Cost-Sensitive Fake Profile Detection in Online Social Networks Using Random Forest Feature Selection and LightGBM

Hedia Zardi
Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia
h.zardi@qu.edu.sa (corresponding author)

Raneem Alreshoodi
Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia
392207752@qu.edu.sa

Received: 3 October 2025 | Revised: 23 October 2025 and 8 November 2025 and | Accepted: 10 November 2025

 DOI: https://doi.org/10.48084/etasr.15297

ABSTRACT

The proliferation of fake profiles on Online Social Networks (OSNs) presents serious risks to privacy, security, and trust. Traditional detection methods often struggle with large-scale data and fail to keep up with evolving tactics of malicious actors, highlighting the need for scalable, interpretable machine learning solutions. This study introduces a cost-sensitive and interpretable framework for identifying fake profiles by combining Random Forest (RF) feature selection with advanced gradient boosting models, specifically eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM). The framework was tested on the MIB Twitter dataset using a user-level split to prevent data leakage and ensure a realistic evaluation. Results show that LightGBM achieved the highest Cost-Sensitive Accuracy (CSA) of 0.96, surpassing XGBoost by 2.8% and RF by 4.6%, while training approximately 20% faster than XGBoost. These findings demonstrate that LightGBM strikes the best balance among predictive accuracy, cost sensitivity, and computational efficiency. By focusing on CSA as a key performance metric, this work highlights the importance of reducing false negatives in OSNs, where undetected fake accounts can cause more harm than false positives. Overall, the proposed framework offers a practical, scalable, and interpretable solution for real-time detection of fake profiles on online social networks. This study demonstrates that combining feature selection with cost-sensitive boosting effectively improves trust and security on large online social platforms.

Keywords-fake profile detection; OSNs; CSA; user-level leakage; machine learning; scalable fake account detection

I. INTRODUCTION

OSNs like Twitter, Facebook, and Instagram have revolutionized digital communication by enabling users to share information, build communities, and express opinions worldwide. However, their rapid growth has also facilitated the creation and misuse of fake profiles, which are often exploited for malicious activities such as identity theft, misinformation, spamming, fraud, and political manipulation. The presence of these accounts undermines user trust, distorts public discourse, and poses serious security risks for both individuals and platforms [1, 2]. Traditional detection methods, such as manual verification and rule-based filtering, have become less effective as malicious actors become more sophisticated and adaptable. Consequently, research has shifted to Machine learning (ML) techniques, which offer scalability, automation, and adaptability in detecting fraudulent behavior [1-4]. Public datasets have further supported ML-based fake-profile detection [5]. Early studies using traditional classifiers like Support Vector Machines (SVMs), Decision Trees (DTs), and RFs showed reasonable accuracy. However, they often faced challenges related to high-dimensional behavioral data, redundancy, and limited interpretability. Recent research has focused on ensemble and boosting algorithms such as XGBoost and LightGBM, which improve accuracy and efficiency by modeling complex nonlinear relationships in user behavior. Still, these models can encounter such problems as feature redundancy, overfitting, and uniform treatment of misclassification costs [6].

To address these limitations, this study introduces a cost-sensitive and interpretable detection framework that combines RF-based feature selection with LightGBM classification. The approach enhances interpretability, reduces redundancy, and emphasizes CSA, a metric that penalizes false negatives more heavily to reflect the real-world costs of undetected fake accounts. The framework is tested on the MIB Twitter dataset using a user-level split to prevent data leakage, ensuring a

realistic and unbiased evaluation. In summary, this work offers a hybrid framework described above, a CSA-based assessment aligned with OSN security priorities, and an empirical comparison of Logistic Regression (LR), RF, XGBoost, and LightGBM that highlights trade-offs among accuracy, efficiency, and interpretability. The growing number of fake profiles on OSNs has attracted significant research attention. These profiles are frequently used to spread misinformation, commit fraud, or sway online communities. Researchers have explored this issue from different perspectives, which can generally be categorized into three groups: fake profile identification, machine learning techniques, and feature selection with boosting algorithms. This classification agrees with recent survey results on fake profile detection [2, 7].

