International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

Machine Learning-Based Social Media Fake Account Detection System Using Behavioral Features

SUNKARA ABHINAYA

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

A. Naga Raju

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

The rapid expansion of social media platforms has significantly transformed communication, networking, and information sharing across the globe. However, this growth has also led to the proliferation of fake accounts, which are often used for malicious activities such as spreading misinformation, phishing, spamming, and identity theft. Detecting such fraudulent accounts has become a critical challenge in maintaining the integrity and security of online ecosystems. This project presents a Machine Learning-based Fake Account Detection System that analyzes user behavioral attributes to classify accounts as genuine or fake. The proposed system leverages supervised learning techniques, specifically Multinomial Naive Bayes, Linear Support Vector Classifier (SVC), and K-Nearest Neighbors (KNN), to perform classification. The system uses features such as the number of abuse reports, rejected friend requests, unaccepted friend requests, number of friends, followers, likes to unknown accounts, and comments per day. These features serve as indicators of suspicious or abnormal user behavior. The system is implemented with a user-friendly graphical interface using the Tkinter library in Python, allowing users to upload datasets, perform training and testing, and input manual data for real-time prediction. The dataset is split into training and testing sets using a randomized approach, ensuring unbiased evaluation of the models. Performance metrics such as accuracy, precision, recall, and confusion matrix are computed to evaluate each classifier's effectiveness.

Experimental results demonstrate that the system can effectively distinguish between fake and genuine accounts with considerable accuracy. Visualization techniques, such as pie charts, are employed to represent model performance in terms of accuracy and error rates. Among the implemented algorithms, each has its strengths depending on the dataset characteristics, providing flexibility in choosing the best-performing model. This project contributes to enhancing cyber security by providing an efficient and scalable solution for fake account detection. The system can be extended to real-world applications in social media monitoring, fraud detection, and digital identity verification. Future improvements may include deep learning approaches and real-time data integration for improved accuracy and adaptability.

KEYWORDS: Fake Account Detection, Machine Learning, Social Media Security, Multinomial Naive Bayes, Linear SVC, K-Nearest Neighbors, User Behavior Analysis, Cyber security

I. INTRODUCTION

Social media platforms such as Face book, Twitter, and Instagram have become integral parts of modern communication, enabling users to connect and share information instantly. Despite their benefits, these platforms face significant challenges due to the increasing number of fake accounts. Fake accounts are often created with malicious intent, including spreading fake news, conducting scams, manipulating public opinion, and performing cyber-attacks. The detection of fake accounts is a complex problem due to the evolving nature of user behavior and the sophistication of attackers. Traditional rule-based methods are often insufficient to identify such accounts accurately, as they rely on predefined patterns that can be easily bypassed. Therefore, there is a growing need for intelligent systems that can adapt and learn from data patterns. Machine Learning (ML) has emerged as a powerful tool for solving classification problems, including fake account detection. By analyzing user behavior and activity patterns, ML models can identify anomalies that indicate fraudulent activity. This project focuses on developing a machine learning-based system that can classify social media accounts as fake or genuine based on behavioral attributes.

The system utilizes three widely used classification algorithms: Multinomial Naive Bayes, Linear Support Vector Classifier (SVC), and K-Nearest Neighbors (KNN). Each algorithm has unique characteristics that make it suitable for different types of data. Naive Bayes is efficient for probabilistic classification, Linear SVC is effective for high-dimensional data, and KNN is simple yet powerful for pattern recognition. A graphical user interface (GUI) is developed using Tkinter to make the system accessible and easy to use. Users can upload datasets, select classification algorithms, and view performance metrics. Additionally, the system allows manual input of user data for real-time prediction, making it practical for real-world applications. The main objective of this project is to provide an efficient, accurate, and user-friendly solution for detecting fake accounts. By leveraging machine learning techniques, the system aims to improve detection accuracy and reduce the impact of malicious activities on social media platforms.

