Identifying Fake Profile in Online Social Network

Himanshi Gupta Nagariya, Neha Dhanotiya, Shruti Joshi and Sarika Jain

National Institute of Technology, Kurukshetra, India

Abstract

Online Social Networks involve a huge amount of people from all over the world and it has become a big part of their life. People use social networks to share their feelings, to make new friends, to set up new businesses, to connect with friends and family and what not. The Online Social Networks provides a great advantage to individuals in different ways but it also suffers with some disadvantages. There are many people who use these networks to cause harm to others by making fake accounts on these networks. For detection of such fake and genuine accounts we can use machine learning algorithms. The machine learning algorithms are applied for the prediction and classification of datasets through the different models that are prepared. It sometimes become difficult to differentiate between the results of different models and so we to use a hybrid approach of machine learning algorithm can make this task easy. In our work we compared the 8 different combinations of classification algorithms and calculated their accuracy on the dataset of an Online Social Network. We used the combination of Random Forest, Support Vector Machine, Logistic Regression, KNN, and Decision Trees. After comparing the result of each hybrid approach, we concluded that the best accuracy was obtained by combination of SVM and Logistic Regression and Neural Network. So, we proposed a model for the detection of fake account with the hybrid approach giving the best accuracy among all the combinations.

Keywords

Online Social Network, Fake Account Detection, Feature Extraction, Spammer.

1. Introduction

Machine Learning is a branch of artificial intelligence (AI) which is able to provide a system the ability to act without being programmed explicitly. It is used in many fields like Google cars, recommendation engines, friend suggestions in social media networks, shopping apps, cybercrimes etc.

Machine Learning has made a phenomenal change in the way how data was extracted and interpreted by replacing the old statistical techniques. Classifications of machine learning techniques are: Reinforcement, Supervised and Unsupervised Machine Learning.

Our work is concerned with the Classification algorithms that come under the Supervised Machine Learning. Classification is a supervised learning approach in which the machine takes the input data learns from that data and then further classifies the testing data according to its training data.

Although classification algorithms (Support Vector Machine, Logistic Regression, Decision Tree, Random Forest, Artificial Neural Network) can be used separately and individually but in our system we are developing a hybrid model combining two or three machine learning models has helped in increasing the accuracy of the model and its predicative power. The fact that which hybrid model will perform better is unknown, but it is also affected by the dataset provided and also the feature selection. The concept to develop a hybrid model is in a two- stage manner, first using clustering or classification techniques for pre-processing

ACI'21: Workshop on Advances in Computational Intelligence at ISIC 2021, February 25–27, 2021, New Delhi, India
EMAIL: himanshi100497@gmail.com (H. Nagariya);
nehadhanotiya1612@gmail.com (N. Dhanotiya);
shrutijoshijma@gmail.com (S.Joshi);
jasarika@nitkkr.ac.in (S. Jain)
ORCID: 0000-0002-7432-8506 (S. Jain)

of data and in second stage the output of the first stage to build second stage predictive classifier. It can be made using different algorithms of supervised or unsupervised learning but in our work, we developed the model using classification algorithms of supervised learning. Our main contribution is to propose a hybrid approach of machine learning algorithms and to compare the hybrid of different classification algorithms. Eight different experiments were conducted, and the accuracy thus obtained was compared.

The total number of users in online social networking sites is continuously increasing and with that the number of fake accounts is also increasing. As in September 2019, monthly active users on Facebook are 2.45 billion worldwide. According to Alexa, after Google and YouTube the third most visited website is Facebook. In a survey it is found that there are a greater number of female accounts in the world than the total population of female. From this, we can infer how many fake profiles have been created. According to Statistics April 2018 stats report, Facebook has more than 336 million active Twitter accounts, but Facebook is the leader with 2,196 million users worldwide. In September 2019, monthly active users on Facebook are 2.45 billion, of which India has the most. 270 million users. People who log on to Facebook daily are approximately 1.62 billion. And among these 83 million accounts are fake on Facebook. This statistics was given by Facebook in their Wall Street reports (SOURCE: Zephoria Digital Marketing). Figure 1 shows the monthly active users in the year 2019 on various OSNs.

