DETECTION OF DECEPTIVE ACCOUNTS USING MACHINE LEARNING ALGORITHM AND DEEP NEURAL NETWORK

[1]Jaladhi Pandya, [2]Prof. Gayatri Pandi(Jain)
[1]Research Scholar, [2]Assistant Professor
[1]Information Technology Department (IT),
[1]L.J. Institute of Engineering and Technology, Ahmedabad, Gujarat, India.

Abstract: Internet is become a world for social media and communication. We are usually using a social media like a Facebook, twitter and LinkedIn for communication and knowledge widening. The social media sites are misused for communication by creating fake accounts. The accounts do not have genuine data. Some people are using other's personal data for creating fake profile that account or profile called fake accounts. Spam legitimate users are posting inappropriate or illegal content. Social media sites are being used to spy and send malicious links and carry out financial frauds. Social media conman since they give permission to striker to build a bond with their targets. We are using different new features, which are more effectiveing and robusting compared to already exciting features like selection, generalize batter, interpretable. In our proposed system, we are using machine learning algorithm for detecting fake accounts on social media which are get more efficient and accurate as compared to the existing ones.

IndexTerms **- Social Media, Fake accounts, Real account detection, Machine learning.**

I. INTRODUCTION

Current generation we are using so many different social media. There is different social media like a Facebook, Twitter, etc. A social networking service work for as a platform to build social networks or social relation among people who, share interests, activities, backgrounds, or real life connections.

A social network generally offered to participants who register this site with their unique representation. One of the most common ways of performing a large scale data gathering attack is the use of fake accounts.

Where hostile users are present themselves in profiles impersonating fictitious or real persons. At attempt has been made in this paper to use of Machine learning.

Our day cannot start without mobile. In our mobile, laptop and desktop we are using different types of social media. For using that social media we need to create an account which is connect with us to the world and social network.

These social media is used by so many people and the purpose is to explore a world and connect with the society from which we get awareness.

But somehow people using these social media for harming someone and spreading negativity. So we need to detect fake accounts in social media.

Spamming is the fruition of messaging systems to send an unsought or blackballed, especially advertisement, fake messaging using fake account.

Spammers are targeting users of instant messaging service or private number or SMS. Some people they want harm others so they are creating fake accounts on different social media.

In our society most probably every person are using a different social media which is useful to communicate each other. There are some people which would like to harm intentionally.

So we need to detect that accounts and block that accounts because they are harming other person. They are sending some unnecessary detail and harmful videos and images which harm to our society. So that we are need to detect spam account. On sites such as Twitter, fake followers have been commonplace for years and are essentially a part of the culture.

Users are at least likely to have their accounts put into jeopardy as a result of having social media fans that are not real because the practice is so prevalent. In fact, many people use fake social media fans on Twitter because everyone is doing it – follower counts have become grossly inflated and in order to appear influential, many adopt this approach.

On other networks, the benefits are a bit more tangible. Those who are pinging websites from their pages on Facebook, for instance, need to be recognized in order to be successful.

The number of likes combined with activity has a direct correlation to exposure in select search results on the social media network. Whether you are a local business or a national movement, the number of likes and followers that you have plays a direct role in how many people within a given audience see your page.

II. LITERATURE REVIEW

1. Friend similarity criteria were calculated from the adjacency matrix of the network graph and new features were extracted from the PCA method.The cross validation technique is used in this paper. The classifier was trained and tested, which showed that the Medium Gaussian SVM classifier [1]. In this method, fake accounts must work in the network so that it will be possible to recognize them as legitimate or fake.

2. In this paper authors use seven machine learning algorithms, kNearest Neighbor (k-NN), Decision Tree (DT), Naive Bayesian (NB), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost) to classify spam and non-spam users and also used feature graph-based and content-based features that have been proved to be powerful for spam account detection on Twitter. Authors are detecting a fake accounts using twitter dataset. In this dataset there are 168 real accounts' and 157 spam users [2].

3. In this paper, the classifier is being trained regularly as new training data set is feed into the classifier. Dataset is divided into training and testing data. Classification algorithm is trained using training dataset and testing data set is used to determine the efficiency of algorithm. Authors using different feature like number of friends, number of, followers, language. They are using a publicly available dataset of 1337 fake users and 1481 real accounts [3].

