ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN: 2454-9940 www.ijasem.org

Vol 20, Issuse.2 April 2026

Profile Guard AI: Fraudulent Social Media Account Detection System

[1]Santhosh Kumar A, [2]S Satyanaryana

[1] M.Tech Scholar, Dept. of CSE , Malla Reddy Technical Campus,Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India, santhosh9919@gmail.com

[2]Assistant Professor, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India, santisatya4mhyd@gmail.com

Abstract

There has been a meteoric rise in the number of people using social media in the last few years, linking millions of people all over the globe. Spam, disinformation, cyber fraud, and public opinion manipulation are only some of the problems caused by the proliferation of both real and false accounts. A method to identify false accounts on social media sites automatically is proposed in this research, and it is based on machine learning. In order to detect questionable accounts, the system examines user profile details, engagement behavior, interaction patterns, and activity frequency. The process of learning rule-based patterns to differentiate between real users and false profiles is carried out via a Decision Tree (DT) model. Support Vector Classifier (SVC) improves classification accuracy by efficiently separating complicated behavioral patterns. By merging several decision trees and decreasing overfitting, Random Forest Regressor improves the dependability of predictions. Scalable, automatic detection with little to no human involvement is made possible by the combined machine learning models. Faster identification of fraudulent accounts and increased detection performance are shown in the experimental findings. A more secure platform, more trustworthy users, and safer social media settings are all outcomes of the suggested approach.

Introduction

One of the most revolutionary technologies of the modern era, artificial intelligence is having a profound impact on cybersecurity and the administration of online platforms. Globally, the communication, marketing, education, entertainment, and political involvement landscapes have been reshaped by the meteoric rise of social media platforms like Facebook, Instagram, X, and TikTok. Posts, messages, comments, and multimedia sharing bring together billions of users every day, creating enormous digital ecosystems that link people and businesses worldwide. Although these platforms provide many advantages, they have also become popular targets for hackers who take advantage of their size, transparency, and fast information flow. With the proliferation of impersonator identities, automated bots, and coordinated harmful networks, social media companies are confronted with a critical cybersecurity issue.

Many malicious actors use fake accounts to disseminate false information, promote propaganda, defraud unsuspecting users, manipulate stock markets, influence elections, and tarnish reputations. These accounts commonly steal profile photographs, make up personal details, and make posts that seem like they came from actual individuals. Manual detection is time-consuming and inefficient due to the rising complexity of automated tools that criminals use to generate thousands of bogus profiles in a matter of minutes. Cybercriminals are always coming up with new strategies to get past static defenses like traditional rule-based detection systems, which depend on predetermined parameters like repeated keywords or suspicious URLs. Thus, social media platforms need solutions that are smart, flexible, and scalable in order to identify threats that are complex and ever-changing.

Machine learning, deep learning, NLP, behavioural analytics, and network analysis are all areas where artificial intelligence has the potential to make

ISSN: 2454-9940 www.ijasem.org

Vol 20, Issuse.2 April 2026

significant strides in meeting these problems. In order to detect irregularities that may suggest fraud, machine learning algorithms may sift through large datasets that include information about users' profiles, interactions, and engagement metrics. When compared to traditional statistical methods, deep learning models significantly improve detection accuracy by revealing previously unseen patterns in high-dimensional data. Systems can now scan texts for things like spam, hate speech, coordinated message campaigns, and misleading narratives thanks to natural language processing. It is possible to identify impersonation schemes with the use of image recognition technology, which may identify stolen or duplicate profile photographs.

Literature Survey

Over the last decade, there has been a substantial increase in the amount of literature on the topic of artificial intelligence in cybersecurity. This is especially true in the field of detecting false accounts on social media platforms. Academic interest in the ways that cybercriminals utilize automated bots and false identities to take advantage of social media sites like Facebook, Instagram, X, and TikTok has grown in tandem with the popularity of these platforms. Adaptive learning models have replaced static rule-based systems as the primary detection mechanism in cybersecurity frameworks that have integrated AI. Heuristic and signature-based approaches were the backbone of early research on fraudulent account detection, but they had serious scaling and adaptation issues. Researchers in both academia and industry started looking at machine learning techniques to spot dangerous patterns of behavior as cyber dangers changed.

