STM JOURNALS

Journal of
Mobile Computing, Communications & Mobile Networks

ISSN: 2349-901X
Volume 11, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoMCCMN

https://journals.stmjournals.com/jomccmn

Review JoMCCMN

Detecting Fake Accounts on Social Media Using Machine Learning

Abhishek Agrawal[1,*], Priyanka Baste[1], Snigdha Take[1], Sachin Tupe[1], S.N. Botekar[2]

Abstract

The growing frequency of fake accounts on social media platforms underscores the critical necessity for effective detection methods. In response to this challenge, our study leverages state-of-the-art machine learning techniques to identify and counter deceptive entities effectively. By conducting a thorough analysis of social media data, our approach unveils intricate patterns indicative of fraudulent accounts, enabling proactive measures against them. Through the application of advanced algorithms, we present a comprehensive methodology for the identification and mitigation of such accounts, addressing current online security challenges and anticipating future trends. Moreover, our research not only contributes significantly to enhancing cybersecurity within the realm of social media but also lays the groundwork for future advancements in this field. By integrating machine learning methods, our study offers a nuanced and highly effective approach, thereby bolstering the overall security landscape of online platforms. This paper serves to underscore the considerable strides made in social media security and provides a robust foundation for further research and development initiatives aimed at safeguarding users from malicious activities.

Keywords: Machine learning (ML), natural language processing (NLP), receiver operating characteristic (ROC), area under the curve (AUC), true positive rate (true positive rate), false positive rate (FPR), support vector machine (SVM).

INTRODUCTION

In the vibrant tapestry of social media, the rise of fake accounts has woven a complex challenge, necessitating a harmonious blend of technological prowess and analytical finesse. As digital platforms evolve, the intricacies of misleading tactics also grow. In this realm, our endeavor intertwines the artistry of machine learning and the precision of data analysis to unravel the intricate threads of fraudulent accounts [1]. The increasing spread of false information and the decline in trust online highlight the importance of our mission. This paper embarks on a journey through the algorithmic corridors of pattern recognition, leveraging the power of machine learning, natural language processing (NLP), and deep learning to illuminate the concealed pathways of deceptive behaviors. As we navigate this dynamic landscape, we not only confront the current challenges of fake account detection but also cast our gaze towards the horizon of emerging trends. Our exploration, guided by the synergy of technology and pattern elucidation, aims not only to fortify the defenses against deceptive entities but also to sculpt a resilient foundation for the future integrity of online platforms [2, 3].

***Author for Correspondence**
Abhishek Agrawal
E-mail: abhishek0201agrawal@gmail.com

[1]Student, Department of Computer Engineering, Jawahar Education Society's Institute of Engineering, Nashik, Maharashtra, India.
[2]Professor, Department of Computer Engineering, Jawahar Education Society's Institute of Engineering, Nashik, Maharashtra, India.

Received Date: March 22, 2024
Accepted Date: June 25, 2024
Published Date: June 29, 2024

Citation: Abhishek Agrawal, Priyanka Baste, Snigdha Take, Sachin Tupe, S.N. Botekar. Detecting Fake Accounts on Social Media Using Machine Learning. Journal of Mobile Computing, Communications & Mobile Networks. 2024; 11(2): 21–32p.

PROJECT SETUP (INCLUDING LITERATURE REVIEW)

In the sprawling landscape of social media, the infiltration of fake accounts disrupts the very essence of digital connection. The motivation behind our project is fueled by the realization that these deceptive entities not only deceive individuals but sow seeds of doubt in the collective trust we place in online platforms. As technology rapidly evolves, the sophistication of fake accounts advances in parallel, demanding innovative solutions. We aim to lead innovation by leveraging machine learning, NLP, and deep learning to forge a revolutionary approach. The significance lies not only in detecting and curbing fake accounts also but in pioneering a proactive defense mechanism against evolving digital deception. Our motivation extends beyond mere identification; we aspire to empower users with a shield against misinformation, fostering a digital realm where authenticity prevails. In this pursuit, our project stands at the intersection of technology and societal integrity, aiming to reshape the narrative of trust in the digital age as shown in Table 1.

METHODOLOGY

Our approach to fake account detection marries the prowess of machine learning and NLP, creating a synergistic framework for robust identification. This section outlines the technical intricacies and stepwise implementation of our methodology [4].

Data Collection

Collect a broad dataset containing both authentic and counterfeit social media profiles, encompassing diverse user behaviors, posting frequencies, and patterns of engagement.

Pre-processing

Prepare the dataset by addressing missing values, eliminating unnecessary features, and standardizing data. Apply text preprocessing methods to handle textual data, such as tokenization and stemming as shown in Figure 1.

Feature Engineering

Extract relevant features such as account creation date, post frequency, user engagement, and linguistic patterns. Utilize domain specific features inspired by the literature review, enhancing the model's discriminatory power.

Model Selection

Employ a hybrid model combining supervised and unsupervised machine learning techniques. Utilize deep learning architectures for capturing intricate patterns in user behavior. Implement algorithms such as random forest or gradient boosting for ensemble learning, leveraging their efficiency in classification tasks as shown in Figure 2.

Table 1. Project setup (including literature review).

