[1]Sumitra Menaria
[2]Dr. Viral H. Borisagar

Machine Learning Based Detection of Fake Accounts on Online Social Media: A Feature Engineering

Journal of Electrical Systems

Abstract: - The prevalence of online social networks (OSNs) among today's youth has led to a close connection between social activities and these websites. However, the growth of OSNs and the user data on these platforms has also attracted attackers and imposters who engage in detrimental activities like stealing user data, spreading false information, and creating fake accounts. Individual account analysis and coordinated activity detection are two types of techniques that scholars have developed for identifying bogus accounts and suspicious behavior to address these issues. The article proposes a feature engineering method to efficiently detect fraudulent social media profiles and bots by using dimension reduction, feature selection, and data preprocessing techniques. Additionally, support vector machines, neural networks, Ada Boost Classifiers, random forests, and decision trees are used as machine learning categorization methods. The project's goals are to make fake social media accounts easier to spot and shed light on the role that false identities play in sophisticated persistent threats**..**

Keywords: Fake profile, Feature Engineering, Machine Learning, Classification, Pre-processing

1. Introduction

Websites and apps that enable users to produce and share content or engage in social networking are referred to as online social media (OSM). Facebook, Twitter, Instagram, LinkedIn, and Snapchat are a few of the most widely used social media websites. Users can interact and converse with one another through these platforms, exchange information, and establish or join groups based on shared affiliations or hobbies. Keeping in contact with friends and family to promoting and advertising goods and services, online social media has become a crucial component of contemporary communication [1]. Online social media has become increasingly popular in recent years for several reasons, including:

1. Connectivity: Online social media provides a way for people to connect with each other regardless of their location. With the help of social media platforms, people can keep in touch with friends and family members who live far away [2]..

2. Convenience: Social media is easy to access and use. People can log in to their accounts and check updates on their phones, tablets, and computers anytime and from anywhere [2].

3. Entertainment: Social media platforms offer various types of content, such as photos, videos, and memes, which can be entertaining and enjoyable to users [3].

4. Information sharing: Social media is a platform where people can share news, events, and information with others. This feature of social media has made it a popular source of news and information for many people [4].

5. Business Promotion: Social media is used by many businesses for promotion, brand building, and customer engagement. It provides a cost-effective way for businesses to reach out to their target audience and build their brand [5].

Social media accounts that are established with the purpose to deceive or incorrectly portray a person or business are referred to as fake accounts. These accounts may be used for disseminating false information, phishing, and carrying out social engineering assaults, among other things [6]. A significant effect on society has resulted from the rise in phoney accounts on social media in recent years, including the dissemination of false information, the swaying of public opinion, and the decline in confidence in social media platforms. Fake accounts are created for a variety of reasons, such as:

[1] *Ph.D. Research Scholar, Computer Science & Engineering, Gujarat Technological University, Ahmedabad, Gujarat, India
[2] Assistant Professor, Computer Engineering, Vishwakarma Government Engineering College, Chandkheda, Ahmedabad, Gujarat, India.
Copyright © JES 2024 on-line : journal.esrgroups.org

1. Spamming: Fake accounts are created to spam users with unwanted messages and links.
2. Impersonation: Fake accounts are created to impersonate a real person or brand for malicious purposes.
3. Trolling: Fake accounts are created to provoke others and cause conflict.
4. Scamming: Fake accounts are created to scam people out of money or personal information.
5. Propaganda: Fake accounts are created to spread false information or propaganda.

Fake accounts are a significant problem on social media sites because they can hurt peo- ple, businesses, and entire communities. They have the power to disseminate untruths, cause commotion, and undermine public confidence in social media platforms. Social media firms have taken action to detect false accounts by putting in place a number of measures, including verification procedures, automated detection systems, and guide- lines that forbid the creation of phoney accounts [5]. The increasing number of fake accounts on online social media can lead to several issues, including: Misinformation, Cyberbullying, Identity theft, Spam and phishing, Reputation damage, Ad fraud and Political manipulation.

These issues highlight the need for effective methods to detect and remove fake accounts from online social media platforms.

2. Literature Survey

The research problem is the need to develop effective methods for detecting fake accounts on social media to prevent their negative impact on society. The objective of the study is to evaluate the performance of machine learning classification algorithms such as support vector machine, XGB classifier, Ada Boost Classifier, random forest, and decision tree in detecting fake accounts on social media.

