Rohini Ashok Gamane, 2025, 13:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Social Media Fake Account Identification Using Machine Learning Approach

[1]Rohini Ashok Gamane, [2]Vaibhav Dabhade

[1]ME Computer Engineering, Dept. of Computer Engineering, MET's Institute of Engineering, Savitribai Phule Pune University, Nashik, India

[2]Associate Professor, Dept. of Computer Engineering, MET's Institute of Engineering, Savitribai Phule Pune University, Nashik,India

Abstract- The widespread use of social media has resulted in a surge of fake accounts, posing serious risks to individuals, organizations, and society at large. Identifying fake accounts effectively is essential to preserving the integrity and credibility of social media platforms. This study introduces a machine learning-based approach to detect fake social media accounts. We employed five machine learning algorithms—Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Artificial Neural Networks (ANN)—to classify accounts as fake or genuine. The dataset used in this study consisted of features extracted from social media profiles, such as user behavior, profile details, and network characteristics. Experimental results revealed that the ANN algorithm outperformed the others, achieving a high accuracy of 95.6% in detecting fake accounts. The proposed approach offers significant benefits for social media platforms by enabling more efficient detection and prevention of fake accounts. Furthermore, the findings of this study can guide the development of advanced fake account detection systems, contributing to a safer and more reliable online environment.

Keywords - Fake account detection, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, Artificial Neural Networks (ANN), classification, behavior analysis.

I. INTRODUCTION

In today's digital age, where social media is deeply ingrained in daily life, the rise of fake accounts presents critical challenges to the authenticity of online interactions. These fraudulent accounts contribute to the spread of misinformation, cyberbullying, political manipulation, and identity theft, underscoring the urgent need for robust detection mechanisms. Consequently, identifying fake social media accounts has become a vital research focus, leveraging advanced machine learning techniques to enhance online security and uphold trust on social media platforms.Machine learning algorithms, particularly those designed for classification tasks, offer powerful solutions for distinguishing genuine accounts from fake ones. Prominent algorithms like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Artificial Neural Networks (ANN) have been extensively studied, each offering distinct advantages in terms of accuracy, computational efficiency, and interpretability, depending on the nature of the detection task.

Support Vector Machines (SVM) excel in handling high-dimensional data, which is common in social media datasets comprising features such as user behavior, account age, and interaction patterns. By identifying the optimal hyperplane to separate classes, SVM effectively classifies accounts as real or fake, even in complex feature spaces. Conversely, K-Nearest Neighbors (KNN) relies on simplicity, classifying new accounts based on the majority class of their nearest neighbors, making it intuitive and easy to implement.Random Forest, an ensemble learning method, combines multiple decision trees to improve accuracy and reduce overfitting. Its ability to rank feature importance provides valuable insights into the characteristics that differentiate authentic accounts from fake ones. Logistic

Regression, though simpler, performs effectively when the relationship between features and classifications is linear, making it a reliable choice for straightforward detection scenarios.Artificial Neural Networks (ANN) offer a more intricate approach, employing multiple layers and neurons to capture non-linear relationships within the data. Their adaptability and capacity to process large datasets make them particularly effective for complex detection tasks. However, balancing the complexity of ANNs with their interpretability is a challenge, particularly in privacy-sensitive applications like social media verification.The integration of these machine learning techniques forms a robust framework for identifying fake social media accounts. A comparative analysis of SVM, KNN, Random Forest, Logistic Regression, and ANN highlights their respective contributions to enhancing detection accuracy and efficiency. As the digital landscape continues to evolve, combining these methodologies represents a pivotal step toward fostering safer online interactions, restoring user trust, and encouraging responsible social media usage.

II. LITERATURE REVIEW

The paper "Virtual vs. Real Self: Gendered Presentation and Everyday Performance of Virtual Selfhood—A Case Study of Pakistan" [1] by Aksar, Firdaus, and Pasha investigates how gender shapes identity construction and performance in digital environments. Through qualitative methods, the study examines how individuals navigate online personas, balancing societal expectations with personal expression. It highlights the diverse strategies employed by various genders to shape virtual identities, emphasizing the cultural context's role in digital selfhood. This research contributes to understanding gender dynamics in social media and the complexities of online identity formation.