Author(s)	Title of study	Methodology	Key finding
Gayathri et al. [5].	Detecting fake accounts in media application using machine learning	Machine learning	Fake accounts, fake identities, social media, data science, friends, followers, fake profiles
Rao et al. [6]	Detecting fake account on social media using machine learning algorithms	Machine learning, support vector machine (SVM), neural network (NN)	Promotion, communications, agenda creation, advertisements, news creation
Bhambar et al. [4]	Detecting fake accounts on social media using neural network	Machine learning, SVM, NN	Facebook, fake accounts, feature selection, clustering, classification
Singh and Khan [7]	Detection of fake profile in social media	Machine learning	Privacy violation, fake profile

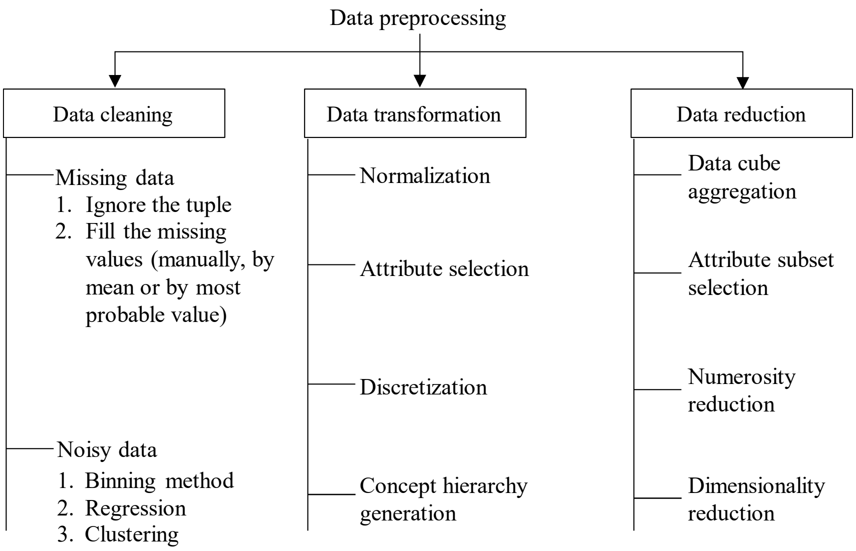

Figure 1. Data pre-processing.

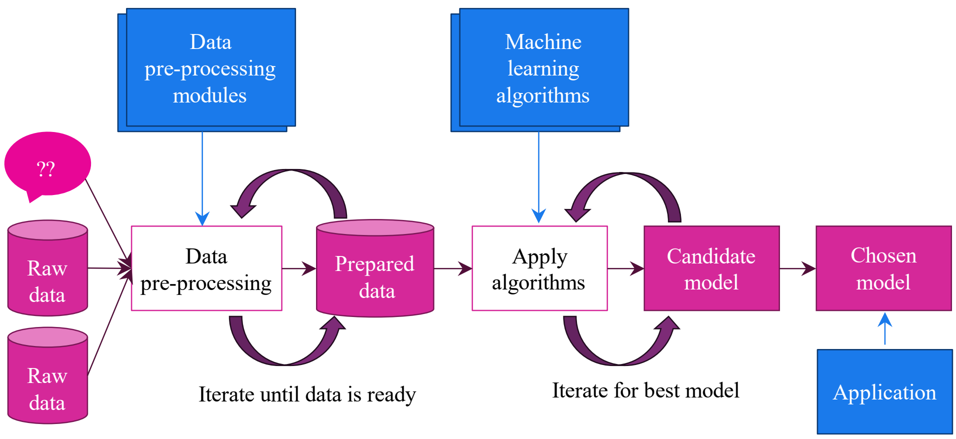

Figure 2. Model selection.

Training and Evaluation

Divide the dataset into training and testing subsets. Train the model on the training data and evaluate its performance using the testing subset. Apply cross-validation methods to ensure the model generalizes well as shown in Figure 3.

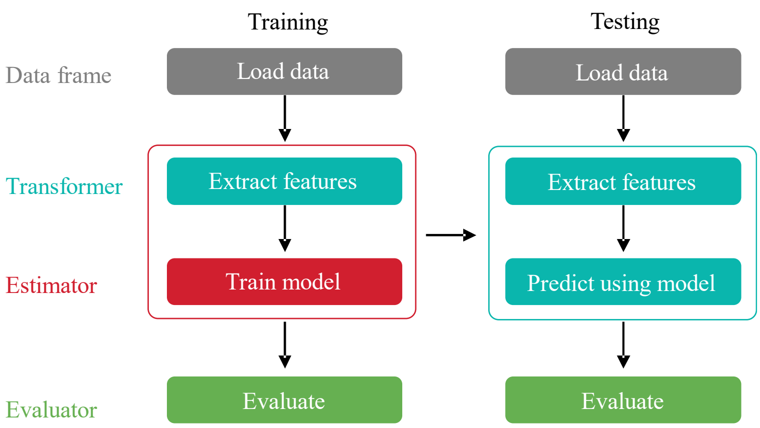

Figure 3. Data training and evaluation.

Integration of Natural Language Processing

Apply NLP techniques to analyze textual content in posts and comments. Leverage sentiment analysis and linguistic pattern recognition to enhance feature extraction as shown in Figure 4.

Validation and Testing

Validate the model on real-world data collected from social media platforms. Conduct rigorous testing to ensure the model’s robustness against diverse fake account scenarios.

