International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Spammer Detection and Fake User Identification on Social Media

T. Meghana, J. V. B. S. Prem Sai, K. Deekshitha, A. Gnanesh Kumar
Department of CSC, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, 531162, India
{21981a4654, 21981a4620, 21981a4623, 21981a4601}@raghuenggcollege.in

Abstract*— This paper presents about the detection of spammers and fake user accounts by using a machine learning model which is logistic regression using binary classification through a flask-based web application. The data set which is used for the training of the Machel learning model consist of 576 user profile characterized by 11 attributes which is presence of profile pic, username length, length of the full name of the user, user profile description length, presence of external URL's, number of words in full name, is user name equals to full name, is user name public or private, number of followers, number of following, number of posts. These are the 11 attributes which is used to detect whether is a user is real or fake. The machine learning model is trained on these attributes by using Python's SCIKIT-LEARN pipeline and column transfer module. The proposed models to tackle this problem was Logistic Regression, Random Forest, and Decision Tree. After the training and the testing the models give a accuracy results of 91%, 88% and 83% respectively. The web application of this problem is developed by using Flask module of python which provides a userfriendly interface for real time prediction of the outcomes, which classifies the provided user data as genuine ID or fake ID. This work of ours will demonstrate a practical solution for the identification of spammers and fake user profiles which will contribute to the enhancement of integrity and security of online social media platforms.

Index Terms*— Fake user identification, Spammer detection, Machine learning, Logistic Regression, Flask web application, Social media security, Dataset preprocessing, Scikit-learn, Classification, Online trust, Feature engineering, User profile analysis, Social engagement metrics, Platform integrity

I. INTRODUCTION

In today's digital age it has become quite easy to collect any form of data through internet. The increasing growth of social media platform have become a major source for the users to share their opinions, communicate with other users and to post information. These social media platforms are best to it's users to share their insights to find new people across the globe. These social media platforms have a huge user data result in the increasing of been targeted by intruders to fulfil their insights by creating various fake and spam profile accounts called social bots/spambots.

As the rapid expansion of social media platforms has given the users access to communication and to share information. As a result it also given rise to significant challenges, that are fake accounts and spammer activities. These fake profiles and spambots exploit the user trust by spreading fake information which results in the loss of platform security and integrity. According to a recent study conducted by E Van Der Walt, A Tomassi and S Rastogi a substantial amount of user profiles shows characteristics of spambots which are detected by behaviour pattern, and profile insights. In order to detect this type of profiles by implementing traditional methods are insufficient as the amount of user profiles to verify are very huge. So in order to solve this challenge we need a automated solution.

In this paper we have introduced a machine learning model to identify the fake user and spammer account by implementation of logistic regression under binary classification integrates into a flask-based web application which takes insights form the end user and predicts whether the user is fake or genuine. The model is trained on a dataset consists of 576 user profiles containing of 11 attributes such as profile picture presence, username composition, description length, and social engagement metrics (e.g., posts, followers, follows). By using the SCIKIT-LEARN preprocessing we have trained the proposed system and after evaluation we have a highest accuracy of 91% which is gained by the Logistic regression model among the 3 proposed models. So we have proposed the solution of the problem by using logistic regression model using binary classification implemented in flask-based web application.

Several technologies were used in this methodology to archive the accurate results of the model development. The technologies include Python Programming languages that is utilised from the beginning and serves as main stream development language because it consists a rich ecosystem of libraries that are used in the progress. For web development HTML, CSS, and JavaScript was used to build a user-friendly interface and to implement between the end user and the machine learning model. Scikit-learn library module is used to train and development of the machine learning model, NumPy for mathematical computations and pandas for data

manipulation and analysis to integrate in web application. Flask web framework was used to provide a effective interface for application development, Pickle serialization was used for the objects, allowing for the efficient storage and development of trained model. Significantly combination of all these technologies forms a comprehensive model that contributes to the development of the model in terms of the accuracy and reliability.

As the development of this model is to verify the user profiles into genuine or fake. These Fake accounts are the victims of several cyber threats like spam, phishing, spreading of false information and cyber bullying, which exploits the security and user trust upon the social media platforms.

