Investigating Credential-Based and Behavioral Characteristics for the Development of an Intelligent Framework Toward Fake Account Identification

Harjot Kaur
Department of Computer Science and Applications Sant Baba Bhag Singh University
Jalandhar, Punjab, India Itsme.jass,jot@gmail.com
Dr. Nirmal Kaur
Associate Professor Department of Computer Science and Applications Sant Baba Bhag Singh University
Jalandhar, Punjab, India Nkparhar.sbbs@gmail.com

How to Cite this Article:

Kaur, H. (2026). Investigating Credential-Based and Behavioral Characteristics for the Development of an Intelligent Framework Toward Fake Account Identification. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.811

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.

https://doi.org/10.55041/ijcope.v2i4.811

***Abstract*—** The rapid growth of online platforms has significantly increased the challenge of maintaining secure and trustworthy digital environments. One of the major concerns associated with this growth is the widespread creation of fake accounts, which are often used for spam dissemination, financial fraud, misinformation campaigns, and manipulation of online interactions. Such activities negatively affect platform integrity, user trust, and overall cybersecurity. Traditional fake account detection systems primarily depend on long-term behavioral monitoring and post- registration activity analysis. Although these methods can provide meaningful insights, they often introduce delays in detection, allowing malicious accounts to exploit platform vulnerabilities before being identified.

This study presents an intelligent fake account detection framework that combines credential-level characteristics with selected behavioral indicators to support more effective account classification. The proposed model considers multiple measurable features, including username patterns, profile completeness, account age, follower-to-following ratio, suspicious bio content, and activity regularity. These features are processed using a Logistic Regression–based classification model, where each feature contributes to the final decision through an assigned weight that reflects its relative significance.

The weighted feature combination is transformed into a probability score using the sigmoid function, enabling the system to estimate the likelihood of an account being fake. Based on this probability, accounts are categorized into low-risk, medium-risk, and high-risk groups, thereby supporting practical and interpretable decision-making. The probabilistic nature of the model enhances transparency and allows flexible risk-based intervention strategies.

The proposed approach aims to provide a balanced solution by integrating early account-level indicators with lightweight behavioral signals. This combination improves detection capability while maintaining computational efficiency, scalability, and practical applicability for modern online platforms. Although the study does not claim universal detection capability, it demonstrates the potential of combining multiple feature types within an interpretable machine learning framework for intelligent fake account identification.

Keywords— Fake Account Detection, Logistic Regression, Credential Features, Behavioral Features, Probability Classification, Cybersecurity, Social Media Security, Machine Learning.

Introduction

Online platforms have become an essential part of modern digital infrastructure, supporting communication, commerce, education, entertainment, and financial transactions. With the rapid increase in the number of users across social media networks, e-commerce systems, and digital services, the dependency on digital identities has grown significantly. However, this widespread adoption has also introduced serious cybersecurity challenges, among which the creation and misuse of fake accounts remains one of the most persistent threats.

Fake accounts are commonly used for malicious purposes such as spam distribution, phishing, misinformation campaigns, financial fraud, artificial engagement, and manipulation of online interactions. These accounts not only compromise the integrity of digital platforms but also reduce user trust and may negatively affect the reliability of online ecosystems. As automated account generation tools and bot technologies continue to evolve, detecting suspicious accounts has become increasingly complex.

Traditional fake account detection mechanisms largely rely on behavioral monitoring techniques. These methods analyze factors such as posting frequency, interaction patterns, login behavior, and network relationships to identify suspicious activity. While such approaches can provide valuable insights, they generally require sufficient post-registration activity data before a reliable decision can be made. This dependency often introduces a detection delay, during which malicious accounts may exploit system vulnerabilities and cause significant damage before being flagged.

To address this limitation, recent research has increasingly focused on integrating early account- level indicators with selected behavioral features to improve detection effectiveness.

Inspired by this direction, the present study proposes an intelligent detection framework that evaluates measurable characteristics associated with both account credentials and lightweight behavioral signals. Features such as username structure, profile completeness, account age, follower-following ratio, suspicious bio content, and activity regularity are considered as important indicators for distinguishing fake accounts from genuine ones.

The analytical foundation of the proposed framework is based on Logistic Regression, a widely used and interpretable machine learning algorithm for binary classification tasks. The model assigns weights to each selected feature, combines them into a linear score, and converts the result into a probability using the sigmoid function. This probabilistic output enables the system to classify accounts into different risk levels, thereby supporting practical decision-making and scalable deployment.

