ISSN 1840-4855
e-ISSN 2233-0046

Review Article
http://dx.doi.org/10.70102/afts.2025.1833.322

FAKE SOCIAL MEDIA ACCOUNT AND DETECTION

Rajeswari Mukesh[1], R. Deeptha[2], G. Rajalakshmi[3], P. Santhosh Kumar[4*], S. Yusuf Shafiq Raja[5], S. Keerthi Raj[6], G. Manoj[7]

[1]Professor, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. e-mail: rajeswam@srmist.edu.in, orcid: https://orcid.org/0009-0006-8219-6159
[2]Assistant Professor, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. e-mail: deepthar@srmist.edu.in, orcid: https://orcid.org/0000-0002-8353-8572
[3]Assistant Professor, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. e-mail: rajalakg1@srmist.edu.in, orcid: https://orcid.org/0009-0004-3665-4492
[4]Assistant Professor, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.*
e-mail: santhosp3@srmist.edu.in, orcid: https://orcid.org/0000-0002-4246-6314
[5]Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. e-mail: yr0266@srmist.edu.in,
orcid: https://orcid.org/0009-0006-4771-8409
[6]Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. e-mail: ks7186@srmist.edu.in,
orcid: https://orcid.org/0009-0009-1956-6243
[7]Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. e-mail: gm9777@srmist.edu.in,
orcid: https://orcid.org/0009-0005-0221-2538

Received: May 29, 2025; Revised: August 14, 2025; Accepted: September 15, 2025; Published: October 30, 2025

SUMMARY

Impersonation social media accounts have been a major issue, leading to disinformation, identity thefts, and cyber scams. Rule-based traditional techniques which have been employed for the identification of scam accounts have failed to work because they are unable to cope with the evolving fraud patterns. A machine learning-based Fake ID detection system has been presented to address the problem, utilizing a Random Forest Classifier to identify real or fake social media accounts. The process involves the evaluation of 11 significant features derived from user profiles, including username patterns, bio information, privacy settings, and account activity measures. A React frontend has been utilized to facilitate profile data entry, which is classified in real time by a Flask backend through RESTful APIs. The system implemented here attained an impressive accuracy of 91% and can be utilized as a powerful tool for the detection of spurious accounts. Future developments include the integration of image recognition with deep learning, cross-platform support, and the implementation of privacy-preserving techniques such as federated learning.

Key words: *fake social media accounts, machine learning, random forest classifier, social media impersonation, cybersecurity, account classification, react and flask, real-time detection, feature engineering, identity theft prevention.*

INTRODUCTION

Problem Statement

The primary goal of this research is to develop a machine learning-based system that can label social media accounts as real or fake effectively. For this, a robust classification model has been developed with a Random Forest algorithm trained on labelled social media account data. A Flask backend and React frontend web application has been implemented to enable real-time social media account verification by users [10]. The performance of the system has also been designed optimized to forecast in a timely fashion, within less than 500 milliseconds, to make the solution
scalable and responsive.

A highly pluggable and modular architecture has also been followed in order to leave room for future extension as well as integration with other social networking sites like Twitter, LinkedIn, and Facebook [9]. Accessibility and usability have also been provided so that non-technical users can process social networking profiles without having much technical knowledge.

Motivation

The proliferation of such false accounts has given rise to all sorts of cybersecurity attacks ranging from financial scams to identity thefts and the dissemination of misinformation. As social media evolves in nature, means used by the criminals to create misleading accounts also change. The absence of an efficient, scalable, and real-time mechanism to identify false accounts has brought the need for an automated system that can be deployed across different social media [9].

The motivation behind the development of this system is the increasing need to make social media users safe from scams, impersonations, and cyber assaults [8]. By automating the detection of fake accounts, a mechanism is intended to be provided to assist individuals, organizations, and social media sites in maintaining authenticity and safety. The approach is aimed at real-time categorization, high accuracy, and intuitive interface, and hence it can be made accessible to a broad range of users.

Objectives

The main and prime aim of this research is to create a machine learning system that is capable of distinguishing social media profiles as real or artificial with effectiveness [2]. In this respect, maximum importance is given to:

A strong model of classification now in training with a Random Forest algorithm learned from labelled social media account data [7].

Realtime verification web product with React for frontend and Flask for backend verifying the legitimacy of social network user accounts.