4. In this paper they present a classification method for detecting the fake accounts on Twitter. The Research have preprocessed our dataset using a supervised discretization technique named Entropy Minimization Discretization (EMD) on numerical features and analyzed the results of the Naïve Bayes algorithm. In this paper research using twitter social account for detecting fake accounts. There is using confusion matrix which most common evolution metrics [4].

5. They consider that the hereby proposed methodology offers a safe way to identify the real user or users behind a trolling account given some previous conditions: i) the real user/s behind the fake profile has/have a "real" and active account in the social network, ii) the real account of the user/s behind the fake profile is/are somehow connected to the fake profile. These conditions are in theory easy to fullfil due to the assumption that a real person behind a trolling profile wants to keep track of the activities and parallel conversations surrounding the trolling profile [5].

6. Supervised machine learning algorithms require a dataset of features with a label classifying each row or outcome. Features are thus the input used by supervised machine learning models to predict an outcome. These features can be the attributes found via APIs that describes a single piece of information about an SMP account, like the number of friends. Features can also be engineered by combining attributes from an SMP account, past engineered features, and/or domain knowledge. An example of an engineered feature is the combination of the number of friends and followers to present their relationship as a ratio for input to a machine learning model [6].

III. METHODOLOGY FOR DECEPTIVE ACCOUNTS

There is malicious activity done in social media or social sites and hackers for hacking they are creating fake accounts. In other papers there is not get that much accuracy so that we are trying to getting as much as accuracy and also efficiency. So we are using machine learning algorithm random forest and in neural network deep neural network which give us good accuracy comparatively other machine learning algorithm.

3.1Proposed Method

We are using machine learning classifier random forest. Random forest gives best accuracy comparatively other classifier and also more efficient compare to other classifier. Also use an artificial neural network for detecting fake accounts. Using neural network we can improve efficiency and accuracy. So, we are using RF and DNN model for accurate result. We are using some different features like following list, common friends, family member, message frequency, and Language type.

Random forest is kind of ensemble classifier which is using decision tree algorithm in random fashion. First of all creates a fictitious accounts and then inject the data and for a detection of fake account we are using machine learning algorithm. We are using the machine learning algorithms for improve an efficiency and accuracy. First we select the profile which would be tested from extracting our database. We are creating a database where we are tested on that profile. Then extracting features what we are needed to help an evolution of that there is a profile is fake or real and then we are using a classification algorithm and then we evaluate the result.

3.2 Data and Sources of Data

For this study we are downloading from public source. We are download dataset from github [7] where we can download dataset freely. In this data set there is so many colum like following count, status count, listed count, utc offset, Id, name, profile name. In this dataset there are 1338 fake accounts and 1482 real accounts.

