International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Literature Review on Detecting Fake Accounts on Social Media Using Cyber Bullying

Mr.Devaraj F V*[1],Samruddhi A Navale*[2],Siddarth K*[3],Varun D*[4],Yashas T*[5]

*[1]Asst Prof.Dept.of CS&E,JNNCE,Shivamoga.

*[2,3,4,5]Dept. of CS&E ,JNNCE,Shivamog

ABSTRACT

Enhancement in the technology trend of using social networking is increasing day by day as of now there are more than 50 crores active users are using different social media platforms for the interaction which had affected their life so just like a coin has two face in a similar way misuse of these platforms is going which cause the the rapid rise of cybercrime and exploitation eg harassing someone by sending malicious messages, spreading abusive messages through fake accounts on the social media etc.. In this new era insulting a person physically or emotionally is done by cyberbullying and by using fake accounts, so as a preventive measure to ensure the above things should not happen there is a need of detecting cyberbullying and the fake accounts. In our study to stop cyberbullying and fake accounts we'll use different Machine Learning algorithms for detecting the cybercrime and fake accounts so as to report these issues to the system immediately and to stop the crimes to increase in future and develop a secure online environment.

INTRODUCTION

Social networking sites have connected us to different parts of the world however, people are finding illegal and unethical ways to use these communities. We see that people, especially teens and adults, are finding new ways to bully one another over the internet. About 25% of parents in a study conducted by Semantic reported that their child had been involved in a cyberbullying incident. Other than cyberbullying, the spread of false information is increasing at an alarming rate. The number of users in social media is increasing exponentially. Instagram, Twitter has recently gained immense popularity among social media users. The major sources of fake news are fake accounts. Business organizations that invest a huge sum of money on social media influencers must know whether the following gained by that account is organic or not. Hence there is a huge need for the detection of these fake accounts which are increasing

PROPOSED SYSTEM

The applied technique consists of the following points namely re-processing, mining required parameters, and a separate phase is listed below.

1] The very first part is to convert the jumbled or the impure information into pure information and to convert the strings into small tokens this process is known as tokenization.

2] In this part, we will convert the pure information collected from the first part to the smaller format that means converting the capital letters to small letters.

3] This is a very crucial part of this technique where we remove certain special characters such as '\b' or '\n' since we need meaningful characters and such characters don't provide any meaningful content.

4] The next part is to convert this data into Machine learning format so as to give input to our Models.

5] The final part of this technique is to provide input data to our machine learning algorithm so as to classify the data as toxic, sever_toxic, identity_hate, threat, obscene, insult.

6] The accuracy of different algorithms will be Compared to get the best possible result. For fake profile detection, this paper proposes the detection process starts with the selection of the profile that needs to be tested.

7] After selection of the profile the suitable attributes ie., features are selected on which the

classification algorithm is being implemented, the attributes extracted are passed to the trained classifier. Different Classifier algorithms such as Gradient Booster, random forest Decision trees, Support Vector Machine, and Neural Networks such as RNN and CNN can be used. The model generated by the learning algorithm should both fit the input data correctly and also correctly predict the class labels of the learning algorithm to build the model with good generality capability.

8] The complete dataset of the fake account is used for the training purpose this data after preprocessing is fed to the different machine learning algorithms and the accuracy is compared and according to the results the Random Forest has given us the best results and for the testing purpose, the live data is fetched from the Twitter.

EXISTING SYSTEM

As Compared to the existing System there are many Lacunas –

1] Lack of Security -There is a lack of Security in the existing systems but our system will deal with the proper security provision to the users.

2] No Transparency- As the existing system doesn't provide the proper transparency in their system as they are not able to deal with the Sharing of their reports to the Cybercrime Department.[3]

3] One Feature is Implemented - Other systems deal with only one part but our System will provide different features to give the best solution.

MOTIVATION

Social media platforms have become a

significant part of modern communication, allowing users to connect, share, and engage with a global audience. However, these
platforms also harbor threats like cyberbullying and the proliferation of fake accounts.
Cyberbullying can lead to severe psychological distress, depression, and even self-harm,
particularly among teenagers and young adults.

Meanwhile, fake accounts are frequently used for malicious activities such as spreading
misinformation, phishing, and online

harassment. Detecting and mitigating these issues are crucial for ensuring a safe and trustworthy digital space. Advancements in artificial intelligence (AI), machine learning
(ML), and natural language processing (NLP) offer promising solutions for automatically identifying harmful behaviors and fraudulent accounts, making this an essential area of
research.

PROBLEM STATEMENT

The rise of cyberbullying and fake accounts on social media has created an unsafe environment for users, leading to mental health concerns and the spread of misinformation. Manual
moderation of such content is inefficient due to the vast amount of data generated daily.
Existing detection methods often fail to accurately classify cyberbullying content and differentiate between genuine and fake profiles. Therefore, there is a need for an intelligent,
automated system that can effectively detect and mitigate cyberbullying incidents while
distinguishing real users from fake ones. The proposed research aims to develop an AI-
driven system using deep learning, NLP, and graph-based approaches to improve detection accuracy and enhance the overall security of social media platforms.

LITERATURE SURVEY

---

1. Cyberbullying Detection Using Machine Learning Techniques (2021) – This study explores various ML algorithms, including Support Vector Machines (SVM) and deep
learning models, for detecting cyberbullying in textual content. It highlights the challenges of dataset imbalances and context understanding.

