International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue:11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

FAKE ACCOUNT IDENTIFICATION USING MACHINE LEARNING APPROACHES INTEGRATED WITH ADAPTIVE PARTICLE SWARM OPTIMIZATION

Dr. A. NISHA JEBASEELI
Assistant Professor of Computer Science, CDOE-Bharathidasan University, Tiruchirappalli - 620024.

Abstract

It is customary for humans, bots, and other automated systems to generate new user accounts by utilizing pilfered or otherwise deceitful personal information. They are employed in deceitful activities such as phishing and identity theft, as well as in spreading damaging rumors. An somebody with malevolent intent may generate a substantial number of counterfeit accounts, ranging from hundreds to thousands, with the aim of disseminating their harmful actions to as many authentic users as possible. Users can get a wealth of knowledge from social networking networks. Malicious individuals are readily encouraged to take use of this vast collection of social media information. These cybercriminals fabricate fictitious identities and disseminate meaningless stuff. An essential aspect of using social media networks is the process of discerning counterfeit profiles. This study presents a machine learning approach to detect fraudulent Instagram profiles. This strategy employed the attribute-selection technique, adaptive particle swarm optimization, and feature-elimination recursion. The results indicate that the suggested adaptive particle swarm optimization method surpasses RFE in terms of accuracy, recall, and F measure.

Keywords: Fake account, Machine Learning, Feature Selection, Adaptive Particle Swarm Optimization, Recursive Feature Elimination.

1. INTRODUCTION

Social networking has become a popular online activity nowadays, with millions of users and billions of minutes spent on these sites. Online social network services encompass a spectrum of platforms, ranging from those that prioritize social contact, such as Facebook or Myspace, to those that prioritize the transmission of knowledge, such as Twitter or Google Buzz, to those that add social interaction capabilities to existing systems, like Flicker. Conversely, enhancing security concerns and upholding the privacy of online social networks (OSNs) remain a vital objective and perceived undertaking. Social media platforms are commonly utilized for the transmission and reception of data. Individuals universally employ social media platforms for various purposes, such as sharing high-value visual content, tracking the activities of public figures, or engaging in communication with both local and distant acquaintances. This location provides an exceptional environment for engaging in social interactions and sharing knowledge. The user's text is "[1].". They employ them to engage with the content and information shared by other users on the network. The advent of social media has facilitated the interconnection and information sharing among individuals from diverse backgrounds, fostering the dissemination of creative content. Additionally, it offers a forum for individuals to exhibit their expertise and establish connections with like-minded individuals worldwide. The user's text is "[2]." Various platforms serve as social media, such as Instagram, LinkedIn, Twitter, Facebook, Snapchat, and others. With the advancement of technology, social networking has become an essential part of most people's everyday lives.

A counterfeit account refers to an account on any social media platform that presents misleading or deceitful information. Fabrication using fraudulent accounts, employing fictitious particulars, misleads the general populace in disseminating erroneous data or gathering monetary or personal data [8]. Individuals create profiles on diverse social networking platforms to share and exchange digital content. Users often create profiles with fabricated or anonymous information to disseminate fraudulent information while concealing their identities. Users commonly modify their accounts or establish accounts under another person's identity (identity theft). There are distinct financial benefits associated with the creation of fraudulent accounts. These phone identities are managed by bots or automated programs, which facilitate the widespread and rapid spread of fake news on the internet. Within the network, counterfeit accounts frequently establish connections and actively monitor the postings of influencers. By implementing measures to combat impersonation, popular online social networking platforms such as Twitter, Facebook, and WhatsApp have the ability to delete or suspend user accounts associated with fraudulent activity. Shortly after the bombing at the Boston Marathon, some 32,000 new accounts

were registered, with roughly twenty percent of them having their profiles terminated by Twitter (Gupta et al., 2013). Most malicious profiles are formed with the purpose of engaging in spamming, phishing, and acquiring a larger number of followers. The fraudulent accounts possess all the requisite instruments to perpetrate cyber offenses. Fraudulent accounts provide significant dangers in terms of identity theft and data breaches. When users access the URLs sent by this phony profile, all of their personal information is transmitted to remote servers, where it can potentially be used against them. Spurious profiles purporting to represent companies or individuals have the potential to damage their reputation and result in diminished popularity and followership [9].