Creating a dataset of social media profiles and their associated labels is part of the study approach. (i.e., fake or real). After preprocessing the data to eliminate noise and unimportant characteristics, it is divided into training, validation, and testing groups. Using measures like accuracy, precision, recall, and F1 score, the machine learning classification algorithms are tested on the validation set after being taught on the training set. The generalization performance of the top-performing algorithm is then measured using the trial set. The contribution of the research is to provide a comparative analysis of the performance of different machine learning classification algorithms in detecting fake accounts on social media. This can help researchers and practitioners in the field to identify the most effective algorithms for detecting fake accounts and develop more accurate and efficient methods for preventing their negative impact on society. Additionally, the research can contribute to the development of better social media policies and practices for detecting and removing fake accounts. This can enhance the user experience on social media platforms and improve the credibility and trustworthiness of online content. To achieve the research objectives, the study will focus on the following research questions:

1. Key features and characteristics of fake accounts on social media

2. Various Machine learning classification algorithm and their performance in detect- ing fake accounts on social media

3. Implications of the findings for social media users and stakeholders

Already, exhaustive studies have been done to detect fake profiles on OSNs different methods can be used. here we have presented literature review of few papers.

The goal of Wang et al.'s systematic study [7] was to locate and assess the most recent machine learning techniques for spotting fake social media profiles. The major- ity of studies used supervised learning algorithms and had a Twitter data emphasis, according to their analysis of 30 studies released between 2012 and 2018. They dis- covered a number of characteristics that are useful in spotting fake identities, such as user profiles, network structures, temporal trends, and content. In order to increase the precision and effectiveness of fake account identification on social media, the study emphasises the need for using a wide variety of features, taking into consideration the temporal aspects of account activity, and creating more complex machine learning models.

A machine learning and data mining-based strategy was put forth by Singh et al.

[3] to identify phoney social media profiles. They gathered information from Twitter, Facebook, and Instagram and trained various classifiers using a variety of character- istics including user profiles, sharing habits, network properties, and content. They used accuracy, precision, recall, and F1 score measures to assess the effectiveness of their method, and they were successful in identifying fake Twitter profiles with a high accuracy of 96%. Their research emphasizes the value of combining characteristics and machine learning methods to identify fake profiles on various social media sites.

Dey et al.[8] addressed different machine learning-based methods for spotting fake social media profiles in this review paper. They offered a thorough study of the prior literature that covered a variety of methodologies, including clustering, categorization, and abnormality detection. The review also found areas for future research and emphasized the shortcomings of the methods currently in use. The writers came to the conclusion that while machine learning-based methods have demonstrated promise in identifying fake accounts, more study is required to enhance their performance and reliability.

The research by Chen, J. et al. [9] used a collection of 6,825 Weibo profiles to develop a machine learning-based method for identifying fake identities on social media. The authors evaluated their suggested model thoroughly using six feature sets and three distinct categorization methods. The outcomes demonstrated that the suggested strategy outperformed other cutting-edge techniques, achieving high accuracy, precision, memory, and F1 score. The research proved how well different feature sets and categorization methods worked to improve the model's performance.

In their article, Zhang, X. et al. [2] suggested a machine learning method for identifying phoney social media accounts. Users' behavior, network structure, and temporal trends were among the characteristics that were extracted from a collection that contained both fake and genuine profiles. They next evaluated the effectiveness of various machine learning models and discovered that the gradient boosting tree method produced the finest outcomes, with an accuracy and F1-score of 0.964 and 0.963, respectively. The writers claim that a variety of uses, including the detection of malicious internet activity and the detection of fake news, can benefit from their methodology.

In their research, Seneviratne, N. M. et al.[10] suggested using machine learning to identify phoney social media profiles. To teach the algorithm, the researchers used traits like user behavior trends, metadata, and network characteristics. They used measures like accuracy, precision, recall, and F1 score to assess the model's success. According to the research, the suggested technique is more effective than current ones at spotting fake identities. The writers also emphasized the significance of ongoing model tracking and updating in order to accommodate the changing tactics used by fake account makers.

