Identification of Fake vs. Real Identities on Social Media using Random Forest and Deep Convolutional Neural Network

[1]Priyanka Shahane, [2]Deipali Gore

[1] M.E. Scholar, Department of Computer Engineering, PES MCOE, Pune, India
[2]Assistant Professor, Department of Computer Engineering, PES MCOE, Pune, India.

Abstract : **Identity deception on different social media platforms is increasing rapidly with huge growth of these platforms. As these fake identities are being used by criminals for different malicious purposes, it has become necessity of time to identify them. The fake identities are categorized into two main types i.e. fake identities produced by bots and fake identities produced humans . This system removes fake identities produced by bots during preprocessing and focuses mainly on identification of fake identities produced by humans as very less research has been made till now on the fake identities produced by humans. For classification we test for two different algorithms i.e. Random Forest (RF) and Deep Convolutional Neural Network (DCNN). The classification is based on various features such as user name, location, friends count, followers count and so on. Here, dataset used is that of Twitter.**

Index Terms **- social media; identity deception; cyber crimes; machine learning; random forest; deep learning; deep convolutional neural network; activation functions.**

I. INTRODUCTION

Social media platform such as Twitter is one of the most crucial means of communication and information dissemination over internet. Much can be learned about people's behavior by analyzing their profiles on the social media. This helps offenders to create fake identities in order to commit various cyber crimes such as skewing perceptions, manipulation of credit worthiness of accounts, terrorist propaganda, cyber bullying, fraud, identity impersonation, dissemination of pornography, misdirecting people to some malicious website, spreading malwares and so on.

These fake identities are generally produced by bots or humans. The fake identities produced by bots usually target huge number of people at a time, whereas, fake identities produced by humans generally target particular individual or small group of people. This system represents an approach to detect fake identities produced by humans on Twitter.

In order to classify identities as fake or real we test for two different machine learning algorithms i.e. Random Forest (RF) & Deep Convolutional Neural Network (DCNN). Furthermore, DCNN is implemented using linear, sigmoid and tan h activation functions. Here, both the algorithms are trained using different cross validation techniques such as 5 fold, 10 fold and 15 fold cross validation.

Finally, the system is evaluated based on various performance measures such as accuracy, f1 score, precision and recall in order to predict which activation function as well as cross validation technique gives better classification.

II. LITERATURE SURVEY

In machine learning, classification is based on learning from training database. This learning can be classified into three types as: supervised, semi-supervised and unsupervised. In supervised method of learning class labeled data is present in the beginning. Whereas, in unsupervised learning class labeled data is not available in the beginning. Semi-supervised method of learning is integration of both supervised and unsupervised learning where some of the labels of class are known.

The problem of identification of fake identities can be solved by different classification techniques such as Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Multi Layer Perceptron (MLP), Naïve Bayes (NB), K Nearest Neighbour (KNN), Artificial Neural Network (ANN), Adaboost, Gradient Boosting and so on. Here are some examples,

Estee et. al. [1] trained the classifier using features that were previously implemented for bots identification in order to detect fake identities produced by humans on Twitter. Here, the classifier is trained using supervised learning method. They have used three different classifiers i.e. SVM with linear kernel, Adaboost and RF. SVM is is implemented using R software with svmLinear library. Here, the classification boundary is based on feature vectors. Boosting model is implemented using R software with Adaboost function. It is used with decision trees and different weight is assigned for each feature in order to predict outcome. These weights are modified iteratively in order to analyze effectiveness of classification for each iteration and the process is repeted until best result is achieved. RF model is implemented using R software with RF library. This model creates number of trees and highly probable output is used to detect identity deception. Among these 3 classifiers RF gave the best result.