Algorithmic Details

We utilize both supervised and unsupervised learning in our approach. Specifically, for supervised learning, we employ the random forest algorithm due to its capacity to handle diverse feature sets and maintain robustness against overfitting. Additionally, we integrate a deep learning architecture, utilizing a convolutional neural network (CNN), to capture intricate patterns in user behavior that might not be discernible through traditional algorithms [8].

Feature Importance

Feature selection is a critical aspect of our methodology. We prioritize features such as account creation date, post frequency, user engagement metrics, and linguistic patterns. Inspired by the literature review, we introduce novel domain specific features, including sentiment scores derived from NLP analyses and linguistic complexity metrics.

Training Parameters

During the training phase, we fine-tune parameters to optimize model accuracy and efficiency. For the random forest component, we experiment with tree depth and minimum split criteria. In the CNN, we concentrate on adjusting parameters such as learning rate, batch size, and dropout rates to improve the model's capacity for generalization [9].

Handling Imbalanced Data

Given the inherent imbalance in social media datasets, we address this issue through a combination of oversampling the minority class (fake accounts) and under sampling the majority class (genuine accounts). This ensures that our model is exposed to a representative distribution, preventing biases towards the prevalent class as shown in Figure 5.

Model Evaluation Metrics

To evaluate our model, we employ a set of metrics tailored for binary classification tasks. These metrics, such as accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic (ROC) curve (AUCROC), together offer a thorough evaluation of our model's ability to differentiate between authentic and fraudulent accounts as shown in Figure 6.

Data Filtering

Data filtering eliminates irrelevant or misleading information, reducing noise in the dataset. This ensures that the learning algorithms focus on meaningful patterns essential for accurate fake account

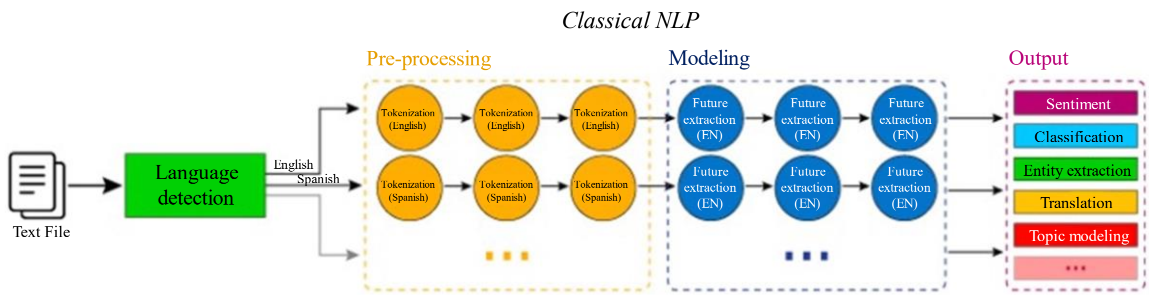

Figure 4. Classical natural language processing.

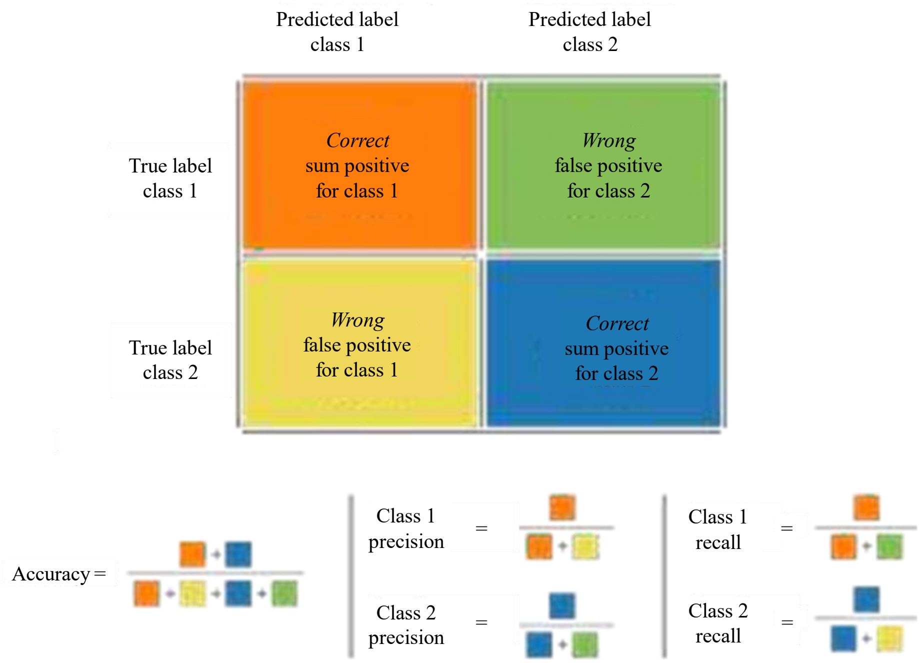

Figure 5. Handling imbalanced data.

