International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Detection of Cyberbullying Attacks in the Social Media Platform Using SVM Method

Suriya A[1], Fathima G[2]

[1]Suriya A M.sc, Department of Computer science and Engineering, Dr. MGR Educational and Research Institute, Chennai, India
[2]Fathima G Faculty, Center of Excellence in Digital Forensics, Dr. MGR Educational and Research Institute, Chennai, India

Abstract:
Social media platforms such as Twitter, Facebook, Instagram and also WhatsApp, they have become the preferred online platform for interaction and communication. These platforms are leads to malign activities such as Cyberbullying. Cyberbullying is a type of psychological abuse with a help of using digital Medias and digital devices in now a days and it create a significant impact on a society. Cyberbullying is becoming widely increased in now a days in young age Children's and women. Need for a Cyberbullying detection is important In now a days. The complete solution for Cyberbullying is may not found and it would not be hundred percentage prevented but the detection of Cyberbullying is reduce the impact of Cyberbullying in the society in now a days. Cyberbullying leads to serious mental health issues (anxiety, depression, sleeping disorders). Cyberbullying has been mostly increased among the young age people. In India 45% Children's were experiencing the Cyberbullying according to the Times of India report. In this paper we had proposed the demo method for the CB(cyberbullying) detection using machine learning using the SVM in the WhatsApp social media platform. In this method it will detected the cyberbullying words and it will alert the user with the notification. The SVM algorithm performed well during the detect and predict the cyber bully words. The accuracy level of the SVM algorithm is in this proposed model is 96%. This accuracy level shows that svm outperformed well in the detection of cyberbullying words.

Keywords: Cyberbullying, Alert, detection, SVM, WhatsApp social media.

Introduction:

Social media platforms such as Twitter, Facebook, Instagram and also WhatsApp, have become the preferred online platform for interaction and communication. Particularly social media networks namely Twitter and Facebook. These platforms leads to malign activities such as Cyberbullying. These platforms leads to malign activities such as Cyberbullying. Cyberbullying is a type of psychological abuse with the help of using digital medias and digital devices in now a days and it create a significant impact on a society. Mainly the Cyberbullying mostly increased among the young age peoples. In India, for example ,14% of all online harassments occurs on Facebook and Twitter, in that the 37% of incidents involves and affects the youngsters [1].furthermore the cyberbullying leads to serious mental health issues namely anxiety, depression, stress. In some peoples commits suicide due to this anxiety and depression [2]. This motivates the need for an approach to detect and identify cyberbullying in social media.

In this paper, we mainly focus on the cyberbullying detection on the WhatsApp social media and it's a demo model for the future enhancement. In this paper cyberbullying detection was done in the WhatsApp social media and alert the user, who had used the cyberbullying words or any harmful words, with the notification.

Review of literature:

Celestine Iwendi, Gautam Srivastava and et al., [3] had proposed the system for cyber bullying detection .In that proposed model they had compared the four deep learning algorithms namely Bidirectional Long Short-Term memory(LSTM),Gated Recurrent Units(GRU),Recurrent Neural Network(RNN) for the cyber bullying detection. For that they had pre-processing the data used for the detection. That pre-processing steps involved that the text cleaning, stemming , tokenization, and lemmatization. After the data pre-processing, they had passed the pre-processed data into deep learning algorithms. In that the BLSTM(Bidirectional Long-Short memory) achieve the highest accuracy against the cyber bullying detection.

Ammar Almomani, Khalid Nahar and et al.,[4] had proposed the system for image cyber bullying detection. In that proposed system they had used a hybrid approach for image cyber bullying that means using a deep learning model for a extracts and using machine learning model for classification. The extracts process had completed using deep learning models namely InceptionV3, ResNet50, and VCG16.They had used the classifiers like Logistic Regression and Support Vector Machines for classification process. They had feeding the extracted features into this classifiers. They had combined the both deep learning mode and classifiers for the high accuracy.