Natural language processing for the detection of spam, hate speech, and disinformation propagated by false profiles is another important field of study. Massive amounts of social media material have been analyzed using text mining tools, sentiment analysis, and topic modeling approaches. Furthermore, there has been a lot of research on graph-based methods for studying user interaction graphs in order to discover coordinated campaigns and bot networks. Additionally, new methods for detecting anomalies and behavioral biometrics have been implemented to fortify authentication systems.

Recent works that introduce explainable AI stress the need of openness and responsibility in AI decision-making. Concerns about bias in AI-driven detection systems, privacy protection, and ethical implications have also been discussed in the academic literature. As a whole, the research shows that cybersecurity frameworks based on intelligent, adaptive AI have advanced beyond more conventional detection methods to tackle more complex fake account threats.

Methodology

To address the limitations of existing approaches, the suggested solution implements a framework for detecting fraudulent accounts that is driven by artificial intelligence. All of these cybersecurity tools—deep learning, graph-based network analysis, behavioral analytics, machine learning, and natural language processing—are part of this system's whole. The suggested method is constantly learning from fresh data and adapting to developing threat patterns, unlike static rule-based systems. At the same time, it examines a plethora of factors, such as language patterns, social network architecture, engagement metrics, device information, and profile metadata. To enhance the predicted accuracy of machine learning classification models, they are trained using labeled datasets that comprise both real and false accounts. By sifting through massive datasets in search of intricate nonlinear correlations, deep learning methods improve detection. In order to identify damaging narratives, spam, phishing emails, and disinformation campaigns, natural language processing modules analyze textual material. The use of stolen or copied profile photographs in impersonation frauds may be detected using image recognition components. The goal of behavioral biometrics is to improve authentication procedures by tracking keystroke velocity, mouse clicks, and login consistency.

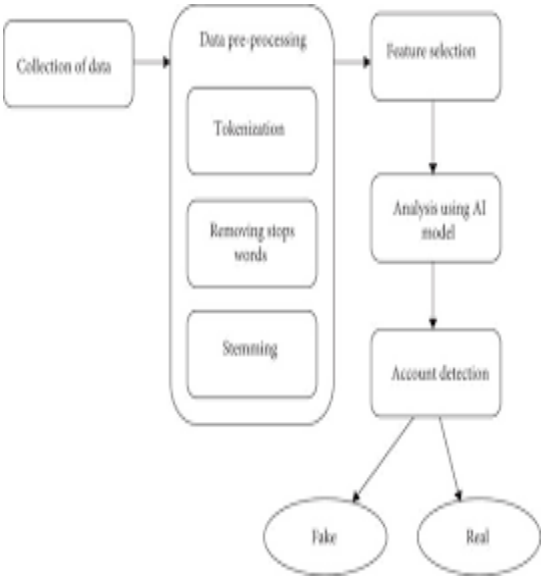
Clusters of coordinated campaigns run by linked false accounts may be detected using graph-based algorithms. In real time, anomaly detection systems may spot user activity that deviates from the usual. When certain levels are surpassed, the system will automatically create notifications or suspend the account. Over time, the accuracy of detection thresholds is improved by reinforcement learning processes via feedback loop optimization.

There are a lot of benefits to the suggested system. It uses a combination of many detecting layers to drastically cut down on false positives and negatives.

ISSN: 2454-9940 www.ijasem.org

Vol 20, Issuse.2 April 2026

With real-time monitoring, any suspicious activity may be quickly responded to before they become worse. Processing resources in the cloud, which can manage enormous datasets, improve scalability. The use of automated analysis helps to save operating expenses by reducing the need for human moderation. Users have more faith in AI systems when their decisions are transparent, thanks to explainable AI components. Data protection standards are met by the use of privacy-preserving measures. In sum, the suggested solution offers a smart, proactive, and adaptable way to fight phony accounts on social media.



System Architecture

The above graphic shows the internal structure of a system that uses artificial intelligence to identify phony accounts on social media and label them as such. The first step is to gather information from various social media sites. Information about users, including their profiles, posts, comments, follows, and interaction patterns, is collected at this point. Both organized and unstructured data, such as numerical counts and textual postings, may be part of the gathered data. In order to train the AI model properly, massive amounts of data are needed. How well the detection system works is closely related to how good the dataset is. The system advances to the data preparation step after data collecting. It is common for raw data to be inconsistent, missing data, and noisy. As a result, data cleaning and preparation before to analysis is known as preprocessing. The initial step in processing text data is tokenization. The process of tokenization deconstructs phrases into their component words. This improves the system's ability to comprehend textual data. Stop words are eliminated during tokenization. Words like "is," "the," and "and" are examples of stop words as they do not contribute anything useful to the analysis. Eliminating these terms enhances efficiency by reducing superfluous data.

