Fake Profile Detection Using Machine Learning

Dr.S.N.Tirumala Rao[1], Palavai Venkata Deepthi[2], Somu Thejita Lasya[3], Mangalagiri Saraswathi[4]
[1] Professor, [2, 3 & 4] Student
[1]nagatirumalarao@gmail.com, [2] palavaideepthi2002@gmail.com, [3] tejithalasya@gmail.com, [4] mangalagirisaraswathi8@gmail.com
Department of Computer Science and Engineering,
Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh, India

ABSTRACT: In today's world, social media is an essential component of everyone's social life on the internet. It's become simpler to make new acquaintances and stay updated on their activities. Numerous fields are impacted by online social networks, including business, education, employment, community involvement, and research. Employers use these social networking sites to find and hire qualified applicants who are enthusiastic about their job. Spreading misinformation via social media is another problem. Incorrect accounts that propagate unsuitable and incorrect information may give rise to conflicts. These made-up accounts are likewise intended to attract followers. More harm is done to people by false profiles than by other internet crimes. Consequently, it's critical to be able to recognize a fake profile.

KEYWORDS: Fake profile, Social media, Machine Learning, Gaussian Naïve Bayes classifier, Random Forest, Logistic Regression, XG(Extreme Gradient) Boosting, Fake profile detection

I. INTRODUCTION

You have access to a wide range of contacts and possibilities via the Internet. Popular social networking services like Facebook, Instagram, Snapchat, and WhatsApp are certainly recognizable to you. Apart from these conventional modes of social engagement, the contemporary generation engages in a plethora of alternative types of interaction. Social networking services are quite simple to use for educators to instruct children and These days, teachers use these websites extensively to provide online lectures, give homework, host conversations, and more—all of which greatly enhance student learning. Employers may find candidates who are knowledgeable and passionate about their profession by using social networking sites. These websites make it easy to check candidates' backgrounds.

While some of these platforms are free, others charge a membership fee, which they use for business purposes, and yet others rely on advertising to generate revenue. However, there are drawbacks, and false profiles are one of them. They often arise from a straightforward absence of in-person interactions, and this frequently results in invites that, in the absence of these phony [1] identities on social media, we wouldn't typically receive. Several studies in this field have been conducted due to the widespread usage of social networks.

The majority of phony profiles are created to gain more followers by sending out spam and phishing attacks [2]. False accounts are outfitted with all the tools required for performing crimes online. False accounts pose a risk of data breaches and identity theft. False accounts purporting to be from individuals or groups may harm their reputation and acquire fewer likes and follows.

II. LITERATURE SURVEY

Gururaj et al.[3]. In this study, they suggest using natural language processing (NLP) to identify users who may be suspicious based on their regular exchanges with one another. They used their anomalous actions to illustrate each user's behavior. Support vector machine (SVM) classifiers are another machine learning approach that may be used to identify harmful comments in a blog. The first work in this study focuses on using anomalous activity, behavior profiles, communications, and the comment area to identify the malicious user.

Kodati, S., et al. [4] This study describes a strategy for identifying human-created phony accounts. This introduces a novel hybrid classification method for effectively identifying phony profiles on social media. Here, the feature vector is reduced by Spearman's rank-order correlation. Redundancy is eliminated, and the correlation technique's best characteristics are chosen. According to the result analysis, the suggested hybrid SVM outperformed the other machine learning methods already in use, achieving a 98% accuracy rate in the identification of phony Twitter profiles. It also demonstrated superior performance and efficiency.

Hajdu et al. [5] In this study, we employ artificial neural networks and machine learning to ascertain the likelihood of a real Facebook friend request. We also describe the relevant libraries and classes. We also go over the sigmoid function and the calculation and application of the weights. Lastly, we take into account the social network page's specifications, which are crucial to the offered solution.

Khaled et al.[6] This research suggests a novel classification approach to enhance social network fake account detection. An NN model was trained using SVM-trained model decision values, and the NN model was tested using SVM-testing decision values. To accomplish our aim, we ran the "MIB" baseline dataset through a pre-processing step in which the feature vector was reduced using several feature reduction approaches. As compared to the other two classifiers, "SVM-NN" has demonstrated superior accuracy results with all feature sets, with an approximate classification accuracy of 98%, according to the analysis findings.

Kulkarni et al. [7] examine the authenticity of an internet presence. SVM, Naïve Bayes, and Decision Trees are just a few of the classifier algorithms this programmed model has used to identify profiles. Select the suitable feature, or the feature that makes use of the classification technique, after deciding which target profile needs to be assessed. Using the newly acquired dataset and the extracted feature, a classifier was trained on the data. A notification is issued to a target profile asking for a genuine identity when the classifier determines that the profile is fraudulent.

III. PROPOSED SYSTEM

Our Model is Proposed based on certain criteria as follows:

- Dataset Analysis
- Data Preprocessing Techniques
- Creation and Evaluation of Model
- Acquisition of Model Accuracy

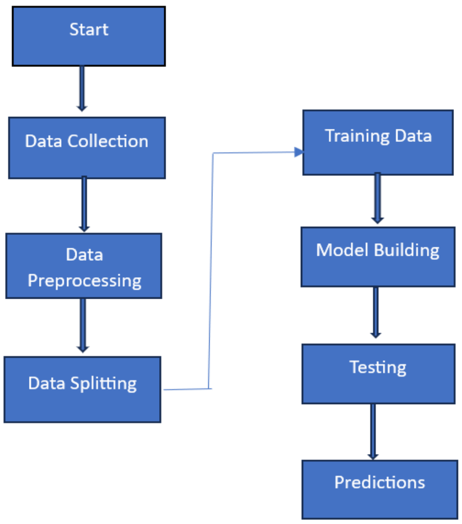

Fig. 1. Proposed model

A. Dataset Analysis

We have taken the datasets from the Kaggle website from the internet. We have collected two datasets which are users.csv and fusers.csv [8]. The users and the fusers datasets contain 35 columns which is shown in Fig. 2. and it has both categorical and numerical data[9].