The primary purpose of this study is how a balanced combination of credential-level information and selected behavioral indicators can improve fake account detection in a way that remains efficient, explainable, and suitable for modern online environments. Rather than replacing existing security systems, the proposed framework is intended to serve as an intelligent screening mechanism that strengthens early-stage account verification and risk assessment.

Background and Problem Context

The continuous expansion of digital platforms has transformed the way individuals, organizations, and institutions communicate, transact, and share information. Social media networks, online marketplaces, financial applications, and digital learning systems all rely heavily on user-generated accounts as the primary medium of access and interaction. While this account-based ecosystem enables convenience and scalability, it has also created significant security vulnerabilities, particularly in the form of fake account creation and misuse.

Fake accounts are often generated using automated scripts, bot frameworks, or coordinated manual efforts. These accounts may be used for a wide range of malicious activities, including spam dissemination, phishing attacks, misinformation campaigns, financial fraud, artificial boosting of engagement metrics, and circumvention of platform restrictions. As the sophistication of such malicious entities continues to increase, online platforms face growing pressure to deploy detection mechanisms that are not only accurate but also scalable and computationally efficient.

Conventional fake account detection systems have traditionally relied on post-registration behavioral analysis. These methods focus on factors such as posting frequency, interaction networks, login timing, follower-following relationships, and activity regularity. Although such indicators can be highly informative, they require sufficient activity data to accumulate over time before suspicious patterns become visible. This introduces a delay between account creation and effective detection, during which harmful actions may already take place.

On the other hand, approaches that focus exclusively on account-level or credential-based attributes—such as username structure, profile completeness, or basic registration information— support earlier-stage screening but may not always provide sufficient discriminative capability when used alone. Sophisticated attackers can often generate realistic-looking credentials that resemble genuine user profiles, thereby reducing the effectiveness of standalone credential-based screening.

This creates an important practical challenge in fake account identification: the need to balance **early- stage detection** with **sufficient classification accuracy**. Relying only on behavioral monitoring may delay response, while relying only on credential features may limit robustness. Therefore, there is a need for an intelligent framework that combines the strengths of both perspectives.

The problem context addressed in this study lies in designing a detection mechanism that integrates measurable account-level features with selected lightweight behavioral indicators. By combining these complementary signals within an interpretable machine learning framework, the system can generate probability-based risk assessments that support timely and practical decisions. Logistic Regression is particularly suitable in this context because it enables transparent feature contribution analysis, fast computation, and scalable deployment across large online platforms.

The present study is therefore positioned within the broader challenge of building efficient, explainable, and practically deployable fake account detection systems that can improve security while remaining suitable for real-world digital environments.

Related Work

The field of fake account detection has evolved considerably over the past decade, progressing from rule-based filtering methods to advanced machine learning and hybrid analytical frameworks. Early research primarily focused on identifying suspicious behavior through observable interaction patterns and social spam activities. Gianluca Stringhini et al. (2010) introduced one of the fundamental study in this area by analyzing spam campaigns on social networks through behavioral indicators such as message repetition and abnormal interaction frequency. Their work demonstrated that behavioral features can effectively distinguish malicious users, although it required sufficient post-registration activity data.

Subsequent studies expanded detection beyond individual behavior to include broader bot characterization. Clayton A. Davis, Onur Varol, and Emilio Ferrara contributed significantly through the development of **BotOrNot/Botometer**, a widely recognized system that combines metadata, network structure, content features, and temporal behavior to estimate bot likelihood. Their findings highlighted the importance of combining multiple feature categories for robust detection, particularly in large- scale social environments.

As machine learning techniques matured, researchers began exploring classification models that integrate profile-level attributes with selected behavioral signals. Kai- Cheng Yang et al.(2020) proposed scalable bot detection methods using lightweight metadata and account-level indicators, demonstrating that efficient models can achieve strong generalization while reducing computational overhead. Their work is particularly relevant to practical deployment scenarios where large user volumes demand fast inference.

More recent studies have emphasized hybrid and explainable detection strategies. Emilio Ferrara (2023) discussed the growing challenge of detecting increasingly human-like bots in the age of advanced AI systems, highlighting the limitations of relying solely on any single feature type. The study strongly supports the use of integrated detection frameworks that combine account metadata, lightweight behavioral indicators, and interpretable machine learning techniques.

Recent literature also reflects increasing attention toward lightweight and probability-based machine learning frameworks for fake account detection. Studies in machine learning-based bot detection have shown that supervised classifiers can effectively leverage profile attributes, interaction ratios, and temporal regularity to improve detection accuracy. At the same time, these works note that highly complex models may reduce interpretability and increase deployment overhead.

