e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

PREVENTION OF RISK ASSESSMENT IN SOCIAL NETWORKUSING DEEP LONG SHORT TERM MODEL

Ramathilagam A[*1], Santhiya R[*2], Harini S[*3], Sivasankari M[*4]

[*1]Professor, Computer Science And EngineeringPSR Engineering College Sivakasi, Tamil Nadu, India.

[*2,3,4]UG Scholar, Computer Science And Engineering PSR Engineering College Sivakasi, Tamil Nadu, India.

ABSTRACT

Online social systems have become an important part of daily life. People use it to share their personal contentwith their friends in chat. Unfortunately, social networks provide little support to prevent harassment and spam comments from users' walls. To overcome this problem, the system introduces a deep short-term transformation (LSTM) algorithm. Thisalgorithm is accepted and works well to block spam text. This method is implemented on Java platform with Netbeans compileras front-end and MySQL database as back-end.

Keywords: Cyber Harassment, Online Social Network, LSTM, Message Blocking.

I. INTRODUCTION

Online social networks (OSNs) enable communication and exchange of interests and information between members by creating public or private profiles. OSNs have become part of our daily lives and are used by millions of people. People use Open Social Networks (OSNs) to conduct business, share personal information, and connect with friends and family [1]. OSN users also build relationships with friends, colleagues, and others. These links form a graph that controls the spread of information in social networks. Facebook and Instagram, for example, already have 1.55 billion monthly active users, 1.31 billion mobile users, and 1.01 billion daily users despite the sharp rise in OSN usage.

Unfortunately, most users are unaware of this effect and it can be dangerous. In addition, some users upload sensitive information from their accounts without mentioning the necessary privacy, as they do not attach much importance to personal information, and this causes people to exercise their risk freedom [2]. For this reason, social networks used today are exposed to many privacy and security threats. For example, these attacks, which persuade users to click on malicious links in order to spread the links throughout the network, collect and delete users' personal information using the OSN infrastructure. The user's personal information and his friends' personal information will be the target of these attacks[3].Online dating has now become an important part of daily life. In the past, people used social networks toshare private content with their circle of friends.

Unfortunately, social networks don't do much to help prevent bullying or complaints from users' walls. In this system, it is recommended to use the deep short-term model (LSTM) method to solve this problem. The program effectively blocks unwanted texts. This technology is implemented on the Java platform using MySQL databaseas the backend and Netbeans compiler as the frontend.

Fake profiles are another attack created to spread malicious content. There is also a growing black market for corrupt practices on OSN, where users can buy fake money, products, Twitter and Instagram followers, and Facebook likes for very little money. Although many solutions to specific types of attacks have been proposed recently (see, for example, Therefore, the aim of this paper is to evaluate the risk of each user, taking into account each user's online activities and online social structure [4]. The purpose of behavioral detection is to compare the user's behavior with the behavior of other users on the network. The basic principle is that user behavior is considered risky (i.e. high risk) in direct proportion to the difference between so-called "good behavior". In order to implement this policy, two important issues need to be addressed. The first of these is the description of the user's behavioral profile, which describes the user's behavior and communication, which is important for risk assessment.

The second question is how to teach "good behavior". We must note that the OSN population exhibits many positive behaviors in achieving this goal. However, just like in the real world, we expect that similar users (e.g. in terms of activity level, gender, education, country, etc.) will generally follow similar rules (morals and

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

customs) with comparable patterns of behavior [5]. Based on the above ideas, this article introduces the LSTM algorithm in this system. The program effectively blocks unwanted text. This document is organized as follows: Section 2 provides an overview of relevant documents. Section 3 describes procedures for investigating harassment. Section 4 presents the experimental study and shows obtained results. Finally the paper ends with a conclusion.

II. RELATED WORKS

The research has suggested a number of methods for detecting cyberbullying and online harassment.

According to Huang et al. [5], taking into account the social connections among Twitter users can enhance the categorization outcomes for cyberbullying. For every tweet, they create a relationship graph and use the quantity of linkages, edges, and nodes to extract social traits.

A technique called embedding-enhanced Bag-Of-Words (EBoW), which integrates word embedding-based bullying features, latent semantic features, and bag-of-words features, was proposed by Zhao et al. [6]. To determine whether the tweet was bullying or not, they employed a linear support vector machine. They then evaluated the effectiveness of their approach against that of the bag-of-words model, a bag-of-words model with semantic enhancements, Latent Semantic Analysis (LSA), and Latent Dirichlet Allocation (LDA).

9484 tweets were first made available in a dataset by Despoina et al. [7]. They then divided the traits into three categories: 1) User-based features based on the number of tweets, age of the account, number of lists subscribed to, and state of account verification; 2) Text-based features based on the sentiment, quantity of uppercase text, emoticons, count of URLs, and number of hashtags used. 3) Network-based features that evaluate user popularity and reciprocity, i.e., which people return the favor of receiving follower connections from other users.