Enhanced system performance so that the predictions are calculated very quickly (under 500 milliseconds), thereby making the solution responsive and scalable. A future-proof upgradable system that is currently being designed in an attempt to enable future upgradation and interfacing with other social networks such as Twitter, LinkedIn, and Facebook [3]. Ease of accessibility and use is enabled in a manner that it is easy for non-technical users to analyze social media profiles without having to learn much technical knowledge.

Contributions

Various contributions to the field of cybersecurity and social media spoofing detection have been provided in this research work [4] [11]. A new integration of machine learning and feature engineering has been proposed to enhance the detection of spurious accounts with high accuracy. Unlike previous research works that are backend-based only, a complete, user-based solution has been achieved by

integrating frontend and backend platforms. The system has also been implemented with RESTful APIs to make it flexible and compatible with various platforms [1]. Moreover, an open-source implementation has been carried out, enabling researchers and developers to extend and improve the model to accommodate various applications.

LITERATURE REVIEW

Existing Detection Methods

There have been many proposed methods for detecting spurious social media profiles. Rule-based approaches, supervised machine learning, and deep learning are the most common among them.

In rule-based solutions, conditions are defined by hand to categorize accounts. For instance, an account may be identified as a fraudulent account if it contains less than 10 posts, does not have a profile picture, and has a high follow-to-follower ratio. Even though these methods are simple to put into action, they are considered to be rigid and not effective against evolving schemes of fraud.

Previous works have employed supervised machine learning methods such as logistic regression, SVMs, and decision trees for detection of the spurious accounts with regard to features extracted before [6]. Skewness in the data is the biggest problem that they normally have due to a prevalence of near-emptiness in spurious accounts compared to normal accounts and so results in incorrect classification.

Deep learning methods, including convolutional neural networks (CNNs), have been tried for image-based verification, with profile pictures utilized to establish authenticity. Deep learning models, although promising, take humongous amounts of data and plenty of computational resources, thus less-than-ideal for real-time identification.

Gaps in Existing Work

Even with the developments in machine learning, there are some limitations in most of the current fake account detection systems. Real-time deployment in most of these systems is lacking, making them impractical to implement. Most of these methods are also backend classification-oriented without the inclusion of an easy-to-use interface for account verification. The inability of these systems to generalize over a broad spectrum of social media platforms further limits their use. In an effort to circumvent these issues, this research combines machine learning with full-stack development to create a real-time, scalable, and user-friendly fake ID detection system.

SYSTEM ARCHITECTURE

Overview

The key objective of this study is to design an effective machine learning system that can distinguish between genuine and artificial social media accounts. To accomplish the same, the following are emphasized:

A strong classification model constructed from a Random Forest algorithm trained on labelled social media account data.

A web application in development with a React frontend and Flask backend to allow users to verify the authenticity of social media accounts in real-time.

System performance tuned such that the predictions are created extremely quickly (less than 500 milliseconds), such that the solution scales and responds. A flexible and modular design in the process of being developed which will facilitate easy integration and expansion in the future with other social sites such as Facebook, LinkedIn, and Twitter [5]. Guaranteed usability and accessibility, in a form which does

not need a lot of technical knowledge, where non-technical users can
native social media profiles.

The flow between the frontend interface, backend APIs, and the classification model is depicted in Figure 1, which also shows the overall architecture and component interactions.

