AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 3(2026)

FAKE SOCIAL MEDIA ACCOUNTS DETECTION USING MACHINE LEARNING

[1]V. Prasad, [2]Syed Umrabi, [3]Chirala Vasanthi, [4]Venkata Supriya Sambaru, [5]Sai Apurupa Ande

[1]Assistant Professor, Dept of CSE-AI&ML, St. Ann's College of Engineering and Technology Chirala-523187, India.

[2,3,4,5]U. G Student, Dept of CSE-AI&ML, St. Ann's College of Engineering and Technology Chirala-523187, India.

ABSTRACT- ***The rapid growth of social media platforms has resulted in an increase in fake accounts that are used for spam, identity theft, phishing, misinformation, and fraudulent activities. Detecting such accounts manually is difficult because of the large number of users and continuously changing account behaviour. This paper presents a Machine Learning-based Fake Social Media Accounts Detection system that automatically classifies social media accounts as real or fake using account-related features. The proposed system performs data preprocessing, visualization, feature encoding, and model training using Logistic Regression, Decision Tree, and Random Forest algorithms. After comparing the performance of these models, the best-performing model is selected and deployed in a Flask web application. The application allows users to register, log in, provide account details, and receive instant predictions. The prediction history is stored using SQLite for future reference. The proposed system provides a simple, efficient, and user-friendly solution for detecting fake social media accounts and demonstrates how Machine Learning can support online security.***

KEYWORDS*: Machine Learning, Fake Social Media Accounts, Flask, Random Forest, Data Preprocessing, Classification, SQLite, Social Media Security.*

INTRODUCTION

The widespread adoption of social media platforms has transformed the way people communicate, share information, and interact globally. Platforms such as Instagram, Facebook, Twitter, and LinkedIn have become essential for personal, educational, and business purposes. However, the increasing popularity of these platforms has also led to a rapid rise in fake social media accounts. These accounts are often created to spread misinformation, perform phishing attacks, send spam

messages, manipulate public opinion, and impersonate genuine users. In this project, a Machine Learning-based Fake Social Media Accounts Detection system is developed using Flask. The system uses account features such as platform, profile picture availability, followers count, following count, account age, number of posts, and verification status to classify an account as either real or fake. Before training the model, the dataset undergoes preprocessing steps including duplicate removal, handling missing values, label encoding, and feature scaling. Three classification algorithms — Logistic Regression, Decision Tree, and Random Forest — are trained and evaluated. The best-performing model is selected for deployment. A Flask web application is developed to provide an interactive interface where users can enter account details and receive prediction results. The application also stores prediction history using an SQLite database for future reference.

RELATED WORK

Several researchers have proposed Machine Learning techniques for detecting fake social media accounts. Earlier methods relied on manually designed features such as profile information, friend relationships, posting behaviour, and account activity. Classification algorithms like Support Vector Machine (SVM), Naïve Bayes, Decision Tree, and Logistic Regression were commonly used to identify fake profiles. Although these approaches achieved moderate accuracy, their performance depended heavily on feature engineering and often failed to adapt to changing user behaviour.

Recent studies have focused on improving detection accuracy by using advanced machine learning techniques and better feature selection methods. These approaches help in identifying complex patterns associated with fake accounts and enhance the overall effectiveness of detection systems. In comparison, the proposed system combines efficient data preprocessing with multiple Machine Learning algorithms to develop a lightweight and practical solution. The Random Forest model is selected after comparing different classifiers, providing reliable predictions while maintaining low computational complexity. By integrating the trained model into a Flask web application, the proposed system offers a user-friendly platform that enables users to detect fake social media accounts quickly and efficiently.

LITERATURE SURVEY

The detection of fake social media accounts has become an important research area

because of the rapid increase in online fraud, spam, and identity theft. Researchers have proposed several machine learning techniques to identify fake accounts by analyzing user profile information, account activity, and behavioural patterns.

One common approach uses Logistic Regression for binary classification of social media accounts. This algorithm predicts whether an account is real or fake based on extracted features such as followers, following, posts, and verification status. Although Logistic Regression is simple and computationally efficient, its performance may decrease when dealing with complex or non-linear datasets.