In their article, Lee, J. et al. [11] suggested a novel method that combined network analysis and machine learning to identify fake social media accounts. The suggested

approach comprises of two stages: feature extraction based on network analysis and categorization based on machine learning. The writers used network measures like degree centrality and between ness centrality to extract characteristics from the data they had gathered from Twitter. They then classified the profiles as fake or real using a number of machine learning methods, including SVM, Random Forest, and Ada-Boost. The proposed method achieved high accuracy in detecting fake accounts and outperformed existing methods. The study suggests that integrating network analysis and machine learning can be an effective approach to detecting fake accounts on social media.

A machine learning method was put forth by Ramanathan et al. [12] to identify phoney social media profiles. They gathered information from Twitter and created a dataset with real and phoney profiles. They taught machine learning models like random forest and support vector machine to categories accounts as legitimate or fake, using feature engineering to extract pertinent characteristics like amount of friends, retweets, and likes. They found that their method can successfully identify fake social media profiles after achieving high precision in their tests.

In their research, Karunarathne et al. [4] suggest using machine learning to identify phoney social media profiles. The suggested method achieves high precision in identifying fake accounts by combining feature engineering, feature selection, and various categorization models. The research measures the effectiveness of

the suggested method using real-world information from Twitter and Facebook. The findings demonstrate that the suggested strategy beats current cutting-edge techniques in terms of accuracy, precision, recall, and F1-score. According to the study's findings, the suggested method is efficient at identifying phoney profiles on social media and can be helpful in reducing the dissemination of false information and fake news.

The study by Pan, J. et al. [13] suggests a clever method for spotting fraudulent social media profiles by utilizing machine learning techniques. The suggested method makes use of a variety of data sources, including user profiles, social network architecture, and user activity, to build an extensive feature set for fake account identification. For the categorization of fake accounts, a number of machine learning methods are used, including Random Forest, Support Vector Machine, and Neural Network. Utilizing a dataset compiled from Twitter, the suggested method is assessed, and the findings of the experiments demonstrate that it has a high level of precision in identifying phone accounts. The suggested method can successfully identify fake profiles on social media and can be helpful for preserving the integrity of online social networks, according to the paper's conclusion.

Although significant progress has been made in detecting fake accounts on social media using machine learning algorithms, there are still several gaps in the existing literature, including:

1. Lack of standard datasets: There is a lack of standard datasets for evaluating the performance of different fake account detection methods. This makes it difficult to compare the results of different studies [4], [2].

2. Limited generalizability: Many existing studies use data from a specific social media platform or a specific geographic region. Thus, the generalizability of the results to other platforms and regions is limited [4], [11], [12].

3. Need for explain ability: Many existing studies rely on complex machine learning models, which can be difficult to interpret. There is a need for more explainable models to provide insights into how the models make decisions [7], [8], [11].

4. Limited focus on the impact of fake accounts: While many studies focus on detecting fake accounts, there is a need for research that explores the impact of fake accounts on social media, including the spread of misinformation and propaganda [13], [14], [4].

5. Need for real-time detection: Many existing studies focus on offline detection of fake accounts. However, there is a need for real-time detection of fake accounts to prevent their spread and impact on social media [13], [12], [14].

3. Proposed Approach

The development of a fake account detection system for social media involves a multi-step methodology designed to ensure accuracy and reliability. The process begins with the selection of a suitable dataset, which is crucial for training the model. This dataset may be sourced from publicly available datasets, direct data collection from social media platforms using APIs, or through the generation of synthetic data to supplement existing resources. Once the dataset is acquired, extensive data pre-processing is undertaken. This involves data cleaning to remove noise and irrelevant information, feature extraction to identify key indicators of fake accounts such as account age, activity patterns, and follower/following ratios, normalization to standardize numerical features, and data transformation techniques like one-hot encoding for categorical data.

Choosing the right machine learning algorithms is the next critical step. Both supervised learning algorithms, such as logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks, and unsupervised learning algorithms, like k-means and DBSCAN, are considered based on the problem's complexity and data nature. Ensemble methods, which combine multiple models to enhance performance, are also evaluated. The effectiveness of the chosen algorithms is assessed using a variety of evaluation metrics, including accuracy, precision, recall (sensitivity), the F1 score, and the AUC-ROC curve, each providing insights into the model's ability to distinguish between real and fake accounts.

