International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Machine Language Based Fake Social Media Account Detection Analysis

Rubi Kumari [a], ***Kirti Bhatia*** [b], ***Rohini Sharma*** [c]

[a] *Student, Sat Kabir Institute of Technology and Management, Bahadurgarh, India*
[b] *Assistant professor, Sat Kabir Institute of Technology and Management, Bahadurgarh, India*
[c] *Assistant professor, GPGCW, Rohtak, India*

ABSTRACT

The proliferation of social media platforms, such as Instagram, has led to an increasing presence of fake and automated accounts, posing significant challenges to user privacy, security, and platform integrity. Detecting and addressing these accounts is crucial for maintaining a trustworthy and authentic user experience. This research focuses on the detection of fake and automated accounts on Instagram through the development of an advanced and effective detection system. By analyzing account characteristics and leveraging machine learning algorithms and data analysis techniques, a novel approach is proposed to accurately identify and differentiate fake and automated accounts. The research objectives involve investigating unique account features, exploring existing detection techniques, and designing a comprehensive detection framework for enhanced accuracy and efficiency. A prototype system will be implemented to demonstrate real-time detection and classification capabilities. Extensive experiments using diverse datasets will evaluate the system's performance, including comparisons with existing detection methods. The research outcomes aim to provide valuable insights into combating the issue of fake and automated accounts on Instagram, contributing to a more secure and trustworthy social media environment.

Keywords: Instagram Account, Fake Account Detection

1. Introduction

Fake and automated accounts on Instagram present various risks and undermine the authenticity of the platform. These accounts are created with malicious intent, engaging in activities such as spamming, phishing, identity theft, and the dissemination of disinformation. Furthermore, automated accounts, commonly referred to as bots, can manipulate engagement metrics, distort social interactions, and deceive users by appearing as legitimate accounts. Consequently, user trust is eroded, and the credibility of the platform is compromised.

The presence of fraudulent and automated profiles on social media platforms, including Instagram, has emerged as a critical challenge. These profiles engage in malicious activities, such as spamming, phishing, identity theft, and the dissemination of misinformation, leading to significant risks for users and undermining the authenticity of the platform. Furthermore, automated profiles, commonly known as bots, can manipulate engagement metrics, distort social interactions, and deceive users by posing as genuine accounts. The prevalence of such profiles compromises user trust, hampers the user experience, and threatens the credibility of Instagram as a reliable social media platform.

Detecting and mitigating fraudulent and automated profiles on Instagram poses a complex and evolving problem. Traditional methods of manual user reporting and rule-based approaches are insufficient to handle the scale and sophistication of these accounts. The dynamic nature of fraudulent profiles and evolving tactics employed by automated accounts require advanced detection mechanisms that can accurately differentiate them from genuine user accounts. the primary problem addressed in this research is the development of an advanced and effective detection system for identifying fraudulent and automated profiles on Instagram. This system should overcome the challenges posed by the diversity of Instagram's user base, adapt to evolving tactics employed by malicious actors, and accurately differentiate between genuine and fraudulent activity. Additionally, the system should be evaluated and refined to enhance its accuracy, reduce false positives and false negatives, and improve the overall efficiency of profile detection on Instagram.

We have Developed an advanced detection framework with Design and implement a comprehensive detection framework for identifying suspicious profiles on Instagram. The framework utilizes machine learning algorithms, data analysis techniques, and behavioral patterns to enhance the accuracy and efficiency of profile detection.

2. Related Work

The existing research on fraudulent and automated profile detection on social media platforms has made significant strides. However, several challenges persist. Firstly, the rapid growth and diversity of Instagram's user base necessitate the development of detection techniques that can effectively identify fraudulent and automated profiles across various demographics and account types. Secondly, the evolving tactics employed by malicious actors continually pose new challenges, requiring adaptive and robust detection systems that can keep pace with emerging threats. Thirdly, the detection of sophisticated automated accounts that mimic human behavior remains a significant challenge, as these accounts blur the lines between genuine and fraudulent activity.

Authors in [1] conducted a study on Twitter to identify indicators of fake accounts. They found that fake accounts often exhibit abnormal posting patterns, including high-frequency posting, repetitive content, and low follower-to-friend ratios. By analyzing these indicators, they developed a classification model that achieved accurate detection of fake accounts.

