MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY

APRIL 2018

DETECTION OF FAKE PROFILES ON SOCIAL MEDIA

MINOR PROJECT

Submitted By

Kratika Pandey (151112014)
Aditya Kumar Sharma (151112113)
Kapeesh Kumar Sharma (151112088)
Vidyanand Kumar (151112098)

Under the guidance of

PROF. DHIRENDRA PRATAP SINGH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that **Kratika Pandey, Aditya Kumar Sharma, Kapeesh Kumar Sharma and Vidyanand Kumar,** students of B.Tech. 3rd year (Computer Science And Engineering), have successfully completed their Project "Detection Of Fake Profile On Social Media" in their Minor Project in Computer Science and Engineering .

Dr. SANYAM SHUKLA **Dr. DHIRENDRA PRATAP SINGH**

(COORDINATOR) **(PROJECT GUIDE)**

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We hereby declare that the following report which is being presented in the Minor Project Documentation entitled “DETECTION OF FAKE PROFILE ON SOCIAL MEDIA” is the partial fulfillment of the requirements of the third year (sixth semester) Minor Project in the field of Computer Science and Engineering. It is an authentic documentation of our own original work carried out under the guidance of Dr. Dhirendra Pratap Singh. This work has been carried out entirely at Maulana Azad National Institute of Technology, Bhopal. The following project and its report, in part or whole , has not been presented or submitted by us for any purpose in any other institute or organisation.

Kratika Pandey **151112014**
Aditya Kumar Sharma **151112113**
Kapeesh Kumar Sharma **151112088**
Vidyanand Kumar **151112098**

ACKNOWLEDGEMENT

With due respect, we express our deep sense of gratitude to our respected guide Dr. Dhirendra Pratap Singh, for his valuable help and guidance. We are thankful for the encouragement that he has given us in completing this project successfully. His rigorous evaluation was of great assistance.

We are also grateful to our respected HOD, Dr. Meenu Chawla, for the additional help and support extended in allowing us to use the departmental laboratories and other services.

We are thankful to all the other faculty, staff members and laboratory attendants of our department for their kind cooperation and help.

We would like to express our deep appreciation towards our family members and batch mates for providing the much-needed support and encouragement.

ABSTRACT

In the present generation, social network has become an integral part of our life. Online media and networking sites have had a huge impact on our lifestyle. Making friends and keeping contact with them is easier now. But, with all the advancements and growth, problems like fake profiles, online impersonation have also grown. According to Facebook's record in May 2015, there were over 170 million fake accounts. In this project, we came up with a framework through which we can classify a profile as genuine or fake. We have used two different classification techniques, namely, Support Vector Machine and Neural Network to classify the profiles into fake and genuine classes. As, this is an automatic detection method, it can be applied easily by online social networks which has millions of profile whose profiles can not be examined manually.

Table Of Contents

List of Figures

1. INTRODUCTION

A social networking site is a website where each user has a profile and can keep in contact with friends, share their updates, meet new people who have the same interests. These Online Social Networks (OSN) use web2.0 technology, which allows users to interact with each other. Social networking sites are growing rapidly and changing the way people keep in contact with each other. The online communities bring people with same interests together which makes users easier to make new friends.

1.1 HISTORY

Early social networking on the World Wide Web began in the form of generalized online communities such as Theglobe.com (1995), Geocities (1994) and Tripod.com (1995). In the late 1990s, user profiles became a central feature of social networking sites, allowing users to compile lists of "friends" and search for other users with similar interests. Facebook, launched in 2004, became the largest social networking site in the world in early 2009. Facebook was first introduced as a Harvard social networking site, expanding to other universities and eventually, anyone. The term social media was introduced and soon became widespread.

1.2 SOCIAL IMPACT

In the present generation, the social life of everyone has become associated with the online social networks. Adding new friends and keeping in contact with them and their updates has become easier. The online social networks have impact on the science, education, grassroots organizing, employment, business, etc. Researchers have been studying these online social networks to see the impact they make on the people. Teachers can reach the students easily through this making a friendly environment for the students to study, teachers nowadays are

getting themselves familiar to these sites bringing online classroom pages, giving homework, making discussions, etc. which improves education a lot. The employers can use these social networking sites to employ the people who are talented and interested in the work, their background check can be done easily.

1.3 ISSUES

The social networking sites are making our social lives better but nevertheless there are a lot of issues with using these social networking sites. The issues are privacy, online bullying, potential for misuse, trolling, etc. These are done mostly by using fake profiles.

1.4 OBJECTIVE

In this project, we came up with a framework through which we can detect a fake profile using machine learning algorithms so that the social life of people become secured.

1.5 CHAPTER ORGANIZATION

Chapter 2 discusses about the various social media attacks and how social media is misused by people through social engineering to make money or for the purpose of defaming other people.

