FAKE PROFILE DETECTION ON SOCIAL NETWORKING WEBSITES

Rajeesh Raj[1], Nadera Beevi S[2], Natheera Beevi M [3]

Department of Computer Applications[1,2,3], TKM College of Engineering, Kollam, Kerala

***Abstract*— The increased reliance on computer tech- nology in today's digital age has made the common citizenexposed to crimes such as data breaches and probable identity theft.These breaches or crime often target social media networks such as Instagram, Facebook, LinkedIn and Twitter. The motivates the needs for social networks to improve their cyber safety. The project intends to buildan artificial intelligence solution to prevent the danger of a bot in the form of a fake profile on social media .The deep learning algorithms can determine the possibility of a social media profile to be authentic/not. The attributes of the social media that drive a breach are also identified in the work and web browser plugin is built to identify these fake profiles. Researches have observed that 20 percentage to to 40 percentage profiles in online social networks like Facebook, Instagram, linkedIn and Twitter are fake profiles. The owners of fake accounts extract the personal information about other people and spread the forget data on social networks. The Work employs a combination of SVM and Random Forest model and Neural Network model to determine fake profiles in social network with improved accuracy.**

I. INTRODUCTION

People have often utilised social networking sites as a means of communication. Users of social networking sites can share their informa- tion and daily activities which attract a numberof people towards these sites. The most widely used social networking sites are 'Facebook',' In- stagram','Twitter' etc. Fig. 1 shows the increasing number of people using the social media from the year 2004 to 2018. Social Media's allow the users to add friends and share various kind of informa- tion such as personal, social, political, business etc

. Moreover, they can also share photos, videos, travels and another day to day affairs. However, some people don't use these sites with good in- tent. Therefore, they create fake accounts on social networking sites. Fake accounts do not have any

real identity. Basically, the person who creates fakeaccounts is known as Attacker. The attacker uses incorrect information or statistics about some real world person to create a fake account. Using the- ses fake accounts, attacker spread false informationwhich affects other users.

A data mining process called classification assigns items in a collection to target categories or groups. The purpose of classification is to compute the tar-get class for each example in the data as precisely as possible. There are 2 phases in classification. The first one is learning in which the training data is analysed and in the second phase the algorithm is evaluated with the test data and resulting perfor- mance is measured. The output class is predicted based on input applied to the algorithm. There are various classification techniques available. Neural

Networks and SVM is most promising methods for classification. for classification.

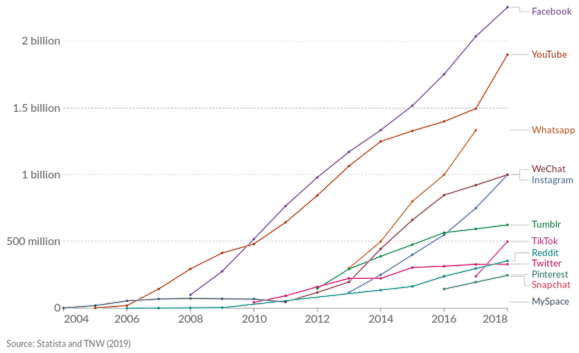


Fig. 1. Number of people using Social Media Platforms 2004- 2018

II. RELATED WORKS

For spam profile identification in OSN websites,Ahmad et al. proposed a Markov clustering (MCL) based technique[1]. They identified three features such as active friends, page-likes and URL shared from the profile data of users.They concluded that fake accounts share URLs more often than real users.

Meligy et al employed regular expressions and deterministic finite automata to distinguish be- tween real and fake profile. They achieved good performance scores of F1 on on Twitter, Facebook,and Google+ datasets[2].

On Twitter, a NN-based algorithm was presentedby Khaled et al. for detecting bogus profiles and bots. Techniques for dimensionality reduction and feature selection were used . The study relied on the MIB dataset. Several models were utilised for feature reduction. They discovered that combining the SVM and NN models produced the greatest accuracy of 98.3

DOI: 10.5281/zenodo.6758955
***ISBN**: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

percent when utilising Spear- man's rank-order correlation approach to acquire the features [3].