3.3 Proposed Approach

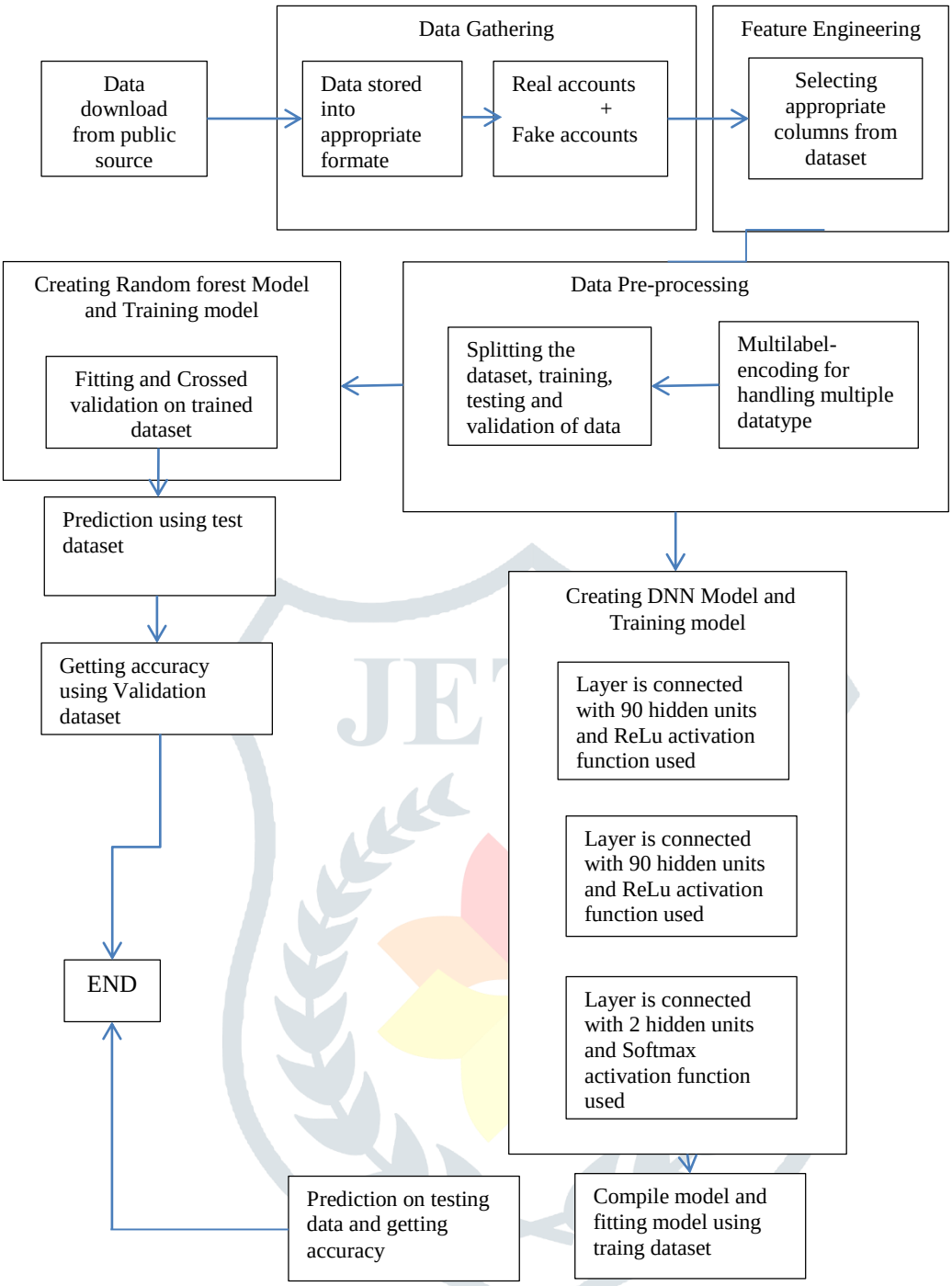

Fig.1 Process flow

The detailed flowchart of this model is given below:

Step 1:
First, in this data set there will be some data set are real and some data set are fake. And gathering the real and fake account dataset .we have to go through data pre-processing. Data pre-processing is necessary because suppose there is data which is unnecessary and also redundant data or garbage data then data pre-processing clean this type of data. And after this and also after extracting feature, this data pass through Classification algorithm.

Step 2:
After extracting database or dataset we should extract features and then applying on the database. There are different types of feature which have a different type of characteristics. In this paper we are using simplest features like who is following the account, what is the relation between that follower and the user account.

Step 3:
We need to select profiles or account which we are wanted to testing and training on that database. After this we are creating model of random forest and deep neural network. And after that fit the trained dataset in to RF and DNN model.

Step 4:

In this phase we are doing prediction using tested dataset. This phase is the evolutions phase that, where we properly extract data from the database? Where there is extracting features properly? So this is the evolution of overall of process flow and every phase work properly.

Step 5:
After step 4, this phase we are comparing with exciting approach and get more accuracy compare to exciting model.

3.4 Implementation

In this section we are showing about how we implemented and how we coding for detection of fake accounts.

3.4.1 Data Preprocessing:

We download data from the public source. In this dataset there is two different dataset fake accounts dataset and real accounts dataset. We download about 3770 accounts dataset with different categories i.e., id, name, followers count, status count, listed count. In shown below image this is about the exciting models and which they give us accuracy and which we can compare our model that which gives that we can get more accuracy compare to exciting.

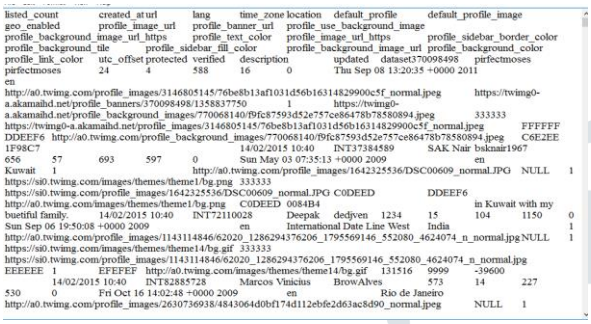

Fig.2 Fake accounts dataset

Fig.3. Real accounts dataset

We are trained our dataset so first we need to encoding our dataset for appropriate form so we can do easily trained our model and model can understood that what is our dataset. So there is multiple data types are there. And we need to multiple dataset encoding at a same time. So we are using multiple labels encoding for our multiple dataset. After that we are doing training our dataset using sklearn library. We train and test using sklearn library. After that we are doing validation using sklearn library.