Another widely used algorithm is the Decision Tree classifier. It classifies accounts by creating decision rules based on feature values. Decision Trees are easy to understand and interpret, making them suitable for classification problems. However, they are prone to overfitting when trained on large datasets without proper optimization.

In this project, three machine learning algorithms — Logistic Regression, Decision Tree, and Random Forest — are trained and compared. The model with the highest accuracy is selected for deployment in the Flask web application. This approach improves prediction reliability while maintaining simplicity and efficiency.

EXISTING METHOD

Existing fake social media account detection systems mainly rely on manual verification or simple rule-based approaches. These methods use predefined conditions such as checking profile completeness, follower count, posting frequency, and account verification status. Although these techniques are straightforward to implement, they often fail to detect sophisticated fake accounts that imitate genuine user behaviour.

Some existing systems use a single machine learning algorithm for classification. While these models provide acceptable accuracy, they may not generalize well when the dataset contains different account patterns or imbalanced class distributions. Additionally, many systems do not include proper preprocessing techniques.

Another limitation of existing methods is the lack of a user-friendly interface. Many research models remain as standalone scripts without providing a web application for end users. As a result, users cannot easily interact with the prediction model or store previous prediction results.

These limitations highlight the need for an improved system that combines effective

preprocessing, multiple machine learning algorithms, and a web-based interface. The proposed system addresses these issues by integrating data preprocessing, model comparison, Random Forest classification, and a Flask-based web application for real-time fake account detection.

PROPOSED METHOD

The proposed system aims to detect fake social media accounts using Machine Learning techniques integrated with a Flask-based web application. The proposed system automatically learns from historical data and predicts whether a social media account is genuine or fake based on its profile information.

Initially, the dataset is collected and loaded using the Pandas library. The preprocessing stage includes removing duplicate records and handling missing values.

After preprocessing, the dataset is divided into training and testing sets using an 80:20 ratio. The machine learning algorithms — Logistic Regression, Decision Tree, and Random Forest — are trained on the training dataset. The model with the highest accuracy is selected for deployment.

The selected model is saved using Joblib and integrated into the Flask web application. The application provides functionalities such as user registration, user login, admin login, prediction, and prediction history. Users enter account information including platform, profile picture availability, follower count, following count, account age, number of posts, and verification status. The trained model processes these features and predicts whether the account is Real or Fake. The prediction results are displayed instantly, and the history of predictions is stored in an SQLite database for future reference.

The proposed system provides an efficient, accurate, and user-friendly solution for fake social media account detection.

SYSTEM ARCHITECTURE

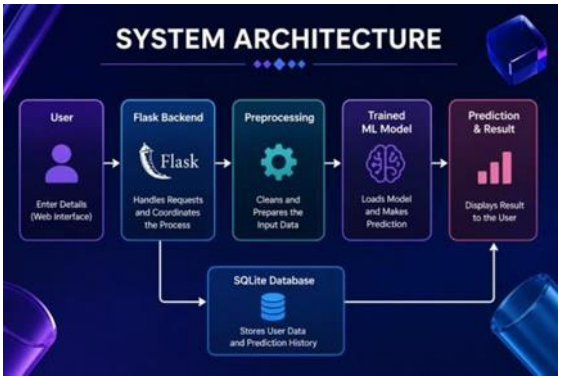

Fig 1: Fake Social Media Accounts Detection System Architecture

MODULE DESCRIPTION

The proposed Fake Social Media Accounts Detection System is divided into several modules that work together to provide accurate prediction results. Each module performs a specific task, beginning from

user registration to fake account prediction and result storage.

Module 1: User Registration

The User Registration Module allows new users to create an account by entering their personal information such as full name, username, email address, phone number, and password. The system validates the entered details before storing them securely in the SQLite database. This module ensures that only registered users can access the prediction system and provides secure authentication for future logins.

Module 2: User Login

The User Login Module authenticates registered users using their username and password. After successful authentication, users are redirected to the dashboard where they can access the fake account prediction feature. Invalid login credentials are rejected, preventing unauthorized access to the application.

Module 3: Admin Login

The Admin Login Module provides access to administrative functions. The administrator can log in using predefined credentials and monitor registered users, manage prediction records, and supervise the overall functioning of the system. This module helps maintain security and proper management of the application.

