Fake Accounts Detection in Online Social Networks using Hybrid Machine Learning Models

Fatema A. Sarhan and Ebrahim Mattar+

College of Engineering, University of Bahrain, P. O. Box 32038, Kingdom of Bahrain
ebmattar@uob.edu.bh+

***Abstract* - Online social networks (OSNs) have become popular worldwide and are excellent spaces for exchanging ideas, keeping up with the news, and promoting goods. However, because of their increasing popularity, social networks have become a target for suspicious exploitation, such as the disseminating of false or harmful information, which makes them less dependable and trustworthy. Unwanted content may circulate on the social network by way of the creation of fake and malicious accounts. Therefore, predicting fake accounts is a significant issue. In this research, we applied various machine learning methods to this issue and assessed and compared their results.**

Keywords* - *Fake Accounts, SVM, ANN, RF, KNN, Machine Learning, Hybrid.

I. INTRODUCTION

Online social networks (OSNs) play a significant part in the day-to-day activities of Internet users, including reading news, sharing material, writing product evaluations, publishing messages, and discussing current events. Popular social media platforms can be abused and present new trust, security, and privacy problems. Fake accounts are frequently made to spread dangerous material, such as spam, malware, hazardous unified resource locators (URLs), or unsolicited communications, to legitimate users [1]. Machine learning has fundamentally changed how data is extracted and handled by replacing the older statistical approaches. Reinforcement learning, supervised learning, and unsupervised learning are the three categories into which machine learning techniques fall. Even though the various classification algorithms (Support Vector Machine, Logistic Regression, Decision Tree, Random Forest, and Artificial Neural Network) can be used independently, a hybrid model that combines two or three machine learning models can help to increase the model's accuracy and predictive power [2]. Recent studies have discovered that machine learning strategies provide improved outcomes for spotting fake accounts. A neural network is made up of numerous connected processing components. It makes choices similar to a human brain [3].

For classification, supervised machine learning algorithms called support vector machines (SVM) are employed. To classify the data, it locates the hyperplane. Based on a variety of account criteria, neural networks and SVM are appropriate for detecting fake accounts on social networking sites since they can accept a significant amount of random input. The Naive Bayes classifier is based on the Bayes theorem. It forecasts the likelihood that a particular variable belongs to a specific class [3]. In this paper, we will discuss some similar research that has employed a hybrid approach to address the
issue of fake account identification. We will then demonstrate our own proposed solution and present a comparison.

II. RELATED WORK

By using SVM trained model decision values to train a NN model and also by using SVM testing decision values to test the NN model, Sarala [4] suggested a new classification approach to enhance the detection of fake accounts on social networks. They employed the "MIB" baseline dataset to accomplish their objective and ran it through the preprocessing stage, where four feature reduction techniques were applied to minimize the feature vector. Multiple Linear Regression, K-Means Spearman's Rank Order Correlation, and Wrapper SVM. There were two learning algorithms utilized during the classification phase. They found that Naïve Bayes had better accuracy across the board when compared to the other two classifiers, with a classification accuracy of about 98%. When compared to Decision Tree and Naïve Bayes, the NN algorithm's classification accuracy was shown to be the lowest. This happened because the SVM method, which uses the gradient descent technique, obtains the global minimum of the optimized function whereas the NN algorithm does not.

Biyani [6] aimed to illustrate the core process of classifying tweets into the genuine or spam category by setting a baseline and illustrating how language models are related to the algorithm and can improve outcomes. Python code was used to test the support vector machine (SVM), decision tree, and logistic regression machine learning algorithms to see which one was most effective in telling Twitter apart from spam accounts. The outcomes demonstrate

that the results generated by the support vector machine algorithm are more accurate than those generated by other methods.

A hybrid technique is employed in Bhambar's [3] planned effort to create the most effective classifier possible using neural networks and SVM. In order to increase the algorithm's precision and decrease its time complexity, K-mediod clustering is also used. They gathered real-time Facebook or Twitter data set from Facebook or Twitter users for the suggested study.