A. Fake Profile Detection in OSNs

Fake profiles can be classified into compromised accounts, cloned identities, sockpuppets, Sybil accounts, and bots. Compromised accounts are legitimate profiles that are hijacked without their owners' knowledge, while cloned identities duplicate existing users to deceive others across platforms. Sockpuppets manipulate discussions by pretending to be independent users, whereas Sybil accounts are created in large numbers to skew reputation systems. Lastly, bots, ranging from simple spam bots to advanced social bots, pose increasing threats by mimicking human behavior. Traditional static detection methods are becoming less effective as these threats evolve, underscoring the need for dynamic, adaptable models [8, 9].

B. Machine Learning Approaches for Fake Profile Detection

Machine learning has become the leading approach for detecting fake profiles on online social networks because of its ability to recognize hidden patterns in user activity and behavior. Supervised methods like SVMs, DTs, and RFs have been widely used, often providing strong baseline results for distinguishing genuine accounts from fraudulent ones. Recently, research has expanded to include advanced ensemble models and deep learning techniques, which excel at handling high-dimensional data and complex feature interactions [10]. At the same time, graph-based learning has gained interest because fake profiles often exploit the structural properties of social networks. Methods like random walks [3], graph embeddings, and community detection [11, 12] are increasingly integrated into detection frameworks. These approaches are beneficial for identifying sophisticated malicious actors, such as Sybil accounts and coordinated botnets, that cannot be reliably detected using only user-level features [13].

C. Feature Selection Using Random Forest

Feature selection is essential for improving classifier performance by reducing dimensionality, computational overhead, and overfitting. RF-based methods are especially popular because RF provides feature importance scores intrinsically, which can guide selection [14]. Beyond the basic usage, researchers have explored hybrid approaches that enhance RF-based selection. For example, authors in [15] proposed a model-free feature selection method that combines feature screening with random forest recursive elimination, enabling scalability to ultra-high-dimensional data while boosting efficiency. In another direction, authors in [16] introduced the all-relevant feature selection paradigm, which uses RF as a wrapper over synthetic contrast features to identify both strong and weak predictors. These developments suggest that RF-based feature selection remains a robust foundation for integration with boosting methods in fake profile detection.

D. Boosting Algorithms: LightGBM and XGBoost

Boosting algorithms, especially XGBoost and LightGBM, have become leading techniques across many classification tasks due to their capacity to capture complex nonlinear patterns. LightGBM often trains more quickly and uses less memory than XGBoost, while still delivering comparable accuracy [17]. Recent research has expanded these approaches into fraud and anomaly detection. For example, in [18], researchers combined BERT embeddings with LightGBM for fake news detection, and in [19], graph features were used with boosting models to identify financial fraud. These findings demonstrate the adaptability and scalability of boosting algorithms, making them highly effective for identifying fake profiles in extensive online social networks. Recent research also highlights the importance of cost-sensitive learning for fraud and intrusion detection, where misclassifying malicious entities has a different impact than misclassifying legitimate ones. Authors in [20] introduced a cost-sensitive boosting strategy that adjusts learning weights to focus on correctly identifying minority or high-risk classes, enhancing performance in imbalanced situations. Similarly, authors in [21] showed that incorporating misclassification cost factors into detection models is vital for intrusion detection systems, as missing malicious activity can cause severe operational damage. These studies emphasize the practical need to address cost asymmetry and support the use of CSA as a key evaluation metric for detecting fake profiles.

E. Research Gap and Contribution

While boosting algorithms are powerful, they need careful hyperparameter tuning and can be prone to overfitting on small datasets. Their interpretability is also somewhat limited compared to simpler models like RF. Therefore, using Random Forest feature selection before applying boosting methods provides a good balance of accuracy, robustness, and interpretability. In addition, although boosting algorithms and feature selection methods have achieved promising results, many prior studies have evaluated models solely on accuracy, without considering the real-world costs of misclassification or user-level leakage. Unlike prior studies that rely only on accuracy, our work incorporates CSA and leakage-free validation, making the evaluation more realistic and practically relevant.

Table I presents representative studies, their datasets, methods used, and key performance metrics, enabling precise comparison with our approach and highlighting recent progress in fake profile and fraud detection.

