# MINOR PROJECT

Conversational AI - Data Science

# Detecting Fake Profiles On Social Media

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

*Submitted By : Mrigankshi Gupta*

*Submitted To : Dr Sahil Sharma*

(MAY 2022)

Github Link : https://github.com/mriganskhi26/Detecting-Fake-Profiles-On-Social-Media

# ABSTRACT

In the present generation, on-Line social networks (OSNs) have become increasingly popular, people's social lives have become more associated with these sites. They use on-Line social networks (OSNs) to keep in touch with each other, share news, organize events, and even run their own e-business. The rapid growth of OSNs and the massive amount of personal data of its subscribers have attracted attackers, and imposters to steal personal data, share false news, and spread malicious activities. On the other hand, researchers have started to investigate efficient techniques to detect abnormal activities and fake accounts relying on accounts features, and classification algorithms. However, some of the account's exploited features have negative contribution in the final results or have no impact, also using standalone classification algorithms does not always achieve satisfactory results. The approaches to detect fake social media accounts can be classified into the approaches aimed on analysing individual accounts, and the approaches capturing the coordinated activities spanning a large group of accounts. The project sheds light on the role of fake identities in advanced persistent threats and covers the mentioned approaches of detecting fake social media accounts. In this project, a new algorithm, SVM-NN, is proposed to provide efficient detection for fake Twitter accounts and bots, data preprocessing, feature selection and dimension reduction techniques were applied. Machine learning classification algorithms were used to decide the target accounts identity real or fake, those algorithms were support vector machine (SVM), neural Network (NN), Ada Boost Classifier, Random Forest and Decision Tree .

# CONTENTS

# Introduction

Online Social Networks (OSNs), such as Facebook, Twitter and LinkedIn, have become increasingly popular over the last few years. People use OSNs to keep in touch with each other’s, share news, organize events, and even run their own e-business. Facebook community continues to grow with more than 2.2 billion monthly active users and 1.4 billion daily active users, with an increase of 11% on a year-over-year basis. Online Social Networks (OSNs) have also attracted the interest of researchers for mining and analyzing their massive amount of data, exploring and studying users behaviours as well as detecting their abnormal activities. Researchers have made a study to predict, analyze and explain customers loyalty towards a social media-based online brand community, by identifying the most effective cognitive features that predict their customers attitude. The implications of researchers attempt may helps an OSN operator detecting fake accounts efficiently and effectively, hence improve the experience of their users by preventing annoying spam messages and other abusive content. The open nature of OSNs and the massive amount of personal data for its subscribers have made them vulnerable to Sybil attacks . In 2012, Facebook noticed an abuse on their platform including publishing false news, hate speech, sensational and polarizing, and others. In general, attackers follow the concept of having OSNs user accounts are “keys to walled gardens” , so they deceive themselves off as somebody else, by using photos and profiles that are either snatched from a real person without his/her knowledge, or are generated artificially, to spread fake news, and steal personal information. These fake accounts are generally called imposters . To enhance their effectiveness, these malicious accounts are often armed with stealthy automated tweeting programs, to mimic real users, known as bots. OSNs are employing different detecting algorithms and mitigation approaches to address the growing threat of fake/malicious accounts . This phenomena raised the flag for the need of new techniques to detect such actions and avoid them . For the purpose to detect fake accounts on the social media platforms the dataset generated was pre-processed and fake accounts were determined by machine learning algorithms. The classification performances of the algorithms Random Forest, Decision Tree, Neural Network and Support Vector Machines are used for the detection of fake accounts. The accuracy rates of detecting fake accounts using the mentioned algorithms are compared and the algorithm with the best accuracy rate is noted.

# Related Work

Instead of analysing individual profiles and their connections, many researchers focus on characterizing malicious activities involving a coordinated use of numerous accounts – for instance, in the context of black markets of bots and fake accounts for online social networks.

