International Journal of Engineering and Management Research e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-3 (June 2023)
https://ijemr.vandanapublications.com https://doi.org/10.31033/ijemr.13.3.37

Fake Profile Detection on Social-Media

Shamim Ahmad[1] and Dr. Manish Madhava Tripathi[2]
[1]Student, Department of Computer science & Engineering, Integral University Lucknow, INDIA
[2]Professor, Department of Computer science & Engineering, Integral University Lucknow, INDIA

[3]Corresponding Author: khanshamimahmad981@gmail.com

Received: 29-05-2023 Revised: 13-06-2023 Accepted: 30-06-2023

ABSTRACT

In this generation, social media platforms such as Twitter, Instagram, Linkedln, and others play an important role in our everyday lives. The whole world is actively involved. However, it must also address the problem of false profiles. The majority of fake pattern are generated by human or robots or cyborgs built to spread for misinformation, data piracy and identity theft. Therefore, In this article, we will discuss a model name as Detection of fake profile on social media model, which will be differentiate between fake and real profile on twitter based on visible features like, friend counts, follower counts, status counts, and more by using various machine learning classification methods. The dataset will be used twitter profile and we will Taking the Machine learning classification model like, Neural Network (NN), Random Forest, XG-Boost, and LSTM for determining the authenticity of a social media profile and for used implementation language is Python3 along with all the required libraries like, pandas, NumPy, and Sklearn etc.

Keywords-- Random Forest, XG-Boost, NN, Social Media, Fake Profile

I. INTRODUCTION

Social media platforms have become an integral part of our daily lives, offering a means that connect with friends, family and people around the whole world. However, along with the benefits, social media platforms also face the challenges of dealing with fake profile or accounts. Fake profile can be created for various purpose such as, spreading misinformation, conducting scams, or engaging in cyberbullying etc. has turned out to be a problem.

In current scenario, there are 450 million active users in monthly and 259.9 million active users daily on Twitter. Facebook also adds about 500,00 new users every day and six new users every second. Loads of information are shared over twitter every single day.

The main motivations for creating false profiles include spamming, phishing, and gaining more followers. The fraudulent accounts are fully capable of committing online crimes. The fake accounts pose a serious hazard, including identity theft and data intrusions. These counterfeit accounts give people numerous URLs that, when transfer all of the user's data from each site visited to distant servers so they can be used against the user. Additionally, the fake profiles that are purportedly made on behalf of businesses or individuals can hurt their reputations and reduce the number of likes and follows they receive. Social media manipulation is a challenge in addition to all of these. Conflicts occur as a result of the distribution of false and unsuitable information from the untrue accounts. Fake profiles are frequently made under fictitious identities, and they spread defamatory and abusive posts and images to influence society or advance anti-vaccine conspiracy theories, among other things. These days, fraudulent personas are an issue on every social networking platform.

In this specific context here, we talk about detecting fake profiles on Twitter. We deploy various machine learning models. The dataset of twitter profiles E13 and TFP for genuine, and INT, TWT, FSF for fake is taken into use. To combat the creation of fake profiles, common defenses are:

1. Methods such as user verification must be incorporated while creating accounts on social media.
2. To detect abnormal activities, user behavior analysis must be employed. Bot detection solution consisting of analyzation based on real-time AI will be beneficial.
3. An automated bot protection tool must be used. As a technical contribution, we designed a multi-layer neural network model, a random forest model, an XG boost model, and an LSTM model. The mentioned models are supervised machine learning model.

II. LITERATURE BACKGROUND

This section encompasses the contingent of literature on the 'Detection of fake profile on social media'.

