–2055, 2019.

[17]. D. P. Mandic and J. A. Chambers, Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability. Hoboken, NJ, USA: Wiley, 2001.

[18]. 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.

[19]. Q. Huang, V. K. Singh, and P. K. Atrey, ''Cyber bullying detection using social and textual analysis,'' in Proc. 3rd Int. Workshop Socially-Aware Multimedia (SAM), 2014, pp. 3–6.

[20]. A. Squicciarini, S. Rajtmajer, Y. Liu, and C. Griffin, ''Identification and characterization of cyberbullying dynamics in an online social network,'' in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Aug. 2015, pp. 280–285.

[21]. V. Balakrishnan, S. Khan, and H. R. Arabnia, ''Improving cyberbullying detection using Twitter users' psychological features and machine learning,'' Comput. Secur., vol. 90, Mar. 2020, Art. no. 101710.

[22]. K. S. Alam, S. Bhowmik, and P. R. K. Prosun, ''Cyberbullying detection: An ensemble based machine learning approach,'' in Proc. 3rd Int. Conf. Intell. Commun. Technol. Virtual Mobile Netw. (ICICV), Feb. 2021, pp. 710–715.

[23]. S. Pericherla and E. Ilavarasan, ''Transformer network-based word embeddings approach for autonomous cyberbullying detection,'' Int. J. Intell. Unmanned Syst., May 2021.

[24]. V. Nahar, S. Al-Maskari, X. Li, and C. Pang, ''Semi-supervised learning for cyberbullying detection in social networks,'' in Databases Theory and Application (Lecture Notes in Computer Science), vol. 8506. Cham, Switzerland: Springer, 2014, pp. 160–171.

[25]. J.-M. Xu, K. S. Jun, X. Zhu, and A. Bellmore, ''Learning from bullying traces in social media,'' in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., 2012, pp. 656–666.

[26]. J. Chen, S. Yan, and K.-C. Wong, ''Verbal aggression detection on Twitter comments: Convolutional neural network for short-text sentiment analysis,'' Neural Comput. Appl., vol. 32, no. 15, pp. 10809–10818, Aug. 2020.

[27]. R. Zhao and K. Mao, ''Cyberbullying detection based on semanticenhanced marginalized denoising auto-encoder,'' IEEE Trans. Affect. Comput., vol. 8, no. 3, pp. 328–339, Jul. 2017.

[28]. Z. Zhang, D. Robinson, and J. Tepper, ''Detecting hate speech on Twitter using a convolution-GRU based deep neural network,'' in Proc. Eur. Semantic Web Conf. (ESWC), in Lecture Notes in Computer Science, vol. 10843, A. GangemiAnna, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, and M. Alam, Eds. Cham, Switzerland: Springer, 2018, pp. 745–760.

[29]. S. Paul and S. Saha, ''CyberBERT: BERT for cyberbullying identification,'' Multimedia Syst., no. 0123456789, Nov. 2020.

[30]. M. P. Akhter, Z. Jiangbin, I. R. Naqvi, and M. AbdelMajeed, ''Abusive language detection from social media comments using conventional machine learning and deep learning approaches,'' Multimedia Syst., Jun. 2021.

[31]. X. Zhang, J. Tong, N. Vishwamitra, E. Whittaker, J. P. Mazer, R. Kowalski, H. Hu, F. Luo, J. Macbeth, and E. Dillon, ''Cyberbullying detection with a pronunciation based convolutional neural network,'' in Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2016, pp. 740–745.

[32]. H. Rosa, D. Matos, R. Ribeiro, L. Coheur, and J. P. Carvalho, ''A 'deepe,' look at detecting cyberbullying in social networks,'' in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–8.

[33]. M. A. Al-Ajlan and M. Ykhlef, ''Optimized Twitter cyberbullying detection based on deep learning,'' in Proc. 21st Saudi Comput. Soc. Nat. Comput. Conf. (NCC), Apr. 2018, pp. 1–5.

[34]. Q. Huang, D. Inkpen, J. Zhang, and D. Van Bruwaene, ''Cyberbullying intervention based on convolutional neural networks,'' in Proc. 1st Workshop Trolling, Aggression Cyberbullying (TRAC), 2018, pp. 42–51.

[35]. R. R. Dalvi, S. B. Chavan, and A. Halbe, ''Detecting a Twitter cyberbullying using machine learning,'' Ann. Romanian Soc. Cell Biol., vol. 25, no. 4, pp. 16307–16315, 2021. [13] R. Zhao, A. Zhou, and K. Mao, ''Automatic detection of cyberbullying on social networks based on bullying features,'' in Proc. 17th Int. Conf. Distrib. Comput. Netw., Jan. 2016, pp. 1–6.

[36]. L. Cheng, J. Li, Y. N. Silva, D. L. Hall, and H. Liu, ''XBully: Cyberbullying detection within a multi-modal context,'' in Proc. 12th ACM Int. Conf. Web Search Data Mining, Jan. 2019, pp. 339–347.

[37]. K. Reynolds, A. Kontostathis, and L. Edwards, ''Using machine learning to detect cyberbullying,'' in Proc. 10th Int. Conf. Mach. Learn. Appl. Workshops (ICMLA), vol. 2, Dec. 2011, pp. 241–244.

[38]. S. Agrawal and A. Awekar, ''Deep learning for detecting cyberbullying across multiple social media platforms,'' in

Advances in Information Retrieval (Lecture Notes in Computer Science), vol. 10772, G. Pasi, B. Piwowarski, L. Azzopardi, and A. Hanbury, Eds. Cham, Switzerland: Springer, 2018, pp. 141–153.

[39]. A. M. Vivolo-Kantor, B. N. Martell, K. M. Holland, and R. Westby, ''A systematic review and content analysis of bullying and cyber-bullying measurement strategies,'' Aggression Violent Behav., vol. 19, no. 4, pp. 423–434, Jul. 2014.

[40]. H. Sampasa-Kanyinga, P. Roumeliotis, and H. Xu, ''Associations between cyberbullying and school bullying victimization and suicidal ideation, plans and attempts among Canadian schoolchildren,'' PLoS ONE, vol. 9, no. 7, Jul. 2014, Art. no. e102145.

[41]. M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, ''Improving cyberbullying detection with user context,'' in Proc. Eur. Conf. Inf. Retr., in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 7814, 2013, pp. 693–696.

[42]. K. Miller, ''Cyberbullying and its consequences: How cyberbullying is contorting the minds of victims and bullies alike, and the law's limited available redress,'' Southern California Interdiscipl. Law J., vol. 26, no. 2, p. 379, 2016.

BIOGRAPHIES

BRINDA BM is an Assistant professor, Department of Computer Science and Engineering in Paavai College of Engineering.

FARITHA BEGUM N is an undergraduate student, Department of Computer Science and Engineering in Paavai College of Engineering.

SWARNA N is an undergraduate student, Department of Computer Science and Engineering in Paavai College of Engineering.

SWATHI S is an undergraduate student, Department of Computer Science and Engineering in Paavai College of Engineering