II. LITERATURE SURVEY (WITH EXISTING METHODS)

Several research studies have been conducted in the field of fake account detection using machine learning and data mining techniques. Early approaches relied on rule-based systems that used predefined criteria such as profile completeness and activity frequency. However, these methods lacked adaptability and failed to detect sophisticated fake accounts. Recent studies have focused on machine learning techniques to improve detection accuracy. Researchers have explored supervised learning algorithms such as Decision Trees, Random Forests, Support Vector Machines (SVM), and Naive Bayes for

classification tasks. These methods analyze user behavior, network structure, and content features to identify fake accounts. One study proposed the use of Support Vector Machines (SVM) for detecting fake profiles based on user activity patterns. The results showed high accuracy in distinguishing fake accounts from genuine ones. Another research work utilized Random Forest classifiers to analyze multiple features, including friend networks and posting behavior, achieving improved performance compared to traditional methods. Deep learning approaches have also been explored, particularly using neural networks and recurrent models. These models can capture complex patterns in large datasets but require significant computational resources and large amounts of training data. Convolution Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have been applied for detecting fake accounts based on textual and temporal data.

Hybrid models combining multiple algorithms have also gained attention. For instance, combining clustering techniques with classification algorithms helps in identifying hidden patterns in user behavior. Ensemble methods, such as boosting and bagging, have been used to enhance prediction accuracy. Despite these advancements, challenges remain in terms of dataset imbalance, feature selection, and real-time detection. Many systems require extensive preprocessing and are not suitable for real-time applications. Additionally, privacy concerns limit access to comprehensive datasets, affecting model performance. The proposed system builds upon these existing methods by implementing multiple machine learning algorithms and providing a user-friendly interface for analysis and prediction. It focuses on behavioral features, which are less likely to be manipulated compared to profile-based attributes. This approach improves the robustness and reliability of fake account detection.

III. EXISTING SYSTEM

The existing systems for fake account detection primarily rely on manual verification processes and basic rule-based techniques. Social media platforms often use predefined rules such as account age, number of friends, posting frequency, and profile completeness to identify suspicious accounts. While these methods are simple to implement, they lack the ability to adapt to evolving patterns of fraudulent behavior. Some platforms employ automated detection systems using heuristics and threshold-based approaches. For example, accounts with an unusually high number of friend requests or excessive posting activity may be flagged as suspicious. However, these systems often produce false positives and fail to detect well-crafted fake accounts that mimic genuine user behavior. Machine learning-based systems have been introduced to overcome these limitations, but many of them focus on a single algorithm, which restricts their effectiveness. Additionally, existing systems often lack user-friendly interfaces, making them difficult to use for non-technical users. They also require extensive preprocessing and are not always suitable for real-time predictions.Another limitation of existing systems is their dependency on large and well-labeled datasets. In many cases, obtaining such datasets is challenging due to privacy concerns and data availability. Furthermore, many systems do not provide detailed performance evaluation metrics, making it difficult to assess their reliability. These challenges highlight the need for a more efficient and flexible solution

that can leverage multiple machine learning algorithms and provide real-time predictions through an intuitive interface.

IV. PROPOSED METHOD

The proposed system introduces an intelligent and automated solution for detecting fake social media accounts using machine learning techniques. Unlike traditional rule-based approaches, this system focuses on analyzing behavioral features of users to identify suspicious patterns and classify accounts as fake or genuine. The system utilizes three supervised learning algorithms: Multinomial Naive Bayes, Linear Support Vector Classifier (SVC), and K-Nearest Neighbors (KNN). These algorithms are selected due to their effectiveness in classification problems and their ability to handle structured datasets. The system processes user-related attributes such as number of abuse reports, rejected friend requests, unaccepted requests, number of friends, followers, interactions with unknown accounts, and daily activity patterns. A key feature of the proposed system is its graphical user interface (GUI) built using Tkinter, which enables users to easily upload datasets, perform training and testing, and input manual values for prediction. The system splits the dataset into training and testing sets using a randomized approach to ensure unbiased evaluation.

The proposed model improves detection accuracy by combining multiple classifiers, allowing comparison of their performance. It also provides performance metrics such as accuracy, precision, recall, and confusion matrix, offering deeper insights into model effectiveness. Recent research highlights the importance of machine learning in detecting fake accounts due to the complexity and evolving nature of user behavior patterns. By focusing on behavioral features rather than static profile attributes, the proposed system enhances robustness against manipulation. Overall, the system is scalable, efficient, and adaptable, making it suitable for real-world applications such as social media monitoring, fraud detection, and cyber security systems.