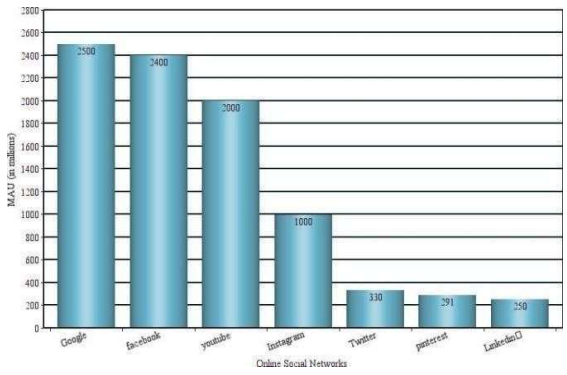


Figure 1: Monthly active users in different OSNs in year 2019

1.1. Motivation

As the number of people using OSN increases, so does the fake social media accounts creation. The main motivational factor in identifying those fake accounts is the cyber-crime rate, as these accounts were created primarily to commit cyber robbery or to commit cybercrime anonymously or unidentified is a significant increase from last few years. Fake account owners also try to take advantage of people's kindness by composing fake messages and spreading false news through these fake accounts in order to steal money from sinless people. In addition, people want to create multiple accounts that don't belong to anyone, created just to raise votes in an online voting system, and receive referral incentives, as in online games.

The detection of fake accounts in OSN attracts many researchers, so several algorithms for detection of fake accounts have been developed using machine learning techniques and various functions to connect to the account. Spammers can also find ways to support such techniques. These security technologies provide sophisticated detection mechanisms that require the continuous development of new approaches to spam detection. The main hazards in detection of fake accounts are to achieve accuracy and response time in the analysis of characteristics.

1.2. Challenges

Modeling a Fake Profile Detection System is an old problem but due to the many challenges this problem presents there still exist a lot of gaps that have been identified and need to be worked upon. The many challenges this system presents have been listed below:

- The data is not readily available: accounts on online social networks are highly private and protected, so the networking sites do not reveal any account information to maintain the confidential nature and keep the trust of their users.
- There is a lot of overlapping between genuine and fake accounts: At times the feature set of legitimate and fake

accounts overlap, and this poses a considerable setback when it comes to training the neural network by making it learn the pattern to differentiate between them.

- The number of parameters to process: The enormous number of parameters between learning and decision making is a major obstacle in developing systems for detecting fake accounts.
- Selection of optimal features (variables) is a big challenge: When it comes to optimal feature selection, it needs to be really dealt with care as the performance of whole system depends on which features it's taking into consideration for classification of fake and genuine accounts. And at times it's really perplexing to decide on these optimal features.
- Ability to handle noise in the data: Noise means missing or incorrect data which poses challenges while processing the dataset. There is no means by which we can make up for this lost information as such systems aren't partition tolerant, so this adversely affects the outcome.
- Heterogeneity in features.
- Single user multiple accounts.
- Many of the times it resembles a legitimate transaction: At times the fake account activities are stacked up in close resemblance with the legitimate ones. Hence, it becomes difficult to comprehend them and abort them before they make it to completion.

1.3. Gaps Identified

- We can extend the evaluation of propose feature by testing on different social networking sites like Facebook, Twitter etc. as most the previous researches were done on any one social site among Facebook, Twitter, LinkedIn, Myspace etc.
- The existing system does not work for the real time accurately on changing the features.
- Identification of rumor sources on social media by using the content-based features.

1.4. Related Work

Fake Profile Detection is an old problem and there has been a lot of work done in providing an optimal solution. But the fact that the mannerism of fake accounts keeps on evolving with time and there are enormous numbers of challenges and gaps still left to tackle, this problem still has a lot of significance. In order to study the work already done on Fake Account Detection we searched articles and research papers on two major sources: i) general online indexing websites, ii) publisher databases. Examples of former are Research Gate, Towards Data Science, IEEE Explorer and Google Scholar and examples of latter are Scopus, Springer, ACM Digital Library and Elsevier databases.

