Machine Learning-Based Identification of False Profiles

[1]J.V.N. Raju, [2]K. Anuhya Chowdary, [3]G. Vinay Kumar, [4]K. Mounika, [5]S.Lakshmi
[1]Associate Professor, [2,3,4,5]Student
[1]Department of Information Technology,
Dhanekula Institute of Engineering & Technology, Vijayawada, India
[1]jvnraju.rao@gmail.com, [2]anuhyakambapu@gmail.com, [3]vinaygandham037@gmail.com
[4]kinthalamounika3@gmail.com, [5]lakshmisaikam2003@gmail.com

***Abstract*—** In the current digital era, phony profiles are a problem for social media sites like Instagram. An important problem that has surfaced on these platforms is the increase in phony profiles, which can result in fraud, false information, and privacy violations, among other problems. Because of the huge number of users, manually identifying bogus profiles takes a lot of effort and is frequently ineffective. This study explores the use of machine learning approaches for Instagram false profile detection. The study uses Random Forest methods and artificial neural networks (ANN) to identify patterns suggestive of fraudulent accounts by utilizing a large dataset. The accuracy with which these approaches can identify bogus profiles is evaluated through a thorough process of experimentation and evaluation. The outcomes demonstrate how well ANN and Random Forest work to differentiate between real

Keywords **-** Fake profile detection,Social media fraud detection, Machine learning, ANN.

I. INTRODUCTION

Fake profiles are becoming a major problem because to the exponential expansion of social media platforms like Instagram.Online interactions, information sharing, and communication have all undergone radical changes as a result of social media platforms' explosive expansion. With over a billion monthly active users, Instagram has become one of these platforms' most well-known and significant users. Unfortunately, because of its immense popularity, bad actors have also been drawn to it. They create false identities for a variety of reasons, such as disseminating false information, engaging in fraudulent activity, and swaying public opinion.

These difficulties include misinformation, fraud, and privacy breaches. Given the huge number of users, manually identifying these phony profiles is time-consuming and frequently unsuccessful. As such, there is a strong demand for automated techniques to detect and stop the spread of false profiles. The project "FAKE PROFILE IDENTIFICATION USING MACHINE LEARNING" attempts to create an advanced system that uses machine learning algorithms to identify phony Instagram profiles in response to this issue. Through the examination of several factors taken from user profiles, such as metadata, activity trends, and content attributes, the research aims to develop classifiers that can correctly identify real accounts from fake ones. By empowering administrators to take preventive measures against fraudulent profiles, this project has the potential to greatly improve the credibility and integrity of social media networks. Furthermore, the effort is consistent with the overarching goal of promoting an online environment that is more dependable and safe for users everywhere. This project aims to provide useful insights and workable solutions to counteract the negative impacts of phony profiles through thorough testing and analysis, ultimately encouraging authenticity and transparency in the digital sphere.In this work, we provide an overview of the approaches and strategies used to use machine learning to detect phony Instagram profiles. Then, we explore the steps involved in gathering and preparing the data, emphasizing the difficulties and

factors unique to Instagram datasets. Next, we examine several characteristics taken from user accounts and postings that are used as input for machine learning classifiers.

Our project's primary goal is

- To efficiently identify the input data as real or fake account.
- Putting into practice machine learning algorithms like Random Forest,ANN.
- The prediction made by the ANN algorithm has a high accuracy of 96.7% when compared to an existing model, and it can accurately distinguish whether the account is real or fake.
- To improve the classification algorithms' overall performance.

II. LITERATURE REVIEW

The new communication medium known as online social networks (OSNs) has made it possible to create and manage social ties among large numbers of people. The vast volume of personal subscriber data on OSNs and their rapid growth have drawn the attention of attackers. Then they pose as propagandists, disseminating fake information, destructive actions, and even stolen personal information.With over 500 million tweets generated every day, the majority of which are malicious posts, Twitter is one of the largest microblogging social networking services. Determine who in social networks is promoting dangers by analyzing classification of user profiles on social networks is necessary. There have historically been a variety of classification techniques for identifying phony social network profiles, but these techniques needed to increase in classification accuracy. In this study, we thus concentrate on machine learning techniques. This research proposes a hybrid Support Vector Machine (SVM) method for the detection of bogus profiles on Twitter. In this case, dimension reduction techniques, feature selection, and bots are utilized along with a hybrid SVM algorithm based on machine learning to classify false and real Twitter account profiles. The suggested hybrid SVM technique uses fewer characteristics and properly classifies 98% of the accounts.

