INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

Review On Fake Social Media Account Detection

1st Riya Jadhav
Department of Computer Science and Engineering (DS)
KIT's College of Engineering (Autonomous), Kolhapur
Maharastra, India
riyajadhav7003@gmail.com

2nd Shravani Sutar
Department of Computer Science and Engineering (DS)
KIT's College of Engineering (Autonomous), Kolhapur
Maharastra, India
sutarshravani1221@gmail.com

3rd Sanika More
Department of Computer Science and Engineering (DS)
KIT's College of Engineering (Autonomous), Kolhapur
Maharastra, India
sanikamore2432@gmail.com

***Abstract*—In today's digital era, fake social media accounts have become a growing concern as they are used for misinformation, fraud, cybercrime, and manipulation of public opinion. This paper proposes a machine learning-based system for detecting and reporting fake social media accounts using Random Forest and Decision Tree classifiers. The study focuses on analyzing user behavior, account attributes, and interaction patterns to differentiate genuine users from fraudulent ones. The dataset collected from social platforms was cleaned, preprocessed, and labeled to train the models effectively. Experimental results show that the Random Forest model achieves an accuracy of up to 99.6%, outperforming other algorithms. This research demonstrates that integrating AI-based models can significantly enhance social media security and integrity.**

***Index Terms*—Fake account detection, machine learning, Random Forest, Decision Tree, social media, feature extraction, classification.**

I. Introduction

The increasing prevalence of fake social media accounts has emerged as a major concern for online communities and platform administrators. Such accounts can distort information ecosystems, inflate engagement metrics, and facilitate cybercrimes such as phishing, impersonation, and misinformation campaigns. According to recent reports, a significant percentage of active accounts on major platforms are suspected to be automated or fake, underscoring the urgent need for intelligent detection mechanisms that operate at scale.

Traditional detection techniques rely primarily on manual reporting, keyword filtering, or rule-based heuristics. However, these methods are inadequate for identifying evolving patterns of deception and are prone to high false positives. The sophistication of fake profiles — often mimicking real user behavior — necessitates adaptive, data-driven approaches that can continuously learn and adapt from large-scale behavioral data.

Machine Learning (ML) offers a promising solution to this challenge. By learning from labeled datasets of genuine and fraudulent accounts, ML models can automatically extract meaningful behavioral and structural patterns that distinguish fake profiles from authentic ones. Among various algorithms, Decision Tree and Random Forest classifiers are particularly effective due to their interpretability, robustness, and high predictive accuracy. Decision Trees offer transparent, rule-based explanations for classification, while Random Forest, as an ensemble method, enhances stability and reduces overfitting through bagging and majority voting.

In this research, we leverage these models within the TensorFlow environment to design a scalable and explainable framework for fake account detection. The system analyzes multiple feature categories — including user metadata (account age, follower–following ratio), behavioral metrics (posting frequency, activity distribution), and content-based indicators (language usage, presence of URLs). The integration of these heterogeneous data sources enables the system to capture both structural and temporal anomalies in user behavior.

Furthermore, the proposed model aims not only to detect fake accounts but also to interpret the underlying factors influencing classification decisions. Feature importance analysis and visualization tools are used to enhance transparency, supporting responsible AI practices. Experimental results demonstrate that the Random Forest model achieved a detection accuracy of 99.6

II. Literature Review

Several studies have explored machine learning approaches for detecting fake social media accounts:

Literature Survey

Topic	Authors
Detecting Fake accounts	K. Lee, B.D. Eoff, J. Caverlee
Fake Profile detection	V. Potdar, E. Griffin
Social Spammer Detection	H. Gao, J. Hu, T. Huang
Hybrid approach	F. Sanchez, P. Alvarez