TABLE I. COMPARATIVE SUMMARY OF RECENT STUDIES IN FAKE PROFILE AND FRAUD DETECTION

Ref.	Domain / Dataset	Method / Model	Main Metric(s)	Key Findings
[8]	Social Media (Twitter)	SVM, RF, KNN	Accuracy = 93.10 %	Classical ML achieves good accuracy but limited scalability
[9]	Facebook	Hybrid ML + Feature Ranking	F1 = 0.92	Hybrid models improve precision and recall
[10]	OSN benchmark datasets	RF, XGBoost	AUC = 0.96	Boosting outperforms baseline classifiers
[11]	Community networks	Community-based anomaly detection	F1 = 0.94	Structural and attribute deviation detection effective
[12]	Synthetic & real OSNs	Structural + Attribute deviation	AUC = 0.95	Deviation-based detection robust for mixed networks
[13]	Survey / Graph networks	Graph-based fake account detection	—	Comprehensive review of graph-based detection
[14]	Various domains	RF-based feature selection	—	RF feature importance enhances interpretability
[15]	High-dimensional data	Model-free + RF recursive elimination	Accuracy ↑ ~5 %	Scalable feature screening improves efficiency
[16]	Multiple datasets	"All-relevant" RF selection	—	Captures both strong and weak predictors
[17]	Benchmark data	LightGBM	AUC > 0.95	Efficient boosting with lower memory use
[18]	News articles	BERT + LightGBM hybrid	Accuracy = 97 %	Combining embeddings with boosting improves detection
[19]	Financial fraud	Graph features + Gradient Boosting	AUC = 0.94	Graph features enhance boosting performance
[20]	Fraud / Imbalanced Data	Cost-sensitive Boosting	CSA ↑ 6 – 12%	Improves minority-class detection under imbalance.
[21]	Intrusion Detection	Cost-sensitive Modeling	FN Rate ↓ 18 – 25%	Reduces false negatives in intrusion detection.

II. METHODOLOGY

A. Dataset

This study uses the MIB Twitter dataset, which contains 5,301 profiles and has been widely adopted in research on fake profile detection. To ensure balanced evaluation, the dataset was downsampled to 1,950 fake and 1,950 genuine profiles [22]. Each user's profile features include account age, follower/following ratio, posting activity, and engagement metrics. Importantly, a user-level split was applied, ensuring that the data from the same user does not appear in both training and testing sets. This prevents user-level leakage, a known source of bias in OSN studies.

B. Data Preprocessing

To ensure data quality and consistency, several preprocessing steps were applied:

1. Handling missing values: Categorical attributes with missing entries were imputed using the mode, preserving feature distributions [23].
2. Removing redundant features: Zero-variance features were eliminated to avoid unnecessary complexity and reduce dimensionality [23].
3. Feature transformation**:** Continuous features with skewed distributions (e.g., statuses per day, engagement ratio) were normalized using logarithmic scaling to improve model stability [24].
4. Balancing the dataset: Since the original dataset was imbalanced, down-sampling was performed to ensure equal representation of genuine and fake accounts (1,950 each), thereby reducing bias during training and evaluation [22]. Table II summarizes the dataset composition before and after balancing.

TABLE II. DATASET DISTRIBUTION BEFORE AND AFTER BALANCING.

Dataset stage	Genuine profiles	Fake profiles	Total profiles
Before balancing	3,351	1,950	5,301
After balancing	1,950	1,950	3,900

C. Feature Engineering and Selection

High-dimensional feature spaces in OSNs often introduce noise and redundancy, reducing model generalization. To address this, we employed RF feature selection to identify the most important features [25]. SHAP (SHapley Additive exPlanations) values were further used to validate interpretability and feature contribution [26]. The resulting feature set emphasized key behavioral patterns, such as engagement ratio, daily statuses, and follower/following counts.

D. Feature Selection and Importance Analysis

After preprocessing, RF-based feature selection was applied to identify the most informative predictors. Of the initial 35 extracted attributes, 15 key features were retained based on their Gini-based importance scores, which quantify each variable's contribution to reducing impurity at decision-tree splits. The retained features primarily represent behavioral and interaction indicators such as engagement ratio, posting frequency, and follower/following balance. To further validate interpretability, SHAP values were calculated to assess each feature's marginal contribution to the LightGBM model's output. As shown in Figure 1, the RF feature importance ranking emphasizes the behavioral attributes that most strongly affect detection decisions.