The issue was posed as a classification challenge by T. Marwa et al. [8], who looked into the efficacy of deep learning in identifying online harassment in a large human-labeled corpus created especially for the purpose of harassment research. Long short-term memory (LSTM), bidirectional long short-term memory (BLSTM), and convolutional neural network (CNN) are the models taken into consideration in order to do this, and they are comparedwith other categorization models. The outcomes obtainedare really positive.

To combat online harassment, K. Rizwan et al. [9] used natural language processing and machine learning. This work presented a real-time machine learning algorithm that actively detects harassment and notifies the user to take appropriate action. In order to detect it, Naïve Bayes classification is employed.

An informed strategy for reducing such difficulties was provided by M. T. Shahria et al. [10]. A model that relies onConvolutional Neural Networks (CNNs) has been developed to identify instances of sexual harassment in the workplace, enabling prompt resolution or verdict. We built our own dataset from social media videos for the model since the existing dataset on sexual harassment at work was inadequate. In conclusion, it has been suggested to apply transfer learning techniques to three well-known Keras pre-trained models in order to improve model accuracy insight.

III. PROPOSED METHODOLOGY

This section presents the suggested methodology. Figure 1 depicts the whole system architecture, which includes feature extraction, dataset collection, LSTM classifier, and harassment prediction.

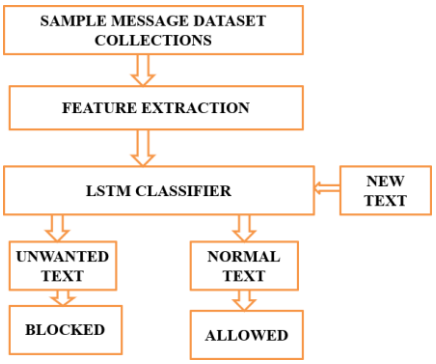

Figure 1: proposed system architecture.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

After gathering the dataset, features are extracted using numerical values that have already been preprocessed to change the feature value range between 0 and 1. Subsequently, the suggested system employed Long Short-Term Memory techniques for the dynamic blocking procedure. This algorithm uses supervised learning techniques to classify data and automatically blocks undesirable text messages. More categories are created by this method, and more data is saved under each category.

This system is implemented using multiple modules,including

- Network scenario
- Filtering rules
- Online setup assistant for FRS thresholds
- Blocked unwanted message

A. Network scenario

On social media, creators can be seen on the Internet using data from social graphs. This suggests that in order to control the design, designers must be involved in determining the nature, strength and trust of the relationship.The concept of creator's specifications formalizes each of these options.

B. Filter Rules (FR)

We take into consideration three primary concerns that we believe should influence a message filtering decision while designing the language for the FRs definition. Firstly, just as in real life, the meaning and significance of a communication can vary depending on who publishes it, even on open social networks. FRs ought to enable users to impose restrictions on message producers as a result. A FR may be applied to creators based on a number of factors, themost important of which is placing restrictions on the characteristics of their profile. This allows for the formulation of regulations that, for example, apply exclusively to young artists or creators who share a particular political or religious viewpoint.

C. Online setup assistant for FRS thresholds

As was indicated in the preceding section, we create and implement an Online Setup Assistant (OSA) procedure within FW to address the issue of defining thresholds tofilter rules. The user is shown a selection of messages chosen from the dataset by OSA. The user informs the system whether they choose to accept or reject each message. It is possible to calculate personalized thresholds that indicate the user's stance toward accepting or rejecting particular contents through the gathering and processing of user judgments on a sufficient set of messages spread across all classes. These messages are chosen using the subsequent procedure. The LSTM classifies a given number of non-neutral messages that are extracted from a portion of the dataset and do not belong in the training or test sets, are classified by the LSTM in order to have, for each message, the second level class membership values.

D. Blocked unwanted message

Like FRs, the wall owner can identify persons to be blocked based on their associations in the OSN and their profiles with our BlacklList (BL) criteria. Because of this, wall owners can use BL rules to, for instance, block people from their walls who they do not personally know (i.e., withwhom they have only indirect ties) or users who may be friends with someone about whom they may have negative opinions of them. This prohibition may be imposed for a set amount of time or during a particular window of time. Furthermore, the way users behave on the OSN may also be factored into the banning criteria. More specifically, we have concentrated on two primary metrics among the potential indicators of users' inappropriate behavior. The first is basedon the idea that if a user enters a behavior loop (BL) more than once in a predetermined amount of time, say, and their conduct does not improve, they may be entitled to remain inthe BL for an additional period of time. This idea is applicable to users who have at least one prior insertion into the BL under consideration.

