International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT23903126

Novel Approaches to Detect Phony Profile on Online Social Networks (OSNs) Using Machine Learning

Ms Farah Shan*[1], Versha Verma[2], Apoorva Dwivedi[3], Dr. Yusuf Perwej[4], Ashish Kumar Srivastava[5]

[1*]Assistant Professor, Department of Computer Science & Engineering, Maharana Pratap College of Engineering, Kanpur, India

[2]Research Scholar, Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, U.P, India

[3]Assistant Professor, Department of Computer Science & Engineering, Invertis University, Bareilly, U.P, India

[4]Professor, Department of Computer Science & Engineering, Ambalika Institute of Management & Technology, Lucknow, U.P, India

[5]Assistant Professor, Department of Computer Science & Engineering, Shri Ramswaroop Memorial University, Lucknow, U.P, India

ARTICLEINFO

Article History:

Accepted: 03 June 2023
Published: 20 June 2023

Publication Issue
Volume 9, Issue 3
May-June-2023

Page Number
555-568

ABSTRACT

Currently, almost everyone spends more time on online social media platforms engaging with and exchanging information with people from all over the world, from children to adults. Our lives are greatly influenced by social media sites like Twitter, Facebook, Instagram, and LinkedIn. The social network is evolving into a well-liked platform for connecting with individuals across the globe. Social media platforms exist as a result of the enormous connectivity and information sharing that the internet has made possible. Social media's rising popularity has had both beneficial and detrimental consequences on society. However, it also has to deal with the issue of bogus profiles. False profiles are often constructed by humans, bots, or cyborgs and are used for phishing, propagating rumors, data breaches, and identity theft. Thus, we are emphasizing in this post the significance of setting up a system that can identify false profiles on social media networks. To illustrate the suggested concept of machine learning-based false news identification, we used the Twitter dataset for phony profile detection. The suggested model involves pre-processing to improve the dataset's quality and minimize its dimensions by modifying its contents and features. To forecast the bogus profiles, the widely used machine learning algorithms are used.

Keywords : Fake Profile, Twitter, Phony Identities, Detection, Social Network Analysis, Kaggle Dataset, Machine Learning.

I. INTRODUCTION

Social media has a big impact on how we live now. Social networking services are widely utilized in daily life as a medium for communication between individuals [1]. The constant sharing of information and everyday activities by users of this site draws a lot of people to it. From 2007 through 2023, Facebook or Twitter or Twitter will become more popular. Users may add friends and exchange a variety of information, including personal [2], social, economic, educational, political, and corporate information. Additionally,

they may communicate daily by exchanging images, videos, and other materials. Some users, meanwhile, don't utilize these websites objectively [3].

Twitter has 695.1 million monthly users and 289 million daily active users. Additionally, Facebook adds six new members per second, or around 550,000 new users every day. On Twitter, a tonne of information is shared every day [4]. One may access the most well-liked articles, the newest hashtag, news, and details about their most recent vacation on Twitter. People can respond, like, comment, trade ideas, and express their opinions within the given 280 characters. There are frequently rumors, but there are also serious concerns that are looked at. These rumors cause tension among the various social groups. Recent revelations have raised questions about privacy [5], exploitation, cyberbullying, and incorrect information. False profiles are used in all of these actions. We view accounts that were formerly genuine but were subsequently hacked as fraudulent accounts. Such accounts that contain personal data that does not belong to the individual who established the account are sometimes referred to as phony accounts. A falsified account is one that incorporates made-up personal information. Deception, the dissemination of potentially harmful material, and a desire to meet new people are some of the causes of establishing phony users. These bogus documents are difficult to spot [6].

A variety of machine learning algorithms, including neural networks (NN) [7], naive bayes, Markov models, and Bayesian networks, have been developed to identify bogus accounts on social networking sites. To find criminal users who could deceive individuals, ML [8] algorithms were used, although platforms had an equal number of profits and losses. It all depends on who is using it and what their objectives are. Social media may be useful for interacting with people, studying, having fun, and learning new things [9]. Those with bad intentions, however, could harm other individuals. One of the problems is creating a phony account and utilizing it to harass individuals or disturb others. A neural network is made up of several linked processing components. It makes choices similarly to a human brain. For classification, supervised machine learning algorithms called support vector machines (SVM) [10] are employed. To categorize the data, it locates the hyper plane. Based on a variety of account criteria, neural networks [11] and SVM are appropriate for detecting phony accounts on social networking sites since they can take a significant quantity of random input [12]. On the Bayes theorem, the Naive Bayes classifier [13] is based. It forecasts the likelihood that a particular variable belongs to a specific class.

II. RELATED WORK

Fake accounts are expanding as a result of the popularity of social media sites. The creation of such a phony account or identity is done for a variety of nefarious reasons [14]. The use of these false personas is particularly detrimental to society and can lead to participation in a number of online and offline crimes [15]. The method put forward by [16] is an intriguing solution to the issue of identifying bogus accounts in social networks. In order to build a detection system, they adopt a point of view based on the victims of these phony accounts in this article. In this instance, they described "Integro" as software that was built on a graph-based algorithm with a raking scheme while also using some end-user activities. They claim that the user can see everything that happens throughout this procedure, and that "Integro works on social networks that only approve bidirectional friendships." To be more precise, they utilize user behaviors and easily accessible user data. After obtaining this data, they employ it to train a victim classificatory that enables "Integro" to identify phony accounts from that starting point and locate future victims. According to Chu et al. [17], Twitter accounts run by humans, bots, or cyborgs (i.e., bots and people working together) should be distinguished. An Orthogonal Sparse Bigram (OSB) text classifier that employs pairs of words as features is

used to identify spamming records as part of the formulation of the detection issue. The gathers the Twitter dataset for research [18] from April 2010 until July 2010. The dataset is very unbalanced, with a fake-to-real user ratio of 1:10; this imbalance is corrected to an equal ratio by deleting tweets that are not in standard English. Three categories within a month, within two months, and within four months are used to categories tweets.