Authors in [2] investigated fake accounts on Facebook and Twitter. They identified several characteristics commonly associated with fake accounts, including the use of generic profile pictures, high friend-request activity, and low engagement with other users. Their findings indicated that these indicators can be effective in distinguishing fake accounts from genuine ones.

Machine learning algorithms have been widely employed in fake account detection due to their ability to analyze large datasets and identify patterns. Researchers have utilized various machine learning techniques, including supervised learning, unsupervised learning, and ensemble methods, to develop classification models for fake account detection.

Authors in [3] proposed a machine learning-based approach for fake account detection in online social networks. They considered factors such as account age, posting frequency, friend connections, and profile completeness as input features for their classification model. By training the model on a labeled dataset, they achieved accurate identification of fake accounts.

Authors in [4] incorporated machine learning algorithms for the detection of fake profiles on social media platforms. They analyzed features such as account creation time, posting behavior, network properties, and textual content to develop a classification model. Their results demonstrated the effectiveness of machine learning techniques in distinguishing fake profiles from genuine ones.

Social network analysis techniques have been utilized to detect fake accounts by analyzing the network structure, connections, and interactions among users. Researchers have examined the network properties and engagement patterns of fake accounts to identify distinct characteristics that differentiate them from genuine accounts.

Authors in [4-10] conducted a comprehensive analysis of fake accounts on Facebook and Twitter, focusing on their social network characteristics. They found that fake accounts tend to have higher friend-request activity, engage less with other users, and exhibit clustering behavior in the network. These findings suggest that social network analysis can be a valuable approach for fake account detection.

3. Methodology

3.1 Data Collection

The selection and acquisition of data for fake account detection can vary depending on the social media platform and research objectives.

A. API Access

Many social media platforms provide APIs that allow researchers to access user data, including profile information, posts, and social connections. By utilizing APIs, researchers can collect data directly from the platform in a structured and controlled manner.

B. Web Scrapping

In cases where APIs are limited or unavailable, researchers resort to web scraping techniques to extract data from social media platforms. Web scraping involves automating the retrieval of information from web pages by parsing their HTML structure. This approach allows researchers to gather data, such as user profiles, posts, and engagement metrics.

C. Ground Truth Database

Some studies make use of pre-existing datasets that have been labeled or annotated by experts as fake or genuine accounts. These datasets serve as benchmarks for evaluating the performance of detection algorithms. . For fake account detection, we have two datasets, each of which contains account information in the JSON data format. The "fakeAccountData.json" dataset consists of 994 records with fake account information, while the "realAccountData.json" dataset contains 200 records with real account information.

3.2 Data Processing

In this phase, we have processed feature extraction. We marked some prominent features of the data base as such:

- **userFollowerCount**: This feature represents the number of followers a user account has. It indicates the popularity or reach of the account in terms of the number of people following it.

- **userFollowingCount**: This feature indicates the number of accounts that the user is following. It gives an idea of the user's activity and interests on the platform.

- **userBiographyLength**: This feature represents the length of the user's biography or profile description. It can provide information about the user's self-description, interests, or promotional content.

- **userMediaCount:** This feature indicates the number of media items, such as photos or videos, uploaded by the user. It reflects the user's engagement with content creation on Instagram.

- **userHasProfilPic**: This binary feature indicates whether the user account has a profile picture. Having a profile picture can be considered a sign of authenticity and active engagement on the platform.

- **userIsPrivate**: This binary feature indicates whether the user's account is set to private. Private accounts restrict access to their content, indicating a more personal or exclusive nature of the account.

- **usernameDigitCount**: This feature represents the number of digits present in the username. It can Proposed work provide insights into naming patterns, such as the use of numbers for uniqueness or anonymity.

- **usernameLength**: This feature indicates the length of the username. It can be a potential indicator of the account's authenticity, as longer, more elaborate usernames might be less common for fake or automated accounts.

- **isFake**: This binary label serves as the target variable and indicates whether the account is classified as fake (1) or not fake (0). It is the ground truth information used for training and evaluating detection models.