Xuan et al. studied different ML based classi- fiers to detect rogue accounts on location-based so- cial networks (LBSNs) such as Twitter, Facebook, and others. They chose 398 valid user accounts at random from the 3997 legitimate user accounts to balance the dataset. They achived 0.89 F1 score with SVM algorithm [4].

Al-Zoubi et al. identified fake profiles using En- glish and Arabic posts on Twitter. They employed public features and ML classifiers to develop the model and concluded that Naive Bayes algorithm have the maximum accuracy of 95.70 % [5].

TABLE I

FREQUENTLY USED FEATURES FOR FAKE ACCOUNT DETECTION

Feature Types	Features
Account	Username [59], [60], [8], [61], [6], [62], [63], [64], [65]
	Biography [66], [67], [68], [59], [60], [8], [61], [6], [29], [64], [69], [70], [71], [65]
	Profile Photo [68], [59], [72], [73], [60], [8], [74], [6], [61], [75], [64], [76], [63], [65]
	Header Photo [73], [6], [29], [76]
	Homepage [59], [6]
	Theme Color [77], [76], [65]
	Birth date[78], [71]
	Location [59], [60], [8], [6], [62], [64], [63], [71], [70], [65]
	Creation [67], [8], [61], [74], [64], [63], [70], [65]
	Number of tweets [66], [6], [75], [69], [79], [80], [71], [65]
	Number of followers [66], [82], [59], [72], [60], [8], [61], [74], [6], [75], [62], [29], [79], [69], [71], [65]
	Following count [66], [60], [75], [69], [81], [71], [65]
	Number of likes [82], [83], [79]
	Listed count [84], [59], [74], [6], [29], [69], [64], [76], [70]
Textual	Sender [78], [85], [29]
	Mentions [67], [82], [84], [74], [29], [69], [80], [71]
	Hashtags [66], [67], [84], [59], [8], [74], [6], [86], [29], [69], [71], [70]
	Link [66], [67], [78], [85], [84], [59], [87], [88], [74], [86], [89], [29], [62], [69], [80], [79], [71]
	Number of retweets [84], [59], [69], [80], [71], [70]
	Number of replies [70]
	Sent date[85], [83], [75], [90], [81]
	Location [85], [6]

value of f- measure and recall required for detec- tion of fake account in Online Social Networking.

The Fig. 2 shows architecture of the proposed

system, consists of the following phases:

- Collect the data from social media platforms
- Clean the data/data scrubbing/Pre-processing
- Feature Engineering (Extract the essential fea-ture)and select a model
- Design ML Model/Evaluate the Model
- Train the model and Testing Data
- Deploying the model using Flask API
- Summarizes the result

III. METHODOLOGY

The proposed work uses the techniques like Neural Networks, Support Vector Machine and Random Forest for classification of real and fake accounts. The feature set that influences the de- tection of fake accounts are categorized as shown in Table 1, includes Nontextual (account-based) features and Textual features. This proposed work combines the weighted feature set with machine.

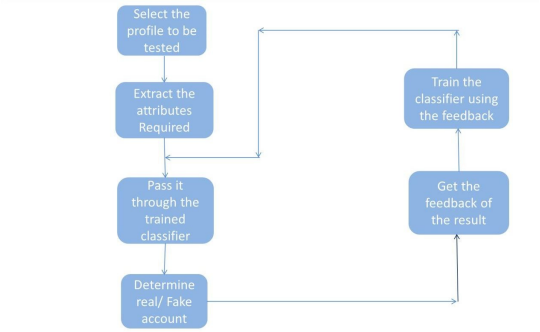


Fig. 2. System Architecture

learning techniques to obtain the best results. The proposed work is expected to generate the higher

The dataset for the study is collected from Kaggle and consists of 2 csv files, one related with fake users and other with genuine users. For the model to work upon, there is a need for data collection. The dataset can be col- lected from various online platforms and can also be created by using Crawler. We have collected two datasets through online from well-known websites Kaggle and GitHub. But we worked on the dataset which is collected by Kaggle and in that we are using two CSV files corresponding to fake and genuine users. Fig. 3 shows the sample of csv file.