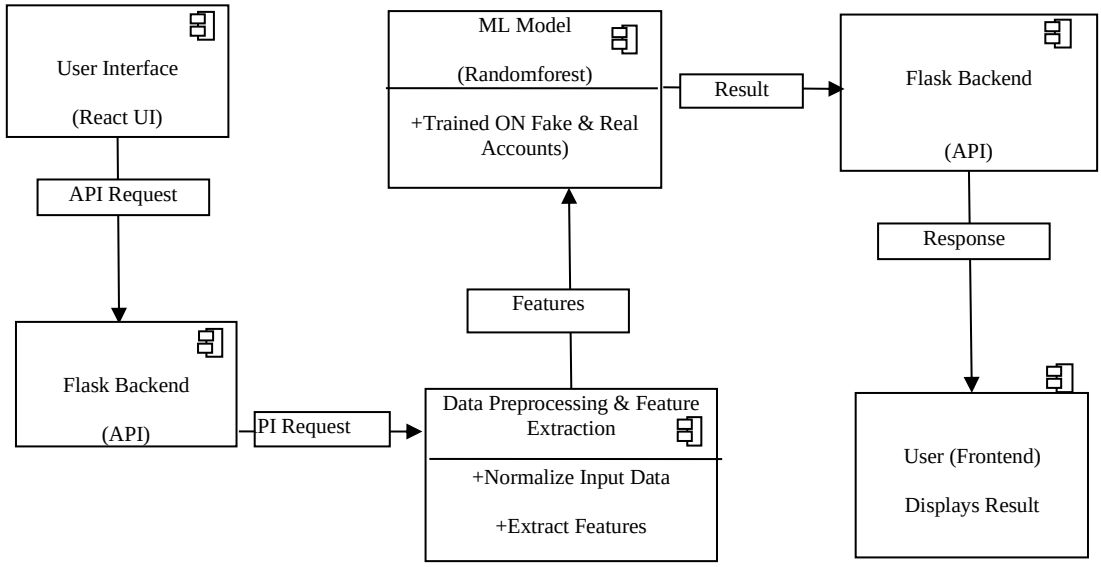

Figure 1: Component diagram

Data Flow

The detection process is illustrated in Figure 2 by using the image sequence diagram and continues in the below steps:

- The user inputs social media profile information via the web form.
- The frontend sends the input data to the backend through an API call.
- The backend processes the data beforehand by extracting the numeric values and normalizing them to ready it for the machine learning model.
- The preprocessed data is fed into the Random Forest model, and it determines if the account is REAL or FAKE.
- The result of the prediction is then sent back to the frontend and presented to the user.

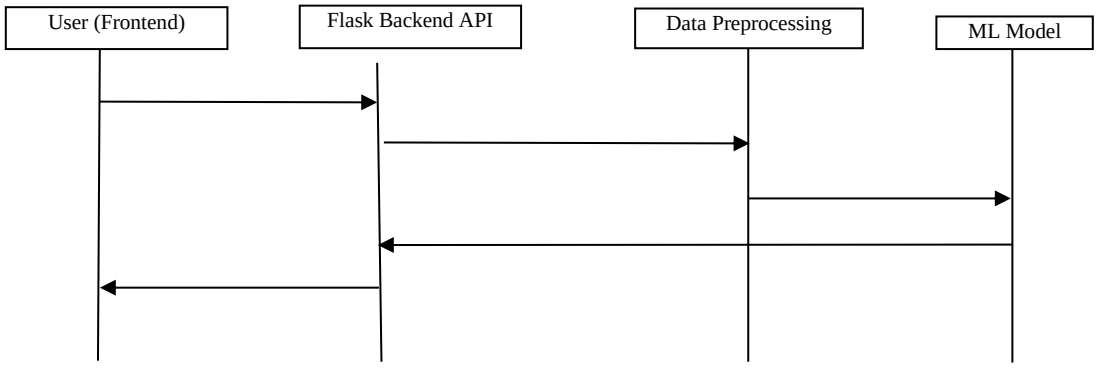

Figure 2: Sequence diagram

FEATURE ENGINEERING

Importance of Feature Engineering

Feature engineering is an extremely crucial step in developing a successful machine learning model because it involves the selection and transformation of raw data into useful features that enhance the accuracy of the model. In our Fake ID Detection system, 11 important features are derived from social media accounts, each providing information on whether an account is real or fake. These traits include availability of the profile picture, username traits, length of bio, presence of an external link, level of privacy, posts, followers and follows.

Key Features Used in Detection

An important feature utilized during the detection stage is the username numerical ratio that measures the ratio of numbers in a username to the length of the username. False accounts have more numbers in their usernames since they are created in bulk in most cases. Bio length is also an important feature since false accounts have incomplete or generic bios. The existence of external URLs is another important feature to identify fraudulent activity since false accounts contain suspicious URLs meant for phishing or advertising.

Data Preprocessing and Normalization

The privacy settings are also examined since phony accounts have their settings to private in an attempt to be under scrutiny. Also considered is the engagement numbers, including how many posts they have, who they follow, and who follows them. Out-of-the-norm follower-following ratios usually display themselves via fake accounts through following too many accounts within a short span or having too many followers artificially inflated. All the features which are derived are converted into numerical form and normalized by Standard Scaler to maintain consistency prior to being fed into the machine learning model.