3.3 Implementation of various Machine Learning (ML) Approaches for Fake Account Detection

Automated accounts often engage in activities that are not genuine or lack natural authenticity. These activities are commonly referred to as fake engagements. Fake engagements refer to artificial interactions or actions carried out by automated systems, bots, or other means with the intention of creating a false impression of popularity, influence, or engagement. Fake engagement refers to artificial or fraudulent activities aimed at inflating the engagement metrics of social media content, such as likes, comments, shares, and followers. It involves the use of automated systems, bots, or paid services to generate fake interactions, giving the impression of popularity and influence.

Fake engagement can take various forms, including:

- Fake Likes: Automated systems or fake accounts are used to generate likes on social media posts artificially. These likes often come from accounts that are created solely for this purpose.

- Fake Comments: Similar to fake likes, automated systems or bots are employed to leave generic or irrelevant comments on posts to make them appear more engaging.

- Fake Shares: Automated systems or fake accounts are used to share content, giving the illusion of widespread interest and popularity.

- Fake Followers: Services offering fake followers provide users with a large number of followers for a fee. These followers are typically inactive or low-quality accounts created to boost follower counts.

4. Evaluation of Features

1) userFollowerCount: An analysis of the fake engagement done on the userFollowerCount accounts suggests that high follower count have fake engagement (Fig. 1). While the orange suggests fake engagements, there are 800 datapoints showing the same result.

2) userFollowingCount: An analysis of the user Following account suggests that user with high following tend to have fake engagement. And commonly looking at the userFollowingCount and userFollowerCount suggests that user account with both parameters high means that it has high chance of being automated account. A manhatten plot showing number of accounts exhibiting fake engagement for the userFollowingCount.

3) Username Length

The reason for including the number of digits present in the account username as a feature is because it has been observed that during the generation of fake accounts, some individuals or automated systems create multiple accounts by adding different numbers to the same base name. By considering the count of digits in the username, we can capture this pattern and potentially identify fake accounts that follow this naming convention. For example, if we notice a significant number of accounts with similar usernames but varying digits appended to them, it could indicate a suspicious activity associated with fake account creation.

Including this feature helps in capturing the naming patterns and provides additional information for detecting potentially fake or generated accounts. However, it is important to note that this feature alone may not be conclusive evidence of an account being fake, and it should be used in conjunction with other relevant features and analysis techniques for more accurate detection (Fig. 2 and 3).

4) User Follower Count

Fake accounts tend to have a lower follower count compared to real accounts. This characteristic can be used as an indicator to differentiate between fake and real accounts. One way to visualize and verify this trend is by using a Manhattan plot.