Sen et. al. [2] used supervised learning method for training classifier based on features extracted from RandLike_data and FakeLike_data. They have experimented with various classification methods such as XGBoost, AdaBoost with RF as a base initiator, SVM with RBF kernel, RF, LR and simple neural network with feed forward architecture i.e. MLP to detect fake likes created on instagram. MLP with two hidden layers and 200 neurons per layer is implemented here. Activation function used for both the layers is sigmoid and dropout of output layer is kept 0.2 in order to prohibit over fitting. Here, MLP gave better result compared to other classification techniques.

Sedhai et. al. [3] used semi-supervised learning method in order to train his classifiers i.e LR, NB and RF. The classification techniques used by these three classifiers are discriminative, generative and decision tree based model respectively. Here, Twitter dataset is used. Twitter Id is considered as spam only if at least 2 of these 3 classifiers detect it as a spam. This framework is called as S^3D (Semi- Supervised Spam Detection) and it has reached best classification result with respect to any single classifier.

Xiao et. al. [4] used supervised learning to extract best features from LinkedIn data. They have implemented three different classifiers i.e. LR with L1 regularization, SVM with radial kernel and RF a nonlinear decision tree based algorithm having ensemble learning approach. LR identifies parameters with the help of maximum likelihood estimation. In paper L1 penalization is used to regularize LR model. This method maximizes probability distribution of class y using known feature vector x and minimizes count of irrelevant features with the help of penalty term to bound coefficients of L1 norm. SVM looks for optimal hyperplane as a decision function in high dimensional space. Whereas, RF combines number of weak classifiers (decision trees) in order to generate strong classifier. Here, RF gave best result for identification of fake profiles.

Ikram et. al. [5] used supervised binary SVM classifier implemented by scikit learn (an open source machine learning library for python) for classification of like farm users and normal (baseline) users. SVM is compared with other well known classifiers based on supervised method of learning such as AdaBoost, Decision tree, KNN and RF. Here, two class SVM gave best result for identification of like farms on Facebook dataset.

Dickerson et. al. [6] performed training on Indian Election Dataset which was extracted from Twitter. They have implemented 6 different classifiers such as Extremely Randomized Trees, RF, Gradient Boosting, AdaBoost, Gaussian Naïve Bayes and SVM. The scikit-learn, a toolkit used for machine learning which is supported by Google and INRIA is used to build classifiers. Here, AdaBoost gave best outcome on reduced features set where only the features that do not involved sentiment analysis were considered. Whereas, Gradient Boosting gave best outcome on full feature set.

Fuller et. al. [7] used dataset taken from law enforcement personal at military bases which is also known as "person of interest statements" or Form 1168. Person of interest statements are reports written by a witness or subject in an official investigation. Three common classification methods that they have tested are, ANN, LR and Decision Tree. Among all these methods ANN gave the best performance. ANN is a collection of nodes arranged in layers. It has three main layers: input layer, hidden layer and output layer. The nodes in hidden layer combine inputs from previous layers into a single output value. This output is then passed on to next layer. The weight is associated with each unit in the network, it is determined by training a network on portion of data. Then network performance is evaluated on holdout sample.

Peddinti et. al. [8] designed a classifier which converts 4 class classification task into binary classification task such that one classifier classifies each identity into two categories i.e. anonymous and non anonymous, while, other classifier classifies each identity as non identifiable or identifiable. Then results of these two classifiers are combined in order to classify each identity as 'anonymous', 'identifiable' or 'unknown' for Twitter data. Both the binary classifiers use RF with 100 trees as base classifier. The choice of classifier and number of trees is based on cross validation performance and out of bag error. These classifiers are meta classifiers sensitive to cost, where misclassifying identities as anonymous or identifiable imposes higher cost.

Oentaryo et. al. [9] used unsupervised and supervised methods of learning and implemented 4 prominent classifiers: NB, RF, SVM and LR. The dataset used is generated by Twitter users from Singapore in period of 1 January 2014 to 30 April 2014 and it is extracted via Twitter REST and streaming API. Here, LR performed best for classification of identities as Consumption bots, Broadcast bots, Human and Spam bot.