Many of these issues are concentrated on the social network Twitter, where bogus and spam accounts are heavily policed. The suggestion in [19] takes a different methodological tack. In this instance, they focused their efforts on grouping spam accounts together by organizing these groupings according to shared features. Spammer profiles may also be automated, false accounts with the purpose of spreading false information who follow several accounts to broaden their audience. They utilized a crawler to collect the necessary data over the course of two months by looking for spam-triggering terms in a large number of actual tweets. After that, they used Principal Component Analysis to extract 15 features, used the k-means method to locate groups of accounts that shared the same spam tweet semantics, and used this information to train three separate classifiers. Twitter supporter markets are analyzed by Stringhini et al. [20]. They list the characteristics of Twitter aficionados who promote and classify the patrons of various company sectors. According to the authors, there are primarily two types of profiles that pursue the "client": hacked accounts and phony accounts (also known as "sybils"), whose owners do not assume that the number of their followers is growing. Clients of follower markets may include politicians or celebrities who want to appear to have a larger fan following, or they may include crooks who want to make their profile appear more real so they can transmit malware and spam more quickly. The [21] carefully verifies 5,386 real followers and 13,000 bogus followers that were acquired. The attributes used by an ML model to distinguish between fake and real users include the number of followers, follower sees, favorites, followers, and listed users. For false and real users, the values of Cumulative Distribution Frequency are extremely distinct.

In [22], which also covers other feature reduction and selection strategies, a support vector machine-based neural network solution to fraudulent account identification in twitter is covered in length. The method described in [23], where they examine a method to detect hostile bots that ultimately are phony accounts, is another suited strategy for detecting fraudulent accounts and fake profiles. They claim that the issue with these bots as phony accounts is that they may greatly increase their reach by publishing false news and forming fictitious connections with actual people. This study established a methodology where each account is taken into consideration on its own, and the veracity of each tweet and follow action is also examined. The work was only focused on Twitter and used Twitter's URL features for gathering information. They utilize a Learning Automata algorithm to evaluate whether an account is a bot or a user of the social network after collecting the necessary information from the URLs. In this instance, it may be seen of as a technique to employ machine learning to identify harmful bots, but with a completely different approach to information gathering. Thomas et al.'s [24] investigation looks at black market profiles used to spread Twitter spam.

An algorithm was presented by Xiao et al. to identify groups of phony accounts on social media before they do harm or interact with real people [25]. It uses the k-means clustering method to group the accounts into groups, identifies cluster level properties [26], scores the accounts in each group, and assigns labels to the groups based on the group's average score. In order to identify false accounts in social networks, a model built on the similarity between the user's buddy networks is suggested in [27]. The adjacency matrix of the relevant social network graph is used to generate similarity

metrics like mutual friends, cosine, Jaccard, L1-measure, and weight similarity. The suggested model is assessed using the Twitter dataset and the SVM with medium Gaussian. The author of [28] offers a graph-based strategy to reduce bogus accounts in social networks. Similar to this, the effectiveness of a collection of approaches is verified. To detect false profiles in online social networks, an ML [29] and NLP system is described in [30].

Additionally, the NB algorithm and SVM classifier are included to improve the accuracy of bogus profile detection. [31] presents a graph-based strategy that gathers anomalous characteristics from several accounts to create a graph, which is then used to identify the densely linked person and locate fraudulent profiles. The paper discusses a study that looked at identifying phony accounts made by people as opposed to bots in [32]. Research is being done to see if data from earlier research to identify bot accounts can be effectively used to identify phony human accounts. Engineered traits that had previously been utilized to successfully identify fraudulent accounts made by bots are added to a corpus of human accounts. Comparable to this, in [33], a deep neural network-based solution is given that identifies the bogus profiles with the region attributes by using text and user features. In order to identify fraudulent profiles, the technique [34] takes into account how similar friends are across various accounts. Additionally, a two-layer strategy is proposed that categorizes profiles based on meta data [35] and topological information.

III. Problem Definition

The development of phony social media profiles is rising together with the number of persons utilizing OSNs, as can be shown. The primary driving force behind the identification of these fake accounts is the fact that they are typically created to engage in cyber extortion or to commit cybercrimes covertly or under an alias. As a result, the rate of cybercrime has significantly increased over the past year. Additionally, the creator of phony accounts occasionally seeks to profit off of people's goodwill by disseminating misleading information or making fraudulent [36] claims in order to steal money from unwitting victims. Additionally, individuals are establishing several accounts that do not belong to anybody and were simply made to increase the number of votes in online voting systems and online games in an effort to earn referral bonuses. In-depth analysis is done on the issue of fraudulent account identification in social networks. Numerous approaches have been addressed in literature; each has advantages and disadvantages. They endure suffering, nonetheless, in order to do better. As a result of the literature review, it is possible to adapt profile level trust, prior information [37] level confidence, and profile level trust to solve the issue of fake account detection. The effectiveness of false detection and access restriction may be enhanced by implementing the trust assessment with different degrees [38].

IV. Why Do Phony Profiles Get Made?

In situations of Advanced Persistent Threat, fake identities on social media are frequently used to transmit malware or a link to it. Additionally, they are utilized in harmful actions like sending spam [39] and junk emails, as well as in some programmers to advertise them by artificially increasing the number of users. A report claims that a Facebook-supported gaming application offers rewards to users/players who invite an increasing number of their friends to join the game. People therefore create false accounts out of a need for incentives. A huge number of false accounts may [40] be made by politicians or celebrities in an effort to highlight their vast fan bases, or they may be created by cybercriminals in an effort to make their accounts appear more authentic. To increase the rating of a product or application [41], the owner or corporation may use applications like online surveys

where fraudulent accounts are employed to obtain greater feedback.

