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

Fake Account Detection on Social Media Using Machine Learning and Deep Learning

[1] PARUCHURI SRIVALLI , [2] PODAMEKALA LAHARI , [3] PULETIPALLI SAFEENA
[4] MRS.VIDHYA , [5] MR J.JAYAPRAKASH , [6] MRS.CHINCHU NAIR
[1] [2] [3] Students, [4] Assistant Professor, [5] Professor, [6] Assistant Professor Paruchurisrivalli29@gmail.com, laharipodamekala@gmail.com, psafeena16@gmail.com Dr. MGR Educational and Research Institute, Maduravoyal , Chennai 600095, **TN**

ABSTRACT:

Social networking websites have become an essential part of life, making it easy for individuals to stay connected and exchange information. They provide numerous features, including the ability to chat with others, share news, plan events, and many more. But with the increasing number of users and the volume of personal information, the bad guys have also seen opportunities to exploit these networks. They exploit security loopholes to pilfer personal data, propagate false information, and partake in other malicious activities.

As a result, researchers have been focusing on developing effective methods to detect suspicious activities and identify fake accounts. While some features of social media accounts can help with these efforts, they may sometimes have little to no effect, or even negatively impact the results. Additionally, relying on standalone classification algorithms doesn't always yield optimal outcomes.

This paper suggests using the Decision Tree algorithm to effectively detect fake Instagram accounts by employing four feature selection and dimensionality reduction techniques. In previous research, algorithms such as Decision Trees, Random Forest, Logistic Regression, and Convolutional Neural Networks (CNN) were explored for classification. Among these, CNNs performed exceptionally well, accurately identifying fake accounts and producing satisfying results. Given their high performance, we applied CNNs for Instagram account classification in our study. With deep learning offering various types of neural networks, CNNs have proven to be the most effective for this type of task.

KEYWORDS: Decision Tree, Random Forest, Logistic Regression, and CNN (convolution neural network)

1. INTRODUCTION

Online Social Networks (OSNs) are now an integral aspect of contemporary life, with millions of users interacting on social media websites such as Instagram, Facebook, Twitter, and more. As much as the sites enable interaction and connectivity, they also suffer from security threats, including the spread of impersonation accounts. These fraudulent accounts are frequently employed for nefarious activities such as spam, phishing, and the dissemination of disinformation, posing serious threats to both users and the platforms themselves. Consequently, identifying fake accounts has emerged as a top priority in social media security research.

Social networking sites like Facebook, Twitter, LinkedIn, and Google+ are very popular these days, offering users the means to communicate, exchange information, organize activities, and even operate e-businesses. In 2012, Facebook saw rampant abuse of its platform, with some users posting false news, hate speech, and sensational content. Yet, the vast amount of data created by OSNs has also generated interest among researchers for its value in data mining, analysis of user behavior, and anomaly detection.

For instance, Facebook has more than 2.2 billion monthly active users and 1.4 billion daily active users with a year-on-year growth rate of more than 11 percent. In the second quarter of 2018, the company's revenue stood at $13.2 billion, with advertisements alone generating $13 billion in revenue.
However, even with its enormous user base, Facebook has also been plagued by fake accounts. In 2015, the firm announced that roughly 14 million of its monthly active users were fake accounts created in contravention of its terms of service. Facebook released a report in early 2018 detailing its internal policies to fight unwanted content, such as graphic violence, hate speech, and fake accounts. From October 2017 to March 2018, Facebook took down 837 million spam posts and disabled around 583 million fake accounts. Even with these attempts, estimates place approximately 88 million fake accounts on the platform, which create problems for advertisers, developers, and content creators who depend on proper user metrics. Fake accounts also affect financial institutions. In the United States, banks have started checking the Twitter and Facebook profiles of loan applicants, which makes fake accounts a problem for the financial industry as well.

2.PROBLEM STATEMENT

The problem of classifying online social network accounts, such as Instagram, as real or fake with better accuracy involves analyzing various features and behaviors associated with the account.

3.LITERATURE SURVEY

[1] Political ads on Facebook are the newest trend of campaign activities that have been widely used in national and local elections across the globe.Current devices offer valuable insights into trends within ongoing campaigns. This paper focuses on the reasons for Facebook advertising's popularity and explores the methodological and regulatory challenges it presents for academic researchers and electoral authorities.