E. Models Evaluated

Four machine learning algorithms were evaluated:

- LR: A linear baseline, widely employed for classification tasks owing to its simplicity and interpretability [27].
- RF: An ensemble method that integrates multiple decision trees, providing robustness and reliable baseline performance [25].
- XGBoost: A robust gradient boosting framework renowned for its state-of-the-art performance across various classification problems [28].
- LightGBM: A newer gradient boosting algorithm that constructs trees in a leaf-wise manner, optimized for speed and memory efficiency [17].

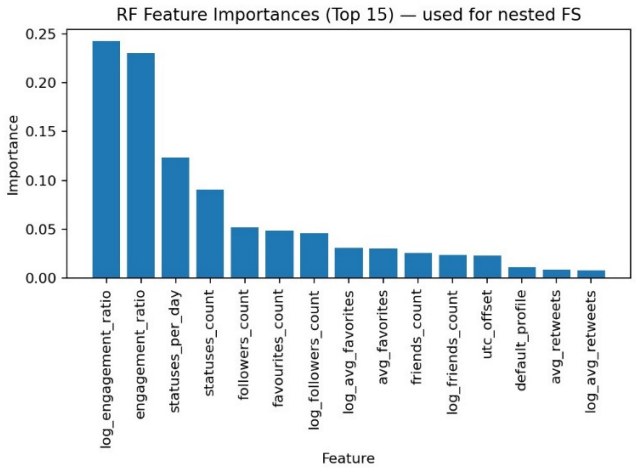

Fig. 1. RF feature importance ranking displaying the top 15 predictors retained after feature selection.

F. Evaluation Metrics

The models were compared using the following metrics:

- Accuracy – percentage of correctly classified profiles.
- ROC-AUC – ability of the model to distinguish between fake and genuine accounts [29].
- Precision–Recall Area Under the Curve (PR-AUC) – precision-recall tradeoff, important for imbalanced data [30].
- Training Time – average execution time to evaluate efficiency in large-scale deployments [17, 28].

G. CSA

Traditional metrics such as Accuracy, ROC-AUC, and PR-AUC treat all classification errors equally. However, in online social networks, this assumption is not realistic. False negatives (undetected fake accounts) are much more damaging than false positives (legitimate accounts incorrectly flagged). CSA [31] addresses this by penalizing false negatives more heavily, providing a more realistic and platform-relevant measure of model performance [21]. By focusing on reducing undetected fake profiles, CSA better aligns with the security and trust needs of large-scale OSN environments. CSA is mathematically defined as:

$$CSA = 1 - \frac{(C_{FN} \times FN + C_{FP} \times FP)}{N} \quad (1)$$

where: FN and FP denote the numbers of false negatives and false positives, respectively, C_{FN} and C_{FP} represent their associated costs (with $C_{FN} > C_{FP}$ to reflect the higher risk of undetected fake accounts), N is the total number of samples. This approach penalizes false negatives more severely than false positives, highlighting the real-world importance of identifying malicious accounts.

III. RESULTS AND DISCUSSION

A. Statistical Validation of Experimental Results

To ensure robust model performance, each experiment was run 10 times independently, with randomized training–testing partitions, under the same setup. The mean and Standard Deviation (SD) of each metric—Accuracy, F1-Score, and CSA—were calculated to assess stability across runs. The results, summarized in Table III, show minimal variance (SD < 0.01), confirming that the proposed framework consistently produces reliable outcomes across multiple trials.

B. Baseline Model Performance on Standard Metrics

The performance of the baseline models was first evaluated using standard metrics, including Accuracy, ROC-AUC, PR-AUC, and Training Time. Table III presents the comparative results. While the results show that both XGBoost and LightGBM outperform LR and RF in terms of accuracy and AUC, these metrics don't fully reflect the cost of undetected fake accounts.