0	id	2818 non-null	int64
1	name	2818 non-null	object
2	screen_name	2818 non-null	object
3	fav_number	2818 non-null	int64
4	statuses_count	2818 non-null	int64
5	followers_count	2818 non-null	int64
6	friends_count	2818 non-null	int64
7	favourites_count	2818 non-null	int64
8	listed_count	2818 non-null	int64
9	created_at	2818 non-null	object
10	url	463 non-null	object
11	lang	2818 non-null	object
12	time_zone	1069 non-null	object
13	location	2271 non-null	object
14	default_profile	1728 non-null	float64
15	default_profile_image	8 non-null	float64
16	geo_enabled	721 non-null	float64
17	profile_image_url	2818 non-null	object
18	profile_banner_url	987 non-null	object
19	profile_use_background_image	2760 non-null	float64
20	profile_background_image_url_https	2818 non-null	object
21	profile_text_color	2818 non-null	object
22	profile_image_url_https	2818 non-null	object
23	profile_sidebar_border_color	2818 non-null	object
24	profile_background_tile	489 non-null	float64
25	profile_sidebar_fill_color	2818 non-null	object
26	profile_background_image_url	2818 non-null	object
27	profile_background_color	2818 non-null	object
28	profile_link_color	2818 non-null	object
29	utc_offset	1069 non-null	float64
30	protected	0 non-null	float64
31	verified	0 non-null	float64
32	description	2547 non-null	object
33	updated	2818 non-null	object
34	dataset	2818 non-null	object
35	isFake	2818 non-null	float64

Fig. 2. Dataset Description

These datasets tell us about the profile of each user, the user being either genuine or fake. Fig. 3. shows which data we are considering for further steps

#	Column	Non-Null Count	Dtype
0	statuses_count	2024 non-null	int64
1	followers_count	2024 non-null	int64
2	friends_count	2024 non-null	int64
3	listed_count	2024 non-null	int64
4	favourites_count	2024 non-null	int64
5	lang	2024 non-null	int32
6	default_profile	2024 non-null	int64
7	profile_use_background_image	2024 non-null	int64
8	isFake	2024 non-null	int64

Fig. 3. Consider data

B. Data Pre-Processing Techniques

The first step in this study is data preprocessing, where raw data is transformed into a format suitable for analysis by the machine learning model. This involves various tasks such as data cleaning, normalization, and feature encoding. Raw data from social media platforms typically contains noise, missing values, and inconsistencies, which must be addressed before feeding it into the model. Additionally, preprocessing may involve extracting relevant features from the data that are indicative of fake profiles, such as activity patterns, profile completeness, and interaction behavior.

Machine learning algorithms do not allow missing values in the data, so handling missing values is crucial. Among the methods for dealing with missing values are: Swapping out for a mean, median, or mode, Back-fill or forward-fill, and Deleting row.

In our dataset, the target variable is "isFake". It contains the values of real users and fake users. Fig. 4. and Fig. 5. shows how the values are distributed. However, there is a class imbalance in Fig. 4. When there is value of a class is more than another value, it can lead to class imbalance. To remove class imbalance we are using SMOTE which is a technique used to address class imbalance in machine learning. The class is balanced after applying SMOTE, as shown in Fig. 5.

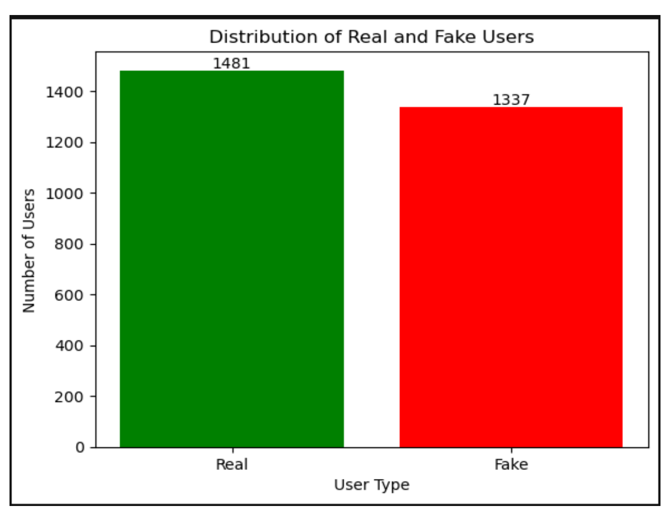

Fig. 4. Class imbalance

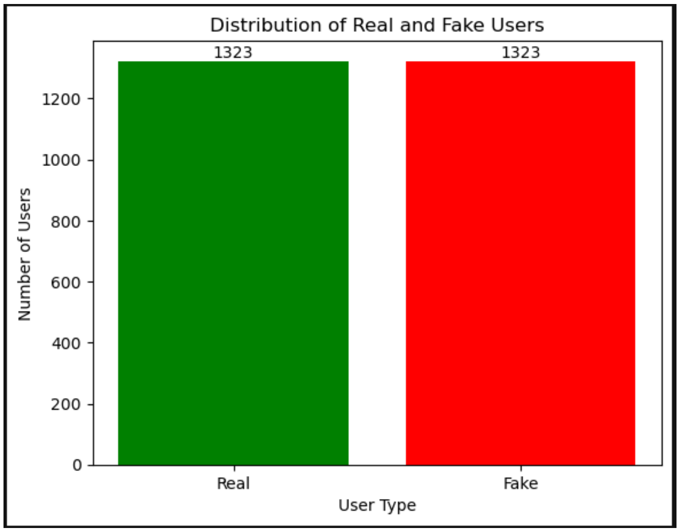

Fig. 5. Class balance

The next step in our study is to check whether the dataset has null values or not. The matplotlib library in Python provides a convenient way to visualize missing values [10] in a dataset. It offers various types of plots to help you understand the distribution and patterns of missing values within your dataset. Fig. 6. shows which columns have missing values.