Reem Albayari, Sherief Abdallah and et al.,[5] had proposed the cyber bullying detection model for the Arabic text. In that they had proposed the deep learning models for the detection. They had also conduct the performance evaluation and comparison for various deep learning algorithms namely (LSTM, GRU, LSTM-ATT, CNN, BLSTM, CNN-LSTM, LSTM-TCN) on different dataset. And also they had proposed the hybrid DL(combines the characteristics of the baseline models CNN, BLSTM, GRU) model for the result of the model's evaluation. The proposed hybrid model increased the accuracy of the of the all studied datasets. And that proposed hybrid model significantly reduce the cyberbullying.

T. Nitya Harshitha, M. Prabu and et al.,[6] had propped the hybrid deep learning model or the cyber bullying detection on social media. In that proposed model they had used the hybrid RANDOM FOREST based CNN model for the text classification. And also they had combining the strengths of the both approaches. The had collected the dataset from the twitter and the Instagram. They had compared the performances of the various ML and DL algorithms, and the RF-based CNN model had outperformed them in accuracy and execution speed. In the execution they got the result of the RF-based CNN model achieved an accuracy of 96%and delivered the results 3,4 seconds faster than the standard CNN models.

Alanoud Mohammed Alduailaj, Aymen Belghith [7]had proposed the detection of Arabic cyberbullying tweets using the machine learning. In that they had proposed the several detection methods, but they are mainly focused on the word-based data and user attributes. Mainly they had used the Support Vector Machine(SVM) classifier algorithm for the classification purposes. And they had taken dataset from the YouTube and Twitter. They had done the cyberbullying detection on the Arabic social media. And also they had used the Farasa tool which is suite of tools for Arabic Natural Language. They had received the result of, the SVM model had detecting the cyberbullying content with the percentage of 95.72%.

Arnisha Akhter , Uzzal Kumar Acharjee and et al.,[8] had proposed the machine learning model for Bengali cyberbullying detection. In that proposed model they had used the robust ML model for cyberbullying detection in the Bengali language on social media. For this proposed model they had used the publicly available Bengali dataset(44,001comments). In that model they had done the effective preprocessing to make the Bengali text data into useful text format. For the feature extraction they had used the TfidfVectorizer(TFID)for get the useful information of text data and resampling the dataset by Instance Hardness Threshold(IHT). This proposed model achieved the high accuracy rate of 98.57%and 98.82% in binary and multilabel classification to detect the cyber bullying on social media in the Bengali language.

Belal Abdullah Hezam Murshed, Suresha and et al.,[9] had proposed the cyber bullying detection social media platform using topic modelling and deep learning. In this proposed model they had integrates the Fuzzy Adoptive Equilibrium Optimization(FAEO), and Extended convolutional Neural Network(ECNN) to enhance the accuracy of cyberbullying detection. They had evaluated the proposed FAEO-ECNN thoroughly with two short text datasets: Real-world CB Twitter(RW-CB-Twitter) and Cyberbullying Menedely(CB-MNDLY). They had used the state of the art models(SoAT) namely long short term memory(LSTM), Bi-directional LSTM(BLSTM), RNN, and CNN-LSTM. This proposed FAEO-ECNN model was outperformed the SoTA models in Cyberbullying on social media platforms. In result this proposed model obtained the 92.91%of accuracy.

Research methodology:

To figure out or detect the cyber bullying detection using the machine learning algorithm namely as SVM (support vector machine). In this model detect and alert the user while using the cyberbully words or any other harassment words that alert method is done by using the platform named Twilio which is best platform for unique communication features. This model had mainly focused on the WhatsApp social media platform. First in this model extracted the dataset, and use the extracted dataset for the detection purposes. Before the data extraction the data pre-processing is had done. The data pre-processing steps are important to data extraction. The data pre-processing done under the three main sub-phases that is the noise removal such as url's removal, hashtag/punctuation removal, and emotion transformation process. From that pre-processed dataset we had took the particular word for our deployment model And the next level is the training the dataset for the detection of cyberbullying. In this training step the ML techniques had used. In this step to train the algorithm for the cyber bullying detection (SVM). And then next the validation process is had done. And then last alert the user with the notification message with the cyber bullying detection.