Vishwanath et. al. [10] used unsupervised method of learning for Facebook dataset. The classification is performed using KNN algorithm. In KNN data is classified based on majority voting of its neighbors, with test data being assigned to a most common class among its K-nearest neighbors where K is a small positive integer. Here, Facebook ID's are classified into four classes i.e. Black market, Colluding, Compromised and Unclassified.

From this literature survey we found that Random Forest and Neural Networks are giving best results for identification of fake profiles on social media. Thus, we test for these two classification techniques in our system.

III. SYSTEM ARCHITECTURE

The flow of our system is as follows:

1. Data Acquisition:

First of all, data is extracted from Twitter using Twitter API based on keywords such as "school" and "homework" as these are the keywords that are mostly used by minors and minors are more susceptible to cyber crimes. Here we have extracted about 3000 accounts from Twitter.

2. Preprocessing:

The various preprocessing steps that we have applied are,

A. Lexical analysis:

Lexical analysis separates the input alphabet into,

a) Word characters: For e.g., letters a-z and

b) Word separators: For e.g., space, newline, tab

B. Stopword removal:

Stopword removal refers to the removal of words that occur most frequently in the documents. The stopwords includes,

a) Articles (a, an, the,…)

b) Prepositions (in, on, of,…)

c) Conjunctions (and, or, but, if,…)

d) Pronouns (I, you, them, it,…)

e) Possibly some verbs, nouns, adverbs, adjectives (make, thing, similar…)

C. Stemming:

Stemming replaces all the variants of a word with a single stem word. Variants include plurals, gerund forms (ing forms), third person suffixes, past tense suffixes, etc. Here we used the Porter's algorithm for stemming.

D. Index term selection:

Index term selection refers to the selection of appropriate features from large amount of data that contribute most to our prediction variable or output.

E. Data cleaning:

During data cleaning step bots are removed from the dataset based on certain parameters such as presence of name, profile image, number of followers, number of tweets, use of punctuation etc. Also, accounts of known celebrities are removed from the given corpus.

3. Create fictitious accounts:

Then fictitious accounts are created with the help of various random human data generator APIs and manually by us. The number of fictitious accounts created by us is around 4000. The basis for creation of fictitious accounts is that the people generally lie on their age, gender, image, location and the name most. For example, if location given is that of Arctic ocean or some volcano where human being cannot survive then it can be considered as fake.

4. Validate data:

For validity of research, it is decided to make sure that the fabricated fictitious accounts are as much as possible aligned with the accounts extracted using Twitter API in data acquisition step (original corpus). This was done to make our research results as realistic as possible. For that we have implemented following two statistical tests,

A. Mann-Whitney U test:

This test proofs that the means of the two sets are similar per attribute.

B. Chi square test:

This test proofs that the datasets are not correlated and therefore independent.

This means that both the deceptive and original corpus must have similar data and show same distributions.

5. Inject fictitious accounts:

Fictitious accounts that pass Mann Whitney U test and Chi square test are injected into the system. Thus, now our corpus will consists of fake and real accounts by humans extracted via Twitter API as well as the fake accounts that we have created manually and the total number of accounts becomes about 7000.

6. Create new features:

Here some new features are created using features that we have extracted in preprocessing step which made identification of fake identities much easier. For example, ratio of tweets containing URL to the total number of tweets is higher for fake identities as the URLs are used by offenders to misdirect people to malicious websites.

7. Classification:

We have tested for two different algorithms i.e. Random Forest and Deep Convolutional Neural Network (Linear, Sigmoid and Tan h activation function) for classification of Fakes vs. Real identities. Both the algorithms are trained using supervised learning method. Here we have experimented with three different cross validation techniques i.e. 5 fold, 10 fold and15 fold where around 70 percent data is given for training and remaining 30 percent data goes for testing.

