CYBERGUARD: FAKE PROFILE DETECTION USING MACHINE LEARNING

M.Vijaya Lakshmi , Aishwarya Rao , Kalyani Boddulah,
OjaswiCheekati, and Varsha Vadla

G. Narayanamma Institute of Technology and Science (For Women), Shaikpet,
Hyderabad, India

ABSTRACT

Social media platforms that foster high levels of user engagement, such as Facebook, Instagram, and Twitter, have huge impact on people's lives everywhere. They still have issues with false profiles, which may be created by automated systems, computer programs, or individuals. These false accounts facilitate illicit activities like phishing and identity theft as well as the dissemination of rumors. To solve this issue, our study employs many machine learning techniques to differentiate between authentic and fraudulent Twitter profiles. Analysis is done using key data including the quantity of friends and followers, activity trends, and more.

The algorithms such as Random Forest, XG Boost, LSTM, and neural networks emphasizes how crucial itis to choose important criteria while evaluating social media accounts. After training, the models generatea value of 1 for bogus profiles and 0 for real ones, allowing for the identification and the removal of bogus profiles to lessen cybersecurity risks.

KEYWORDS

Random forest, LSTM, XG Boost, Twitter, social media, fake profiles, machine learning, and neural networks.

1. INTRODUCTION

Nowadays, social interactions in the present age are mostly mediated by online social networks. Adding new friends and keeping up with their updates has been easier. Online social networks have an impact on many different domains, such asbusiness, education, employment, community activity, and research.

Research has been done to look at how these online social networks affect certain people. This allows teachers to quickly establish a rapport with their students and promote a positive learning atmosphere. These websites are become more and more familiar to instructors, who use them to make online lesson plans, assign assignments, hold conversations, and do other tasks that greatly enhance student learning. Employers may utilize these social networking sites to find and choose skilled, driven applicants.

Disseminating false information through social media is another issue. Conflicts arise when false accounts spread inappropriate and inaccurate information. The key aims of this research initiative are the following:

Similar to that, these phony accounts are made with the goal of acquiring followers.

Compared to other internet crimes, false accounts hurt individuals more. Since the user is now aware of it, it is imperative to detect a fake profile. bogus accounts are usually used online to spread spam, misleading information, and other bogus accounts.

This study looks at the technical efforts made in the past and present to detect false profiles. The majority of phony profiles are created with the intent to get more followers, send out spam and phishing attacks, and amass more followers. False accounts are outfitted with all the tools requiredto perpetrate crimes online. A significant risk of identity theft and data breaches is posed by false accounts. All user data is transferred to remote servers and can be used against users who visit the URLs supplied by these fraudulent accounts. False accounts that impersonate individuals or groups may also harm their reputation and reduce the number of likes and followers they receive. [1].

2. LITERATURE SURVEY

Today's online social networks are plagued by a number of problems, such as phony profiles and online impersonation. As of yet, no one has created a viable remedy for these problems. In this research, we aim to propose a unique model for computerized false profile early detection to protecthuman social life. We can also use our automatic detecting technology to make it easier for websitesto alter the wide variety of profiles, which is difficult to do manually. Many fake document emphasistechniques rely on the analysis of an individual's interpersonal organization profiles to identify the characteristics or combinations of them that aid in distinguishing between authentic and fraudulent documents. Specifically, after obtaining several features from the posts and profiles, a classifier that can identify fraudulent data is constructed using machine learning techniques.

The issue of phony social media accounts is tackled by Padmavati et al. through the use of Deterministic Finite Automata (DFA).