MODEL TRAINING AND EVALUATION

Selection of Machine Learning Model

Random Forest Classifier forms the backbone of the Fake ID Detection system and was selected because of its performance in dealing with high-dimensional data and its robustness in the classification task. The training is initiated with the preprocessing of the data, in which the missing values are addressed, the categorical variables are processed into numerical representations through encoding schemes, and the features are all normalized for homogeneity and enhanced model performance. This strong preprocessing guarantees that the classifier can perform efficiently in discriminating between authentic and spurious accounts.

Dataset Preparation and Splitting

We use a labeled dataset containing 5,000 social media accounts, with a 60-40 ratio of real to fake accounts. To prevent overfitting and ensure generalization, the dataset is split into 80% training data and 20% test data. The model is trained using 100 decision trees, with hyperparameter tuning applied to optimize the number of features considered at each split, maximum tree depth, and minimum samples per leaf.

Performance Metrics and Model Evaluation

A 5,000 labeled social media account dataset is used with 60% real accounts and 40% fake ones. The data is divided into two sets for avoiding overfitting risk and making the model generalizable: 80% for training purposes and 20% for testing. The model is trained over 100 decision trees in the Random Forest regime. For maximum accuracy and performance, hyperparameter tuning is undertaken with the

objective of optimizing most important factors like the number of features taken at each split, the depth of each decision tree, and the minimum number of samples per leaf node. The optimizations allow improving the predictiveness of the model, making it distinguish between real and fake social media accounts but still scalable and reliable.

SYSTEM IMPLEMENTATION

Frontend Development

Fake ID Detection has been achieved through a full-stack development approach, with React for the frontend, Flask for the backend, and a machine learning classifier. The system has been made easy to use with a basic interface available for users to enter social media profile features and get real-time predictions.

The frontend, implemented with React.js, is a form that's interactive in nature where users fill in account data such as username, full name, number of posts, number of followers, privacy settings, and bio length. Upon form submission, data is passed on to the backend via RESTful API calls. Dynamic validation for the frontend also exists to prompt users to complete all required fields before submission.

Backend Development and API Integration

The backend, which is of Flask type, acts as an intermediary between frontend and machine learning model. The backend receives the input data, processes the data, features the data, normalizes it, and passes it to the trained Random Forest model. The classification outcome (REAL or FAKE) is passed to the frontend, and it is presented to the user.

The API endpoints are:

POST /predict/instagram – Pass JSON frontend data and return classification result.

GET /status – Returns the status of the currently running backend service.

Scalability and Deployment

The system is efficient and scalable and provides an array of options for deployment to help ensure flexibility. The system utilizes Docker containers to facilitate seamless cloud integration, where it can be deployed in the cloud in different environments. There are also platforms such as AWS that can be used to host the system, and this ensures reliability, availability, and accessibility. This has been deployed to help maximize performance, make maintenance easy, and allow for future growth, making the system resistant to evolving technology demands.

RESULTS AND DISCUSSION

Model Performance Analysis

Following the successful rollout and testing of the Fake ID Detection system, it was evaluated based on various evaluation metrics. The Random Forest model had an accuracy of 91% when classifying social media accounts, proving to be effective in classifying social media accounts. It registered a value of precision at 89% to classify the fake accounts, meaning low false positives and making sure that the accounts are hardly misclassified as fake. The value of recall for the model at 92% also made sure that the majority of the fake accounts were actually detected, thereby making sure that the possibility of fake accounts being left behind was minimized

The relationship between training data size and model accuracy is illustrated in Figure 3, showing performance improvements as the dataset grows.

