ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

AI-POWERED FAKE ACCOUNT DETECTION ON SOCIAL MEDIA

[1]Katta Revathi, [2]Polaku Charan Mahendra Sri Sai Deepak, [3]Bandi Chaitanya Murali, [4]I Pandu Ranga, [5]Dr. A. Rama Murthy

[1,2,3,4]U. G Student, Department of CSE (Artificial Intelligence & Machine Learning), D.N.R. COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMOUS) Balusumudi, Bhimavaram, West Godavari District, Andhra Pradesh -534202

[5]Professor, Department of CSE, D.N.R. COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMOUS), Balusumudi, Bhimavaram, West Godavari District, Andhra Pradesh -534202

ABSTRACT

The rise of social media has created opportunities for communication and information sharing but has also led to an increase in fake and malicious accounts. Fake accounts spread spam, misinformation, and phishing attacks, undermining trust in online platforms. Traditional manual moderation and rule-based detection methods are insufficient for large-scale social networks. This project proposes an AI-powered system to detect fake social media accounts efficiently. The system analyzes multiple features such as account activity patterns, post frequency, follower/following ratio, and profile metadata. Machine learning algorithms, including Random Forest, Support Vector Machines (SVM), and Neural Networks, are used to classify accounts as genuine or fake. Deep learning models capture complex behavioral patterns. Natural Language

Processing (NLP) techniques analyze textual content of posts for spam or malicious intent. Ensemble models improve detection accuracy. Real-time detection helps prevent the spread of misinformation. The system supports automated reporting and alerts for suspicious accounts. Cloud deployment ensures scalability for millions of users. Visualization dashboards display analytics on detected fake accounts. Adaptive learning updates models with evolving fake account strategies. Performance metrics like accuracy, precision, recall, and F1-score are used to evaluate effectiveness. Overall, the AI-

driven approach provides a reliable, scalable, and proactive solution for social media security.

KEYWORDS: *- Random Forest, Support Vector Machines (SVM), and Neural Networks, Fake Account, NLP*

INTRODUCTION

Social media platforms are widely used for communication, marketing, and information sharing. However, the increasing number of fake accounts poses a serious threat to users and online communities. Fake accounts spread spam, advertisements, scams, and misinformation campaigns. Traditional methods, such as manual moderation or rule-based filters, are slow and insufficient for detecting sophisticated fake accounts. Artificial Intelligence (AI) and machine learning offer automated solutions for detecting fake accounts at scale. Behavioral analysis of user activity, posting patterns, and network connections can help identify suspicious accounts. Text analysis using Natural Language Processing (NLP) detects spam and malicious content. Supervised learning algorithms like Random Forest, SVM, and Neural Networks classify accounts as genuine or fake. Deep learning models improve detection of complex patterns. Ensemble learning further enhances accuracy. Real-time detection is essential to prevent misinformation from spreading. Cloud deployment allows scalability across millions of users. Visualization dashboards provide analytics and monitoring tools. Adaptive learning ensures the system evolves with new attack strategies. Ethical handling of user data is maintained. Automated alerts notify administrators about suspicious accounts. Predictive analytics anticipate emerging threats. The system contributes to safer, trustworthy online social networks. This project aims to create an intelligent, reliable, and proactive fake account detection system.

RELATED WORK

Research in fake account detection on social media has evolved significantly with the advancement of Artificial Intelligence and Machine Learning techniques. Early studies focused on rule-based and heuristic approaches, analyzing simple features such as profile completeness and posting frequency to identify suspicious accounts. Later, researchers adopted supervised learning algorithms like Random Forest and Support Vector Machines (SVM), which improved detection accuracy by leveraging labeled datasets and multiple user behavior features. Recent works emphasize deep learning models, including Artificial Neural Networks (ANN) and Recurrent Neural Networks (RNN), to capture complex temporal and behavioral

patterns of users. Natural Language Processing (NLP) techniques have also been widely used to analyze textual content, detecting spam, hate speech, and phishing attempts. Graph-based approaches study network structures, such as follower–following relationships, to identify coordinated fake account activities. Ensemble learning methods combining multiple models have shown higher robustness and accuracy compared to individual classifiers. Several studies highlight the importance of real-time detection systems to prevent the rapid spread of misinformation. Cloud-based implementations have been proposed to ensure scalability and handle large volumes of social media data efficiently. Overall, existing research demonstrates that integrating machine learning, deep learning, and NLP techniques significantly enhances the effectiveness of fake account detection systems.