The literature review "Machine Learning-Based Social Media Bot Detection" [2] by Aljabri et al. provides an extensive overview of methodologies for detecting bots on social media using machine learning. The authors categorize research by supervised and unsupervised methods, feature selection techniques, and model evaluation. They emphasize the growing need for effective bot detection strategies amid rising incidents of manipulation in social interactions. This review consolidates current advancements, serving as a foundational resource for future research in social media analytics and bot detection.

In the study "Detection and Verification of Cloned Profiles in Online Social Networks Using MapReduce Based Clustering and Classification" [3], Saravanan and Venugopal propose a scalable method for identifying cloned profiles impersonating legitimate users. Utilizing a MapReduce framework, the authors employ clustering and classification algorithms to distinguish between authentic and cloned accounts. The empirical results demonstrate the method's effectiveness across large-scale datasets, addressing critical security and privacy concerns in social media. This study offers valuable insights for enhancing user safety and profile verification techniques.

Banerjee and Chua, in their paper "Understanding Online Fake Review Production Strategies" [4], analyze the tactics used by individuals and organizations to generate fake reviews on online platforms. By categorizing these strategies and examining their impact on consumer trust and brand reputation, the study provides a comprehensive understanding of deceptive practices in e-commerce. The authors emphasize the need for countermeasures and policies to curb fake reviews, promoting transparency and accountability in online business.

The research "KC-GCN: A Semi-Supervised Detection Model Against Various Group Shilling Attacks in Recommender Systems" [5] by Cai et al. introduces an advanced model to detect group shilling attacks in recommendation systems. Using a graph convolutional network (GCN) framework, the model employs semi-supervised learning to enhance detection accuracy and adaptability. The evaluation across multiple datasets demonstrates superior performance compared to traditional methods. This study addresses the critical need for safeguarding recommender systems against malicious

manipulation, advancing cybersecurity in recommendation frameworks.

The paper "Fake Profile Identification in Social Networks Using Machine Learning and NLP" [6] by Latha and Sumitra applies machine learning and natural language processing (NLP) to detect fake profiles on social networks. Their framework extracts significant features from user content and account characteristics to distinguish genuine accounts from impostors. Experimental results show high accuracy, highlighting the potential of NLP in enhancing profile verification and online safety.

Sudhakar and Bhuvana Chendrica Gogineni, in "Fake Profile Identification Using Machine Learning" [7], explore the application of machine learning algorithms for detecting fake profiles on social media. Their structured approach to feature engineering and evaluation identifies key attributes that differentiate fraudulent accounts. The findings demonstrate high accuracy, offering a practical framework for implementation in social media systems and contributing to user trust and security.

The study "Identification of Fake Accounts in Social Media Using Machine Learning" [8] by Kotra and Kothapelly evaluates the effectiveness of machine learning algorithms in detecting fake social media accounts. Through detailed experiments on real-world datasets, the authors highlight the capability of these algorithms to distinguish fake accounts from legitimate ones. Their work underscores the potential of machine learning in improving online safety and provides valuable insights for future research in account verification.

The paper "Collaborative Filtering Recommendation Using Fusing Criteria Against Shilling Attacks" [9] by Li et al. proposes a collaborative filtering approach that integrates multiple criteria to improve the resilience of recommendation systems against shilling attacks. Their experiments demonstrate enhanced recommendation accuracy and reliability compared to traditional systems. This research contributes to developing robust mechanisms that protect recommendation systems from manipulation, ensuring a trustworthy user experience in digital marketplaces.

In the study "Fake Profile Identification in Social Networks Using Machine Learning and NLP" [10], Sasikala et al. present a hybrid approach combining machine learning and natural language processing for detecting fraudulent profiles. The methodology involves feature extraction from user data and the application of classification algorithms to improve detection accuracy. Experimental results validate the effectiveness of this approach, offering advanced solutions for enhancing user authenticity verification in social networks.