The major machine learning techniques we used in detection of fake accounts are Neural Network, Support Vector Machine, Random Forest, and Hybrid Models for Comparative analysis of Fake Account Detection.

Yang et al. trained SVM using the ground- truth obtained by Ren Ren for detecting fake accounts. By making use of simple features like frequency of friend requests, accepted requests and per-account clustering coefficient they trained the classifier and got 99% true-positive rate (TPR) and 0.7% false- positive rate (FPR). Íntegro draws out low-cost features from user-level activities to train the classifier for the identification of undetermined victims in social graph and used feature-based detection.

A different approach for hybrid was introduced by Mateen et al., by using content- based features like total number of tweets, hash tag ratio, URL's ratio and some graph- based features also and used the dataset of Twitter. They also made a comparison J48, Decorate and Naïve Bayes in which Decorate was the best performer. Somya et al.'s approach was quite different from others for detection as they tried to detect the account as fake on the user's homepage using Chrome extension which runs on the user site. Along with this they used Petri net based solution for the identification of source of malicious content running on Pn2 simulator environment.

Using a support vector machine and a neural network, Khaled et al [20] obtained 98% accuracy and compared the accuracy obtained by the hybrid of SVM and NN.

BalaAnand et al. [3] achieved 90.3% accuracy using a random forest classifier, support vector machine, and k-nearest neighbor method. For their work, Gupta et al [7] selected a dataset on Twitter and used a labeled dataset with a specific user and tweet feature. They used a hybrid of naive algorithms to classify, cluster, and make highly accurate decisions.

1.5. Organization

In our work we have implemented various algorithm to find the most efficient algorithm. To do so we have conducted several experiments and compared their results. Further, in this paper we have three sections which are briefly define below:

This section is followed by Section 2, System Architecture. In this section, flow diagram and architecture of our work is introduced and is described in brief.

In Section 3, Experimental Results, of up to now what modules we have implemented is shown along with pseudocode and discussed the various results produced by our system and have shown the outputs generated on various inputs in the form of the graph for the better understanding and algorithm of the technique is also mentioned in this section.

In Section 4, Conclusion, we provide an understanding of the overall conclusion of the proposed solution i.e. the combination of the techniques which is efficient than others and is given better accuracy.

2. System Architecture

Although fake profile detection is a robust field, but it has many challenges and gaps which we have discussed and have based our work on. There are a lot of existing solutions to fake profile detection but all of them have some or the other drawback. There is a lot of work already done in this field and a lot more needs to be done like improving upon the response time, prevention from fake accounts instead of detecting and dealing with their aftermaths. Our work is aiming to deliver a system which will have the highest accuracy and hence will be effective in prevention from such fake profiles by implementing and comparing different algorithms. This is done by ensemble machine learning technique which speeds up the training of neural networks and helps them to take decisions faster. Efficient parameter selection is also one of the major objectives of this work for which we are selecting six features manually which will give a better control on the output of neural networks. The proposed solution makes use of the hybrid of the machine learning techniques and combines their advantages and uses one to cancel out the loopholes of the other and hence delivering an efficient and cost-effective system.

In our proposed system we are aiming to design a hybrid system using artificial neural network, support vector machine and logistic regression that will be able to precisely and accurately detect fake profiles in online social network. Goal of the work is to maximize the accuracy and to minimize the time required by using hybrid approach of the Neural Network, Support vector machine and Logistic Regression.

Figure 2 depicts the flowchart of our system. The dataset which we have is partitioned into two sets, Train Dataset and Test Dataset in the ratio 4:1.The train dataset then goes into Support Vector Machine and Logistic Regression Classifier where classes are predicted. Then these classifiers are appended to a voting classifier where final decision of class is made. The output from voting classifier i.e. train data and the predicted class from voting classifier is fed to Neural Network classifier as input. After training has been completed, we get a Trained System on which Test dataset is ran to find the accuracy of the system.