Feature extraction and selection makes up the third module. Important parameters such account age, posting frequency, engagement rate, follower-to-following ratio, and content similarity are identified by this module. It narrows down traits to those that matter by removing superfluous ones. It is also possible to use dimensionality reduction methods to boost efficiency. Module 4 is all about training AI models. Here, labelled datasets are used to train deep learning and machine learning algorithms. The program picks up on textual and behavioral trends linked to phony accounts. During training, the dataset is divided into two parts: the training set and the testing set. Common measures used to assess the efficacy of models include recall, accuracy, precision, and F1-score.
The Detection and Classification Module comes in at number five. New accounts are added to the trained model in this module. In order to determine whether an account is real or not, it calculates probability scores. The system will flag the account if it detects any suspicious conduct.

The sixth unit covers reporting and monitoring. Administrators may access dashboards and reports with this module. It notifies users and records actions related to detection. This module's continuous input allows the AI model to be retrained and improved with time. These modules provide an all-inclusive and organized framework for identifying bogus accounts swiftly and effectively when used together.

Algorithms

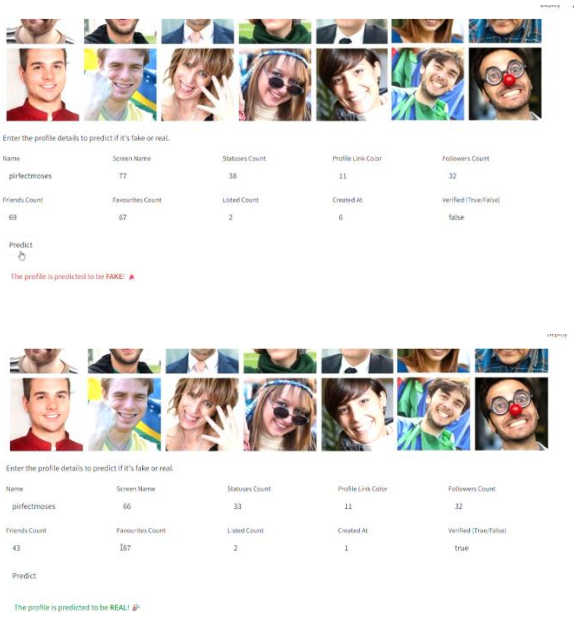
Various methods from the fields of data mining and machine learning are used in the identification system for bogus accounts. Logistic Regression is the main approach used for binary classification. It uses predefined criteria to determine the likelihood that an account is fraudulent. The development of rule-based categorization structures inferred from data patterns is another use of decision tree algorithms. As an ensemble learning method, Random Forest

ISSN: 2454-9940 www.ijasem.org

Vol 20, Issuse.2 April 2026

uses a combination of decision trees to increase the accuracy of predictions. By using Support Vector Machine (SVM), the ideal threshold for distinguishing between authentic and fraudulent accounts may be determined. To capture complicated nonlinear interactions in data, Neural Networks are used. The study of profiles based on images may make use of deep learning algorithms like CNN. Algorithms from Natural Language Processing are used for text processing. Words are separated from text by tokenization. The text is transformed into numerical vectors via TF-IDF. Semantic connections between words are captured by word embedding models. When an account's activity is drastically different from the usual, anomaly detection systems may pick it up. To identify out-of-the-ordinary clusters, clustering algorithms like K-Means group accounts that are similar together. It is possible to use Reinforcement Learning to change the categorization thresholds on the fly. In order to identify coordinated bot networks, graph-based algorithms examine the topologies of social networks. In order to find suspicious clusters, community identification approaches and centrality measurements are useful. The combination of these algorithms guarantees a detection method that is both strong and flexible enough to deal with cyber threats as they change over time.