Experimental design plays a pivotal role in validating the detection system. This includes the train-test split to evaluate performance on unseen data, k-fold cross-validation to ensure model robustness and mitigate overfitting, and hyperparameter tuning through grid search or random search to optimize model performance. Comparative analysis of different algorithms and configurations is conducted to identify the best-performing model. Finally, real-world testing is implemented to monitor the model's performance in a live environment, allowing for adjustments based on feedback and new data. Through this comprehensive and systematic approach, a robust fake account detection system for social media can be developed, capable of accurately identifying malicious accounts as shown in Fig 1.

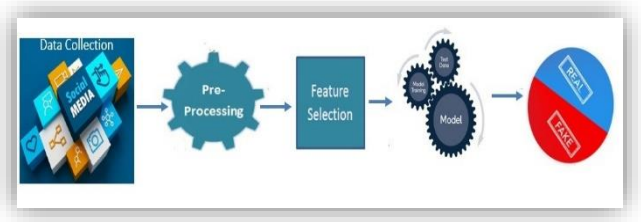

Fig. 1 Analysis Carried Out on Social Media Profiles

A. Dataset:

A publicly available dataset containing 3474 real users and 3351 phony users was utilized for this study. This dataset encompasses a diverse range of variables associated with social media profiles. These variables include indicators such as the friend-to-follower ratio, the content of tweets, the age of the account, and API usage. Each of these characteristics can serve as potential indicators of fake or suspicious accounts.

For instance, a high friend-to-follower ratio might suggest that an account is attempting to appear more legitimate by following many users, while low engagement or highly repetitive tweet content can be red flags for automated or bot-like behavior. Additionally, the account age can provide insights into legitimacy, with older accounts generally being less likely to be fake. Extensive use of APIs by an account can also hint at automation, which is common in fake accounts.

Conversely, characteristics such as the ratio of bidirectional links (mutual followers) and the presence of API URLs (indicative of legitimate third-party application usage) can signal authentic accounts. Real users are more likely to have balanced mutual follow relationships and may use APIs for genuine purposes like social media management tools.

To facilitate effective model training and evaluation, the dataset was divided into training and testing subsets with a 70:30 split ratio. This approach ensures that the model is trained on a substantial portion of the data (70%) while reserving a separate segment (30%) for robust evaluation. By doing so, the model can be trained to accurately differentiate between real and fake users based on the provided attributes, thereby enhancing the detection accuracy on social media platforms.

This methodology not only helps in refining the model's ability to identify fake accounts but also ensures that the evaluation phase tests the model's performance on unseen data, providing a realistic measure of its efficacy in real-world scenarios. By leveraging this comprehensive dataset and careful data partitioning, the model aims to improve the reliability and effectiveness of fake account detection on social media.

B. Pre-processing and Feature Selection

For training and testing purpose feature selections were performed after proper pre- processing of data, we considered following features:

1. **Friends/followers ratio:** A high Friends/followers ratio could indicate that the account is artificially inflating its followers or engaging in follow-back schemes to boost its numbers.

2. **Retweet and link ratio:** If high amount of account's tweets are retweets and contain links, it could suggest that the account is automating its content or promoting spammy or low-quality content.

3. **Bidirectional link ratio:** Legitimate accounts are more likely to have mutual followers, so a low bidirectional link ratio (i.e., accounts that follow each other) could suggest that the account is not engaging with other users in a genuine way.

4. **API ratio:** The ratio of tweets sent from API (i.e., automated tools) to total number of tweets could suggest that the account is using bots or other automated methods to generate its content.

5. **API URL ratio:** If tweets sent from API also contain URLs at a high ratio, it could suggest that the account is using automated tools to post spammy or low-quality links.

6. **Age of the account:** While not necessarily indicative of spam or suspicious behavior on its own, a very new account (e.g., created within the last few weeks or months) could suggest that the account is being used for spam or other nefarious purposes.