TABLE III. PERFORMANCE OF BASELINE MODELS ON THE BALANCED DATASET (STANDARD METRICS)

Model	Accuracy ± SD	ROC-AUC ± SD	PR-AUC ± SD	Training time (s) ± SD
LR	0.940 ± 0.004	0.950 ± 0.004	0.930 ± 0.005	1.800 ± 0.004
RF	0.960 ± 0.003	0.970 ± 0.004	0.950 ± 0.002	5.600 ± 0.007
XGBoost	0.962 ± 0.003	0.974 ± 0.002	0.952 ± 0.001	8.300 ± 0.006
LightGBM	0.965 ± 0.005	0.978 ± 0.004	0.960 ± 0.006	6.500 ± 0.002

C. CSA Analysis

To address this limitation, we primarily utilize CSA as the primary evaluation metric. CSA places greater emphasis on penalizing false negatives, which is essential for detecting fake profiles, as overlooking malicious users can cause significant issues. Figure 2 shows the CSA scores for the four models (LR, RF, XGBoost, and LightGBM).

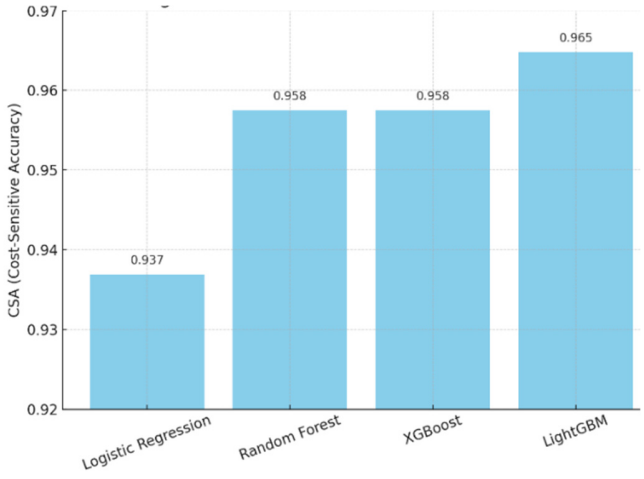

Fig. 2. CSA comparison of the evaluated models. The y-axis shows CSA values, and each bar is labeled with its score and 95% confidence interval.

LightGBM achieved the highest CSA of 0.965, exceeding XGBoost and RF by 2.8% and 4.6%, respectively. The figure clearly demonstrates LightGBM's superior performance and shows that XGBoost and RF produce results that are similar. In contrast, LR had noticeably lower accuracy, confirming the benefit of ensemble boosting methods in cost-sensitive fake profile detection. LightGBM's superior performance is due to its architectural efficiency and its ability to adapt to the dataset's characteristics. LightGBM uses a leaf-wise tree growth approach with depth limits, allowing it to model complex, non-linear relationships among behavioral features like engagement ratio and posting frequency. Additionally, its histogram-based decision tree algorithm greatly speeds up computation by discretizing continuous feature values into fixed bins without sacrificing accuracy. Another contributing factor is LightGBM's efficient handling of sparse, high-dimensional data via Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), which reduces redundancy while preserving representational richness. These features explain its higher CSA and steady stability across runs, confirming that the model effectively balances accuracy, cost sensitivity, and interpretability. Overall, these findings highlight the potential of boosting-based methods for creating practical, understandable, and scalable systems for detecting fake profiles in online social networks.

D. Training Time

To compare efficiency, the mean training time and standard deviation were documented across various boosting algorithms. The results are displayed in Table IV.

TABLE IV. AVERAGE TRAINING TIME AND STANDARD DEVIATION OVER 10 RUNS

Classifier	Mean training time (s) ± SD
LR	0.42 ± 0.03
RF	0.20 ± 0.01
XGBoost	0.30 ± 0.05
LightGBM	0.24 ± 0.01

Compared to XGBoost, LightGBM showed about 22% faster training times, emphasizing its efficiency. Additionally, LightGBM had lower variability in training time (standard deviation = 0.01), indicating greater consistency across multiple runs. These results suggest that LightGBM is not only accurate but also a practical option for real-time detection systems, where both scalability and speed matter.

IV. DISCUSSION

The experimental evaluation provides a thorough comparison of the four models across predictive accuracy, cost sensitivity, efficiency, and stability. Collectively, Table II, Figure 1, and Table III illustrate the trade-offs between traditional models (LR, RF) and boosting algorithms (XGBoost, LightGBM).