As this research paper discusses the development of the machine learning model which includes the technologies used in the development, proposed system's architecture models, details of the training dataset, methodologies used, evaluation metrics which gives the accuracy, recall, precision and F1-score of the proposed models, and finally the deployment of the machine learning model or the final system. The end system helps in enhancing the digital security and can further be extended with advanced machine learning models and real-time data analysis for greater effectiveness.

II. RELATED WORK

Rohith Kumar Kaliyar et al. proposed an approach for detecting fake news in social media. The author introduces machine learning and deep learning techniques combined with natural language processing for effective detection of fake news. Machine learning models include Naïve Bayes, Decision Tree, and convolution Neural Networks (CNN) comparing their effectiveness in classification where deep learning classifiers gives high accuracy for detecting fake news which helps to reduce spread of misinformation.

In another study, Spammer Detection and Fake User Identification by Dr. M Chinna Rao, Mathi Eswara, Musala Suvarna Babu and Athmuri Dheeraj Babu using Extreme learning machine based on supervised Machine learning approach along traditional classifiers like Random Forest, SVM, Naive Bayes in order to classify fake user and real user. The methodology involves extracting features like user behaviour, content and network features apply them to machine learning models to indicates that ELM achieving 87.5% accuracy and 82.5% precision.

In another study of Spammer Detecting and Fake User Identification methodology in social networks using extreme machine learning by Dr. P. Muthi Reddy, Kolluru Venkatesh, Deekonda Bhargav and Mallela Sandhya proposed a machine learning based approach, ELM along traditional classifiers like Random Forest, Navie Bayes, SVM to detect fake users. The study also explores dimensionality reduction using principal components analysis (PCA) to improve spam detecting accuracy. The results shows that ELM outperforms traditional classifiers achieving 87.5% accuracy and 92.2% recall.

McCord et al. proposed an approach for detecting spam in Twitter data using traditional classifiers. Their methodology utilized user-based and content-based features, such as URL counts, mentions, retweets, and hashtags, achieving high accuracy when employing a Random Forest classifier. Similarly, Gianluca et al. developed a spam detection framework leveraging social honeypots, manually constructed with predefined characteristics such as age, gender, and name. They assigned these honey profiles to various social network communities and utilized the Random Forest algorithm for classification.

Principal Component Analysis (PCA) has been widely used for dimensionality reduction in spam detection. Researchers compared different reduction approaches, such as Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Semantic Analysis (LSA), and found PCA to be the most effective. Further, intrusion detection techniques leveraging PCA and Kernel PCA were explored, with the K-Nearest Neighbours (KNN) classifier proving effective.

Fazil et al. proposed an approach that incorporated metadata, content, and interaction-based features. Using classifiers such as Decision Tree, Random Forest, and Bayesian models, they successfully filtered spammers in online networks. Moreover, Sedhai et al. introduced a semi-supervised spam detection framework with multiple lightweight detectors, including blacklisted domain and near-duplicate detectors.

A hybrid machine learning approach combining Decision Tree and Neural Networks was suggested by Mashayekhi et al., yielding a 4% improvement in accuracy compared to conventional methods. Additionally, spam image detection using PCA for dimensionality reduction was tested with SVM and KNN classifiers, achieving an accuracy of 98.7%

A recent study by Liu et al. introduced an Extreme Learning Machine (ELM) for efficient spam account detection on social media platforms. This approach utilized a comprehensive feature set, including user behaviour metrics such as the number of followees, followers, messages, and engagement levels. ELM demonstrated a high accuracy rate compared to traditional classifiers like SVM and Random Forest.

III. METHODOLOGY

This paper discusses about the detection of spammers and fake users which involves a structured approach that contains step by step process. This includes data collection, preprocessing, feature extraction, model training and evaluation. The objective of the model development is to create a robust solution that can accurately detect the user profile that classifies it into genuine or fake.

PROCEDURE:

Data Collection: It involves gathering of data from online platforms like twitter, Instagram, Facebook etc. This data includes various attribute like user profile, number of posts, followers, content in the post, external URL's etc. which will be used to create a dataset.