A. Random Forest:

a) Algorithm:

1) Randomly select k features from total m features, where k is less than m in order to construct n decision trees.
2) Take the test vector and use each randomly created decision tree to predict the outcome and then store predicted outcome.
3) Calculate the votes for each predicted outcome.
4) Consider the highly voted predicted outcome as the final prediction of random forest algorithm.

b) Block diagram:

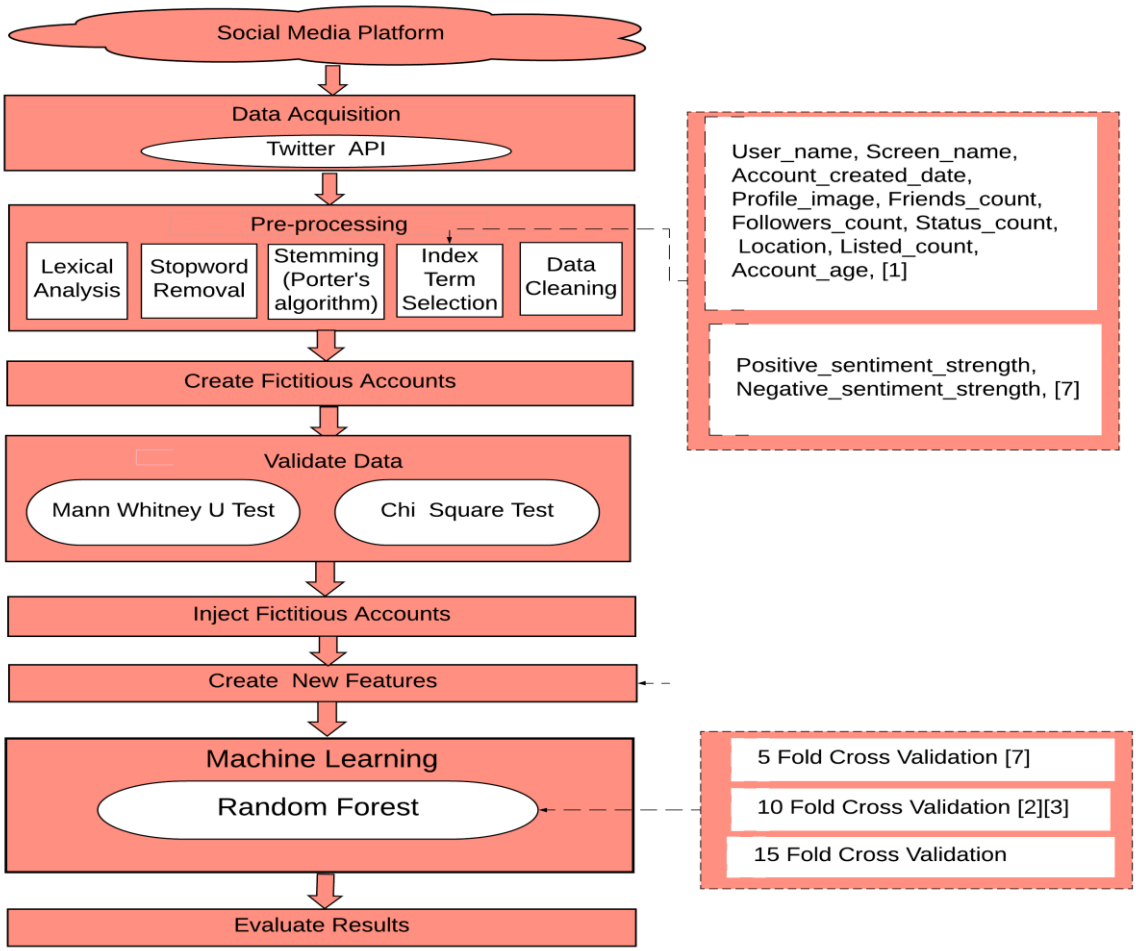

Figure 1.Identification of Fake vs. Real Identities on Twitter using Random Forest algorithm.

c) Activation function:

$W = \sum_{i=0}^{n}(inp[i]) = (hid[i])$ (1)

W > T: 1;

W < T: 0

Where,

W is a weight assigned based on equality of input and hidden identities.