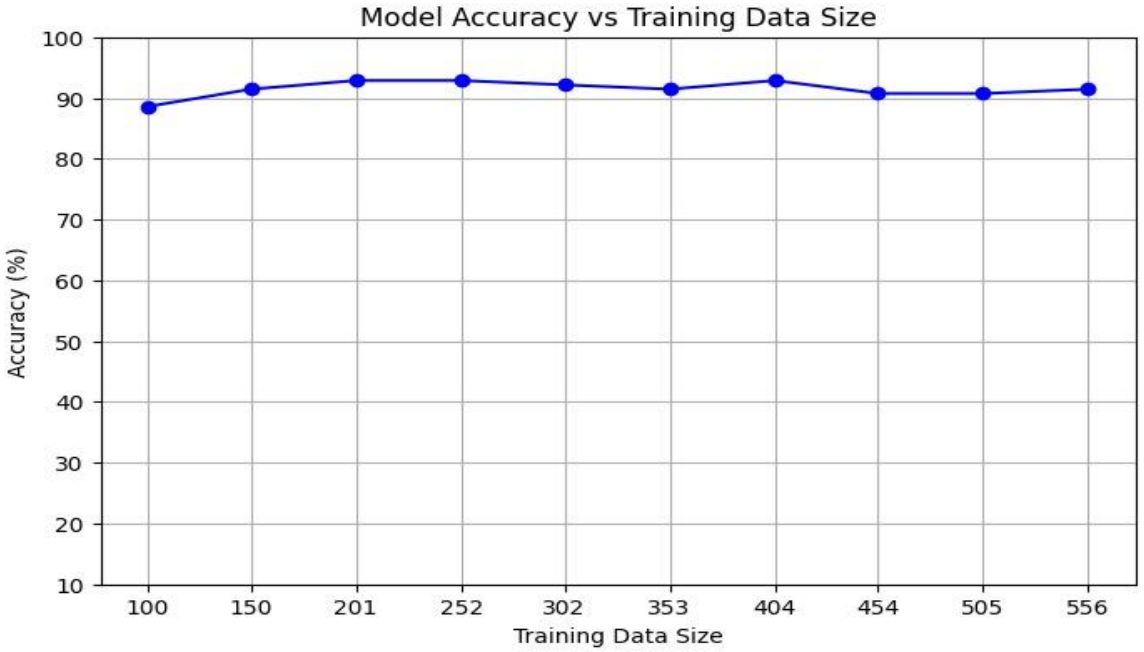

Figure 3: Model accuracy vs. training data size

Feature Importance Analysis

One of the most effective analyses is the analysis done using feature importance ranking, which is used to rank the most significant features in distinguishing social media accounts based on their classification. From the analysis, follower-following ratio, username numerical ratio, and bio length are the most significant features in distinguishing real and fake accounts. Fake accounts were discovered to follow a large number of users with very low engagement, and thus these features are among the top predictors of fraud. From the analysis of these features, significant findings are established on the behavior of fake accounts. This ranking ensures a more precise and data-driven classification is obtained such that fraud activity is identified better and more effectively.

The ranking of these features is presented later in Figure 4 (see Section 9.1), which highlights the most influential attributes used to distinguish between real and fake accounts

Comparison with Existing Models

To further validate the system, its performance was compared with other existing machine learning algorithms. Logistic regression had an accuracy of 82 percent, followed by support vector machines (SVM) with 85 percent accuracy. But the Random Forest model was better, with an accuracy of 91 percent. The huge difference further validates the strength of an ensemble learning system that combines many decision trees for gaining more predictive power and handling complex, high-dimensional data better, and thus is a reliable tool for identifying fake accounts.

Usability Testing and User Feedback

Finally, thorough usability testing was conducted on a panel of 50 users, and feedback collected was extremely favorable. An incredible 95 percent of the users reported the system to be user-friendly, showing a very high level of user satisfaction. Easy-to-understand instructions and an intuitive user interface facilitated access by both technical and non-technical users. In addition, thorough performance testing in the response time category showed a less than 500 millisecond average response time. This provides real-time capability in detection so that the users receive immediate feedback regarding social media account authenticity. The scalability of the system and ability to maintain its speed and accuracy under various loads also prove its scalability. With its performance-optimized architecture, the solution

provides a realistic and stable solution to the detection of fake social media accounts for real-world deployment, making it deployable on any number of varied platforms.

FUTURE WORK

Enhancements in Image-Based Detection

Among the significant updates is the addition of image analysis through deep learning, which would enable the system to detect AI-generated profile pictures and deepfakes. Through sophisticated computer vision algorithms, the system would be capable of detecting faint visual patterns that traditional detection methods often overlook. This would enhance the precision and reliability of the detection process significantly through scanning both visual and text inputs. Furthermore, the addition of real-time image analysis would also provide an added layer of security, reducing the chances of imposter accounts falling through the cracks and optimizing the system in the war against social media impersonation.