Inspired by the importance of detecting fake accounts, researchers have recently started to investigate efficient fake accounts detection mechanisms. Most detection mechanisms attempt to predict and classify user accounts as real or fake by analysing user level activities or graph-level structures. This project presents some filtering algorithms (SVM, Random Forest, Decision Tree, Neural Networks) that rely on classification to decide whether the profile is genuine or fake. There are several approaches that help detecting fake accounts (eg: on Facebook, LinkedIn, Twitter, etc) that are described below.

### • Feature Based Detection

This approach relies on user-level activities and associated account details (user logs and profiles). Unique features are extracted from recent user activities (e.g. frequency of friend requests, fraction of accepted requests), then those features are applied to a classifier that has been trained offline using machine learning techniques. The authors were able to classify the data with 3% false positive rate and 1% false negative rate. The authors used ground-truth to train an SVM classifier in order to detect fake accounts. Using simple features, such as:

- frequency of friend requests
- fraction of accepted requests

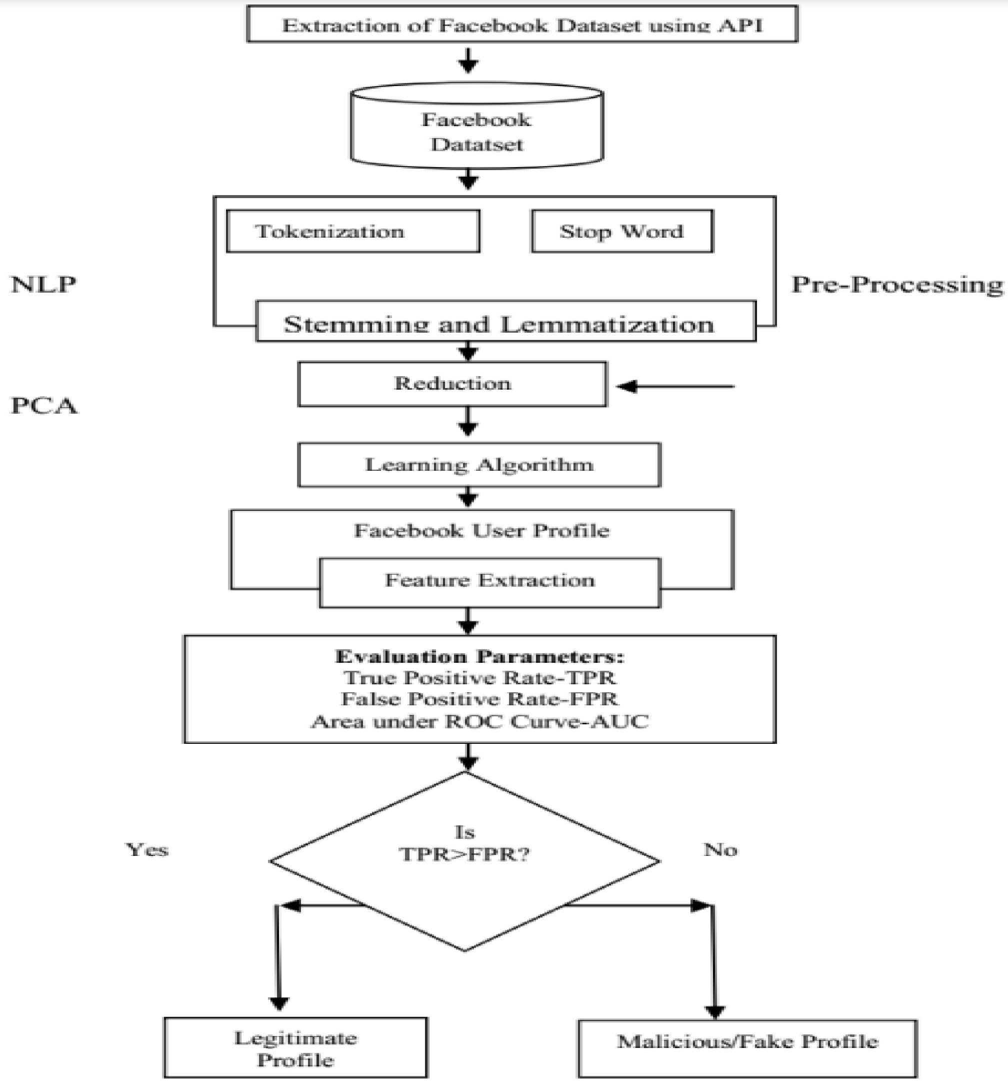
The authors were able to train a classifier with 99% true-positive rate (TPR) and 0.7% false-positive rate (FPR). They were also able to correctly classify more than 95% of the accounts of the original training set.