E. Algorithm

LSTM MODEL: We have evaluated our dataset and implemented an LSTM-based model to categorize sexual harassment in a typical workplace setting. The Long Short-Term Memory (LSTM) architecture is a kind of recurrent neural network (RNN) intended to solve the vanishing gradient issue with conventional RNNs. Because the gradients in the vanishing gradient issue get very small,it becomes difficult to update the weights of the network, which makes it challenging to train RNNs on long data sequences.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Cell State: The ability of LSTM to preserve a componentof long-term memory known as the cell state is essential to its effectiveness. As information is transferred from one time step to another, the cell state functions as a conveyor belt that traverses the length of the network.

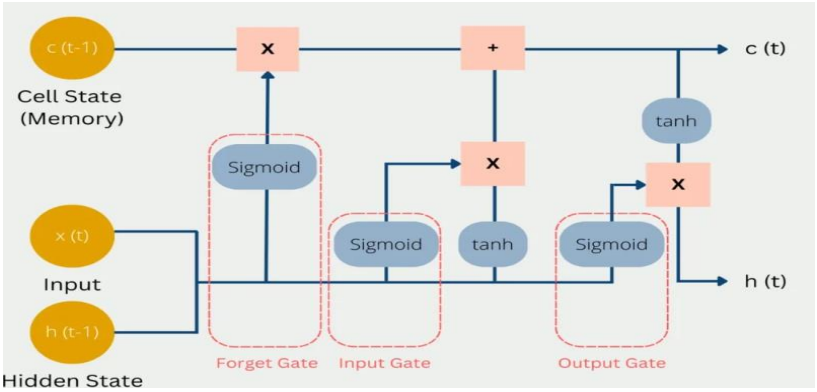

Figure 2: LSTM Architecture

Gates: The input, forget, and output gates are the three gates that LSTMs utilize to regulate the flow of information. These gates are in charge of deciding which data should be output as the final forecast, which should be discarded, and which should be held in the cell state.

Forget Gate: This gate determines which data from the cell state should be retained or deleted. For each number in the cell state, it outputs a number between 0 and 1, taking as inputs the previous concealed state and the current input. "Keep this" is indicated by a 1 and "discard this" with a 0.

Input Gate: The input gate determines what new data tostore in order to update the cell state. The two layers of itare the sigmoid layer, which selects which values to update, and the tanh layer, which builds a vector of potential new values that could be added to the cell state.

Output Gate: The output gate selects the value of the next concealed state. The cell state is filtered to create the concealed state. What portions of the cell state areappropriate for output exposure is determined by the output gate.

Secret State: The current input, the hidden state from the previous time step, and the current cell state all influence thehidden state at each time step. The state that is concealed acts as memory or classification.

IV. RESULT AND DISCUSSION

The use and development of LSTM-based models requires training them to use specific information inthe data. Then test the model using the labels of the test data.

Fig.3 Home Page

The Above Figure Shows That, The Home Page Consists Of Admin, User Login And Creating The New User.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Figure 4: LSTM Architecture

The Above Figure Shows That, Admin Only Can Access All The User Details. So The Admin Login Can Be Used.

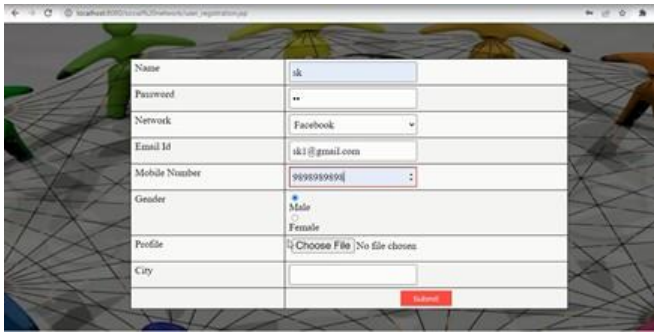

Fig.5 User Registration

In This, Users Should Register With Their Personal Details And Click Submit. So That The Informations Is Stored In The Database.

Fig.6 User Request For Security

In This, The User Must Click 'Create' So That Protection For Their Account Should Be Provided Or Else They Use Their Account Normally.

Fig.7 Adding Block List

The Above Figure Illustrates That The Admin Can Add The Words, Which Is Need To Be Blocked.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Fig.8 Viewing Friend List

This Is The User Page, By Clicking The View Message Module , You Can View The Messages That Were Sent By Your Friends.

Fig.9 Sending Block List Words

The Above Figure Illustrates The Message Box, If The User Send Th Message With Unwanted Words To Another User ,That Message Will Not Send To The User.

