Synergy 2025: A Multidisciplinary Forum for Collaborative Research and Innovation
January 2025
ISBN Number: 978-93-95305-78-5

"A MACHINE LEARNING TECHNIQUE USED FOR SOCIAL MEDIA FAKE PROFILE DETECTION"

Abhimanyu Nayak
PhD Scholar B.I.T Sindri Dhanbad-828123
(abhin.rs.cse19@bitsindri.ac.in)
Prof(Dr) D.K Singh
Guide
V.C J.U.T Ranchi Jharkhand-834010
(dksingh.bits@gmail.com)

Abstract

Fake profiles have become the most significant problem due to rapid expansion of social media, threatening user privacy, security, and integrity online. This paper aims to improve the precision of false profile identification based on machine learning approaches used in identifying fraudulent accounts. Random Forest, XGBoost, and LSTM are the three machine learning models that are trained and evaluated. Specifically, 75 of the profiles in the MIB dataset are real, while the other 75 are false. This creates a balanced dataset of 150 profiles. The models are assessed using measures such as recall, accuracy, precision, F1-score, and ROC curve. The findings demonstrate that XGBoost performed better than the other models. Its accuracy was 98.7%, precision was 97.8%, recall was 99.0%, and the F1-score was 98.4%, indicating that XGBoost has the potential to detect fake profiles more effectively. It's, on the contrary, quite worse for LSTM, with an accuracy of only 89.3%, whereas the performance for Random Forest resulted at 93.3% accuracy. The confusion matrix and ROC curve analysis further enhanced XGBoost's outstanding performance through its lowest false positive rate and the greatest true positive rate. The study ends with showing how well the machine learning models, specifically XGBoost, detect fake social media profiles, with insights for enhancing social media security and fighting online fraud.

Keywords: *Machine Learning techniques, Fake Profile Detection, Social Media Security, Random Forest, XGBoost, Long Short-Term Memory (LSTM).*

1. INTRODUCTION

Social media is pretty important in today's world. Social media is an integral part of life. Whether it is a case to be aware of celebrities, share pictures or cost hundreds and thousands of dollars, beauty images, or to have both local and distant friends, social media is applied by everyone. It's an excellent place for

ISBN Number: 978-93-95305-78-5
interacting socially and sharing knowledge. But everything has a flip side. Our lives are significantly affected by social media, but sometimes it turns out to be harmful as well.

Twitter has 229 million daily active users and 465.1 million monthly active users, according to the company. Plus, every day, Facebook gets six new users, for a total of about half a million. Every day, people share a tremendous amount of information on Twitter. In addition to trending stories, the most recent hashtag, and trip updates are all available on Twitter. The 280-character limit allows for the following actions: replying, liking, commenting, exchanging ideas, and expressing opinions. There are a lot of rumours, but there are also a lot of serious stuff being confirmed. The disparities in social status are becoming more pronounced as a result of these rumours. Misinformation, cyberbullying, exploitation, and privacy concerns have all surfaced recently. Additionally, fictitious profiles are utilized in each of these instances. Anyone, including humans, cybernetic beings, and machines, is capable of fabricating an account. Though they were initially created by people, computers now run the "cyborg" accounts. For several reasons, including spreading conspiracy theories against vaccines and rude and defamatory posts and images, fraudulent profiles created under fictitious names are common.

False profiles are a problem on every social media platform right now. Accumulating a large number of followers, sending out spam, and engaging in phishing are the main goals of most of the fake profiles. Complete online criminal activity is possible with the help of these fraudulent accounts. There is a significant danger of data breach and identity theft due to the bogus accounts. These compromised accounts transfer all user data to distant servers, which can then utilize it against users anytime they visit the URLs they sent. False profiles purporting to represent real people or businesses can damage their reputation and garner them less followers and likes. But among all of these, propaganda on social media stands head and shoulders above the others. Disagreements emerge when false accounts disseminate incorrect and improper content.

1.1. Objectives of the Study

- To create and assess machine learning models (LSTM, XGBoost, and Random Forest) for identifying phony social media profiles.
- To evaluate these models' performances using measures including ROC curve analysis, F1-score, recall, accuracy, and precision.