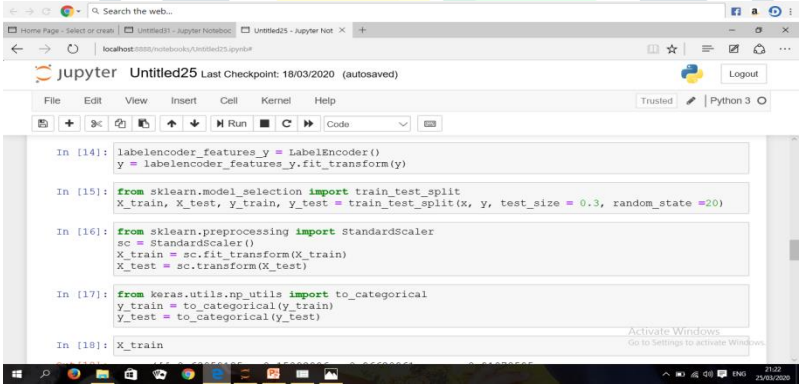

Fig.4. Label Encoding and training data

3.4.2 Training Random Forest Model:

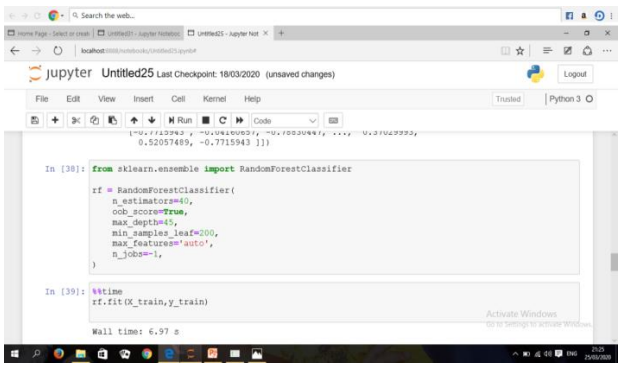

Fig.5 Proposed Models: Random Forest

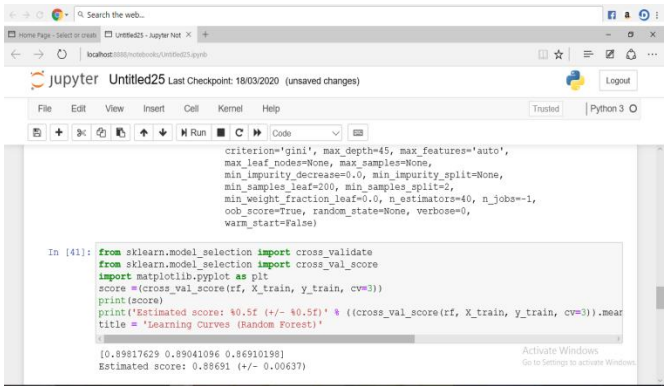

Fig.6 Cross validation trained dataset in Random Forest

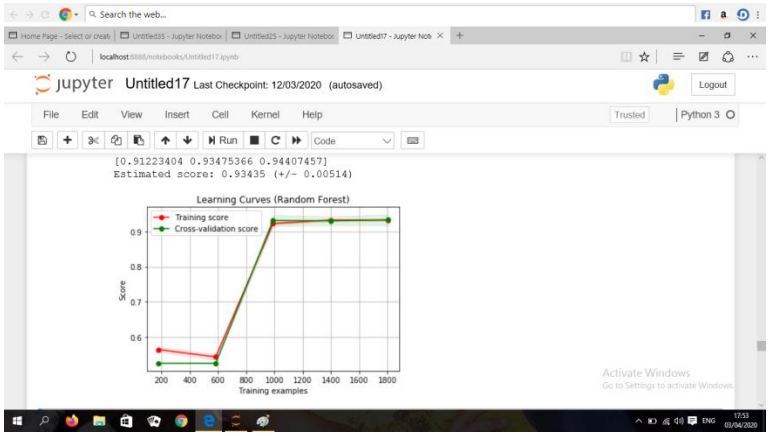

Fig.7 Graph of Cross validation trained dataset in Random Forest

3.4.3 Testing Random Forest Model:

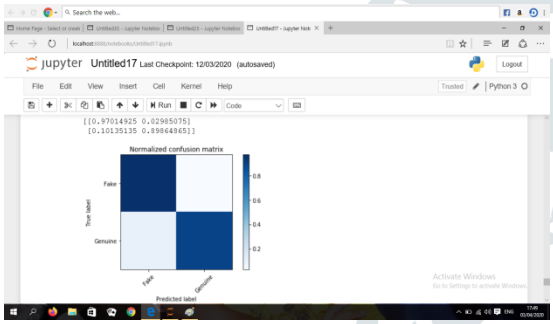

Fig.8 Normalized confusion matrix

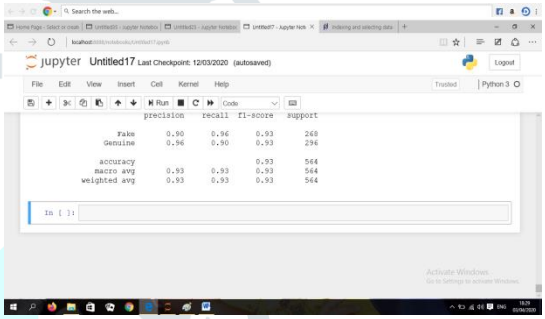

Fig.9 Classification report for RF model

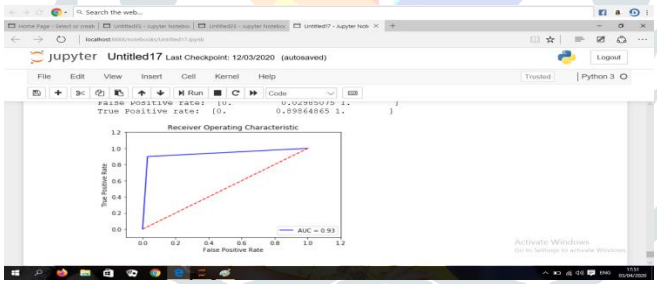

Fig.10 Ratio between True positive rate and False positive rate

3.4.4 Training DNN Model

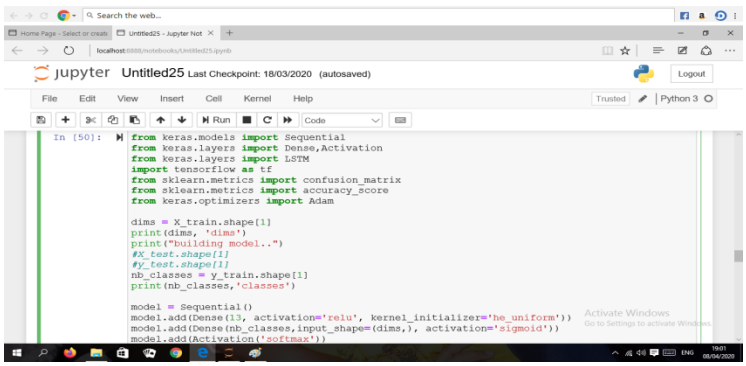

Fig.11 Training DNN model

3.4.5 Testing DNN Model

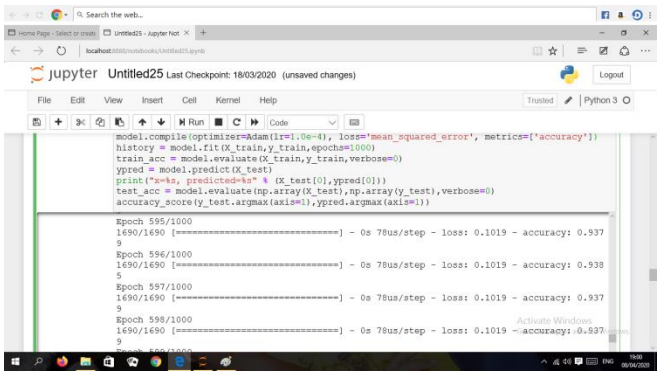

Fig.12 Fitting, Testing and predicting DNN model

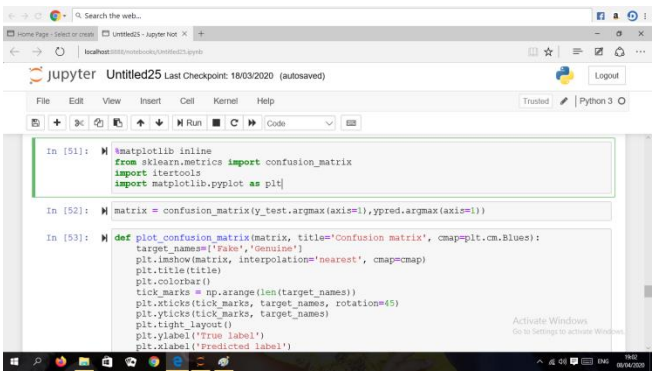

Fig.13 Creating confusion matrix for DNN model

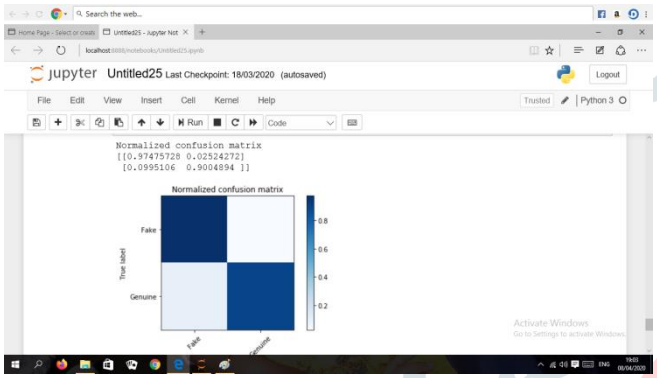