III. METHODOLOGY

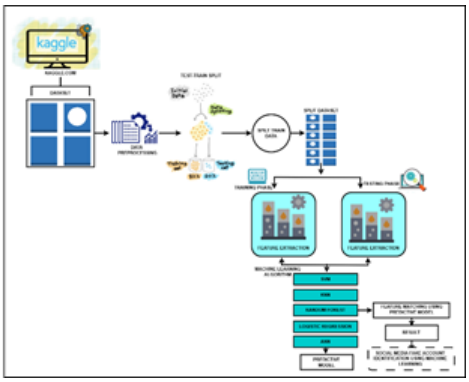

Fig.1 Proposed Methodology Architecture

The process of identifying fake accounts on social media using machine learning involves several essential steps, starting with data collection and preprocessing. A dataset is first compiled, containing a mix of genuine and suspected fake accounts. This dataset includes features such as user profile attributes (e.g., account age, follower count, and post frequency), behavioral patterns (e.g., activity levels and engagement metrics), and content analysis (e.g., language usage and sentiment in posts). During preprocessing, the data is cleaned by removing duplicates, handling missing values, and normalizing numerical features to ensure consistency. Feature selection techniques, such as correlation analysis or recursive feature elimination, are then applied to identify the most relevant attributes for distinguishing between real and fake accounts.machine learning algorithms such as

Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Artificial Neural Networks (ANN) are employed for classification. The preprocessed dataset is used to train these models, with cross-validation techniques applied to evaluate performance and minimize overfitting. Hyperparameter tuning is conducted to optimize the models for better accuracy.The models are then assessed using performance metrics, including accuracy, precision, recall, and F1-score, to determine their effectiveness in detecting fake accounts. The results are analyzed to identify the algorithm that performs best under the given conditions. This methodology provides valuable insights into leveraging machine learning to enhance social media safety and maintain platform integrity.

Dataset

The dataset for social media fake account detection, obtained from Kaggle, comprises user data designed to differentiate between fake and legitimate accounts. It includes a diverse range of features such as account age, post count, follower and following numbers, as well as engagement metrics like likes and comments. These attributes are instrumental in identifying patterns commonly associated with fake accounts, such as disproportionately high follower-to-following ratios or minimal engagement levels. The dataset is divided into training and testing subsets to facilitate model evaluation, enabling the application of machine learning algorithms to accurately classify accounts based on these characteristics.

Support Vector Machines(SVM):

Support Vector Machines (SVM) are a reliable classification method used to detect fake social media accounts by identifying the optimal hyperplane that maximizes the margin of separation between classes (fake vs. genuine accounts). SVM performs well in high-dimensional spaces, making it ideal for analyzing complex features derived from user profiles and behaviors. By leveraging kernel functions, SVM effectively handles non-linear relationships, ensuring accurate classification even in cases of noisy or intricate data.

K-Nearest Neighbors (KNN):

K-Nearest Neighbors (KNN) is a simple yet effective algorithm that identifies fake accounts based on the proximity of data points in the feature space. An account is classified as fake or genuine by examining the majority class among its 'k' nearest neighbors, using labeled examples as a reference. KNN's straightforward approach makes it a valuable tool for initial analyses of account authenticity, though it may face challenges with scalability and sensitivity to noise in large datasets.

Random Forest:

Random Forest is a robust ensemble learning algorithm used for detecting fake accounts by constructing multiple decision trees and aggregating their outputs. It effectively processes diverse features such as user activity, social connections, and profile characteristics, making it highly suitable for social media applications. Renowned for its resistance to overfitting, Random Forest delivers reliable predictions even in high-dimensional or complex datasets. Feature selection and hyperparameter tuning can further enhance its performance, solidifying its role in various machine learning tasks.

Logistic Regression:

Logistic Regression is a statistical method used to classify social media accounts as fake or genuine by estimating the probability of each class based on features like account creation date, posting frequency, and interaction patterns. While it is a linear classifier, Logistic Regression is computationally efficient and interpretable, making it a practical choice for scenarios requiring transparency. However, it may struggle with capturing complex, non-linear patterns in data, limiting its effectiveness in highly intricate classification tasks.