[2] One way to stop a "Sybil attack" is to get identities certified by a trusted agency. Because of resource level parity, and coordinated behavior of entities, it is not possible to have Sybil attacks. The author proves that in the absence of such a logically centralized authority, there will always be Sybil attacks.

[3] Online social networks (OSNs) form a vital component of today's web. Famous people like politicians and activists use social media to speak to millions of people in a single go. Sadly, people can manipulate OSNs to push a fake grassroots campaign to spread misinformation and propaganda. A targeted OSN is infiltrated on a large scale to initialize such campaigns. This paper explains how vulnerable OSNs are to large-scale infiltration by social bots (computer programs that control OSN accounts and mimic real users).

[5] we assess the susceptibility of OSNs to large-scale social bot infiltration: computer programs that manage OSN accounts and simulate actual users. We employed a conventional web-based botnet architecture and developed a Social bot Network a set of adaptive social bots that are managed in a command-and-control paradigm. We ran such an SBN on Facebook---a 750 million user OSN for approximately 8 weeks. We gathered data concerning the behavior of users following a mass-scale infiltration when social bots were employed to engage with a mass of Facebook users.

More research in this area, particularly in developing more robust and accurate machine-learning models. OverallThe structured literature review offers useful information about existing research on the use of machine learning for Instagram fake account detection and some potential areas for future development and research.

[6] The study delves into various algorithms and techniques, including supervised learning methods like Support Vector Machines (SVM) and Random Forest, unsupervised learning methods like K-Means and Hierarchical Clustering, and deep learning methods like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).

[7] The review highlights the strengths of each approach, such as the ability of supervised learning methods to detect fake accounts with high accuracy, and the limitations, such as the requirement of labeled data. The study also discusses the challenges faced in fake account detection research, including the evolving nature of fake accounts, the scarcity of labeled data, and the need for more robust and generalizable models. Furthermore, the review identifies future directions for research, including the exploration of new machine learning algorithms, the development of more advanced feature extraction techniques, and the investigation of transfer learning and domain adaptation methods.

4.EXISTING SYSTEM

Since the domain of machine learning is growing at a very rapid pace, computational techniques are nowadays categorized into traditional techniques and machine learning-based techniques. Here, studies on sentiment analysis that are relevant and comparisons of how machine learning algorithms, i.e., Decision Trees, Random Forest, and Logistic Regression, offer betterment compared to traditional techniques are explained.
The current method in this project takes a structured flow with common sentiment analysis methods being used for development. Nevertheless, it demands heavy. Memory resources and results outputted might not always be precise Machine learning algorithms, unlike traditional methods, have the ability to learn through large datasets and adapt to varying patterns in data, which results in more dependable and accurate results. This makes them especially useful in applications such as sentiment analysis, where subtleties in text could be difficult to interpret for less advanced, rule-based systems.

5.PROPOSED SYSTEM

We suggest an application based on a robust algorithm, e.g., Convolutional Neural Networks (CNNs), to break the barrier of conventional and current approaches. The aim of this research is to create an efficient and speedy method for precise sentiment detection. For constructing this system, we utilized the CNN algorithm in a Python setup, which improves efficiency and performance.

Traditional methods for detecting fake accounts have typically relied on rule-based systems and heuristics, which struggle to capture the complex behaviors associated with fake accounts. These approaches often fail to scale with the increasing volume and complexity of data on Online Social Networks (OSNs).

However, recent advancements in machine learning and deep learning have opened up new, more effective ways to detect anomalies in user behavior and account profiles, improving the accuracy of fake account detection.

we propose a hybrid approach that combines traditional ML algorithms with deep learning techniques to tackle the challenges of detecting fake accounts on platforms like Instagram. We apply four feature selection and dimensionality reduction methods to enhance data quality before evaluating several classification algorithms. This research aims to show that integrating deep learning models, particularly CNNs, leads to superior accuracy and efficiency in identifying fake accounts.

6.METHODOLOGY

Dataset and Features

For this analysis, we trained our model using labeled Instagram account data, for which the labels were obtained by Instagram's Security and Trust and Safety teams. Our process starts with choosing a clustering method for accounts. For this, we clustered Instagram accounts by registration IP Addresses and registration dates (Pacific Time). This was the chosen method because it enabled us to have a high volume of manually labeled data that we could use for analysis.The features selected for this study were:

Account Age: The time since the account was created.