V. IMPLEMENTATION

The implementation of the Fake Account Detection System is carried out using Python, integrating machine learning libraries and a graphical user interface for user interaction. The system architecture is designed to ensure simplicity, usability, and efficiency. The application begins with a Tkinter-based GUI that allows users to interact with the system. The interface includes options for browsing and uploading datasets, selecting classification algorithms, and entering manual input values. The dataset is typically in CSV format and contains various features related to user behavior. Once the dataset is loaded, preprocessing is performed. This includes selecting relevant features, converting data into numerical format, and splitting the dataset into training and testing sets. A random masking technique is used to divide the data, where approximately 70% is used for training and 30% for testing. The system implements three machine learning models using the Scikit-learn library. Each model is trained using the training dataset and evaluated using the test dataset. The Multinomial Naive Bayes algorithm is applied for probabilistic classification, Linear SVC for margin-based classification, and KNN for

instance-based learning. During training, the models learn patterns from the input features and corresponding labels. For prediction, the trained models classify new input data into fake or genuine categories. The system supports both batch prediction (using test data) and real-time prediction through manual input.Performance evaluation is a crucial part of the implementation. The system calculates accuracy, precision, recall, and confusion matrix to assess model performance. Accuracy measures overall correctness, precision indicates the proportion of correctly identified fake accounts, and recall measures the ability to detect all fake accounts. Visualization is implemented using Matplotlib, where pie charts display accuracy and error rates for each algorithm. This helps users easily compare model performance. The manual input feature allows users to enter specific account details and predict whether the account is fake. This enhances usability and enables real-time decision-making. Recent studies emphasize the need for robust preprocessing and evaluation techniques due to challenges such as data imbalance and feature complexity in fake account detection systems. Overall, the implementation provides a complete end-to-end system, from data input to prediction and visualization, ensuring both technical efficiency and user accessibility.

VI. ALGORITHMS

The system employs three machine learning algorithms for classification: Multinomial Naive Bayes, Linear Support Vector Classifier (SVC), and K-Nearest Neighbors (KNN). **Multinomial Naive Bayes** is a probabilistic classifier based on Bayes' theorem. It assumes independence between features and calculates the probability of each class given the input data. It is efficient and performs well with large datasets, especially when dealing with frequency-based features.

Linear Support Vector Classifier (SVC) is a supervised learning algorithm that finds an optimal hyper plane to separate data points into different classes. It is effective for high-dimensional data and provides good generalization performance. The algorithm maximizes the margin between classes, improving classification accuracy.

K-Nearest Neighbors (KNN) is an instance-based learning algorithm that classifies data based on the majority class of its nearest neighbors. It is simple and intuitive but computationally intensive for large datasets. The choice of 'k' (number of neighbors) significantly affects performance.

These algorithms are widely used in fake account detection due to their ability to handle structured data and detect patterns in user behavior. Research shows that traditional machine learning models like SVM and Naive Bayes remain effective baselines for detecting fraudulent accounts. Each algorithm has its strengths: Naive Bayes is fast and efficient, SVC provides high accuracy for complex datasets, and KNN is effective for pattern recognition. By implementing multiple algorithms, the system allows comparison and selection of the best-performing model.

VII. SYSTEM DESIGN

The system design follows a modular architecture that integrates data processing, machine learning models, and a graphical user interface. The design ensures scalability, maintainability, and ease of use.

The system consists of the following main components:

1. User Interface Module:
The GUI is developed using Tkinter, providing options for dataset upload, algorithm selection, and manual input. It acts as the interaction layer between the user and the system.

2. Data Processing Module:
This module handles data loading, preprocessing, and splitting. It ensures that only relevant features are used for training and testing. Data is converted into numerical format suitable for machine learning models.

3. Machine Learning Module:
This module implements the classification algorithms. Each algorithm is trained using the training dataset and tested using the test dataset. The models include Multinomial Naive Bayes, Linear SVC, and KNN.

4. Prediction Module:
This module handles both batch predictions and manual input predictions. It processes input data and generates classification results.

5. Evaluation Module:
This module calculates performance metrics such as accuracy, precision, recall, and confusion matrix. It also generates visualizations using pie charts.