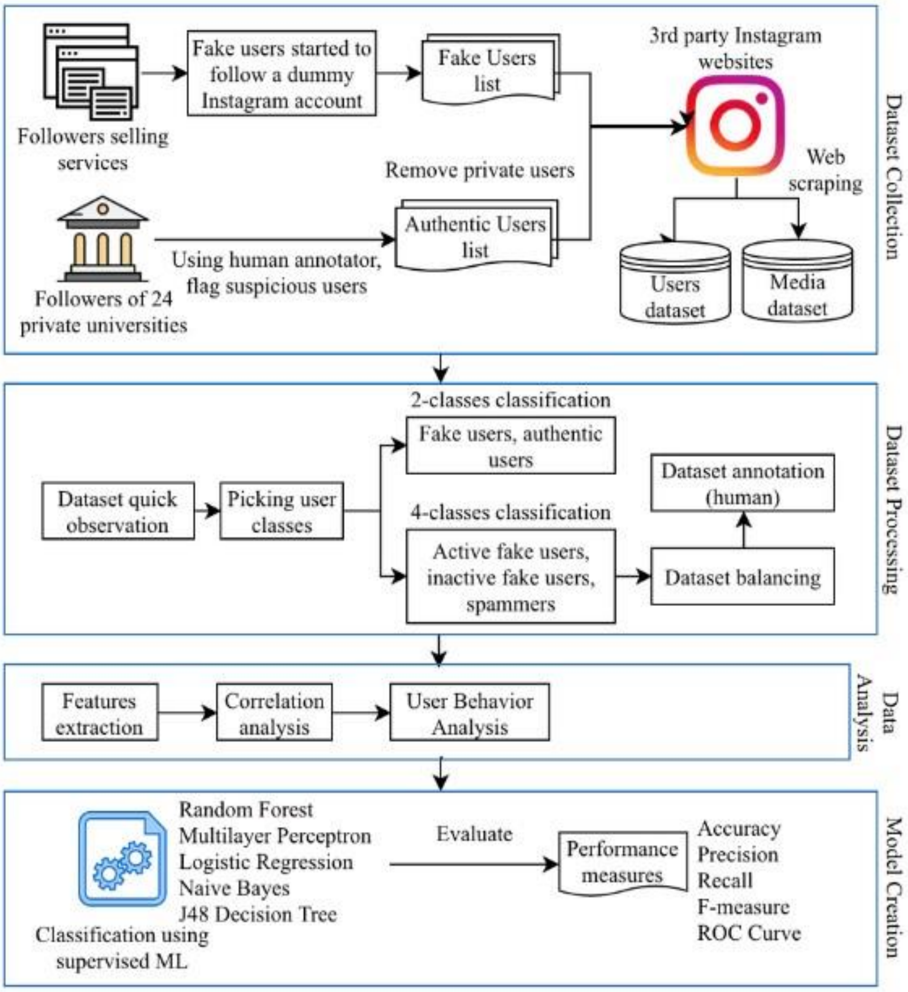

Figure 6. Model evaluation metrics.

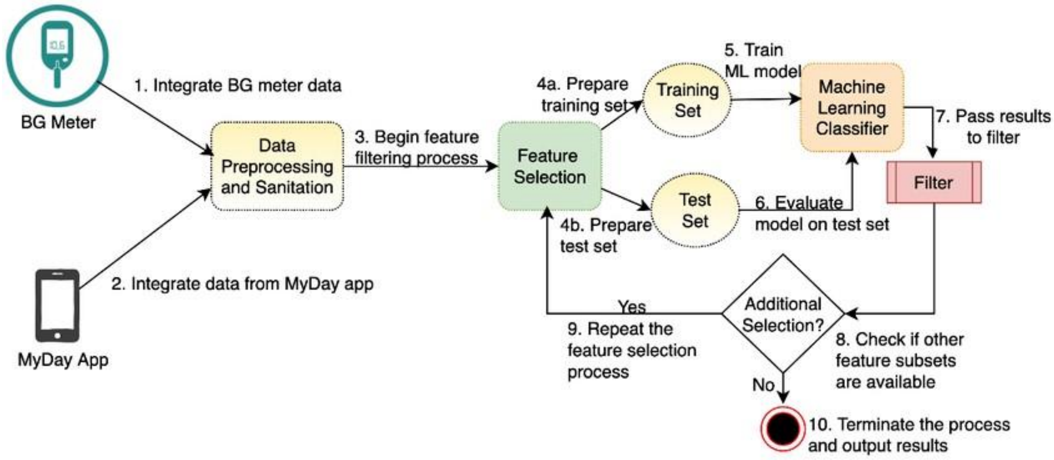

Figure 7. Data filtering.

identification. Relevant features, such as post frequency, engagement metrics, and linguistic patterns, are prioritized during data filtering. This process enhances the system's ability to discern between authentic and fraudulent profiles by focusing on crucial information. Techniques are applied to identify linguistic anomalies, such as irregular language patterns or specific keywords associated with fraudulent activities. This linguistic filtering contributes to the system's proficiency in flagging potentially malicious profiles based on linguistic cues.

Data filtering identifies unusual behavioral patterns within the dataset, including irregular engagement behaviors. This enhances the system's understanding of potential indicators of fake accounts, contributing to more accurate detection. Abnormal patterns in user engagement metrics, such as suspiciously high or low engagement levels, are filtered. This step contributes to the overall effectiveness of the fake account detection system by scrutinizing profiles with unusual engagement behaviors. The data filtering process is designed to be adaptable, considering the dynamic tactics employed by malicious actors. Regular updates to the filtering criteria ensure that the system remains effective in identifying emerging patterns associated with fake accounts as shown in Figure 7.