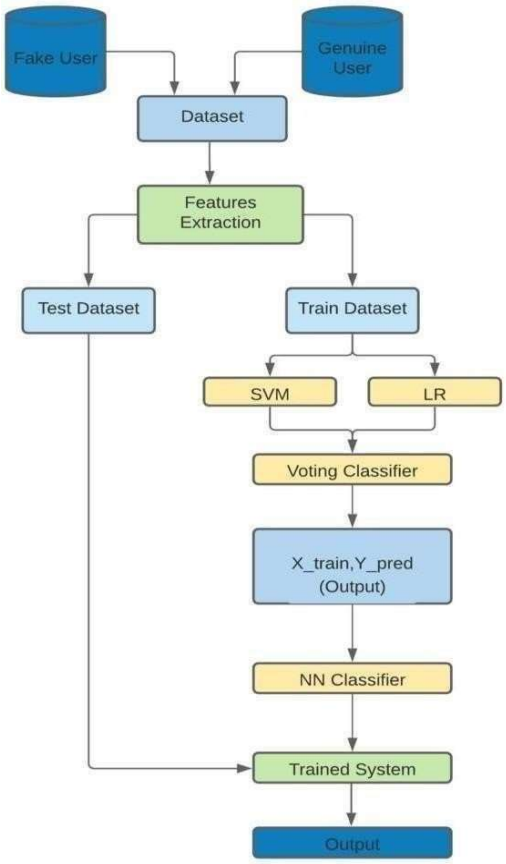


Figure 2: Flowchart of proposed system

Figure 3 depicts the architecture of the proposed system in which the first step is collection of data of any social networking sites in which you want to detect the fake accounts. In our proposed work we collect the data from the web sources. And then the data is preprocessed by using feature extraction techniques in our work we manually select the features. And then training of data is there and then pass the result in voting classifier and then training and testing of data in neural network classifier and then we got the result in the form of fake and real accounts.

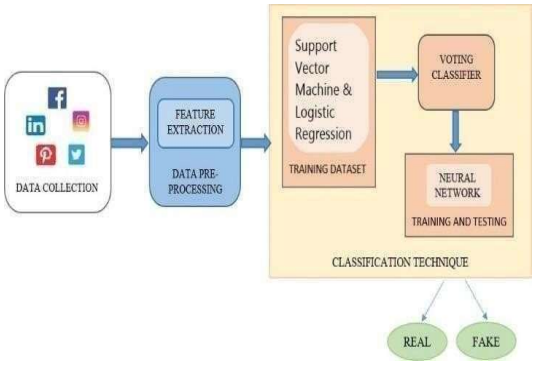


Figure 3: Architecture of proposed system

2.1. Algorithm for System

Algorithm

INPUT: The dataset from CSV files
OUTPUT: Accuracy

1. Read dataset: Read genuineusers.csv and fake users.csv and append them in a list, named x, and make list y for labelling class.
Return x,y

2. Feature Extraction: Convert non-integer features in dataset to integer. Store and overwrite selected 6 features in list x.
Return x

3. Split data into training data and test data using 5 cross validation and store them separately in x_train, x_test, y_train, y_test.

4. Scaling of the X_data. (x_train, x_test)

5. Use ensemble classifier, voting classifier, with SVM and Logistic Regression.

6. Store result in y_pred variable and
Return y_pred.

7. Repeat step 3 with y pred and x_test

8. Output from step 3 is given to Neural Network and then store the output in y_pred.

9. Testing: Evaluating our trained model against the test data. The output is visual graph consisting of True_Positive_Rate and False_Positive_Rate with accuracy, i.e, ROC curve.

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

11. Exit

3. Experimental Result

No proposal can be modeled into a system without some experiments to support it. In this section we have included the results and outputs produced during experiment with our system and by our system under various inputs and parameters.

3.1. Implementation Details

Each phase of our proposed system is briefly described in this section along with description, results at each stage are also provided.