Nagariya [2] suggested a system that employed a feature-based dataset and hand-selected features. This method is based on the account information and user-level behaviors. They use a hybrid strategy to compare various classification algorithms, submit the results to a voting classifier, and then submit the results to a neural network. By testing and training the dataset on various hybrid approaches to classification algorithms, they have maintained the highest accuracy in identifying fake accounts. The outcomes demonstrate an improvement in accuracy because of the various classification algorithms. When using SVM, LR, and NN hybrid approaches, they achieved accuracy of 99.56%, which was the highest.

For the purpose of detecting fake Twitter accounts, Benabbou [1] suggested a system based on the BiGRU model and the pre-trained model GloVe. To extract the syntactic and semantic characteristics of comments, this approach concentrated on the tweet content level. With an accuracy of 99.44%, the comparison findings showed that this strategy surpasses the LSTM, CNN, LSTM+BiGRU, and CNN+BiGRU models. GloVe captures both global and local textual information, which explains why it performs better than Word2vec. Their algorithm, however, is unable to differentiate between several varieties of false accounts when it comes to categorizing accounts. To improve and extend the suggested fake account detection system, a system that classifies the various types of fake accounts by taking into account additional features related to user behavior and comments may be the subject of a research proposal project in the future [1].

Singh [5] utilized neural networks and user profiles to try to address the issue of fake accounts on the social media website. In contrast to the prior approach, which made use of machine-neural networks to eliminate fake profiles, the method proposed in this research is predicted to prevent the generation of fake profiles. The article only suggests the concept of restricting each person to having a single profile on a social media platform in order to stop the establishment of a false profile by mapping the creator's facial signature and profile data verification age. Since there needs to be testing before this technology is fully implemented, accuracy cannot be praised.

III. PROPOSED WORK

For classifying real from fake accounts, this proposed work employs techniques such as artificial neural networks (ANN), K nearest neighbor (KNN), random forest (RF), and support vector machine (SVM). The feature set that affects Instagram's ability to identify fake accounts will be applied. This proposed work is anticipated to produce the higher accuracy values needed for the detection of fake Instagram accounts. In the proposed study, machine learning techniques' accuracy will be compared.

A. System Architecture

Two different Instagram datasets were used in our model. Both datasets were fed into the system for pre-processing, with the data split into training and test sets, and then a machine learning model was applied. In our model, we are using four machine learning models: SVM, KNN, ANN, and RF. We will compare and demonstrate the results. Figure 1 illustrates the system architecture.

B. Dataset

We have used two publicly accessible datasets. The first dataset was gathered by Jafari [7], who made an Instagram account and subsequently purchased 700 fake followers. To stop new users from following the page, the account privacy was adjusted to private after that. He utilized a C# web scraper to get data from all 700 accounts. From his private Instagram account, real users were taken. The second dataset was compiled by Bakhshandeh [8]. Bakhshandeh [8] personally identified the spammer or fake accounts contained in this dataset after carefully reviewing each instance, and as a result, the dataset has a high level of accuracy, even though Bakhshandeh [8] acknowledged that there may be a few accounts in the spammer list that were incorrectly identified. A crawler was used to gather the data between March 15 and March 19, 2019.

C. Data Preprocessing

The first dataset, which is not balanced as depicted in Figure 2, comprises 12 features with 692 fake users and 93 real users. Since none of the accounts in the dataset have a channel, the "has channel" attribute has no impact on the dataset and therefore no impact on the results. With 348 actual accounts and 348 fake accounts, which are in this instance balanced as illustrated in Figure 3, the second dataset comprises 11 features. 20% of the dataset was used for testing, while 80% was used for training, for both datasets.