These days, social networking sites like Facebook, Twitter, Weibo, and others are incredibly popular. Furthermore, the majority of malevolent users use these websites to influence reputable individuals for a variety of objectives, such as advertising their goods, clicking on spam links, defaming others, and so on. The number of people using these social networking sites is constantly rising, and the prevalence of false accounts has become a significant problem. Rather than employing a standard spam terms list, this study uses a blacklist to detect bogus accounts.Both topic modeling and keyword extraction techniques are used to generate blacklists. We use the 1KS–10KN dataset as well as the Social Honeypot dataset in an evaluation exercise. We examine the differences in accuracy between our blacklist-based strategy and the conventional spam terms list-based approach. To distinguish authentic accounts from fraudulent ones on Twitter, Decorate, a meta-learner classifier, is utilized. Our method has a true positive rate of 0.95 and an accuracy of 95.4%.

Our lives are greatly impacted by social media platforms such as Facebook, Instagram, Twitter, and others. People are actively involved in it everywhere in the world. However, it must also deal with the issue of fraudulent profiles. Individuals, programs, or devices frequently create fake accounts. They work for illicit activities like phishing and identity theft as well as the propagation of rumors. This project makes use of a number of machine learning approaches to distinguish between real and phony Twitter profiles based on various attributes like friend and follower counts, status updates, and more. Twitter profile dataset, which categorizes phony accounts as INT, TWT, and FSF and real accounts as TFP and E13. The author discusses XG Boost, Random Forest, LSTM, and neural networks in this section. The salient characteristics are chosen to assess a social media page's authenticity. Additionally covered are the architecture and hyperparameters.Finally, results are produced following the training of the models. Therefore, for genuine profiles, the output is 0, and for bogus profiles, it is 1. When a false profile is discovered, it can be disabled or removed, avoiding problems with cyber security.

An increasing number of people use social media platforms (SMPs) to maintain profiles, but they conceal their identities for nefarious reasons. Unfortunately, not much research has been done to date—especially on SMPs—to identify phony identities produced by people. Conversely, There are numerous examples of instances when machine learning models have been effective in identifying phony accounts made by computers or bots. These machine learning algorithms were reliant on using designed features, including the "friend-to-followers ratio," in the case of bots.These functionalities were developed using attributes that are immediately available in the account profiles on SMPs, such as "friend-count" and "follower-count." These same designed qualities are applied by the research covered in this paper to a collection of fictitious user profiles with the aim of improving the identification of fictitious user identities on SMPs.

The issue of identifying bots—automated social media profiles run by software that pose as real users—has far-reaching consequences. Bots have been used, for instance, to influence political elections through the manipulation of online conversation, to control the financial market, or to promote conspiracy theories about vaccines that have led to health crises. The majority of methods up to this point have been designed to identify bots at the account level by analyzing a significant volume of social media postings and using data from temporal dynamics, network structure, and sentiment analysis.In order to detect bots at the tweet level, we present a deep neural network in this study that is built on contextual long short-term memory (LSTM) architecture and makes use of both content and metadata: When processing a tweet text, LSTM deep nets receive auxiliary input in the form of contextual features that are collected from user metadata.Our other contribution is the proposal of a method based on synthetic minority oversampling to produce a big labeled dataset (approximately 3,000 examples of smart Twitter bots) from a little quantity of labeled data (ideal for deep nets training). We show that our architecture can achieve high classification accuracy (AUC > 96%) in distinguishing humans from bots from a single tweet.By utilizing the same architecture, we are able to detect bots at the account level with almost flawless classification accuracy (AUC > 99%). Our approach requires less training data and performs better than the prior state of the art, all while exploiting a modest and interpretable set of characteristics.

