ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Smart Identification of Fake Online Profiles using Machine Learning

[1]KATREDDY SRI SAI KARTHIKEYA, [2] KONDETI DURGA SRI VENKATA VARA PRASAD,[3] KOTA SURYA DURGA PRAVEEN, [4] NELLI SUSHMA KARUNA, [5]Dr. B.V.S. VARMA,

[1234]Student, Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

[5] Professor, Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

Abstract—

This generation's social life are heavily reliant on online social networks. Thanks to these platforms, we can now observe our social life in a new light. Making new acquaintances and keeping in touch with old ones via various social and personal activities is a breeze these days. The contributions of online social networks (OSN) are widespread, spanning fields as diverse as health care, business, technology, research (in all its forms), employment, data collecting, and information gathering. False profiles are a major issue on many social networking sites. The goal of these profile builders is to promote or popularize removed propaganda under someone else's name for financial gain, or to damage and defame the genuine person by impersonating them. Many research have been conducted on the topic of these phony accounts and how to prevent them. In order to detect false profiles, several methods have been considered, including graph-level activity and feature analysis. Compared to the problems that are cropping up now, these approaches are antiquated. In this research, we provide a machine learning-based method for effective false profile identification. In order to make the data presentation more feasible, the benchmark data set is first compiled and combined with manual data. Additionally, a data cleaning approach is used. The next step is to construct a model using the preprocessed data, which includes all the necessary details like the profile's name, ID, number of followers, etc. Incorporating a cross-validation procedure, we run several training algorithms on the provided data and compare their results. According on the results of the trials, the RF classifier outperformed the other classification techniques. For efficient profile authenticity prediction, the Random Forest classifier is used.

Index Terms—Random Forest, Machine Learning, Classification, and Fake Profile Identification

INTRODUCTION

There are many doors that may be opened to you and people that you can meet on the internet. It's likely that you're already aware with some of the most well-known social media platforms. Our generation engages in a plethora of other types of engagement, not limited to these [1-2]. Teachers can easily educate students about social media and Modern educators have mastered the use of these platforms to great effect, enhancing student learning via online lectures, assignments, conversations, and more. Social media makes it easy for companies to research potential employees' backgrounds and find those who are both qualified and excited about working for their company. Some of these platforms charge for membership and utilize the money for business purposes, while others rely on advertising to generate revenue [3-4].

The majority of these platforms, however, are free to use. But there are drawbacks as well, and one of them is the prevalence of false profiles. They often arise from individuals not interacting with us in person, which in turn causes us to get invites we wouldn't normally receive if these false profiles weren't on social media [5]. Many research have been conducted in this arena because of the widespread usage of social networks. Among them, research by Devakunchari (2018) found that 82% of internet users had fallen victim to online impersonation. In addition, 9% of people have fallen for a deception, and 22% have been duped into divulging sensitive information. A lot of research has gone into finding false accounts on the OSN's platform; these accounts aren't helpful for anything other than naïve assaults since they're so easy to hack. When it comes to controlling this issue, there is no practical method that could be 100% accurate [6]. A. Issue Description Online impersonation and false accounts are only two

of the many problems that have recently emerged in social media. No one has come up with a possible remedy for these problems as yet. We want to develop a new model for early identification of computerized phony profiles as part of this research so that people's social life may be protected. Additionally, our automated detecting technology may make it easier for websites to modify the diverse profiles, something that is difficult to do manually.