Fig.14 Confusion report for DNN model

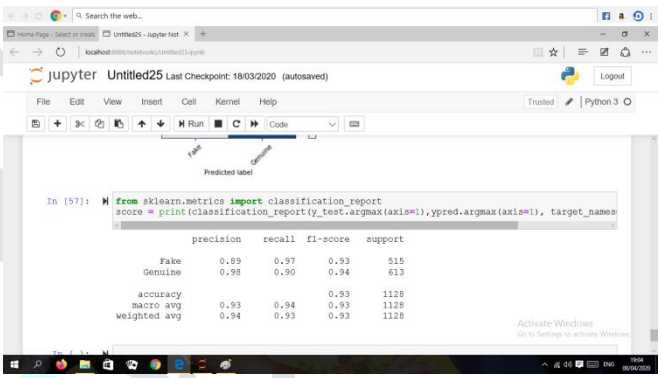

Fig.15 Classification report for DNN model

IV. RESULTS AND DISCUSSION

4.1 Results of detection for fake accounts

Table 4.1: Accuracy Table

Model	Accuracy
RF	0.93
DNN	0.94

In our proposed approach we got 94% and 93% accuracy and low percentage of error value. When we are comparing exciting approach there is we got 88%, 90% and 87% accuracy and comparatively we can get more accuracy. And we can get more accuracy so that automatically increase e.

V.CONCLUSION

So using a Random forest and deep neural network we are getting more accuracy and also efficiency. And we are reducing the malicious activity and detecting the fake accounts on social media. Our main aim is to serve better composition Spam account detection. By using Machine Learning technique the proposed system give definitely help in better composition of fake account detection. The proposed method is take advantage of deep neural network for better composition..

REFERENCES