V. Machine Learning

Applications for machine learning [42] that resolve problems and automate in numerous sectors have increased at an extraordinary rate. This is mostly due to the growth in the amount of information that is available, significant improvements in machine learning [43] techniques, and advancements in computing power. There is no denying that machine learning has been used to a variety of complex and modern network management and operation problems. For specialized networking industries or particular network technologies, many Machine Learning surveys have been undertaken. Data screening and inference are made possible by machine learning. It encompasses applying knowledge and enhancing it via experience and time, going beyond merely acquiring or obtaining information [44]. Machine learning's main goal is to find and utilize hidden patterns in "training" data. It is possible to categories or map new data to existing categories using the learnt patterns. All artificial intelligence subjects fall under the umbrella of machine learning, which calls for a multidisciplinary approach that incorporates knowledge of probability theory, mathematics, trends detection, dynamic modelling (DM), cognitive psychology, adaptive control, computational neuroscience, and theoretical computer science.

5.1 C4.5

Quinlan developed C4.5, a decision tree-generating method, to create Classification Models (Quinlan, 1993). To get around these drawbacks, the fundamental ID3 algorithm was extended. The ID3 method was enhanced by the C4.5 algorithm, which made a number of changes. In order to increase prediction accuracy, C4.5 [45] actually performs a recursive partition of observations in branches to build a tree. To do this, a variable and matching threshold for the variable that divides the input data into two or more subgroups are found using mathematical techniques. This process is continued until the entire tree has been built at each leaf node.

5.2 Support Vector Machine (SVM)

By supplying relevant data with a feature and creating a classifier that shines on hidden data, support vector machine classification seeks to distinguish between two groups. The most basic type of support vector machine is maximum range classification. Binary classification using linearly separable training data is a typical approach to the main classification issue.

5.3 Artificial Neural Network (ANN)

Artificial neural networks are computer simulations of organic neural networks. Another name for a neural network is an ANN. The idea behind ANN [46] is mostly inspired by biology, where the neural network is crucial to the functioning of the human body. The human body is used for practice in the neural network. A neural network may be conceptualized as a group of linked input/output units, each with a distinct weight.

5.4 Random Forest

A supervised learning approach called random forest is employed for both classification and regression. But it is mostly employed for categorization issues. As we all know, trees make up a forest, and stronger forests result from having more trees. Similar to this, the random forest method builds decision trees on data samples, obtains [47] predictions from each one, and then uses voting to determine the optimal option. Because it averages the outcomes, the ensemble technique is superior than a single decision tree in that it lessens over-fitting.

5.5 Decision Tree (DT)

One of the most effective supervised learning methods for both classification and regression applications is the decision tree. It creates a tree structure resembling a flowchart where each internal node represents a test on an attribute, each branch a test result, and each leaf node (terminal node) a class label. A stopping requirement, such as the maximum depth of the tree or the least number of samples needed to split a node, is reached by repeatedly separating the training data into subsets depending on the values of the attributes. One of the most potent algorithms is this one. Additionally, Random Forest, one of the most potent machine learning algorithms, uses it to train on various subsets of training data.

5.6 K- Nearest Neighbor (KNN)

By concentrating on the closest neighbor whose value is already known, the Nearest Neighbor approach locates the unknown data point. Try to locate the nearest spot. The Nearest Neighbor mechanism can be disassembled in one of two ways. Structure and function are less frequently utilized when classification methods based on nearest neighbors are applied. In the scheme, K-NN [48] is referred to as a less approach. To determine how many NN must be verified for each sample data point in the class description, the KNN method employs the NN for the value of k. NN techniques may be divided into two categories: KNN-dependent structure and KNN-less structure.

5.7 Naive Bayes (NB)

The naive Bayes classifier is a probabilistic classifier built on the Bayes theorem that operates on the presumption that each feature contributes equally and independently to the target class [49]. The NB classifier makes the assumption that each feature is separate from the others and does not interact, therefore each feature independently and equally influences the likelihood that a sample belongs to a given class. The NB classifier works well on big datasets with high dimensionality and is computationally quick and easy to deploy. The NB classifier is noise-resistant and well-suited for real-time applications.

VI. Dataset

In the current generation, everyone's social life is now entwined with online social networks. It has been simpler to add new friends and stay in touch with them and their updates. Online social networks have an influence on a variety of fields, including research, education, grassroots activism, commerce, and employment [50]. These online social networks have been the subject of research to see how they affect people. This Twitter Dataset's main objective is to aid in the study of "real-setting" deepfake social media text detection. As each sample in this dataset is labelled with the appropriate text generation technique ('human', 'GPT-2', 'RNN', or 'Others'), we are able to comprehend how our detector behaves with respect to each generative method in addition to [51] evaluating the general accuracy of your deep-fake text detector. Starting the research depicted in figure 1 requires a collection of deep-fake social media postings [52]. The tweets from each account pair (human and bot/s) were randomly picked based on the least productive in order to provide a dataset that was balanced across both categories ('human' and 'bot'). For instance, to obtain the equivalent quantity of data, X tweets were selected from the set of N tweets if the bot (human) account had X tweets and the matching human (bot) account had N tweets (with N>X). 25,836 tweets in total (half human, half bot) were gathered. the screen_name of the account, indicating that tweets from each account were sent to the train, validation, and test sets. First, the entire dataset was divided into train and test sets, with the train set receiving 90% of the total tweets and the test set receiving the remaining 10%. The validation set was created by selecting 10% of the training tweets from the train set.