III. PROPOSED SYSTEM

The suggested approach identifies phony Instagram profiles by using sophisticated machine learning techniques like Random Forest and neural networks (ANN). By training on a dataset that includes labeled instances of both real and false profiles, these algorithms are able to identify patterns and traits that separate the two. The machine learning algorithms identify important characteristics of Instagram profiles, like the number of friends and followers, status updates, and other patterns of behavior. These attributes were selected with effort to represent the unique qualities that set real profiles apart from fakes. The suggested approach makes it possible to identify bogus profiles proactively, in contrast to the current system.

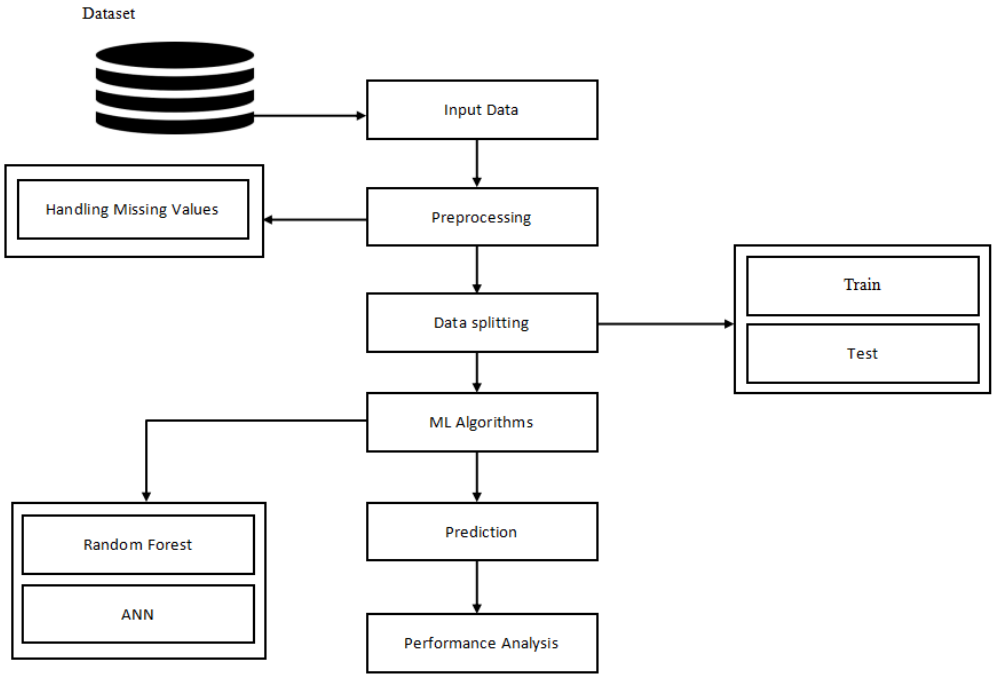


Figure 2 : Flow diagram for fake account identification

Input Dataset

The process of choosing the right data type, source, and instrumentation for data collection is known as data selection.We have gathered the Instagram dataset from Kaggle and retrieved the publicly available dataset for the suggested system. CSV format is used to store the data for machine extraction.

CSV FILE CONVERSION:

Because of its magnitude, Microsoft Excel has a vast dataset. These Excel files were exported into CSV format so that the intended applications could use the information. Following the Excel file's conversion to CSV, the following steps were done: Start the file import process. Modify the text Notepad and other spreadsheet programs like Microsoft Excel and Google Sheets can be used to do this task. You could use Excel (Windows) in its place.

• Select a file.
• Select "Save As" from the drop-down option.
• The author has the option to rename the file and choose the.csv suffix (comma-delimited) if they so choose.
• Select the Save option.
• After these procedures were finished, the author found the final dataset for our proposed system, stored in a CSV file.

