Machine Learning Techniques for Fake Account Detection in Social Networks

[1] Abdul Muthalib Mohammed, [2] Dr.K. Santhi Sree

[1]MCA Student, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, Telangana, India

[2]Professor, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, Telangana, India

abdulmuthallib18@gmail.com , drksanthisree@gmail.com

DOI: https://zenodo.org/records/16410847

Abstract: **In the present generation, the social life of everyone has become associated with the online social networks. Adding new friends and keeping in contact with them and their updates has become easier. The online social networks have impact on the science, education, grassroots organizing, employment, business, etc. Fake profiles play an important role in advanced persisted threats and are also involved in other malicious activities. Social networks fake profile creation is considered to cause more harm than any other form of cybercrime**

Keywords—Fake Profile, KNN, Logistic Regression, Naïve Bayes, Random Forest, SVM

I. INTRODUCTION

Twitter is a famous online social networking platform that provides a medium for people to communicate, share information, and connect with each other. However, with the increasing use of Twitter, there has been an alarming rise in the number of fake accounts, which poses a serious threat to the security and privacy of genuine users. These fake accounts can be used to disseminate spam, malware, and propaganda, and can also be employed in cyber-attacks.

In this project, we aim to investigate the problem of fake account detection on Twitter using machine learning techniques. Specifically, we will explore the use of machine learning algorithms including Logistic Regression, Linear Support Vector Classifier (Linear SVC), Naive Bayes, K-Nearest Neighbors (KNN), and Random Forest to build a classifier that can detect fake accounts with high accuracy.

II. LITERATURE SURVEY

Several studies have focused on detecting fake accounts using various machine learning techniques. Some of the key contributors include:

[1] Sarah Khaled et al. used the MIB dataset from Twitter and introduced an SVM-NN approach with moderate accuracy.

[2] Ala M. Al-Zoubi et al. used the dataset from Twitter for spam profile detection. Firstly, they analysed the public information available in twitter. Based on this analysis they have identified ten features for spam profile detection. The following are some factors to take into account when analysing tweets: suspicious words, the default image, the text-to-links ratio, the following-to-followers ratio, repeated words, the comments ratio, the tweet time pattern, the different descriptions from tweets, the different following interest from tweets, and the number of tweets per day.

[3] Adikari and Dutta (2014) explored fake profile detection on LinkedIn and achieved 84% accuracy using constrained profile information.

III. METHODOLOGY

A. Proposed Work:

The rapid expansion of social media platforms has led to a rise in fake profiles, which contribute to misinformation, cyber threats, and fraudulent activities. Key aspects of fake profiles include abnormal engagement patterns, high friend request rejections, and suspicious follower ratios. This framework leverages machine learning models to detect fake profiles efficiently. The dataset comprises various user attributes such as abuse reports, friend request rejections, number of followers, and engagement patterns. To enhance detection accuracy, the system applies preprocessing techniques like feature scaling, missing value imputation, and dimensionality reduction. Key classification models, including Logistic Regression, K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector Machine (SVM), and Random Forest, are employed

B. System Architecture:

The system architecture consists of multiple components:

- Data Ingestion: Module for collection and storage of user data from Kaggle Website

- Feature Extraction: Extraction and selection of relevant features for effective model training.
- Model Training: Implementation and optimization of multiple machine learning algorithms.

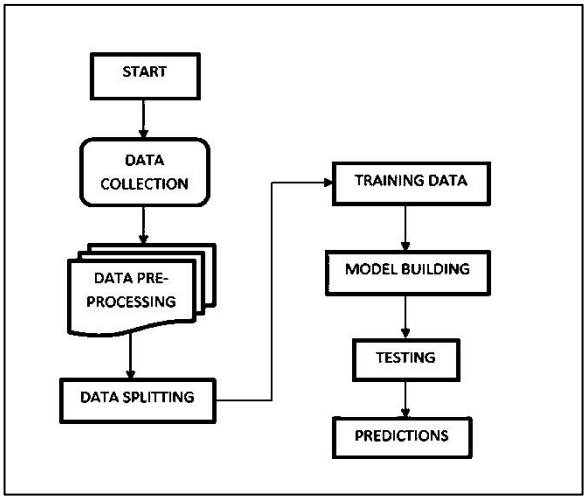

Fig 1 Proposed System Architecture

Fig 1 Illustrates data collection, preprocessing (handling null values, duplicates), splitting into train-test sets, and making predictions using machine learning models.

IV. ALGORITHMS

For detecting fake profiles on twitter account, we used different machine learning models such as KNN, Logistic Regression, Naive Bayes, Random Forest and SVM.