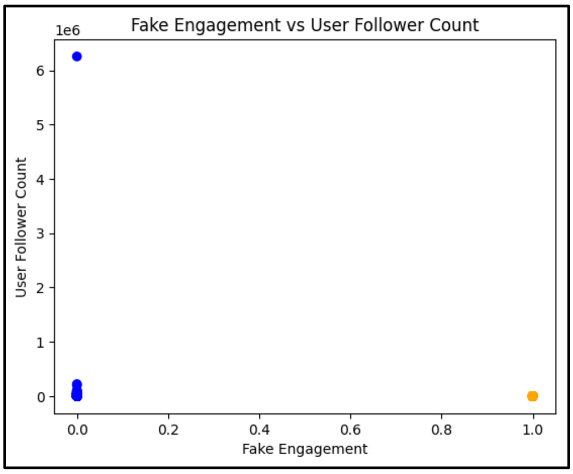


Fig. 1 Manhattan plot for Fake Engagement based on Used Follower Count

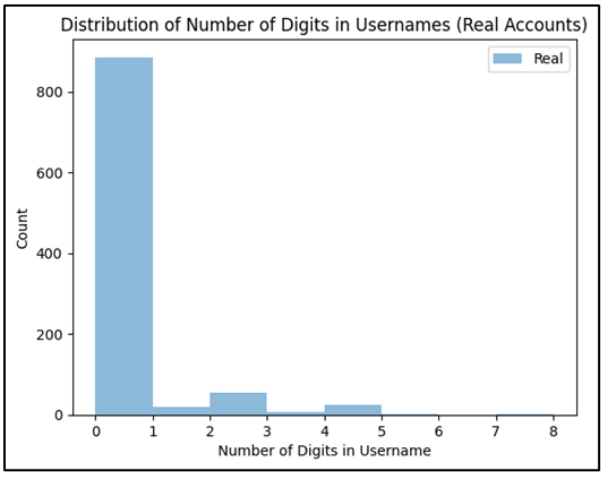


Fig. 2 Real Account Having Number of digits in their username

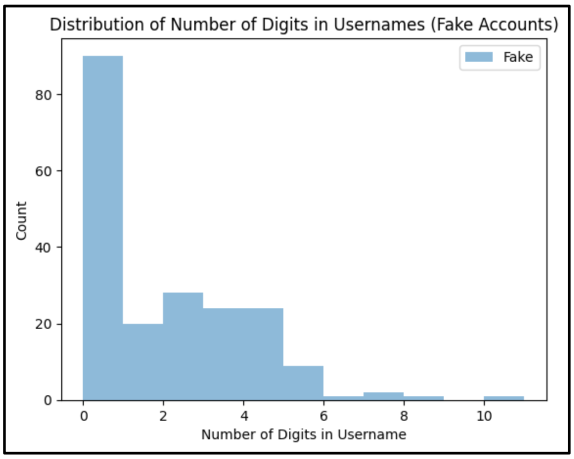


Fig. 3 Fake Account Having Number of digits in their username

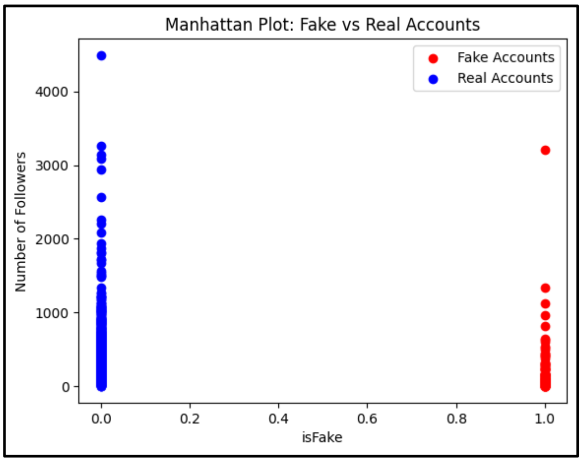


Fig. 4: User Follower Count for Fake vs Real Account

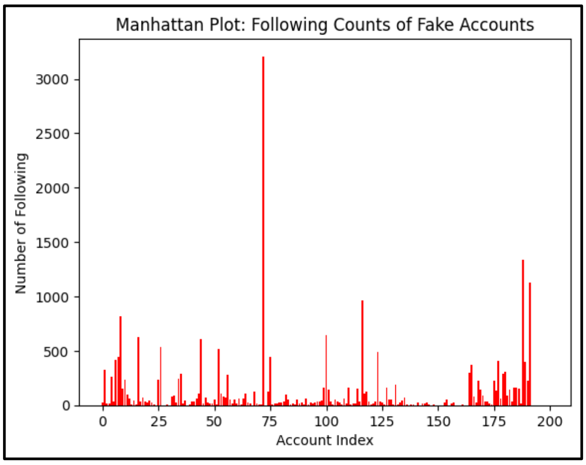


Fig. 5: Manhattan plot to visualize the User Follower Count for Fake vs Real Account

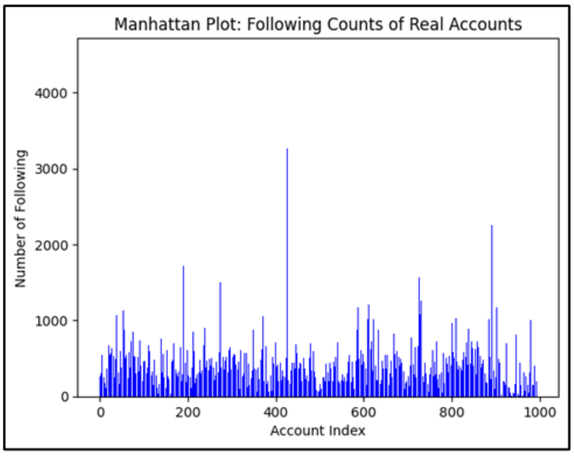


Fig. 6: Manhattan plot to visualize the User Follower Count for Fake vs Real Account