Chapter 3 focuses on various data structures, algorithms and techniques used for detection of fake profiles. These include support vector machines and neural networks.

Chapter 4 discusses implementation and result involving accuracy of the algorithm facilitated by the confusion matrix and ROC curve.

2. LITERATURE REVIEW

Fake profiles are the profiles which are not genuine i.e. they are profiles of persons who claim to be someone they are not, doing some malicious and undesirable activity, causing problems to the social network and fellow users.
Why do people create fake profiles ?
• Social Engineering.
• Online impersonation to defame a person.
• Advertising and campaigning a person, etc.

2.1 SOCIAL ENGINEERING

Social engineering is the art of manipulating people so they give up confidential information. The types of information these criminals are seeking can vary, but when individuals are targeted the criminals are usually trying to trick you into giving them information.
The social engineering techniques are like Pretexting, Diversion theft, phishing, baiting, quid pro quo, tailgating, etc.
Eg: Creating a profile of some person X not in some online social networking site like facebook. Adding the friends of the X in facebook and making them believe that its the profile of X. They can get the private information meant for only X by communicating with Xs friends in facebook.

2.2 ONLINE IMPERSONATION TO DEFAME A PERSON

The other reason why people create fake profiles is to defame the persons they do not like. People create profiles in the name of the people they don’t like and post abusive posts and pictures on their profiles misleading everyone to think that the person is bad and thus defaming the person.

2.3 ADVERTISING

Facebook disclosed that it found about 3,000 ads that ran during the 2016 US election that were run by fake accounts linked to Russia. There is evidence that they were designed to influence voters by promoting polarizing topics. It was a wild abuse of Facebook's automated ad platform. The review posted by a genuine user is always desirable but reviews when posted by fake profiles are completely undesirable.

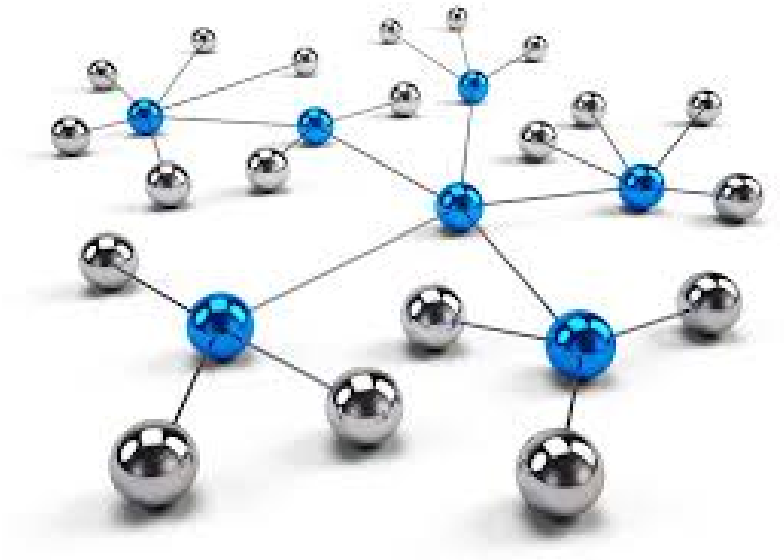

Fig 2.1: Social Influence via social network [10]

Fig 2.1. shows a social graph where the blue nodes shown are real profiles, the silver nodes show fake profiles and the edges show the connections between them. If the fake profiles start advertising a brand or campaigning for some politician then the users connected to the fake profiles are misled in believing them. Inturn the profiles who didn't add the fake profiles are affected using the mutual connections.

2.4 SOCIAL BOTS

A socialbot is a type of bot that controls a social media account. Like all bots, a socialbot is an automated software. The exact way a socialbot replicates

depends on the social network, but unlike a regular bot, a socialbot spreads by convincing other users that the socialbot is a real person. These are semi-automatic or automatic computer programs that replicate the human behavior in OSN. These are used mostly by hackers now-a-days to attack online social networks. These are mostly used for advertising, campaigning purposes and to steal users personal data in a large scale. The botmaster may or may not have inputs from a human attacker. The social bots look like human profiles with a randomly chosen human name, randomly chosen human profile picture and the profile information posted randomly from a list prepared from before by the attacker. These social bots send requests to random users from a list. When someone accepts the request, they send requests to the friends of the user who accepted the request, which increases the acceptance rate due to existence of mutual friends.