Table 1 Notations Used

Symbol	Description
f	*number of friends*
w	*number of followers*
n	*total number of tweets*
r	*number of retweets*
l	*number of tweets containing links*
a	*age of the account in days*
b	*number of accounts that follow target and are followed back*
t	*total number of tweets*
at	*number of tweets sent using the API*
ut	*number of tweets sent using the API and containing a URL*

Steps for generating vector:

Let X be a vector of features for the account under investigation, represented as follows:

- X[0]: Friends-to-followers ratio
- X[1]: Percentage of tweets that are retweets
- X[2]: Percentage of tweets that contain links
- X[3]: Age of the account in days
- X[4]: Bidirectional link ratio (mutual followers ratio)
- X[5]: API usage ratio
- X[6]: API URL ratio

Define the Classification Label: Let y be the binary classification label for the account:

- y=1: If the account is classified as fake
- y=0: If the account is classified as genuine

Generating Values for X and y:

- Calculate the friends-to-followers ratio for the account and assign it to X[0].
- Determine the percentage of tweets that are retweets and assign this value to X[1].
- Calculate the percentage of tweets that contain links and assign it to X[2].

- Measure the age of the account in days and assign this value to X[3].
- Compute the bidirectional link ratio and assign it to X[4].
- Calculate the API usage ratio and assign it to X[5].
- Determine the API URL ratio and assign it to X[6].
- Based on the predefined classification criteria, assign y as 1 if the account is deemed fake, otherwise assign y as 0.

By following these steps, we can systematically generate the feature vector X and the classification label

y for each account under investigation. This structured approach allows for the effective training and evaluation of a model designed to distinguish between genuine and fake social media accounts.

Let y be the binary classification label for the account under investigation, where 'y=1' if the account is classified as fake and 'y=0' if the account is classified as genuine. Then, we can use the following rules to generate the values for 'X' and 'y':

1. Friends/followers ratio:

- Let *'f'* be the number of friends for the account under investigation.
- Let *'w'* be the number of followers for the account under investigation.
- $X[0] = \begin{cases} 1 & if \quad \frac{f}{w} \geq 50 \\ 0 & Otherwise \end{cases}$.

2. Retweet/link percentage:

- Let *'n'* be the total number of tweets for the account under investigation.
- Let 'r' be the number of tweets that are retweets for the account under investigation
- Let 'l' be the number of tweets that contain links for the account under investigation.
- Define X[1] and X[2] as follows:
- $X[1] = X[2] = \begin{cases} 1 \; if \; \frac{r}{n} > 0.9 \; and \; \frac{l}{n} > 0.9 \\ 0 \qquad\qquad\qquad Otherwise \end{cases}$

3. Age of the account:

- Let 'a' be the age of the account in days.
- Define X[3] as:
- $X[3] = \begin{cases} 1, \; if \; a < 15 \\ 0, \; Otherwise \end{cases}$

4. Bidirectional link ratio:

- Let 'b' be the number of accounts that follow the account under investigation and are followed back by the account under investigation.
- Let 'f ' be the number of friends for the account under investigation.
- Define X[4] as follows:
- $X[4] = \begin{cases} 1, \; if \; \frac{b}{f} < 0.2 \\ 0, \; Otherwise \end{cases}$

5. API ratios:

- Let 't' be the total number of tweets for the account under investigation.

- Let ‘at’ be the number of tweets that were sent using the API for the account under investigation.
- Let ‘ut’ be the number of tweets that were sent using the API and contain a URL for the account under investigation.
- Define X[5] and X[6] as follows
- $X[5] = \begin{cases} 1, & if\ \frac{at}{t} > 0.8 \\ 0, & Otherwise \end{cases}$
- $X[6] = \begin{cases} 1, & if\ \frac{ut}{a} > 0.8 \\ 0, & Otherwise \end{cases}$

Classification label:

Define y as follows:

$$y = \begin{cases} 1, & if\ X[0] = 1\ and\ X[1] = 1 \\ 0, & Otherwise \end{cases}$$

With the feature vector X and label y, we can train various classification models like Support Vector Machines, XGBoost, AdaBoost, Random Forest, and Decision Trees. These models can learn patterns in the features to distinguish between genuine and fake accounts.

1. Feature Vector X: A 7-dimensional vector representing different characteristics of the account.
2. Label y: A binary value indicating if the account is fake (1) or genuine (0).

Equation for Machine Learning:

Given a dataset of m accounts, each represented by a feature vector $X^{(i)}$ and label $y^{(i)}$ for i=1,2,……,m, the objective is to learn a function h:X→y that can predict y for unseen accounts based on their feature vector

X.