Module 4: Data Preprocessing

The Data Preprocessing Module prepares the dataset before training the machine learning models. It removes duplicate records, handles missing values, converts categorical attributes into numerical values using Label Encoding, and scales numerical features using StandardScaler. Proper preprocessing improves model performance and prediction accuracy.

Module 5: Machine Learning

This module is responsible for training and evaluating different machine learning algorithms. Logistic Regression, Decision Tree, and Random Forest algorithms are trained using the prepared dataset. Their performance is compared using accuracy, and the best-performing model is selected and saved as a pre-trained model (trained_model.pkl) for deployment.

Module 6: Prediction

The Prediction Module accepts user inputs such as platform, profile picture availability, follower count, following count, account age, number of posts, and verification status. The saved preprocessing objects and trained machine learning model process these inputs and classify the account as either Fake Account or Real Account. The prediction result is displayed immediately on the web page.

Module 7: Prediction History

Every prediction made by the user is automatically stored in the SQLite database. The Prediction History Module allows users to review previously generated prediction results, making it easier to review and track account analyses performed through the application.

RESULT AND DISCUSSION

The proposed Fake Social Media Accounts Detection Using Machine Learning system was successfully developed and implemented using the Flask framework and Machine Learning algorithms. The application provides a simple graphical user interface where users can register, login, enter social media account details, and receive prediction results instantly.

Initially, the dataset was collected and preprocessed by removing duplicate records, handling missing values, encoding categorical attributes, and scaling numerical features. After preprocessing, the dataset was divided into training and testing datasets. Three Machine Learning algorithms, namely Logistic Regression, Decision Tree, and Random Forest, were trained and evaluated. Based on the accuracy comparison, the Random Forest classifier achieved the best performance and was selected as the final prediction model.

The trained model was integrated into the Flask web application using Joblib. When users provide account information such as platform, profile picture availability, followers, following, account age, number of posts, and verification status, the application processes these inputs and predicts whether the account is Real or Fake. Every prediction is stored in the SQLite database, allowing users to review previous prediction results through the Prediction History page.

The developed system provides fast prediction with a simple and user-friendly interface. It reduces manual effort and demonstrates how Machine Learning can effectively detect fake social media accounts.

Figure 2: Home Page

The Home Page is the entry point of the application. It provides navigation options for User Signup, User Login, and Admin Login. The interface is designed to be simple and attractive, allowing users to

access different modules of the system easily.

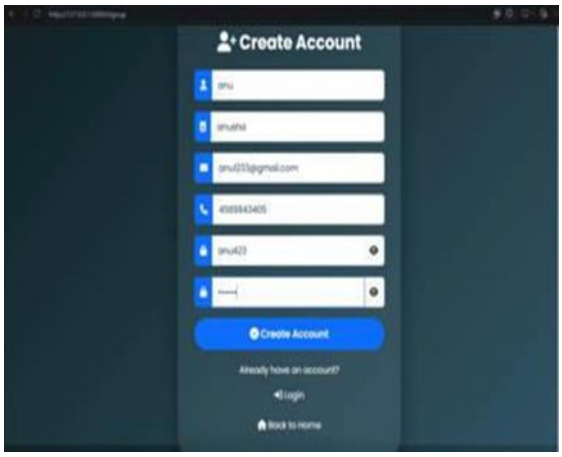

Figure 3: User Signup Page

The User Signup page allows new users to register by entering their personal details, including name, username, email address, phone number, and password. After successful registration, the information is securely stored in the SQLite database.

Figure 4: User Login Page

The User Login page authenticates registered users using their username and password. After successful authentication, users are redirected to the prediction dashboard where they can access all application features.