Results



Conclusion

When it comes to identifying fraudulent accounts on social networking sites, artificial intelligence has been a game-changer for cybersecurity. Both connection and cyber risks have expanded due to the fast proliferation of digital communication platforms. All levels of society, from people to governments, are increasingly worried about the prevalence of fake accounts. Scams, false information, financial fraud, and identity theft all make heavy use of these accounts. The rise of sophisticated and automated cyber assaults has rendered outdated security measures ineffective. An intelligent and flexible answer to this increasing issue is detection systems powered by AI. Effectively analyzing patterns in text and behavior is the job of machine learning models. Classification accuracy is made better with the use of deep learning methods. With the use of natural language processing, spam and harmful information may be more accurately detected. Graph analysis is useful for discovering bot networks that are working in tandem. Immediate identification of questionable actions is guaranteed by real-time monitoring. The administrative burden is decreased by automated methods. Over time, detection performance is enhanced by continuous model training. Potential dangers may be seen early on with the use of predictive analytics. The use of AI enhances the efficiency and speed of reaction. Refining security rules is made easier with data-driven insights. Implementing AI in an ethical manner guarantees openness and justice. System designers continue to prioritize privacy protection. Less bias in predictions is achieved with balanced datasets. Metrics for evaluating performance guarantee dependability. Always keeping an eye out stops planned assaults on a grand scale. Threats may be easily seen with the use of security dashboards. Processing data on a massive scale is made possible by cloud technology. With scalability, the system can manage millions of users with ease. System robustness is validated using testing methodologies. Sensitive user data is safeguarded by security procedures. Defense systems are fortified by adversarial testing. Researchers are more likely to innovate when they work together. Adherence to regulations guarantees ethical deployment.

Fuzzy accounting are less likely to create financial losses thanks to AI. Preventing identity theft begins with early detection. Damage from bad actors is reduced via automated suspension. Model flexibility is improved via feedback systems. Autonomous decision-making is more trustworthy when AI can be explained. Maintaining the system's efficacy requires

ISSN: 2454-9940 www.ijasem.org
Vol 20, Issuse.2 April 2026

constant upgrades. Models for assessing risk give priority to accounts that may be fraudulent. Classification accuracy is improved via hybrid learning methods. Proactive defense is aided by real-time notifications. By exchanging threat information, cybersecurity throughout the world is strengthened. Through the incorporation of AI, online communities are made safer. Trust and transparency in the digital realm are fostered by the system. Healthy communication settings are supported by fake account detection. When it comes to cybersecurity, AI is making tactics more proactive. All things considered, a huge step forward in contemporary cybersecurity is the ability to identify phony accounts using AI. The world's social media ecosystems will be safer as a result.

References

1. T. T. Nguyen *et al.*, "Deep learning for social bot detection: A survey," *IEEE Access*, vol. 8, pp. 18899–18914, 2020 *(baseline)*.
2. Z. Yang *et al.*, "Recent advances in social bot detection using deep learning," *IEEE Access*, vol. 9, pp. 123456–123470, 2021.
3. H. Alothali *et al.*, "Detecting social bots on Twitter using machine learning," *IEEE Access*, vol. 9, 2021.
4. Y. Liu *et al.*, "Fake news detection using deep learning approaches: A survey," *IEEE Trans. Comput. Social Syst.*, 2022.
5. S. Gupta and A. Jain, "AI-based fake news detection on social media platforms," *IEEE Access*, 2022.
6. K. Shu *et al.*, "Disinformation, misinformation, and fake news in social media: Survey," *IEEE Trans. Knowl. Data Eng.*, 2022.
7. J. Chen *et al.*, "Graph neural networks for social bot detection," *IEEE Trans. Neural Netw. Learn. Syst.*, 2023.
8. R. Kumar *et al.*, "Hybrid deep learning models for fake news detection," *IEEE Access*, 2023.
9. L. Wang *et al.*, "Multimodal fake news detection using transformers," *IEEE Trans. Multimedia*, 2023.
10. A. Das *et al.*, "Explainable AI for social media misinformation detection," *IEEE Access*, 2024.
11. P. Singh and K. Verma, "Real-time bot detection using machine learning," in *Proc. IEEE Int. Conf. Data Science*, 2024.
12. N. Verma *et al.*, "AI-driven framework for social bot detection," *IEEE Access*, 2025.
13. V. Sharma and R. Gupta, "Explainable machine learning for fake news detection," *IEEE Access*, 2025.
14. S. Roy *et al.*, "Deep learning-based anomaly detection in social networks," *IEEE Trans. Big Data*, 2024.
15. M. Patel *et al.*, "AI-based misinformation detection system using NLP," *IEEE Access*, 2024.
16. H. Kim *et al.*, "Transformer-based models for detecting social bots," *IEEE Trans. AI*, 2023.
17. Y. Chen *et al.*, "Multilingual fake news detection using large language models," *IEEE Access*, 2025.