For example, a simple logistic regression model would aim to find parameters θ such that:

$$h_\theta(X) = 1/(1+e^{-\theta TX})$$

This results in a feature vector X of length 7 and a binary classification label y. We can then use X and y to train a classification model to predict whether a given social media account is genuine or fake. By combining these features into a feature vector and using them to train a machine learning model, it is possible to automatically identify fake social media accounts with a high degree of accuracy. The advantage of this approach is that it can be applied to large datasets quickly and accurately, without requiring manual review of each individual account. This makes it a more scalable solution for identifying fake accounts on social media, which is an important challenge for the platform to address in order to maintain the trust and safety of its users.

we used the feature vector X with different machine learning algorithms, such as Support Vector Machine Classifier, XGB Classifier, Ada Boost Classifier, Random Forest Classifier, and Decision Tree Classifier, to predict whether a given social media account is genuine or fake with higher accuracy. We can train the model using the training data and evaluate its performance on the testing data to select the best algorithm for the given problem. The study’s conclusions have important ramifications for partners and consumers of social media. For information to remain accurate and users’ protection to be maintained, fake profiles on social media must be identified. The study’s findings can aid social media sites in creating strategies that will effectively stop the proliferation of malicious activity and fake profiles. This could enhance the general user experience and boost users’ confidence in social media

4. Results

Table 2 presents the performance evaluation metrics for the model, which involved the measurement of confusion matrices and accuracy. A confusion matrix is a table that is used to evaluate the performance of a classification model by comparing the predicted class with the actual class of a set of test data. Accuracy, as mentioned, is the ratio of correctly classified instances to the total number of instances. It is a commonly used metric for classification models, and it provides an overall performance measure for the model. However, it may not be the most appropriate metric in cases where the classes are imbalanced.

Table 2 Performance Metrics of Machine Learning Algorithms

Name of Algorithm	TP	FP	TN	FN	Accuracy
Support Vector Machine	1007	50	934	57	94.78%
Ada Boost Classifier	1046	11	981	10	98.97%
Decision Tree Classifier	1040	17	978	13	99.37%
Random Forest Classifier	1053	4	979	12	99.22%
XGB Classifier	1051	6	984	7	99.73%

Precision and recall are two other commonly used metrics for evaluating classification models. Precision is the proportion of true positive instances out of all predicted positive instances, and recall is the proportion of true positive instances out of all actual positive instances. The F1-score, which is the harmonic mean of precision and recall, provides a more balanced evaluation metric than accuracy in cases where the classes are imbalanced. The presented table provides the performance evaluation met- rics for five different algorithms, namely the Support Vector Machine, Ada Boost Classifier, Decision Tree Classifier, Random Forest Classifier, and XGB Classifier. The table presents the number of true positives, false positives, true negatives, and false negatives for each algorithm, along with the accuracy percentage. The second table provides the recall and F1-score for each algorithm. It is evident from the table that all algorithms had high performance metrics, with the XGB Classifier having the high- est recall and F1-score. These metrics can be used to select the best algorithm for the given classification problem.

Figure 2 is a visual representation of the accuracy values presented in Table 2. The figure shows the accuracy values for each of the algorithms, allowing for easy comparison between them. The purpose of this figure is to provide a quick overview of the accuracy performance of each algorithm, as opposed to reading through the table.

Table 3 Performance Metrics of Machine Learning Algorithms

Algorithm	Recall (%)	F1-score (%)
Support Vector Machine	94.6	94.9
Ada Boost Classifier	99.1	99.1
Decision Tree Classifier	98.8	98.6
Random Forest Classifier	98.9	99.2
XGB Classifier	99.3	99.3

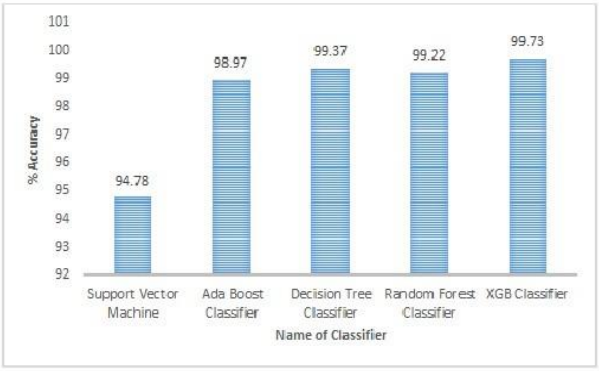

Fig. 2 Accuracy Comparison of various Machine Learning Algorithms

On the other hand, Figure 3 represents the recall and F1-score comparison for the algorithms. The figure shows the recall and F1-score values for each of the algorithms, allowing for easy comparison between them. The purpose of this figure is to provide a quick overview of the recall and F1-score performance of each algorithm, as opposed to reading through the table.