Here, input (inp) corresponds to the identities whose class label is to be detected and hidden (hid) identities corresponds to the training data whose class label is known.

T is a threshold kept on calculated weight to detect fake identities.

B. Deep Convolutional Neural Network:

a) Algorithm:

1) First of all we inject number of Twitter accounts that we have extracted via Twitter API to the system for classification purpose. Now Input layer consists of all samples like I = (input sample 1, input sample 2,….., input sample n).

2) Now first convolution layer is dependent on training database which can generate the output samples based on current classification weight which will be given as a input to next layer.

3) Then second convolution layer is dependent on background knowledge i.e. classification rules. The output samples of this layer are then provided to output layer where different activation functions can be applied on it for classification purpose.

4) Finally output layer gives the final output labeled in the form O = (Fake accounts, Real accounts). During whole process it follows Feed Forward architecture.

b) Block diagram:

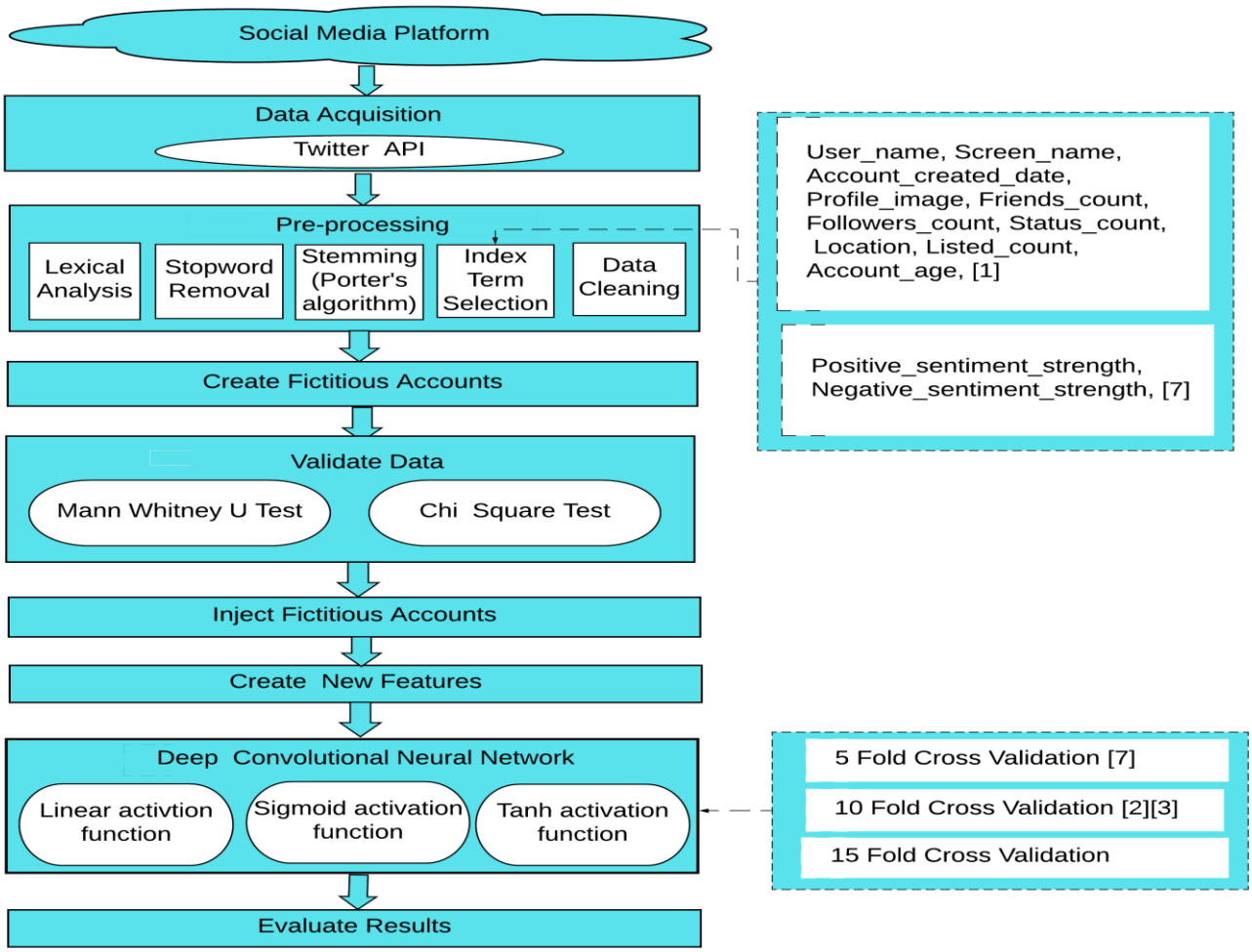

Figure 2.Identification of Fake vs. Real Identities on Twitter using Deep Convolutional Neural Network.

c) Activation functions:
For DCNN we have tested for three different action functions i.e. Linear, Sigmoid and Tan h.
1) Linear activation function:

$y = a + v$ (2)

Where,
$v = \sum w_i x_i$
x_i is a set of features.
w_i are weights associated with features.
a is bias.

2) Sigmoid activation function:

$y = 1/(1 + e^{-v})$ (3)

Where,
$v = \sum w_i x_i$
x_i is a set of features.
w_i are weights associated with features.

3) Tan h activation function:

$y = \tanh(x) = 2/(1 + e^{-2v}) - 1$ (4)