[1]Zulfikar Alom, DiSTA, Barbara Carminati DiSTA,Elena Ferrari DiSTA; Ceyhum Akyol, “Detecting spam accounts on Twitter,” IEEE International Conference,2017

[2]Gayathri A , Radhika S , Mrs. Jayalakshmi S.L.; “Detecting Fake Accounts in Media Application Using Machine Learning ,“ International Journal of Advanced Networking & Applications, 2018

[3]Buket Ershin; Ozlem Aktas; Deniz Kilinc; Ceyhum Akyol; “Twitter Fake Account Detection,” IEEE International Conference,2017

[4]Mahdi Washhaa,∗, Aziz Qaroushb, Manel Mezghania, Florence Sedesa; “A Topic-Based Hidden Markov Model for Real-Time Spam Tweets Filtering,”; Elsevier, 2017

[5]Patxi Gal´an-Garc´ıa, Jos´e Gaviria de la Puerta, Carlos Laorden G´omez, Igor Santos, Pablo Garc´ıa Bringas, "Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying," Springer International Publishing ,2014

[6]ESTÉE VAN , DER WALT , JAN ELOFF "Using Machine Learning to Detect Fake Identities: Bots vs Humans," IEEE, 2016

[7]harshitkgupta; access on 11 Dec 2019, https://github.com/harshitkgupta/Fake-Profile-Detection-using-ML/tree/master/data