- The 'Rajdavindar' et.al, (2022) that is proposed the literature in the title of paper, which is the need for developing a system that is able to identify social sharing profile by the help of Graph and User based features that have been mostly used. And suggest to improve the accuracy by the using of Decision Tree (DT), Naives Baiye's(NB),and SVM Further. [2]

- In recent study 'Onkar kadam' et.al, (2021), that's evaluates a constructs model that is used to determine whether a given articles is true or not by the help of (NB) and (SVM)classification and to identify fake news with the help of semantic analysis, and to determine the truth and false in text format as well as how and why it occurs.
  The second think is that 'Shruti Shinde' et.al, (2021), that describe the existing solution for detecting malicious profile on social media by the purpose of identifying fake users and to find the Maximum accuracy 95% by the using of Random Walk Methods and evaluate the parameters such as, Precision, F1-Score, Recall, that is used to detecting the fake Accounts. [3,5]
- In another study 'Ananya B' et.al, (2021), To Evaluate a model and applied a lot of Machine Learning Methods to detect authenticity on Social Media Platform and to find the highest accuracy in detecting false profiles, So, the prediction made 93% fake and 96% genuine account correctly. [6]
- The 'T. Om P' et.al, (2021), that is used the Machine learning algorithms to detect the fake profile accurately and collecting the dataset like Training dataset (70%), Validating dataset (10%) and Testing Dataset (20%) for each features matrix fed into such as [LR, XGB, ADB, GBM] used to detect the fraudulent account and better achieving accuracy up to 95%. [7]
- The 'Jyoti Singh' et.al, (2019), which is proposed a system that can make simpler to connect with others in secure and effective way with the help of ML Algorithms such as, SVM, NB, that is used to identify fraudulent profile as high as around 95% more ever, and we can improve the fraudulent profile detection by using of NPL technique in future. [13]

So, In this circumstances, there are A lot of work has gone into building to detect fake or genuine account in the Machine Learning research.

III. EXISTING SYSTEM

Although there are a great deal factors that are acknowledged to comprehend the entire work process. No model has ever been able to assess them effectively. Fake misinformation spread detection decision support tools must be used to address this problem. The factor including:

1- Feature-based techniques- These techniques use different characteristics from user profiles and use machine learning algorithms to label them as authentic or false. The qualities of a feature may be written (such as a bio or list of interests), network-related (such as the number of friends or interactions), or visual (such as how good the quality of the profile photo or the results of a reverse image search). Support vector machines (SVMs), decision trees, and random forests are among the kinds of classifiers that can be applied.

2- Social graph analysis- is a method that focuses on examining the interpersonal relationships and network characteristics of user profiles. Comparing fake profiles to real ones, fake profiles typically have different network patterns. To find abnormalities in the social graph, features including centrality measures, clustering coefficients, and community discovery techniques can be used.

3- Hybrid strategies- To increase the precision of falsified information profile identification, several systems combine several techniques. For instance, a hybrid system might discover suspect profiles by combining network analysis and language analysis. To get better outcomes, ensemble approaches can be used, such as combining the results of different classifiers.

IV. METHODOLOGY

For detecting fake real profile on social media then this project categorised many step that display in our proposed methodology in Fig 1.

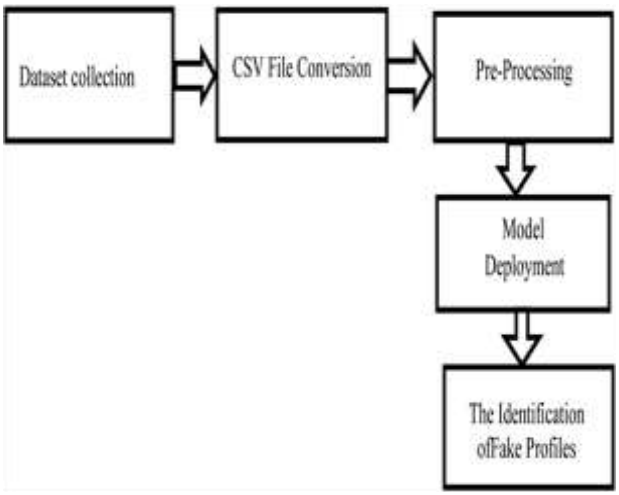

Figure 1: Flow Chart

For the detection of fake profile, we used a number of supervised techniques, each with a different level of accuracy, to detect bogus Twitter profiles. Based solely on observable qualities, each model can identify a fake profile. The accuracy and loss graphs for each of these supervised models are plotted using the same dataset.