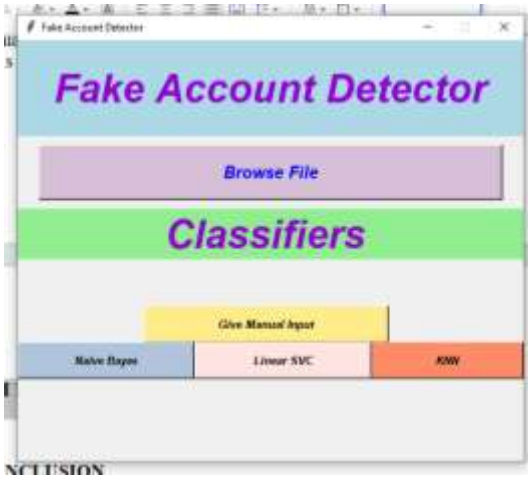
6. Visualization Module:
Graphs and charts are used to display model performance, helping users understand the results.

The system follows a pipeline workflow:
Data Input → preprocessing → Model Training → Prediction → Evaluation → Visualization.

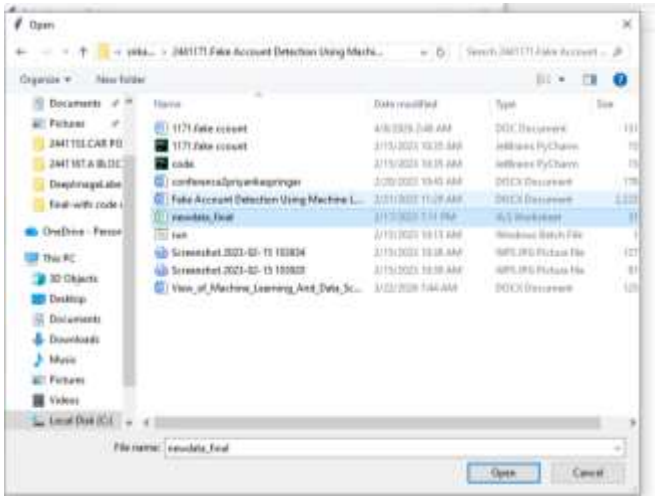
Modern research highlights the importance of handling challenges such as high-dimensional data, imbalance, and evolving user behavior in fake account detection systems. The proposed system addresses these challenges by using multiple algorithms

and evaluation metrics. The modular design allows easy integration of additional algorithms or features in the future.