3.1.1. Data Collection

For the model to work upon, there is a need for data collection. The dataset can be collected 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. Figure 5 shows the sample of csv file. And the code for reading both the files are:

```
genuineusers=pd.read_csv("users.csv")
fakeusers= pd.read_csv("fusers.csv")
,
```

3.1.2. 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 we import the useful libraries which will rescale or clean our data and the libraries we import are numpy, panda, scikit-learn and from sklearn we import preprocessing to clean our data.

Now in the next part for data pre-processing we use feature extraction technique first we try the principal component analysis technique and then we use the genetic algorithm and then after we select the features manually and we compare the result obtained from three ways and we get better result from the manually selection of features and the features we select manually are:

- statuses_count
- followers_count
- friends_count
- favourites_count
- listed_count
- lang_code

The language code feature is of string type we convert it into integer. After calling extract feature function it prints the extracted feature name and describes the entire extracted feature in summarized by printing mean, quartile, count, std, min, max etc.

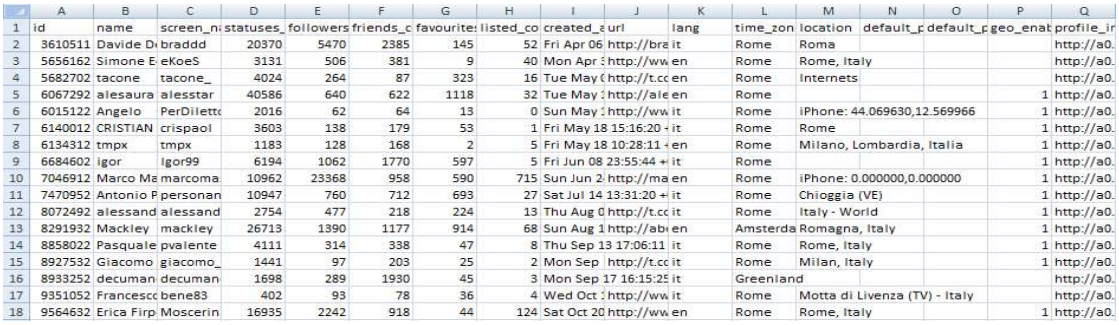
Figure 4 shows the data distribution in each column or feature in terms of count, mean, standard deviation, minimum and maximum values, and average of 25%, 50% and 75% of the data points when taken in ascending order.

	statuses_count	followers_count	...	listed_count	lang_code
count	2818.000000	2818.000000	...	2818.000000	2818.000000
mean	1672.198368	371.105039	...	2.818666	2.851313
std	4884.669157	8022.631339	...	23.480430	1.992950
min	0.000000	0.000000	...	0.000000	0.000000
25%	35.000000	17.000000	...	0.000000	1.000000
50%	77.000000	26.000000	...	0.000000	1.000000
75%	1087.750000	111.000000	...	1.000000	5.000000

Figure 4: Data distribution in each column

3.1.3. Training of Classifiers

As we are using the hybrid approach of the techniques in our proposed system, so we have done experiments with six techniques i.e. SVM, RF, LR, DTC, NN, KNN and finalize the techniques that gives the best result and they are Support Vector Machine, Logistic Regression and Neural Network. First we train our data using support vector machine independently and then we train our data on Logistic Regression independently and after



