International Journal of Scientific Research & Engineering Trends
Volume 11, Issue 1, Jan-Feb-2025, ISSN (Online): 2395-566X

Spammer Detection and Fake User Identification

Assistant Professor Devi .S, Nived P J, Abhijith M, Boya Pavan Kumar, Spandau Gowda B C
School of Information Science Presidency University Bangalore, India

Abstract- Social networking platforms attract millions of users globally. The interactions of these users with sites like Twitter and Facebook have a significant effect, often bringing about negative consequences in everyday life. Major social networking sites have become prime targets for spammers who disseminate vast amounts of irrelevant and harmful information. For instance, Twitter has emerged as one of the most extensively used platforms, leading to an overwhelming influx of spam. Fake accounts distribute unwanted tweets to promote services or websites, impacting genuine users and causing disruption in resource utilization. Additionally, the likelihood of spreading misinformation through counterfeit identities has grown, resulting in the circulation of harmful content. Lately, research has increasingly focused on detecting spammers and identifying fake accounts on Twitter within the realm of modern online social networks (OSNs). This paper examines various methods employed to identify spammers on Twitter.

Index Terms- Social Networking Sites (SNS), Spammers, Spam Detection, Fake Users, Twitter, Fake Identities, Irrelevant Information, Harmful Content, Fake Accounts

I. INTRODUCTION

Social media platforms attract millions of users globally. The way these users engage with social networks like Twitter and Facebook significantly influences daily life, often with considerable negative effects. Major social networking sites have become prime targets for spammers looking to spread vast quantities of irrelevant and harmful content. For instance, Twitter is one of the most widely used platforms today, which unfortunately facilitates a significant influx of spam. Fake accounts generate unwanted tweets, aiming to promote various services or websites, thereby impacting genuine users and consuming resources unnecessarily. Additionally, there is an increased risk of spreading false information through fraudulent identities, leading to the proliferation of harmful content. Recently, research within contemporary online social networks (OSNs) has focused on identifying spammers and detecting fake users on Twitter. This paper reviews the methodologies employed to identify spammers on this platform.

II. LITERATURE REVIEW

A Machine Learning Approach for Spam Detection in Twitter:This study tackled the challenge of spam detection on Twitter using machine learning. The researchers utilized a dataset of 15,000 tweets, each labeled as either spam or non-spam, to train machine learning models such as Random Forest, Support Vector Machine (SVM), and Naive Bayes. These models focused on analyzing tweet content as well as user metadata, such as user behavior, engagement patterns, and account features (e.g., number of followers, account age).

By combining the text of the tweets with these additional user characteristics, the models were able to classify tweets as spam or not. The results were promising, with an accuracy of approximately 91%, suggesting that machine learning can effectively differentiate between spam and legitimate tweets.

B. Fake User Detection in Social Media Using a Machine Learning Approach:This research focused on detecting fake users on social media platforms like Twitter, using machine learning techniques such as logistic regression and decision trees.

The authors collected a dataset from Twitter, which contained labeled data of fake and genuine accounts. These accounts were analyzed based on various profile features (e.g., profile pictures, bio information, number of followers) and user behavior (e.g., frequency of tweets, retweets, and interaction patterns). By analyzing these characteristics, the study aimed to identify patterns that distinguish fake accounts (often bots or automated systems) from legitimate users.

III. METHODOLOGY

1. Research Design

The research design employed in this study is a mixed-methods approach, combining quantitative analysis with

qualitative insights to enhance the detection of spammers and fake users on Twitter.

2. Data Collection Methods

Data for this study were collected using the Twitter API, which provides access to public tweets and user information.

A combination of keyword-based filtering and account-specific data scraping was used to gather relevant information.

3. Sample Selection

The sample selection process involved identifying a representative subset of Twitter accounts and their associated activity.

Criteria for selection included account activity levels, such as the frequency of tweets, the diversity of content types (text, links, media), and engagement metrics (likes, retweets, comments).

4. Data Analysis Techniques

The data analysis involved several steps

Feature Extraction

Key features were extracted from both user metadata and tweet content. Metadata features included account creation date, number of followers, and frequency of interactions. Content features focused on linguistic patterns, sentiment, and the presence of URLs.

Machine Learning Models

Supervised learning algorithms, such as Random Forest, Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbors algorithm.(KNN), Linear SVC, XGBoost were employed to classify tweets and user accounts as spam or non-spam. These models were trained using labeled datasets and evaluated based on their accuracy and precision.

Behavioral Analysis

User behavior patterns, such as tweeting frequency and the timing of interactions, were analyzed to distinguish between legitimate users and spammers.