A. Data Preprocessing

Data pre-processing is used to achieve the better result from any machine learning model and data processing is used to clean the data from raw data. The libraries such as numpy, panda and scikit-learnare used for cleaning the data.

The features are manually selected which includes:

- statuses-count
- followers-count
- friends-count
- favourites-count
- listed-count
- lang-code

DOI: 10.5281/zenodo.6758955
***ISBN**: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

The extract feature function converts string typelanguage code feature to integer type.The extractedfeature are shown in Fig. 3.

	id	statuses_count	followers_count	friends_count	favourites_count	listed_count	default_profile	default_profile_image	geo_enabled
count	2.818000e+03	2818.000000	2818.000000	2818.000000	2818.000000	2818.000000	1728.0	8.0	721.0
mean	5.374889e+08	1672.198368	371.105039	395.363023	234.541164	2.818666	1.0	1.0	1.0
std	2.977005e+08	4884.669157	8022.631339	465.694322	1445.847248	23.480430	0.0	0.0	0.0
min	3.610511e+06	0.000000	0.000000	0.000000	0.000000	0.000000	1.0	1.0	1.0
25%	3.620867e+08	35.000000	17.000000	168.000000	0.000000	0.000000	1.0	1.0	1.0
50%	6.162253e+08	77.000000	26.000000	306.000000	0.000000	0.000000	1.0	1.0	1.0
75%	6.177673e+08	1087.750000	111.000000	519.000000	37.000000	1.000000	1.0	1.0	1.0
max	1.391998e+09	79876.000000	408372.000000	12773.000000	44349.000000	744.000000	1.0	1.0	1.0

Fig. 3. Sample of CSV file

B. Algorithm

INPUT: The dataset from CSV files.OUTPUT: Performance Matrics.

1. Read dataset: Read the 2 csv files and append them in a list, named x, and creates another list y for labelling class. Return x,y.

2. Feature extraction: The extracted 6 features are stored in list x. Return x

3. Split data into training data and test data using5 cross fold validation and store them separately in x-train, x-test, y-train, y-test.

4. Scaling of the X-data for preprocessing

5. Use GridSearchCV with SVM and Random forest for predicting the result.

6. Store result in y-pred variable and Return y-pred.

7. Repeat step 3 with y pred and x-test

8. Store the output in y-pred.

9. Testing - Evaluating our trained model againstthe test data.

The output is visual graph consisting of True- Positive-Rate and False-Positive-Rate with accu- racy measures.

10. Print the classification accuracy on testing dataset. Plot the confusion matrix.

11. Exit

The training of SVM and Neural Network modelis shown in Fig. 4 and Fig. 5 respectively.

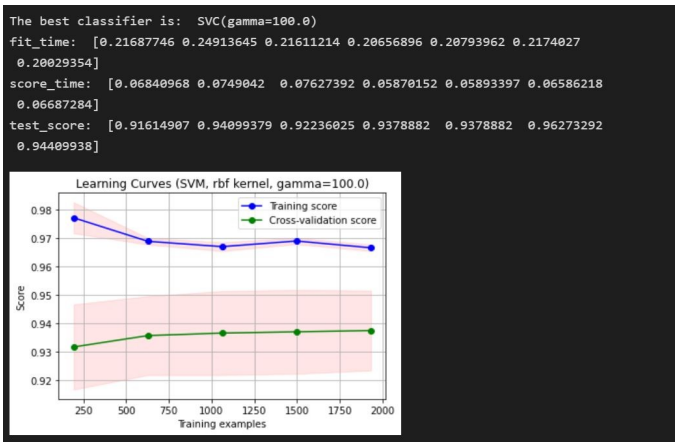


Fig. 4. Training of SVM model

After getting the score of each fold final averagescore of 0.91 is estimated. Fig. 4 shows the training of SVM and Neural Network models.