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	id	name	screen_n	statuses_	followers	friends_c	favourite	listed_co	created_a	url	lang	time_zon	location	default_p	default_p	geo_enat	profile_ir
2	3610511	Davide D	braddd	20370	5470	2385	145	52	Fri Apr 06	http://bra	it	Rome	Roma				http://a0.
3	5656162	Simone E	eKoeS	3131	506	381	9	40	Mon Apr 3	http://ww	en	Rome	Rome, Italy				http://a0.
4	5682702	tacone	tacone_	4024	264	87	323	16	Tue May (	http://t.cc	en	Rome	Internets				http://a0.
5	6067292	alesaura	alesstar	40586	640	622	1118	32	Tue May 1	http://ale	en	Rome				1	http://a0.
6	6015122	Angelo	PerDilett	2016	62	64	13	0	Sun May 1	http://ww	it	Rome	iPhone: 44.069630,12.569966			1	http://a0.
7	6140012	CRISTIAN	crispaol	3603	138	179	53	1	Fri May 18	15:16:20 +	it	Rome	Rome			1	http://a0.
8	6134312	tmpx	tmpx	1183	128	168	2	5	Fri May 18	10:28:11 +	en	Rome	Milano, Lombardia, Italia			1	http://a0.
9	6684602	igor	igor99	6194	1062	1770	597	5	Fri Jun 08	23:55:44 +	it	Rome				1	http://a0.
10	7046912	Marco Ma	marcoma	10962	23368	958	590	715	Sun Jun 2	http://ma	en	Rome	iPhone: 0.000000,0.000000			1	http://a0.
11	7470952	Antonio P	personan	10947	760	712	693	27	Sat Jul 14	13:31:20 +	it	Rome	Chioggia (VE)			1	http://a0.
12	8072492	alessand	alessand	2754	477	218	224	13	Thu Aug 0	http://t.cc	it	Rome	Italy - World			1	http://a0.
13	8291932	Mackley	mackley	26713	1390	1177	914	68	Sun Aug 1	http://ab	en	Amsterda	Romagna, Italy			1	http://a0.
14	8858022	Pasquale	pvalente	4111	314	338	47	8	Thu Sep 13	17:06:11	it	Rome	Rome, Italy			1	http://a0.
15	8927532	Giacomo	giacomo_	1441	97	203	25	2	Mon Sep	http://t.cc	it	Rome	Milan, Italy			1	http://a0.
16	8933252	decuman	decuman	1698	289	1930	45	3	Mon Sep 17	16:15:25	it	Greenland					http://a0.
17	9351052	Francesc	bene83	402	93	78	36	4	Wed Oct 2	http://ww	it	Rome	Motta di Livenza (TV) - Italy				http://a0.
18	9564632	Erica Firp	Moscerin	16935	2242	918	44	124	Sat Oct 20	http://ww	en	Rome	Rome, Italy			1	http://a0.

Figure 5: Sample of CSV file

analyzing the result of both the classification techniques we merge both the techniques to check the accuracy of both of them together and hybrid approach of both the techniques gives us the best result and after training the data from both the voting classifier is used to get the best result from both and then passing value for any one of them and then we use 5 fold cross validation technique to avoid the situation of overfitting as in k-fold cross validation technique dataset in divided into k folds where 1 fold is used for validation or testing while others are used for training and in these way we can avoid the situation of overfitting. After getting the score of each fold final estimated score is printed and in these we got 0.91 and the accuracy on testing dataset is 99.56.and after that the confusion matrix is plotted which will gives us the 261 true positive value and 7 false negative value and 29 false positive and 267 true negative value and then we plot the normalized confusion matrix which gives us all the four (TP,TN,FP,FN) values in percentage form along with precision, recall, f1 score and support and all these are evaluation criteria. For fake recall we got is 0.98 and for genuine it is 1.00 and f1 score for both is 0.99 and overall accuracy is0.99.

3.1.4. Training of Neural Network

Figure 6 shows the training of neural network. Each line corresponds to each round of forward and backward propagation called epoch. For this instance, we have taken our epoch to be 10, total number of layers to be 3, it took approximately minutes and seconds to train the system with final accuracy and loss value to be respectively.