D. Classification Algorithms

D1. Support Vector Machine: Support vector machines (SVMs) are a group of supervised learning techniques for

classifying data, performing regression analysis, and identifying outliers. This method maps the data into a massively dimensional input space and then creates a perfect disjointed hyperplane within it. The most common software problem is quadratic; however, for neural network designs, an inclination tutoring technique degrades the authenticity of the majority of native minima. An SVM model is just a hyperplane in multidimensional space that represents several classes. SVM will generate the hyperplane in an iterative manner in order to reduce error. SVM aims to classify datasets in order to find a maximum marginal hyperplane (MMH) [6].

For the first dataset, we have implemented SVM and applied SVC, a linear kernel, a Gaussian rbf, poly, and a sigmoid kernel. With accuracies of 91.1%, 92.4%, 91.9%, 92.4%, and 89.2%, respectively. For the second dataset the accuracies were 91.4%, 88.6%, 91.4%, 85%, and 88.6% respectively.

D2. Random Forest: The first dataset got an accuracy of 92.99%, and the second one got an accuracy of 90.71% by applying Random Forest algorithm.

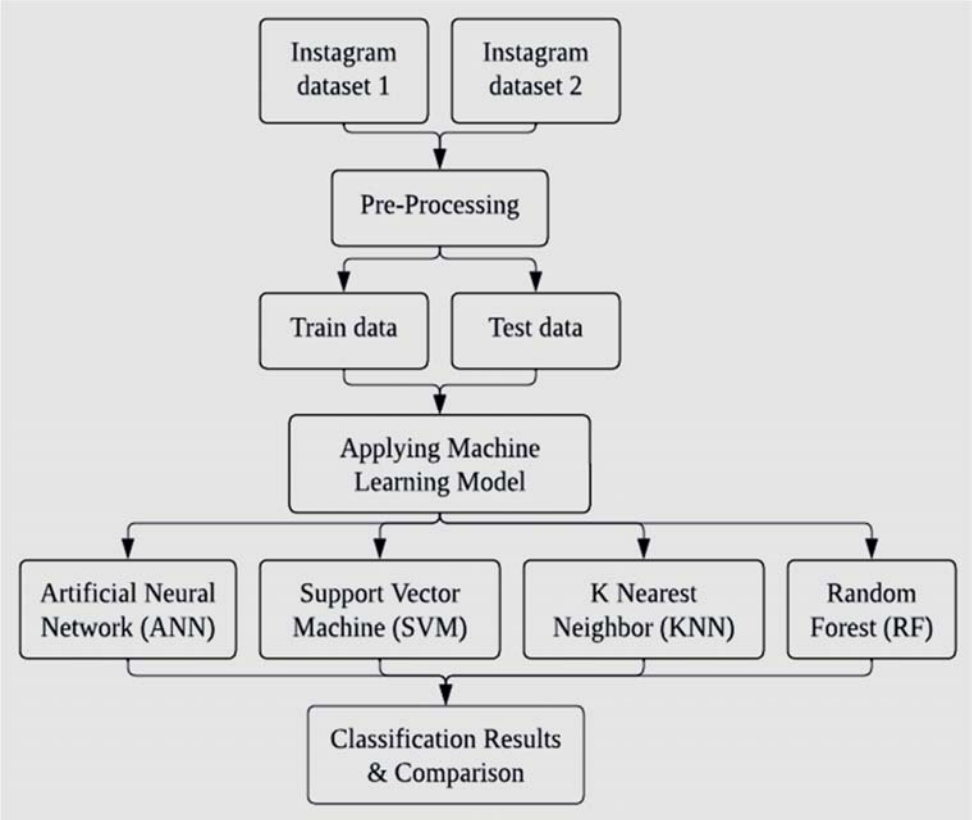

Figure 1. Model Architecture

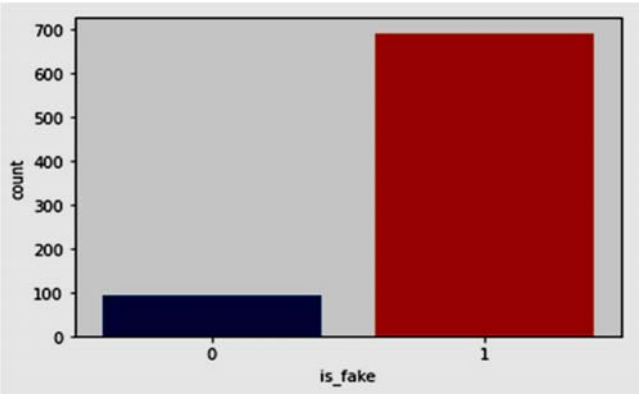

Figure 2. 1st Dataset fake and real users' distribution.