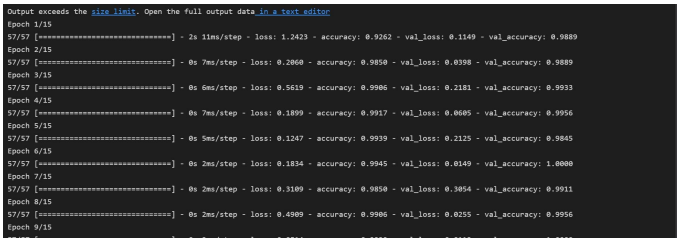


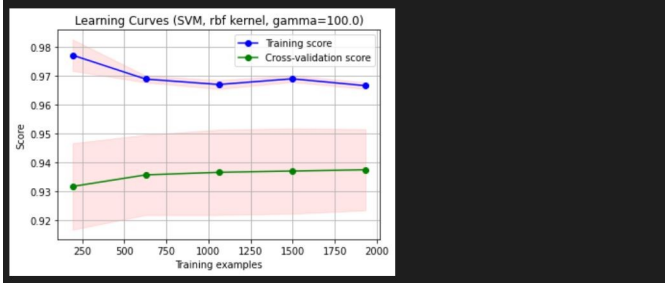
Fig. 5. Training of Neural Network

IV. RESULTS AND DISCUSSIONS

In the proposed work the fake accounts can be detected by using machine learning techniques on Social Media. In the study we achieved accurate

DOI: 10.5281/zenodo.6758955
***ISBN**: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

```
The best classifier is:  SVC(gamma=100.0)
fit_time:  [0.21687746 0.24913645 0.21611214 0.20656896 0.20793962 0.2174027
 0.20029354]
score_time:  [0.06840968 0.0749042  0.07627392 0.05870152 0.05893397 0.06586218
 0.06687284]
test_score:  [0.91614907 0.94099379 0.92236025 0.9378882  0.9378882  0.96273292
 0.94409938]
```



DOI: 10.5281/zenodo.6758955
***ISBN**: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

results by using Neural Networks.The following section discusses the various performance matrics used in the study

A. Classification report

The classification report displays the precision, recall, F1, and support scores for the model. The metrics are defined in terms of true and false positives and negatives .

$$Precision = TP/(TP+FP) \quad (1)$$

$$F1Score = 2(Recall\ \ precision)/Recall+Precision. \quad (2)$$

$$Recall = TP/TP + FN \quad (3)$$

Fig. 8. Evaluating the SVM model using Learning Curve

The classification report of SVM and Neural Net- work is shown in Fig. 6 and Fig. 7 respectively. It is clear that SVM predicts results with an accuracy of 0.90 and Neural Network with 99.46. While developing Neural Network model 3 input layers are used and 10 epochs are executed.

	precision	recall	f1-score	support
Fake	0.99	0.82	0.90	296
Genuine	0.83	0.99	0.90	268
accuracy			0.90	564

Fig. 6. Classification report of SVM

C. Receiver Operating Characteristics Curve(ROC)

ROC curve, also known as Receiver Operating Characteristics Curve, is a metric used to measure the performance of a classifier model. The ROC curve depicts the rate of true positives with respect to the rate of false positives, therefore highlighting the sensitivity of the classifier model. The ROC curves obtained for SVM and hybrid classifier is shown in Fig. 9 and Fig. 10 respectively.The train and validation accuracy of Neural Network model is shown in Fig.11.

```
Test loss: 0.03712095692753792
Test accuracy: 99.46808218955994
```

Fig. 7. Accuracy of Feed forward neural networks.

B. Learning Curve

A learning model shows how the error in the prediction of class changes as the size of the training set increases or decreases. The Learning Curve while evaluting the SVM model is shownin Fig. 8.