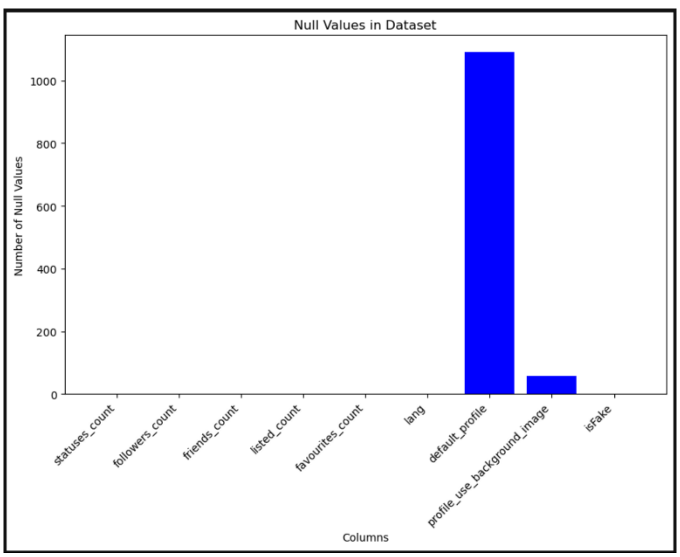

Fig. 6. Null values

Fig. 7. depicts that there are no null values in our dataset and we have successfully removed all the null values from our datasets using the data preprocessing methods as mentioned above, we have removed the null values in our datasets.

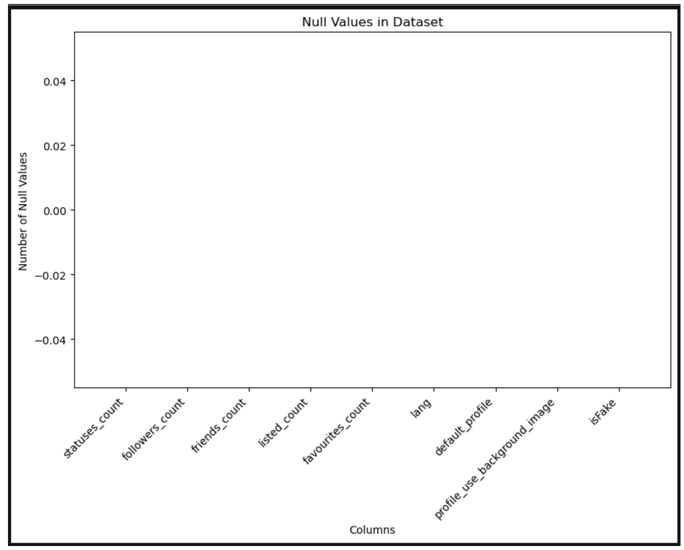

Fig. 7. After the removal of null values

Outlier detection is a crucial step in identifying anomalous instances within the dataset, which may signify the presence of fake profiles. Outliers could manifest as profiles exhibiting unusual activity levels, suspicious posting behavior, or inconsistencies in profile information. Leveraging outlier detection techniques, such as clustering-based approaches or statistical methods, enables the identification and subsequent handling of such instances. Outliers are detected and removed from the dataset as shown in Fig. 8. and Fig. 9.