- One of the techniques that falls under the category of ensemble learning techniques is random forest, sometimes referred to as

random-decision forest. Due to its ease of application in both classification and regression issues, this approach is employed in machine learning. Unlike the decision tree approach, random-forest generates many decision trees, and the final outcome is collectively the result of all the decision trees formed

Another ensemble learning approach for regression is called XG Boost. Using this, the stochastic gradient boosting algorithm is implemented. the knock-on effect of random forest is that it works best when all inputs are present, or when there are no missing values. We employ a gradient boosting approach to get around this.

V. EXPERIMENTAL RESULT

Dataset

Here is the dataset from MIB that we will be using. Another benefit is that it divided the dataset between real and fraudulent profiles. The dataset used consisted of TFP and E13 for authentic profiles and INT, FSF for false ones. Additionally, the information is kept in CSV file format for machine reading.

Charts and Graph

Here, all the process are obtained after testing and training the model, and were obtain the following results. The model accuracy, model loss vs the epochs graph plotted for the neural network LSTM, and model accuracy comparison for random forest, XG boost, and other methods.

The trained neural network model loss graph and accuracy model graph are as follows:

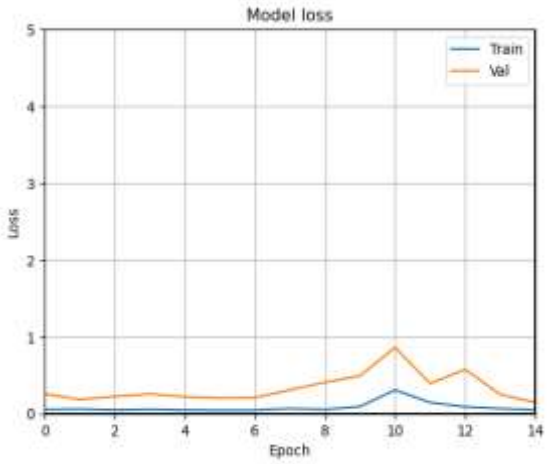

Figure 2: Model loss

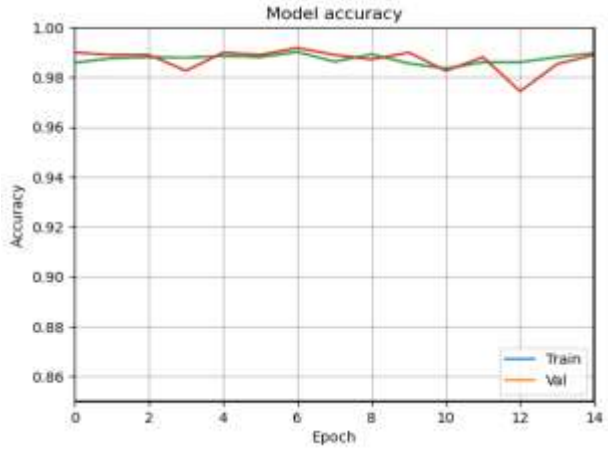

Figure 3: Model Accuracy

Accuracy Model Comparison Over Classifier

The Fig 4 describe the many model accuracies that are gained by the considered classifiers. Gaining high accuracy is the important thing to be noticed. The XG boost, which is equal to 99.66% produces the highest level of precision. Additionally, random forest and decision tree both have an accuracy of about 99 % And we now got Ada boost, at last.

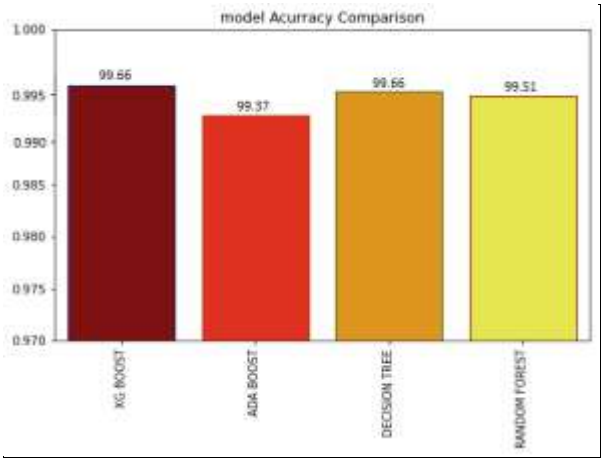

Figure 4: Different models Accuracy

VI. CONCLUSION AND FUTURE WORK

We used the Neural Network, Random Forest, and XG Boost machine learning techniques in this architecture to train our system to recognise deceptive Twitter profiles based on visible data. We finally come to the conclusion that the maximum accuracy comes from training, validating, and testing our models on the MIB data set exceeded is 99.66% using the XG Boost technique, then ANN, and random forest. Future work can be done by integrating photos of profiles with the categorical and numerical data and applying a CNN, additional work may be done, and also include additional

parameters, combining multiple models, and putting together a real time model might be produced better outcomes.