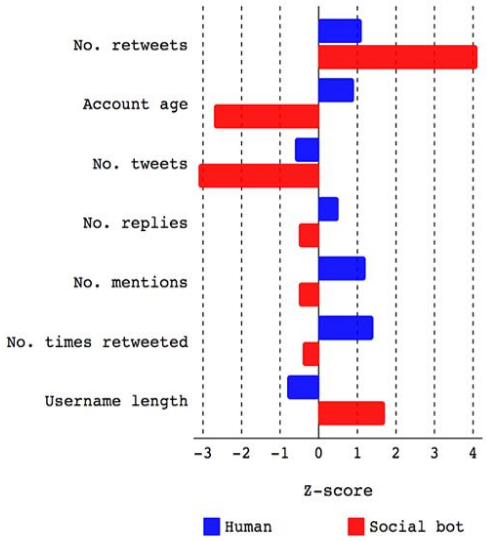

Fig. 2.2: Features discriminating social bots from humans [12]

2.5 FACEBOOK IMMUNE SYSTEM

Facebook has its own security system to protect its users from spamming, phishing, etc. and this is called facebook immune system. FIS does real time checks on every single click and every read and write operation done by it.

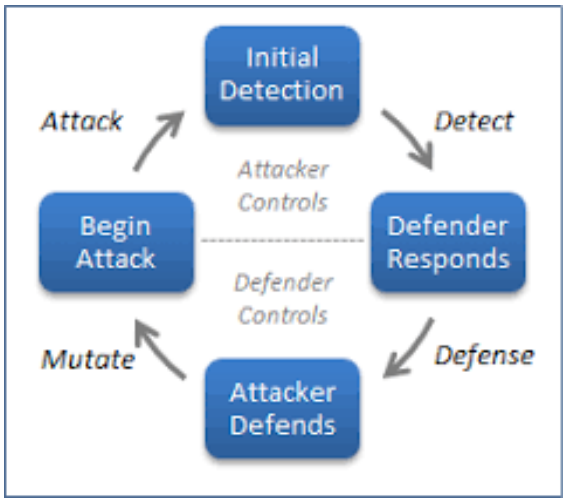

Fig. 2.3: The Adversarial Cycle [17]

2.5.1 HOW DO ATTACKS WORK?

Most attackers are in it to make money. They make money by distributing unwanted ads (spam) or capturing accounts they can reuse or resale (phishing). Attackers need resources to make a profit - fake accounts, real accounts, IP addresses, email accounts, and computing power. All these assets can have a significant cost associated with them, and an attack, like any business venture, needs profit to keep going.

2.5.2 HOW DO ATTACKERS USE ASSETS?

Attackers try to use Facebook accounts, Pages, Groups, Events, and Apps to steal login information, spam people, and ultimately make money. They need email accounts, cookies, and a wide range of IP addresses to circumvent reputation-based defenses. Additionally, they use phone numbers, stolen credit cards, and CAPTCHA solutions in an attempt to circumvent authentication checks.

2.5.3 A SNAPSHOT

As of early Nov 2011 here are some basic numbers that describe the Immune System:

- Runs on 2,000 servers.
- Checks 640,000 user actions per second at peak
- Runs over 1,000 rules in real-time across 150 different product channels.
- Aggregates and analyzes 5,000 different signals.
- Contains 200 different models.
- Does 20 billion classifications checks every day.
- Is used by our extensive support staff around the world 24/7 to detect and resolve threats.
- Is built and maintained by a team of 25 engineers, the Facebook Site Integrity team.

3. PROPOSED WORK

3.1 OVERVIEW

Each profile (or account) in a social network contain lots of information such as name, gender, number of friends, number of followers, number of likes, location etc. Some of these information are private and some are public. We have used information that are public to determine the fake profiles in social Network as private information is not accessible. However, if our proposed scheme is used by the social networking companies itself then they can use the private information of the profiles for detection without violating any privacy issues. We have considered these information as features of a profile for classification of fake and real profiles.

The steps that we have followed for detection of fake profiles are as follows.

1. Features are selected to apply classification algorithms.

The classification algorithm is discussed in the section 3.4 and 3.5. Attributes are selected as features if they are not dependent on other attributes and they increase efficiency of the classification. The features that we have chosen are discussed in section 4.2.

2. After selection of attributes, the dataset of profiles that are already classified as fake or genuine are needed for the training purpose of the classification algorithm.

We have used a publicly available dataset of 1337 fake users and 1481 genuine users consisting of various attributes including name, status count, number of friends, followers count, favourites, languages known etc.

3. The selected attributes are extracted from profile for the purpose of classification.

4. After this the dataset of fake and real profiles are prepared. From this dataset, 80% of both profiles (genuine and fake) are used to prepare a training dataset and 20% of both profiles are used to prepare a testing dataset.

5. The training dataset is then fed to the classification algorithm. It learns from the training dataset and is expected to give correct class labels for the testing dataset.

6. The labels from the testing dataset are removed and are left for determination by the trained classifier.
The result of classification algorithm is shown in 4.4. We have used two classification algorithms and have compared the efficiency of these algorithms.

3.2 PROPOSED FRAMEWORK

The proposed framework in the figure 3.1 shows the sequence of processes that need to be followed for continuous detection of fake profiles with active learning from the feedback of the result given by the classification algorithm.