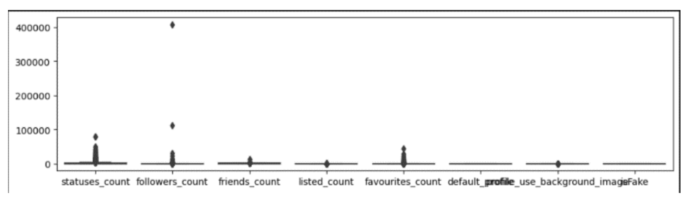

Fig. 8. Before Outlier Removal

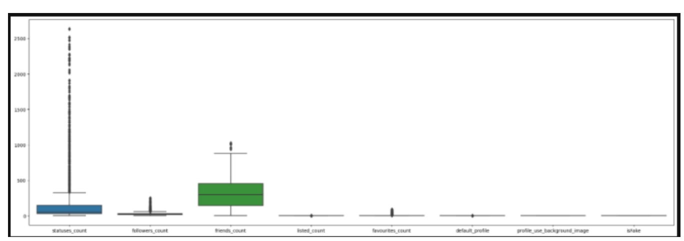

Fig . 9. After Outlier Removal

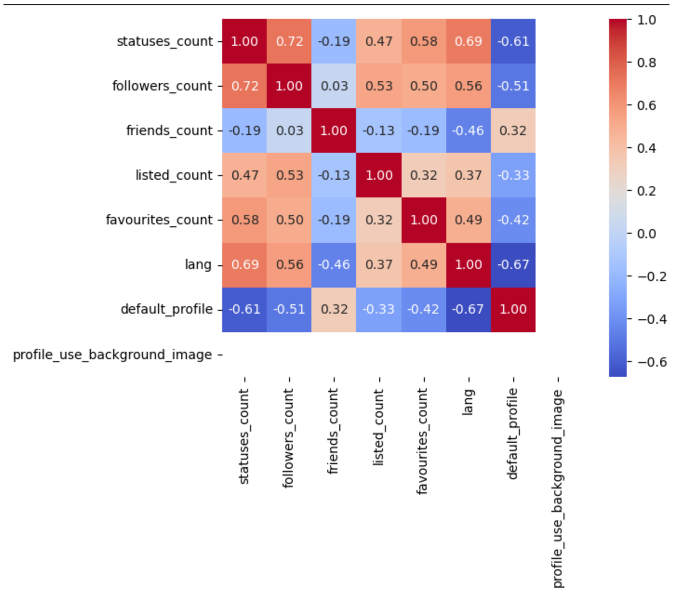

Fig. 10. Correlation Heatmap

A correlation heatmap [11] visual representation of the dataset is plotted to get a brief understanding and visualization regarding which features are strongly correlated and which features are weakly correlated, as shown in Fig. 10.

C. Creation and Evaluation of Model

This step involves choosing an appropriate machine learning algorithm and training it on the ready data. To reduce the discrepancy between the model's anticipated output and the actual output found in the training set, the parameters are optimized. The model is tested on a different validation dataset to gauge its performance after training. The performance criteria and problem type determine the evaluation measures that are employed. Evaluation criteria that are frequently used include F1 score, recall, accuracy, and precision. The model can be further improved by changing its parameters or using an alternative algorithm in light of the evaluation findings.

1. Feature Selection:

Feature selection is a crucial process that enhances model performance and lowers computing costs by determining which subset of characteristics is most pertinent for categorization. It is crucial to use discriminative features for false profile identification that capture the unique traits that set real profiles apart from fraudulent ones. Finding the characteristics with the greatest predictive potential is made easier by methods including information gain, correlation analysis, and model-based selection.

2. Data Splitting:

The dataset is divided to assess how well the machine learning model performs. The model is trained using labeled data from the training set, which helps it discover the fundamental patterns that point to fraudulent profiles. The model's performance is then evaluated on the testing set, which consists of unobserved data. This assessment guarantees the model's good generalization to new cases and gives a precise evaluation of how effectively it detects phony profiles.

3. Model Evaluation

Random Forest: The below Fig. 11 shows the pictorial representation of random forest[12]

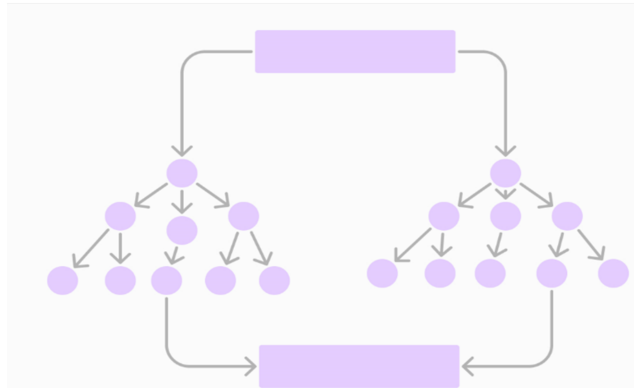

Fig. 11. Random Forest

Gradient Boosting: Below Fig. 12 shows that combining different weaker learning models to construct a powerful prediction model.

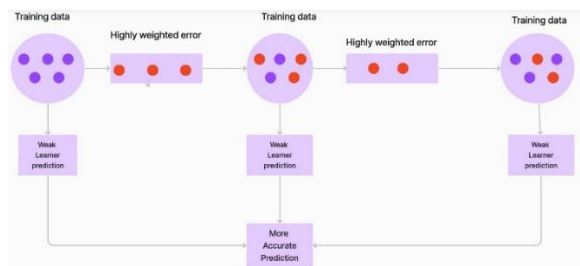

Fig. 12. Gradient Boosting

Logistic Regression: Below Fig. 13 shows the classification between two classes and Fig.14 below shows the classification between the multiple classes.

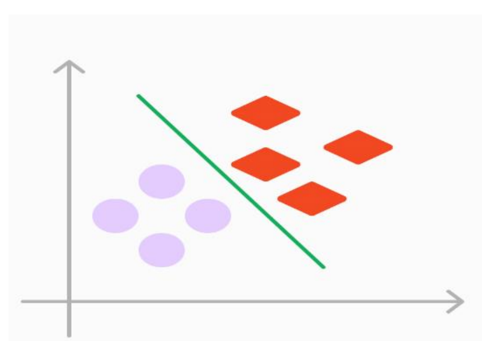

Fig. 13. Binary classification

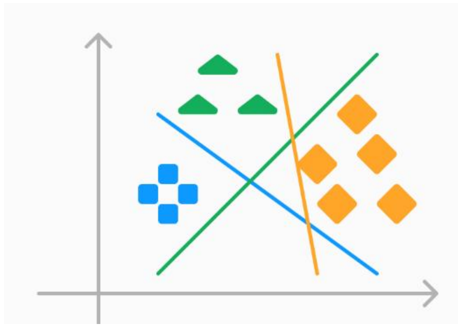

Fig. 14. Multiple regression

Gaussian Naive Bayes: Below Fig. 15 shows how the naive Bayes works

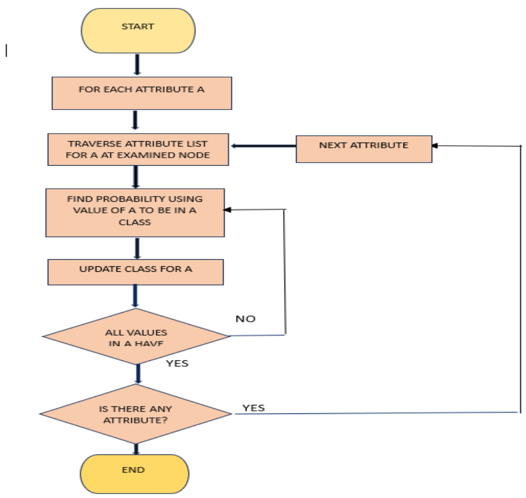

Fig. 15. Gaussian Naïve Bayes Algorithm

IV. RESULT ANALYSIS

The models that we used in this study are Random Forest, Gradient Boosting Classifier, Logistic Regression, and Gaussian Naïve Bayes.