SYSTEM ARCHITECTURE

The system kicks off by grabbing information from Facebook using the application programming interface (API). This information forms a Facebook dataset, a mix of real and possible fake profiles. To make sense of all this, we use NLP. This involves breaking down the words and getting rid of common ones (like "and" or "the"). It helps us focus on what really matters.

Next, we go into stemming and lemmatization, a process that sorts through the data efficiently and keeps words in their basic form. To simplify things further, we use principal component analysis (PCA) to reduce the complexity of the data, making it easier to handle.

Now, it is time for learning. We are introducing various algorithms that will learn from the dataset. These algorithms become our "smart tools," adept at discerning patterns and improving their ability to differentiate between real and fake profiles.

Then comes the part where we look closely at a specific Facebook user profile. We pick out features like how often they post, how others engage with them, and the language they use. These features help us understand the profile better.

To check how well our system is doing, we use some technical measures like true positive rate (TPR), false positive rate (FPR), and AUCROC. We set a condition: if the TPR is higher than the FPR, we label it as a real profile; otherwise, we might flag it as possibly fake.

Journal of Mobile Computing, Communications & Mobile Networks
Volume 11, Issue 2
ISSN: 2349-901X

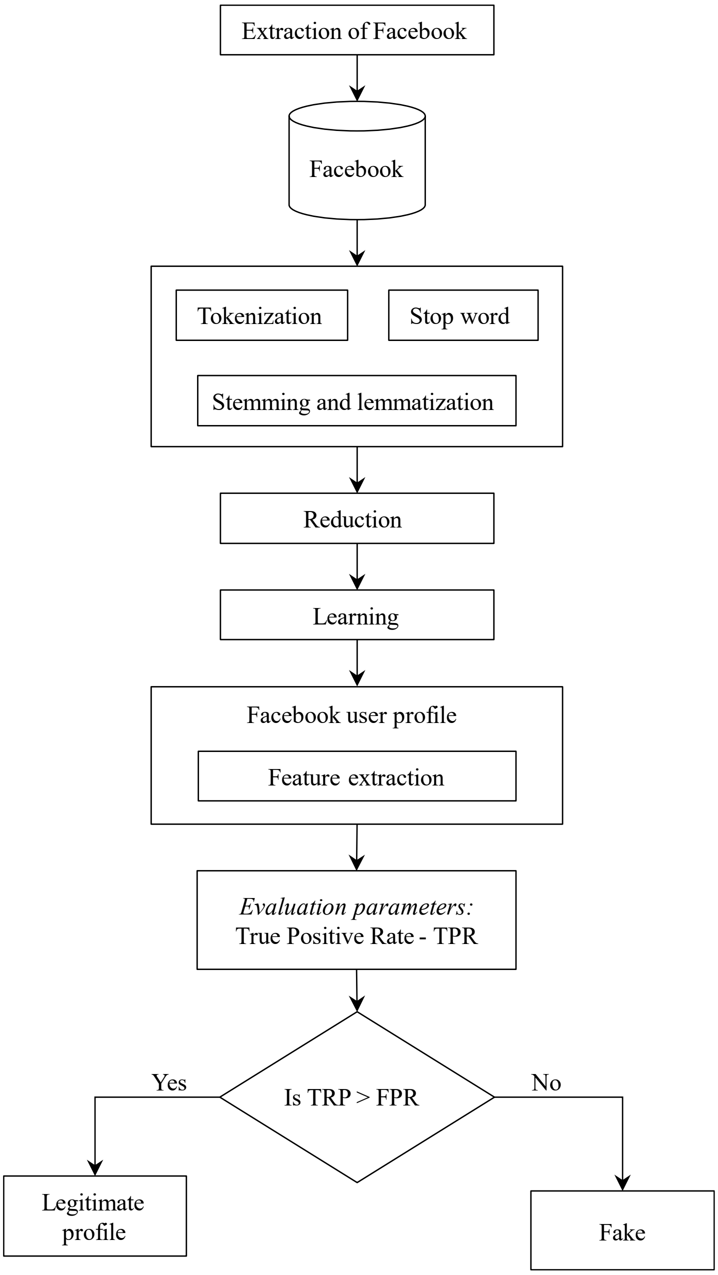

Figure 8. System architecture.

In a nutshell, this system architecture is like a learning detective. It gathers data, breaks it down, learns from it, and then applies that learning to decide if a Facebook profile is likely real or could be fake. It is a tech-savvy way to keep social media spaces safer as shown in Figure 8.

FLOWCHART

The procedural flow outlined in the flowchart delves into the intricate process of building and refining a machine learning model for fake account detection. The journey begins with the extraction of essential libraries, establishing the foundation for subsequent analyses. Following this, the importation of the dataset marks a pivotal step, incorporating diverse Facebook data crucial for training and validating the model. The subsequent phase involves data pre-processing, where the dataset undergoes meticulous cleaning, handling of missing values, and normalization to ensure uniformity in the feature space. The dataset is then strategically split into training and testing sets, laying the groundwork for the evaluation of model performance.