Fake account detection using machine learning techniques addresses the growing concern of fraudulent activities on online platforms, where individuals create deceptive profiles for various malicious purposes. These fake accounts can be used for identity theft, spreading misinformation, engaging in scams, or manipulating public opinion. To combat this issue, machine learning plays a crucial role in developing robust and automated methods for identifying and mitigating the impact of fake accounts.

Key Components:

1. **Data Collection:**
   - Building a reliable dataset is a fundamental step in training machine learning models for fake account detection. The dataset should encompass diverse examples of both genuine and fake accounts, including various characteristics such as profile information, activity patterns, and user interactions.
2. **Feature Extraction:**
   - Extracting meaningful features from user profiles and their activities is essential for training effective machine learning models. Features may include account creation date, posting frequency, content type, friends or followers network, and other behavioral patterns.
3. **Labeling and Annotating:**
   - Each instance in the dataset needs to be labeled as either a genuine or fake account. Annotating the data involves marking specific attributes or patterns that indicate fraudulent behavior, such as excessive use of generic profile pictures, irregular posting behavior, or suspicious friend/follower connections.
4. **Supervised Learning Models:**
   - Machine learning models, particularly supervised learning algorithms, are trained on labeled data to distinguish between genuine and fake accounts. Common algorithms include Decision Trees, Random Forests, Support Vector Machines (SVM), and more advanced techniques like ensemble methods.
5. **Behavioral Analysis:**
   - Beyond static features, machine learning models may employ behavioral analysis to detect anomalies or inconsistencies in user activities. This can involve analyzing the temporal patterns of posting, interactions, or sudden changes in account behavior.
6. **Unsupervised Learning and Clustering:**
   - Unsupervised learning techniques, such as clustering, can be employed to identify patterns in the data without labeled examples. Clustering algorithms can help group together accounts with similar characteristics, making it easier to detect outliers that may indicate fake accounts.
7. **Deep Learning Approaches:**
   - Deep learning models, such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, can capture sequential patterns and dependencies in user activities, enhancing the ability to identify sophisticated fake accounts with dynamic behaviors.
8. **Evaluation Metrics:**
   - Metrics like precision, recall, F1 score, and accuracy are used to evaluate the performance of the machine learning models. These metrics provide insights into the model's ability to correctly identify fake accounts while minimizing false positives and false negatives.
9. **Deployment and Integration:**
   - Once trained and validated, the model can be deployed and integrated into online platforms and social networks to automatically identify and flag suspicious accounts. This integration helps in real-time monitoring and prevention of fake account proliferation.
10. **Continuous Improvement:**

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: 10 September 2023 ***Revised:*** 20 October 2023 ***Accepted:*** 30 October 2023

- Given the dynamic nature of online threats, continuous monitoring and periodic model updates are essential to adapt to evolving tactics used by malicious actors. This ensures that the fake account detection system remains effective over time.

2. BACKGROUND STUDY

Various machine learning algorithms can be employed for fake account detection, each with its strengths and weaknesses. The choice of algorithm often depends on the characteristics of the data and the specific requirements of the detection task. Here are some commonly used machine learning algorithms for fake account detection:

1. **Decision Trees:**
   - Decision trees are used to make decisions based on a series of conditions. In fake account detection, decision trees can analyze different features such as account creation date, posting frequency, and profile completeness to classify accounts as genuine or fake.
2. **Random Forests:**
   - Random Forests are an ensemble learning technique that combines multiple decision trees to improve accuracy and robustness. They are effective in handling large and diverse datasets, making them suitable for fake account detection where the data can be complex.
3. **Support Vector Machines (SVM):**
   - SVM is a supervised learning algorithm that can be used for binary classification tasks. SVM aims to find a hyperplane that separates genuine and fake accounts in feature space. It is particularly useful when dealing with high-dimensional data.
4. **Logistic Regression:**
   - Logistic Regression is a simple yet effective algorithm for binary classification. It models the probability of an account being fake based on linear combinations of input features. It is interpretable and computationally efficient.
5. **K-Nearest Neighbors (KNN):**
   - KNN is a non-parametric and instance-based learning algorithm. In fake account detection, KNN can classify an account based on the majority class of its k-nearest neighbors in the feature space.
6. **Naive Bayes:**
   - Naive Bayes is a probabilistic algorithm based on Bayes' theorem. It assumes that features are independent, which might not hold true in all cases. Naive Bayes is computationally efficient and can be effective in certain scenarios.
7. **Neural Networks:**
   - Deep learning models, including neural networks, can capture complex relationships and patterns in data. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are often used for feature extraction and sequential data analysis, respectively, in the context of fake account detection.
8. **Ensemble Methods:**
   - Ensemble methods, such as AdaBoost or Gradient Boosting, combine the outputs of multiple weak learners to create a strong classifier. These methods can enhance the model's generalization and performance.
9. **One-Class SVM:**
   - One-Class SVM is a variation of traditional SVM designed for anomaly detection. It can be useful when there is a scarcity of labeled examples of fake accounts, as it learns to identify patterns of genuine accounts and flags deviations as potential fakes.
10. **Autoencoders:**
    - Autoencoders are a type of neural network used for unsupervised learning. They can learn a compressed representation of input data and are useful for detecting anomalies or outliers, making them applicable to fake account detection.

3. LITERATURE REVIEW

In their study [5], the authors propose the use of a detection technique called 3PS (Publicly Private Protected system) to identify phony accounts on online social networks. This technique involves analyzing the behavioral patterns of user account activity. It focuses on identifying fraudulent or harmful users who possess several online social network accounts, make friend requests, and distribute numerous posts with destructive intent. The system takes into account the

posts, follows, status updates, followers, and posts. A fraudulent or malevolent user can be identified by differentiating between the threshold value of a characteristic linked to the user's personal profile and the examination of network similarities.

A system was developed in [6] to detect counterfeit Twitter accounts. The system included logistic regression with PSO, naïve Bayes, and KNN algorithms to classify user profile characteristic features. The feature selection techniques employed were information gain, correlation-based feature selection (CFS), and minimum relevance maximum redundancy (MRMR). The dataset, comprising 6973 profile data, was collected by manual methods and the Twitter REST API. It was then divided into training and testing sections.

In [7], a framework was utilized that conducts a dual analysis, specifically predicting the trustworthiness of messages and the reliability of social network user profiles. The process encompasses both offline research utilizing deep learning algorithms and online analysis involving real users to determine the trustworthiness of a Twitter presence. The United Nations should use the methodology proposed by S. Uppada et al. [8]. User interaction trends are tracked by utilizing statistical methods such as rating users and articles, as well as identifying false profiles. The integration of forensic approaches and photo polarity analysis is employed to generate attributes associated with fake news images. The SENAD technique achieved an accuracy level of 76.3% in detecting fake news, whereas the CredNN model achieved a higher accuracy level of 93.5%.

The researcher suggests a novel approach for identifying bots by utilizing deep neural networks and active learning. The package includes modules for extracting features, doing active learning, gathering and classifying data, and detecting patterns. Furthermore, this system provides a state-of-the-art RGA deep neural network model for identification, incorporating ResNet, BiGRU, and attention mechanisms. The test results clearly showed that the proposed DA Bot framework outperforms existing detection methods in terms of its effectiveness in recognizing social bots [9].