LITERATURE SURVEY

Online impersonation and false accounts are only two of the many problems that have recently emerged in social media. No one has come up with a possible remedy for these problems as yet. We want to develop a new model for early identification of computerized phony profiles as part of this research so that people's social life may be protected. Additionally, our automated detecting technology may make it easier for websites to modify the diverse profiles, something that is difficult to do manually. Several methods that concentrate on fake documents use an individual's interpersonal organization profile to find the characteristics or a combination of them that aid to distinguish between actual and counterfeit records. In particular, after collecting a large number of variables from the profiles and posts, a classifier is constructed using machine learning methods to identify fake data. Using a Deterministic Finite Automata (DFA) methodology, Padmavati et al. [7] tackle the issue of social media fraudulent accounts. Using an accounting pattern, the article examines the characteristics of the current user and their acquaintances. Pattern matching with friend requests is based on regular expressions that are constructed using certain characteristics, such as the working and living community, and so on. One major issue with this strategy is how long it takes to generate regular expressions for someone with friends in several groups. The authors argue that there's room for improvement in how the strategy works in practice. The problem of fake social media accounts was investigated by Mohammadreza et al. in their study using graph analysis and classification algorithms. The preferred social media site was Twitter. Based on the similarity of the user's pals, they devised a plan. Utilizing the buddy similarity criteria from the network graph is the first step before Principal Component Analysis (PCA) is used to extract new features [8]. The Synthetic Minority Oversampling (SMOTE) approach is then used to balance the data before it is sent to the classifier. After using the cross-validation technique, a medium Gaussian SVM classifier with an AUC of 1 was chosen. One problem with this approach is that fake accounts can only be used within the network; otherwise, their friends' accounts would reveal them. According to the authors, a new method will be available soon that can detect a fake account either when the user signs up for the service or even before they do anything on the network. To identify false profiles, Srinivas Rao et al. [9] used machine learning and NLP in their research. Facebook profiles were used as the dataset by the writers. The procedure consists of three stages: natural language processing pre-processing, principal component analysis, and learning algorithms. The data was pre-processed using stemming, lemmatization, tokenization, and stop word removal. We use principal component analysis (PCA) to get the raw data from the table. Profiles are then classified using two ML algorithms called SVM NB. When these methods were utilized, the detection accuracy increased, according to the observation made after their technique was evaluated. Stringhini researched Twitter fan bases in online marketplaces. They categorize the business sectors' patrons and identify the features of Twitter enthusiast ads. Bills that follow the "client" often fall into one of two kinds, as stated by the authors: either compromised accounts or phony accounts (also called "sybils"), whose providers fail to account for the fact that their fan base is expanding [10].

Notable people or politicians may use devoted markets to inflate their fan bases, while cybercriminals may use them to make their files seem constantly genuine in order to spam and distribute malware unpredictably. The counterfeit currency used to transmit spam on Twitter is examined by Thomas. The users were categorized as genuine or fake by Nancy Agarwal et al. according to their emotions, which included joy, sorrow, anger, fear, and so on. Using posts made by Facebook users, they put it through its paces. To train the detection algorithm, twelve emotion-based features are used [11]. The author's research stems from the discovery that real users express a wide variety of emotions in their posts, in contrast to the uniformity of emotion shown by fake users allocated to certain occupations. Further, we use a noise-reduction technique. In the end, a variety of machine learning methods, such as NB, JRip, SVM, and RF, were used to train the detection model. At the ICRITO conference in 2021, Ananya et al. used machine learning to detect fake social media accounts [12]. Kaggle, an open-source website that stores data sets for public usage, is where they got the data. An immensely famous Chinese social media platform called Weibo is the source of the data. Subsequently, they used five supervised learning models for training and cross-validation to determine which one produced superior test scores.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Because of their superior performance compared to the other four methods, they settled on the Gradient Boosting Classifier and the Random Forest Classifier. They settled on a random forest classifier after much deliberation; it outperformed the gradient boosting classifier by 1%. In their future work, they want to develop an automated system that can learn and use additional properties than those listed in this research. The use of machine learning to detect phony Instagram accounts was detailed by Preethi Harris et al. [13]. Using the Kaggle service, we extracted Instagram profile data. All sorts of classification methods were used to train the model, including SVM, KNN, RF, NB, and XG Boost. The RF classifier emerged as the most appropriate model for the dataset, yielding the best prediction results, after computation of the confusion and accuracy matrices. The Fake profile IDs are thereafter entered into a data dictionary. In 2018, Abhishek Narayanan et al. published a research on identifying false accounts using data collected from Twitter. First, the data was used to extract features, and then SVM, RF, and LR, three machine learning algorithms, produced a highly praised outcome for the random forest. Random forest classifier emerged up with an 88% accuracy rate in predicting Twitter phony accounts after accuracy tests and confusion matrices. Compared to other methods, it was faster and more effective. The goal of their future efforts is to make social media browsing safer for users [14]. Mauro Conti et al. (2012) laid forth potential solutions to the problem in an article [15]. A comparison to the population of actual users was the first step in creating a profile. Then, in order to identify false profiles, they turned to graph topologies. They looked at the user's connections, or friends list, to see whether there were a lot of random people or if there were any common buddies. One way to detect a phoney profile is to use structural analysis of social networks. Along with improving the current approach, their future work aims to expand the categorization to include online interactions like tags, friendship requests, and the rate of request acceptance, among other things.