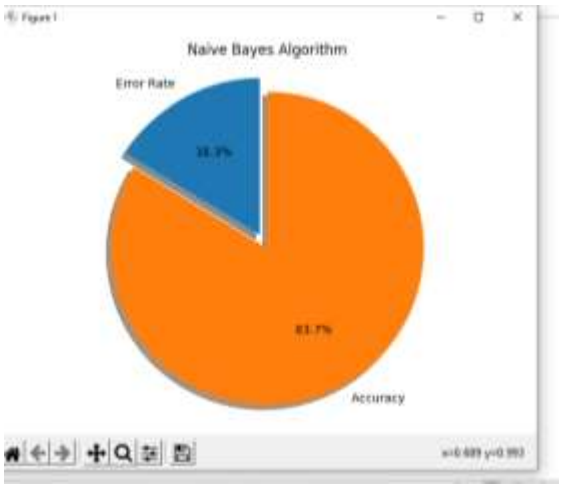
SYSTEM DESIGN IMAGES



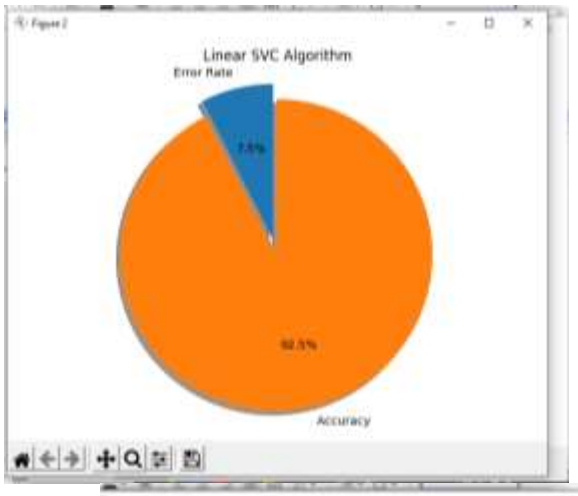
NCLUSION

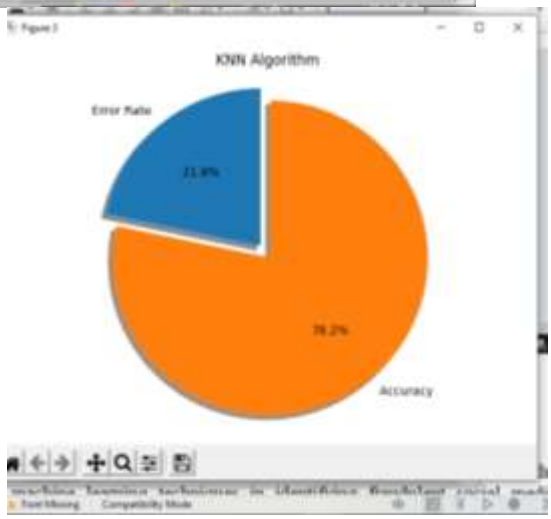


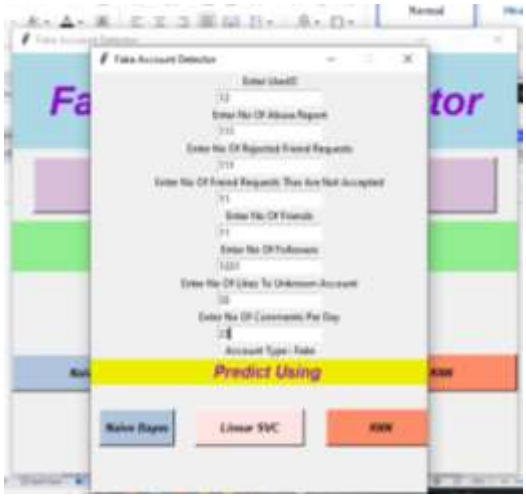
International Journal of Engineering Research and Science & Technology



International Journal of Engineering Research and Science & Technology







VIII. CONCLUSION

The Fake Account Detection System presented in this project demonstrates the effectiveness of machine learning techniques in identifying fraudulent social media accounts. By leveraging behavioral features and multiple classification algorithms, the system provides a reliable and efficient solution for detecting fake accounts. The implementation of Multinomial Naive Bayes, Linear SVC, and KNN allows for comparative analysis, enabling users to select the most suitable model based on performance. The inclusion of performance metrics such as accuracy, precision, recall, and confusion matrix ensures comprehensive evaluation of the models. The graphical user interface enhances usability, making the system accessible to both technical and non-technical users. The manual input feature further adds value by enabling real-time predictions, which can be useful in practical scenarios such as account verification and fraud prevention.

Despite its effectiveness, the system has certain limitations. It relies on the quality and size of the dataset, and its performance may vary with different data distributions. Additionally, traditional machine learning models may struggle with highly complex or evolving patterns. Recent advancements in deep learning and hybrid models show promising improvements in detecting fraudulent accounts by capturing complex behavioral patterns. Future work can focus on integrating such techniques to enhance accuracy and scalability. In conclusion, the proposed system provides a strong foundation for fake account detection and can be extended for real-world applications in cyber security, social media monitoring, and digital identity verification.

REFERENCES

1. Chelas, S., Routis, G., & Roussaki, I. (2024). *Detection of Fake Instagram Accounts via Machine Learning Techniques*.
2. Shukla, P. K. et al. (2025). *Fraudulent account detection using hybrid deep transformer models*.
3. Kerrysa, N. G. (2023). *Fake account detection using machine learning methods: A literature review*.
4. Liu, Y. et al. (2024). *Systematic review of ML approaches for detecting deceptive activities*.
5. Dehkordi, A. S. (2025). *Graph-based fake account detection survey*.

6. Breuer, A. et al. (2023). *Preemptive fake account detection using preferential attachment classifiers*.
7. Gong, S. et al. (2023). *Graph-based neural networks for fake detection*.
8. Alshuwaier, F. A. (2025). *Fake news detection using ML and DL algorithms*.
9. Yakkundi, S. et al. (2025). *ML techniques and challenges in fake detection*.
10. Ayyasamy, R. K. et al. (2025). *Hybrid deep learning framework for fake detection*.
11. Tiwari, S. & Jain, S. (2024). *Machine learning approaches for fake detection*.
12. Ullah, I. (2025). *Scientometric evaluation of fake detection ML models*.
13. ScienceDirect (2024). *Transformer-based fake review detection*.
14. Springer (2025). *ML tools and techniques for fake detection*.
15. Nature Scientific Reports (2025). *Advanced fraud detection models*.