Despite these advancements, many existing systems still depend heavily on long-term behavioral monitoring or large multi-source feature sets. This dependence may introduce detection delays and increase computational complexity. The present study builds upon these insights by exploring a balanced framework that combines credential-level characteristics with selected lightweight behavioral features within a Logistic Regression–based probabilistic model, enabling interpretable and scalable risk-based fake account classification.

Research Gap

Although significant progress has been made in fake account detection, several limitations still remain across existing approaches. A large portion of the current literature primarily relies on long- term behavioral monitoring, where suspicious accounts are identified based on posting patterns, interaction frequency, login timing, follower- following relationships, and network structures. While these methods often provide strong detection accuracy, they inherently depend on the availability of sufficient post-registration activity data. This creates a delay between account creation and reliable detection, during which malicious accounts may already perform harmful actions.

At the same time, studies that focus only on credential-level or profile-based attributes provide earlier access to useful signals such as username structure, profile completeness, and registration- related metadata. However, when used independently, these features may not always provide enough discriminative strength to identify sophisticated fake accounts that are designed to imitate genuine users.

Another limitation observed in recent literature is the growing complexity of detection systems. Many advanced frameworks combine multiple data sources, including network graphs, temporal behavior, content analysis, and device-level metadata. Although these approaches improve performance, they often increase computational overhead, reduce interpretability, and make deployment more difficult for large-scale real-world platforms.

A further gap exists in the limited use of lightweight probabilistic models that balance interpretability, efficiency, and practical risk-based decision support. While several studies employ complex ensemble or deep learning architectures, comparatively less emphasis has been placed on transparent machine learning methods that allow clear understanding of how individual features contribute to the final decision.

This study addresses these gap by exploring a balanced detection framework that combines credential-level characteristics with selected lightweight behavioral indicators within a Logistic Regression–based probabilistic classification model. By integrating complementary feature types, the proposed approach seeks to improve early detection capability while maintaining computational efficiency, model transparency, and scalable deployment suitability for modern online platforms.

Rather than replacing existing security mechanisms, the framework is intended to provide an interpretable and practical risk-assessment layer that supports timely fake account identification and informed intervention strategies.

Proposed Framework

The proposed framework introduces an intelligent fake account detection mechanism that combines credential-level characteristics with selected behavioral indicators to support accurate and interpretable account classification. The framework is designed as a lightweight analytical pipeline that evaluates measurable account attributes and transforms them into a probability-based risk score using Logistic Regression.

The overall workflow begins when an account is submitted for analysis. At this stage, the system extracts a set of predefined measurable features representing both account-level properties and selected behavioral signals. These

features are normalized on a scale of 0 to 1 to ensure consistency in model processing.

The selected feature set includes indicators such as account age, profile completeness, presence of a profile image, username structure, follower-to- following ratio, suspicious keywords in the bio, posting regularity, interaction abnormality, engagement consistency, and selected temporal activity patterns. Each feature captures a specific aspect of account authenticity and contributes differently to the final classification process.

To account for varying significance among these indicators, the framework assigns an individual weight to each feature. Stronger signals, such as suspicious bio keywords or abnormal follower ratios, receive relatively higher positive influence, while genuine profile indicators, such as complete profile information or the presence of a profile image, may contribute negative weights that reduce suspicion.

Mathematically, the extracted feature vector is processed through the Logistic Regression linear function:

Z=w1x1+w2x2+w3x3+…+wnxn+bz=w_1x_1+w_2x_2+w_3x_3+…+w_nx_n+bz=w1 x1+w2x2+w3x3+…+wnxn+b

Where:

- x1,x2,…,xnx_1,x_2, …, x_nx1,x2,…,xn represent normalized account features.
- w1,w2,…wnw_1,w_2,…, w_nw1,w2,…,wn denote their learned weights.
- bbb is the bias term
- zzz is the combined linear suspicion score.

A higher positive value of zzz indicates stronger evidence of fake account characteristics, while lower or negative values suggest a higher likelihood of genuine behavior.

To convert this score into an interpretable probability, the framework applies the sigmoid function:

P(fake)=11+e−zP(fake) =\frac{1} {1+e^ {-z}}P(fake)=1+e-z1}

This transformation maps the linear score into a probability value between 0 and1, representing the likelihood that the analyzed account is fake.

For practical deployment, the probability output is further mapped into a three-level risk classification strategy:

- **High Risk:** P≥0.70P \geq 0.70P≥0.70
- **Medium Risk:** 0.40≤P<0.700.40 \leq P < 0.700.40≤P<0.70
- **Low Risk:** P<0.40P < 0.40P<0.40

This layered risk-based classification supports more flexible decision-making compared to strict binary labeling. High-risk accounts may be flagged for immediate action, medium-risk accounts can be placed under monitoring, and low-risk accounts may continue normal platform access.