The method described in [10] utilizes a framework that relies on a chrome extension to detect counterfeit Twitter accounts. They also compared several jobs employing machine learning techniques to validate user data collected through manual and web crawling methods. The dataset was acquired from the Twitter profile by utilizing a web crawler and the Twitter API. The collected profile data encompasses the user's name, ID, count of status updates, friends list, count of favorites, and count of URLs referenced in tweets. Subsequently, the data is partitioned into test and training datasets, with a distribution ratio of 80:20. And successfully completed the WEKA machine learning platform. The trust score derived from the features is computed and subsequently utilized by the Chrome extension to detect the presence of a malicious profile. By employing user-specific data and utilizing a trust score, Chrome extensions generate a measure of suspicion for each user. The utilization of random forest and bagging techniques is employed to detect fraudulent profiles and assess the overall effectiveness of the Chrome plugin.

A semi-supervised clustering approach, which incorporates partial background information, along with the deep walk technique on rater graphs, offers a hierarchical framework for identifying fraudulent associations among potential reviewers. Furthermore, by employing temporal affinity, semantic features, and sentiment analysis, this approach can be extended to identify clusters of individuals who engage in spamming activities on social media platforms. Identification of probable spammer groups only based on the topological nature of the underlying graph. The reviewer ID, represented by feature vectors, is partitioned into several clusters of potential spammers using an adapted version of the semisupervised clustering algorithm known as Pairwise. This is done after getting a representation for each node in G using the Deep Walk technique. K-Means with constraints. The methodology was validated using a subset of the main data set, which included 2207 fake reviewer IDs from 23 distinct reviewers, in order to identify clusters of fraudulent reviewers in reviewer graphs [11].

Propose a methodology utilizing dynamic knowledge graphs to detect fraudulent input in the context of [12]. The extraction of the first four types of entities is achieved by utilizing a recently created neural network model called conditional two-way long-term short-term memory. This model employs the embedding of phrase vectors/double words, taking into account the specific features of online product reviews. Subsequently, time series-related characteristics are integrated into the process of constructing the knowledge graph in order to produce dynamic graph networks. The researchers initially employed the ST-BLSTM method to extract several types of entities. Subsequently, they introduced the latest MI-based criteria to assess the relationships between these entities. They created a dynamic graph network by combining the aspects related to the sequence of time. The

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: 10 September 2023 *Revised:* 20 October 2023 *Accepted:* 30 October 2023

second study demonstrated the need of assessing the veracity and quality of the product by quantifying trustworthiness, honesty, high product quality, and reliability evaluations using innovative approaches. The technique highlighted how the analysis of portrait data accurately represents the underlying causes of fraudulent reviews and offers vital insights. The technique showcased how analyzing photo data reveals the underlying causes of fake reviews and provides valuable insights into different types of reviewers.

In [20], a combination of a long-short memory neural network and an AdaBoost model is proposed for the geolocation-based profile recognition prototype. This combination is used to analyze user account and geolocation data. The two sub-models of this system consist of the geolocation detection model and the profile detection framework. Geolocation refers to the process of determining the exact geographical location of a device or user. The direct message (DM) function accepts geolocation data and utilizes Long Short-Term Memory (LSTM) to assess the sequence of geolocation information. This analysis is used to provide prediction scores for identifying the user. The Account-DM module collects input regarding account features and performs analysis on these features using the AdaBoost algorithm. The SVM linear classifier produces the ultimate assessment of user IDs by utilizing the prediction scores from the profile-DM and geolocation-DM as input. According to the study, this technique can accurately and efficiently detect bogus reviews. The study utilized a large dataset from Yelp.

4. PROPOSED SYSTEM

In this research, we have given various machine learning methodologies and attribute selection strategies that are employed to detect fraudulent profiles in online social media. In addition, we are incorporating the Adaptive Particle Swarm optimization and recursive feature reduction approaches to enhance the accuracy of detecting bogus profiles.