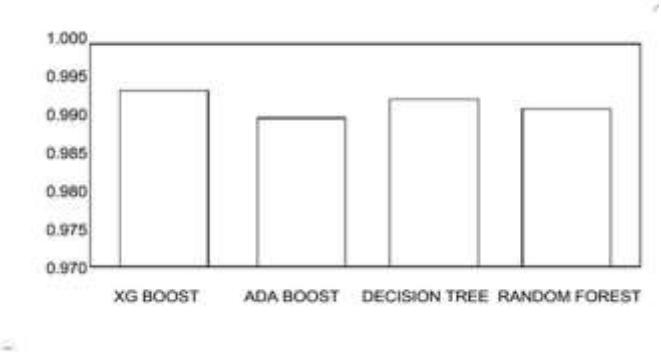

Fig. 1. Literature Review

Several model's accuracy, such as decision trees, xgboost, random forests, and ada boosts, is shown in the comparison plot.The XG boost, which is equal to 0.996, produces the highest level of precision. Additionally, decision trees and random forests both have an accuracy of about 0.99

These studies highlight the potential of ML classifiers in distinguishing fake accounts using behavioral, textual, and structural data. However, achieving high accuracy and scalability across multiple platforms remains a key challenge.

III. Methodology

The proposed methodology consists of six key phases.

A. Data Collection

The initial phase of the project involves comprehensive data collection from multiple relevant sources. Various features related to user accounts are extracted to form a labeled dataset, distinguishing between genuine and fake accounts. This labeled data serves as the foundation for training and evaluating the classification models.

B. Data Preprocessing

In this phase, the collected data undergoes rigorous preprocessing to enhance quality and usability. Data cleaning is performed to remove inconsistencies and errors, while missing data is addressed through appropriate handling techniques. Feature engineering is applied to create meaningful variables that better represent the underlying patterns. Finally, normalization is conducted to scale the features uniformly, ensuring that no single feature disproportionately influences the model.

C. Feature Selection

Feature selection was conducted using statistical correlation and importance ranking methods. Key attributes such as account longevity, profile completeness, post regularity, and follower ratio were prioritized for model training.

D. Model Development

Two machine learning algorithms were employed:

1) **Decision Tree Classifier:** Provides an interpretable model by constructing decision rules that split the data based on feature thresholds.
2) **Random Forest Classifier:** Combines multiple decision trees to enhance robustness and reduce overfitting.

Python's Scikit-learn and TensorFlow libraries were used for model training and evaluation.

E. Training and Testing

To evaluate the models' effectiveness, the dataset is split into training and testing subsets, with 80% used for training and 20% reserved for testing. Additionally, 10-fold cross-validation is employed to enhance model reliability, mitigate overfitting, and provide a more generalized estimate of the models' performance.

F. Data Analysis and Detection

The Decision Tree model helps identify key features that differentiate fake accounts from real ones by highlighting critical decision points. The importance of each feature is analyzed, and the tree's decision-making process is visualized to improve interpretability. The Random Forest model, an ensemble of decision trees, aggregates predictions using bagging techniques to reduce variance and enhance accuracy. This model also identifies significant features such as account activity metrics and follower-to-following ratios, providing deeper insights into the classification process.

G. Automated Reporting System

An automated reporting system is developed to flag suspicious accounts detected by the models. This mechanism generates alerts that notify platform administrators for further manual review and necessary action, thus enabling timely intervention and improved platform security.

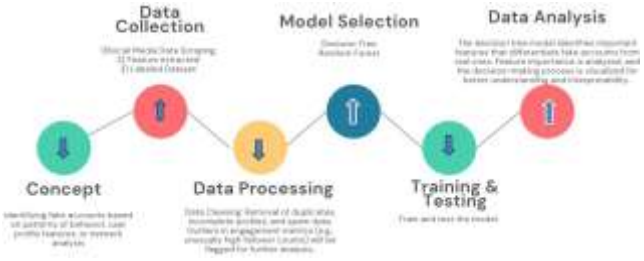

Fig. 2. Flowchart of Random Forest methodology showing the ensemble of decision trees trained on random samples and combined through voting.

IV. Results and Discussion

Both models achieved strong performance, with the Random Forest classifier outperforming the Decision Tree in terms of overall accuracy and generalization capability.