A. Random Forest Classifier

Random forest gives an accuracy of 100%. A ROC curve, which is shown below in Fig. 16. That the model can accurately categorize positive situations without producing any false positives, would look like a straight line in the graph's upper left corner.

The matrix[13] shows the performance of a random forest classifier on a test set. Fig. 17. indicates that the model made 174 true positive values, 332 true negative values, and 0 values for true negative and false positive. Fig. 18. and Fig. 19. Show the evaluation metrics[15] of random forest.

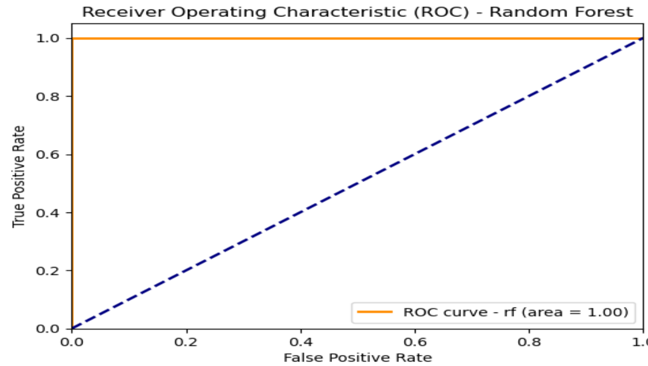

Fig.16. ROC Curve - Random Forest Classifier

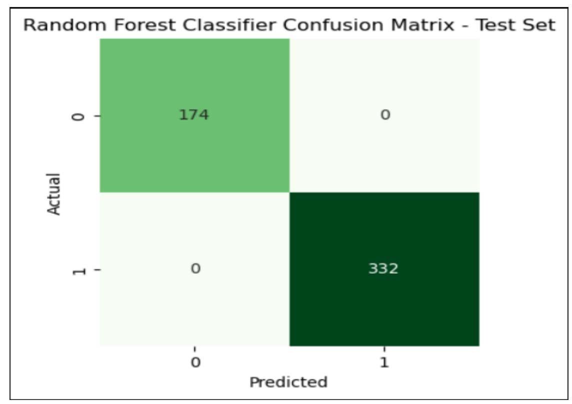

Fig. 17. Confusion matrix

	precision	recall	f1-score	support
0	1.00	1.00	1.00	174
1	1.00	1.00	1.00	332
accuracy			1.00	506
macro avg	1.00	1.00	1.00	506
weighted avg	1.00	1.00	1.00	506

Fig. 18. Metrics of Random Forest

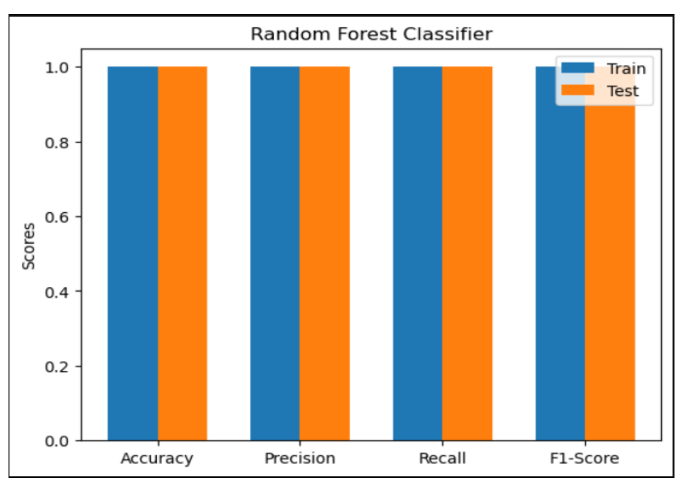

Fig. 19. Metrics of Random Forest

B. Logistic Regression

Logistic Regression gives an accuracy of 99%. The ROC Curve in image fig. 20. Shows that the model is performing well.

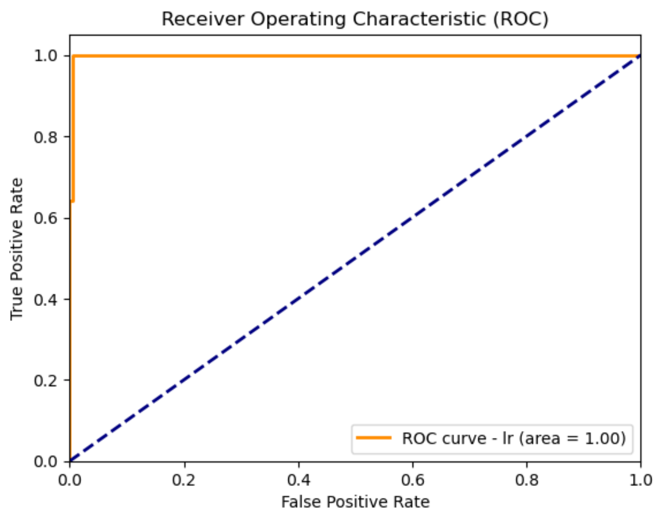

Fig. 20. ROC Curve Logistic Regression

The matrix shows the performance of a Logistic Regression on a test set. Fig. 21. indicates that the model made 173 true positive values, 332 true negative values, 1 value for true negative, and 0 value for false positive.