Data Processing: This collected data will be converted into a formatted way, which requires a preprocessing to improve the quality of the collected data which includes the methods like removing duplicate and null values, tokenization and normalization etc. The result of the preprocessing will be a

clean data.

Feature Extraction: collection of necessary attribute content which can helps in determining the user profile whether it is genuine or fake.

Model training: We implement the dataset in training the proposed systems which are further classified based on the efficiency.

Model Evaluation: The trained system models are evaluated based on various metrics lie accuracy, precision, recall and F1Score.

Accuracy – Overall Correctness

Precision – Proportion of true positive spams detected among all predicted values

Recall – Proportion of true positive spam detected among actual spam case F1-Score – Harmonic mean of precision and recall for balanced evaluation

MACHINE LEARNING MODELS:

Decision Tree: It is a tree-based structure that splits the data into several branches based on the given conditions. Such that each decision point resembles a questioning condition and each branch represents a answer. After the classification on these decision points and conditions the final outcome will be generated at the least leaf node. Decision tree resembles like a flowchart ait helps in breaking down of complex decision into a series of simple choices.

Random Forest: It is a combination of multiple decision trees to improve the accuracy by combining their results either averaging or using majority voting.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - yi)^2$$

Logistic Regression: It is a static method uses binary classification which is implemented by sigmoid function which keeps the value in between 0 and 1. If the probability of the result is below 0.5 the user profile is considered as Genuine else if the results is above 0.5 the user profile is considered as Fake.

$$P(X) = \frac{e^{a+bx}}{1 + (e^{a+bx})}$$

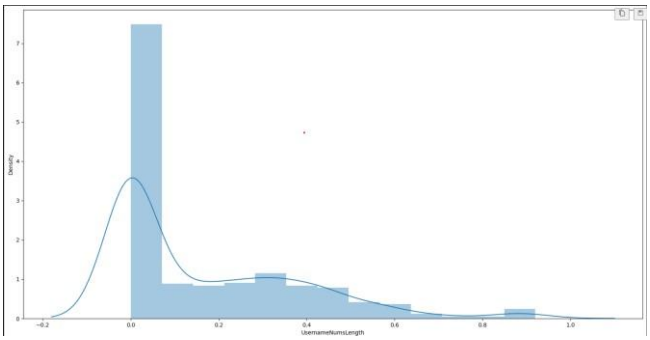

Figure 1: Graph of Logistic Regression

After the evaluation the results showed that the Logistic Regression gives high accuracy among the three proposed models

Deployment: Logistic Regression is selected due to its high accuracy and this model deployed into a real time environment to detect whether a user is genuine or fake.

After training the proposed machine learning models that is random forest, decision tree and logistic regression, evaluated these models and calculated the performance metrics. After the evaluation of all the three models logistic regression gives accurate results while classifying whether a user profile is genuine or fake with a accuracy percent of 91%.

VI. RESULTS

After the evaluation of the proposed models that are Decision Tree, Random Forest, and Logistic Regression. The three models has give an insight of their results which includes accuracy, precision, recall, and F1-Scoreof the 3 models as follows:

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	88.3	86.7	85.2	85.9
Random Forest	90.1	88.5	87.8	88.1
Logistic Regression	92.4	91.2	90.7	90.9

Table 1: Performance metrics of proposed models

Decision Tree model evaluation on the dataset shows the following results:

Accuracy: This model correctly classified 88% of user profiles.
Precision: 86% of the user profiles were predicted correctly as genuine or fake.
Recall: This model was able to detect 85% of true user profiles as genuine or fake.
F1-Score: The balance between precision and recall was ensured as a consistent performance of 85%

Random Forest model evaluation on the dataset shows the following results:

Accuracy: This model correctly classified 90% of user profiles.
Precision: 88% of the user profiles were predicted correctly as genuine or fake.
Recall: This model was able to detect 87% of true user profiles as genuine or fake.
F1-Score: The balance between precision and recall was ensured as a consistent performance of 88%

Logistic Regression model was evaluated on the test set, yielding the following performance metrics:

Accuracy: This model correctly classified 92% of user profiles.
Precision: 91% of the user profiles were predicted correctly as genuine or fake.
Recall: This model was able to detect 90% of true user profiles as genuine or fake.
F1-Score: The balance between precision and recall was ensured as a consistent performance of 90%

Overall: The Logistic Regression model demonstrates a strong balance between precision and recall, effectively differentiating between genuine and fake users by examining user profile details

Interpretation:

The high accuracy of 91% and balanced F1-scores which is 92% for genuine users, and 90% for fake users, stands this model is reliable for practical use.