PROPOSED WORK

As indicated in Figure 1 of the system model, the following framework details the procedures that need to be followed to identify fake profiles; active learning occurs as a result of feedback from the classification algorithm's output. The process consists of the following steps: First, the data is collected and cleaned. Then, to ensure that the model is ideal for the data, it is trained using the data set. To improve the model's accuracy even more, pipe-lining is used. Finally, the model is tested using a test data set.

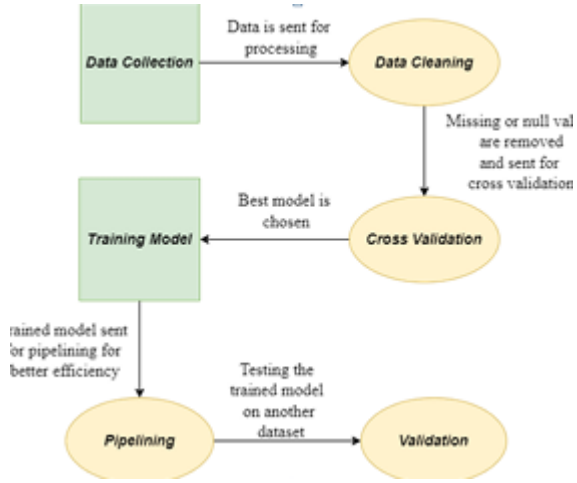

Fig. 1. Proposed System Model

Both stores data sets for public consumption and are open-source. Like any social media platform, we have combined the two sets of data to include common criteria like user name, follower count, and more. The purpose of this was to test the hypothesis that more data may lead to more precise predictions. The data set is enhanced with 200-row inserts by hand and is then used throughout the procedure. Both preexisting datasets and those that were manually integrated with them make up the data sets used. By doing so, we may discover if more quantities are helpful in this circumstance and use the data set for prediction. A dataset including both false and real profiles is required. There are two sets of data in the dataset: training and testing. The classification algorithm learns from the training data and validates its output using the testing data set. In order to test how well the model performs when trained with a larger dataset, we use two datasets. The number of elements in one data set is 556 while in the other data set it is 776. The

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

second data set is formed by manually merging the first data set with 200 items of data from another data set. You can see the features used to train the system to detect false profiles in Figure 2:

Attributes	Description
Profile Picture	User has a profile picture or
Full name words	Number of words in toker
Bio/Description length	Description length in charac
External URL	Has external URL or not
Private	Private account or not
Posts	Number of posts
Followers	Number of followers
Follows	Number of follows

Fig. 2. List of attributes and description

Cross-validation

Several training algorithms are applied to the provided data and then evaluated on the same data in a cross-validation procedure. In the next step, we show you the average scores; greater scores indicate that you may make better use of the data. Because every data set is unique, cross-validation may help you find the best training methods for your data. It is an easy and effective method of maximizing the current model.

Section C. Classification Algorithms • Data set random sample selection A wide variety of subjects and algorithms are used in this project. In order to have a better grasp of the process throughout the implementation, this chapter provides a briefing on such themes. Regression and classification are the two main types of machine learning. These are used in accordance with the dataset and the desired output type of the user. As soon as the data collection is defined and constrained, classification may begin. If you need a yes/no, true/false sort of output, this is the way to go. On the other hand, regression is often used in weather prediction when dealing with continuous data. Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier, and Gaussian Naive Bayes are the four machine learning methods used in this research. One supervised ensemble learning method is Random Forest, which builds many decision trees during training and then utilizes a mean voting mechanism to choose the best ones for making predictions. Trees are trained and evaluated for prediction after being randomly divided into several data samples. Within the trees, there is a voting method that finalizes the prediction score. Secondly, there's Gradient Boosting, a classification technique that combines many decision trees to form an additive prediction model. This model is somewhat similar to the random forest model; however, it is based on the premise that the optimal following model would be the one that produces the fewest mistakes when paired with the preceding one. In order to build a robust prediction model, it combines many ineffective learning methods. 3). Logistic Regression: This method incorporates both a predictive analytic technique and the concept of probability. This statistical approach examines data in situations where one or more independent variables determine the result. By evaluating potential outcomes, evaluation-based logistic regression may choose parameters that improve the chances of finding the right case values. In order to foretell a valid change in the duty of actuality of the relevant aspect, it produces the formulaic coefficients. 4. Naive Bayes using a Gaussian Distribution: The usage of supervised machine learning is also included. Additionally, this is a specific use of the Naive Bayes approach in which the attributes are given continuous values. This method is based on the premise that all characteristics are normally distributed, or that they follow a Gaussian distribution. With only the means and standard deviations, this model fits D. Random Forest Model Selected Here, we used a Random Forest Classifier that we picked during cross-validation. This algorithm can do both regression and classification; it is a supervised learning algorithm. When working with a set quantity of data, classification is the way to go, but when dealing with continuous data, such in the stock market, regression is the way to go. Using a vote system to choose the best decision trees, the Random Forest algorithm creates a forest. It creates several decision trees by partitioning the data set into smaller parts. Consequently, the result is more precise with greater data. Here is how the Random Forest Classifier works: The process begins with building a decision tree for each sample, which yields the projected outcome. Next, all of the results are voted on. Finally, the result with the most votes is selected as the final forecast. E. Laminating pipe Process The repetitive nature of pipeline-lining stems from the fact that each step is iteratively repeated in order to refine the algorithm and enhance the model's