Number of Followers/Following: The number of followers in account

Engagement Rate: Number of interactions (likes, comments, shares) relative to the number of posts.

Profile Information: Completeness of the profile (bio, image, website links).

Posting Frequency: Number of posts per week/month.

User Behaviour Patterns: Irregular activity such as bot-like behaviors (e.g., rapid increases in followers).

We implemented four different classification algorithms to evaluate their performance in detecting fake accounts:

Decision Tree (DT)

A decision tree model was trained to classify accounts as fake or real based on feature values. This model is well-regarded for its interpretability and ease of use, making it a popular choice for tasks where understanding the decision-making process is important.

Gini Impurity Formula:

$$Gini = 1 - \sum p^2$$

where pi is the probability of class iii

Entropy Formula:

$$Entropy = -\sum p_i \log_2 p_i$$

Information Gain:

$$IG = \overset{child}{Entropy} - \sum \left(\frac{N}{N_{parent}} \times Entropy(child) \right)$$

When a decision tree model predicts a value of 0 for an Instagram the account, suggests that the account is likely to be fake. This means the account's features and behavior do not align with those of a legitimate user. Essentially, the account exhibits traits commonly found in spam, bots, or impersonation accounts. Therefore, the model flags the account for further investigation or potential removal, helping to ensure the platform remains secure and trustworthy.

On the other hand, if the decision tree model predicts a value of 1, it indicates that the account is likely real. This suggests that the account aligns with the typical characteristics of a genuine user, exhibiting behaviors consistent with those of an authentic person. As a result, the model concludes that the account is most likely legitimate, and can be trusted, supporting the platform's efforts to maintain a safe and reliable environment for users.

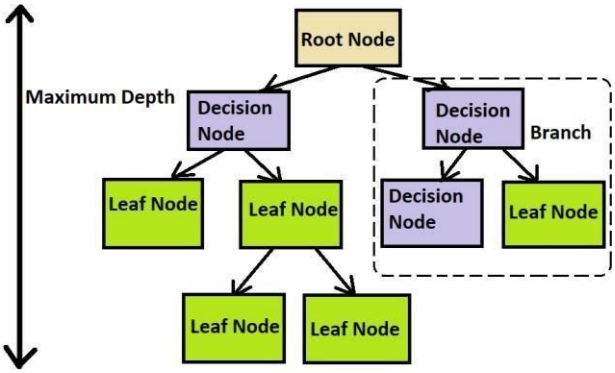


Fig 1:Decision Tree Model

Random Forest (RF)

Random Forest is an ensemble method that enhances the performance of decision trees by averaging predictions from multiple trees. This approach helps improve the model's generalization, making it more robust and accurate compared to using a single decision tree.

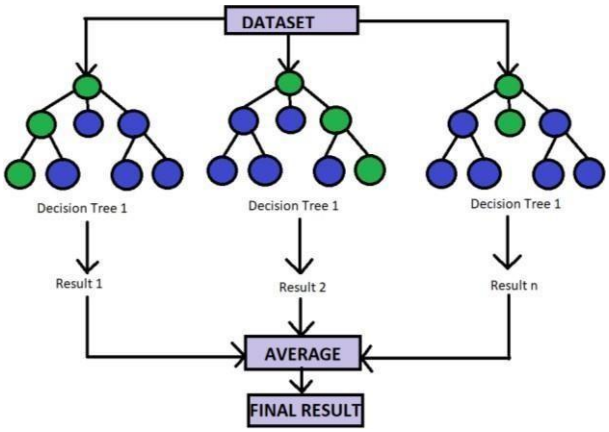


Fig 2: Random Forest Model

Majority Voting (for classification):

$$y\hat{} = \text{mode}(y\hat{}_1, y\hat{}_2, \ldots, y\hat{}_n)$$

Averaging (for regression):

$$y = {}_n\sum^{1n}{}_{i=1} y_i$$

The mode function plays a key role in determining the predicted class label by identifying the most frequent value among the predictions made by multiple decision tree models. In the case of classifying Instagram accounts as fake or genuine, the mode function will return either 0 (fake) or 1 (genuine). If the majority of decision trees classify an account as fake (0), the mode function will output 0, designating the account as fake. On the other hand, if most models predict the account as genuine (1), the mode function will output 1, classifying the account as genuine. This method ensures that the final classification reflects the majority opinion from the decision tree models, leading to more reliable results.