Where,
$v = \sum w_i x_i$
x_i is a set of features.
w_i are weights associated with features

8. Evaluate results:

Results are evaluated based on various performance measures such as accuracy, precision, recall and f1 score.

- Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$ (5)
- Precision = $\frac{TP}{TP+FP}$ (6)
- Recall = $\frac{TP}{TP+FN}$ (7)
- F1 score = 2 * (Precision * Recall)/(Precision + Recall) (8)

Where,
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

IV. RESULTS AND ANALYSIS

In order to evaluate the performance of a system we have analyzed accuracy, f1 score, recall and precision with which fake identities on social media can be detected,

A. Using different cross validation techniques:

a) 5-fold cross validation

b) 10-fold cross validation

c) 15-fold cross validation

B. Using different activation functions for DCNN:

a) Linear activation function

b) Sigmoid activation function

c) Tan h activation function

C. Using different index terms:

a) Positive sentiment strength

b) Negative sentiment strength

Table 1: Results for classification of fake vs. real identities using RF with 5 fold, 10 fold and 15 fold cross validation.

RF	5-Fold	10-Fold	15-Fold
Accuracy	85.40	86.10	89.40
Precision	84.40	86.50	89.40
Recall	85.70	86.65	89.70
F1-Score	86.50	86.90	89.50

Table 2: Results for classification of fake vs. real identities using DCNN (Linear Activation Function) with 5 fold, 10 fold and 15 fold cross validation.

DCNN (Linear)	5-Fold	10-Fold	15-Fold
Accuracy	93.10	93.40	95.00
Precision	92.40	93.15	94.10
Recall	92.70	93.20	94.30
F1-Score	92.50	93.60	94.70

Table 3: Results for classification of fake vs. real identities using DCNN (Sigmoid Activation Function) with 5 fold, 10 fold and 15 fold cross validation.