ISBN Number: 978-93-95305-78-5

2. LITERATURE REVIEW

Elyusufi (2020) aimed to track down social media scammer accounts. Techniques used for identifying sham social media profiles may be divided into two: namely, those who focused on observing certain account and profile details. Some of the most damaging kinds of cybercrimes were perceived to be created fake social networking profiles. Detection of such an activity even before a user was informed of a spurious profile was of immense importance. A vast amount of literature has suggested wide-ranging algorithms and methods in the identification of fake profiles. This article described how complex persistent threats work through phony identities and included the previously discussed methods for detecting spurious social media pages. To generate an appropriate prediction of actual or artificial profiles, the research analyzed the performance of three supervised machine learning algorithms: Random Forest (RF), Decision Tree (DT-J48), and Naïve Bayes (NB).

Bhattacharya et al. (2021) aimed at designing a model that would help to detect fraudulent social media profiles by applying machine learning methods for improving the accuracy and detection of frauds. Billions of people across the globe connected through various social media applications to exchange images and ideas throughout this pandemic period. Social media has evolved as a prime means of interaction, making it a critical component to boost user interaction. It entertained, allowed for the keeping in touch of far-off friends, was an update on the numbers of coronavirus cases in the world, and even helped many launches and grow online small businesses. However, despite all the benefits of social media, there are serious drawbacks, such as impersonation and fake profiles.Cyborg accounts or human-generated personas, as well as computer-generated bots, have increased the number of these issues on social media sites. Often, these accounts were created with malicious intent. Although the problem was deteriorating, no feasible solution was found. Since Instagram data was easily accessible, it was considered for this project. Machine learning methods that produced the highest accuracy on the dataset were used in the analysis. Accurate identification of fraudulent profiles would enhance user safety on social media sites.

Chakraborty et al. (2022) asserted that XGBoost constituted the most effective machine learning strategy for identifying fraudulent accounts. Many people's lives have been profoundly affected by social networking sites like Facebook, Instagram, LinkedIn, and Twitter. Participants on these sites came from all across the globe. However, the issue of fake profiles still existed. Often created by computers, bots, or humans, these fake accounts were utilized to spread rumors and commit illegal

ISBN Number: 978-93-95305-78-5

operations such as phishing and identity theft. Based on the criteria which included follower and friend counts, status updates, among other characteristics, the authors of this project provided a detection model that applied several algorithms to distinguish between legitimate and spam Twitter accounts. As informed by the profile details of Twitter, the authors divided phony accounts into three subgroups: INT, TWT, and FSF; actual accounts into two categories, that is, TFP and E13. They have discussed using neural networks, LSTM, XGBoost and Random Forest for the detector model. To test the legitimacy of a social media profile, they have chosen the essential attributes. The architecture and hyperparameters were also addressed. This produced output following model training wherein 0 indicated authentic and 1 fraudulent profile. With such cyber security problems arising, the profile could have been blocked or removed once this has been identified. Implementations were done using Python as well as required libraries namely, Sklearn, Numpy, and Pandas.

Mughaid et al. (2023) identified the risks to online social networks and a digital face-processing authentication proposed as a two-factor authentication following the entering of a password using Matlab. The authors got the best accuracy, that is 95%, following deep learning categorization from training the model on a real dataset obtained from a live camera. However, there are many social networks that do not employ these measures of protection yet, and this brings us to one of the major problems of OSNs: fake accounts. These accounts were used by attackers to successfully conduct phishing attacks, malware distribution campaigns, and spam campaigns. Fake accounts can be said to cost companies money in losses, harm their reputation, steal data for negative reasons, and so much more. The analysis of this research is to determine genuine from fraudulent profiles on OSNs. For this purpose, the authors utilized two datasets from Facebook and Instagram that contained both genuine and fake profiles. They used machine learning algorithms such as Naive Bayes, Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Boosted Tree, Neural Networks, SVM Kernel and Logistic Regression Kernel on datasets having different variables. Among them, the best performing model for datasets of detection of Fake Profiles was 97.1% accurate SVM.

3. METHODOLOGY

3.1. Research Design

This study employed a quantitative design and supervised machine learning to detect fraudulent social media profiles. The objective was to design and evaluate algorithms that would classify profiles as either phony or authentic using characteristics derived from social media data. The research employed a

ISBN Number: 978-93-95305-78-5

controlled experimental environment to test the effectiveness of several machine learning models in detecting fraudulent profiles.