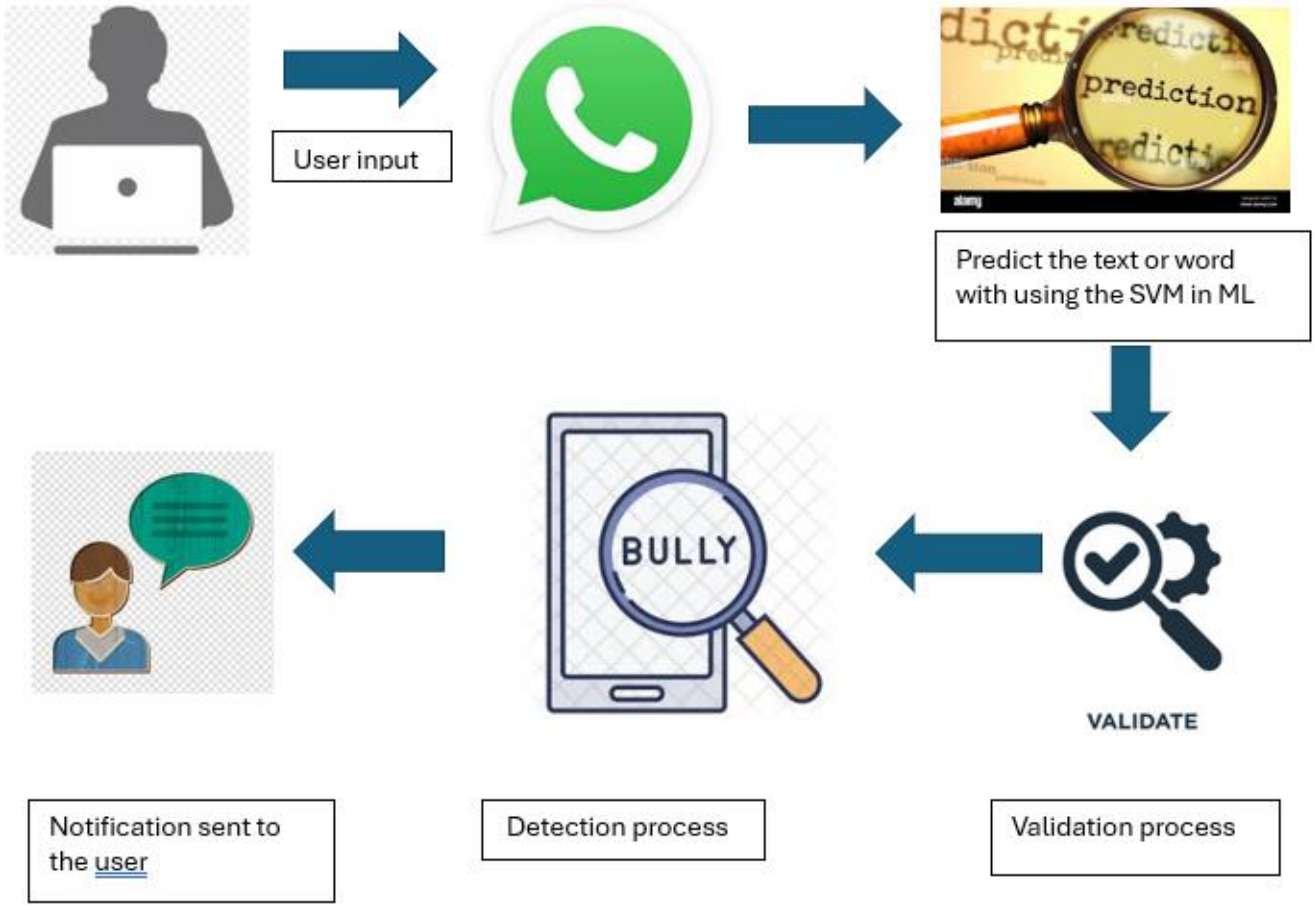

Figure .1: The detection and alert process

The figure 1 shows that the step by step process about this model. In this model if the user use the particular bully word in the WhatsApp then the notification sent to the user that what the word exactly he or she using and when is he or she is using and with the prediction of bullying. For the notification send method the Twilio platform that we had used in this model, In that platform sent the message to which number registered in it. For this model we had used the jupyter notebook. Jupyter note book is the web based environment that allows users for live coding. In that jupyter notebook we had run our python code. Once the code is running the jupyter notebook redirect to the WhatsApp (i.e.registerd WhatsApp number) and the bully word was automatically typed in the chat of the registered WhatsApp number chat, and it was detected by the SVM (Support Vector Machine) and its predicted that is bully word. And after that the notification will sent to the user with the help of Twilio platform that it is bully word.

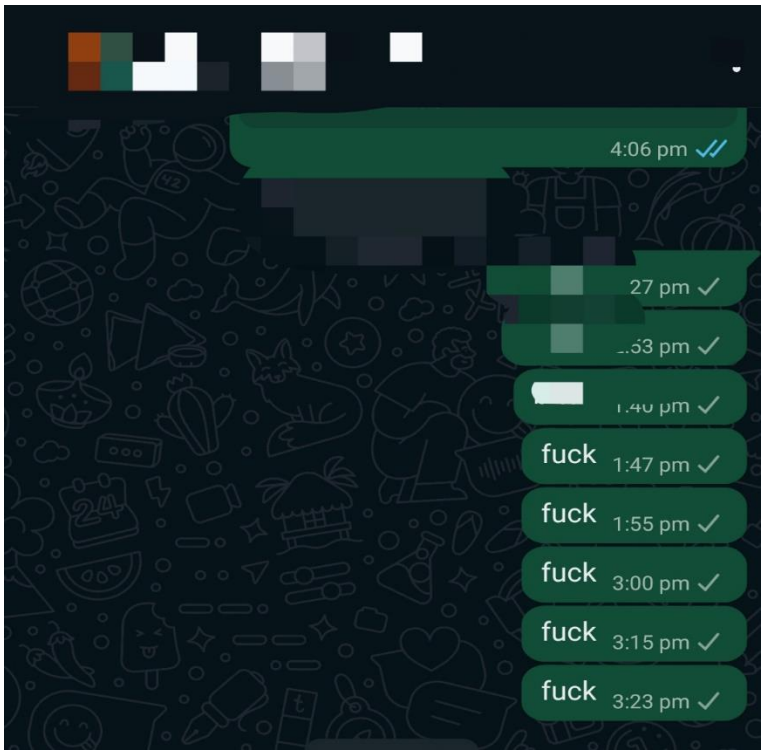

Figure. 2: The message automatically send to the WhatsApp chat

The figure 2 shows that particular bully word send to the registered WhatsApp chat and it was detect and predicted by the Support Vector Machine (SVM) algorithm and after that the notification send to the registered mobile number.

Figure. 3: The notification send to the user

The figure 3 shows that the notification that sent to the user when he or she used the particular cyber bully word. In this figure figures out which word the user is using and mentioned the date and time and also predict that whether it is bully. This detection notification will alert the user he or she was using the bully word.

Conclusion

In this proposed cyber bullying detection model using SVM in machine learning ensures that the cyber bullying words that had clearly detected and predicted. And Support Vector machine (SVM) algorithm has performed as well out during the detection and prediction. And also the Twilio platform is worked well. In the future this proposed model have been added to the other social media platforms. For example this proposed model should be added as a BOT in the YouTube comment section to detect the cyber bullying. Same like also added as a bot to the Instagram comment section to detect the cyber bullying. In

the future it will helpful to detect the cyber bullying and it will much helpful to the investigators to solve the cyber bullying related cases. And also in the future this proposed model will help to reduce the ratio of the people who affected by the cyber bullying.