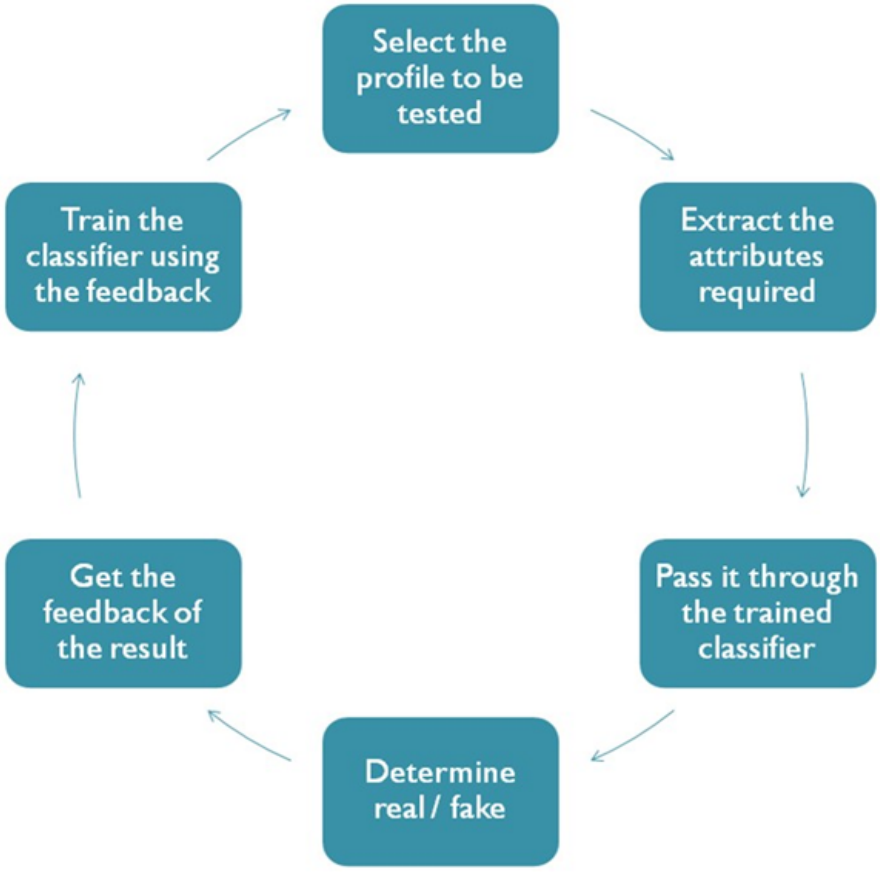

Fig 3.1 Framework for detection of fake profiles

This is a framework that can easily be implemented by social networking companies as they have access to user information.

1. Classification starts from the selection of profile that needs to be classified.

2. Once the profile is selected, the useful features are extracted for the purpose of classification.

3. The extracted features are then fed to trained classifier.

4. Classifier is trained regularly as new data is fed into the classifier.

5. Classifier then determines whether the profile is genuine or fake.

6. The result of classification algorithm is then verified and feedback is fed back into the classifier.

7. As the number of training data increases the classifier becomes more and more accurate in predicting the fake profiles.

3.3 CLASSIFICATION

Classification is a technique of categorizing an object into a particular class based on the training data set that was used to train the classifier. We feed the classifier with data set so that we can train it to identify related objects with as best accuracy as possible. Classifier is an algorithm used for classification. In this project we have used two classifiers namely Neural Networks and Support Vector Machines and have thereby compared their efficiencies.

3.4 NEURAL NETWORKS

The conventional method by which a computer works is that you provide instructions or algorithms to the computer and it generates output based on it. But what if you do not know the algorithm to solve a problem ? Will your computer still be able to provide solutions. If we use conventional techniques

then the computer will not be able to solve the problem unless you provide some instructions.

Here comes the concept of Neural Networks. We can still solve such a problem by training a network as such our program will learn on its own and will provide solution close to a certain accuracy. The term Neural Networks was coined in 1943 but could not be implemented then due to lack of technology. Neural Networks learn by example. Neural Networks are based on biological neurons i.e. brain cells and the way information is processed inside the brain. There are mainly two types of neural networks :

(1) Single Layer Networks also called a Perceptron.

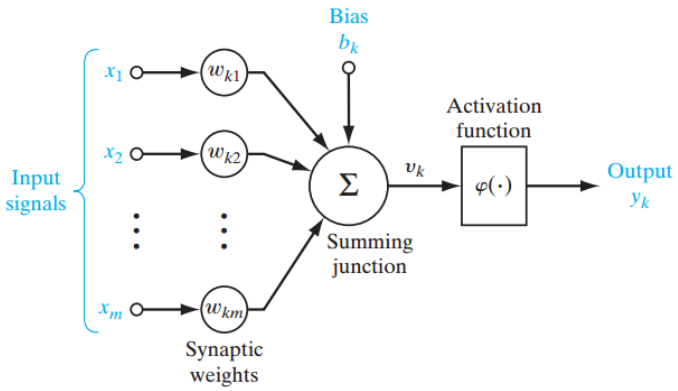

Fig 3.2 Single Layer Perceptron [14]

The above figure shows a perceptron.