### • Feature Reduction

High dimensional data could be a serious problem for many classification algorithms because of its high computational cost and memory usage. Thus, reducing the dimension space would remove noisy (i.e. irrelevant) and redundant features and lead to a better classification model and simple visualization technique. Feature reduction techniques can be categorized into two types:

- Dimensionality reduction where the data in high dimensional space is transformed into a space of fewer dimensions.
- Feature subset selection where the original features set is disjoint into a selected subsets of features to build simpler and faster models, increases the models performance, and gain a better understanding of the data.

Among the most commonly used feature reduction techniques is the Principle Component Analysis (PCA). PCA is a technique used to identify features (dimensions) that best explain the predominant normal user behaviour. PCA projects high-dimensional data into a low-dimensional subspace (called the normal subspace) of the top-N principal components that accounts for as much variability in the data as possible.



The presented process used Facebook profile to notice false profiles. The working method of the proposed procedure includes three principal phases:

1. NLP Pre-processing
2. Principal Component Analysis (PCA)
3. Learning Algorithms

## NLP Pre-Processing

Text pre-processing is an essential part of any NLP method and the significance of the NLP pre-processing are:

a. To minimize indexing records dimension of the textual content records:-
i. Stop words bills 20-30% of total phrase counts in a special textual content records.
ii. Stemming may just diminish indexing size as much as 40% - 50% .

b. To make stronger the efficiency and effectiveness of the IR method:-
i. Stop words aren't valuable for shopping or textual content mining and so they may just confuse the retrieval system .
ii. Stemming used for matching the similar words in a text record.

## Principal Component Analysis (PCA)

Principal Component Analysis purpose is to extract the fundamental understanding from the table, to symbolize it as a suite of new orthogonal variables known as major accessories, and to show the sample of similarity of the observations and of the variables as elements in maps.

## Learning Algorithms

In this proposed system we are using two machine learning algorithms named as Support Vector Machine (SVM) , Neural Networks(NN) and Random Forest.