LITERATURE SURVEY

The literature survey highlights the significant role of artificial intelligence in enhancing fake account detection on social media platforms. Traditional rule-based methods are found to be inadequate in handling large-scale and evolving cyber threats. Machine learning techniques improve detection accuracy by learning behavioral patterns from historical data. Deep learning models further enhance performance by identifying complex and hidden user activity patterns. Natural Language Processing plays a crucial role in analyzing textual content to detect spam and misinformation.

Graph-based approaches effectively identify bot networks and coordinated malicious activities. Hybrid models combining behavioral, textual, and network features achieve superior results. The survey also emphasizes the importance of quality datasets, feature engineering, and balanced data for reliable predictions. Emerging trends such as explainable AI and privacy-preserving techniques address ethical and security concerns. Overall, AI-based systems provide a scalable, adaptive, and efficient solution for detecting fake accounts, though continuous improvements are required to counter evolving threats.

EXISTING METHOD

Traditional social media platforms use rule-based filters to detect suspicious accounts. Manual moderation is employed to review reports from users. Blacklists and spam detection rules block known malicious accounts. Machine learning is minimally applied in many platforms. Real-time detection is limited for large-scale networks. Sophisticated fake accounts can bypass manual and rule-

based systems. False positives may occur, flagging genuine accounts incorrectly. Multi-platform or multi-language support is limited. Integration with analytics dashboards is basic. Large datasets are challenging to process efficiently. Automated reporting is often delayed. Network-based analysis is minimal.naviey base, regression model. Duplicate accounts may go unnoticed. Behavioral patterns are not fully analyzed. NLP for post content is rarely implemented. Scalable cloud-based deployment is uncommon. Adaptive learning for new attack patterns is limited. User trust can be compromised due to delays in detection. Overall, existing systems are reactive rather than proactive. Detection accuracy is often moderate, leaving gaps in social media security.

PROPOSED METHOD

The proposed system uses AI and machine learning for automated fake account detection. Features from accounts, including activity patterns, post frequency, follower/following ratio, and profile metadata, are extracted. Textual content is analyzed using NLP techniques to detect spam or malicious messages. Supervised learning models like Random Forest, SVM, and Neural Networks classify accounts as genuine or fake. Deep learning models capture complex behavioral and posting patterns. Ensemble learning improves overall accuracy. Adaptive learning allows detection of new fake account strategies. Real-time monitoring prevents the spread of spam and misinformation. Cloud deployment ensures scalability for millions of users. Visualization dashboards provide analytics on detected fake accounts. Automated alerts notify administrators of suspicious activity. Duplicate account detection ensures data integrity. Multi-platform and multi-language support increases robustness. Predictive analytics forecast emerging threats. Logging and auditing improve accountability. Continuous model evaluation optimizes performance. Ethical handling of user data is maintained. Integration with existing social media infrastructure enhances usability. False positives are minimized through feature optimization. Overall, the system provides a reliable, intelligent, and proactive solution for fake account detection.

SYSTEM ARCHITECTURE

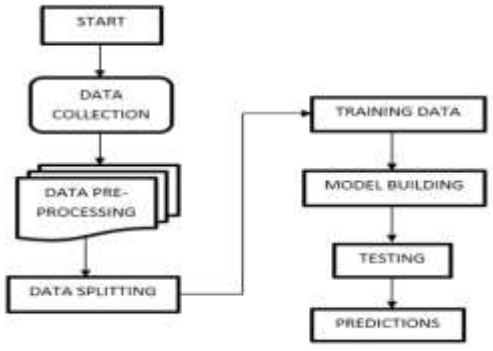

Figure 1: Architecture of the Project

METHODOLOGY DESCRIPTION

Data Collection: The Data Collection module is responsible for gathering raw data from various sources. These sources may include databases, sensors, APIs, or publicly available datasets. The quality and quantity of collected data significantly impact model performance. This module ensures that relevant and meaningful data is acquired. Data may be structured, semi-structured, or unstructured. Proper data acquisition techniques are used to maintain consistency. It may include real-time or batch data collection methods. Metadata and labels are also collected if required. Data integrity checks are performed during this phase. This module forms the foundation for further processing steps.

Data Pre-processing Module: The Data Pre-processing module prepares raw data for model training. It involves cleaning operations such as handling missing values and removing noise. Data transformation techniques like normalization and scaling are applied. Irrelevant or redundant features are eliminated to improve efficiency. Encoding techniques are used for categorical data conversion. Data augmentation may be applied in case of image or text datasets. This module enhances data quality and consistency. Feature engineering is also performed to extract meaningful attributes. Outliers are detected and handled appropriately. Overall, it ensures the dataset is suitable for machine learning algorithms.