Logistic Regression (LR)

Logistic regression, a linear model commonly used for binary classification, was employed as a baseline classifier to evaluate the performance of more complex models. The logistic regression equation is utilized to estimate the probability of an account being fake:

The sigmoid function has an output range between 0 and 1,

$$B = B - \alpha\frac{\partial Loss}{\partial B}$$

where

α alpha is the learning rate.

meaning that its result, p, will always fall within this interval. In the context of binary classification, such as identifying whether an Instagram account is fake or genuine, the value of p represents the probability that the account is genuine. A value of p close to 0 suggests a low likelihood of the account being genuine, while a value close to 1 indicates a high probability that the account is genuine.

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Where

$z = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$
w are the model's learned weights
x is the input features
b is the bias term

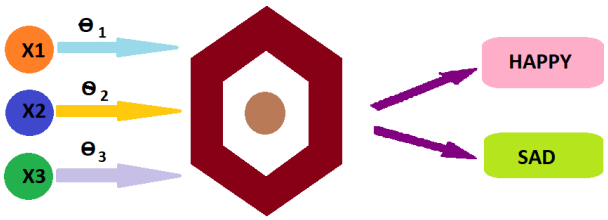
Prediction Rule:

$$y = \sigma(z) = \frac{1}{1+e^{-(w\cdot x+b)}}$$

If y≥0.5, classify as 1; otherwise, classify as 0.

A matrix is used to store the feature values for each sample (account), where each row corresponds to a different sample, and each column represents a specific feature. The features may include:

- Number of followers
- Number of posts
- Engagement rate
- Account age
- Number of mutual connections
- Profile completeness
- Bio length



Inputs:X1,X2,X3||Weights:01,02,03 ||Outputs:Happy,Sad

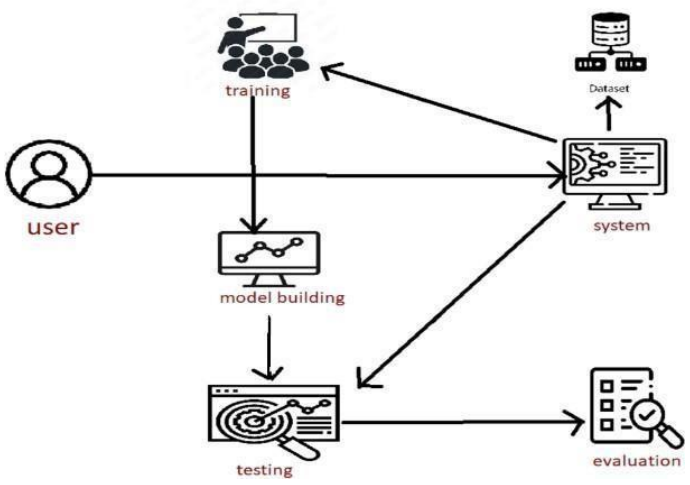


Fig 3: Logistic Regression Model

Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs), a type of deep learning model, are especially effective at identifying patterns in data. In this case, CNNs were used to analyze the features extracted from Instagram profiles and predict the likelihood of an account being fake.

$$Z = W \cdot X + B$$
$$A = f(Z)$$

where

f(Z)) is the activation function (ReLU, Sigmoid, etc.).

Loss Function (Binary Crossentropy):

$$Loss = -\frac{1}{m}\sum (y \log(\hat{y}) + (1-y)\log(1-\hat{y}))$$

Backpropagation & Gradient Descent:

$$W = W - \alpha \frac{\partial Loss}{\partial W}$$

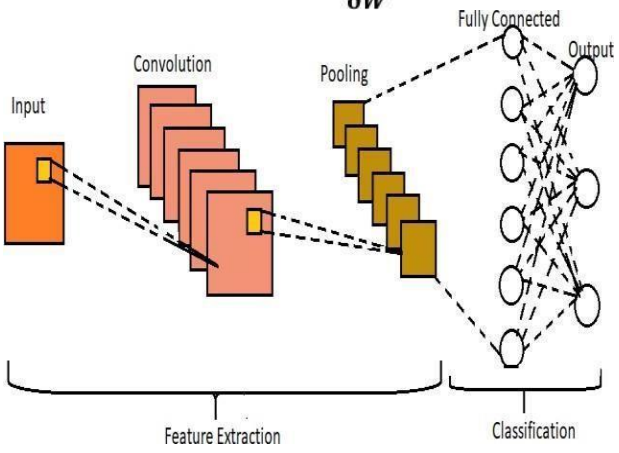


Fig 4: Convolutional Neural Networks Model

7.ARCHITECTURE

8.RESULT

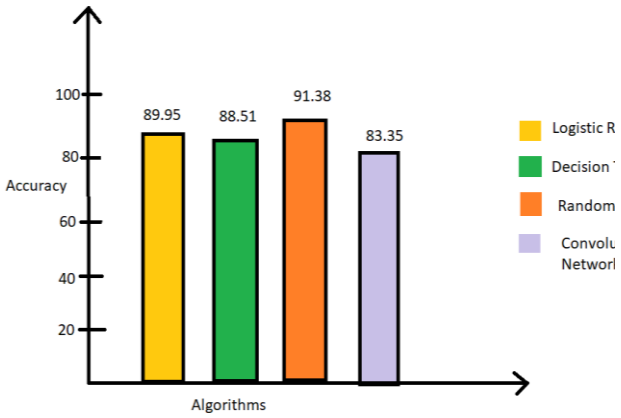


Fig 8.1: Proposed Accuracy

9.CONCLUSION

we introduced a hybrid approach for detecting fake accounts on social media platforms by combining traditional machine learning methods with advanced deep learning techniques. Our findings show that Convolutional Neural Networks (CNNs) are particularly effective at identifying fake accounts, outperforming Decision Trees, Random Forests, and Logistic Regression in terms of accuracy and reliability. Future research could focus on optimizing the model further and adapting it for use on other social media platforms like Twitter and Facebook. Moreover, implementing real-time deployment and integrating the model with OSNs could significantly improve the detection and prevention of fake accounts and other malicious activities.

10.FUTURE ENHANCEMENT

Preliminary experiments have yielded promising results, with accuracy scores of 89.95% (Logistic Regression), 88.51% (Decision Tree), 91.38% (Random Forest), and 82.35% (CNN). We aim to improve these results through future refinements, targeting 92-95% accuracy for Logistic Regression, 90- 93% for Decision Tree, 93-95% for Random Forest, and 85-90% for CNN. These algorithms can be applied to other social media platforms to detect fake accounts, and exploring new techniques can further enhance accuracy, providing a safer user experience.

REFERENCES

[1]Kaur and S. Singh, 2018"A survey of data mining and social network analysis based anomaly detection techniques," Egyptian Informatics Journal, vol. 17

[2]X., David, MF., Theodore, H. 2017"Detecting Clusters of Fake Accounts in Online Social Networks. In: 8th ACM Workshop on Artificial Intelligence and Security", vol. 13

[3]Y. Boshmaf, M. Ripeanu, K. Beznosov, and E. Santos-Neto, 2023"Thwarting fake accounts by predicting their victims," in Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security. ACM, ", vol.7

[4]S.-T. Sun, Y. Boshmaf, K. Hawkey, and K. Beznosov, 2021"A billion keys, but few locks: the crisis of web single sign-on," in Proceedings of the New Security Paradigms Workshop". ACM, vol 8

[5]Breuer, A., Eilat, R., & Weinsberg, U. (2020, April)." Friend or faux: Graph-based early detection of fake accounts on social networks. In Proceedings of The Web Conference ", vol 11

[6]S. Fong, Y. Zhuang, and J. He,2022"Not every friend on a social network can be trusted: Classifying imposters using decision trees," in Future Generation Communication Technology (FGCT), International Conference on. IEEE, vol.12

[7]Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, 2019"The social bot network: when bots socialize for fame and money," in Proceedings of the 27th Annual Computer Security Applications Conference". ACM, vol13

[8]J. Ratkiewicz, M. Conover, M. Meiss, B. Gonc¸alves, S. Patil 2018" A. Flammini, and F. Menczer, "Truthy: mapping the spread of astroturf in microblog streams," vol 827