### *Neural Network and Support Vector Machine (SVM)*

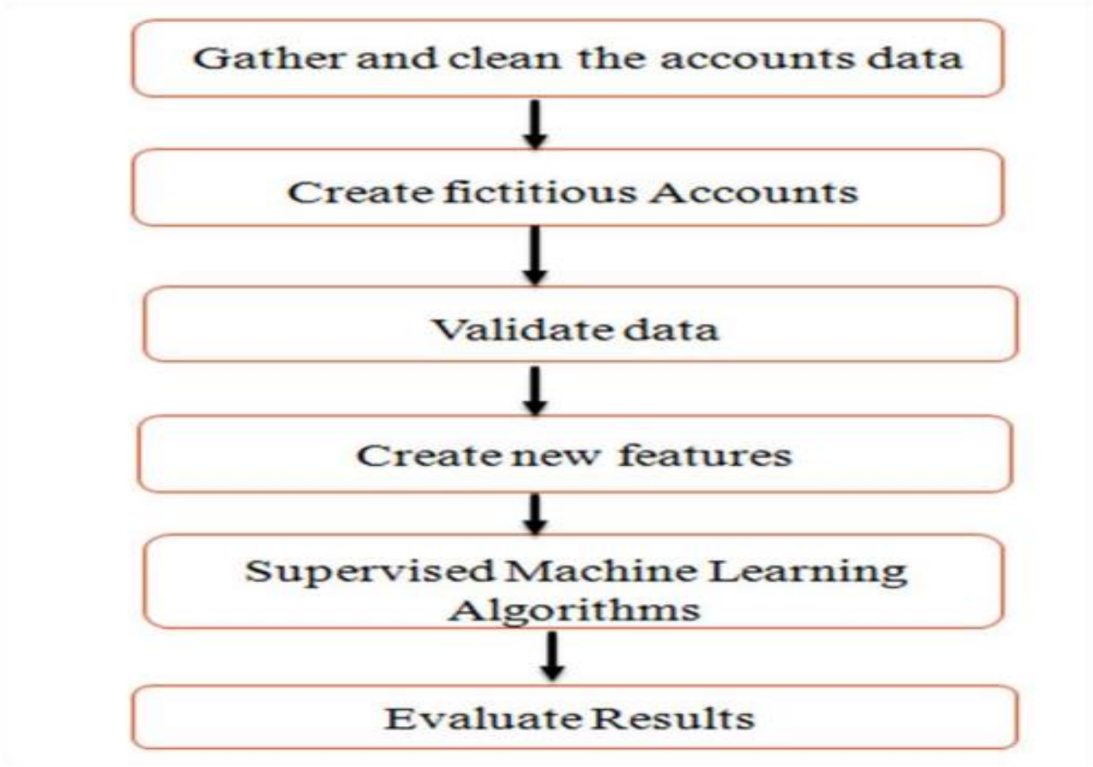
Researchers extracted the profile features using PCA, and then applied Neural Networks and Support Vector machines to detect legitimate profiles. ”Variance maximization” was selected as a mathematical way of deriving PCA results.
The result of this research showed that using PCA as a dimension reduction produced better accuracy results than using all the features without any selection. Even though feature-based detection scales to large OSNs, it could be circumvented. Attackers used to change content and activity patterns of their actions to avoid spam detection techniques . Feature-based detection does not provide any formal security guarantees and often results in a high false positive rate in practice .
Support Vector Machine (SVM) is a binary classification algorithm that finds the maximum separation hyper plane between two classes. It is a supervised learning algorithm that gives enough training examples, divides two classes fairly well and classifies new examples. It offers a principle approach to machine learning problems because of their mathematical foundation in statistical learning theory.

### *Random Forest*

Random Forest is versatile method performing both classification and regression tasks. It has nearly same hyperparameters as a decision tree or a bagging classifier. It creates many variations of trees .The best outcome will be used to predict identity deception .Each outcomes from the classifier represents different sections of a tree.

# Methodology

Proposed system is equipped with various Machine Learning tasks and the architecture followed is as shown below. The proposed system collects the dataset which are pre-processed by providing a framework of algorithms using which we can detect fake profiles in various social media platforms by comparing the accuracy of three machine learning algorithms and the algorithm with very high efficiency is found for the given dataset.


Gather and clean the accounts data
Create fictitious Accounts
Validate data
Create new features
Supervised Machine Learning Algorithms
Evaluate Results

**Proposed Methodology**

For the model preparation process, the numerous ways in which an algorithm might model a problem are dependent on its interaction with the experience or environment, which aids in selecting the most effective algorithm for the given input data in order to get the best outcome.