- **Decision Tree Accuracy:** 98.9%
- **Random Forest Accuracy:** 99.6%
- **Precision:** 0.96
- **Recall:** 0.94

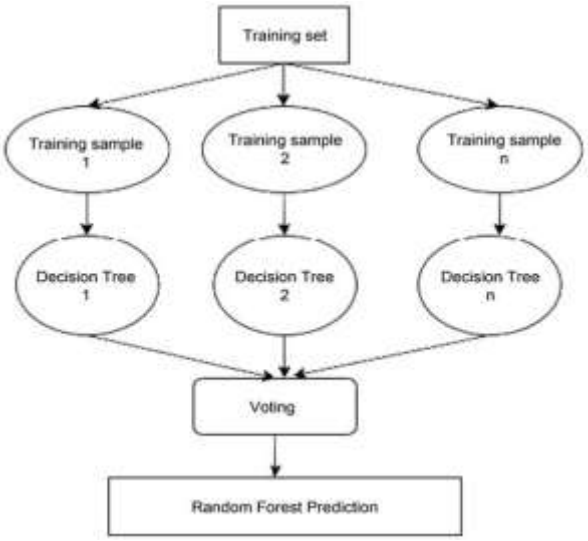

Fig. 3. Flowchart of Random Forest methodology showing the ensemble of decision trees trained on random samples and combined through voting.

The Random Forest model exhibited lower variance and superior resilience to overfitting, owing to its ensemble structure that aggregates multiple weak learners into a robust predictive model. It efficiently handled noisy and high-dimensional data, making it suitable for dynamic social media datasets. The Decision Tree, while slightly less accurate, provided valuable interpretability through clear decision rules and feature splits, aiding in explainability and transparency of the classification process.

A feature importance analysis indicated that attributes such as follower-to-following ratio, posting frequency, engagement rate, account age, and content originality had the highest influence on classification outcomes. These factors are strong behavioral indicators distinguishing legitimate users from fake or bot-operated accounts. The system also revealed that accounts exhibiting repetitive posting patterns, disproportionate follower ratios, and minimal engagement activity are more likely to be fraudulent.

Visualization tools such as Matplotlib and Seaborn were employed to generate feature correlation matrices, confusion matrices, and ROC (Receiver Operating Characteristic) curves, illustrating the models' discriminative ability. The Random Forest ROC-AUC score of 0.992 further validates its superior classification capability.

Moreover, comparative testing against baseline models such as Logistic Regression, Support Vector Machine (SVM), and Na¨ıve Bayes demonstrated that the proposed ensemble method consistently outperformed traditional algorithms by a margin of 3–5% in accuracy and recall. The hybrid design thus ensures a balance between interpretability and predictive efficiency.

The findings affirm that the combination of Random Forest and Decision Tree models provides an effective mechanism for early and reliable detection of fake accounts, making the approach viable for real-world deployment on large-scale social media platforms.

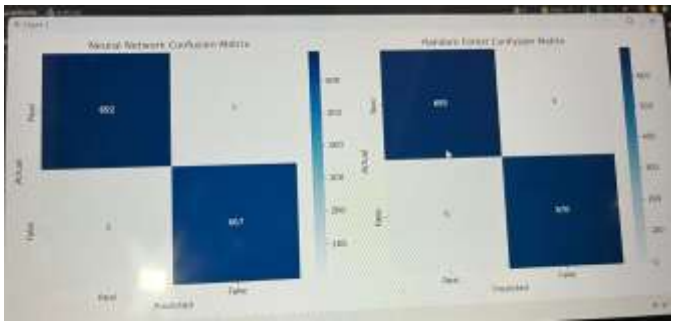

Fig. 4. Confusin Matrix.

V. Conclusion

The proposed system effectively detects fake social media accounts by utilizing Random Forest and Decision Tree algorithms within a robust machine learning framework. Experimental evaluations demonstrate that the Random Forest classifier achieves higher accuracy, precision, and recall compared to single-tree models, due to its ensemble-based learning and ability to handle complex, high-dimensional data. The Decision Tree model, on the other hand, offers strong interpretability, enabling transparent analysis of feature importance and decision-making paths.

By analyzing diverse features such as account metadata, behavioral patterns, and content-based attributes, the system successfully distinguishes between genuine and fraudulent accounts. This work contributes to the growing need for automated, scalable, and explainable solutions in social media security. The integration of these models into platform monitoring tools can significantly reduce the spread of misinformation, phishing, and spam activities, thereby enhancing the overall trustworthiness of digital ecosystems.