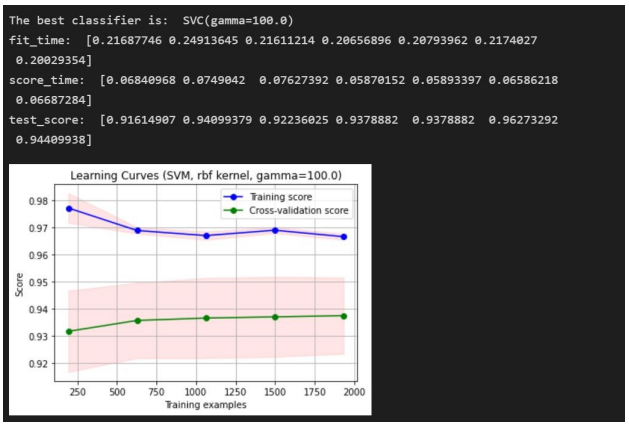


Fig. 8. Evaluating the SVM model using Learning Curve

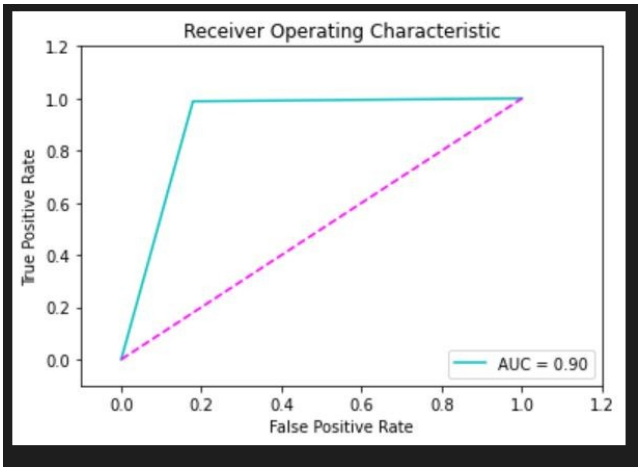


Fig. 9. AUC Curve SVM

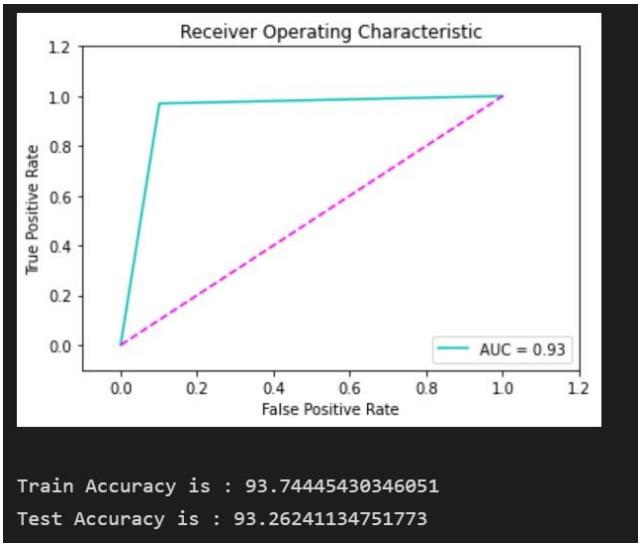


Fig. 10. ROC curve of hybrid classifier

DOI: 10.5281/zenodo.6758955
***ISBN**: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

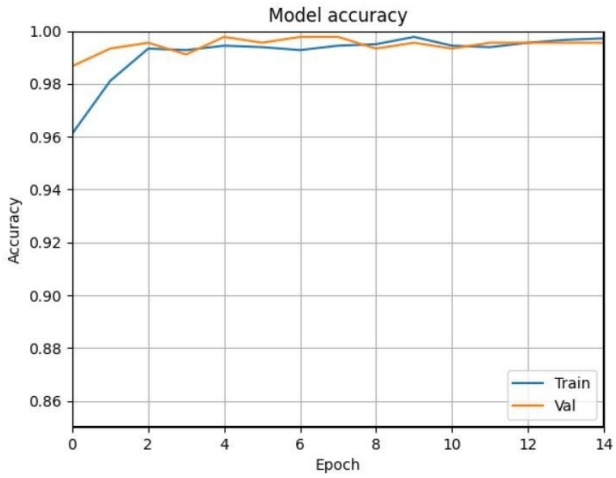


Fig. 11. Train and validation accuracy of Neural Network model

D. Confusion Matrix

Confusion Matrix is the another matric to iden- tify the performance of classifiers.It represents a tabular form of TP, TN, FP and FN. The matrix of SVM and Random Forest is shown in Fig.11 and Fig. 12.After analysing the result both models are combined to obtain an accuracy of 93.4 as shown in Fig. 10.