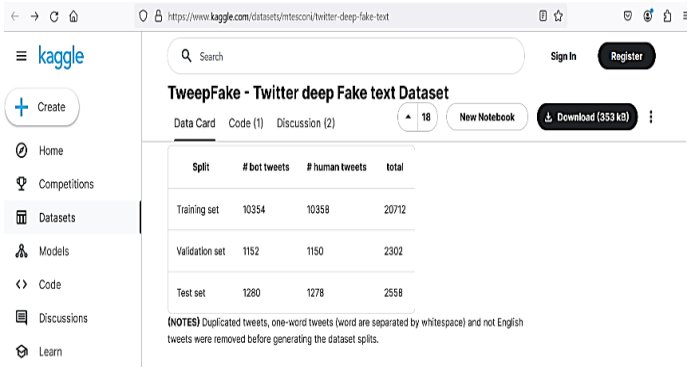

Figure 1. The Twitter Deep Fake Text Dataset

VII. Proposed Approach

We have provided an experimental model to construct the necessary machine learning model for the early-stage fake profile identification approach. Figure 2 illustrates this concept. This section provides a description of the proposed model's specifics. With the supplied false profile dataset, we provide supervised learning method assessment [52]. The Twitter test samples are contained in this.csv file (with comma-separated values). There is at least one tweet in this collection from each of the 40 accounts. Every Twitter account only employs one generating technique. There are a total of 29 characteristics in the dataset.

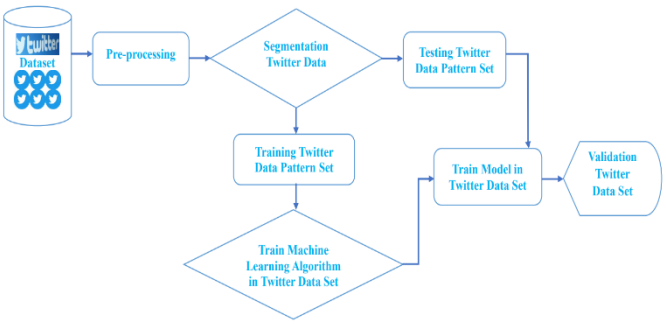

Figure 2. The Proposed Twitter Dataset Module

There are 29 attributes and one class label in the dataset, which is a sizeable number of attributes. We are working to simplify and hone the core characteristics. Identification of the individual or profile is accomplished using the characteristics ID, Name, and screen_name. As a result, we only choose one of these three traits from the group here; we choose the ID in comparison to the other two attributes [53]. Moreover, for profile identification, the properties statuses_count, followers count, and favorites count are crucial. The dataset also includes a listed count, which isn't very useful in our opinion, thus we minimize this property. In addition, it's crucial to know how old a profile is, therefore the characteristic created at converts the date and time into the number of days. Since a social network profile will undoubtedly have a unique URL, the URL element is removed. We take time zone into consideration in comparison to the other two variables since the attributes lang, time_zone, and location may all be integrated into one. In this case, the two properties default profile and default_profile_image is combined into one as a true or false Boolean. Additionally, the attributes geo_enabled, profile_image_url, and profile_banner_url are converted into Booleans. likewise, the Boolean value for profile_background_image_url_https is combined into one. Following the conversion of profile_background_tile into a Boolean value, the non-essential characteristics profile_sidebar_fill_color, profile background _image URL, profile_link_color, and utc_offset is also eliminated [54]. Additionally, Protected, Verified, and Description are utilized as Booleans. The following characteristic, Updated, is used to determine how recently a profile was updated [55]. Dataset is eliminated as a superfluous attribute. Preprocessing is done to clean up the data and raise its quality in order to enhance learning outcomes. The preprocessed data is then utilized to make decisions. In order to construct training and testing data, previously organized data is retained in a local database. In this case, the training set consists of 90% of the data instances that were randomly chosen. Furthermore, a fourfold test dataset is constructed utilizing the idea of n cross validation. Four folds of the 10% of randomly chosen data are utilized to evaluate the data mining methods. Additionally, experiments also make advantage of the 70-30% ratio. The experimental system shown in figure 2 was created to learn about data patterns and appropriately identify the data. Three

supervised learning algorithms the C4.5 decision tree, the ANN (Artificial Neural Network) [56], and the KNN (k-nearest neighbor) algorithm are taken into consideration in order to train the system. These models take training datasets into account and generate trained models appropriately. For instance, ANN creates opaque models, whereas KNN and C4.5 algorithms create transparent models. After mastering these models, applying them to the fourfold created test dataset [57], performing classification, and generating accurate classification results for the test datasets. The results of this model's performance are shown in table 1 and table 2 below.

Table 1. The Model's Performance of Classification Outcomes for the Test Datasets

Machine Learning Algorithms	Performance Summary for 90% - 10	
	Imperfection Rate	Precision
k-Nearest Neighbors (KNN)	13.33%	86.67%
C4.5 Decision Tree	10.54%	89.46%
Artificial Neural Network (ANN)	1.81%	98.19%

Table 2. The Model's Validation of Classification Outcomes for the Test Datasets

Machine Learning Algorithms	Validation Summary for 70% - 30%	
	Imperfection Rate	Precision
k-Nearest Neighbors (KNN)	18.13%	81.87%
C4.5 Decision Tree	17.44%	82.56%
Artificial Neural Network (ANN)	5.31%	94.69%

VIII. Results Assessment

This section describes and contrasts the machine learning algorithm's performance in the context of identifying bogus profiles. For their comparative performance research, the performance metrics listed below are measured.