REFERENCES

1. B. A. H. Murshed, J. Abawajy, S. Mallappa, M. A. N. Saif and H. D. E. Al-Ariki, "DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform," in IEEE Access, vol. 10, pp. 25857-25871, 2022
2. Iwendi, C., Srivastava, G., Khan, S. *et al.* Cyberbullying detection solutions based on deep learning architectures. *Multimedia Systems* **29**, 1839–1852 (2023).
3. Ammar Almomani, Khalid Nahar and et al. "Image cyberbullying detection and recognition using transfer deep machine learning". *International journal of cognitive computing in engineering*, volume 5, pages 14-26.
4. Reem Albayari, Sherief Abdallah and et al. "Cyber bullying detection model for Arabic text using deep learning". *Journal of information and knowledge management*. Jan-2024.
5. T. Nitya Harshitha, M. Prabu and et al., "a hybrid deep learning model for proactive detection of cyberbullying on social media", *methods article*, volume 7-2024
6. lduailaj, Alanoud Mohammed, and Aymen Belghith. 2023. "Detecting Arabic Cyberbullying Tweets Using Machine Learning" *Machine Learning and Knowledge Extraction* 5, no. 1: 29-42.
7. Arnisha Akhter, Uzzal Kumar Acharjee, Md. Alamin Talukder, Md. Manowarul Islam, Md Ashraf Uddin, A robust hybrid machine learning model for Bengali cyber bullying detection in social media, *Natural Language Processing Journal*, Volume 4,2023.
8. Murshed, B.A.H., Suresha, Abawajy, J. *et al.* FAEO-ECNN: cyberbullying detection in social media platforms using topic modelling and deep learning. *Multimed Tools Appl* **82**, 46611–46650 (2023).
9. T. H. Teng and K. D. Varathan, "Cyberbullying Detection in Social Networks: A Comparison Between Machine Learning and Transfer Learning Approaches," in *IEEE Access*, vol. 11, pp. 55533-55560, 2023.
10. Md.Tofael ahmed, nahida akter and et al. Multimodal cyberbullying meme detection from social media using deep learning approach. *International journal of computer science & Information technology (IJCST)* vol 15, no 4, august 2023.
11. Ali, M.U., Lefticaru, R. (2024). Detection of Cyberbullying on Social Media Platforms Using Machine Learning. In: Naik, N., Jenkins, P., Grace, P., Yang, L., Prajapat, S. (eds) Advances in Computational Intelligence Systems. UKCI 2023. *Advances in Intelligent Systems and Computing*, vol 1453. Feb-2024.
12. Islam, M.R., Bataineh, A.S., Zulkernine, M. (2024). Detection of Cyberbullying in Social Media Texts Using Explainable Artificial Intelligence. In: Wang, G., Wang, H., Min, G., Georgalas, N., Meng, W. (eds) Ubiquitous Security. UbiSec 2023. *Communications in Computer and Information Science,* vol 2034. Mar-2024.
13. Bechir, S.B., Mekki, A., Ellouze, M. (2024). Transfer Learning Model for Cyberbullying Detection in Tunisian Social Networks. In: Mosbah, M., *et al.* Advances in Model and Data Engineering in the Digitalization Era. MEDI 2023. *Communications in Computer and Information Science*, vol 2071. Mar-2024.

14. M. Sen, J. Masih and R. Rajasekaran, "From Tweets to Insights: BERT-Enhanced Models for Cyberbullying Detection," *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, Manama, Bahrain, 2024, pp. 1289-1293.
15. A. A. Zotkina and A. I. Martyshkin, "Detection of Cyberbullying in Texts Posted by Users of Social Networks Using Machine Learning," *2024 International Russian Smart Industry Conference (SmartIndustryCon)*, Sochi, Russian Federation, 2024, pp. 639-643.