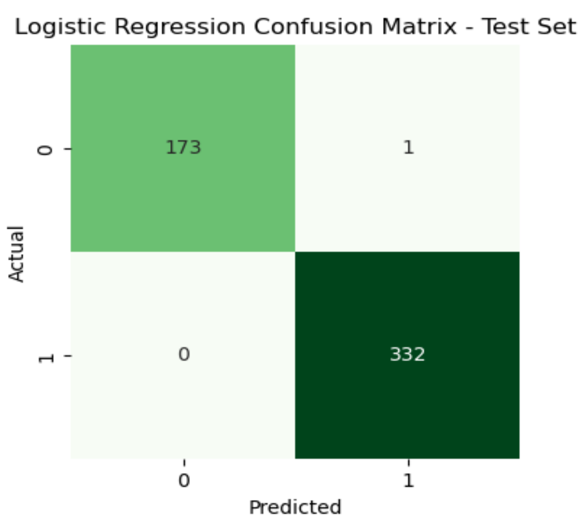

Fig. 21. Confusion matrix

	precision	recall	f1-score	support
0	1.00	0.99	1.00	174
1	1.00	1.00	1.00	332
accuracy			1.00	506
macro avg	1.00	1.00	1.00	506
weighted avg	1.00	1.00	1.00	506

Fig. 22. Metrics for logistic regression

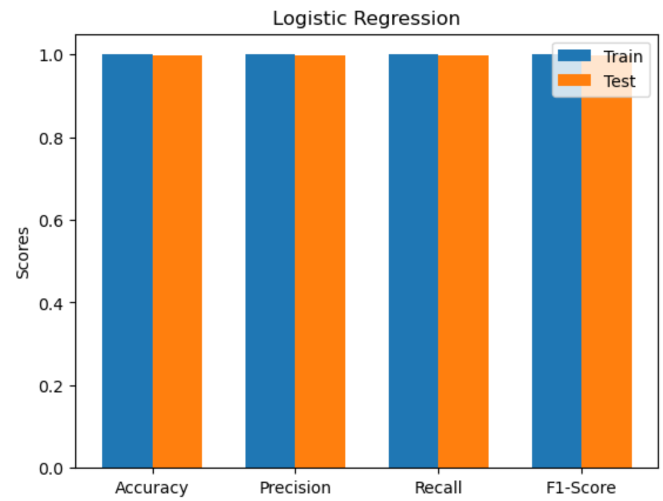

Fig. 23. Evaluation metrics

Fig. 22. And Fig. 23. Shows about the evaluation metrics of logistic regression.

C. Naive Bayes

Naïve Bayes gives an accuracy of 99%. The ROC Curve in image fig. 24. Shows that the model is performing well.

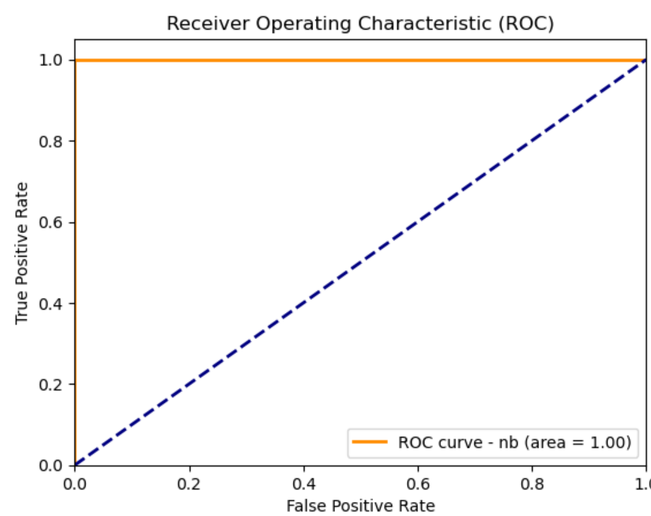

Fig. 24. ROC Curve Naïve Bayes

The matrix shows the performance of a Naïve Bayes on a test set. Fig. 25. indicates that the model made 173 true positive values, 332 true negative values, 0 value for true negative, and 1 value for false positive. Fig. 26. And Fig. 27. Shows about the evaluation metrics of Naïve Bayes.

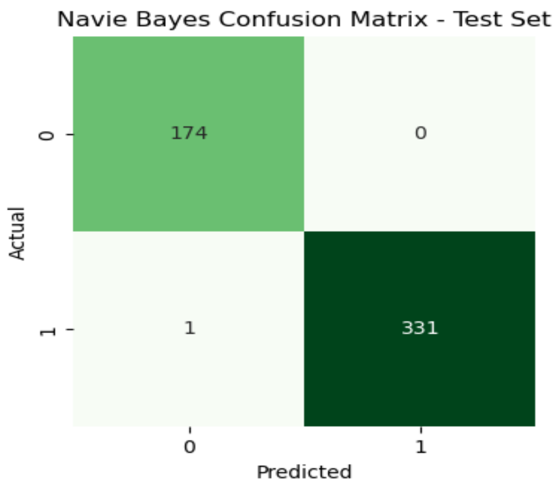

Fig. 25. Confusion matrix

	precision	recall	f1-score	support
0	0.99	1.00	1.00	174
1	1.00	1.00	1.00	332
accuracy			1.00	506
macro avg	1.00	1.00	1.00	506
weighted avg	1.00	1.00	1.00	506