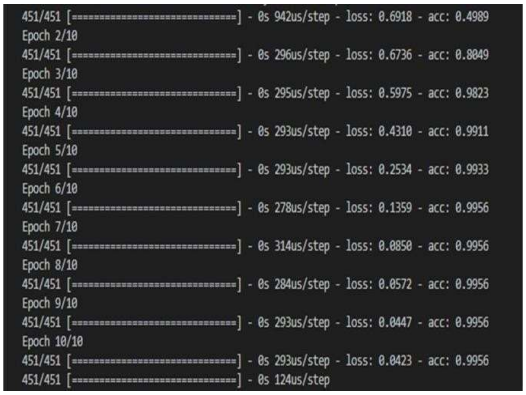


Figure 6: Training of Neural Network

Now the output produced by several hybrid techniques. We have collected two datasets say, D1 and D2 and the difference between these datasets is in their size, D2 is large as compared to D1. D2 contains approx. 3500 rows while D1 contains approx. 1500 rows. The results that we have obtained with different algorithms on both datasets are different and D2 gives less system with less accuracy as compared to D1.

Table 1:
Comparison of Accuracy (Support Vector Machine, Decision Tree Classifier, Logistic Regression Random Forest, Neural Network)

Hybrid Techniques	D2	D1
SVM+RF+NN	91.94%	96.01%
SVM+LR+NN	97.3%	99.56%
RF+LR+NN	93.32%	95.79%
SVM+DTC+LR+NN	96.34%	99.33%
SVM+DTC+RF+NN	92.87%	95.79%
SVM+DTC+NN	91.48%	96.45%
SVM+RF+KNN+NN	92.31%	97.12%

As we can see there is an accuracy difference between both datasets used by different algorithms so further, we will be working and showing results for only dataset, D1. We are using two csv files one is of genuine users and other one is of fake users.

Figure 7 shows the accuracy of each of our experimental model in ascending order and the model with highest accuracy being our trained system.

Figure 7: Accuracy of models in ascending order

Figure 8 shows the confusion matrices for our proposed hybrid model which gives us the summary of true positive, true negative, false positive and false negative without normalization.

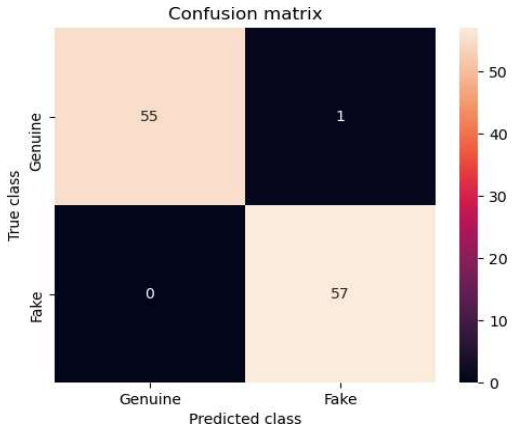


Figure 8: Confusion matrix of proposed hybrid model

Table 2 shows the results of the seven experiments that we performed using different combination of classification algorithms like Support Vector Machine, Random Forest, Logistic Regression, KNN with Neural Network. In the above table we can see that SVM, Log Reg, and NN is giving the maximum of true positive true negative resulting in maximum accuracy of all.

Table 2:
Results of combination of several techniques

Hybrid Techniques	TP	FP	FN	TN	Accuracy (%)
SVM+RF+NN	56	3	0	54	96.01
SVM+LR+NN	55	1	0	57	99.56
RF+LR+NN	55	4	0	54	95.79
SVM+DTC+LR+NN	54	2	0	57	99.33
SVM+DTC+RF+NN	55	4	0	54	95.79
SVM+DTC+NN	55	5	0	53	96.45
SVM+RF+KNN+NN	55	4	0	54	97.12

4. Conclusion

If we look at the system designs, majority of implementations for fake account detection is either graph-based or feature-based and they may use the graph analysis techniques or machine learning techniques to identification of accounts as fake or real. In

our proposed framework we use feature-based dataset and selected the features manually. This approach is based upon the user-level activities and the user's account details. We are comparing the hybrid approach of different classification algorithms and pass them in voting classifier and then pass the result in Neural network what we got from the voting classifier. In addition to our satisfying conclusion, we have maintained the highest accuracy in detecting fake accounts by testing and training the dataset on different hybrid approach of classification algorithms. The results show the increase of the accuracy results of the different classification algorithm.