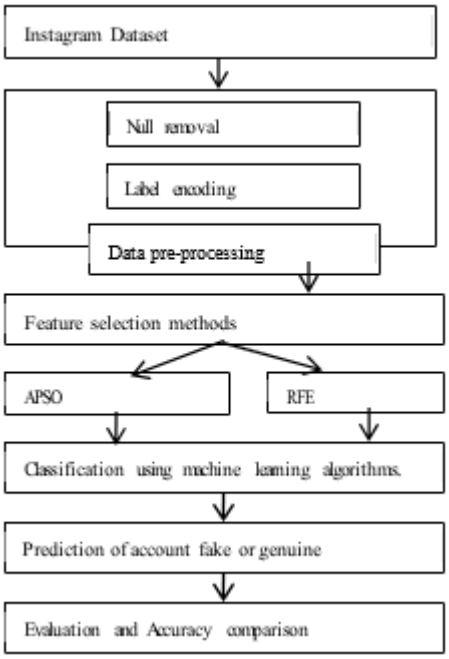

Fig. 1. Proposed System Flow-chart

The following is a summary of the steps for identifying fake profiles:

1.Collect the Instagram dataset from Kaggle.

2.Data pre-processing includes removing null data, label encoding.

3.The data set should be divided into test and training data.
4.Applying feature elimination techniques APSO and RFE.

5.Predict the test result of fake and genuine profiles.
6.Compute accuracy.

7.Compare the performance of classification models.

4.1 Dataset

Instagram accounts authenticity has been determined using data from a Kaggle competition. There are ten different types of information in this dataset, including followers, following, post count, username length, profile image, description length, private, public, and external URL. There are 13 fields to explore in the Kaggle dataset. These characteristics may be used as clues to determine whether a profile is real or a fake. In order to train the machine learning models, all of the characteristics had to be investigated and assessed.

TABLE I. ATTRIBUTES FOR INSTAGRAM PROFILE

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: 10 September 2023 ***Revised:*** 20 October 2023 ***Accepted:*** 30 October 2023

Attributes	Attributes Description
Profile pic	Profile picture is present or not
nums /length username	Ratio of number of numerical chars in username to its length
full name words	full name in word tokens
name==username	Username and full name are same
nums/length full name	Ratio of number of numerical characters in full name to its length
description length	Length of the description in account
external URL	Any link or URL is available or not
private	Profile is private or public
posts	Number of posts
follower	Number of Follower
following	Number of following

4.2 Pre-processing

Incomplete, inconsistent, and missing in particular behaviours or patterns, real-world untrained data are common. They may also be full of mistakes. This means that the machine learning algorithm may utilise the data for the model after they have been gathered and prepared.

Disposable Zero - One of the crucial actions in data wrangling is the elimination of null values.

Any machine learning method will suffer from decreased performance and accuracy due to these missing variables. Therefore, prior to using any machine learning technique, it is crucial to clean the dataset of any null values.

Popular techniques of encoding categorical variables include label encoding. Each label is assigned a unique number following the established system of alphabetical order.

4.3 Feature selection technique

The search efficiency of the given adaptive particle swarm optimization (APSO) is higher than that of the traditional particle swarm optimization. Additionally, it is capable of a quicker convergence rate and a global search throughout the whole search space. The APSO has two primary phases. First, a real-time evolutionary state estimate approach is carried out to identify one of the four designated evolutionary stages, namely exploration, exploitation, convergence, and leaping out in each generation by analysing the population distribution and particle fitness. It provides real-time adaptive tuning of search efficiency and convergence rate by adjusting algorithmic parameters such as inertia weight and acceleration coefficients.

This new approach creates an initial population that contains solutions distributed uniformly across all segments. The proposed algorithm's search capability is enhanced by segmenting the entire search space. The proposed model strengthens the swarms' ability to share information. Every swarm expresses interest in or gathers information from other swarms that are more fit than it is. The following steps are carried out for the algorithm:

1. Set up the PSO parameters, including the maximum number of iterations, population size, and initial particle positions and velocities.