2. Fake Account Detection in Online Social Networks Using Graph-based Features (2020) – The research employs graph-based techniques to analyze user interactions and
detect fake accounts, emphasizing the role of social network structures in classification.

3. Natural Language Processing for

Cyberbullying Detection (2019) – This paper reviews NLP techniques such as sentiment analysis, text classification, and transformers like BERT for identifying harmful speech patterns in online conversations.

4. Deep Learning for Cyberbullying

Detection on Social Media (2022) – The authors investigate the application of deep neural networks, particularly LSTMs and CNNs, to detect cyberbullying based on textual and multimedia content.

5. Hybrid Approaches for Fake Account

Detection (2021) – This study integrates ML and rule-based methods to identify fraudulent profiles, highlighting the effectiveness of hybrid approaches in improving detection rates.

6. Graph Neural Networks for Fake User Identification in Social Media (2022) – The research focuses on using Graph Neural Networks (GNNs) to model user relationships and identify fake accounts through network behavior analysis.

7. Sentiment and Emotion Analysis for

Cyberbullying Detection (2020) – This paper examines the role of sentiment and emotion analysis in cyberbullying detection, demonstrating that toxic language often carries distinct emotional signals.

8. Adversarial Attacks and Defense Mechanisms in Fake Account Detection (2023) – The study explores how attackers manipulate detection models and presents robust defenses against adversarial attacks.

9. Comparative Study of Supervised and Unsupervised Learning in Fake Account Detection (2019) – The authors analyze different learning paradigms and suggest that unsupervised anomaly detection methods can be useful in flagging suspicious accounts.

10. Multimodal Cyberbullying Detection Using Text, Images, and Videos (2022) – This research explores multimodal deep learning techniques to enhance cyberbullying detection, considering images, videos, and audio signals in addition to textual content.

METHODOLOGY

In this project, we aim to detect cyberbullying and fake account detection for the marginal number of attributes the proposed methodology consists of different steps.

1. The first step is the preprocessing and finding the proper set of attributes from the datasets i.e Cyberbullying and the fake account so in this step, we will separate a number of the attributes that will be used to identify these

bullying, contains abusive words, etc from the second dataset we will preprocess and extract the attributes like name, followers count, the following count, listed count, timezone, screen name, favorite to identify that whether the account is fake or not. The data for the fake account will be fetched via Twitter

2. The second step is to create different Machine Learning Models like Support Vector Machine Classifier, Random Forest algorithm, Naïve Bayes, Logistic Regression, K-mean clustering, ADT and BFT tree and Neural Networks and NLP

and applying the following algorithms on the datasets of cyberbullying and the fake account [4] by splitting the datasets into training and test approximately in the ratio of 75:25 or 80:20 to find the algorithm which best suits or fits for our system to achieve the highest accuracy.

3. The third step is to test the messages that are extracted from the chats or the blog which is posted on the blog by the user which causes bullying or use of abusive words to classify the post as toxic, severe toxic, obscene, threat, insult, identity hate and if found then the results will be saved in this step.

4. For Fake account Detection as shown in Fig

4.2 attributes fetched via Twitter API are given as input to the models and the model that will give us the best accuracy will be used that best fits our system and the results that we will get from the third step and the

fourth step will be sent and a report will be generated which will help to identify whether the user has bullied anyone and this will help them to take any action on them.

HARDWARE REQUIREMENTS

1. Intel I3 Processor onwards
2. 4GB Ram
3. 256 GB SSD

SOFTWARE REQUIREMENTS

1. Python 3.10 onwards
2. HTML
3. CSS
4. JS
5. Mysql

CONCLUSION

In this paper, we have presented an idea to find cyberbullying using ML methods. We examined our model in the first cyberbully by comparing the authenticity of the different algorithms with which we can conclude that the SGD (Stochastic Gradient Descent) division gives us the best 92% accuracy and from the results of the model tested on the fake account we found that the most accurate algorithm is Decision Tree accuracy of ~ 98.5% by using fully automated learning algorithms, we have eliminated the need for personal accounting for a fake account, which requires a lot of resources and is a time-consuming process.

FUTURE SCOPE

1. Visual Cyberbullying is more harmful than the written ones thus we also plan to develop ML classifiers detecting cyberbullying from videos and images. This goal could be reached through the contribution of scholars from different fields, because of the technical (i.e., difficulty to create datasets containing this type of entries) and legal (i.e., privacy issues) issues raised by sharing
2. It is also necessary to understand which impact these detection systems could have on users' everyday life. Future works will be challenged to combine these technological systems with the implementation of psychosocial interventions.
3. We can Restrict the access of the fake account users to the authentication servers or the sites.

REFERENCES

1. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE:synthetic minority over-sampling technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.

2. Jolliffe,Principal Component Analysis, 2002.View at: MathSciNet. S. Sperandei, “Understanding logistic regression analysis,” Biochemia Medica, vol. 24, no. 1, pp. 12–18, 2014.

3. R. Kohavi, “A study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” in Proceedings of the in 14th international joint conference on Artificial intelligence,pp. 20–25, 1995.

4. N. E. Willard, “Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, and Distress”, Research Press, 2007.