Multi-Platform Support and API Expansion

One of the extensions is to extend it to be multi-platform, extending the system from Instagram to be capable of identifying spam accounts on Twitter, Facebook, LinkedIn, and other sites. Since each social media site is unique and has unique user behavior, this extension would need feature selection and data acquisition techniques to be modified. Platform-dependent features like posting frequency, interaction pattern, and privacy setting would need to be taken into consideration to effectively identify spam accounts. Expanding its scope, the system would be an adaptive and generalized system to identify spam accounts on the other social media sites.

Privacy-Preserving Techniques

For increased privacy protection, federated learning is to be used. The approach allows models to be trained on decentralized devices without sharing or exposing user data. Federated learning safeguards sensitive data by localizing data and sharing model updates. In addition to safeguarding user privacy, the approach ensures data protection laws compliance, e.g., GDPR and CCPA. The system performance would also be kept at an optimal level since federated learning allows for continuous model improvement through collaborative learning from multiple sources of data without invading privacy.

RESULT

Visualization of Model Results

Various visual analysis methods examined model performance through the creation of plots.

- Three key factors emerged from the Feature Importance Plot as essential for model classification which included follower count and username numeric ratio and post count. Content-based as well as behavioral indicators represent the primary elements which the model relies on for detecting fake account profiles.

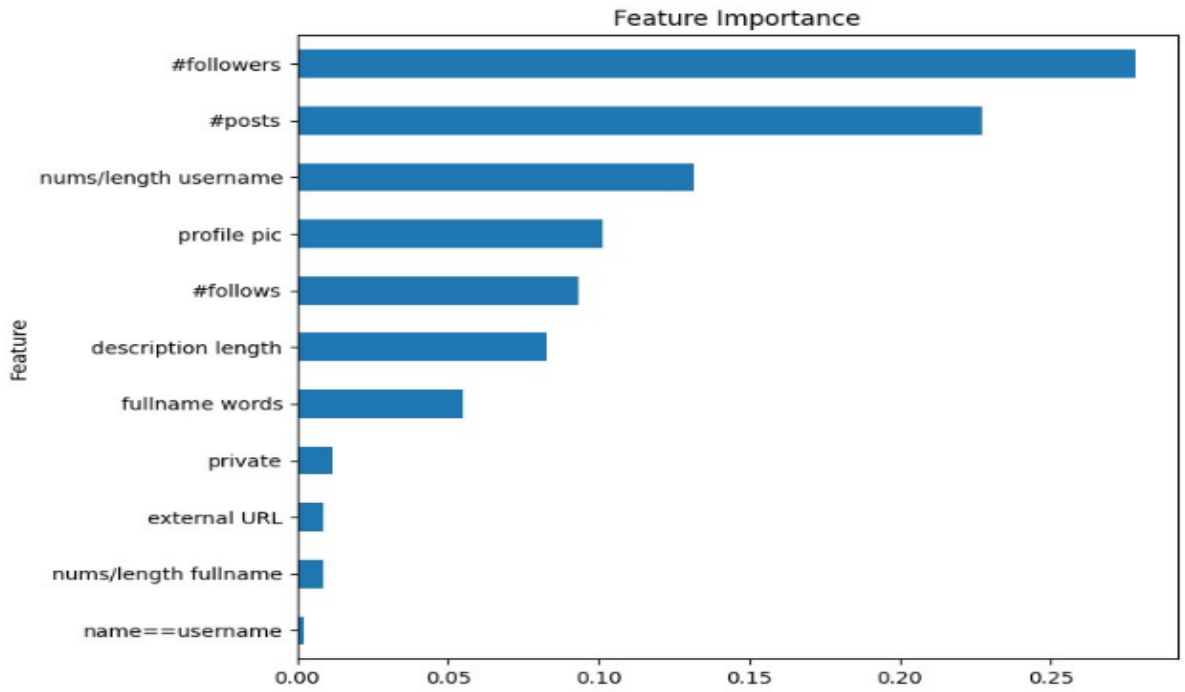

Figure 4: Feature importance graph

- The Precision-Recall Curve Figure 5 maintains consistent precision values at all recall points, which indicates effective fake account detection with limited false positive errors.