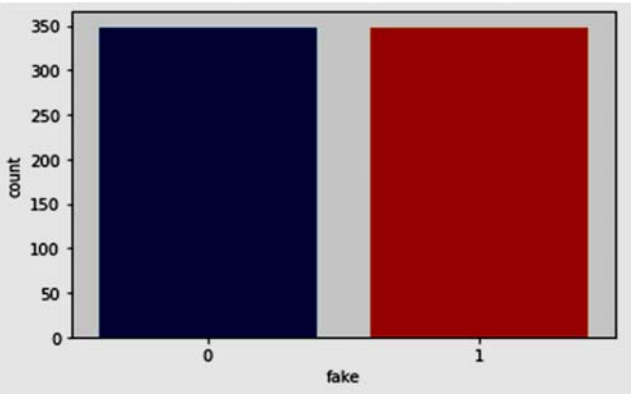

Figure 3. 2nd Dataset fake and real users' distribution.

TABLE I. FEATURES OF DATASETS

Features	
1st dataset	*2nd dataset*
edge_followed_by	#followers
username_has_number	nums/length fullname
full_name_has_number	name==username
is_private	private
has_channel	description length
has_guides	profile pic
edge_follow	#follows
username_length	nums/length username
full_name_length	fullname words
is_joined_recently	#posts
has_external_url	external URL
is_business_account	

D3. K Nearest Neighbor: We have got 91.71% accuracy for the first dataset with a k value of 8 and 91.42% accuracy for the second dataset with a k value of 8. Figure 4 and 5 shows the best accuracies depending on k value.

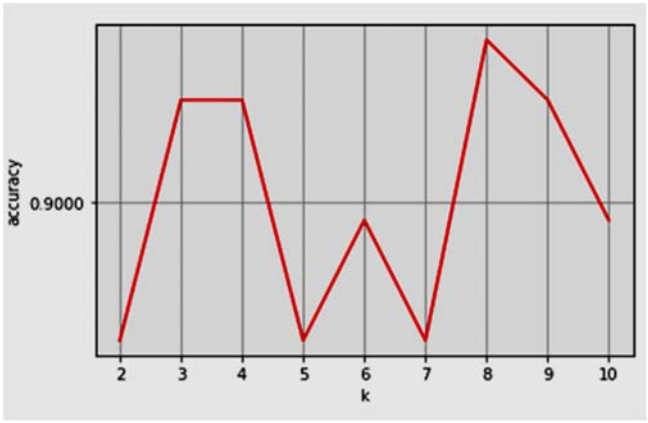

Figure 4. Accuracies depending on k value (1st dataset).

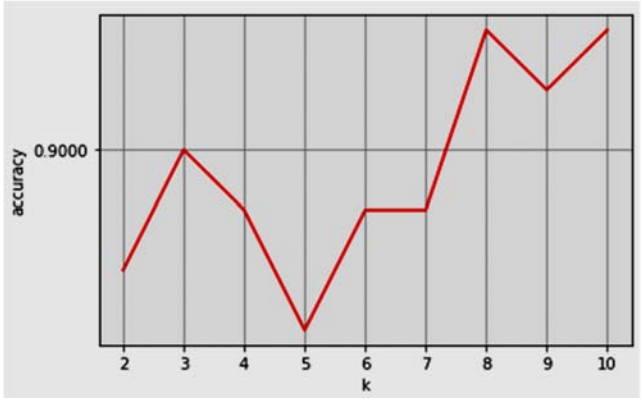

Figure 5. Accuracies depending on k value (2nd dataset).