2. Define an objective function that calculates the fitness of each particle based on its feature subset.

3. Evaluate the fitness of each particle using the objective function.

4. Identify the particle with the best fitness as the global best particle.

5. Begin the main PSO loop.

6. Update the velocity and position of each particle using the PSO update equations. The velocity update of algorithm is as follows:

Vi = Vi*wi + 1 / rank(i)*rand()*(pbest[i] – Xi) + AdaptivePSO();

Xi = Xi + Vi;

Where,

AdaptivePSO(i)

{

Posx□ 0.0

For each individual k of the population

if pFitness[k] is better than fitness[i] posx□posx + 1 / rank(k) * rand()*(pbest[k]-Xi); if(pos>Vmax)return Vmax:

else return posx;

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *10 September 2023* ***Revised:*** *20 October 2023* ***Accepted:*** *30 October 2023*

}
Where, pfitness[k] represent the best local fitness, fitnessì] indicates the current fitness of swam.TheAdaptivePSO() module gives the direction of the swarm by sharing information with all other individuals that have better fitness,Vmax have been set with a small value to prevent jump.

7. Evaluate the fitness of each particle based on its new position, and update its personal best and the global best particle if necessary.

8. Calculate the swarm diversity using a diversity measure.

9. If the diversity is below a certain threshold, increase the exploration probability and decrease the exploitation probability.

10. Update the PSO parameters based on the exploration and exploitation probabilities.

11. Continue the loop until the maximum number of iterations is reached or a stopping criterion is met.

12. Select the best feature subset found by the PSO algorithm based on the global best particle.

To pick the best features for a model, recursive feature elimination (RFE) first finds the optimal number of features and then iteratively eliminates the features with the lowest predictive value. Using the model's attributes to rank features, RFE then tries to get rid of any dependencies and collinearity in the model by iteratively dropping a few features at a time. For RFE to function, a minimum threshold of valid features must be met, however this threshold is not always known in advance. Cross-validation is used in conjunction with RFE to score several feature subsets and pick the highest scoring collection of features, allowing for the determination of the optimum amount of features.

4.4 Classification

T Classification is a technique used in machine learning that allows classes to be learned and applied to a problem. In the study being suggested, an evaluation is made on whether or not the profiles in question are authentic. As part of the suggested effort, a value of 1 indicates a phoney profile, whereas a value of 0 indicates the opposite. Support Vector Machine, KNN, Random Forest, Logistic Regression, and Extra tree are the classification techniques used for this project.

5. PERFORMANCE EVALUATION

Accuracy, Precision, Recall, and F1-score are the four performance metrics that have been used to estimate the effectiveness of the investigated classification algorithms.Every one of the performance indicators has a specific mathematical formulation, which is described below. Accuracy is the percentage of profiles that were successfully identified as false divided by the total number of profiles, and it reflects how close a forecast is to the actual figure.

Accuracy = (T P + T N) / (T P + T N + F P + F N) [1]

Precision: Determines the percentage of projected phoney profiles that are really positive, based on the accuracy of the categorization system. This value may be calculated using Equation 2.

Precision = TP/ (TP+FP) [2]

Recall: How many right hits were remembered (discovered), or how many genuine positives were retrieved.

Recall = TP / (TP+FN) [3]

F1-Score: When calculating the F1-score, both accuracy and recall are included in. Maximum value it may return is 1, minimum value it can return is 0.

F1-score = (0. 2 ∗ precision ∗ recall)/ (precision + recall) [4]

As shown in figure 2, the dataset contains 576 rows and 12 columns.

	profile pic	nums/length username	fullname words	nums/length fullname	name==username	description length	external URL	private	#posts	#followers	#follows	fake
0	1	0.27	0	0.00	0	53	0	0	32	1000	955	0
1	1	0.00	2	0.00	0	44	0	0	286	2740	533	0
2	1	0.10	2	0.00	0	0	0	1	13	159	98	0
3	1	0.00	1	0.00	0	82	0	0	679	414	651	0
4	1	0.00	2	0.00	0	0	0	1	6	151	126	0
...	...	...	...	...	...	...	...	...	...	...	...	...
571	1	0.55	1	0.44	0	0	0	0	33	166	596	1
572	1	0.38	1	0.33	0	21	0	0	44	66	75	1
573	1	0.57	2	0.00	0	0	0	0	4	96	339	1
574	1	0.57	1	0.00	0	11	0	0	0	57	73	1
575	1	0.27	1	0.00	0	0	0	0	2	150	487	1