(a) We provide inputs x1, x2, x3…

(b) Each input has a pre-assigned weight based on the priority of the input. Let the weights be w1, w2, w3…

(c) Each input is multiplied with its corresponding weight and the passed to the summation function.

(d) The output of the summation function is passed through the activation function. The activation function specifies a threshold value. If the output of summation is >= the threshold value then neuron will fire otherwise not.

(2) Multi - Layer Network

In multi layer network apart from the input layer and output layer there are hidden layers which cater on increasing the efficiency of classifying objects. In this network output of one hidden layer is used as input for another hidden layer. This network is also known as Deep Neural Network.

3.4.1 Back Propagation

The algorithm used in implementing Neural Networks is back propagation. The algorithm is as follows :-

(a) There are two types of outputs, model output and desired output. The difference between these two output is calculated which is the error. Mean Square Error is also calculated.

(b) The weights assigned to each input is either increased or decreased with the motive of minimizing the error.

(c) The model output is recomputed, error rechecked until the error cannot be minimized further. Minimum error is found by finding minima using gradient descent.

3.4.2 Neural Network Libraries

Neural Networks is implemented using tensorflow or pybrain in python. In this project we have used pybrain for classifying data into real and fake using neural networks.

3.5 SUPPORT VECTOR MACHINE

A Support Vector Machine (SVM) is a binary classifier that performs classification by finding a hyperplane that maximizes distance between two classes. It is a supervised machine learning algorithm. The algorithm outputs a hyperplane that fairly divides two classes with the help of training data and categorizes new examples. In two dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

Suppose you are given plot of two label classes on graph as shown in FIG 3.2. Can you decide a separating line for the classes?

Fig. 3.3: Draw a line that separates black circles and blue squares [13]

You might have come up with something similar to following image (*fig. 3.3)*. It separates the two classes. Any point that is left of line falls into black circle class and on right falls into blue square class. ***Separation of classes. That's what SVM does.***

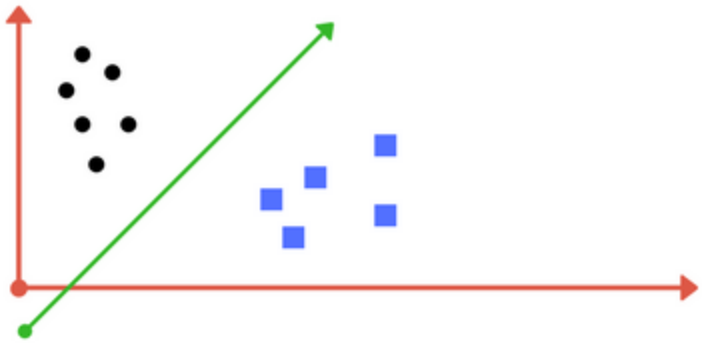

Fig. 3.4: Sample cut to divide into two classes. [13]

3.5.1 Selecting SVM Hyperplanes

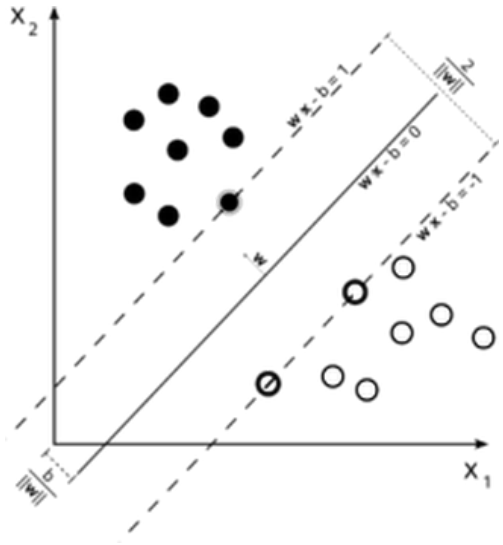

Fig.3.5:Support Vector Machine [15]

Linearly Separable: For the data which can be separated linearly, we select two parallel hyperplanes that separate the two classes of data, so that distance between both the lines is maximum. The region b/w these two hyperplanes is known as “margin” & maximum margin hyperplane is the one that lies in the middle of them.

$$\overline{w}\, x_i - b \geq 1 \quad \text{, if } \theta_i = 1$$

$$\overline{w}\, x_i - b \leq 1 \quad \text{, if } \theta_i = -1$$

Where ω is normal vector to the hyperplane, θi denotes classes & xi denotes features. For proper classification, we can build a combined equation:

$$||\overline{w}||_{min} \text{ for } \theta_i \quad (\overline{w}x_i - b) \geq 1 \quad \forall i = 1,2,.....,n$$

As it maximizes margin between two classes, SVM is a robust model to solve prediction problems.