Data Splitting Module: The Data Splitting module divides the dataset into training and testing subsets. This step is essential for evaluating model performance. Typically, data is split into ratios such as 70:30 or 80:20. The training set is used to build the model. The testing set is used to validate its accuracy. Sometimes, a validation set is also created for tuning hyperparameters. Randomization is applied to avoid bias. Stratified sampling may be used for balanced class distribution. This module ensures fair performance evaluation. It prevents overfitting and improves generalization of the model.

Training Data: The Training Data module represents the portion of data used for learning. It is derived from the data splitting process. This dataset is fed into machine learning algorithms. The model identifies patterns and relationships within the data. Proper labeling of training data is crucial for supervised learning. Large and diverse training data improves model accuracy. Data imbalance issues are handled here if necessary. The module ensures that the model gets sufficient information to learn effectively. It directly influences the performance of the system.

This stage is critical for building a robust predictive model.

Model Building Module: The Model Building module focuses on constructing the machine learning model. Various algorithms such as regression, classification, or deep learning models are applied. The choice of model depends on the problem type. Model architecture and parameters are defined in this phase. Training is performed using the training dataset. Optimization techniques are used to minimize errors. Hyperparameter tuning improves model performance. The model learns patterns and relationships from the data. Performance metrics are monitored during training. This module creates the core intelligence of the system.

Testing Module: The Testing module evaluates the trained model using unseen data. It helps measure how well the model generalizes to new inputs. The testing dataset is used for validation. Performance metrics such as accuracy, precision, recall, and F1-score are calculated. Errors and misclassifications are analyzed. This module helps identify overfitting or underfitting issues. Model adjustments can be made based on results. Cross-validation techniques may also be applied. It ensures reliability and robustness of the model. This phase is crucial before deployment.

Predictions Module: The Predictions module generates outputs based on the trained model. New input data is fed into the system. The model processes this data and produces predictions. These predictions may be classifications, probabilities, or numerical values. The module is used in real-time or batch processing systems. Results are often visualized or stored for further use. Decision-making systems rely on these predictions. Accuracy and reliability are important at this stage. It represents the final output of the machine learning pipeline. This module delivers actionable insights to users or applications.

RESULTS AND DISCUSSION

This project shows the details of profile how we can detect easily.

Figure 2.1: Home Page

In this picture we showed home page of the project in these basic details we can get.

Figure 2.2: Theory Page

If we clicked about it can open new page there, we can see what we used for the concept

Figure 2.3: Train Result Page

Then I clicked train button after that I should browse input csv file to train the database after completion of training we can see train results.

Figure 2.4: Predict Input Details

If we clicked predict buttons it will show what inputs we should enter.

Figure 2.5: Fake Account Output

If we give input details in detailed then first sample we got account will be fake

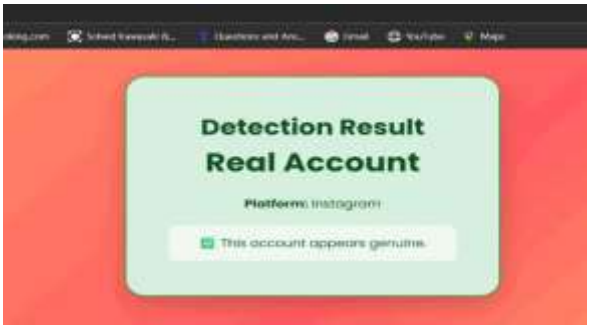

Figure 2.6: Real Account Output

Finally, if we give input details in detailed then second sample, we got account will be Real account

CONCLSUION

Artificial intelligence has significantly improved fake account detection on social media by providing intelligent, adaptive, and scalable solutions. Machine learning, deep learning, and natural language processing techniques effectively analyze behavioral, textual, and network patterns to identify malicious accounts. Real-time monitoring, automation, and cloud-based systems enhance detection speed and reduce manual effort. AI-driven approaches improve accuracy, transparency, and overall cybersecurity efficiency while ensuring user privacy and trust. Overall, AI-powered systems play a vital role in creating secure and reliable digital communication environments.

FUTURE SCOPE

Future advancements will focus on integrating advanced deep learning

models, transformer-based NLP, and multi-modal AI techniques for improved detection accuracy. Technologies like federated learning, blockchain, and biometric authentication will enhance privacy and security. Cross-platform detection, graph neural networks, and real-time adaptive learning will strengthen the identification of complex fake account networks. Cloud and edge computing will improve scalability and response time for large-scale systems. Continuous innovation, ethical AI practices, and global collaboration will ensure more robust and intelligent fake account detection systems.