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

accuracy. The model undergoes pipeline-lining, a method that allows for the comparison and analysis of data with comparable features or that is part of a linear series of data transformations that can be assessed as a whole. This partitions the data into separately useful chunks that may then be assembled into a model. To work, it uses the data to create a model, which can then be tested and assessed to find out how efficient it is. The data used for the preparations may be reused, thanks to this. Specifically, it streamlines the process of creating an optimally accurate and exact machine-learning model. Evaluation Matrix Section F. In order to have a better grasp of the issue, this study made use of several charting approaches, including confusion matrices and correlation matrices. If you need to make sure the qualities are reliable, pull out the correlation graph. A plotting matrix known as a confusion matrix provides a more comprehensible visual depiction of the result.

IMPLEMENTATION AND RESULT ANALYSIS

A. Implementation

What follows is the procedure for implementation. 1. Load, inspect, and clean data. 2. Verify attribute correlations by comparing features to ensure data set relativity. If the correlation heat map of any given data set is close to zero, then we may use it. The data sets are suitable for training the model since the correlation graph was almost 0. Verify that no values are present or are empty. Both datasets are free of missing or null values or incorrect data types; so, no more data has to be supplied before processing may begin. 2) Partitioning the data set into a Train and Test set by percentage. We have used 60% of the data from training data set-1 for training and 40% for testing to ensure the model is ready for implementation. This data set yielded more favorable results compared to the others. But, we have trained the model using 75% of the data from training data set-2 and tested it using 25%. One way to find out which model works best with a given dataset is to use cross validation. Random Forest, Gradient Boosting, Logistic Regression, and Gaussian NB are the models that are being tested. Both Table 1 and Table 2 indicate that the Random Forest Classifier and the Gradient Boosting Classifier have the best validation scores, respectively, for the two datasets. For this reason, we will be training using a Random Forest Classifier model.

TABLE I CROSS VALIDATION RESULTS OF DATA SET 1

MODELS	Training score	Validation score
Random Forest	1.000	0.919
Gradient Boosting	1.000	0.913
Logistic Regression	0.891	0.867
Gaussian Naive Bayes	0.698	0.690

TABLE II CROSS VALIDATION RESULTS OF DATA SET 2

MODELS	Training score	Validation score
Random Forest	1.000	0.929
Gradient Boosting	0.997	0.926
Logistic Regression	0.921	0.916
Gaussian Naive Bayes	0.753	0.747

Grid search is used to apply the Random Forest Classifier using parameters such as the maximum depth of the tree (max depth) and the maximum number of estimators (n-estimators). After training is complete, the process's optimal parameters are selected based on the mean of the training and test scores it achieved with those parameters. Finally, pipelining is completed. There was a significant 4% improvement in accuracy between the two sets of data after pipelining the model; the model for set 2 achieved 94% and the model for set 1 achieved 90%. 6) After that, run the trained model on the test data set and create a confusion matrix to see how well it performed. B. Analysis of Results According to the confusion matrix, which can be shown in Tables 3 and 4, the model that was constructed using data set-2 only identified 6 incorrect profiles, whereas the model that was built using data set-1 identified 8 incorrect profiles. Thus, the model trained using data set-2 outperforms the model trained using data set-1, proving once again that the more data fed into a model, the more accurate and precise the outputs will be. The numerical statistics below provide a complete overview of the classification report generated from the test data. They demonstrate the percentage of precision attained when both models were evaluated on the identical set of 120 items. The table 5 and 6 clearly illustrate that the model from data set 2 predicts with a higher accuracy (95% vs. 93%) than the other model.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Table iii CROSS VALIDATION RESULTS OF DATA SET 1 FOR TEXT(33.0,0.5,ACTUAL VALUES)