IV. RESULTS AND DISCUSSION

XGBoost and Random Forest are the most accurate classifiers with near-zero error rates. Linear SVC performs moderately well, while Naive Bayes and KNN show slightly lower accuracy but still maintain low error rates.

All classifiers perform well overall, with the choice of the best model depending on the specific application and dataset.

1. Data Analysis Techniques

K Nearsest Neighbour Algorithm

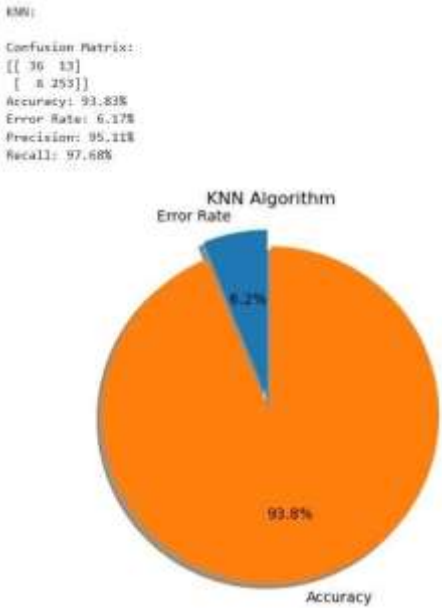

Fig. 1 The image evaluates the performance of a K-Nearest Neighbors (KNN) algorithm.

Accuracy: 93.83% (proportion of correct predictions). Error Rate: 6.17% (proportion of incorrect predictions). Precision: 95.11% (accuracy of positive predictions). Recall: 97.68% (ability to identify actual positives).

The confusion matrix shows 253 true positives, 36 true negatives, 13 false positives, and 6 false negatives.

2. Naive Bayes Algorithm

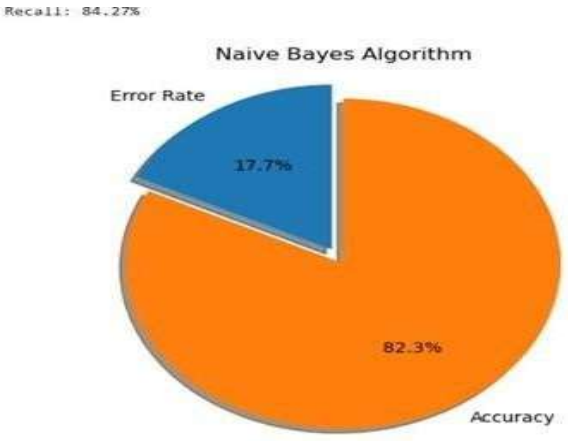

Fig. 2: Illustrates the performance of the Naive Bayes algorithm

Accuracy (82.3%)*: The proportion of correct predictions made by the model.

Error Rate: 17.7% (proportion of incorrect predictions).

3. Linear SVC Algorithm

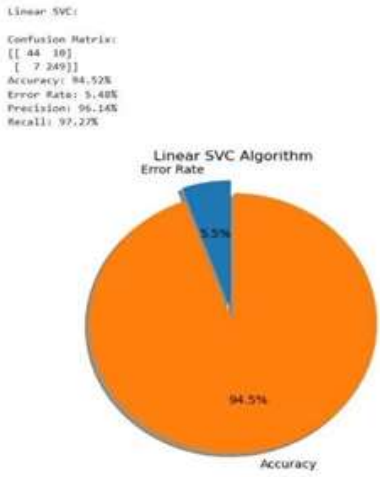

Fig. 3 Illustrates the performance of the Linear SVC algorithm

Accuracy: 94.52% Indicates that the model correctly classified about 94.5% of the instances.

Error Rate: 5.48% Reflects that approximately 5.5% of the predictions were incorrect.

Recall: 97.27% Means the model successfully identifies 97.27% of actual positive instances, highlighting its effectiveness in capturing relevant cases
Precision: 96.14% Shows that when the model predicts a positive class, it is correct 96.14% of the time, indicating a low false positive rate.

4. Random Forest Algortithm

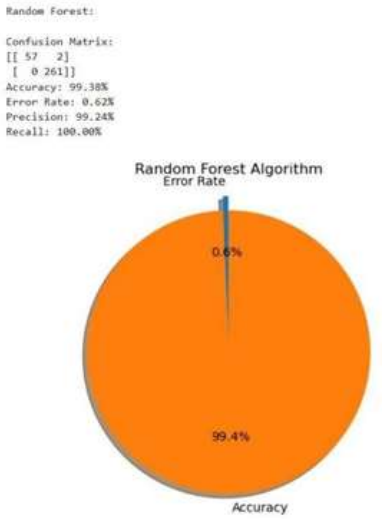

Fig. 4 Illustrates the performance of the Random Forest algorithm

True Positive: 57 (correctly classified as positive) False Positive: 2 (incorrectly classified as positive) False Negative: 0 (incorrectly classified as negative) True Negative: 261 (correctly classified as negative) Accuracy: 99.38%