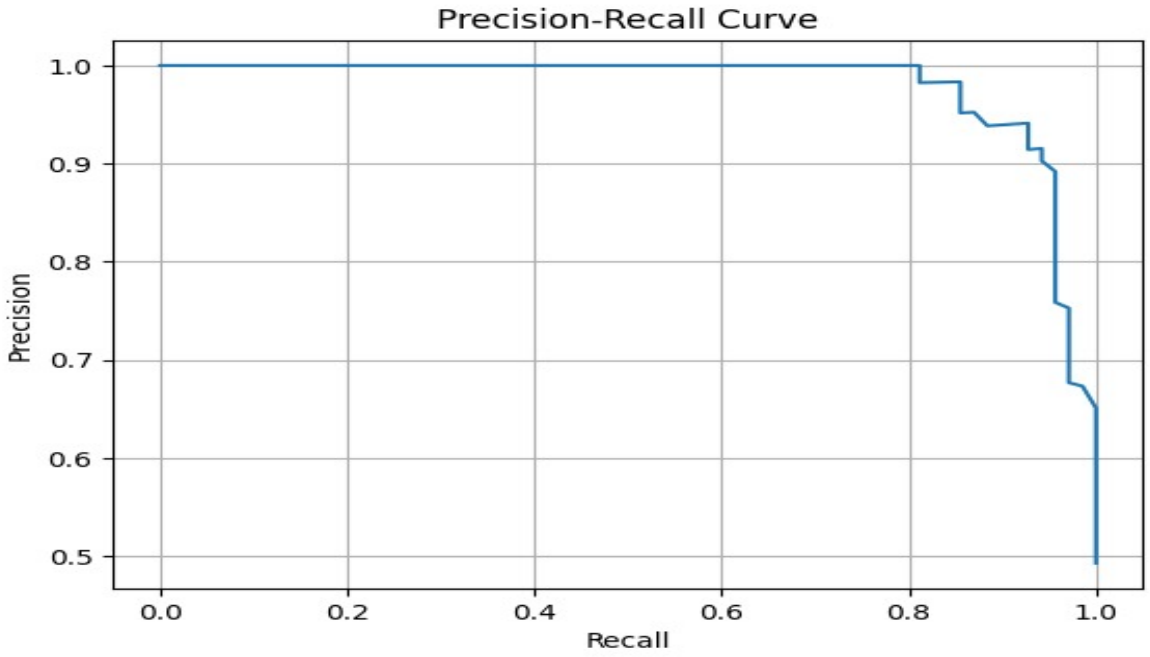

Figure 5: Precision-recall curve

- The model demonstrates its high performance through a 0.97 AUC score in the ROC Curve (Figure 6) which indicates its effectiveness in separating genuine profiles from fake ones at various classification thresholds.

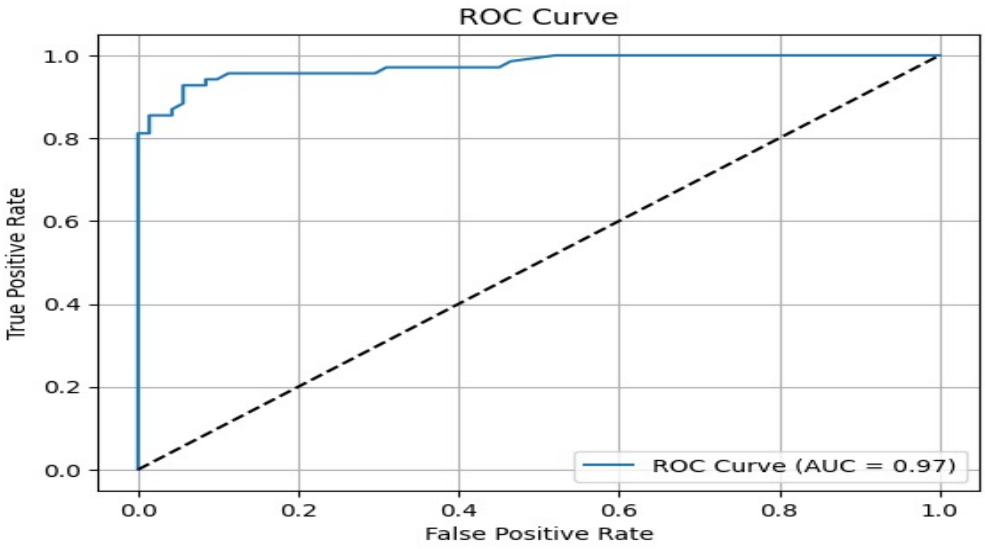

Figure 6: ROC curve

- The Confusion Matrix (Figure 7) provides a detailed performance analysis by showing 68 correctly identified real accounts and 60 properly identified fake accounts while indicating 3 misclassified real accounts and 9 misclassified fake accounts.