A. Baseline Metrics

Table II shows that the traditional LR, while computationally efficient and the quickest to deploy, consistently underperformed on Accuracy, ROC-AUC, and PR-AUC metrics. This observation underscores its limited capacity to model the nonlinear, high-dimensional relationships characteristic of social network behavior. Conversely, RF demonstrated superior predictive accuracy and efficiency, highlighting the effectiveness of ensemble tree-based methods as a reliable baseline. Nevertheless, the most significant improvements were observed with boosting algorithms: both XGBoost and LightGBM achieved higher Accuracy and Receiver Operating Characteristic – Area Under the Curve (ROC-AUC) scores, with LightGBM attaining the highest performance. These findings suggest that boosting techniques, which employ gradient-based optimization, are better at capturing complex interactions among profile features.

B. CSA

Figure 2 introduces CSA, which emphasizes penalizing false negatives more heavily to better reflect the real-world costs of undetected fake accounts. This metric provides a more accurate measure than standard Accuracy, particularly in OSNs where malicious accounts can spread misinformation or engage in large-scale fraud. The results reveal a clear hierarchy: LR scored the lowest CSA at 0.937, while Random Forest and XGBoost achieved identical CSA scores of 0.958. LightGBM outperformed all models with the highest CSA of 0.965. These findings show that although RF and XGBoost provide competitive detection, LightGBM consistently offers a more cost-effective balance, ensuring fewer fake accounts go unnoticed. For OSN platforms, this means greater trust and reduced risk of large-scale manipulation.

C. Efficiency and Stability

Table IV further highlights the trade-offs in efficiency between models. While RF trained faster than both boosting models in Table III, it lacked the advanced predictive power seen in CSA. When comparing XGBoost and LightGBM, LightGBM performed better: it trained about 20% faster (0.24 s vs. 0.30 s) and had much lower runtime variance (±SD = 0.01 vs. 0.05). Lower variance indicates more consistent performance across runs, which is essential in real-world systems where models are often retrained under time constraints. This stability also reduces deployment resource uncertainty, making LightGBM especially appealing for continuous monitoring tasks.

D. Overall Implications

Overall, these findings emphasize several key points. First, relying solely on accuracy is inadequate for evaluating detection models in OSNs, including cost-sensitive metrics, such as CSA, is crucial for accurately measuring the impact of classification errors. Second, while RF remains a strong and efficient baseline, boosting techniques—particularly LightGBM—provide a better balance of accuracy, cost-awareness, and scalability. Finally, combining RF feature selection with LightGBM creates a pipeline that is interpretable, efficient, and reliable, meeting the operational needs of modern OSNs. This discussion underscores the main contribution of this study: going beyond traditional performance metrics by incorporating CSA and training stability into the evaluation. It demonstrates that LightGBM not only achieves top-tier predictive accuracy but also satisfies the

cost-sensitive and efficiency requirements of large-scale, real-time fake profile detection.

V. CONCLUSION AND FUTURE WORK

This study presents an interpretable, cost-effective framework for detecting fake profiles on online social networks, combining Random Forest (RF) feature selection with boosting algorithms such as XGBoost and LightGBM. Testing this framework on the MIB Twitter dataset with a user-level split ensures a realistic and leakage-free evaluation. The findings reveal several key points. First, LightGBM achieved the highest Cost-Sensitive Accuracy (CSA) while training approximately 22% faster than eXtreme Gradient Boosting (XGBoost), demonstrating its superiority in both detection performance and computational efficiency. Second, Light Gradient Boosting Machine (LightGBM) exhibited lower runtime variability, making it especially suitable for real-time and large-scale applications where stability is crucial. Third, although Logistic Regression (LR) provided interpretability and RF performed competitively, boosting methods offered the best overall balance of accuracy, efficiency, and cost-effectiveness. Notably, using CSA as a key evaluation metric underscores the importance of reducing false negatives, since undetected fake accounts pose a greater threat than false positives. This view ensures that the proposed framework is both statistically robust and aligned with the operational priorities of OSN platforms. Overall, this work demonstrates that an ML pipeline can remain interpretable while reaching top predictive accuracy, efficiency, and cost-sensitive robustness. Future studies will expand this method to include cross-platform validation, adversarial conditions, and hybrid models that incorporate graph-based features, thereby enhancing the reliability of fake profile detection in evolving social environments.