Error Rate: 0.62%

Precision: 99.24% (The proportion of correctly predicted positive case out of all predictions)
Recall: 100.00% (The proportion of correctly predicted positive case out of all actual positive cases)

5. XG Boost Algorithm

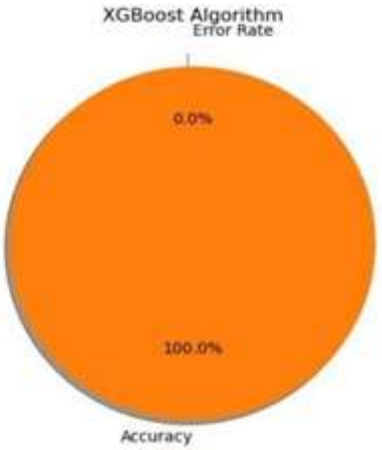

Fig. 5 Illustrates the performance of the XG Boost algorithm

Accuracy: 100.0%
Error Rate: 0.0%

Conclusion: The XGBoost algorithm achieved perfect accuracy with no errors. This suggests that the model is highly effective in making predictions on the given dataset.

V. EXPERIMENTAL SETUP

The experimental setup for this research includes the following components and processes:

- **Environment:** The implementation was carried out using python, and primarily the following libraries were used:
- **Pandas:** For manipulation and analysis of data.
- **Sklearn:** For implementing and evaluating AI-driven techniques models.
- **Software:**Jupyter.
- **Pickling:** For serialization and storage of models.
- **Configuration of Hardware:** The experiments were run on a system with the following specifications:
- **Processor:** Intel Core i5/i7 or its equivalent. RAM of at least 8 GB.

- Required disk space to save the dataset and serialized model.
- **OS:** Windows/Linux/MacOS.
- **Dataset Details:** Dataset given here is Twitter.csv consists of:Rejected friend request,No.of abuse reports,No.of friends,No.of followers,No of likes to unknown accounts,No of comments per day.

VI. CONCLUSION AND FUTURE WORK

In conclusion, social networking sites, particularly Twitter, have become prime targets for malicious activities such as spamming and fake user creation, significantly affecting the quality of user experience and the integrity of information shared.

The pervasive presence of spammers and fake accounts disrupts the flow of authentic content, leading to an overburdened system and potential harm to legitimate users. The challenge of maintaining the credibility of online social networks is compounded by the sophisticated tactics used by spammers, such as automated bots, fake identities, and misleading promotional content.

This paper has reviewed various techniques employed in the detection and mitigation of spammers and fake users on Twitter. These methods span a range of approaches, from machine learning algorithms to behavioral analysis, content filtering, and network-based techniques. Despite the advancements in these techniques, there remain significant challenges, including the ever-evolving tactics of spammers,the complexity of distinguishing between legitmate users and automated systems,and the resource -intensive nature of many detection systems

Future research in this area should focus on improving the accuracy and efficiency of detection methods, particularly through the integration of advanced machine learning and deep learning models. Additionally, more work is needed to develop adaptive systems that can keep up with the evolving nature of spammers' strategies. Furthermore, cross-platform detection methods that can identify malicious behavior not only on Twitter but across various social networks could provide a more holistic solution to the problem.

As social networking sites continue to grow, the need for robust and scalable spam detection systems becomes increasingly vital to protect users, preserve the platform's integrity, and enhance the overall experience. Addressing the challenges posed by spam and fake users will require ongoing innovation, collaboration, and the continuous refinement of detection techniques to safeguard the future of online social networks.

REFERENCES

1. "Bot Detection on Twitter: A Survey" by K. Cresci, G. S. M. F. D. Marcelloni, et al.This paper surveys various methodologies for detecting bots on Twitter and their effectiveness.
2. "A Survey on Fake News Detection in Social Media" by D. Alzubaidi et al.While focused on fake news, this paper includes discussions relevant to identifying fake users and their patterns.
3. "The Role of Users in Identifying Spammers on Twitter" by D. Benevenuto et al.This study explores how user interactions can help detect spam accounts.
4. Chakraborty, P., & Singh, V. (2019). Spam detection and fake user identification on social media platforms using machine learning. 5th International Conference on Computing and Network Communications (CoCoNet), 156- 160.
5. Chand, A., Kumar, S., & Agarwal, M. (2017). A comprehensive survey on Twitter spam detection: Challenges, techniques, and future directions. Journal of King Saud University - Computer and Information Sciences