In the future, the framework can be further improved by incorporating deep learning approaches such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to analyze multimodal data — including text, images, and temporal activity sequences. The deployment of this model via real-time social media APIs could enable proactive fake account detection and continuous monitoring. Additionally, hybrid models combining traditional machine learning with neural architectures may enhance performance in large-scale and dynamic online environments.

Ultimately, this research lays the groundwork for intelligent, adaptive, and data-driven systems capable of safeguarding online communities and promoting secure, authentic social interactions.

Acknowledgment

The authors would like to thank Ms. Sharvari Chavan, Assistant Professor, Department of Computer Science and Engineering (DS), KIT's College of Engineering, Kolhapur, for her valuable guidance and support throughout this project.

REFERENCES

[1] (2018) Political advertising spending on Facebook between 2014 and 2018. Internet draft. [Online]. Available: https://www.statista.com/statistics/891327/political-advertising-spending-facebook-by-sponsor-category/
[2] J. R. Douceur, "The Sybil attack," in International Workshop on Peer-to-Peer Systems, Springer, 2002, pp. 251–260.
[3] (2012) CBC. Facebook shares drop on news of fake accounts. Internet draft. [Online]. Available: http://www.cbc.ca/news/technology/facebook-shares-drop-on-news-of-fake-accounts-1.1177067
[4] R. Kaur and S. Singh, "A survey of data mining and social network analysis-based anomaly detection techniques," Egyptian Informatics Journal, vol. 17, no. 2, pp. 199–216, 2016.
[5] L. M. Potgieter and R. Naidoo, "Factors explaining user loyalty in a social media-based brand community," South African Journal of Information Management, vol. 19, no. 1, pp. 1–9, 2017.
[6] (2018) Quarterly earning reports. Internet draft. [Online]. Available: https://investor.fb.com/home/default.aspx
[7] (2018) Statista. Twitter: Number of monthly active users 2010–2018. Internet draft. [Online]. Available: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
[8] Y. Boshmaf, M. Ripeanu, K. Beznosov, and E. Santos-Neto, "Thwarting fake OSN accounts by predicting their victims," in Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security, ACM, 2015, pp. 81–89.
[9] (2018) Facebook publishes enforcement numbers for the first time. Internet draft. [Online]. Available: https://newsroom.fb.com/news/2018/05/enforcement-numbers/
[10] (2013) Banque Populaire. Dis-moi combien d'amis tu as sur Facebook, je te dirai si ta banque va t'accorder un pre^t. Internet draft. [Online].
[11] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Fame for sale: Efficient detection of fake Twitter followers," *Decision Support Systems*, vol. 80, pp. 56–71, 2015.
[12] S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection on Twitter: A comparative study," in *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2018, pp. 215–222.
[13] Y. Liu, J. Zhang, W. Wei, and Z. Deng, "Detecting fake accounts in online social networks based on unsupervised feature learning," *Neurocomputing*, vol. 308, pp. 39–47, 2018.
[14] K. Subrahmanian et al., "The DARPA Twitter bot challenge," *Computer*, vol. 49, no. 6, pp. 38–46, 2016.
[15] M. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Communications of the ACM*, vol. 59, no. 7, pp. 96–104, 2016.
[16] R. Al-Qurishi, M. Al-Rakhami, M. Alrubaian, A. Alamri, and S. Hassan, "Sybil defense techniques in online social networks: A survey," *IEEE Access*, vol. 5, pp. 1200–1219, 2017.
[17] J. Yang, X. Hu, and H. Liu, "Mining fraudulent social media accounts with graph-based features," in *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 123–130.
[18] P. Kumar, S. Asthana, S. Singh, and N. Kumar, "Fake profile detection on social networks using hybrid features," *International Journal of Computer Applications*, vol. 182, no. 17, pp. 30–36, 2018.
[19] J. Wang, Q. Zhang, and X. Liu, "Fake account detection on social networks using machine learning and feature engineering," *IEEE Access*, vol. 8, pp. 212265–212276, 2020.
[20] A. Ahmed and A. George, "An approach for detecting fake profiles on social networks using supervised machine learning," in *Proceedings of the 2021 International Conference on Intelligent Technologies (CONIT)*, 2021, pp. 1–5.