576 rows × 12 columns

Figure 2: Data Reading

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: 10 September 2023 *Revised:* 20 October 2023 *Accepted:* 30 October 2023

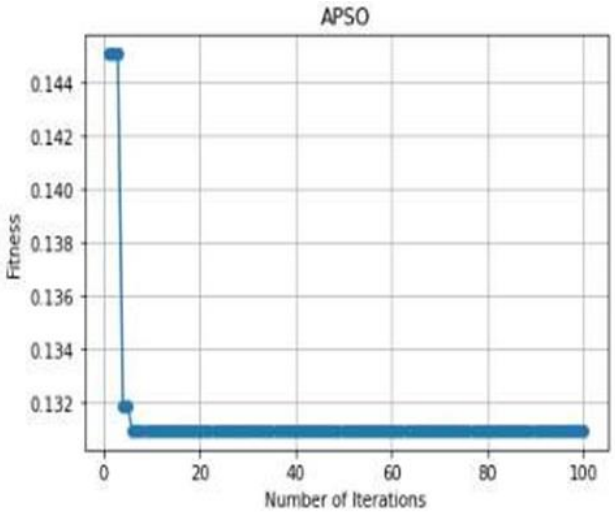

Figure. 3. APSO Feature Selections

Table 2a: Accuracy Analysis of Different Algorithms

Model	Accuracy (in %)		
	Without	RFE	APSO
KNN	84	85	96
RF	92%	85	96
ET	90	90	94

Table 2b: Precision Analysis of Different Algorithms

Model	Precision (in %)		
	Without	RFE	APSO
KNN	84	85	96
RF	92%	85	96
ET	90	90	93

Table 2c: Recall Analysis of Different Algorithms

Model	Recall (in %)		
	Without	RFE	APSO
KNN	84	85	96
RF	92%	85	96
ET	90	90	94

Table 2d: F1-Score Analysis of Different Algorithms

Model	F1-Score (in %)		
	Without	RFE	APSO
KNN	84	85	96
RF	92%	85	96
ET	90	90	94

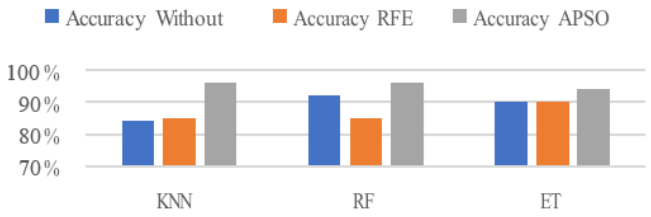

Figure 4: Accuracy Comparative Plot

Figure 5: Precision Comparison Plot

Figure 6: Recall Comparison Plot

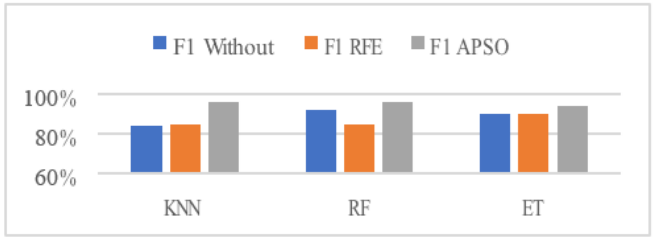

Figure 7: F1-Score Comparison Plot

As shown in table 2, KNN, Random forest and Extra tree works better with APSO as compare with RFE and with 96% of accuracy.

6. CONCLUSION

Fake social networking site profiles are simple to spot using machine learning algorithms. This study proposes a revolutionary machine learning method to recognize phoney Instagram profiles using feature selection method. This research employs a two-stage procedure, first selecting features, and then classifying them. APSO, the original feature selection approach, is used to choose the useful collection of features. And machine learning algorithm, KNN,Random

forest and extra tree are then applied to determine which accounts are fake and which are real. The classification algorithm's performance is heavily influenced by the choices made when choosing its parameters. Thus, KNN and Random Forest has demonstrated its ability to accurately and appropriately identify false profiles.