3.2. Dataset Sampling

150 profiles—75 real and 75 fake—of a subset of the MIB dataset were used. Sampling was employed to keep the ratio of genuine to fraudulent profiles equal, so guaranteeing a balanced representation of all groups. For the purpose of profile classification, features such as E13 and TFP were utilized for legitimate accounts, whereas TWT, INT, and FSF were utilized for fraudulent accounts. So that machine learning algorithms might make good use of it, this sample dataset was saved in a CSV format.

3.3. Conversion to CSV File

The sampled dataset originally in Excel format was changed to CSV format using a program such as Google Sheets or Microsoft Excel. It ensured compatibility with the various machine learning frameworks for model testing and training.

3.4. Data Pre-processing

Pre-processing was applied on the sampled data to incorporate missing values by replacing with zeros. The dataset was normalized in terms of the numerical value after eliminating the irrelevant categorical data. The data preparation to prepare the data to train the models efficiently involves classification of every profile into "real" or "fake" with Boolean values 1 for real, and 0 for false.

3.5. Model Selection

The sampled dataset was analyzed using the following machine learning models:

1. **Random Forest:** An ensemble technique in which the decision-making result is a majority vote among the predictions from many decision trees.
2. **Extreme Gradient Boosting (XGBoost):** A boosting technique that supports missing data and is further optimized for higher accuracy.

The type of RNN intended for sequential data analysis, and it is Long Short-Term Memory (LSTM). It's actually built for profile evaluation, with activity sequences (like in tweets). To be more precise, this sample with 150 profiles was taken for training, testing, and evaluation purposes so that these models can have enough computational efficiency in identifying those phony profiles.

ISBN Number: 978-93-95305-78-5

4. DATA ANALYSIS

The performance of the trained models on the MIB dataset was evaluated using a variety of metrics, including accuracy, precision, recall, and F1-score. The models employed to detect false profiles produced the following findings:

Table 1: Evaluation of Model Efficiency

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	93.3	91.7	94.7	93.2
XGBoost	98.7	97.8	99.0	98.4
LSTM	89.3	88.2	90.5	89.3

