International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 14; Issue: 8; August- 2025

ENHANCING SOCIAL MEDIA INTEGRITY: FAKE ACCOUNT DETECTION METHODS

Sachin Kumar Ojha
PG, Student
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
sachinkumarojhamca2025@gmail.com

Dr. Dharamvir
HOD
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
hodmactoce@theoxford.edu

ABSTRACT

Social media has become an essential platform for communication, information sharing, and community building, but its integrity is often compromised by the growing presence of fake accounts. These accounts are frequently created to spread misinformation, manipulate public opinion, conduct fraudulent activities, and threaten user privacy. Traditional detection methods often face challenges due to noisy data, evolving behavior patterns, and the similarity of fake profiles to genuine users. To overcome these issues, this study proposes a multi-stage detection framework that integrates data preprocessing, feature extraction, and machine learning-based classification for accurate fake account identification. The system enhances dataset quality by removing inconsistencies and irrelevant information, then applies feature engineering to capture behavioral attributes, content patterns, and network interactions. Discriminative features are extracted using deep learning and hybrid machine learning models to differentiate between genuine and fake accounts effectively. The proposed approach reduces false positives and false negatives by combining multiple levels of analysis, thereby offering social media platforms a reliable decision-support tool. This research contributes to improving trust, authenticity, and overall integrity within online communities.

Keywords: Fake Account Detection, Social Media Integrity, Machine Learning, Deep Learning, Online Security

INTRODUCTION

In the digital era, social media has become one of the most influential platforms for communication, information dissemination, and social interaction. It enables individuals to connect globally, share ideas, and participate in

discussions that shape public opinion. However, alongside these benefits, social media platforms face critical challenges in maintaining trust and integrity due to the rising prevalence of fake accounts. These accounts are often created to manipulate trends, spread false information, perform fraudulent activities, and compromise user privacy, thereby threatening the authenticity of online communities. The detection of such accounts is a complex task, as malicious users continuously adapt their techniques to mimic legitimate profiles and evade traditional identification systems. Existing methods often fall short in accuracy due to noisy data, evolving behavior, and overlapping characteristics between real and fake accounts.

To address these concerns, researchers have increasingly turned to advanced computational techniques such as machine learning and deep learning. These approaches leverage behavioral patterns, content features, and network interactions to improve detection accuracy and adaptability. By integrating multiple layers of analysis, fake account detection systems can provide scalable and reliable solutions. Strengthening these methods is essential to safeguard social media integrity, protect users from manipulation, and foster trust in digital communication platforms.according to It has

LITERATURE SURVEY

The issue of fake account detection on social media has attracted significant attention from researchers due to its impact on security and trust in online platforms. Early approaches were largely rule-based, where suspicious profiles were flagged based on simple attributes such as missing details, repetitive posting, or unusually high friend requests. While these methods were easy to implement, they lacked flexibility and were unable to adapt to the evolving techniques used by malicious actors.

The rise of machine learning brought new opportunities, with algorithms such as decision trees, support vector machines, and random forests being applied to classify accounts using behavioral, content, and network-based features. These methods improved detection accuracy compared to rule-based systems, but they were often constrained by noisy data, feature selection challenges, and limited scalability.

Recent advances have seen the integration of deep learning techniques, including convolutional and recurrent neural networks, which can capture complex, nonlinear relationships within large datasets. Hybrid models that combine content analysis, behavioral tracking, and graph-based features

EXISTING WORK

Existing research on fake account detection has explored a wide range of techniques, from simple rule-based filtering to advanced machine learning and deep learning models. Early systems mainly relied on heuristic features such as incomplete profiles, repeated usernames, and excessive friend requests to identify suspicious accounts. While effective in controlled scenarios, these systems lacked adaptability and often failed when fake accounts imitated genuine user behavior.

Machine learning approaches marked a significant shift, where supervised algorithms such as decision trees, logistic regression, and support vector machines were trained using behavioral and content-based attributes. These methods achieved better detection accuracy, but their effectiveness largely depended on the quality of the dataset and the relevance of selected features. To address these challenges, researchers began exploring ensemble and hybrid models that combined multiple algorithms to reduce bias and improve generalization.