Based on the recall and F1-score metrics, the XGB Classifier appears to have the best overall performance, achieving 99.3% for both metrics. This suggests that the XGB Classifier is able to correctly identify true positives and true negatives while minimizing false positives and false negatives.

In terms of accuracy, the XGB Classifier also has the highest value at 99.73%. However, all five algorithms perform well with high accuracy ranging from 94.78% to 99.73%.

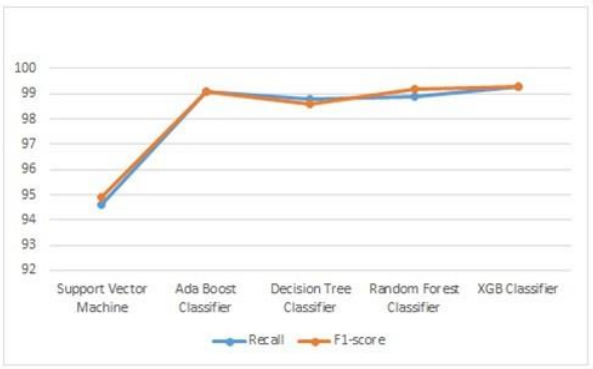

Fig. 3 Recall and F1-score performance of various Machine Learning Algorithms

It is important to note that the choice of algorithm should be based on the specific problem and dataset being analyzed. Other factors such as the interpretability and computational requirements of the algorithm should also be considered.

Overall, the presented data provides valuable insight into the performance of various classification algorithms. Further research may include exploring the reasons for the superior performance of the XGB Classifier and examining the robustness of these algorithms to different datasets and problem domains.

5. Conclusion

The research described in the article suggests that a combination of simple heuristics and machine learning algorithms can be used to effectively identify fake social media accounts. The authors used a dataset of social media accounts, labeled as either genuine or fake, and extracted seven features from each account to use as input to their classification model. The results of our experiments showed that the combination of heuristics and machine learning algorithms significantly outperformed individual algorithms or heuristics alone. The best-performing model achieved an accuracy of 99.73%. Hence we can conclude that machine learning approach could be used to build an automated system for detecting fake social media accounts, which could be useful for identifying and mitigating the impact of misinformation campaigns on social media. However, we also noted that the limitations of our dataset and the evolving nature of fake account creation mean that our approach will need to be continuously updated and refined to remain effective.

The study's research approach has a number of advantages, such as the use of a sizable and varied dataset, the implementation of different machine learning algorithms, and the use of multiple evaluation metrics to gauge the effectiveness of the algorithms.

The approach does have some drawbacks, though. The research, for instance, relied on data that was made openly accessible, which might not correctly reflect the real situation on social media platforms. Furthermore, the research did not take into consideration how different social and cultural variables may affect the ability to identify phoney profiles, which may restrict the applicability of the results.

Limitations of the Study and Future Research Directions:

The study has some limitations that need to be addressed in future research. For exam- ple, the study only focused on the detection of fake accounts and did not consider other forms of malicious activities such as phishing and spamming. Future research could explore the detection of these activities using machine learning algorithms. Addition- ally, the study did not consider the temporal aspect of the data, which may affect the performance of the machine learning algorithms. Future research could consider the dynamic nature of social media and develop real-time detection methods that can adapt to changes in the data. Finally, future research could also consider the impact of user behavior and social network characteristics on the detection of fake accounts, which may improve the accuracy and effectiveness of the detection methods.