3.5.2 SVM Libraries

For implementing SVM on a dataset, we can use libraries. There are many libraries available that can help us to implement SVM smoothly. It contains in-built functions that can be called whenever required.

In Python, we can use libraries like sklearn. For classification, Sklearn provides functions like SVC, NuSVC & LinearSVC. We pass values of kernel parameter, gamma and C parameter etc. By default kernel parameter uses “rbf” as its value. In this project we have used SVC function of Sklearn.

4. IMPLEMENTATION AND RESULTS

4.1 DATASET

We needed dataset of fake and genuine profiles. Various attributes included in dataset are number of friends, followers, status count. Dataset is divided into training and testing data. Classification algorithms are trained using training dataset and testing dataset is used to determine efficiency of algorithm. From the dataset used, 80% of both profiles (genuine and fake) are used to prepare a training dataset and 20% of both profiles are used to prepare a testing dataset.

4.2 ATTRIBUTES CONSIDERED

1. Status Count
2. Followers count
3. Friends Count
4. Favourites Count
5. Listed Count
6. Gender
7. Language Code

4.3 EVALUATION PARAMETERS

Efficiency/Accuracy = Number of correct predictions/ total number of predictions

Percent Error = (1-Accuracy)*100

Confusion Matrix - Confusion Matrix is a technique for summarizing the performance of a classification algorithm. Calculating a confusion matrix can give

you a better idea of what your classification model is getting right and what types of errors it is making.

TPR- True Positive Rate
TPR=TP/(TP+FN)

FPR- False Positive Rate
FPR=FP/(FP+TN)

TNR- True Negative Rate
TNR=TN/(FP+TN)

FNR- False Negative Rate
FNR=1-TPR

Recall- How many of the *true* positives were *recalled* (found), i.e. how many of the correct hits were also found.
Recall = TP / (TP+FN)

Precision- Precision is how many of the *returned* hits were *true* positive i.e. how many of the found were correct hits.
Precision = TP / (TP + FP)

F1 score- F1 score is a measure of a test's accuracy. It considers both the precision *p* and the recall *r* of the test to compute the score.

ROC Curve- The *Receiver Operating Characteristic* is the plot of TPR versus FPR. ROC can be used to compare the performances of different classifiers.