## Support Vector Machine (SVM):

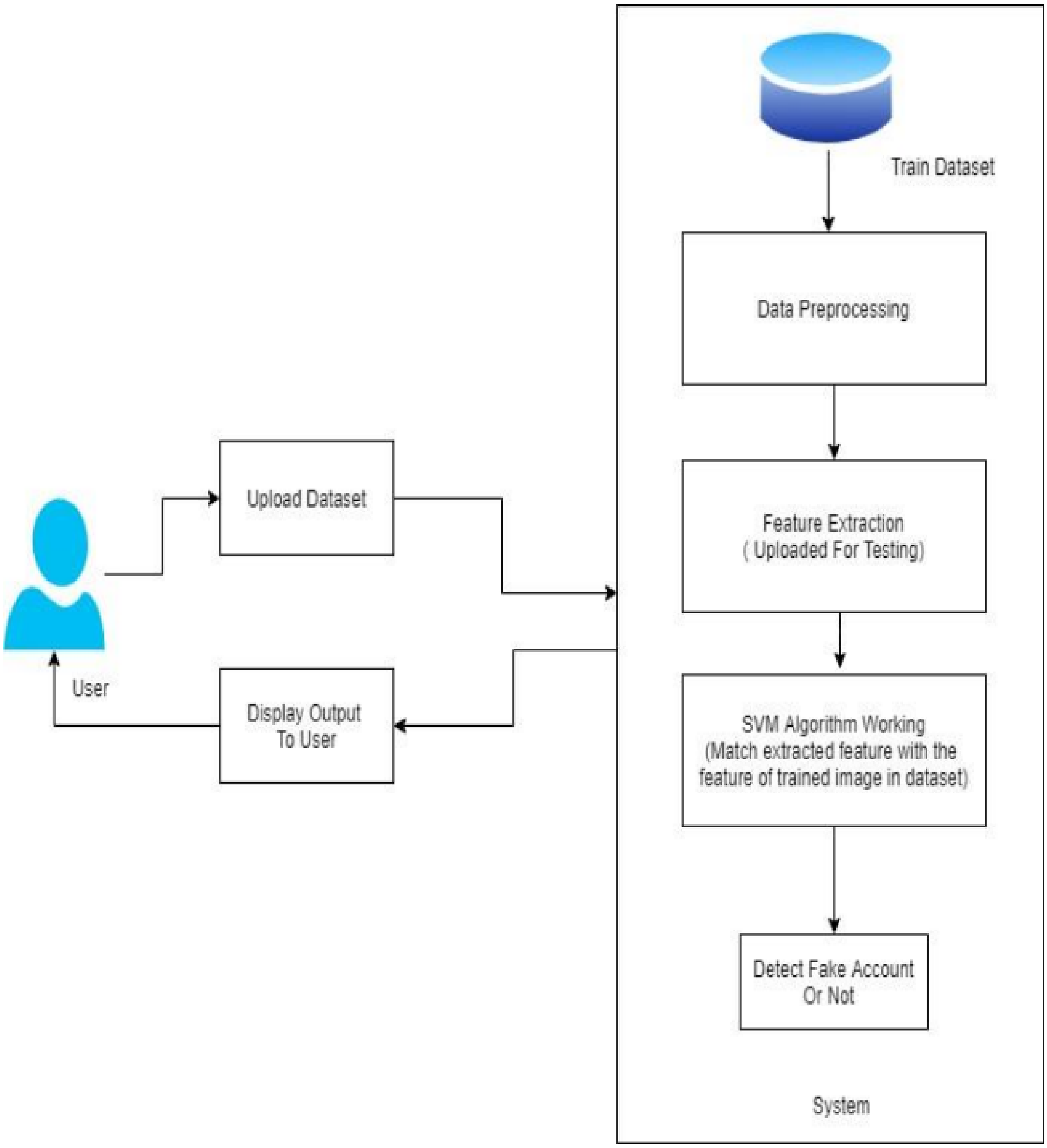
Support-vector machines (SVM) are the supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. For the given labeled training data (supervised learning), the algorithm outputs an optimal hyper plane which categorizes new examples. SVM constructs their solution as a weighted sum of SVs, which are only a subset of the training input. It is effective in cases where number of dimensions is greater than the number of samples given.

## Neural Networks:

A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes. A neural network (NN), in the case of artificial neurons is an interconnected group of natural or artificial neurons that uses a mathematical model for information processing based on connectionist approach.