REFERENCES

[1] Majed Alrubaian, Muhammad Al-Qurishi, Mohammad Mehedi Hassan, and Atif Alamri, A Credibility Analysis System for Assessing Information on Twitter, IEEE (Aug 2018) https://ieeexplore.ieee.org/document/7551232

[2] Van Der Walt, Jan Eloff and Estée. Using machine learning to detect fake identities: bots vs humans. IEEE (2018) https://ieeexplore.ieee.org/document/8265147

[3] Tri HadiahMuliawati, Perdana, Rizal Setya, and Reddy Alexandro. Bot spammer detection in Twitter using tweet similarity and time interval entropy. (Research Gate)

[4] Karatas, Arzum, and SerapŞahin. A Review on Social Bot Detection Techniques and Research Directions. (2015).

[5] Senthil Raja, M., & Arun Raj, L. (2021). Detection of malicious profiles and protecting users in online social networks. Wireless Personal Communications, 1-18.

[6] Bharti, K. K., & Pandey, S. (2021). Fake account detection in twitter using logistic regression with particle swarm optimization. Soft Computing, 25(16), 11333-11345.

[7] G. Sansonetti, F. Gasparetti, G. D'aniello and A. Micarelli, "Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection," in IEEE Access, vol. 8, pp. 213154-213167, 2020, doi: 10.1109/ACCESS.2020.3040604

[8] Uppada, S. K., Manasa, K., Vidhathri, B., Harini, R., &Sivaselvan, B. (2022). Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model. Social Network Analysis and Mining, 12(1), 1-19.

[9] Wu, Y., Fang, Y., Shang, S., Jin, J., Wei, L., & Wang, H. (2021). A novel framework for detecting social bots with deep neural networks and active learning. Knowledge-Based Systems, 211, 106525.

[10] Sahoo, S. R., & Gupta, B. B. (2021). Real-time detection of fake account in twitter using machine-learning approach. In Advances in computational intelligence and communication technology (pp. 149-159). Springer, Singapore.

[11] Rathore, P., Soni, J., Prabakar, N., Palaniswami, M., & Santi, P. (2021). Identifying groups of fake reviewers using a semisupervised approach. IEEE Transactions on Computational Social Systems, 8(6), 1369-1378.

[12] Fang, Y., Wang, H., Zhao, L., Yu, F., & Wang, C. (2020). Dynamic knowledge graph based fake-review detection. Applied Intelligence, 50(12), 4281-429.

[13] Preethi Harris; J Gojal; R Chitra; S Anithra," Fake Instagram Profile Identification and Classification using Machine Learning", 2021 2nd Global Conference for Advancement in Technology (GCAT) | 978-1-6654-1836-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GCAT52182.2021.9587858.

[14] Muñoz, S. D., & Pinto, E. P. G. (2020, December). A dataset for the detection of fake profiles on social networking services. In 2020 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 230-237). IEEE.

[15] Chaudhary, A., Mittal, H., & Arora, A. (2019, February). Anomaly detection using graph neural networks. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (pp. 346-350). IEEE.

[16] Khaled, S., El-Tazi, N., & Mokhtar, H. M. (2018, December). Detecting fake accounts on social media. In 2018 IEEE international conference on big data (big data) (pp. 3672-3681). IEEE.

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

[18] Laleh, N., Carminati, B., & Ferrari, E. (2016). Risk assessment in social networks based on user anomalous behaviors. IEEE Transactions on Dependable and Secure Computing, 15(2), 295-308.

[19] Latha, P., Sumitra, V., Sasikala, V., Arunarasi, J., Rajini, A. R., &Nithiya, N. (2022, March). Fake Profile Identification in Social Network using Machine Learning and NLP. In 2022 International Conference on Communication, Computing, and Internet of Things (IC3IoT) (pp. 1-4). IEEE.