Artificial Neural Networks (ANN):

Artificial Neural Networks (ANN) are advanced models that excel in detecting fraudulent social media accounts by capturing complex, non-linear relationships within large datasets. Consisting of interconnected layers of neurons, ANNs analyze features such as user behavior, profile details, and

social interactions to classify accounts. They achieve high accuracy in identifying fake accounts and are particularly effective in large-scale environments. While their adaptability to complex patterns is a strength, balancing complexity with computational efficiency remains a challenge in some applications.

Algorithm

Step 1: Collect a dataset containing features of real and fake social media accounts. Utilize publicly available datasets from platforms like Kaggle or create custom datasets tailored to your requirements.

Step 2: Clean and preprocess the dataset by addressing missing values, encoding categorical data, and normalizing numerical features to ensure uniformity.

Step 3: Identify and select relevant features such as the presence of a profile picture, friend count, posting frequency, account age, bio content, and activity patterns.

Step 4: Split the dataset into training and testing subsets (e.g., 80% training and 20% testing) to allow for model training and evaluation.

Step 5: Select suitable machine learning algorithms for classification:

Support Vector Machine (SVM): Identifies the optimal hyperplane to classify data points into fake or genuine accounts.

K-Nearest Neighbors (KNN): Classifies accounts based on the majority label of the 'k' nearest neighbors in the feature space.

Random Forest: Constructs multiple decision trees and aggregates their predictions for more accurate classification.

Logistic Regression: Applies a logistic function to model binary outcomes (fake or real accounts).

Artificial Neural Network (ANN): Employs layers of interconnected neurons to capture complex patterns and enable advanced decision-making.

Step 6: Train each model using the training dataset, fine-tuning parameters to achieve optimal performance.

Step 7: Evaluate the models using the testing dataset and performance metrics such as accuracy, precision, recall, and F1-score to measure their effectiveness.

Step 8: Optimize key model parameters, such as the kernel type for SVM, the number of neighbors for KNN, or the number of trees for Random Forest, to enhance performance further.

Step 9: Compare the performance of all models to identify the most accurate and reliable one for detecting fake accounts.

Results and Discussion

Front End

Fig: Login Page

Fig: Data Input 1

Fig: Data Input 2

Fig: Prediction

With Feature Selection

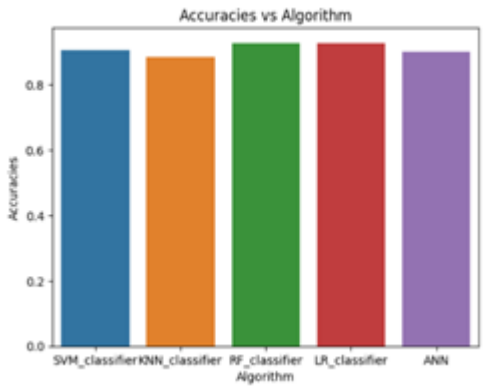

Fig: Accuracies Comparison

The bar chart displays the accuracy levels of various machine learning algorithms used for detecting fake social media accounts. The algorithms compared are SVM, KNN, RF, LR, and ANN. The y-axis represents accuracy values, ranging from 0 to 1, while the x-axis lists the algorithms. According to the chart, the RF classifier achieves the highest accuracy, followed by the LR classifier, SVM classifier, ANN, and finally the KNN classifier. This indicates that RF and LR are the most effective algorithms for this task, whereas KNN is the least effective.

	precision	recall	f1-score	support
0	0.94	0.94	0.94	70
1	0.94	0.94	0.94	70
accuracy			0.94	140
macro avg	0.94	0.94	0.94	140
weighted avg	0.94	0.94	0.94	140

Fig: Classification Report of Hyper Parameter Tuned SVM

The Classification Report outlines the performance metrics of a Hyperparameter-Tuned SVM model for detecting fake social media accounts. It includes precision, recall, F1-score, and support for each class (0 and 1), along with overall accuracy, macro average, and weighted average. The model demonstrates high precision, recall, and F1-scores for both classes, achieving an overall accuracy of 0.94. These results highlight the model's effectiveness in accurately identifying real and fake accounts, with minimal false positives and false negatives.