8.1 Precision

The measurement of algorithm classification correctness can be used to explain precision [58]. The ratio of all patterns properly categorized to all patterns needing Categorized can be used to quantify that. The equation below may also be used to express it.

$$\text{Precision} = \frac{\text{All Patterns Properly Categorized}}{\text{All Patterns Needing Categorized}} \times 100$$

The figure 3 displays the algorithms' precision for the overall precision. Here, the precision of the algorithm is expressed as a percentage (%). According to the results [59], the algorithm's performance is shown to be successful with a 90-10 ratio as opposed to an 70-30 ratio.

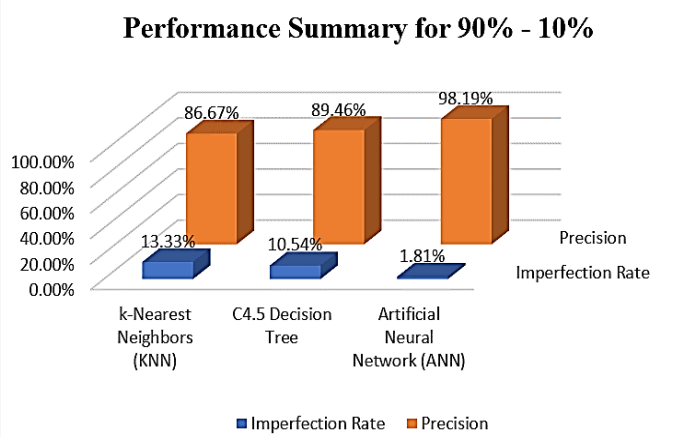

Figure 3. The Model's Performance Summary

In addition, we discovered that the ANN perform better than the other employed algorithms. As a result, both methods may be taken into consideration for the suggested data model's implementation in the near future.

8.2 Imperfection Rate

The algorithm's imperfection rate serves as a performance indicator by showing how frequently [60]

the algorithm is misclassified. This equation may be used to calculate that.

$$\textbf{Imperfection Rate} = 100 - \textbf{Precision}$$

The figure 4 displays the imperfection rate of the implemented methods. The effectiveness is evident for both categories of validation ratios.

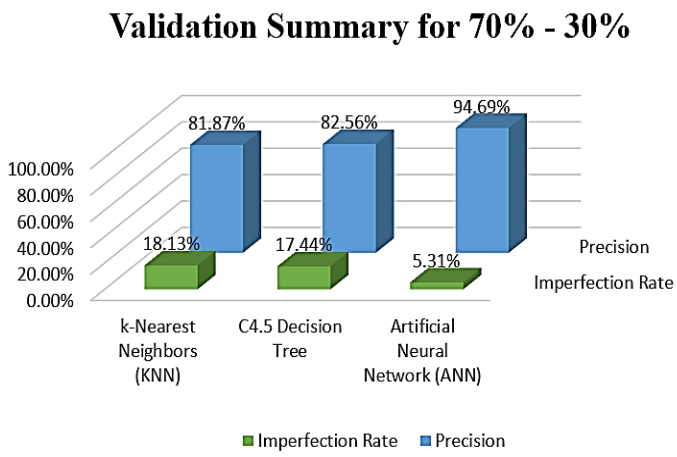

Figure 4. The Model's Validation Summary

The validation ratio and imperfection rate (%) are displayed on the respectively, to demonstrate how well the algorithm performed. The system's performance demonstrates that the ANN [61] report lower error rates than other methods [62].

IX. Conclusion

The majority of people in the world today from young children to elderly individuals spend a significant amount of time on online social networks (OSNs), communicating and exchanging information with others there. Many people have a tendency to abuse the social network platform because of the extensive interconnectivity and extensive information sharing offered by OSN. Making many false identities to get various forms of unethical gains, such as targeting a specific user or trying any other criminality, is one method that individuals abuse social networks. They communicate with one another, share information, plan events, and even manage their own online businesses using online social networks measured. Attackers and imposters have been lured to OSNs because of their rapid expansion and the enormous amounts of personal data they gather from its users, which they exploit to disseminate disruptive activities, steal personal information, and publish false material. The purpose of this study is to examine the strategies and procedures employed for spotting false profiles on various social media sites. Three machine learning algorithms are used: the ANN, C4.5 decision trees, and KNN. Additionally, the fourfold cross-validation technique is employed to get performance, and performance in terms of accuracy and imperfection rate are measured. Two different validation ratios, 90-10% and 70-30%, were employed. The table reports the techniques' performance summaries. We developed a classification approach that is efficient and accurate using the findings we got, and we then utilized it to build a better model for detecting bogus profiles.

X. REFERENCES

[1]. M. Tsikerdekis and S. Zeadally, "Multiple account identity deception detection in social media using nonverbal behavior", IEEE Transactions on Information Forensics and Security, vol. 9, no. 8, pp. 1311-1321, 2014

[2]. Van Der Walt, Estée and Jan Eloff, "Using machine learning to detect fake identities: bots vs humans", IEEE access, vol. 6, pp. 6540-6549, 2018

[3]. Adikari and K. Dutta, "Identifying fake profiles in linkedin", PACIS, pp. 278, 2014

[4]. Yahao Zhang, Ruimin Hu, Dengshi Li and Xiaochen Wang, Fake Identity Attributes Detection Based on Analysis of Natural and Human Behaviors., vol. 8, pp. 78901-78911, 2020

[5]. Yusuf Perwej, Prof. (Dr.) Syed Qamar Abbas, Jai Pratap Dixit, Nikhat Akhtar, Anurag Kumar

Jaiswal, “A Systematic Literature Review on the Cyber Security”, International Journal of Scientific Research and Management (IJSRM), ISSN (e): 2321-3418, Volume 9, Issue 12, Pages 669 - 710, 2021, DOI: 10.18535/ijsrm/v9i12.ec04