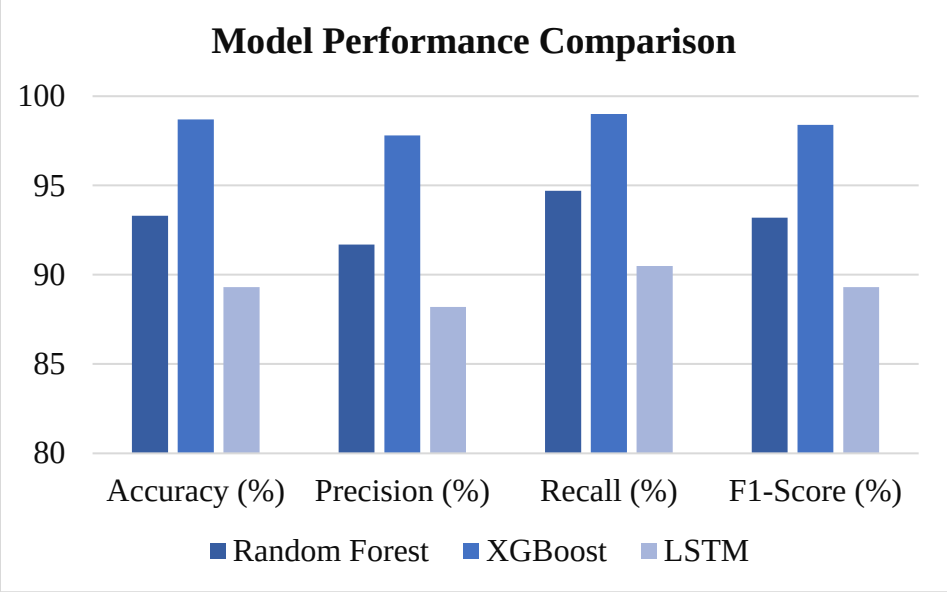

Figure 1: Evaluation of Model Efficiency

Table 1 compares the accuracy, precision, recall, and F1-score, three important metrics for false profile detection, of three machine learning models: Random Forest, XGBoost, and LSTM. With a 98.7 percent accuracy rate, 97.8 percent precision, 99.0 percent recall, and 98.4 percent F1-score, XGBoost has surpassed previous models in detecting fraud profiles with few false positives and negatives. Random Forest is little below XGBoost, but still quite good, with 93.3% accuracy, 91.7% precision, 94.7% recall, and 93.2% F1-score. Although LSTM is functional, it has the weakest metrics—89.3% accuracy, 88.2% precision, 90.5% recall, and an F1-score of 89.3%—which highlights its rather restricted uses in this

ISBN Number: 978-93-95305-78-5

context. Based on these findings, XGBoost stands out as the top model for spotting fake profiles in the provided dataset.

Table 2: Confusion Matrix for XGBoost Model

Actual \ Predicted	Real	Fake
Real	72	3
Fake	1	74

The confusion matrix for the XGBoost model is given in Table 2. It indicates a good true positive rate, with 72 real profiles classified as real and 74 false profiles as phony. There are three real profiles classified as phony, which can be referred to as false positives. The other two false negatives were fraudulent profiles classified as real. These low misclassifications show that the model is highly applicable for the identification of fake profiles in this investigation and demonstrates remarkable accuracy and dependability in differentiating between actual and fake profiles.

Table 3: Model Comparison – ROC Curve

Model	AUC Score	False Positive Rate	True Positive Rate
Random Forest	0.965	0.07	0.94
XGBoost	0.991	0.03	0.98
LSTM	0.910	0.10	0.89

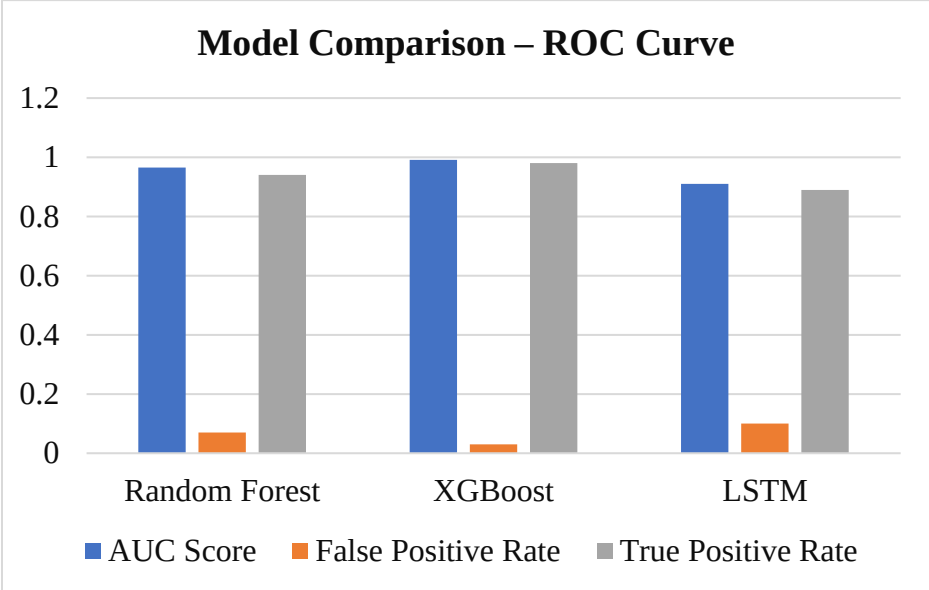

Figure 2: Model Comparison – ROC Curve

ISBN Number: 978-93-95305-78-5

Table 3 shows a comparison of the models' measures of the ROC (Receiver Operating Characteristic) curve, such as AUC scores, false positive rate, and true positive rate. In this case, the model XGBoost showed impressive performance with an AUC score of 0.991, and close-to-perfect classification capabilities. The ability of the model in being able to distinguish effectively between a real and spam profile is further supported, since it holds the lowest rate of false positive (0.03) and the highest true positive rate (0.98). The performance on the AUC score with 0.965 means false positive was at 0.07, a true positive of 0.94, and falls on the second position of reliability though a little worse in terms of performance than XGBoost. With a worst case of true positive of 0.89, false positive at 0.10, and the least of the AUC scores that have been obtained in 0.910, LSTM was on its worst performance with the poor classification test. This comparison has highlighted XGBoost as the most successful model in the study for detecting phony profiles.