## Random Forest:

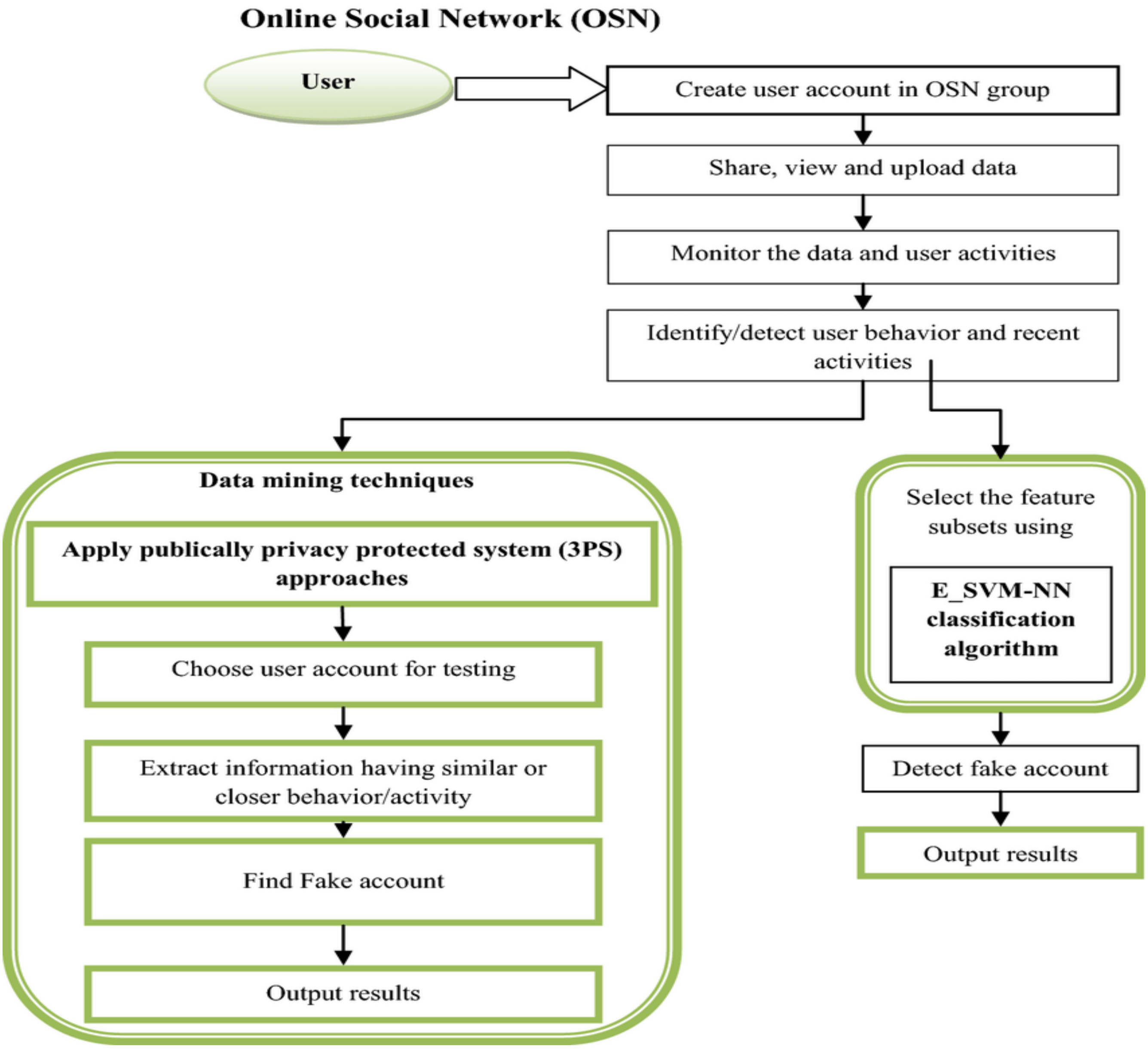
Random forest algorithm is a supervised classification algorithm. This algorithm creates the forest with a number of trees. In general, the more trees in the forest the more robust the forest looks like. In the same way in the random forest classifier, the higher the number of trees in the forest gives the high accuracy results.