D4. Artificial Neural Network: The sequential model for artificial neural networks has been implemented, and three additional layers have been added. The first layer contains the relu activation function, with input dimensions of 12 for the first dataset and 11 for the second. The relu activation function is included in the second layer, while the sigmoid activation function is included in the final layer. The model was then generated using the Adam Optimizer with a batch size of 10 and 100 epochs. In the end, the accuracy for the first dataset was 91.08%, and the accuracy for the second dataset was 92.14%.

IV. EXPERIMENTAL RESULTS

The two datasets behaved differently when machine learning classification models were applied because they had different features and are unrelated to one another. According to Table II, the Random Forest classifier had the best performance for the first dataset, with a 92.99% accuracy rate. According to Table III, applying ANN produced the second dataset's best accuracy, which was 92.14%.

TABLE II. ACCURACIES FOR THE 1ST DATASET

1st dataset Algorithms Accuracies		
Model	*Accuracy*	
SVM	Svc	91.1%
	Linear kernel	92.4%
	Gaussian rbf	91.9%
	poly	92.4%
	Sigmoid kernel	89.2%
RF	92.99%	
KNN	91.71%	
ANN	91.08%	

TABLE III. ACCURACIES FOR THE 2ND DATASET

2nd dataset Algorithms Accuracies		
Model	*Accuracy*	
SVM	Svc	91.4%
	Linear kernel	88.6%
	Gaussian rbf	91.4%,
	poly	85%
	Sigmoid kernel	88.6%
RF	90.71%	
KNN	91.42%	
ANN	92.14%	

V. CONCLUSION AND FUTURE WORK

This study applied various machine learning models to various datasets in an effort to address the issue of fake accounts on social networking websites. We applied SVM, KNN, RF, and ANN to two distinct datasets with various features to demonstrate how they perform in each classification algorithm and what accuracy they acquire. At the end, we have some pretty good accuracy. Although there is still room for development, we intend to employ large datasets with more features in the future by implementing additional machine learning algorithms and other helpful techniques.

ACKNOWLEDGMENT

We want to express our appreciation to University of Bahrain for their support and guidance.

REFERENCES

[1] F. Benabbou, H. Boukhouima, and N. Sael, "Fake accounts detection system based on bidirectional gated recurrent unit neural network," International Journal of Electrical and Computer Engineering (IJECE), vol. 12, no. 3, p. 3129, 2022.

[2] H. G. Nagariya, N. Dhanotiya, S. Joshi, and S. Jain, "Identifying Fake Profile in Online Social Network"

[3] S. Bhambar, K. Khairnar, Y. Nikam, H. Shelar, and Y. K. Desai, "DETECTING FAKE ACCOUNTS ON SOCIAL MEDIA USING NEURAL NETWORK," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 4, no. 5, 2022.

[4] V. Sarala and G. Sandhya, "Spammer Detection and fake user Identification on Social Networks," Journal of Engineering Sciences, vol. 13, no. 8, 2022.

[5] V. Singh, R. Shanmugam, and S. Awasthi, "Preventing fake accounts on social media using face recognition based on convolutional neural network," *Sustainable Communication Networks and Application*, pp. 227–241, 2021.

[6] Y. V. Biyani, "SPAM detection in social media using machine learning algorithm," *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, no. 1, pp. 432–439, 2021.

[7] R. Jafari, "Instagram fake and real accounts dataset," Kaggle, 07-Jan-2021. [Online]. Available: https://www.kaggle.com/datasets/rezaunderfit/instagram-fake-and-real-accounts-dataset.

[8] B. Bakhshandeh, "Instagram fake spammer genuine accounts," Kaggle, 22-Mar-2019. [Online]. Available: https://www.kaggle.com/datasets/free4ever1/instagram-fake-spammer-genuine-accounts.