[6]. C. Yang, R. Harkreader and G. Gu, "Empirical evaluation and new design for fighting evolving twitter spammers", IEEE Transactions on Information Forensics and Security, vol. 8, no. 8, pp. 1280-1293, 2013

[7]. Asif Perwej, Prof. (Dr.) K. P. Yadav, Prof. (Dr.) Vishal Sood, Dr. Yusuf Perwej, “ An Evolutionary Approach to Bombay Stock Exchange Prediction with Deep Learning Technique”, IOSR Journal of Business and Management (IOSR-JBM), e-ISSN: 2278-487X, p-ISSN: 2319-7668, USA, Volume 20, Issue 12, Ver. V, Pages 63-79, 2018, DOI: 10.9790/487X-2012056379

[8]. Yusuf Perwej, “The Bidirectional Long-Short-Term Memory Neural Network based Word Retrieval for Arabic Documents”, Transactions on Machine Learning and Artificial Intelligence (TMLAI), Society for Science and Education, United Kingdom (UK), ISSN 2054-7390, Volume 3, Issue 1, Pages 16 - 27, 2015, DOI: 10.14738/tmlai.31.863

[9]. Yusuf Perwej, Firoj Parwej, Mumdouh Mirghani Mohamed Hassan, Nikhat Akhtar, “The Internet-of-Things (IoT) Security: A Technological Perspective and Review”, International Journal of Scientific Research in Computer Science Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5, Issue 1, Pages 462-482, 2019. DOI: 10.32628/CSEIT195193

[10]. Nikhat Akhtar, Dr. Hemlata Pant, Apoorva Dwivedi, Vivek Jain, Dr. Yusuf Perwej, “A Breast Cancer Diagnosis Framework Based on Machine Learning” , International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Print ISSN: 2395-1990 , Online ISSN : 2394-4099, Volume 10, Issue 3, Pages 118-132, 2023, DOI: 10.32628/IJSRSET2310375

[11]. Venkata K. S. Maddala, Dr. Shantanu Shahi, Dr. Yusuf Perwej, H G Govardhana Reddy, “Machine Learning based IoT application to Improve the Quality and precision in Agricultural System”, European Chemical Bulletin (ECB), ISSN: 2063-5346, SCOPUS, Hungary, Volume 12, Special Issue 6, Pages 1711 – 1722, May 2023, DOI: 10.31838/ecb/2023.12.si6.157

[12]. C. Xiao, D. M. Freeman and T. Hwa, "Detecting clusters of fake accounts in onlinesocial networks", Proceedings of the 8th ACM Workshopon Artificial Intelligence and Security, pp. 91-101, 2015

[13]. Nikhat Akhtar, Devendera Agarwal, “An Efficient Mining for Recommendation System for Academics”, International Journal of Recent Technology and Engineering (IJRTE), Volume-8, Issue-5, Pages 1619-1626, 2020, DOI: 10.35940/ijrte.E5924.018520

[14]. Y. Boshmaf, D. Logothetis, G. Siganos, J. Lería, J. Lorenzo, M. Ripeanu, K. Beznosov, H. Halawa, "Íntegro: Leveraging victim prediction for robust fake account detection in large scale osns", Computers & Security, vol. 61, pp. 142-168, 2016

[15]. Asif Perwej, Kashiful Haq, Yusuf Perwej, “Blockchain and its Influence on Market”, International Journal of Computer Science Trends and Technology (IJCST), ISSN 2347 – 8578, Volume 7, Issue 5, Pages 82- 91, 2019, DOI: 10.33144/23478578/IJCST-V7I5P10

[16]. M. Al-Qurishi, S. M. M. Rahman, M. S. Hossain, A. Almogren, M. Alrubaian, A. Alamri, M. Al-Rakhami, and B. Gupta, “An efficient key agreement protocol for sybil-precaution in online social networks,” Future Generation Computer Systems, vol. 84, pp. 139–148, Jul. 2018

[17]. Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. "Who is tweeting on Twitter: human, bot, or cyborg?." In Proceedings of the 26th annual computer security applications conference, pp. 21- 30. ACM, 2010

[18]. M. M. Swe and N.N. Myo, "Fake Accounts Classification on Twitter", International Journal of Latest Engineering and Management Research (IJLEMR), vol 3, no. 6, pp. 141-146, June 2018

[19]. K. S. Adewole, T. Han, W. Wu, H. Song, and A. K. Sangaiah, "Twitter spam account detection based on clustering and classification methods," The Journal of Supercomputing, vol. 76, no. 7, pp. 4802–4837, Oct. 2018

[20]. Stringhini, Gianluca, Gang Wang, Manuel Egele, Christopher Kruegel, Giovanni Vigna, Haitao Zheng, and Ben Y. Zhao. "Follow the green: growth and dynamics in twitter follower markets." In Proceedings of the 2013 conference on Internet measurement conference, pp. 163-176. ACM, 2013

[21]. A. Khalil, H. Hajjdiab and N. Al-Qirim, "Detecting Fake Followers in Twitter: A Machine Learning Approach" International Journal of Machine Learning and Computing, vol. 7, no. 6, pp. 198-202, December 2017

[22]. S. Khaled, N. El-azi, and H. M. Mokhtar, "Detecting fake accounts on social media", In: Proc. of IEEE International Conference on Big Data, pp. 3672-3681, 2018

[23]. R. R. Rout, G. Lingam, and D. V. L. N. Somayajulu, "Detection of malicious social bots using learning automata with URL features in twitter network," IEEE Transactions on Computational Social Systems, vol. 7, no. 4, pp. 1004–1018, Aug. 2020