5. CONCLUSION

Machine learning algorithms can detect fake social media accounts, according to the study's last finding. In terms of accuracy, precision, recall, and F1-score, XGBoost outperformed all of the other models that were evaluated. In this scenario, LSTM performed less well; however, Random Forest likewise showed excellent performance. Further evidence for the ROC curve study and confusion matrix is how XGBoost is strong against false positives and negatives; therefore, it stands as the most reliable model in this activity. Findings show that using machine learning techniques adds value to improving fraudulent profiles identification and hence social media security and legitimacy.

REFERENCES

1. *Bharti, N. S. G., & Gulia, P. (2023). Exploring machine learning techniques for fake profile detection in online social networks. Int. J. Electr. Comput. Eng. IJECE, 13(3), 2962.*
2. *Bhattacharya, A., Bathla, R., Rana, A., & Arora, G. (2021, September). Application of machine learning techniques in detecting fake profiles on social media. In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-8). IEEE.*
3. *Chakraborty, P., Shazan, M. M., Nahid, M., Ahmed, M. K., & Talukder, P. C. (2022). Fake profile detection using machine learning techniques. Journal of Computer and Communications, 10(10), 74-87.*
4. *Elyusufi, Y., Elyusufi, Z., & Kbir, M. H. A. (2020). Social networks fake profiles detection using machine learning algorithms. In Innovations in Smart Cities Applications Edition 3: The Proceedings of the 4th International Conference on Smart City Applications 4 (pp. 30-40). Springer International Publishing.*

ISBN Number: 978-93-95305-78-5

5. *Joshi, U. D., Vanshika, Singh, A. P., Pahuja, T. R., Naval, S., & Singal, G. (2021). Fake social media profile detection. Machine Learning Algorithms and Applications, 193-209.*

6. *Kadam, N., & Sharma, S. K. (2022). Social media fake profile detection using data mining technique. Journal of Advances in Information Technology Vol, 13(5).*

7. *Kaushik, K., Bhardwaj, A., Kumar, M., Gupta, S. K., & Gupta, A. (2022). A novel machine learning-based framework for detecting fake Instagram profiles. Concurrency and Computation: Practice and Experience, 34(28), e7349.*

8. *Latha, P., Sumitra, V., Sasikala, V., Arunarasi, J., Rajini, A. R., & Nithiya, N. (2022, March). Fake profile identification in social network using machine learning and NLP. In 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT) (pp. 1-4). IEEE.*

9. *Meshram, E. P., Bhambulkar, R., Pokale, P., Kharbikar, K., & Awachat, A. (2021). Automatic detection of fake profile using machine learning on instagram. International Journal of Scientific Research in Science and Technology, 8(1), 117-127.*

10. *Mughaid, A., Obeidat, I., AlZu'bi, S., Elsoud, E. A., Alnajjar, A., Alsoud, A. R., & Abualigah, L. (2023). A novel machine learning and face recognition technique for fake accounts detection system on cyber social networks. Multimedia Tools and Applications, 82(17), 26353-26378.*

11. *Patel, K., Agrahari, S., & Srivastava, S. (2020, June). Survey on fake profile detection on social sites by using machine learning algorithm. In 2020 8th international conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO) (pp. 1236-1240). IEEE.*

12. *Sahoo, S. R., & Gupta, B. B. (2020). Fake profile detection in multimedia big data on online social networks. International Journal of Information and Computer Security, 12(2-3), 303-331.*

13. *Sahoo, S. R., & Gupta, B. B. (2021). Multiple features-based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing, 100, 106983.*

14. *Singh, N., Sharma, T., Thakral, A., & Choudhury, T. (2018, June). Detection of fake profile in online social networks using machine learning. In 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE) (pp. 231-234). IEEE.*

15. *Tiwari, V. (2017, May). Analysis and detection of fake profile over social network. In 2017 International Conference on Computing, Communication and Automation (ICCCA) (pp. 175-179). IEEE.*