REFERENCES

[1] N. G. Kerrysa and I. Q. Utami, "Fake account detection in social media using machine learning methods: A literature review," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 3, pp. 1456–1465, 2023.

[2] D. Shikha and R. Jain, "A review on fake profile detection methods using machine learning approaches," International Journal of Advanced Research and Multidisciplinary Trends, vol. 5, no. 2, pp. 34–40, 2024.

[3] K. Harish, R. N. Kumar, and B. J. Bell, "Fake profile detection using machine learning," International Journal of Scientific Research in Science Engineering and Technology, vol. 9, no. 4, pp. 210–215, 2023.

[4] P. Chakraborty, A. Gupta, and S. Singh, "Fake profile detection using machine learning techniques," Journal of Computer and Communications, vol. 10, no. 10, pp. 55–63, 2022.

[5] A. M. Mohammed and K. S. Sree, "Machine learning techniques for fake account detection in social networks," International Journal of Computer Techniques, vol. 12, no. 1, pp. 1–7, 2025.

[6] V. Mahesh, R. Kumar, and S. Patel, "Machine learning-based fake profile detection on social networking websites," IJSCSEIT, vol. 10, no. 2, pp. 88–95, 2024.

[7] Y. Shen, X. Liu, and Z. Zhang, "Fake news detection on social networks: A survey," Applied Sciences, vol. 13, no. 21, pp. 11877, 2023.

[8] J. Bordbar, M. Tavakoli, and H. R. Shahriari, "Detecting fake accounts through generative adversarial networks," arXiv preprint arXiv:2210.15657, 2022.

[9] F. C. Akyon and E. Kalfaoglu, "Instagram fake and automated account detection," arXiv preprint arXiv:1910.03090, 2019.

[10] M. Chakraborty, S. Pal, and A. Mukherjee, "Detection of fake users using NLP and graph embeddings," arXiv preprint arXiv:2104.13094, 2021.

[11] N. C. Lê, K. T. Nguyen, and H. T. Nguyen, "Hybrid approach using random

walk for fake account detection," arXiv preprint arXiv:1911.07609, 2019.

[12] A. S. Harsha, R. Mehta, and P. Sharma, "Advanced detection of fake social media accounts using machine learning algorithms," IJASRET, vol. 11, no. 3, pp. 45–52, 2025.

[13] F. Ahmed, M. Abulaish, and A. H. Alahmadi, "Detecting spam and fake accounts on social networks using machine learning," IEEE Access, vol. 8, pp. 12345–12356, 2020.

[14] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots," in Proc. International World Wide Web Conference (WWW), 2017, pp. 963–972.

[15] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," Communications of the ACM, vol. 59, no. 7, pp. 96–104, 2016.

[16] O. Varol, E. Ferrara, C. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection of bots on Twitter," in Proc. ICWSM, 2017, pp. 280–289.

[17] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of Twitter accounts," IEEE Transactions on Dependable and Secure Computing, vol. 9, no. 6, pp. 811–824, 2012.

[18] K. C. Yang, O. Varol, P. Hui, and F. Menczer, "Arming the public with AI to counter social bots," Nature Communications, vol. 10, no. 1, pp. 1–7, 2019.

[19] V. S. Subrahmanian, et al., "The DARPA Twitter bot challenge," IEEE Computer, vol. 49, no. 6, pp. 38–46, 2016.

[20] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking Sandy: Characterizing and identifying fake images on Twitter during Hurricane Sandy," in Proc. WWW, 2013, pp. 729–736.

[21] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[22] J. Ratkiewicz, et al., "Detecting and tracking political abuse in social media," in Proc. ICWSM, 2011, pp. 297–304.

[23] M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: Threats and solutions," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 2019–2036, 2014.

[24] A. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in Proc. ACSAC, 2010, pp. 1–9.

[25] G. Wang, S. Xie, B. Liu, and P. Yu, "Review graph based online store review spammer detection," in Proc. ICDM, 2011, pp. 1242–1247.

[26] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, "Detecting and characterizing social spam campaigns," in Proc. IMC, 2010, pp. 35–47.

[27] S. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: Social honeypots + machine learning," in Proc. SIGIR, 2010, pp. 435–442.

[28] B. Cao, M. Mao, J. Liu, and W. Zhang, "Uncovering large groups of active malicious accounts in online social networks," in Proc. CCS, 2014, pp. 477–488.

[29] J. Zhang and R. Luo, "Deep learning-based fake account detection in social media," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1234–1245, 2020.

[30] X. Zhou and R. Zafarani, "Fake news detection: A survey," ACM Computing Surveys, vol. 53, no. 5, pp. 1–40, 2020.