REFERENCES

[1] F. C. D. Da Silva, A. C. B. Garcia & S. W. M. Siqueira. (2022). A systematic literature mapping on profile trustworthiness in fake news spread In: *IEEE 25th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2022*, pp. 275–279. DOI: 10.1109/CSCWD54268.2022.9776232.

[2] R. Singh Boparai & R. Bhatia. Detection of fake profiles in online social networks-a survey. https://ssrn.com/abstract=4159087.

[3] O. Kadam & N. Surse. (2021). *Detection of fake social network account*. DOI: 10.51319/2456-0774.2021.4.0013.

[4] M. J. Ekosputra, A. Susanto, F. Haryanto & D. Suhartono. (2021). Supervised machine learning algorithms to detect instagram fake accounts. In: *4th International Seminar on Research of Information Technology and Intelligent Systems*, pp. 396–400. DOI: 10.1109/ISRITI54043.2021.9702833.

[5] S. Shinde & S. B. Mane. (2021). Malicious profile detection on social media: a survey paper. In: *9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*. DOI: 10.1109/ICRITO51393.2021.9596322.

[6] A. Bhattacharya, R. Bathla, A. Rana & G. Arora. (2021). Application of machine learning techniques in detecting fake profiles on social media. In: *9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*. DOI: 10.1109/ICRITO51393.2021.9596373.

[7] T. Om Prathyusha, N. S. Kumar & E. V. Priya. (2021). *Fake account detection using machine learning*. Available at: www.ijcrt.org

[8] S. D. Munoz & E. Paul Guillen Pinto. (2020). A dataset for the detection of fake profiles on social networking services. In: *International Conference on Computational Science and Computational Intelligence*, pp. 230–237. DOI: 10.1109/CSCI51800.2020.00046.

[9] *IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, Amity University, Noida, India.

[10] C. Zhang, S. Feng, X. Wang & Y. Wang. (2020). ZJU-Leaper: A benchmark dataset for fabric defect detection and a comparative study. *IEEE Trans. Artif. Intell., 1*(3), 219–232. DOI: 10.1109/TAI.2021.3057027.

[11] K. Sreenivasa Rao, S. Gutha, B. Deevena Raju, D. Rao & D. Bdeevena Raju. (2020). Detecting fake account on social media using machine learning algorithms system and method for mapping entities securely view project wireless networks communication view project detecting fake account on social media using machine learning algorithms. *Int. J. Control Autom., 13*(1), 95–100.

[12] S. P. Maniraj, G. Harie Krishnan, T. Surya & R. Pranav. (2019). Fake account detection using machine learning and data science. *Int. J. Innov. Technol. Explor. Eng., 9*(1), 583–585. DOI: 10.35940/ijitee.A4437.119119.

[13] J. Singh & M. Z. Khan. (2019). *Issue 6*, www.jetir.org.

[14] K. Shu, X. Zhou, S. Wang, R. Zafarani & H. Liu.(2019). The role of user profiles for fake news detection. *In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 436–439. DOI: 10.1145/3341161.3342927.

[15] S. Durga & P. Reddy. (2019). Fake profile identification using machine learning. *Int. Res. J. Eng. Technol.* Available at: http://www.sixdegrees.com.

[16] *3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*. IEEE, 2018.

[17] *International Conference on Advances in Computing and Communication Engineering (ICACCE)*. IEEE, 2018.

ACKNOWLEDGMENT

We extend our sincere gratitude to Shamim Ahmad and Dr. Manish Madhava Tripathi for their patient guidance, and useful critiques for this work. Their enthusiastic encouragement helped us to keep the progress on schedule.