The current user and their friends' characteristics are analysed by the paper through the creation of an accounting pattern. Regular expressions are utilized to create the pattern, which is then used to match any friend requests. These expressions are based on several factors, such as the working and living communities. The disadvantage of this strategy is that it takes a while for someone with connections in several areas to generate regular expression. According to the authors, the method might be much more effective in practice. Mohammadreza et al. examined the problem of fake accounts on social networks in their paper using graph analysis and classification techniques. The preferred social media site was Twitter. Based on how similar the user's friends were, they devisedа plan. Prior to use Principal Component Analysis (PCA) to extract novel features from the networkgraph, the buddy similarity criteria is utilized. Once balanced, the data is fed into the classifier by the synthetic minority oversampling technique (SMOTE). Using the cross- validation process, a medium Gaussian SVM classifier was chosen because of its AUC of 1. The drawback of this approach is that fraudulent accounts can only operate inside the network in order to evade detectionthrough friend account browsing.[3]

According to the authors, a novel method may be developed in the future that would be able to identify if an account is authentic or fraudulent at the time of registration, or even before any user activity occurs on the network. In their research, Srinivas Rao et al. attempted to detect fraudulent profiles using machine learning and natural language processing (NLP). Facebook profiles served asthe writers' dataset. The three stages of the procedure include learning algorithms, PCA, and NLP preparation. Tokenization, stop word removal, stemming, and lemmatization were used in

their pre-processing of the data. To extract the basic values from the table, PCA is used. Later, profiles are classified using two machine learning algorithms called SVM NB.[9]

Following their method's examination, it was observed that applying these methods increased the Detection accuracy. String hini searched the markets for followers on Twitter. They categorize the customers of the company sectors and identify the traits of Twitter enthusiast[17] advertisements. The writers claim that two primary types of invoices that follow their clients are those resulting from stolen accounts and false accounts, whose providers fail to see that their following is expanding. Adherent marketplace clients might be politicians or celebrities who wishto appear to have a greater fan base, or they can be cybercriminals who want their files to look consistently authentic so they can distribute spam and viruses in an unpredictable way. Nancy Agarwal et al. used emotions such as happiness, sorrow, anger, fear, etc. to determine if the users were real or fake. They utilize posts made by Facebook users to test it out. The detection algorithmis trained using 12 emotion-based features. [11]

Based on the observation that real users post with a variety of emotions, whereas fictitious users, who are categorized as belonging to specific occupations, post with a fixed set of emotions, the author conducted this study. Furthermore, noise reduction is carried out. Lastly, machine learning methods including NB, JRip, SVM, and RF have been used to train the detection model. Ananya et al. used machine learning techniques in 2021 ICRITO to detect fraudulent social media profiles. The data was obtained from Kaggle, an open-source platform that stores data sets for public access. The data originated from Weibo, a well-known networking site and the Chinese equivalent of Twitter. Afterwards, they trained and compared five supervised learning models to see which produced higher test scores. [12] They selected the Random Forest Classifier and Gradient Boosting Classifier out of the five approaches since they performed better than the others. They ultimately decided on a random forest classifier since it produced results that were 1% better than those of the gradient boosting classifier. For their next effort, they hope to create an automated system that can learn and take additional traits than the ones that are included in this paper. Machine learning might be used to detect phony Instagram profiles, according to Preethi Harris et al. [13]

The Instagram profile information was obtained from the Kaggle website. To train the model, they used classification methods including SVM, KNN, RF, NB, and XG Boost. Upon calculating the accuracy and confusion matrix, the RF classifier was identified as the most appropriate model for the given data set, exhibiting the highest prediction accuracy. Afterwards, a data dictionary is filled in with the IDs of the fake profiles. In their 2018 study, Abhishek Narayanan et al. addressed the identification of phony profiles; Twitter was the source of their data collection.[16] In order to extract features from the data, machine learning methods such as SVM, RF, and LR were initially applied. This process ultimately produced a highly valued outcome for the random forest. Subsequently, randomforest classifier distinguished itself with 88% accurate prediction of bogus profiles on Twitter after undergoing some accuracy testing and confusion matrices. In comparison, it took the least amount of time to accomplish the goals and was more effective. Their next efforts will be focused on protecting consumers' privacy when utilizing social media. [14]