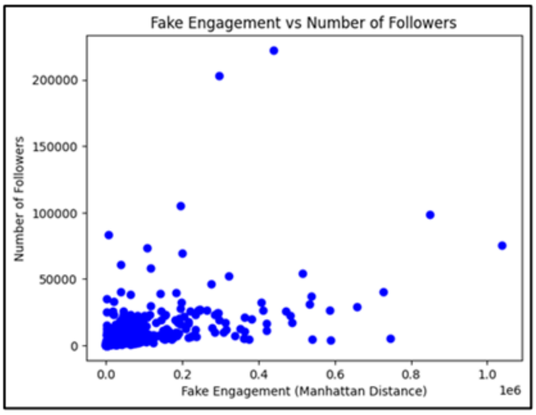
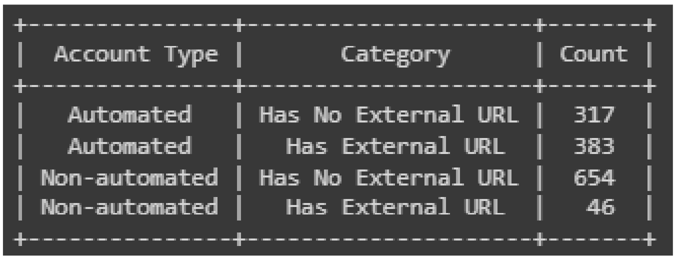
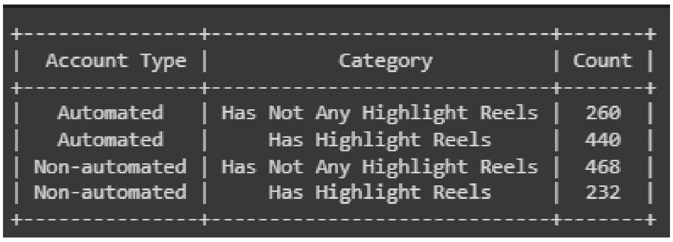


Fig. 7: Fake engagement



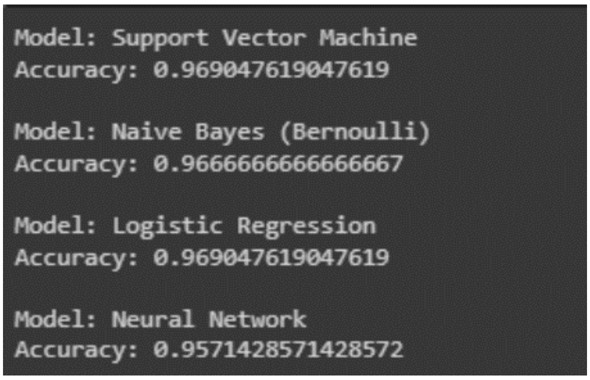
```
+--------------+--------------------+-------+
| Account Type |      Category      | Count |
+--------------+--------------------+-------+
|  Automated   | Has No External URL|  317  |
|  Automated   |  Has External URL  |  383  |
| Non-automated| Has No External URL|  654  |
| Non-automated|  Has External URL  |   46  |
+--------------+--------------------+-------+
```

Fig. 8: Count of Account having External URL in automated and nonautomated account



```
+--------------+---------------------------+-------+
| Account Type |         Category          | Count |
+--------------+---------------------------+-------+
|  Automated   | Has Not Any Highlight Reels |  260  |
|  Automated   |    Has Highlight Reels    |  440  |
| Non-automated| Has Not Any Highlight Reels |  468  |
| Non-automated|    Has Highlight Reels    |  232  |
+--------------+---------------------------+-------+
```

Fig. 9: Count of Account having Highlight URL in automated and nonautomated account



```
Model: Support Vector Machine
Accuracy: 0.969047619047619

Model: Naive Bayes (Bernoulli)
Accuracy: 0.9666666666666667

Model: Logistic Regression
Accuracy: 0.969047619047619

Model: Neural Network
Accuracy: 0.9571428571428572
```