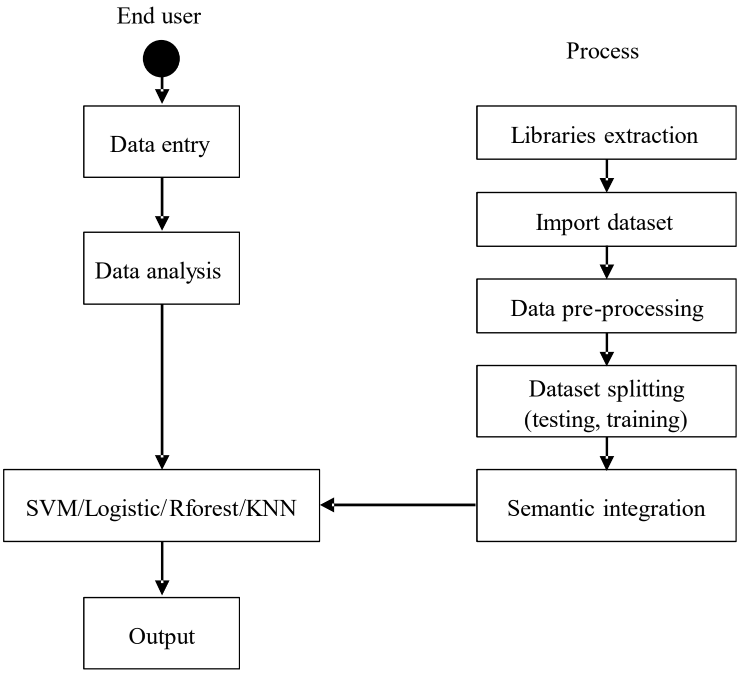

Figure 9. Flowchart of detection and analysis of fake accounts.

Semantic integration emerges as a crucial step, involving the incorporation of linguistic and contextual features derived from NLP techniques. This enhances the dataset, offering the model a deeper grasp of user behaviors. The arrow leading to support vector machines (SVM), logistic regression, random forest, and K-nearest neighbors (KNN) indicates the incorporation of various machine learning algorithms, each enhancing the model's ability to discern patterns. The subsequent stages involving end-user data entry and data analysis highlight the interactive nature of the model, allowing users to input new data for real-time analysis. The iterative loop connecting back to the ensemble of classifiers indicates a continuous learning process, wherein the model refines its predictive capacities based on user feedback and newly incorporated data. This iterative loop, involving SVM, logistic regression, random forest, and KNN, reinforces the adaptability and resilience of the model in identifying fake accounts. The journey culminates in the output stage, where the model produces insightful results, delineating the likelihood of a given Facebook profile being genuine or potentially malicious. This comprehensive flowchart encapsulates the dynamic and iterative nature of the machine learning model, illuminating each sequential step involved in the detection and analysis of fake accounts on social media as shown in Figure 9.

BASELINE METHODS

Rule-based Heuristics

Traditional rule-based heuristics leverage predefined rules to identify patterns associated with fake accounts. For instance, suspicious posting frequencies, such as overly frequent or infrequent posts, and generic profile information, including common names and lack of detailed personal data, are targeted. The simplicity and interpretability of rule-based heuristics make them a widely adopted baseline for evaluating the performance of more complex systems.

Support Vector Machines

SVM stands out as a traditional machine learning method renowned for its efficacy in tasks involving binary classification. In the context of fake account detection, SVM can learn to discriminate between genuine and potentially fraudulent profiles based on extracted features. The model aims to find the optimal hyperplane that separates the two classes, offering a robust baseline for comparison due to its well-established performance in various classification scenarios as shown in Figure 10.

Journal of Mobile Computing, Communications & Mobile Networks
Volume 11, Issue 2
ISSN: 2349-901X

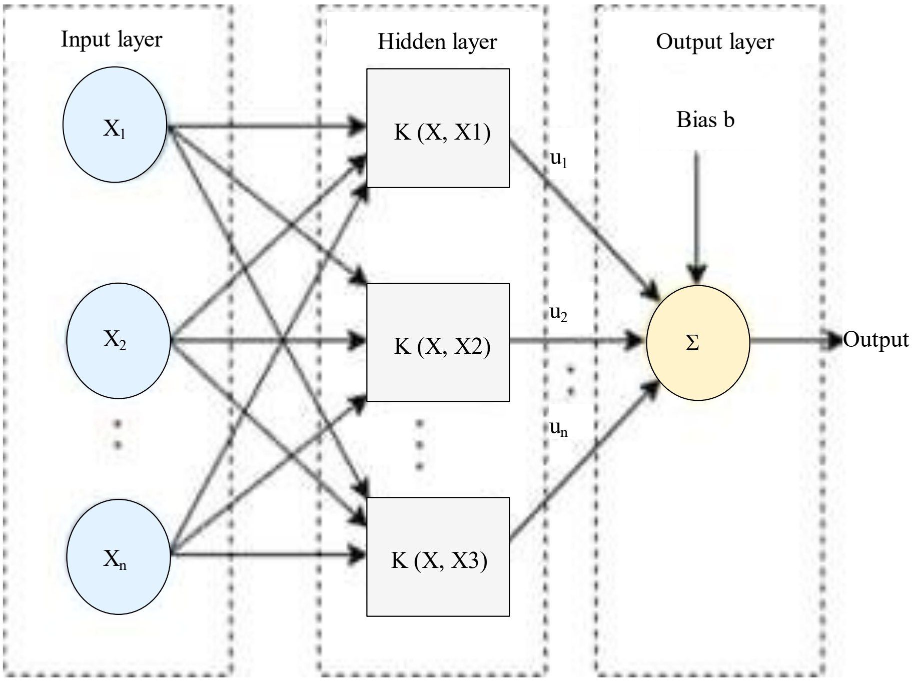

Figure 10. Support vector machine algorithm.