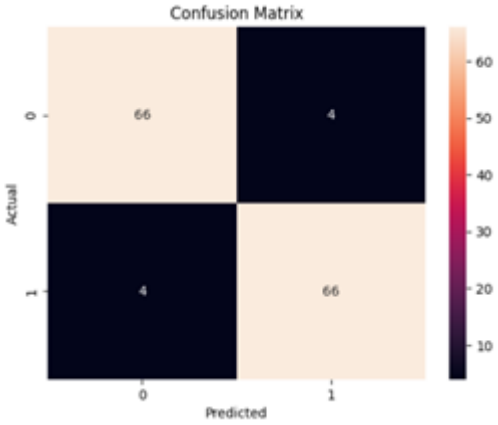

Fig: Confusion Matrix of Hyper Parameter Tuned SVM

The confusion matrix illustrates the performance of the Hyperparameter-Tuned SVM model for detecting fake social media accounts. It displays the counts of true positives (66), true negatives (66), false positives (4), and false negatives (4). The diagonal elements represent correct predictions, while the off-diagonal elements signify errors. The high values along the diagonal and the low error counts off the diagonal indicate that the model achieves high accuracy, effectively distinguishing between real and fake accounts.

Without Feature Selection

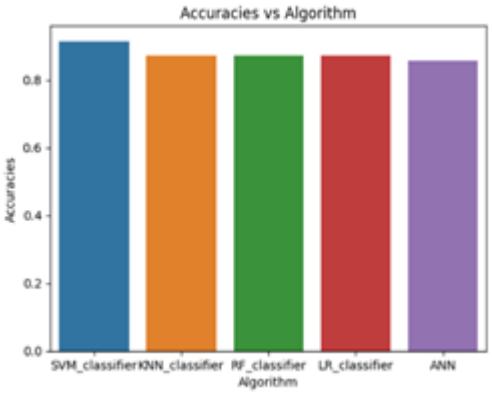

Fig: Accuracies Comparision

The bar chart compares the accuracy of various machine learning algorithms used for fake account detection on social media. The algorithms evaluated include SVM, KNN, RF, LR, and ANN. The y-axis represents accuracy, ranging from 0 to 1, while the x-axis lists the algorithms. The chart shows that the RF classifier achieved the highest accuracy, followed

by LR, SVM, ANN, and KNN classifiers. This indicates that RF and LR are the most effective algorithms for this task without feature selection, while KNN performs the least effectively.

IV. CONCLUSION

In conclusion, the use of various machine learning algorithms—such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Artificial Neural Networks (ANN)—for identifying fake social media accounts reveals the unique advantages and limitations of each method. SVM and Random Forest excel in managing complex data and high dimensionality, while KNN may face computational challenges and Logistic Regression's linear assumptions could limit its performance in more nuanced situations. Meanwhile, ANNs offer powerful pattern recognition but require large datasets and face criticism for their lack of interpretability. Ultimately, choosing the most suitable approach depends on factors like data quality, the need for transparency, and the specific problem at hand, potentially leading to the use of hybrid or ensemble methods to leverage the strengths of multiple algorithms and improve detection accuracy and reliability.