[24]. Thomas, Kurt, Damon McCoy, Chris Grier, Alek Kolcz, and Vern Paxson. "Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse." In Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13), pp. 195-210, 2013

[25]. Xiao C, Freeman DM, Hwa T.,"Detecting clusters of fake accounts in online social networks", In: Proceedings of the 8th ACM workshop on artifcial intelligence and security, AISec'15. Association for Computing Machinery, New York, NY, USA, pp 91–101, 2015

[26]. Shweta Pandey, Rohit Agarwal, Sachin Bhardwaj, Sanjay Kumar Singh, Dr. Yusuf Perwej, Niraj Kumar Singh, "A Review of Current Perspective and Propensity in Reinforcement Learning (RL) in an Orderly Manner", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 9, Issue 1, Pages 206-227, January-February-2023, DOI: 10.32628/CSEIT2390147

[27]. M. Mohammadrezaei, M. E. Shiri, and A. M. Rahmani, "Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms", Security and Communication Networks, pp. 1-8, 2018

[28]. M. Conti, R. Poovendran, and M. Secchiero, "Facebook: detecting fake profiles in online social networks", In: Proc. of the 2012 International Conference on Advances in Social Networks Analysis and Mining, pp. 1071-1078, 2012

[29]. Shobhit Kumar Ravi, Shivam Chaturvedi, Dr. Neeta Rastogi, Dr. Nikhat Akhtar, Yusuf Perwej, "A Framework for Voting Behavior Prediction Using Spatial Data", International Journal of Innovative Research in Computer Science & Technology (IJIRCST), ISSN: 2347-5552, Volume 10, Issue 2, Pages 19-28, 2022, DOI: 10.55524/ijircst.2022.10.2.4

[30]. P. S. Rao, J. Gyani, and G. Narsimha, "Fake Profiles Identification in Online Social Networks Using Machine Learning and NLP", International Journal of Applied Engineering Research, vol. 13, no. 6, pp. 4133-4136, 2018

[31]. D. Yuan, Y. Miao, N. Z. Gong, Z. Yang, Q. Li, D. Song, Q. Wang, and X. Liang, "Detecting fake accounts in online social networks at the time of registrations", In: Proc. of the 2019 ACM SIGSAC Conference on Computer and Communications Security, pp. 1423-1438, 2019

[32]. E. V. D. Walt and J. Eloff, "Using Machine Learning to Detect Fake Identities: Bots vs Humans", IEEE Access, vol. 6, pp. 6540-6549, March 2018

[33]. Y. Liu and Y. F. B. Wu, "Fned: a deep network for fake news early detection on social media", ACM Transactions on Information Systems, Vol. 38, pp. 1-33, 2020

[34]. M. Mohammadrezaei, M. E. Shiri, and A. M. Rahmani, "Identifying fake accounts on social networks based on graph analysis and classification algorithms", Security and Communication Networks, 2018

[35]. Firoj Parwej, Nikhat Akhtar, Dr. Yusuf Perwej, "A Close-Up View About Spark in Big Data Jurisdiction", International Journal of Engineering Research and Application (IJERA), ISSN: 2248-9622, Volume 8, Issue 1, (Part -I1), Pages 26-41, 2018, DOI: 10.9790/9622-0801022641

[36]. Liu Y, Ji S, Mittal P (2016) Smartwalk: enhancing social network security via adaptive random walks. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, CCS'16. Association for Computing Machinery, New York, NY, USA, pp 492–503

[37]. Yusuf Perwej, Md. Husamuddin, Dr. Majzoob K.Omer, Bedine Kerim, "A Comprehend the Apache Flink in Big Data Environments" , IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, P-ISSN: 2278-8727, USA, Volume 20, Issue 1, Ver. IV, Pages 48-58, 2018 , DOI: 10.9790/0661-2001044858

[38]. Jia J, Wang B, Gong NZ Random walk based fake account detection in online social networks. In: 2017 47th annual IEEE/ IFIP international conference on dependable systems and networks (DSN), pp 273–284, 2017

[39]. Xiao C, Freeman DM, Hwa T,"Detecting clusters of fake accounts in online social networks", In: Proceedings of the 8th ACM workshop on artifcial intelligence and security, AISec'15. Association for Computing Machinery, New York, NY, USA, pp 91–101, 2015

[40]. Chang AB," Using machine learning models to detect fake news, bots, and rumors on social media", Thesis and Dissertations, Arizona State University, Library, 2021

[41]. Firoj Parwej, Nikhat Akhtar, Dr. Yusuf Perwej, "An Empirical Analysis of Web of Things (WoT)", International Journal of Advanced Research in Computer Science (IJARCS), ISSN: 0976-5697, Volume 10, No. 3, Pages 32-40, 2019, DOI: 10.26483/ijarcs.v10i3.6434

[42]. A. Al-Sideiri, Z. B. C. Cob, and S. B. M. Drus, Machine Learning Algorithms for Diabetes Prediction: A Review Paper,‖ ACM Int. Conf. Proceeding Ser., pp. 27–32, 2019, doi: 10.1145/3388218.3388231.

[43]. Yusuf Perwej, Dr. Ashish Chaturvedi, "Machine Recognition of Hand Written Characters using Neural Networks", International Journal of Computer Applications (IJCA), USA, ISSN 0975 – 8887, Volume 14, No. 2, Pages 6- 9, 2011, DOI: 10.5120/1819-2380

[44]. Dr. E. Baraneetharan, —Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey, ‖ J. Inf. Technol. Digit. World, vol. 02, no. 03, pp. 161– 173, 2020, doi: 10.36548/jitdw.2020.3.004.