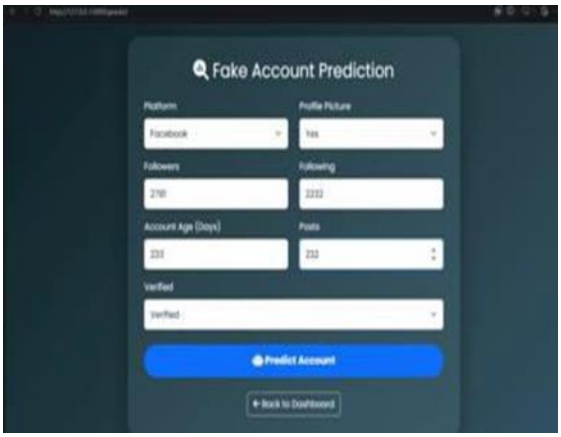

Figure 5: Prediction Page

The Prediction Page allows users to enter social media account information such as platform, profile picture status, followers, following, account age, posts, and verification status. After submitting the form, the trained Machine Learning model predicts whether the account is fake or real.

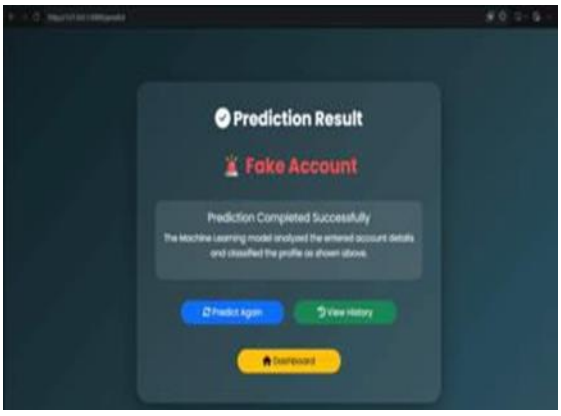

Figure 6: Prediction Result

The Result Page displays the prediction generated by the trained Random Forest model. The result clearly indicates whether the entered account is a Fake Account or a

Real Account, enabling users to interpret the prediction easily.

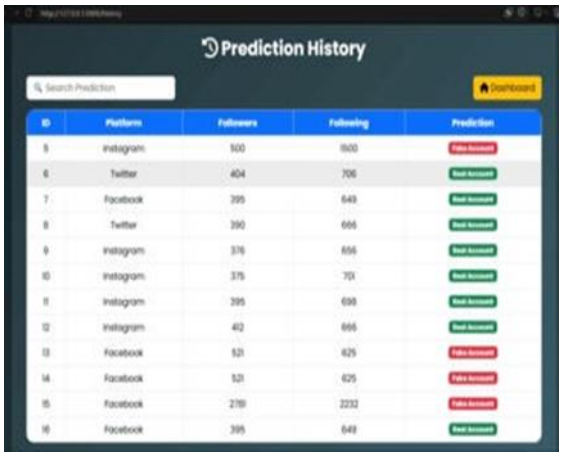

Figure 6: Prediction History

The Prediction History page stores all previous prediction records in the SQLite database. Users can review their past predictions, making the application more useful for monitoring and future reference.

CONCLUSION

The Fake Social Media Accounts Detection Using Machine Learning system was successfully designed and implemented using Machine Learning techniques and the Flask web framework. The primary objective of this project was to automatically classify social media accounts as either Real or Fake based on various account features such as platform, profile picture availability, follower count, following count, account age, number of posts, and verification status.

Overall, the developed system provides a simple, accurate, and efficient solution for fake social media account detection. It demonstrates the practical application of Machine Learning in enhancing online security and reducing the risks associated with fake profiles.

FUTURE SCOPE

Although the proposed system performs effectively, several improvements can be made in future versions to enhance its functionality and prediction accuracy. A larger and more balanced dataset can be used to improve the performance of the machine learning models and reduce prediction bias. Additional account features such as posting frequency, user interaction patterns, profile description, account activity history, and image-based features can also be incorporated for more reliable classification. These improvements would make the system more scalable, reliable, and suitable for real-world social media security applications.

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[4] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[5] M. Grinberg, *Flask Web Development*, 2nd ed. O'Reilly Media, 2018.

[6] W. McKinney, *Python for Data Analysis*, 3rd ed. O'Reilly Media, 2022.

[7] Python Software Foundation, "Python Documentation."

[8] Pallets Project, "Flask Documentation."

[9] Kaggle, "Social Media Fake Account Dataset."

[10] A. K. Jain, R. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[11] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[12] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.

[13] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[14] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.

[15] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012.