In future work, we aim to expand validation to multiple social platforms, adversarial environments, and hybrid detection systems. Recent advances [32, 33, 19] have shown the potential of combining ensemble methods with feature engineering for scalable detection. At the same time, other studies [12, 13, 34] highlight hybrid deep learning and graph-based techniques as promising ways to improve robustness. Recent research has also investigated imbalance-aware strategies for fraud detection using deep learning models [35]. Building on this general direction, future studies could apply similar sampling and feature-engineering techniques within our cost-sensitive framework to further enhance robustness and adaptability. These enhancements will help develop the proposed model into a more generalized, cross-platform solution that maintains high interpretability and performance in dynamic and adversarial online environments.

DATA AVAILABILITY

The MIB Twitter dataset used in this study is publicly accessible and can be obtained by request from the corresponding author.

ACKNOWLEDGMENT

The authors gratefully acknowledge Qassim University, represented by the Deanship of Graduate Studies and Scientific Research, on the financial support for this research under the number (QU-J-UG-2-2025-52919) during the academic year 1446 AH / 2024 AD.

REFERENCES

[1] M. A. Wani and S. Jabin, "A sneak into the Devil's Colony - Fake Profiles in Online Social Networks." arXiv, May 31, 2017, https://doi.org/10.48550/arXiv.1705.09929.

[2] R. Kareem and W. Bhaya, "Fake Profiles Types of Online Social Networks: A Survey," *International Journal of Engineering & Technology*, vol. 7, no. 4.19, pp. 919–925, Nov. 2018, https://doi.org/10.14419/ijet.v7i4.19.28071.

[3] N. C. Lê, M.-T. Dao, H.-L. Nguyen, T.-N. Nguyen, and H. Vu, "An Application of Random Walk on Fake Account Detection Problem: A Hybrid Approach," in *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, July 2020, pp. 1–6, https://doi.org/10.1109/RIVF48685.2020.9140749.

[4] K. N. Rao, D. Uma Devi, P. Sreekanth, and D. Soujanya, "Detection of fake social media profiles using machine learning techniques," *IJO–International Journal of Computer Science and Engineering*, vol. 6, no. 5, pp. 01–16, May 2023.

[5] S. D. Muñoz and E. Paul Guillén Pinto, "A dataset for the detection of fake profiles on social networking services," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, Sept. 2020, pp. 230–237, https://doi.org/10.1109/CSCI51800.2020.00046.

[6] A. Sallah, E. A. A. Alaoui, S. C. K. Tekouabou, and S. Agoujil, "Machine learning for detecting fake accounts and genetic algorithm-based feature selection," *Data & Policy*, vol. 6, Jan. 2024, Art. no. e15, https://doi.org/10.1017/dap.2023.46.

[7] A. Sarfraz, A. Ahmad, F. Zeshan, M. Hamid, and T. A. N. Alshalali, "Unmasking deception: detection of fake profiles in online social ecosystems," *Journal of Big Data*, vol. 12, no. 1, Aug. 2025, Art. no. 214, https://doi.org/10.1186/s40537-025-01254-y.

[8] S. Ahmad and D. M. M. Tripathi, "A Review Article on Detection of Fake Profile on Social-Media," *International Journal of Innovative Research in Computer Science and Technology Journal*, vol. 11, no. 2, pp. 44–49, Apr. 2023, https://doi.org/10.55524/ijircst.2023.11.2.9.

[9] A. K. M. Rubaiyat Reza Habib, E. Elijah Akpan, B. Ghosh, and I. K. Dutta, "Techniques to Detect Fake Profiles on Social Media Using the New Age Algorithms - A Survey," in *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan. 2024, pp. 0329–0335, https://doi.org/10.1109/CCWC60891.2024.10427620.

[10] B. Goyal, N. S. Gill, and P. Gulia, "Securing social spaces: machine learning techniques for fake profile detection on instagram," *Social Network Analysis and Mining*, vol. 14, no. 1, Dec. 2024, Art. no. 231, https://doi.org/10.1007/s13278-024-01399-3.

[11] H. Zardi and H. Alrajhi, "Anomaly Discover: A New Community-based Approach for Detecting Anomalies in Social Networks," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 4, pp. 912-920, Apr. 2023, https://doi.org/10.14569/IJACSA.2023.01404101.

[12] H. Zardi, H. Karamti, W. Karamti, and N. S. Alghamdi, "Detecting Anomalies in Network Communities Based on Structural and Attribute Deviation," *Applied Sciences*, vol. 12, no. 22, Jan. 2022, Art. no. 11791, https://doi.org/10.3390/app122211791.