Fig. 26. Evaluation Metrics

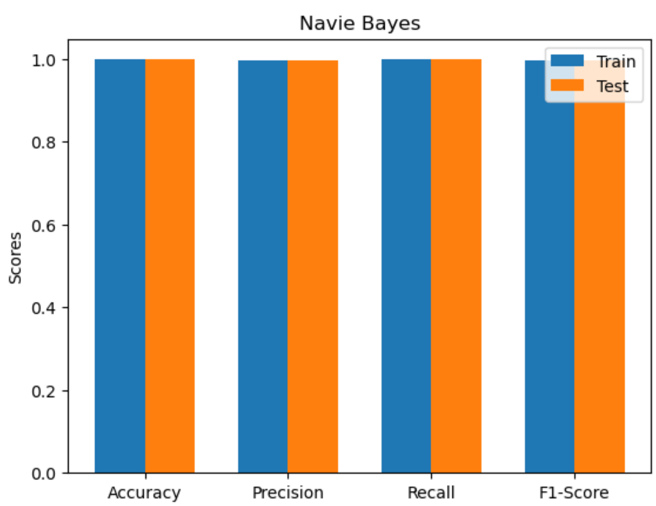

Fig. 27. Evaluation Metrics

D. Gradient Boosting Classifier

Gradient Boosting Classifier gives an accuracy of 99%. The ROC Curve in image fig. 28. Shows that the model is performing well.

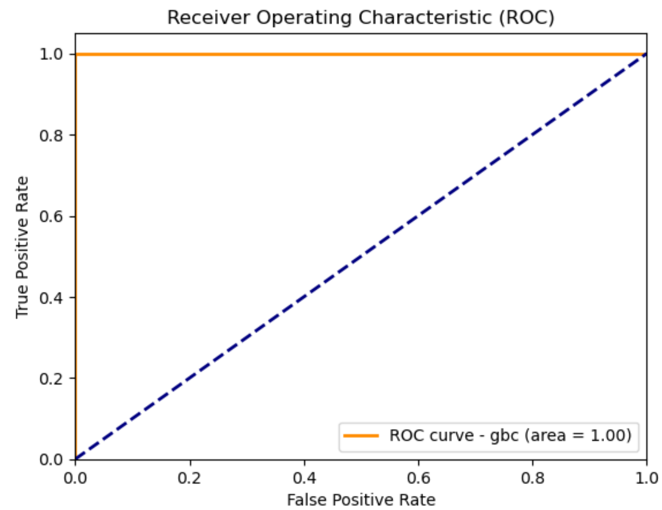

Fig. 28. ROC Curve Gradient Boosting

The matrix shows the performance of a Gradient Boosting Classifier on a test set. Fig. 29. indicates that the model made 173 true positive values, 332 true negative values, 0 value for true negative, and 1 value for false positive. Fig. 30. And Fig. 31. Shows about the evaluation metrics of the Gradient Boosting Classifier.

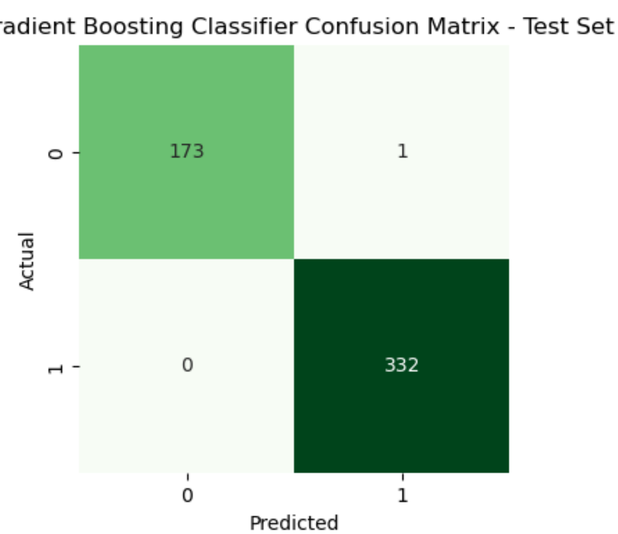

Fig. 29. Confusion Matrix

	precision	recall	f1-score	support
0	1.00	0.99	1.00	174
1	1.00	1.00	1.00	332
accuracy			1.00	506
macro avg	1.00	1.00	1.00	506
weighted avg	1.00	1.00	1.00	506

Fig. 30. Evaluation Metrics

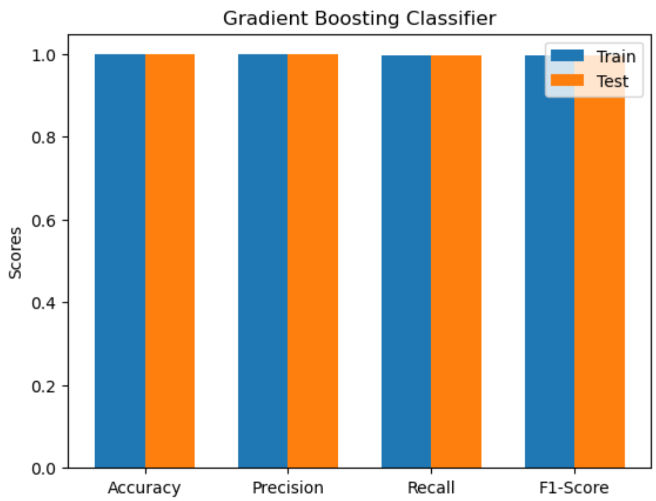

Fig. 31. Evaluation Metrics

Once the model has been trained and evaluated, its efficiency and effectiveness in fake profile detection are analyzed. Performance metrics provide quantitative measures of the model's performance.