A. K-Nearest Neighbors:

A data point is classified using the k nearest neighbour algorithm based on how its neighbours are grouped. The number of nearest data points to include in the majority voting process is indicated by the k value. The value of k is chosen as 3.

B. Logistic Regression:

A classification algorithm that estimates the probability of a given instance belonging to a particular class. It applies the logistic (sigmoid) function to ensure the output remains between 0 and 1. The solver used in this study is LBFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno), an optimization algorithm that efficiently handles large datasets while maintaining computational efficiency. Unlike traditional BFGS, LBFGS discards older gradient values to reduce memory consumption, making it well-suited for high-dimensional data.

C. Naive Bayes:

Naive Bayes classifiers work on the basis of Bayes Theorem. Given the likelihood of an earlier occurrence, the Bayes Theorem determines the likelihood that an event will occur. Gaussian Naive Bayes classifier was employed here. Continuous values connected to each feature in Gaussian Naive Bayes are distributed in a Gaussian manner.

D. Random Forest:

The results from several decision trees applied to various subsets of the input data set are averaged by the Random Forest classifier in order to improve the accuracy of the input data set. The random forest employs predictions from all the trees instead of just one in order to anticipate the result depending on which forecasts received the most votes.

E. Support vector machine:

The SVM approach aims to identify the optimal line or decision boundary that effectively separates classes within an n-dimensional space, allowing it to categorize new data points in the future. SVM utilizes support vectors, which are extreme data points and vectors to determine the hyperplane that effectively separates classes. The training set of data is processed using a kernel function to convert a non-linear decision surface into a linear equation in a high dimensional space. The kernel used by the SVM for training is linear kernel.

V. EXPERIMENTAL RESULTS

A. KNN

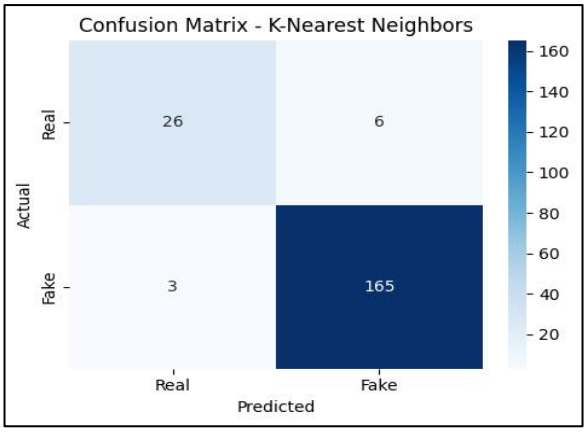

Fig 2 Confusion Matrix for KNN

Confusion Matrix showing a few misclassifications, especially in Class 0.

K-Nearest Neighbors Classification Report:

	precision	recall	f1-score	support
0	0.90	0.81	0.85	32
1	0.96	0.98	0.97	168
accuracy			0.95	200
macro avg	0.93	0.90	0.91	200
weighted avg	0.95	0.95	0.95	200

Accuracy : 0.955

Fig 3 Classification Report for KNN

Precision: 90% , 96% ; Recall: 81% , 98% ; Weighted Accuracy: **95%**, indicating strong overall performance.

B. Logistic Regression

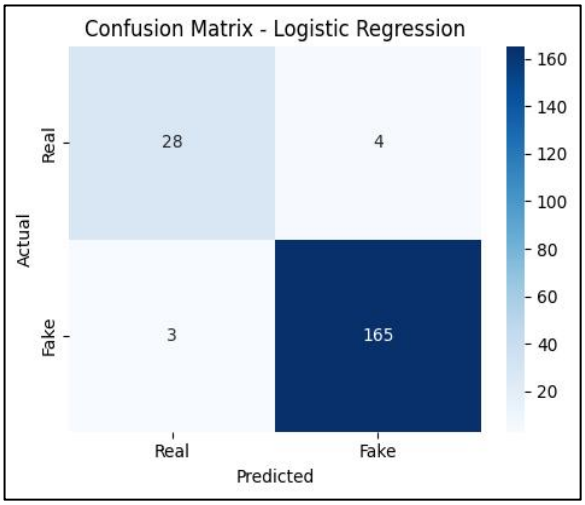

Fig 4 Confusion Matrix for Logistic Regression

Confusion Matrix with minimal misclassifications, proving its reliability.

Logistic Regression Classification Report:

	precision	recall	f1-score	support
0	0.90	0.88	0.89	32
1	0.98	0.98	0.98	168
accuracy			0.96	200
macro avg	0.94	0.93	0.93	200
weighted avg	0.96	0.96	0.96	200

Accuracy : 0.965

Fig 5 Classification Report of Logistic Regression

Precision: 90% , 98% ; Recall: 88% , 98% ; Weighted Accuracy: **96.5%**, showing high prediction accuracy.