A	B	C	D	E	F	G	H	I	J	K	L
profile pic	nums/length username	fullname words	nums/length fullname	name==username	description	external UR	private	#posts	#followers	#follows	fake
1	0.27	0	0	0	53	0	0	32	1000	955	0
1	0	2	0	0	44	0	0	286	2740	533	0
1	0.1	2	0	0	0	0	1	13	159	98	0
1	0	1	0	0	82	0	0	679	414	651	0
1	0	2	0	0	0	0	1	6	151	126	0
1	0	4	0	0	81	1	0	344	669987	150	0
1	0	2	0	0	50	0	0	16	122	177	0
1	0	2	0	0	0	0	0	33	1078	76	0
1	0	0	0	0	71	0	0	72	1824	2713	0
1	0	2	0	0	40	1	0	213	12945	813	0
1	0	2	0	0	54	0	0	648	9884	1173	0
1	0	2	0	0	54	1	0	76	1188	365	0
1	0	2	0	0	0	1	0	298	945	583	0
1	0	2	0	0	103	1	0	117	12033	248	0
1	0	2	0	0	98	1	0	487	1962	2701	0
1	0	3	0	0	46	0	0	254	50374	900	0

training3000

Pre-Processing

Here, the author adds one additional pre-processing step before moving on to the models. Before the data gathered is fed into a model, it is pre-processed. This technique looks at a profile's appearance to determine whether it is authentic or fraudulent. All the details have now been resolved. After removing the category components, only the numerical data is left. After combining a user data set that is both accurate and unreliable, each profile is assigned the additional Boolean variable "fake." The response relevant to profile X is then stored in the Y variable. Lastly, zeros are added to any blank or NAN entries.

Data splitting:

For machine learning to take place, data must be there.

- In order to evaluate the algorithm's performance and ascertain its efficacy, test data are also required.
- During our approach, we categorized 70% of the input dataset as training data and the remaining 30% as testing data.
- Data splitting is the process of breaking available data into two portions, usually for cross-validator purposes.
- A portion of the data is used to construct a predictive model, while the remaining amount is used to evaluate the model's performance.
- Separating the data into training and testing sets is an essential step in evaluating data mining models.

Algorithms:

ANN

The research for identifying false profiles on Instagram datasets has successfully made use of artificial neural networks, or ANNs. In order to extract pertinent features from user profiles and postings, the dataset is first preprocessed. The ANN model receives inputs from these features, which include engagement metrics, posting frequency, and profile attributes. An ANN's design usually consists of several levels, such as input, hidden, and output layers, with weighted connections between the neurons. Using backpropagation and gradient descent optimization techniques, the model learns to map the input features to the associated class labels (fake or real profiles) during training. A variety of activation functions are used to add non-linearity and boost the representational capability of the model, including sigmoid and ReLU. To maximize the ANN's performance, cross-validation and hyperparameter tuning approaches are used.

Random Forest:

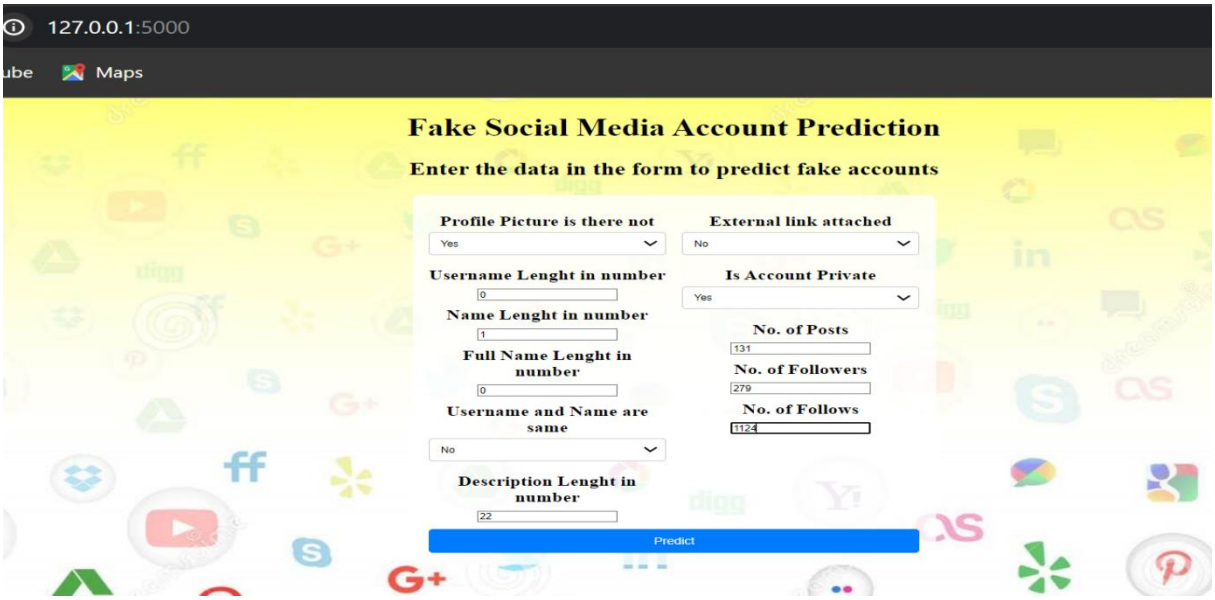
Because of its stability and capacity to manage complicated datasets, Random Forest is a well-liked algorithm in the machine learning project for identifying phony Instagram profiles. In order to extract pertinent features like profile information, posting activity, and engagement metrics, the Instagram dataset is first preprocessed. The Random Forest classifier uses these features as input. Using random feature subsets and data sample selections, the Random Forest method builds several decision trees. Every decision tree learns on its own during training to distinguish between real and fraudulent profiles using the given features. A final prediction is produced during inference when the ensemble of decision trees votes collectively on the class label for a particular profile. Because of its adaptability and efficiency, Random Forest can be used to identify fraudulent activity on Instagram by identifying features that are indicative of fraudulent activity and collecting complex patterns. The Random Forest model's success in correctly recognizing phony profiles is evaluated using evaluation criteria like accuracy, which adds to the project's overall efficacy.

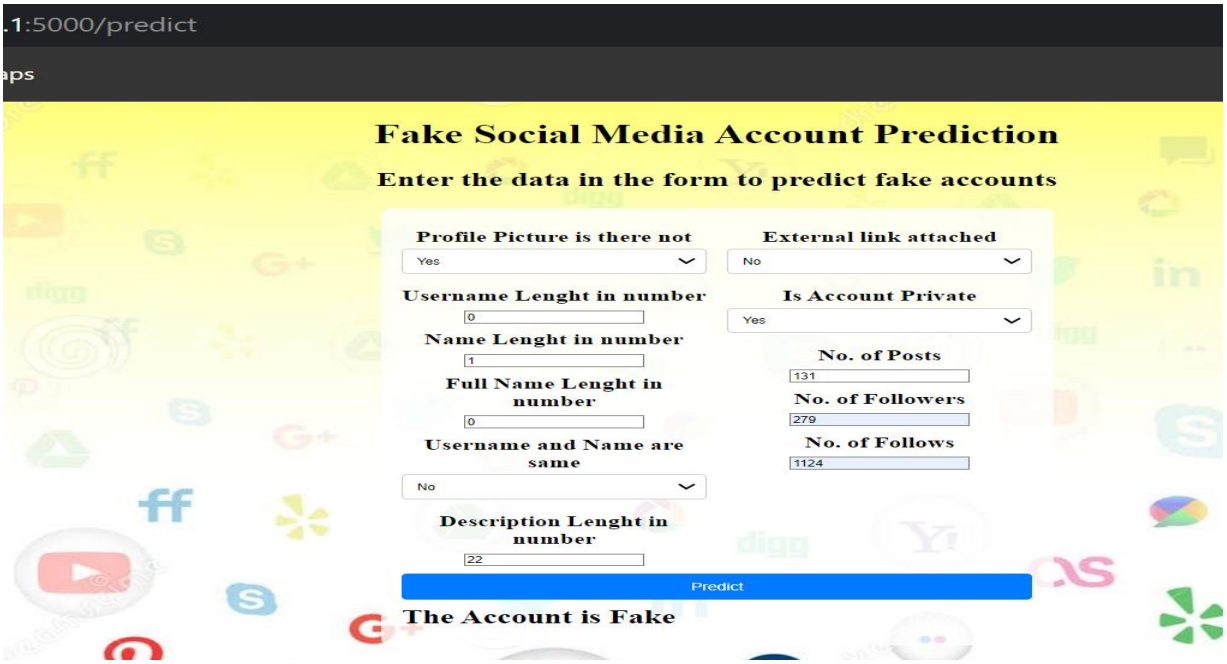
IV. RESULTS:

The whole classification and projection will serve as the foundation for producing the ultimate outcome. Metrics like these are used to evaluate the efficacy of this proposed approach: The accuracy of the classifier indicates its competence. Predictor accuracy quantifies how well a specific predictor can approximate the expected attribute value for a given collection of data. It precisely predicts the class label.

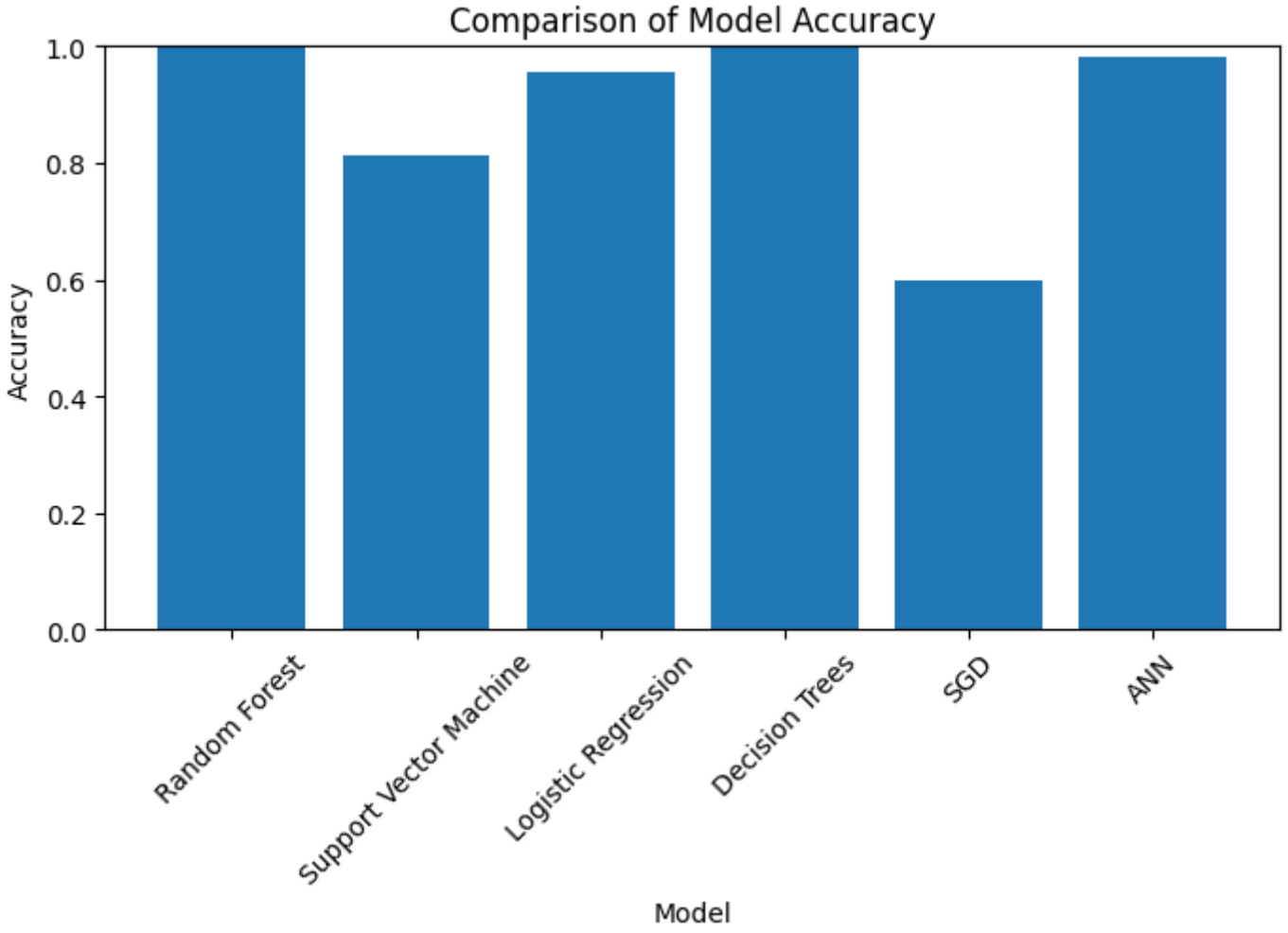
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \times 100\%. \quad (1)$$