we have used the random forest algorithm to achieve a higher accuracy of 100%, in the case of both the genuine and fake user accounts, as compared to the existing system[16] which could achieve an accuracy score of 94% in the case of accurately detecting the genuine users and 97% in case of accurately classifying the fake users

V.CONCLUSION AND FUTURE SCOPE

This study explored the application of four popular machine learning algorithms, Random Forest, XG Boosting algorithm, and Logistic Regression, for predicting whether a social media account of a user is real or fake. The results indicate that of the four algorithms we have used for our predictions, the random forest algorithm is the most effective in generating accurate predictions, with random Forest outperforming the XG boost, Naïve Bayes, and Logistic Regression in terms of accuracy and efficiency. The predictive model developed in this study can be useful for better safety and privacy protection of authentic users from the fraudulent practices of fake users on social media platforms.

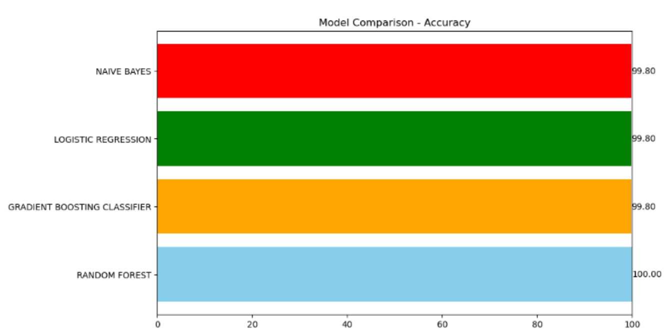

Fig. 32. Accuracy Comparision

Overall, this project highlights the potential of machine learning in enhancing user safety and privacy and allowing safe online engagement of users on social media platforms. Future research can expand this work by incorporating more complex features, exploring more optimized machine learning algorithms, and analyzing each machine learning model's performance on a larger dataset.

REFERENCES

[1] Exploring machine learning techniques for fake profile detection in online social networks: Bharti, Nasib Gill, Preeti Gulia vol. 3, June 2023, pp. 2962~2971.

[2] K.Harish, R.Naveen Kumar, Dr.J.Briso Becky Bell(2023) Fake profile detection using machine learning

[3] Gururaj, H.L., Tanuja, U., Janhavi, V., Ramesh, B.: Detecting malicious users in the social networks using machine learning approach. Int. J. Soc. Comput. Cyber-Phys. Syst. 2(3), 229–243 (2021)

[4] Kodati, S., Reddy, K.P., Mekala, S., Murthy, P.S., and Reddy, P.C.S. (2021) Detection of Fake Profiles on Twitter Using Hybrid SVM Algorithm. E3S Web of Conferences, 309, Article No. 01046.

[5] Hajdu, G., Minoso, Y., Lopez, R., Acosta, M. and Elleithy, A. (2019) Use of Artificial Neural Networks to Identify Fake Profiles. 2019 IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, 3 May 2019, 1-4.

[6] Khaled, S., El-Tazi, N. and Mokhtar, H.M. (2018) Detecting Fake Accounts on Social Media. 2018 IEEE International Conference on Big Data (Big Data), Seattle, 10-13 December 2018, 3672-3681.

[7] Kulkarni, Sumit Milind, and Vidya Dhamdhere."Automatic detection of fake profiles in online social networks." Open access International Journal of Science and Engineering 3.1 (2018):70-73. M. Young, The Techincal Writers Handbook. Mill Valley, CA: University Science,1989.

[8] Fake Profile Identification Using Machine Learning: Asso. Prof. G Swathi, R Vaishnavi, Shaik Noorus Sabiha, P Rakesh Anand, P Nithish Kumar Sreyas Institute of Engineering and Technology(2023).

[9] Joshi, U.D., Singh, A.P., Pahuja, T.R., Naval, S. and Singal, G. (2021) Fake Social Media Profile Detection. In: Srinivas, M., Sucharitha, G., Matta, A. and Chatterjee, P., Eds., Machine Learning Algorithms and Applications, Scrivener Publishing LLC, Beverly, MA, 193-209.

[10] Meshram, E.P., Bhambulkar, R., Pokale, P., Kharbikar, K. and Awachat, A. (2021) Automatic Detection of Fake Profile Using Machine Learning on Instagram. International Journal of Scientific Research in Science and Technology, 8, 117-127.

[11] Sarah Khaled, Neamat El-Tazi, Hoda M. O.Mokhta: Detecting Fake Accounts on Social Media,2018 IEEE International Conference on Big Data (Big Data)

[12] Prof of. Vivek Solvande, Vaishnavi Ambolkar, Siddhesh Birmole, Divya Gawas, Dnyanada Juvale and Fake Profile Identification Using Machine Learning Algorithm,2021

[13] Reddy, S.D.P. (2019) Fake Profile Identification Using Machine Learning. International Research Journal of Engineering and Technology (IRJET), 6, 1145-1150.

[14] Sk.Shama, K. Siva Nandini, P.Bhavya Anjali, K. Devi Manaswi, Fake Profile.Identification in Online Social Network, 2019

[15] S. Sheikhi, An Efficient Method for Detection of Fake Accounts on the Instagram Platform, 2020.

[16] T. Sudhakar, Bhuvana Chendrica Gogineni, J. Vijaya "Fake Profile Identification Using Machine Learning" IEEE International Women in Engineering(WIE) Conference on Electrical and Computer Engineering(WIECON-ECE) 2022.