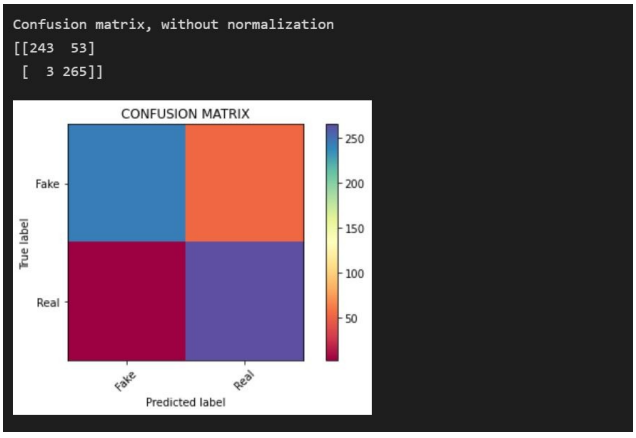


Fig. 12. Confusion matrix of SVM

- Area Under the Curve(AUC) Area UnderCurve or AUC is one of the most widely used metrics for model evaluation. It is generally used for binary classification problems. AUC measures the entire two-dimensional area present underneaththe entire ROC curve.

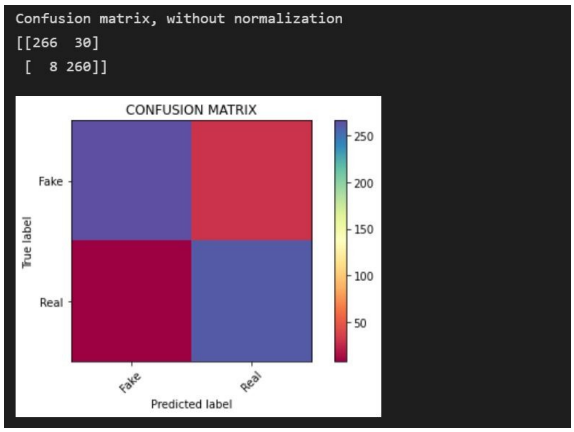


Fig. 12. Confusion matrix of Random Forest

V. CONCLUSION

The proposed framework for fake profile de-tection on social networking sites used feature-based dataset and the required features are selected manually. It is based on the user-level activities andthe user's account details.The study is carried out using the combined classifier SVM and Random Forest and with Neural Network. We achieved an accuracy of 99.4 for Neural Network and 94.2 for the hybrid SVM and Random Forest Classifier.

REFERENCES

[1] F. Ahmed and M. Abulaish, "An MCL-based approach for spam profile detection in online social networks," in Proc. IEEE 11th Int. Conf. Trust, Secur. Privacy Comput. Commun.,2012, pp. 602–608.

[2] A. M. Meligy, H. M. Ibrahim, and M. F. Torky, "Identity verification mechanism for detecting fake profiles in online social networks," Int. J. Comput. Netw. Inf. Secur., vol. 9,no. 1, pp. 31–39, 2017.

[3] S. Khaled, N. El-Tazi, and H. M. Mokhtar, "Detecting fake accounts on social media," in Proc. IEEE Int. Conf. Big Data.,2018, pp. 3672–3681.

[4] Y. Xuan, Y. Chen, H. Li, P. Hui, and L. Shi, "LBSNShield: Malicious account detection in location-based social net- works," in Proc. 19th ACM Conf.Comput. Supported Coop- erativeWork SocialComput.Companion, 2016, pp. 437–440.

[5] A. M. Al-Zoubi, J. Alqatawna, and H. Paris, "Spam profile detection in social networks based on public features," in Proc. 8th Int. Conf. Inf. Commun. Syst., 2017, pp. 130–135.

DOI: 10.5281/zenodo.6758955
***ISBN**: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*