Fig. 10: Different model accuracy on Trained Datase

Results of different machine learning algorithms are as follows:

Support Vector Machine Accuracy: 0.9288702928870293

Naive Bayes (Bernoulli Distribution) Accuracy: 0.9288702928870293

Naive Bayes (Gaussian Distribution) Accuracy: 0.9246861924686193

Logistic Regression Accuracy: 0.9539748953974896

Neural Network Accuracy: 0.9665271966527197

Sample Predictions:

Support Vector Machine: [0 1]

Naive Bayes (Bernoulli): [0 1]

Naive Bayes (Gaussian): [0 1]

Logistic Regression: [0 1]

Neural Network: [0 1]

5. Conclusion

The detection of fake or automated accounts is an important task in various domains, including social media, online platforms, and cybersecurity. Fake accounts can be created for various purposes, such as spreading misinformation, engaging in spamming or phishing activities, manipulating online discussions, or conducting social engineering attacks. Support Vector Machines (SVM), Naive Bayes, Logistic Regression, and Neural Networks are common machine learning algorithms that can be applied to detect automated accounts. These models are trained on labeled datasets that consist of both automated and non-automated account samples. By training on these datasets, the models learn to recognize patterns and make predictions on unseen data. Evaluation of the models is crucial to assess their performance. Accuracy is often used as a metric to measure how well the models classify automated and non-automated accounts. Higher accuracy values indicate better performance in distinguishing between the two types of accounts.

Additionally, libraries such as JSON, Matplotlib, NumPy, and Keras play important roles in supporting the implementation and analysis of the fake detection and automated account detection systems. JSON is used for data storage and exchange, Matplotlib enables data visualization, NumPy provides efficient numerical computing capabilities, and Keras simplifies the development of deep learning models.

References

[1] Alsubaei, Faisal. (2023). Detection of Inappropriate Tweets Linked to Fake Accounts on Twitter. Applied Sciences. 13. 3013. 10.3390/app13053013.

[2] Cresci, Stefano & Pietro, Roberto & Petrocchi, Marinella & Spognardi, Angelo & Tesconi, Maurizio. (2015). Fame for sale: Efficient detection of fake Twitter followers. Decision Support Systems. 80. 10.1016/j.dss.2015.09.003.

[3] Chakraborty, P., Shazan, M. , Nahid, M. , Ahmed, M. and Talukder, P. (2022) Fake Profile Detection Using Machine Learning Techniques. Journal of Computer and Communications, 10, 74-87.

[4] I. Sen, A. Aggarwal, S. Mian, S. Singh, P. Kumaraguru, ve A. Datta, “Worth its weight in likes: Towards detecting fake likes on Instagram,” WebSci, 2018, sf. 205–209.

[5] P. G. Efthimion, S. Payne, ve N. Proferes, “Supervised machine learning bot detection techniques to identify social twitter bots,” SMU Data Science Review, vol. 1, no. 2, p. 5, 2018.

[6] F. C. Akyon and E. Kalfaoglu, “Instagram Fake and Automated Account Detection Insagram Sahte ve Otomatik Hesap Kullanımı Tespiti,” arXiv:1910.03090v1 [cs.IR] 13 Sep 2019].

[7] A. G. Karegowda, A. S. Manjunath, ve M. A. Jayaram, “Comparative study of attribute selection using gain ratio and correlation based feature selection,” 2010.

[8] A. El Azab, A. M. Idrees, M. A. Mahmoud, ve H. Hefny, “Fake account detection in twitter based on minimum weighted feature set,” Int. Sch. Sci. Res. Innov, vol. 10, no. 1, sf. 13–18, 2016.

[9] Y. Li, O. Martinez, X. Chen, Y. Li, ve J. E. Hopcroft, “In a world that counts: Clustering and detecting fake social engagement at scale,” Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2016, sf. 111–120.

[10] Chakraborty, P., Shazan, M. , Nahid, M. , Ahmed, M. and Talukder, P. (2022) Fake Profile Detection Using Machine Learning Techniques. Journal of Computer and Communications, 10, 74-87.