DCNN (Sigmoid)	5-Fold	10-Fold	15-Fold
Accuracy	93.60	94.80	95.10
Precision	92.20	93.00	94.10
Recall	92.00	93.25	94.35
F1-Score	92.60	93.80	94.90

Table 4: Results for classification of fake vs. real identities using DCNN (Tan h Activation Function) with 5 fold, 10 fold and 15 fold cross validation

DCNN (Tan h)	5-Fold	10-Fold	15-Fold
Accuracy	92.40	93.55	94.90
Precision	91.50	92.60	93.90
Recall	91.80	93.10	94.20
F1-Score	92.10	93.05	94.70

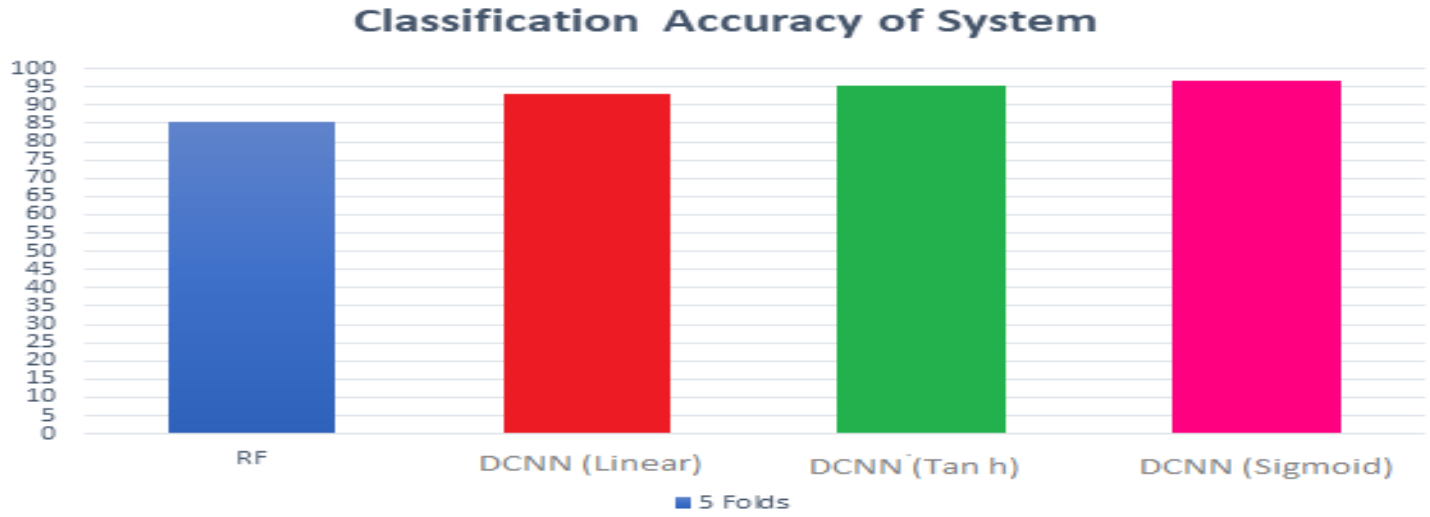

Figure 3. Comparative analysis for accuracy of RF, DCNN (Linear), DCNN (Sigmoid) and DCNN (Tan h) using 5 Fold cross validation.