Sample screen shorts






```
   ANN
19/19 [==============================] - 0s 4ms/step
Test Accuracy of Loaded Model (using predict): 0.9800995024875622
```



V. CONCLUSION

In conclusion, there has been a noticeable progress made in tackling the widespread problem of phony Instagram profiles thanks to the project on false profile identification utilizing machine learning approaches, particularly Artificial Neural Networks (ANN) and Random Forest algorithms. Both ANN and Random Forest classifiers have been successfully trained on the Instagram dataset to discriminate between real and fake profiles through careful feature engineering and data preprocessing. The ANN model exhibits remarkable performance in identifying minute details suggestive of fraudulent activity, thanks to its capacity to discover intricate patterns and relationships within the data. However, the ensemble of decision trees used in the Random Forest approach provides stability and scalability, making correct classification possible even in the presence of unbalanced or noisy data... By utilizing these cutting-edge machine learning techniques, the project offers automated and effective procedures for identifying phony profiles, improving the security and reliability of online social networks. In order to further increase detection accuracy and resilience against changing fraudulent strategies in the dynamic landscape of social media platforms like Instagram, future initiatives might involve investigating ensemble methods integrating ANN and Random Forest classifiers.

REFERENCES

[1]. Automated Fake Profile Detection Using Machine Learning on Instagram, Meshram, E.P.,Bhambulkar, R., Pokale, P., Kharbikar, K., and Awachat, 2021. Journal of Scientific Research in Science and Technology: International, 8, 117–127. https://doi.org/10.32628/IJSRST218330

[2]. Fake Profile Detection Methods in Large-Scale Online Social Networks: A Complete Study, D. Ramalingam and V. Chinnaiah, 2018.165–177 in Computers & Electrical Engineering, vol. 65. https://doi.org/10.1016/j.compeleceng.2017.05.020

[3]. Identification of Fake Accounts on Twitter Using Hybrid SVM Algorithm, Kodati, S.,Reddy, K.P., Mekala, S., Murthy, P.S., and Reddy, P.C.S., 2021.Article No. 01046 of E3 S Web of Conferences, page 309. https://doi.org/10.1051/e3sconf/202130901046

[4]. Jie, H.J., and Wanda, P. (2020) DeepProfile: Utilizing Dynamic CNN to Detect Fake Profiles in Internet Social Networks.52, Article ID 102465 in Journal of Information Security and Applications. https://doi.org/10.1016/j.jisa.2020.102465

[5]. Detecting Fake Profiles in Online Social Networks by Long-Range Dependence Analysis. https://ieeexplore.ieee.org/document/8366718

[6]. Fake Profile Detection on Social Networking Sites Using Supervised Learning Techniques. https://www.sciencedirect.com/science/article/pii/S1877050919333115

[7]. Detecting Fake Profiles in Online Social Networks Using Supervised Machine Learning Algorithms. https://www.researchgate.net/publication/330957815_Detecting_Fake_Profiles_in_Online_Social_Networks_Using_Supervised_Machine_Learning_Algorithms

[8]. Fake Social Media Profile Detection Using Machine Learning Techniques. https://www.sciencedirect.com/science/article/pii/S1877050919311565

[9]. Fake Profile Detection on Social Networking Sites Using Supervised Learning Techniques https://www.sciencedirect.com/science/article/pii/S1877050919333115