Logistic Regression

Logistic regression is another widely used baseline method for binary classification. It calculates the likelihood of an instance belonging to a specific class and is well-suited for situations where clarity of interpretation is crucial. In fake account detection, logistic regression can capture the relationship between features and the likelihood of an account being fake, providing a straightforward benchmark for evaluating the interpretability and accuracy of the proposed system as shown in Figure 11.

Random Forest

Random forest is an ensemble learning method that aggregates predictions from numerous decision trees, adept at capturing intricate data relationships while being resilient to overfitting. As a baseline method, Random forest offers a versatile comparison point due to its adaptability to various types of datasets. Its ability to handle a large number of features and capture intricate patterns makes it a suitable benchmark for evaluating the proposed fake account detection system as shown in Figure 12.

EVALUATION METRICS PRECISION

Precision evaluates how accurate the model's positive predictions are. In the context of fake account detection, precision indicates the proportion of profiles identified as fake that are actually genuine. Mathematically, it is expressed as shown in Figure 13.

Recall (True Positive Rate)

In the context of fake account detection, recall, also referred to as true positive rate (TPR) or sensitivity, measures how effectively the model detects all instances of fake profiles. Specifically, recall indicates the percentage of real fake profiles that the system correctly identifies. Mathematically, it is defined as shown in Figure 14.

F1 Score

The F1 score, being the harmonic mean of precision and recall, offers a balanced assessment that accounts for both false positives and false negatives, proving especially beneficial in scenarios with imbalanced class distributions. In the context of fake account detection, the F1 score balances the trade-off between precision and recall and is calculated as shown in Figure 15.

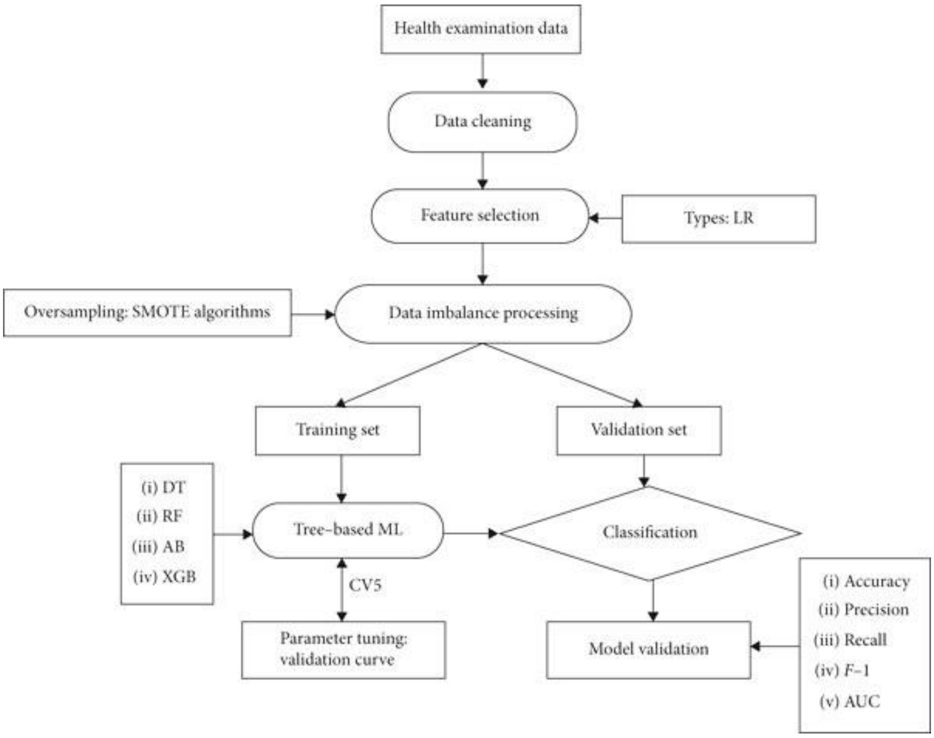

Figure 11. Logistic regression flowchart.

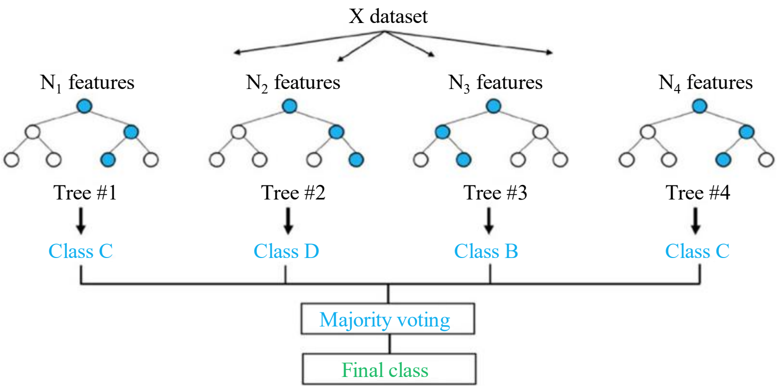

Figure 12. Random forest.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Figure 13. Precision formula.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negative}}$$

Figure 14. Recall formula.

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Figure 15. F1 score formula.