[45]. Yusuf Perwej, Firoj Parwej, Nikhat Akhtar, "An Intelligent Cardiac Ailment Prediction Using Efficient ROCK Algorithm and K- Means & C4.5 Algorithm", European Journal of Engineering Research and Science (EJERS), Bruxelles, Belgium, ISSN: 2506-8016 (Online), Vol. 3, No. 12, Pages 126 – 134, 2018, DOI: 10.24018/ejers.2018.3.12.989

[46]. Yusuf Perwej, Firoj Parwej, "A Neuroplasticity (Brain Plasticity) Approach to Use in Artificial Neural Network", International Journal of Scientific & Engineering Research (IJSER), France , ISSN 2229 – 5518, Volume 3, Issue 6, Pages 1- 9, 2012, DOI: 10.13140/2.1.1693.2808

[47]. A. Telikani, A. Tahmassebi, W. Banzhaf, and A. H. Gandomi, Evolutionary Machine Learning: A Survey,‖ ACM Comput. Surv., vol. 54, no. 8, 2022

[48]. R. Katarya and S. Jain, Comparison of different machine learning models for diabetes detection,‖ Proc. 2020 IEEE Int. Conf. Adv. Dev. Electr. Electron. Eng. ICADEE 2020, no. Icadee, pp. 0–4, 2020

[49]. Nikhat Akhtar, Devendera Agarwal, "An Efficient Mining for Recommendation System for Academics", International Journal of Recent Technology and Engineering (IJRTE), ISSN 2277-3878 (online), SCOPUS, Volume-8, Issue-5, Pages 1619-1626, 2020, DOI: 10.35940/ijrte.E5924.018520

[50]. K. Kersting, N. M. Kriege, C. Morris, P. Mutzel and M. Neumann, "Benchmark data sets for graph kernels", 2016

[51]. Y. Guo, X. Cao, W. Zhang and R. Wang, "Fake colorized image detection", IEEE Transactions on Information Forensics and Security, vol. 73, no. 8, pp. 1932-1944, 2018

[52]. https://www.kaggle.com/datasets/mtesconi/twitter-deep-fake-text

[53]. M. Egele, G. Stringhini, C. Kruegel and G. Vigna, "Towards detecting compromised accounts on social networks", IEEE Trans. Dependable Secure Comput., vol. 14, no. 4, pp. 447-460, Jul./Aug. 2017

[54]. A. El Azab, A. M. Idrees, M. A. Mahmoud and H. Hefny, "Fake account detection in twitter based on minimum weighted feature set", Int. Scholarly Sci. Res. Innov., vol. 10, no. 1, pp. 13-18, 2016

[55]. Yusuf Perwej, Dr. Shaikh Abdul Hannan, Firoj Parwej, Nikhat Akhtar, "A Posteriori Perusal of Mobile Computing", International Journal of Computer Applications Technology and Research (IJCATR), ATS (Association of Technology and Science), India, ISSN 2319–8656 (Online), Volume 3, Issue 9, Pages 569 - 578, 2014, DOI: 10.7753/IJCATR0309.1008

[56]. Asif Perwej, Prof. (Dr.) K. P. Yadav, Prof. (Dr.) Vishal Sood, Yusuf Perwej, " An Evolutionary Approach to Bombay Stock Exchange Prediction with Deep Learning Technique", IOSR Journal of Business and Management (IOSR-JBM), e-ISSN: 2278-487X, p-ISSN: 2319-7668, USA, Volume 20, Issue 12, Ver. V, Pages 63-79, December. 2018, DOI: 10.9790/487X-2012056379

[57]. Prof. Kameswara Rao Poranki, Dr. Yusuf Perwej, Dr. Asif Perwej," The Level of Customer Satisfaction related to GSM in India ", TIJ's Research Journal of Science & IT Management – RJSITM, International Journal's-Research Journal of Science & IT Management of Singapore, ISSN: 2251-1563, Singapore, in www.theinternationaljournal.org as RJSSM, Volume 04, Number: 03, Pages 29-36 , 2015

[58]. H. Zhao, H. Zhou, C. Yuan, Y. Huang and J. Chen, "Social discovery: Exploring the correlation among three-dimensional social relationships", IEEE Trans. Comput. Social Syst., vol. 2, no. 3, pp. 77-87, Sep. 2015

[59]. C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou and G. Min, "Statistical features-based real-time detection of drifted twitter spam", IEEE Trans. Inf. Forensics Secur., vol. 12, no. 4, pp. 914-925, Apr. 2016

[60]. B. Ersahin, O. Aktas, D. Kilinc and C. Akyol, "Twitter fake account detection", Proc. IEEE Int. Conf. Comput. Sci. Eng., pp. 388-392, 2017

[61]. Y. Perwej, Nikhat Akhtar, Firoj Parwej, "The Kingdom of Saudi Arabia Vehicle License Plate Recognition using Learning Vector Quantization Artificial Neural Network", International Journal of Computer Applications (IJCA), USA,

ISSN 0975 – 8887, Volume 98, No.11, Pages 32 – 38, July 2014, DOI: 10.5120/17230-7556

[62]. G. Suarez-Tangil, M. Edwards, C. Peersman, G. Stringhini, A. Rashid and M. Whitty, "Automatically dismantling online dating fraud", IEEE Trans. Inf. Forensics Secur., vol. 15, pp. 1128-1137, 2020

Cite this article as :

Ms Farah Shan, Versha Verma, Apoorva Dwivedi, Dr. Yusuf Perwej, Ashish Kumar Srivastava, "Novel Approaches to Detect Phony Profile on Online Social Networks (OSNs) Using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.555-568, May-June-2023. Available at doi : https://doi.org/10.32628/CSEIT23903126
Journal URL : https://ijsrcseit.com/CSEIT23903126