Mauro Conti et al. addressed potential solutions for the problem in their 2012 publication. The first step was determining if a given profile is representative of the actual user base. Then, they detected phony profiles using graph structures. Researchers looked at the user's connections, or friends list, to see if there were more random friends or a certain amount of mutual friends. To look for a phony profile, social network structural analysis is performed. [12]

3. PROPOSED METHOD

The suggested method represents a substantial improvement in social media platform fake account detection. with the use of gradient boosting methods like Random Forest, LSTM, and XGBoost. Even when certain inputs are absent, the model performs quite well. These algorithms, which use judgment trees built from variables such as spam comments, interaction rate, and simulated behavior, show better robustness against missing data than earlier techniques. XGBoost and Random Forest in particular are quite effective compared to other methods, especially when using default hyper parameter values. All things considered, the use of gradient boosting algorithms signifies a significant advancement in this field by improving the effectiveness of detecting fraudulent accounts.

3.1. Methodology

The research uses XG Boost, Random Forest, and characteristics taken from a multilayered neural network that focuses on profiles in order to construct the model to identify false profiles on social networking sites. The model can simply read the gathered characteristics because they are stored in a CSV file. The ultimate goal of the model's testing, training, and analysis is to determine if a profile is real or fraudulent. The reason Google Colab is used for model building is that it offers free GPU utilization. The Google Colab NVIDIA Tesla K80 GPU can run continuously for 12 hours, which makes it easier to identify fake profiles. The technique emphasizes both the framework's practical and aesthetic features.[4]

3.2. Dataset Collection and Preprocessing

The study's dataset comes from the MIB dataset and includes both fictitious and authentic profiles that are grouped according to various attributes. CSV format is used to store the dataset so that machine extraction is possible. To make using the program easier, large datasets in Microsoft Excel are converted to CSV format. To make sure the dataset is suitable for model input, preprocessing include managing missing data and removing category components.

3.3. Model Development and Deployment

Many machine learning techniques were taken into consideration for this research, with a primary focus on those pertinent to categorization tasks. Two core machine learning categories—classification and regression—each appropriate for various data kinds and intended results. When the dataset is limited and a binary output—such as true or false—is needed, classification is used. Regression, on the other hand, is utilized in situations where the data is continuous, such as in forecasting the weather. We focused on categorization methods in our study, using Random Forest and XGBoost in particular. These algorithms were selected based on how well they performed in classification tests and how well they held up against complicated datasets.[17]

Furthermore, We used recurrent neural networks (RNNs) with Long Short-Term Memory(LSTM) networks, which are very good at processing text and other sequential data. Though they were suggested as possible algorithms, Gaussian Naive Bayes and Logistic Regression were not used in our research because of their limited suitability in our particular situation. Our objective was to create a complete framework that could identify fake accounts on social media networks by utilizing the unique capabilities of XGBoost,Random Forest, and LSTM.

3.3.1. Random Forest

The random forest (also known as random-decision forest) ensemble learning algorithm is one example of this kind of approach. Due to its ease of application for both regression and classification issues, this approach finds application in machine learning. Like the Figure, the random forest does not rely just on one decision tree; instead, it uses predictions from each tree and predicts the result based on the votes of the majority of projections. However, random-forest produces a lot more choice trees than the decision tree technique does, and the result appears to be the sum of almost all of the decision trees produced. use the random forest method to find profiles. The model receives input data and outputs relevant results. Using the bootstrap aggregating procedure, the trees (FB) are fitted to the sample for the given set of 1 2, nX x x= and1 2, NY y y= answers. Periodically, a random sample is selected (B times). The following is the procedure used to determine the results for a particular sample (x') following training:

$$f = \frac{1}{B}\sum_{b=1}^{B} f_b(x')$$

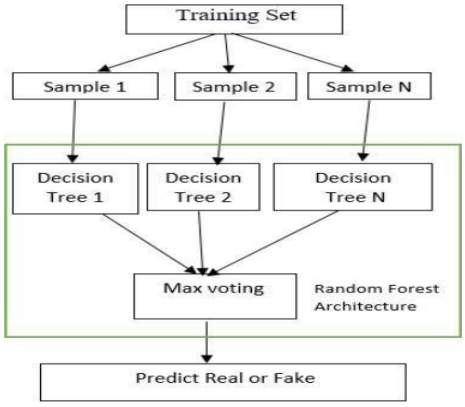

Figure 1: Random Forest Architecture

3.3.2. Extreme Gradient Boost

XG Boost is an additional ensemble learning method for regression. This approach is achievedby subsampling various Stochastic Gradient boosting settings. Random forest's drawback is that it performs best when all inputs are available or when there are no missing data. To get around this, the author uses a gradient boosting technique. As per the boosting process, F0(x) is first initialized. $f0\ (x) = \arg\min\gamma \sum L\ (yi\ ,\gamma\ )$

i=1

Next, the gradient of the loss function is determined repeatedly.

$$\left[\frac{\delta L\left(y_i, F(x_i)\right)}{F(x_i)}\right]$$

Lastly, the boosted model $Fm(X)$ is defined.

$Fm(X)$
$Fm\ (x) = Fm-1\ (x) + \gamma mhm\ (x)$

The pace of learning is α.
The component that multiplies is γn.

3.3.3. LSTM

Using tweet data, Long Short-Term Memory (LSTM), a specific kind of recurrent neural network, is used to assess social media profiles' credibility. LSTM networks are particularly good at identifying patterns and long-term relationships in sequential data, which makes them a good choice for textual content analysis, like tweet analysis. To prepare the twitter data, preprocessing procedures were used, such as removing identifier strings and using normalization strategies including changing tokens to lowercase and removing stop words. An embedding layer was then used to convert the tweet content into numerical vector representations.[11]

Effective analysis of the tweet data was made possible by the LSTM architecture's innate capacity to preserve and update internal state information over time, which was made possible by certain gating techniques. The LSTM ultimately produced a single 32-dimensional vector output, which was then refined further using layers activated by sigmoid functions. The resulting output was indicative of the social media profile's assessed level of trustworthiness. This application of the LSTM-based framework provides a strong method for evaluating profiles and demonstrates how well it processes sequential data to determine the authenticity and reliability of profiles.

4. RESULT ANALYSIS

A thorough analysis of the data is conducted, which includes a review of the efficacy of several machine learning algorithms in detecting false profiles. The report highlights the possible effects on social structures and the economy and highlights how important it is to address the threat posed by false accounts on social media sites. When comparing several machine learning algorithms for the purpose of detecting false profiles, research finds that XGBoost produces the highest accuracy rate of 95.05%.

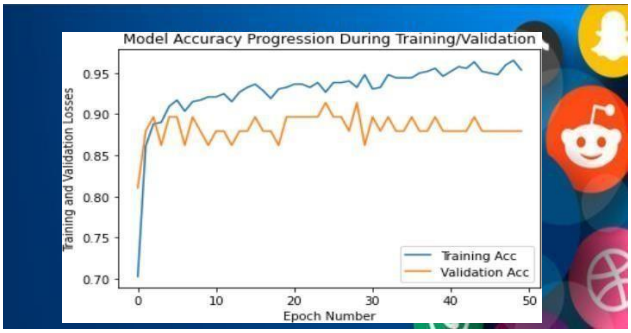

Figure 2.1 Training Progress – Accuracy

The following are the model accuracy and loss graphs for the trained neural network:

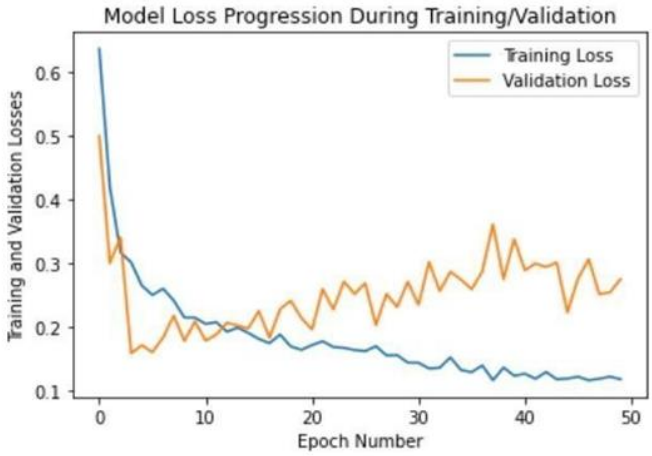

Figure 2.2 Training Progress - Loss

The results of 50 operating epochs are shown in the accuracy and loss graphs above. The accuracy initially varies across the course, peaking at 0.95 after starting at 0.70. Similar to this,the test dataset's loss graph begins at 1 and the validation data's at 5, with a local minimum of less than 0.2. Utilizing the binary cross-entropy function, the loss is calculated. The computer initially assigns a random weight to each attribute before giving it a precise weight.

Confusion matrices help visualize the results of the classification job and offer information about how well the model distinguishes between real and false profiles.

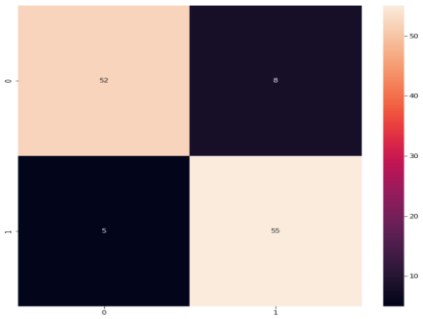

Figure 2.3. Confusion Matrix (Evaluation)

5. CONCLUSIONS

In conclusion, 'CyberGuard' is a sophisticated technology developed to tackle the threat of phony profiles on social networking sites. It uses a variety of sophisticated approaches to discern between real and fake accounts on social media networks by utilizing machine learning. The system's main objective is to examine particular characteristics, such the number of friends and followers, how frequently a profile is updated, and other relevant elements that add up to the legitimacy of a profile. The system makes use of a wide range of machine learning models, such as Random Forest, XG Boost, Long Short-Term Memory (LSTM) networks, and neural networks. It demonstrates a creative and successful strategy to address the widespread problem of phony profiles on social networking sites. It attainsa remarkable 95.05% accuracy rate. Its main goal is to promote confidence and authenticity in the digital realm by quickly detecting and resolving fraudulent accounts, hence boasting online security. This method fosters trust and authenticity in

the digital sphere by utilizing machine learning and data analysis to create a safer and more secure online environment.

ACKNOWLEDGEMENTS

We would like to extend our heartfelt gratitude to our esteemed professor, Dr. M. Vijaya Lakshmi, for her unwavering guidance, insightful feedback, and constant encouragement throughout the duration of this project. Her expertise and dedication were instrumental in the successful completion of our research.

We are also grateful to our ETM Department for providing us with the necessary resources and facilities to conduct this study. The support from our institution has been crucial in enabling us to explore and develop our ideas.

REFERENCES

[1] F. Benabbou, H. Boukhouima, and N. Sael, "Fake accounts detection system based on bidirectional gated recurrent unit neural network," International Journal of Electrical and Computer Engineering (IJECE), vol. 12, no. 3, p. 3129, 2022.

[2] H. G. Nagariya, N. Dhanotiya, S. Joshi, and S. Jain,"Identifying Fake Profile in Online Social Network"

[3] S. Bhambar, K. Khairnar, Y. Nikam, H. Shelar, and Y. K. Desai, "DETECTING FAKE ACCOUNTS ON SOCIAL MEDIA USING NEURAL NETWORK," International Research Journal of Modernization in Engineering Technology and Science, vol. 4, no. 5, 2022.

[4] V. Sarala and G. Sandhya, "Spammer Detectionand fake user Identification on Social Networks," Journal of Engineering Sciences, vol. 13, no. 8,2022.