C. Naive Bayes

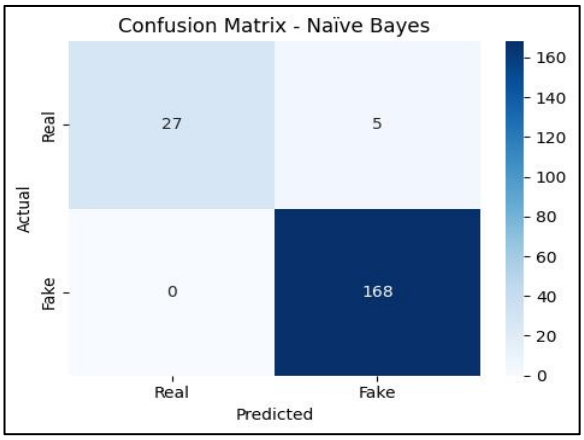

Fig 6 Confusion Matrix for Naïve Bayes

Confusion Matrix with Class 0 having slight misclassification.

Naïve Bayes Classification Report:

	precision	recall	f1-score	support
0	1.00	0.84	0.92	32
1	0.97	1.00	0.99	168
accuracy			0.97	200
macro avg	0.99	0.92	0.95	200
weighted avg	0.98	0.97	0.97	200

Accuracy : 0.975

Fig 7 Classification Report for Naïve Bayes

Precision: 100%, 97%; Recall: 84% , 100% ; Weighted Accuracy: **98%**, highlighting its effectiveness.

D. Random Forest

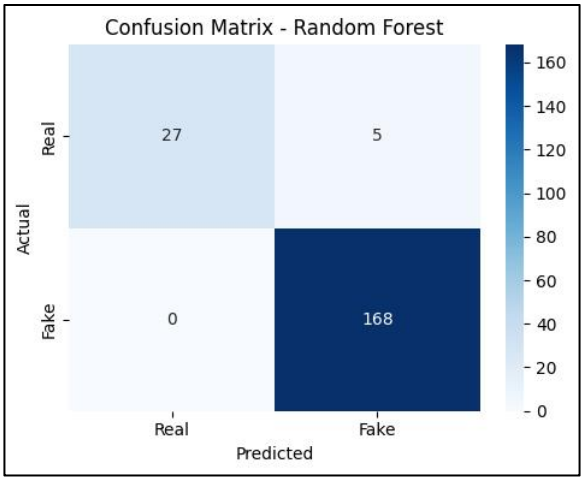

Fig 8 Confusion Matrix for Random Forest

Confusion Matrix with Class 0 experiencing some misclassifications.

Random Forest Classification Report:				
	precision	recall	f1-score	support
0	1.00	0.84	0.92	32
1	0.97	1.00	0.99	168
accuracy			0.97	200
macro avg	0.99	0.92	0.95	200
weighted avg	0.98	0.97	0.97	200
Accuracy : 0.975				

Fig 9 Classification Report of Random Forest

Precision: 100%, 97% ; Recall: 84%, 100% ; Weighted Accuracy: **98%**, indicating a well-balanced model.

E. SVM

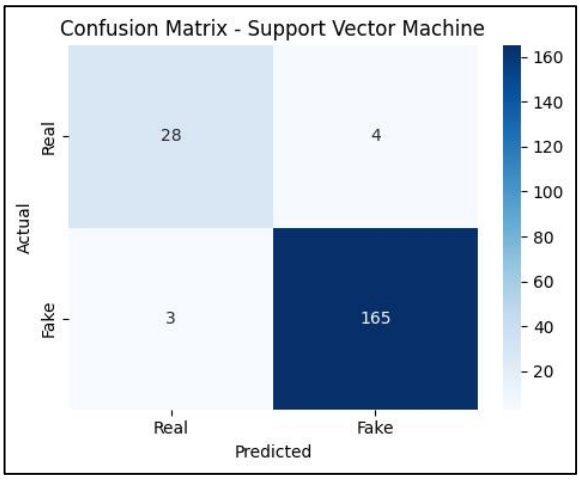

Fig 10 Confusion Matrix using SVM

Confusion Matrix showing performance similar to Logistic Regression.

Support Vector Machine Classification Report:				
	precision	recall	f1-score	support
0	0.90	0.88	0.89	32
1	0.98	0.98	0.98	168
accuracy			0.96	200
macro avg	0.94	0.93	0.93	200
weighted avg	0.96	0.96	0.96	200
Accuracy : 0.965				

Fig 11 Classification Report of SVM

Precision: 90% , 98% ; Recall: 88% , 98% ; Weighted Accuracy: **96.5%**, demonstrating strong classification capability.