References

[1] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: Exploiting tri-relationship for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, ACM, pp. 153–162 (2017)

[2] Zhang, X., Liu, Y., Wang, H., Tang, J.: A machine learning approach to detecting fake accounts on social media. In: Proceedings of the 2019 IEEE International Conference on Big Data, IEEE, pp. 3843–3845 (2019)

[3] Singh, P., Singh, S., Singh, V.: Detecting fake accounts on social media using machine learning and data mining techniques. In: Proceedings of the Interna- tional Conference on Machine Learning, Big Data, Cloud and Parallel Computing, Springer, pp. 77–83 (2020)

[4] Karunarathne, A., Madhushan, T., Silva, K.: Fake account detection on social media using machine learning. In: Proceedings of the International Conference on Artificial Intelligence and Computer Vision, Springer, Cham, pp. 51–55 (2021)

[5] Zafarani, R., Liu, H.: Connecting the dots between news articles and social media. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management, Springer, pp. 1671–1674 (2009)

[6] Cui, Y., Wang, T., Chen, L., Wen, J.-R., Liu, Y.: Enhancing cnn with attention for fake news detection. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 2371–2381 (2019)

[7] Wang, G., Xie, X., Liu, Y., Tang, J., He, D.: Detecting fake accounts in social media using machine learning: A systematic review. IEEE Access 7, 173008– 173018 (2019)

[8] Dey, A., Mukhopadhyay, A., Chattopadhyay, A.: Machine learning-based approaches for detecting fake accounts on social media: a review. Journal of Ambient Intelligence and Humanized Computing, Springer 11, 1153–1169 (2020)

[9] Chen, J., Xu, C., Wang, X.: A machine learning approach to detecting fake accounts in social media. Multimedia Tools and Applications, Springer 80, 19515–19533 (2021)

[10] Seneviratne, N.M., Ekanayake, D.D.: Detecting fake social media accounts using machine learning techniques. International Journal of Computer Science and Network Security 20, 150–156 (2020)

[11] Lee, J., Lee, J., Lee, K., Jung, H.: Detection of fake accounts on social media using machine learning and network analysis. Sustainability, Multidisciplinary Digital Publishing Institute 13(5), 2655 (2021)

[12] Ramanathan, M., Garg, S., Saini, D.: Detecting fake accounts on social media: A machine learning approach. Proceedings of the 2020 12th International Confer- ence on Communication Systems & Networks (COMSNETS), IEEE Access, 1–6 (2020)

[13] Pan, J., Wang, Y., Xie, J., Zhu, W., Liu, J.: An intelligent approach to detect- ing fake accounts on social media using machine learning. Mobile Networks and Applications, Springer 26, 437–448 (2021)

[14] Gilda, S., Kalra, D., Bali, K.: Fake news detection using machine learning: A review. Artificial Intelligence Review, Springer 53, 4133–4166 (2020)

[15] F. N. Pakaya, M. O. Ibrohim, and I. Budi, "Malicious Account Detection on Twitter Based on Tweet Account Features using Machine Learning," in 2019 Fourth International Conference on Informatics and Computing (ICIC), IEEE, pp. 1-5, Oct. 2019, doi: 10.1109/icic47613.2019.8985840.

[16] [16] P. Chakraborty, M. M. Shazan, M. Nahid, M. K. Ahmed, and P. C. Talukder, "Fake Profile Detection Using Machine Learn ing Techniques," Journal of Computer and Communications, vol. 10, no. 10, pp. 74–87, 2022, doi: 10.4236/jcc.2022.1010006.

[17] [17] K. Shreya, A. Kothapelly, D. V, and H. Shanmugasundaram, "Identification of Fake accounts in social media using machine Learning," in 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), IEEE, pp. 1-4, Dec. 2022, doi: 10.1109/icerect56837.2022.10060194.

[18] [18] S. Khaled, N. El-Tazi, and H. M. O. Mokhtar, "Detecting Fake Accounts on Social Media," in 2018 IEEE International Conference on Big Data (Big Data), IEEE, pp. 3672-3681, Dec. 2018, doi: 10.1109/bigdata.2018.8621913.

[19] [19] S. D. P. Reddy, "Fake Profile Identification using Machine Learning," International Research Journal of Engineering and Technology (IRJET), vol. 6, no. 12, pp. 1145–1150, 2019.

[20] [20] F. C. Akyon and M. E. Kalfaoglu, "Instagram Fake and Automated Account Detection," in 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE, pp. 1-7, Oct. 2019, doi: 10.1109/asyu48272.2019.8946437.