		Predicting fake account		
		Genuine	Fake	Total
Actual values	Genuine	56	4	60
	Fake	4	56	60
	Total	60	60	120

TABLE IV CROSS VALIDATION RESULTS OF DATA SET 2 FOR TEXT(33.0,0.5,ACTUAL VALUES)

		Predicting fake account		
		Genuine	Fake	Total
Actual values	Genuine	58	2	60
	Fake	4	56	60
	Total	62	58	120

Additionally, for the model built using data set-1, the random forest classifier technique could efficiently distinguish between real and bogus accounts with a sensitivity of 93%. For the model built using data set-2, the random forest classifier technique was able to effectively recognize 94% of real accounts and 97% of bogus accounts.

TABLE V FINAL EVALUATION SCORE OF MODELS OF DATA SETS 1

-	Precision	Recall	F1-Score	Support
Genuine	0.93	0.93	0.93	60
Fake	0.93	0.93	0.93	60
Accuracy	-	-	0.93	120
Macro average	0.93	0.93	0.93	120
Weighted average	0.93	0.93	0.93	120

TABLE VI FINAL EVALUATION SCORE OF MODELS OF DATA SETS 2

-	Precision	Recall	F1-Score	Support
Genuine	0.94	0.97	0.95	60
Fake	0.97	0.93	0.95	60
Accuracy	-	-	0.95	120
Macro average	0.95	0.95	0.95	120
Weighted average	0.95	0.95	0.95	120

. The comparison of results with base classifiers The suggested model's classification accuracy was compared to that of basic classifiers such DT, KNN, NB, SVM, and ANN for the produced data set in Table 6. The results demonstrate that the suggested model outperforms the state-of-the-art base clasifier models in detecting false profiles. D. Examining the Results in Light of Current Models For the generated dataset, Table 7 shows how the suggested model's classification accuracy compares to three other works by the authors[18, 20]. The results demonstrate that the suggested model outperforms the state-of-the-art classifier methods in detecting false profiles.

TABLE VII REFERENCES RESULT COMPARISON WITH BASE CLASSIFIERS

MODELS	P.RF	DT	KNN	SVM	NB	ANN
Accuracy(%)	95	88	87	88.5	77	83

Recognizably fraudulent LinkedIn profiles are shown as proof by Adhikari and Dutta [18]. Using limited profile data as input, the research reveals that phony profiles can be identified with an accuracy of 84% and a false negative rate of 2.44%. Principal component analysis, neural networks, and support vector machines are some of the methods used. Notable features include, among other things, a wide range of languages spoken, level of education, abilities, recommendations, interests, and accolades. We begin with the traits of profiles that have been found to be fraudulent and posted on strange websites. The goal of the Chu is to differentiate between Twitter accounts that are managed by humans, bots, or cyborgs [19]. As part of the design of the detection issue, an Orthogonal Sparse Bigram text content classifier is used to identify spamming archives. This classifier uses pairs of words as features. Nazir explains in his writings how to recognize fake accounts in online gaming applications that are built on social networking platforms. The research looks at the "Fighters club" Facebook program, which is an online game that supposedly gives players benefits and incentives if they get their friends to participate as well [20]. If the sport offers such incentives, the authors argue, it encourages participants to fabricate their profiles so that the user might feel more pressure to do so and boost his or her own motivation.

TABLE VIII RESULT COMPARISON WITH EXISTING MODELS

MODELS	P.RF	Dutta[18]	Chu[19]	Nazir[20]
Accuracy(%)	95	84	92	90.5

CONCLUSION

95.2% 90.5% We compare two models, one trained with a smaller dataset and the other with a larger dataset, to see which one performs better. The data collection with greater information yielded superior outcomes. Using the Random Forest Classifier model and pipelining, we have provided a system that can detect phony accounts in any online social network with an average efficiency of 95%. Eventually, we'd want to be able to categorize profiles using a bigger dataset that contains a variety of data kinds. When working with bigger datasets, it may be necessary to use data preparation techniques in order to extract useful information. Additionally, we need to devise a system that can detect a phony profile by feeding the mode with the necessary attributes.