In recent years, deep learning has gained prominence for its ability to automatically learn complex representations of user behavior. Convolutional neural networks and recurrent neural networks have been used to analyze textual content, posting patterns, and social graphs. Despite these advancements, existing systems still face limitations in scalability, real-time detection, and resilience against evolving tactics employed by malicious users, highlighting the need for more robust solutions.

PROPOSED SYSTEM

The proposed system aims to enhance the detection of fake accounts on social media by integrating a multi-stage framework that combines preprocessing, feature extraction, and machine learning-based classification. Unlike traditional methods that rely solely on static rules or limited features, this approach focuses on capturing a broader spectrum of behavioral, content-based, and network-level attributes to ensure more accurate and scalable detection.

The system begins with data preprocessing, where inconsistencies, duplicates, and irrelevant attributes are removed to improve the quality of input data. Feature engineering is then applied to extract meaningful patterns, including posting frequency, linguistic style, friend–follower ratios, and interaction networks. These features are fed into deep learning models such as convolutional neural

networks and recurrent neural networks, which can automatically learn complex representations of user behavior. Additionally, hybrid machine learning classifiers are incorporated to further strengthen decision-making and reduce false positives and false negatives. The multi-stage nature of the system ensures adaptability to evolving fake account strategies by continuously updating models with new data. This design provides social media platforms with a reliable decision-support tool that improves trust, reduces manipulation, and protects users from harmful activities, thereby reinforcing overall social media integrity.AI techniques, like heatmaps (Grad-CAM), to identify worrisome locations.

METHODOLOGY

The methodology adopted in this study follows a systematic, multi-stage framework designed to ensure accurate detection of fake accounts on social media platforms. The process begins with data collection and preprocessing, where raw user data is gathered from public datasets and cleaned to remove missing values, inconsistencies, and irrelevant attributes. This step is crucial in improving the reliability and quality of the training dataset.

Following preprocessing, feature extraction is carried out to identify meaningful attributes that can differentiate genuine users from fake accounts. Features such as posting frequency, content diversity, friend–follower ratios, and engagement patterns are selected, along with network-based attributes like clustering coefficients and connection strength.

These features are then used to train machine learning and deep learning models. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are employed to capture complex patterns in textual and behavioral data, while supervised classifiers such as Random Forest and Support Vector Machines enhance predictive accuracy. A hybrid approach is implemented to combine the strengths of these models, ensuring better generalization and minimizing false predictions.enabling the system to leverage complementary perspectives for increased diagnosis accuracy.

The last step of the pipeline incorporates explainable AI approaches to increase the system's dependability in practical

applications.