VI. CONCLUSION

This study explores fake profile detection in online social networks using machine learning models, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Naïve Bayes. The models were evaluated based on accuracy, precision, recall, and F1-score.

- Random Forest achieved an accuracy of 97.5%, making it the most effective model for fake profile detection.
- Accuracy of Navie bayes is also proved to be 97% but since it is dependent mostly on one feature we conclude that RF is most effective
- Support Vector Machine (SVM) and Logistic Regression both produced high accuracy of 96.5%, proving to be reliable classifiers.
- K-Nearest Neighbors (KNN) showed an accuracy of 94%, making it a competitive method for fake profile detection.

Among these models, Naïve Bayes and Random Forest stood out as the top-performing classifiers. Future research can explore deep learning techniques, behavioural analysis, and real-time detection methods to further enhance accuracy and adaptability in identifying fake profiles

VII. REFERENCES

[1] Sarah Khaled, Neamat El-Tazi, and Hoda M.O. Mokhtar, "Detecting Fake Accounts on Social

Media," *2018 IEEE International Conference on Big Data (Big Data)*, 2018.

[2] Ala' M. Al-Zoubi, Ja'far Alqatawna, and Hossam Faris, "Spam Profile Detection in Social Networks Based on Public Features," *International Conference on Information and Communication Systems (ICICS)*, 2017.

[3] Adikari, Shalinda, and Kaushik Dutta, "Identifying Fake Profiles in LinkedIn," *PACIS 2014 Proceedings*, 2014.

[4] Qahtani and Najjar, "Machine learning approach to detect fake accounts on Twitter," *International Conference on Artificial Intelligence and Computer Vision*, 2018.

[5] R. Malik et al., "Detecting fake Twitter accounts using machine learning algorithms," *Information Processing & Management*, vol. 57, no. 1, pp. 102191, 2020.

[6] Gaitonde and Doye, "Machine learning algorithms in detecting fake profiles," *International Conference on Intelligent Computing and Control Systems*, Madurai, 2019.

[7] S. Kumar and R. Kumar, "On Fake Twitter account detection using machine learning techniques: A systematic review," *International Journal of Information Technology*, vol. 13, no. 1, pp. 15-27, 2021.

[8] Patel and S. Patel, "A survey on fake Twitter account detection using machine learning techniques," *International Journal of Computer Applications*, vol. 183, no. 26, pp. 6-10, 2021.

[9] N. Jain et al., "Fake Twitter account detection using machine learning and deep learning techniques: A review," *International Journal of Advanced Research in Computer Science*, vol. 12, no. 1, pp. 1-9, 2021.

[10] M. Baskar et al., "A comprehensive study of fake Twitter account detection using machine learning techniques," *International Journal of Advanced Research in Computer Science*, vol. 12, no. 1, pp. 16-22, 2021.

[11] Agarwal and S. Tripathi, "Fake Twitter account detection using machine learning and deep learning techniques," *International Journal of Scientific & Engineering Research*, vol. 11, no. 4, pp. 752-760, 2020.

[12] A. Almansour and H. Alhothali, "Hybrid deep learning and feature engineering for fake account detection on Twitter," *Journal of Information Security and Applications*, vol. 65, pp. 103113, 2022.

[13] W. Shu, S. Wang, and J. Tang, "Graph-based detection of fake accounts in social media," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 4, pp. 1-20, 2020.

[14] Y. Zhang and X. Zhang, "Behavioral analysis for automated fake profile detection in online social networks," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 2, pp. 312-322, 2019.

[15] T. Nguyen, P. Tran, and D. Pham, "Deep learning approach for fake Twitter bot detection," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 8, pp. 2175-2189, 2021.

[16] R. Gupta and S. Sharma, "Sentiment analysis-based approach for fake account detection in social networks," *International Journal of Data Science and Analytics*, vol. 10, no. 3, pp. 189-200, 2021.

[17] M. Hussain, A. Khan, and I. Ahmad, "Ensemble learning techniques for fake account detection in online social networks," *Neural Computing and Applications*, vol. 32, no. 11, pp. 7631-7642, 2020.

[18] K. Wang, C. Li, and H. Wu, "Impact of network structure and user activity on detecting fake profiles," *Expert Systems with Applications*, vol. 134, pp. 170-182, 2019.

[19] J. Lee and H. Kim, "Anomaly detection framework for identifying fake Twitter accounts," *IEEE Access*, vol. 8, pp. 98765-98778, 2020.