4.4 RESULTS

NEURAL NETWORK

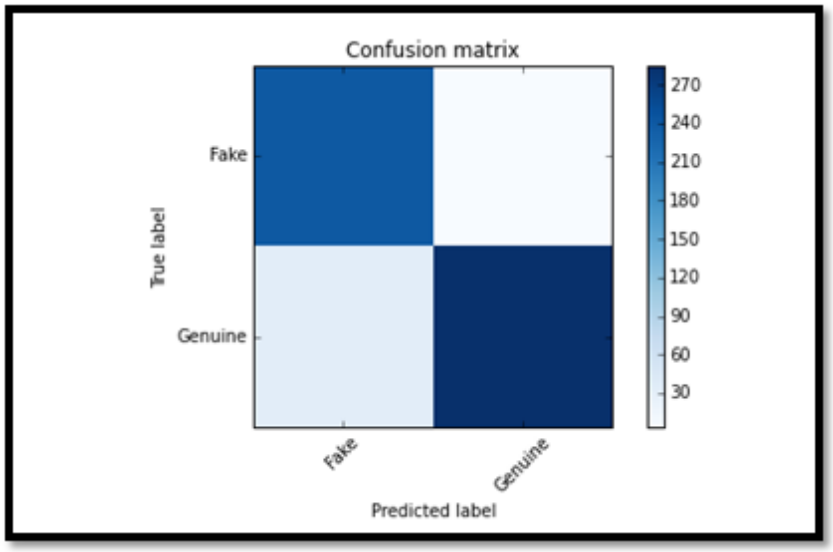

Fig 4.1- Confusion Matrix

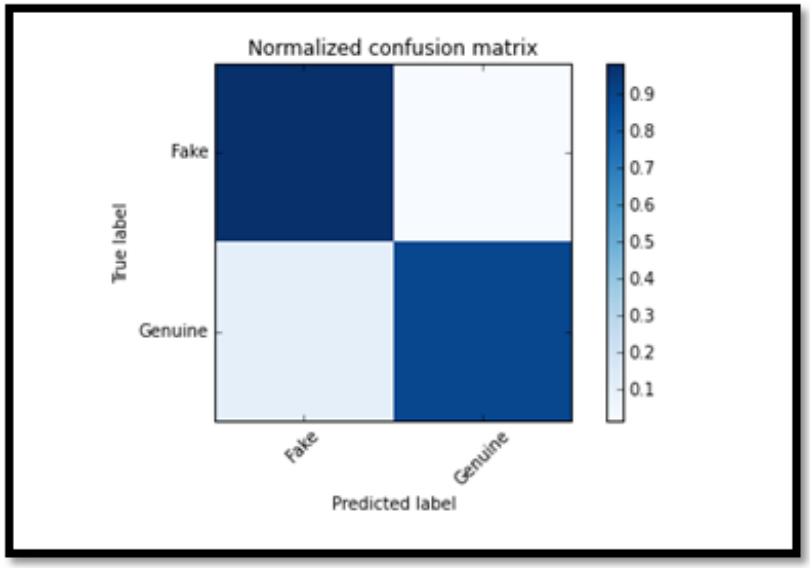

Fig 4.2- Normalized Confusion Matrix

	precision	recall	f1-score	support
Fake	0.88	0.98	0.93	245
Genuine	0.99	0.90	0.94	318
avg / total	0.94	0.93	0.93	563

Fig 4.3- Classification Report

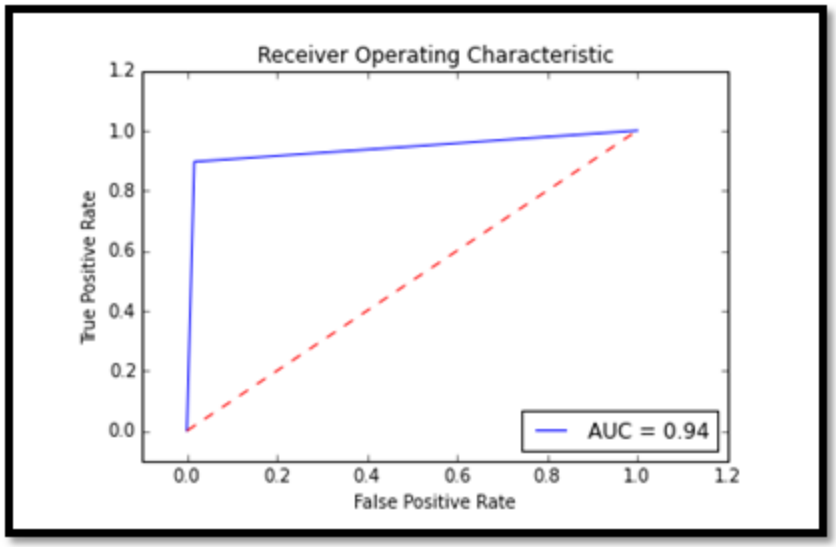

Fig 4.4- ROC Curve

Efficiency of Neural Network in classifying data is 93.4%. We have taken 80% of data for the purpose of training and 20% for classification.