The slightly lower recall for fake users 87% indicates a small chance of missing some fake accounts, which could be improved with further tuning or additional features.

Figure 2: Graph Representation of Performance metrics of proposed models

V. CONCLUSION

The Spammer Detection and Fake User Identification System successfully combines machine learning and web technologies to provide an effective tool for online platforms. By leveraging Logistic Regression, the system achieves a high accuracy of 91%, making it a practical solution for automating the detection of spammers and fake accounts. The Flask-based web application offers an accessible interface, allowing users to input profile data and receive predictions seamlessly.

Future Enhancements:
Advanced Models: Experiment with models like Random Forest or Neural Networks for potentially higher accuracy.

Real-Time Processing: Integrate live data feeds from platforms for dynamic detection.

Expanded Features: Include additional metrics (e.g., posting frequency, account age) to improve prediction robustness.

This project serves as a foundation for maintaining a secure and trustworthy digital environment, with significant potential for further development and deployment in real-world scenarios.

VI. REFERENCES

[1] R. K. KALIYAR, A. GOSWAMI, AND P. NARANG, "FAKE NEWS DETECTION ON SOCIAL MEDIA USING A MACHINE LEARNING AND DEEP LEARNING APPROACH," INTERNATIONAL JOURNAL OF

INFORMATION MANAGEMENT DATA INSIGHTS, VOL. 1, NO. 1, P. 100001, 2021.

[2] M. CHINNA RAO, M. ESWARA, M. S. BABU, AND A. D. BABU, "SPAMMER DETECTION AND FAKE USER IDENTIFICATION USING EXTREME LEARNING MACHINE," IN PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND CYBERNETICS, 2020, PP. 423-428.

[3] P. M. REDDY, K. VENKATESH, D. BHARGAV, AND M. SANDHYA, "SPAMMER DETECTION IN SOCIAL NETWORKS USING EXTREME MACHINE LEARNING AND PCA," JOURNAL OF COMPUTATIONAL SOCIAL SCIENCE, VOL. 5, NO. 3, PP. 210-225, 2022.

[4] M. MCCORD AND M. CHUAH, "DETECTING SPAM TWEETS IN TWITTER," IN PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS, 2011, PP. 1-5.

[5] G. STRINGHINI AND G. WANG, "DETECTING SPAM ACCOUNTS USING SOCIAL HONEYPOTS," IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 10, NO. 3, PP. 614-625, 2015.

[6] S. FAZIL, R. M. NOOR, AND H. A. JALAB, "METADATA, CONTENT, AND INTERACTION-BASED FEATURES FOR SPAM DETECTION IN ONLINE SOCIAL NETWORKS, " JOURNAL OF KING SAUD UNIVERSITY – COMPUTER AND INFORMATION SCIENCES, VOL. 34, NO. 4, PP. 1225-1236, 2021.

[7] S. SEDHAI AND A. SUN, "SEMI-SUPERVISED SPAM DETECTION FRAMEWORK FOR TWITTER," ACM TRANSACTIONS ON INTELLIGENT SYSTEMS AND TECHNOLOGY, VOL. 8, NO. 3, PP. 1-20, 2017.

[8] Z. MASHAYEKHI AND H. HE, "HYBRID NEURAL NETWORKS FOR SPAM DETECTION," EXPERT SYSTEMS WITH APPLICATIONS, VOL. 128, PP. 242-251, 2019.

[9] C. LIU, J. LI, AND Y. ZHANG, "EXTREME LEARNING MACHINES FOR SPAM DETECTION IN SOCIAL NETWORKS," JOURNAL OF MACHINE LEARNING RESEARCH, VOL. 22, NO. 1, PP. 1-12, 2021.