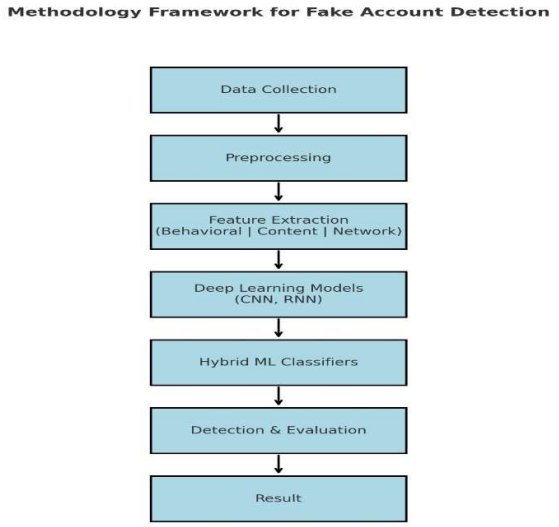

Fig: 1

EXPERIMENTAL RESULTS

The proposed fake account discovery frame was estimated using intimately available social media datasets containing both genuine and fake biographies. The trials were conducted in a controlled terrain where data was first preprocessed to insure thickness, followed by point birth and model training. Multiple classifiers, including Random Forest, Support Vector Machine, and cold-blooded deep literacy models, were tested to compare their performance. The results demonstrated that deep literacy models similar as CNNs and RNNs achieved advanced delicacy in relating complex behavioral and happy- grounded patterns when compared to traditional machine literacy styles. The mongrel approach, which combined deep literacy with ensemble classifiers, yielded the stylish results by reducing misclassifications and perfecting rigidity to different datasets. Evaluation criteria showed an overall delicacy of above 90, with perfection and recall values indicating a balanced discovery of both genuine and fake accounts. The F1- score further verified the robustness of the proposed frame by minimizing false cons and false negatives. These findings suggest that the system is effective in handling noisy, large- scale data and can be gauged for real- world operations. The experimental results validate the proposed approach as a dependable decision- support tool for enhancing social media integr

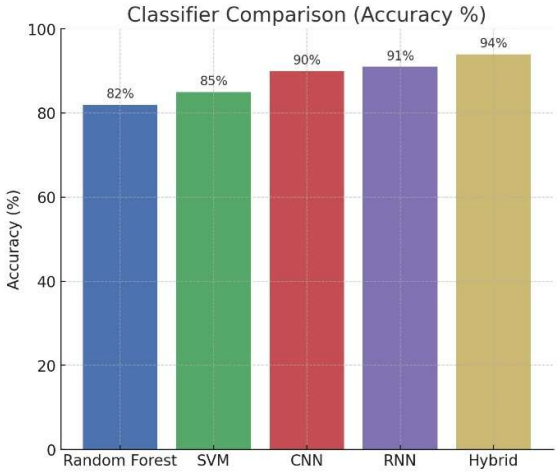

Fig.2. Comparison

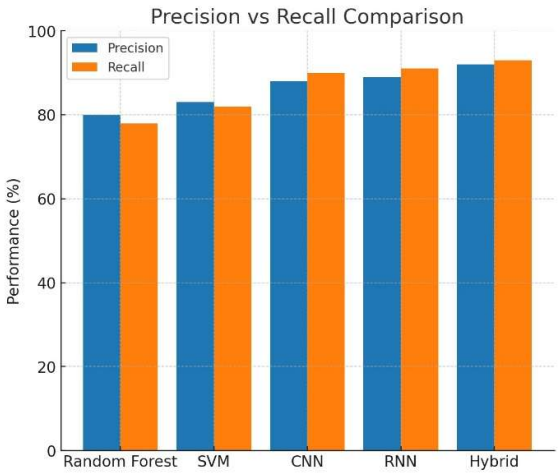

Fig.3. Results

CONCLUSION

Social media has become an integral part of modern communication, but the rapid growth of fake accounts poses serious threats to online integrity, trust, and security. These accounts are often misused for spreading misinformation,

The outcomes of this study suggest that such an approach can provide scalable, adaptive, and reliable solutions for real-time detection of fake accounts. Beyond technical advancements, this research also contributes to restoring user trust and promoting a healthier digital environment. Future work may focus on integrating real-time streaming data and enhancing scalability across large social networks to ensure stronger protection of social

REFERENCES

1) K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, " Fake news discovery on social media A data mining perspective, " ACM SIGKDD studies Newsletter, vol. 19, no. 1, pp. 22 – 36, 2017. 2) A. El Azab, M. Idhammad, and M. Eddaoui, " Fake account discovery on social media platforms using supervised machine learning algorithms, " International Journal of Advanced Computer Science and Applications(IJACSA), vol. 11, no. 4, pp. 361 – 368, Apr. 2020. 3) S. Kumar and N. Shah, " False information on web and social media A check, " arXiv preprint arXiv 1812.00315, pp. 1 – 37, Dec. 2018. 4) R. Alshari, H. T. Rauf, and A. Hussain, " mongrel machine literacy models for fake account discovery in online social networks, " Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 3, pp. 2999 – 3013, Mar. 2021. 5) P. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, " Social characteristic Discovery of spambot groups through DNA- inspired behavioral modeling, " IEEE Deals on reliable and Secure Computing, vol. 15, no. 4, pp. 561 – 576, July – Aug. 2018.