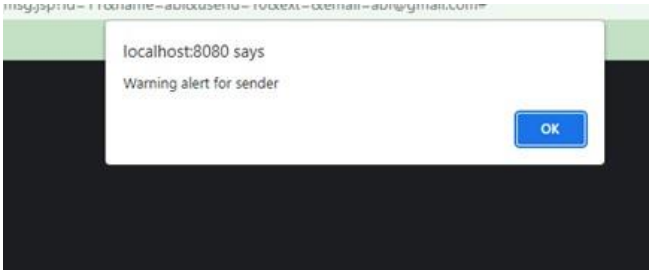

Fig.10 Warning Alert

If The User Sends The Unwanted Messages, Our Model Warns Him Till It Meets The Threshold Value ,If It Repeats The User Will Be Blocked.

V. CONCLUSION

In order to eliminate this social issue, workplaces should be intelligently monitored at all times. And if the monitoring tool itself is able to identify the event and take appropriate action, then this can be guaranteed. Nonetheless, our method sheds light on one possible resolution for this problem. This concept has the potential to lessen this horrible worldwide problem if it can be widely used. Even though our model has a respectable accuracy rate, it could do better with a larger dataset. To create a generalized version of this model, large datasets must be used for training and testing. We plan to carry out additional research in this area and develop a more comprehensive model to forecast sexual harassment in the future. We shall therefore attempt to make the model lightweight.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

VI. REFERENCES

[1] Ferrari, E., Laleh, N., and Carminati, B. (2018). Evaluation of User Abnormal Behaviors in Social Networks to Determine Risk. doi:10.1109/tdsc.2016.2540637 IEEE Transactions on Dependable and Secure Computing, 15(2), 295–308.

[2] Li, F., and Gao, T. (2019). De-Anonymization of Dynamic Online Social Networks Using Persistent Structures. 2019's IEEE Communications International Conference (ICC). Citation: 10.1109/icc.2019.8761563.

[3] Ghandeharioun, A., Jones, N., Jaques, N., Pataranutaporn, P., & Picard, R. (2019). Examining Internet Suicide Risk Using Latent Dirichlet Allocation and Document Embeddings. 2019 Workshops and Demos of the 8th International Conference on Affective Computing and Intelligent Interaction (ACIIW). Reference: 10.1109/aciiw.2019.8925077

[4] PAPAMANTHOU, C., SONG, D., AND MITTAL, P. Link privacy preservation in social network-based systems. In the 2017 NDSS, ISOC.

[5] Q. Huang, V. K. Singh, and P. K. Atrey, "Cyber bullying detection using social and textual analysis," in Proceedings of the 3rd International Workshop on Socially-Aware Multimedia. ACM, 2014, pp. 3–6.

[6] R. Zhao, A. Zhou, and K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," in Proceedings of the 17th international conference on distributed computing and networking. ACM, 2016, p. 43.

[7] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali, "Mean birds: Detecting aggression and bullying on twitter," in Proceedings of the 2017 ACM on Web Science Conference. ACM, 2017, pp. 13–22.

[8] T. Marwa, O. Salima and M. Souham, "Deep learning for online harassment detection in tweets," 2018 3rd International Conferenceon Pattern Analysis and Intelligent Systems (PAIS), 2018, pp. 1-5, doi: 10.1109/PAIS.2018.8598530.

[9] K. Rizwan, S. Babar, S. Nayab and M. K. Hanif, "HarX: Real-time harassment detection tool using machine learning," 2021 International Conference of Modern Trends in Information and Communication Technology Industry (MYCITI), 2021, pp. 1-6, doi: 10.1109/MTICTI53925.2021.9664755.

[10] M. T. Shahria, F. Tasnim Progga, S. Ahmed and A. Arisha, "Application of Neural Networks for Detection of Sexual Harassment in Workspace," 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies(ICAECT), 2021, pp.1-4, doi:10.1109/ICAECT49130.2021.9392429.

[11] "Sexual harassment and violence against garment workers in Bangladesh," ActionAid International, Jul. 25, 2019. [Online]. Available: https://actionaid.org/publications/2019/sexual-harassmentand-violence-against-garment-workers-bangladesh. [Accessed: Nov. 10, 2020].

[12] K. B. Clancy, L. M. Cortina, and A. R. Kirkland, "Opinion: Use science to stop sexual harassment in higher education," Proceedings of the National Academy of Sciences, Sep. 2020, vol. 117, no. 37, pp. 22614-22618.

[13] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," Progress in Artificial Intelligence, Jun. 2020, vol. 9, no. 2, pp. 85-112.

[14] H. Cho, and S. M. Yoon, "Divide and conquer-based 1D CNN human activity recognition using test data sharpening", Sensors, Apr. 2018, vol. 18, no. 4, pp. 1055.

[15] J. Huang, S. Lin, N. Wang, G. Dai, Y. Xie, and J. Zhou, "TSE-CNN: A two-stage end-to-end CNN for human activity recognition," IEEE journal of biomedical and health informatics, Apr. 2019, vol. 224, no. 1, pp. 292-299.