[5] V. Singh, R. Shanmugam, and S. Awasthi, "Preventingfake accounts on social media using face recognition basedon convolutional neural network," Sustainable Communication Networksand Application, pp. 227–241, 2021

[6] Y. V. Biyani, "SPAM detection in social media using machine learning algorithm," International Journal for Research in Applied Science and Engineering Technology, vol. 9, no. 1, pp. 432–439,2021

[7] Instagram Fake Spammer Dataset - Kaggle.

[8] Instagram Fake and Automated Account Detection

[9] Fatih Cagatay Akyon; M. Esat Kalfaoglu

[10] Using Machine Learning to Identify False Identities:Bots vs. Humans. Van Der Walt, E. and Eloff, J. 2018.6540–6549 In IEEE Access. https://doi.org/10.1109/ACCESS.2018.2796018

[11] Ferrara and Kudugunta, S. (2018) Bot detection using deep neural networks. 312–322 in Information Sciences, 467. https://doi.org/10.1016/j.ins.2018.08.019

[12] Fake Profile Detection Methods in Large-Scale Online Social Networks: A Complete Study, D. Ramalingam and V. Chinnaiah, 2018.165–177 in Computers & Electrical Engineering, vol. 65. https://doi.org/10.1016/j.compeleceng.2017.05.02

[13] Y. Minoso, G. Hajdu, R. Lopez, M. Acosta, and A. Elleithy (2019)Artificial Neural Networks are used to spot fake profiles. Farmingdale, New York, 3–4 May 2019: 2019 IEEE Long IslandSystems, Applications, and Technologies Conference(LISAT).

[14] Using a black-list, Swe, M.M., and Myo, N.N. (2018) detected fake accounts on Twitter. Singapore, 6–8 June 2018, IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), 562–566. https://doi.org/10.1109/ICIS.2018.8466499

[15] Jie, H.J., and Wanda, P. (2020) DeepProfile: Utilizing Dynamic CNN to Detect Fake Profiles in Internet Social Networks.52, Article ID 102465 in Journal of Information Security and Applications. https://doi.org/10.1016/j.jisa.2020.102465.

[16] Identification of Fake Accounts on Twitter Using Hybrid SVM Algorithm, Kodati, S., Reddy, K.P., Mekala, S., Murthy, P.S., and Reddy, P.C.S., 2021.Article No. 01046 of E3 S Web of Conferences, page 309. https://doi.org/10.1051/e3sconf/202130901046

[17] Automated Fake Profile Detection Using Machine Learning on Instagram, Meshram, E.P.,

Bhambulkar, R., Pokale, P., Kharbikar, K., and Awachat, 2021. Journal of Scientific Research in Science and Technology: International, 8, 117– 127. https://doi.org/10.32628/IJSRST218330.

[18] Utilizing face recognition, Chakraborty, P., Muzammel, C.S., Khatun, M., Islam, S.F., and Rahman, S. (2020) developed an automatic student attendance system. IJEAT, 9, 93–99. https://doi.org/10.35940/ijeat.B4207.029320

[19] Evaluation of Eye-ball Movement and Head Movement Detection Based on Reading, Sayeed, S., Sultana, F., Chakraborty, P., and Yousuf, M.A. 2021.

AUTHORS

Aishwarya Rao, Kalyani Boddulah, Ojaswi Cheekati, and **Varsha Vadla** are B.Tech students at GNITS, Hyderabad, India. Their shared passion for cybersecurity and machine learning drove them to collaborate on this innovative project.

This research was conducted by a team of four dedicated students -Aishwarya Rao, Kalyani Boddulah, Ojaswi Cheekati, and Varsha Vadla under the guidance of their professor, Dr. M. Vijaya Lakshmi. Together, they developed the CyberGuard project, focusing on fake profile detection using machine learning.

Dr. M. Vijaya Lakshmi is a distinguished professor at GNITS with extensive experience in Machine Learning. Her mentorship and expertise were invaluablein guiding the team through the complexities of this research.