Area Under the Receiver Operating Characteristic Curve

The ROC curve shows how sensitivity (true positive rate) and 1 – specificity (false positive rate) vary with different decision thresholds, while the AUCROC summarizes the model's performance across

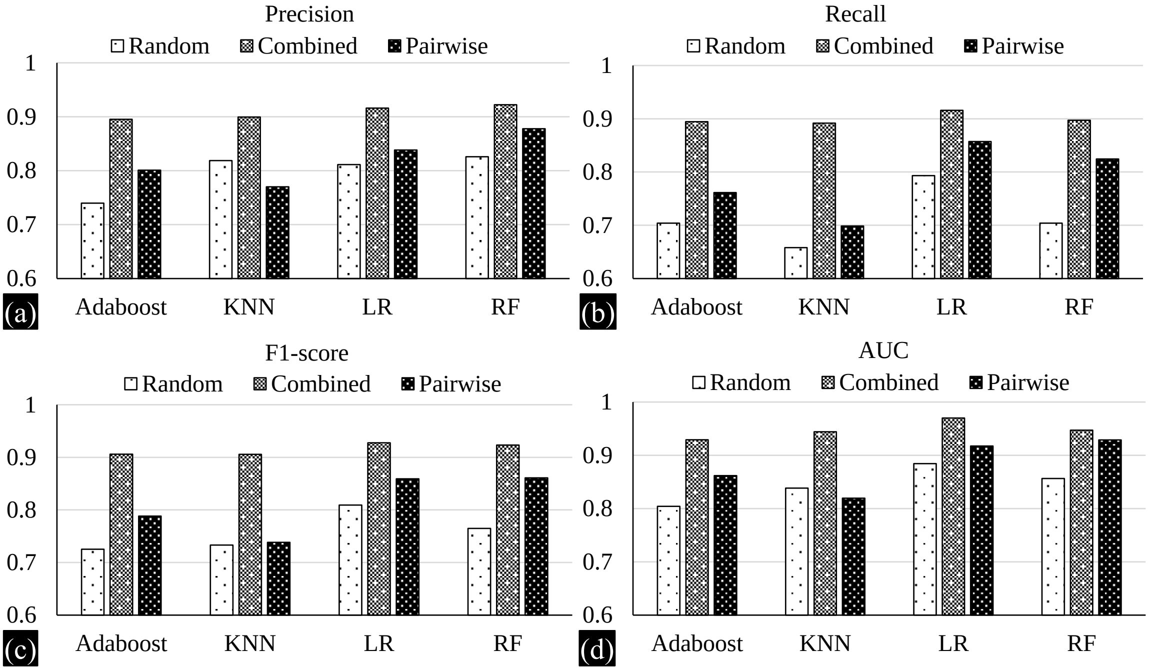

Figure 16. Evaluation metric.

these thresholds. A higher AUCROC value indicates better discriminative ability. In fake account detection, a strong AUCROC suggests effective differentiation between genuine and fake profiles as shown in Figure 16.

CONCLUSION

In conclusion, this research endeavors to address the pervasive issue of fake accounts on social media platforms through the development of an innovative detection system. Leveraging a combination of rule-based heuristics and machine learning techniques, our proposed approach aims to enhance the accuracy and efficiency of identifying fraudulent profiles. The integration of linguistic anomaly detection, behavioral pattern analysis, and user engagement metrics filtering forms a robust foundation for discerning between genuine and fake accounts. While the experimentation and conclusive results are yet to be realized, the conceptual framework established in this paper lays the groundwork for a promising contribution to the field of social media security. The fusion of traditional heuristics with advanced machine learning algorithms provides a holistic perspective, emphasizing the adaptability and comprehensiveness of the proposed fake account detection system. The journey from conceptualization to experimentation marks a crucial step in advancing the understanding and capabilities of fake account detection, underscoring the importance of continuous research and innovation in the realm of online security.

REFERENCES

1. Voitovych O, Kupershtein L, Holovenko V. Detection of fake accounts in social media. Cybersecurity Education Science Technique [Кібербезпека: освіта, наука, техніка]. 2022; 2 (18): 86–98.
2. Romanov A, Semenov A, Mazhelis O, Veijalainen J. Detection of fake profiles in social media – literature review. In: International Conference on Web Information Systems and Technologies Porto, Portugal, April 25, 2017. Volume 2, pp. 363–369. SciTePress.
3. Khaled S, El-Tazi N, Mokhtar HM. Detecting fake accounts on social media. In: 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, December 10–13, 2018. pp. 3672–3681.
4. Bhambar S, Khairnar K, Nikam Y, Shelar H, Desai YK. Detecting fake accounts on social media using neural network. Int Res J Modern Eng Technol Sci. 2022; 4 (5): 3450–3453.

5. Gayathri A, Radhika S, Jayalakshmi SL. Detecting fake accounts in media application using machine learning. Int J Adv Netw Appl. 2019; (Special Issue) : 234–237.
6. Rao KS, Gutha S, Raju DB. Detecting fake account on social media using machine learning algorithms. Int J Control Automation. 2020; 13 (1): 95–100.
7. Singh J, Khan MZ. Detection of fake profile in social media. J Emerging Technol Innov Res. 2019; 6 (6): 38–42.
8. Kadam N, Sharma SK. Social media fake profile detection using data mining technique. J Adv Inform Technol. 2022; 13 (5): 518–523.
9. Bhovar SA. Study of different methodologies to detect fake account on social media using machine learning. Int J Sci Res. 2023; 12 (2): 53–57.