SUPPORT VECTOR MACHINE

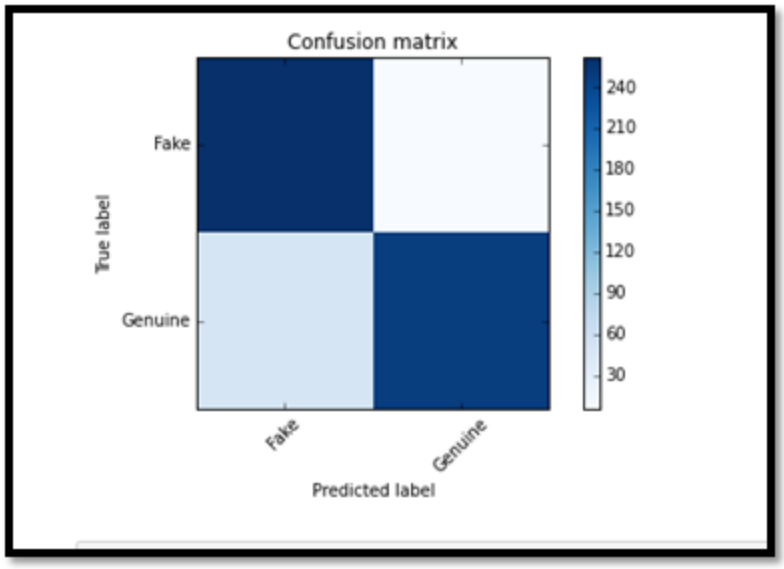

Fig 4.5- Confusion Matrix

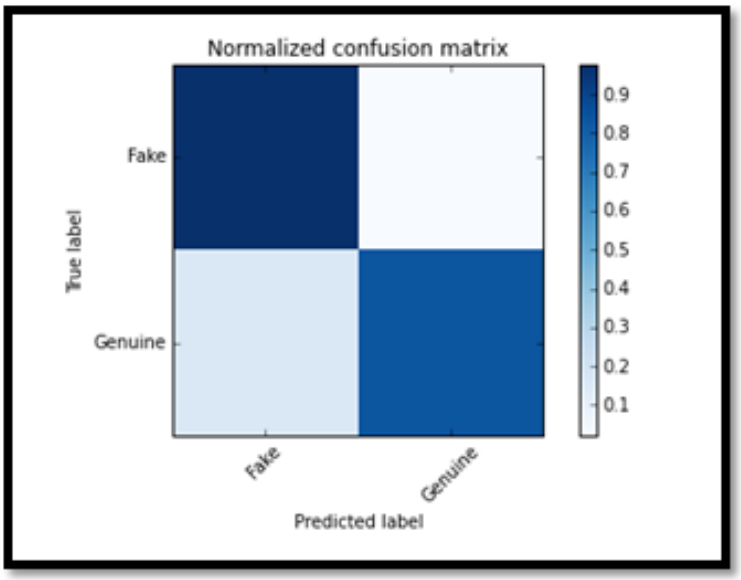

Fig 4.6- Normalized Confusion Matrix

	precision	recall	f1-score	support
Fake	0.85	0.98	0.91	268
Genuine	0.98	0.84	0.90	296
avg / total	0.91	0.90	0.90	564

Fig 4.7- Classification Report

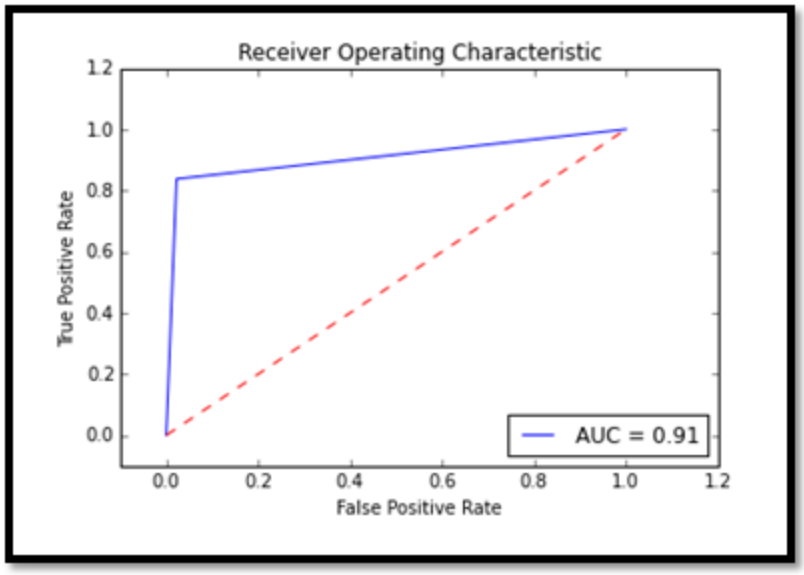

Fig 4.8- ROC Curve

Efficiency of SVM in classifying data is 91%. We have taken 80% of data for training SVM classifier and 20% for classification.

5. CONCLUSION AND FUTURE WORK

We have given a framework which suggests that binary classification through Neural Networks is more efficient than through Support Vector Machine. Using Neural Networks we have achieved efficiency of 93.4%. In the future we wish to classify profiles by taking profile pictures as one of the features.

6. REFERENCES

1. Andrew Ng, "Machine Learning" https://www.coursera.org/learn/machine-learning

2. freelancer.com

3. "Support Vector Machine" https://en.wikipedia.org/wiki/Support_vector_machine

4. "Chapter-2 Support Vector Machine(SVM) theory" https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72

5. Shahzeb Haidar "Facebook Immune System: A summary" http://home.iitk.ac.in/~shaidar/cs300/4B/4B.pdf

6. Jason Brownlee "Support Vector Machines for Machine Learning" https://machinelearningmastery.com/support-vector-machines-for-machine-learning/

7. "Social Bot" https://www.techopedia.com/definition/27811/socialbot

8. "Convolutional Neural Networks" http://cs231n.github.io/convolutional-networks/

9. "Artificial Intelligence - Neural Networks" https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm

10. Digitaltrends.com

11. "A Literature Review of Research on Facebook Use" https://www.researchgate.net/publication/270100677_A_Literature_Review_of_Research_on_Facebook_Use

12. Laila Sydney “Support Vector Machines” http://slideplayer.com/slide/4043773/

13. https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72

14. “Understanding Support Vector Machine algorithm from examples” https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machin

15. “Support Machine Learning” https://en.wikipedia.org/wiki/Support_vector_machine

16. Zeltser.com

17. Digitaltrends.com

18. “Practical Machine Learning Tutorial with Python(sentdex)” https://www.youtube.com/playlist?list=PLQVvvaa0QuDfKTOs3Keq_kaG2P55YRn5v