5. References

[1] Joshi, Shruti, et al. "Identifying Fake Profile in Online Social Network: An Overview and Survey." International Conference on Machine Learning, Image Processing, Network Security and Data Sciences. Springer, Singapore, 2020.

[2] Mohanty, Sachi, et al. Recommender System with Machine Learning and Artificial Intelligence. Wiley-Scrivener, 2020.

[3] Balaanand, Muthu, et al. "An enhanced graph-based semi-supervised learning algorithm to detect fake users on Twitter." The Journal of Supercomputing 75.9 (2019): 6085-6105.

[4] Boshmaf, Yazan, et al. "Integro: Leveraging Victim Prediction for Robust Fake Account Detection in OSNs." NDSS. Vol. 15. 2015.

[5] Erşahin, Buket, et al. "Twitter fake account detection." 2017 International Conference on Computer Science and Engineering (UBMK). IEEE, 2017.

[6] Mateen, Malik, et al. "A hybrid approach for spam detection for Twitter." 2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST). IEEE, 2017.

[7] Gupta, Arushi, and Rishabh Kaushal. "Improving spam detection in online social networks." 2015 International conference on cognitive computing and information processing (CCIP). IEEE, 2015.

[8] Rahman, Sazzadur, et al. "Detecting malicious Facebook applications." IEEE/ACM transactions on networking 24.2 (2015): 773-787.

[9] Sahoo, SomyaRanjan, and Brij B. Gupta. "Hybrid approach for detection of malicious profiles in twitter." Computers & Electrical Engineering 76 (2019): 65-81.

[10] Kaur, Ravneet, and Sarbjeet Singh. "A survey of data mining and social network analysis based anomaly detection techniques." Egyptian informatics journal 17.2 (2016): 199-216.

[11] Jia, Jinyuan, Binghui Wang, and Neil Zhenqiang Gong. "Random walk based fake account detection in online social networks." 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). IEEE, 2017.

[12] Dhawan, Sanjeev. "Implications of Various Fake Profile Detection Techniques in Social Networks." IOSR Journal of Computer Engineering (IOSR-JCE), AETM'16 (2016): 49-55.

[13] Gurajala, Supraja, et al. "Fake Twitter accounts: profile characteristics obtained using an activity-based pattern detection approach." Proceedings of the 2015 International Conference on Social Media & Society. 2015.

[14] Xiao, Cao, David Mandell Freeman, and Theodore Hwa. "Detecting clusters of fake accounts in online social networks." Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security. 2015.

[15] Adikari, Shalinda, and Kaushik Dutta. "Identifying fake profiles in linkedin." arXiv preprint arXiv:2006.01381 (2020).

[16] Al-Qurishi, Muhammad, et al. "A prediction system of Sybil attack in social network using deep-regression model." Future Generation Computer Systems 87 (2018): 743-753.

[17] Masood, Faiza, et al. "Spammer detection and fake user identification on social networks." IEEE Access 7 (2019): 68140-68152.

[18] Cresci, Stefano, et al. "Fame for sale:Efficient detection of fake Twitter followers." Decision Support Systems 80 (2015): 56-71.

[19] Yang, Zhi, et al. "Uncovering social network sybils in the wild." ACM Transactions on Knowledge Discovery from Data (TKDD) 8.1 (2014): 1-29.

[20] Khaled, Sarah, Neamat El-Tazi, and

Hoda MO Mokhtar. "Detecting fake accounts on social media." 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018.

[21] Gupta, Aditi, and Rishabh Kaushal. "Towards detecting fake user accounts in facebook." 2017 ISEA Asia Security and Privacy (ISEASP). IEEE, 2017.

[22] Benevenuto, Fabricio, et al. "Detecting spammers on twitter." Collaboration, electronic messaging, anti-abuse and spam conference (CEAS). Vol. 6. No. 2010. 2010.

[23] Stein, Tao, Erdong Chen, and Karan Mangla. "Facebook immune system." Proceedings of the 4th workshop on social network systems. 2011.