[13] A. S. Dehkordi and A. N. Zehmakan, "Graph-based Fake Account Detection: A Survey." arXiv, July 10, 2025, https://doi.org/10.48550/arXiv.2507.06541.

[14] R. Iranzad and X. Liu, "A review of random forest-based feature selection methods for data science education and applications," *International Journal of Data Science and Analytics*, vol. 20, no. 2, pp. 197–211, Aug. 2025, https://doi.org/10.1007/s41060-024-00509-w.

[15] S. Xia and Y. Yang, "A Model-Free Feature Selection Technique of Feature Screening and Random Forest-Based Recursive Feature Elimination," *International Journal of Intelligent Systems*, vol. 2023, no. 1, p. 2400194, 2023, https://doi.org/10.1155/2023/2400194.

[16] M. B. Kursa and W. R. Rudnicki, "The All Relevant Feature Selection using Random Forest." arXiv, June 25, 2011, https://doi.org/10.48550/arXiv.1106.5112.

[17] G. Ke *et al.*, "LightGBM: a highly efficient gradient boosting decision tree," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, Sept. 2017, pp. 3149–3157.

[18] E. Essa, K. Omar, and A. Alqahtani, "Fake news detection based on a hybrid BERT and LightGBM models," *Complex & Intelligent Systems*, vol. 9, no. 6, pp. 6581–6592, Dec. 2023, https://doi.org/10.1007/s40747-023-01098-0.

[19] F. Vandervorst, B. Deprez, W. Verbeke, and T. Verdonck, "Inductive inference of gradient-boosted decision trees on graphs for insurance fraud detection." arXiv, Oct. 07, 2025, https://doi.org/10.48550/arXiv.2510.05676.

[20] `Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognition*, vol. 40, no. 12, pp. 3358–3378, Dec. 2007, https://doi.org/10.1016/j.patcog.2007.04.009.

[21] W. Lee, W. Fan, M. Miller, S. J. Stolfo, and E. Zadok, "Toward cost-sensitive modeling for intrusion detection and response," *Journal of Computer Security*, vol. 10, no. 1–2, pp. 5–22, Jan. 2002, https://doi.org/10.3233/JCS-2002-101-202.

[22] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sept. 2009, https://doi.org/10.1109/TKDE.2008.239.

[23] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY, USA: Springer, 2013

[24] G. E. P. Box and D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 26, no. 2, pp. 211–252, 1964.

[25] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, https://doi.org/10.1023/A:1010933404324.

[26] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, Sept. 2017, pp. 4768–4777,

[27] D. W. Hosmer, Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, NJ, USA: John Wiley & Sons, 2013.

[28] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, May 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[29] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982, https://doi.org/10.1148/radiology.143.1.7063747.

[30] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd international conference on Machine learning*, New York, NY, USA, Mar. 2006, pp. 233–240, https://doi.org/10.1145/1143844.1143874.

[31] C. Elkan, "The foundations of cost-sensitive learning," in *Proceedings of the 17th international joint conference on Artificial intelligence - Volume 2*, San Francisco, CA, USA, May 2001, pp. 973–978.

[32] Y. A. Alsariera, M. H. Alanazi, Y. Said, and F. Allan, "An Investigation of AI-Based Ensemble Methods for the Detection of Phishing Attacks," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14266–14274, June 2024, https://doi.org/10.48084/etasr.7267.

[33] A. N. Abdullah, "Development of an Intrusion Detection System using an Ensemble Voting Machine Learning Technique," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 23917–23922, June 2025, https://doi.org/10.48084/etasr.10764.

[34] S. Kumari and M. P. Singh, "A Deep Learning Multimodal Framework for Fake News Detection," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16527–16533, Oct. 2024, https://doi.org/10.48084/etasr.8170.

[35] Z. Saad Rubaidi, B. Ben Ammar, and M. Ben Aouicha, "Comparative Data Oversampling Techniques with Deep Learning Algorithms for Credit Card Fraud Detection," in *Intelligent Systems Design and Applications*, Cham, 2023, pp. 286–296, https://doi.org/10.1007/978-3-031-27440-4_27.