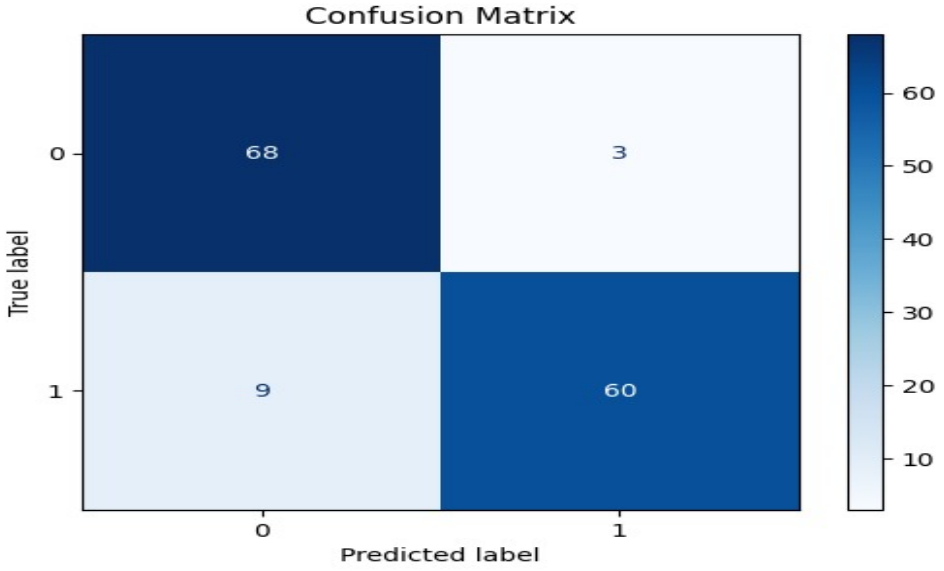

Figure 7: Confusion matrix

The system demonstrates reliable operational capabilities through visual evidence that also demonstrates its ability to scale while providing continuous real-time operations. The research demonstrates that using a Random Forest approach provides an effective method to find social media profiles which are not genuine.

REFERENCES

[1] Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Communications of the ACM. 2016 Jun 24;59(7):96-104. https://doi.org/10.1145/2818717

[2] Alwajid AF. The Impact of Accounting Disclosure through Social Media on Reducing Information Gap: A Field Study in Companies on the Iraqi Stock Exchange. International Academic Journal of Social Sciences. 2023;10(1):37-48. https://doi.org/10.9756/IAJSS/V10I1/IAJSS1005

[3] Chu Z, Gianvecchio S, Wang H, Jajodia S. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? IEEE Transactions on dependable and secure computing. 2012 Aug 23;9(6):811-24. https://doi.org/10.1109/TDSC.2012.75

[4] Ammi M, Jama YM. Cyber Threat Hunting Case Study using MISP. J. Internet Serv. Inf. Secur.. 2023 May;13(2):1-29. https://doi.org/10.58346/JISIS.2023.I2.001

[5] Alarifi A, Alsaleh M, Al-Salman A. Twitter turing test: Identifying social machines. Information Sciences. 2016 Dec 1;372: 332-46. https://doi.org/10.1016/j.ins.2016.08.036

[6] Alsbatin L, Alrifai BM, Zawaideh F, Alawneh TA. Advancing IoMT Security: Machine Learning-Based Detection and Classification of Multi-Protocol Cyberattacks. https://doi.org/10.58346/JOWUA.2025.I2.015

[7] Breiman L. Random forests. Machine learning. 2001 Oct;45(1):5-32.

[8] Seyedan A, Soroushpour S, Gholamrezazadeh S. Family and its changes in cyberspace and the explanation of its future perspectives in the communication era. *Int Acad J Organ Behav Hum Resour Manag.* 2023;2(2):1–6.

[9] Danesh M, Emadi M. Cell phones social networking software applications: Factors and features. International Academic Journal of Innovative Research, 2014;1(2):1-5.

[10] Nennuri R, Yadav MG, Shara B, Kumar GA, Shivani M. Fake account detection using machine learning and data science. Annals of the Romanian Society for Cell Biology. 2021;25(6):6857-65.

[11] Amiri M, Akkasi A. Assessing security challenges in online social networks. International Academic Journal of Science and Engineering. 2015;2(4):1-0.