REFERENCES

[1]. Prabhu Kavin, B., et al. ”Machine learning-based secure data acquisition for fake accounts detection in future mobile communication networks.” Wireless Communications and Mobile Computing 2022 (2022).

[2]. Salman, Fatima Maher, and Samy S. Abu-Naser. ”Classification of Real and Fake Human Faces Using Deep Learning.” International Journal of Academic Engineering Research (IJAER) 6.3 (2022).

[3]. [3] Kenny, Ryan, et al. ”Duped by bots: why some are better than oth ers at detecting fake social media personas.” Human factors (2022): 00187208211072642.

[4]. Purba, Kristo Radion, David Asirvatham, and Raja Kumar Muruge san. ”Classification of instagram fake users using supervised machine learning algorithms.” International Journal of Electrical and Computer Engineering 10.3 (2020): 2763.

[5]. Samala Durga Reddy. “Fake Profile Identification using Machine Learn ing” In Dec 2019 IRJET journal Volume: 06 Issue:12 pp.1145-1150

[6]. Devakunchari Ramalingam, Valliyammai Chinnaiah. “Fake profile detec tion techniques in large-scale online social networks:A comprehensive review”In 2018 Computers Electrical Engineering,Volume 65,Pages 165-177

[7]. Padmaveni Krishnan D.John Aravindhar Palagati Bhanu Prakash Reddy “Finite Automata for Fake Profile Identification in Online Social Net works.” Proc. Of ICICCS 2020, Part Number: CFP20K74-ART

[8]. Mohammadreza Mohammadrezaei, Mohammad Ebrahim Shiri and Amir Masoud Rahmani “Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms” Security and Communication Networks, Volume 2018, Article ID 5923156, 8 pages

[9]. P. Srinivas Rao, Dr. Jayadev Gyani, Dr. G. Narasimha “Fake Profiles Identification in Online Social Networks Using Machine Learning and NLP” International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 6 (2018) pp. 4133-4136

[10]. 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

[11]. Mudasir Ahmad wani, Nancy Agarwal, Suraiya Jabin and Syed Zeeshan Hussain. “Analyzing Real and Fake users in Facebook Network based on Emotions.” In the proceedings of the 2019 11th International Conference on Communication Systems Networks (COMSNETS), pp. 110-117.

[12]. Ananya Bhattacharya, Ruchika Bathla, Ajay Rana, Ginni Arora.” Ap plication of Machine Learning Techniques in Detecting Fake Profiles on Social Media”In the 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) Amity University, Noida, India. Sep 3-4, 2021

[13]. Preethi Harris, Gojal J, Chitra R, Anithra S.”Fake Instagram Profile Identification and Classification using Machine Learning”. In 2021 2nd Global Conference for Advancement in Technology (GCAT) Bangalore, India. Oct 1-3, 2021

[14]. Abhishek Narayanan, Anmol Garg, Isha Arora, Tulika Sureka, Manjula Sridhar, Prasad H B.”IronSense: Towards the Identification of Fake User-Profiles on Twitter Using Machine Learning”In 2018 Fourteenth International Conference on Information Processing (ICINPRO)

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

[15]. Mauro Conti, Radha Poovendran, Marco Secchiero. “FakeBook: De tecting Fake Profiles in On-line Social Networks”In 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,pp.1071-1078

[16]. https://www.kaggle.com/datasets/bitandatom/social-network-fake account-dataset

[17]. https://https://github.com/harshitkgupta/Fake-Profile-Detection-using ML

[18]. Haq, Amin Ul, et al. ”A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms.” Mobile Information Systems, 2018.Adikari, Shalinda, and Kaushik Dutta. “Iden tifying Fake Profiles in LinkedIn.” In PACIS, p. 278. 2015

[19]. Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. “Who is tweeting on Twitter: human, bot, or cyborg?” Proc. Of the 26th annual Computer security applications conference 2010, pp.21-30

[20]. Nazir, Atif, Saqib Raza, Chen-Nee Chuah, Burkhard Schipper, and C. A. Davis. “Ghostbusting Facebook: Detecting and Characterizing Phantom Profiles in Online Social Gaming Applications.” In WOSN. 2010