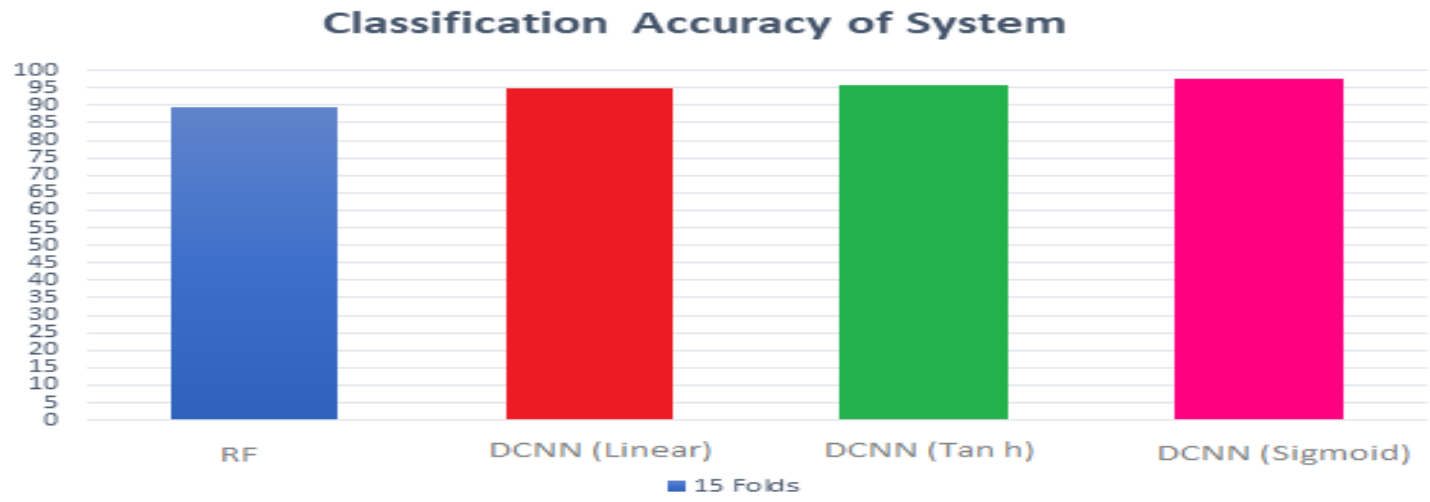

Figure 4. Comparative analysis for accuracy of RF, DCNN (Linear), DCNN (Sigmoid) and DCNN (Tan h) using 15 Fold cross validation.

V. CONCLUSION

- The maximum accuracy with which problem of classification of fake vs. real identities on social media can be solved is 95.10 % and it is achieved by DCNN with Sigmoid activation function.

- Using Sigmoid activation function (DCNN), gave an increase of 5.7 % in accuracy as compared to RF.
- Also, Sigmoid activation function (DCNN) gave an improved accuracy of 0.10 % and 0.20 % respectively as compared to Linear activation function and Tan h activation function.
- The performance of given system varies with dataset used for it.
- Also we found that the classification accuracy of system increases as the number of folds used in system increases.

REFERENCES

[1] Estee Van Der Walt and Jan Eloff, "Using Machine Learning to Detect Fake Identities: Bots vs Humans," IEEE, 2018
[2] Indira Sen et. al. "Worth its Weight in Likes: Towards Detecting Fake Likes on Instagram," ACM,2018.
[3] B. Viswanath et. al. "Towards Detecting Anomolous User Behaviour in Online Social Networks," USENIX, 2014.
[4] Surendra Sedhai and Aixin Sun, "Semi-Supervised Spam Detection in Twitter Stream," IEEE, 2018.
[5] Cao Xiao, David Freeman and Theodore Hwa, "Detecting Clusters of Fake Accounts in Online Social Networks," ACM, 2015.
[6] Ikram et. al., "Combating Fraud in Online Social Networks: Detecting Stealthy Facebook Like Farms," ARXIV, 2016
[7] J. Dickerson,V. Kagan and V. Subhramanian,"Using Sentiment to Detect Bots on Twitter: Are Humans more Opinionated than Bots?," IEEE,2014.
[8] C. Fuller, D. Biros and R. Wilson "Decision Support for Determining Veracity via Linguistic based Cues," ELSEVIER, 2009.
[9] S. Peddinti, K. Ross and J. Cappos "Mining Anonymity: Identifying Sensitive Accounts on Twitter," ARXIV, 2016.
[10] R. Oentaryo et. al. "On Profiling Bots in Social Media," ARXIV, 2016.
[11] Priyanka Shahane, Deipali Gore "A Survey on Classification Techniques to Determine Fake vs. Real Identities on Social Media Platforms,"IJRDT, 2018.
[12] Priyanka Shahane, Deipali Gore, "Detection of Fake Profiles on Twitter using Random Forest & Deep Convolutional Neural Network," IJMTE, 2019.
[13] Priyanka Shahane, Sneha Kasbe, Rohini Kasar, M.P. Navle "Ontology Based Information Retrieval System Using Multiple Queries For Academic Library,"IRJET, 2018.