The proposed framework is intentionally designed to remain computationally efficient, interpretable, and scalable. By combining early account-level indicators with lightweight behavioral features in a probabilistic model, the system aims to provide a balanced solution that improves detection capability without introducing excessive complexity.

Discussion

The proposed framework introduces an intelligent fake account detection mechanism that combines credential-level characteristics with selected behavioral indicators to support accurate and interpretable account classification. The framework is designed as a lightweight analytical pipeline that evaluates measurable account attributes and transforms them into a probability-based risk score using Logistic Regression.

The overall workflow begins when an account is submitted for analysis. At this stage, the system extracts a set of

predefined measurable features representing both account-level properties and selected behavioral signals. These features are normalized on a scale of 0 to 1 to ensure consistency in model processing.

The selected feature set includes indicators such as account age, profile completeness, presence of a profile image, username structure, follower-to- following ratio, suspicious keywords in the bio, posting regularity, interaction abnormality, engagement consistency, and selected temporal activity patterns. Each feature captures a specific aspect of account authenticity and contributes differently to the final classification process.

To account for varying significance among these indicators, the framework assigns an individual weight to each feature. Stronger signals, such as suspicious bio keywords or abnormal follower ratios, receive relatively higher positive influence, while genuine profile indicators, such as complete profile information or the presence of a profile image, may contribute negative weights that reduce suspicion.

Mathematically, the extracted feature vector is processed through the Logistic Regression linear function:

z=w1x1+w2x2+w3x3+…+wnxn+bz=w_1x_1+w_2x_2+w_3x_3+…+w_nx_n

+bz=w1x1+w2x2+w3x3+…+wnxn+b where:

- x1,x2,…xnx_1, x_2, …., x_nx1,x2,…, xn represent normalized account features.
- w1,w2,…, wnw_1,w_2,…, w_nw1,w2,….,wn denote their learned weights.
- bbb is the bias term
- zzz is the combined linear suspicious score.

A higher positive value of zzz indicates stronger evidence of fake account characteristics, while lower or negative values suggest a higher likelihood of genuine behavior.

To convert this score into an interpretable probability, the framework applies the sigmoid function:

P(fake)=11+e−zP(fake) =\ Frac{1} {1+e^{-z} P(fake)=1+e-z1

This transformation maps the linear score into a probability value between 0 and 1, representing the likelihood that the analyzed account is fake.

For practical deployment, the probability output is further mapped into a three-level risk classification strategy:

- **High Risk:** P≥0.70P \geq 0.70P≥0.70
- **Medium Risk:** 0.40≤P<0.700.40 \leq P < 0.700.40≤P<0.70
- **Low Risk:** P<0.40P < 0.40P<0.40

This layered risk-based classification supports more flexible decision-making compared to strict binary labeling. High-risk accounts may be flagged for immediate action, medium-risk accounts can be placed under monitoring, and low-risk accounts may continue normal platform access.

The proposed framework is intentionally designed to remain computationally efficient, interpretable, and scalable. By combining early account-level indicators with lightweight behavioral features in a probabilistic model, the system aims to provide a balanced solution that improves detection capability without introducing excessive complexity.

References

1. Stringhini, G., Kruegel, C., & Vigna, G. (2010). Detecting spammers on social networks.

Proceedings of the Annual Computer Security Applications Conference, 1–9.

2. Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 280–289.

3. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots.

Communications of the ACM, 59(7), 96–104.

4. Yang, K. C., Varol, O., Hui, P. M., & Menczer, F. (2020). Scalable and generalizable social bot detection through data selection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01), 1096–1103.

5. Ferrara, E. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities.

First Monday, 28(6).

6. Ellaky, Z., et al. (2024). Political social media bot detection: Unveiling cutting-edge feature engineering, feature selection, and machine learning approaches. *Array*, 24, 100357.

7. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

8. Cortes, C., & Vapnik, V. (1995). Support- vector networks. *Machine Learning*, 20(3), 273–297.

9. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *Proceedings of the IEEE International Conference on Data Mining*, 413–422.

10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

11. Feng, Y., et al. (2021). Towards learning- based, content-agnostic detection of social bots.

IEEE Transactions on Computational Social Systems.

12. Sarfraz, A., et al. (2025). Unmasking deception: Detection of fake profiles in online social ecosystems. *Journal of Big Data*.