REFERENCES

1. Aksar, A. Firdaus, and S. A. Pasha, ''Virtual vs. real self: Gendered presentation and everyday performance of virtual selfhood—A case study of Pakistan,'' J. Commun. Inquiry, vol. 47, no. 1, pp. 84–114, Jan. 2023, doi: 10.1177/01968599221089236.
2. [M. Aljabri, R. Zagrouba, A. Shaahid, F. Alnasser, A. Saleh, and D. M. Alomari, ''Machine learning-based social media bot detection: A comprehensive literature review,'' Social Netw. Anal. Mining, vol. 13, no. 1, p. 20, Jan. 2023, doi: 10.1007/s13278-022-01020-5.
3. A. Saravanan and V. Venugopal, "Detection and Verification of Cloned Profiles in Online Social Networks Using MapReduce-Based Clustering and Classification," International Journal of Intelligent Systems and Applications in Engineering, vol. 11, no. 1, pp. 195–207, January 2023.
4. S. Banerjee and A. Y. K. Chua, ''Understanding online fake review production strategies,'' J. Bus. Res., vol. 156, Feb. 2023, Art. no. 113534, doi:10.1016/j.jbusres.2022.113534.
5. H. Cai, J. Ren, J. Zhao, S. Yuan, and J. Meng, ''KC-GCN: A semisupervised detection model against various group shilling attacks in recommender systems,'' Wireless Commun. Mobile Comput., vol. 2023, pp. 1–15, Feb. 2023, doi: 10.1155/2023/2854874
6. Latha P, Sumitra V,"Fake Profile Identification in Social Network using Machine Learning and NLP", 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)|978-1-6654-7995-0/22/$31.00©2022IEEE| DOI: 10.1109/IC3IOT53935.2022.9767958, 978-1-6654-7995-0/22/$31.00 ©2022 IEEE
T.Sudhakar,Bhuvana Chendrica Gogineni," FAKE PROFILE IDENTIFICATION USING MACHINE LEARNING", 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE),979-8-3503-1156-3/22/\$31.00 c2022 IEEE
7. Kotra Shreya, Amith Kothapelly," Identification of Fake accounts in social media using machine learning", 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), 978-1-6654-5635-7/22/\$31.00 ©2022 IEEE
8. L. Li, Z. Wang, C. Li, L. Chen, and Y. Wang, ''Collaborative filtering recommendation using fusing criteria against shilling attacks,'' Connection Sci., vol. 34, p. 1, 1678-1696, 2022, doi: 10.1080/09540091.2022.2078280.
9. V. Sasikala, J. Arunarasi, A. R. Rajini, and N. Nithiya, ''Fake profile identification in social network using machine learning and NLP,'' in Proc. Int. Conf. Commun., Comput., Internet Things (IC3IoT), Chennai, India, 2022, pp. 1–4, doi: 10.1109/IC3IOT53935.2022.9767958.
10. H. M. F. Shehzad, A. Yasin, Z. K. Ansari, M. A. Khan, and M. J. Awan, ''Fake profile recognition using big data analytics in social media platforms,'' Int. J. Comput. Appl. Technol., vol. 68,

no. 3, p. 215, 2022, doi: 10.1504/IJCAT.2022.10049746.

11. K. Kaushik, A. Bhardwaj, M. Kumar, S. K. Gupta, and A. Gupta, ''A novelmachine learning-based framework for detecting fake Instagram profiles,''Concurrency Comput., Pract. Exper., vol. 34, no. 28, p. e7349, Dec. 2022,doi: 10.1002/cpe.7349.
12. M. Vyawahare and S. Govilkar, ''Fake profile recognition using profanity and gender identification on online social networks,'' Social Netw. Anal. Mining, vol. 12, no. 1, Dec. 2022, doi: 10.1007/s13278-022-00997-3.
13. B. P. Kavin, S. Karki, S. Hemalatha, D. Singh, R. Vijayalakshmi, M. Thangamani, S. L. A. Haleem, D. Jose, V. Tirth, P. R. Kshirsagar, and A. G. Adigo, ''Machine learning-based secure data acquisition for fake accounts detection in future mobile communication networks,'' Wireless Commun. Mobile Comput., vol. 2022, pp. 1–10, Jan. 2022, doi: 10.1155/2022/6356152.
14. [15] K. Shahzad, S. A. Khan, S. Ahmad, and A. Iqbal, ''A scoping review of the relationship of big data analytics with context-based fake news detection on digital media in data age,'' Sustainability, vol. 14, no. 21, p. 14365, Nov. 2022, doi: 10.3390/su142114365.