Train Dataset
Data Preprocessing
Feature Extraction
( Uploaded For Testing)
SVM Algorithm Working
(Match extracted feature with the
feature of trained image in dataset)
Detect Fake Account
Or Not
System
Upload Dataset
User
Display Output
To User

**Flow Chart**

# Proposed Solution Architecture

Many social networking platforms are available for use although the basic functionality of all the online social media is same. There are many existing solutions for the fake account detection . Since the dataset collection is a tedious task ,very less work has been done to detect the fake accounts in it. So the work proposed in this project discusses results obtained by using various features to detect the fake accounts.



Online Social Network (OSN)
User
Create user account in OSN group
Share, view and upload data
Monitor the data and user activities
Identify/detect user behavior and recent activities
Data mining techniques
Apply publically privacy protected system (3PS) approaches
Choose user account for testing
Extract information having similar or closer behavior/activity
Find Fake account
Output results
Select the feature subsets using
E_SVM-NN classification algorithm
Detect fake account
Output results

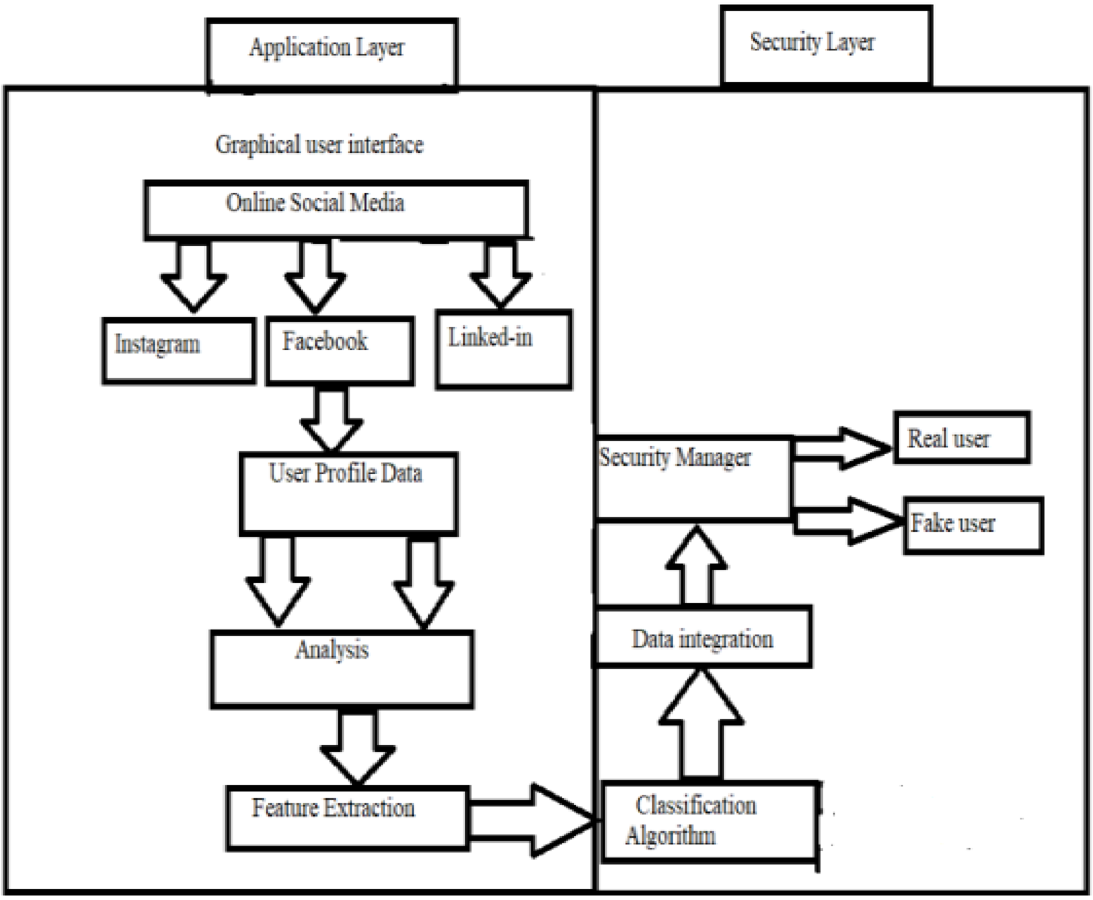
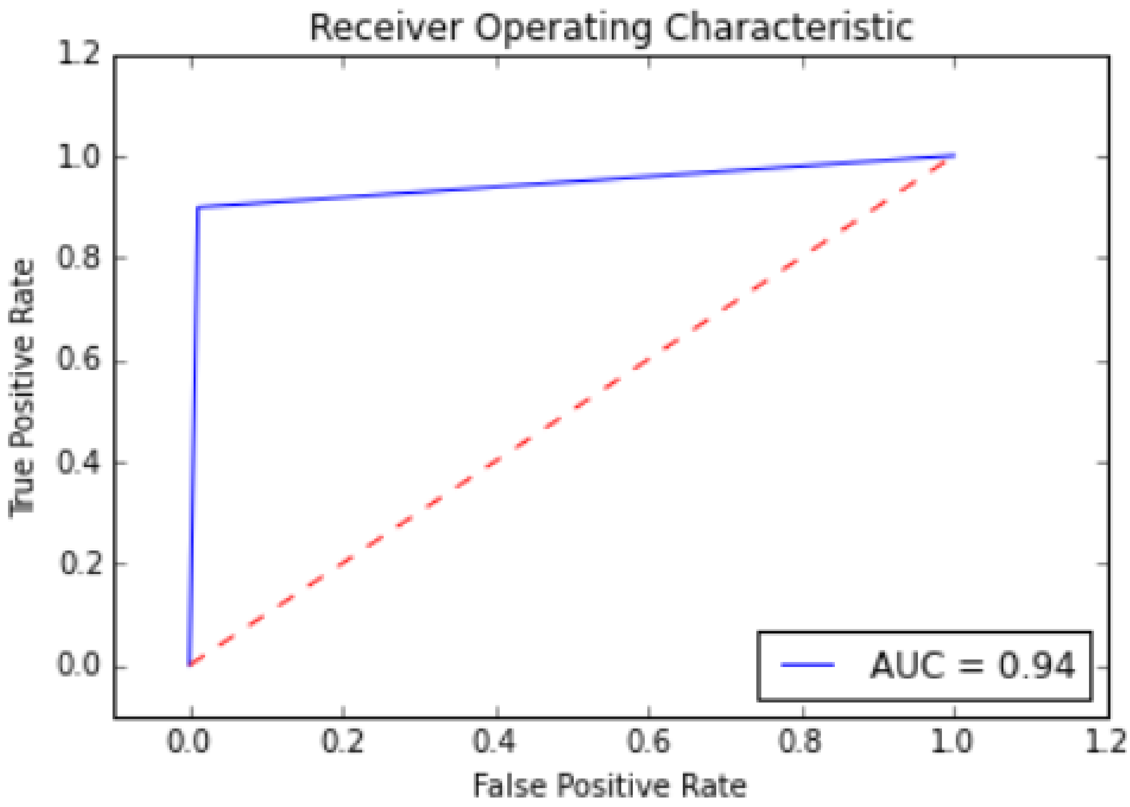


**Fig. 1: Architecture of Proposed System**

# Experimentation and Results

## a. Performance of model using Random Forest Algorithm:

The random forest is a model made up of many decision trees. When training the model using Random forest algorithm, each tree in a random forest learns from a random sample of the data points and the samples drawn with replacement are known as bootstrapping in which some samples will be used multiple times in a single tree.



b. Performance of model using Support Vector Machine Algorithm:

In many supervised learning tasks, labelling instances to create a training set is time consuming and costly thus, finding ways to minimize the number of labelled instances is beneficial. The Support Vector Machine algorithm is used to minimize the instances by improving efficiency. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. We then perform the detection of fake accounts through classification technique by finding the hyper-plane that differentiate the two classes very well.

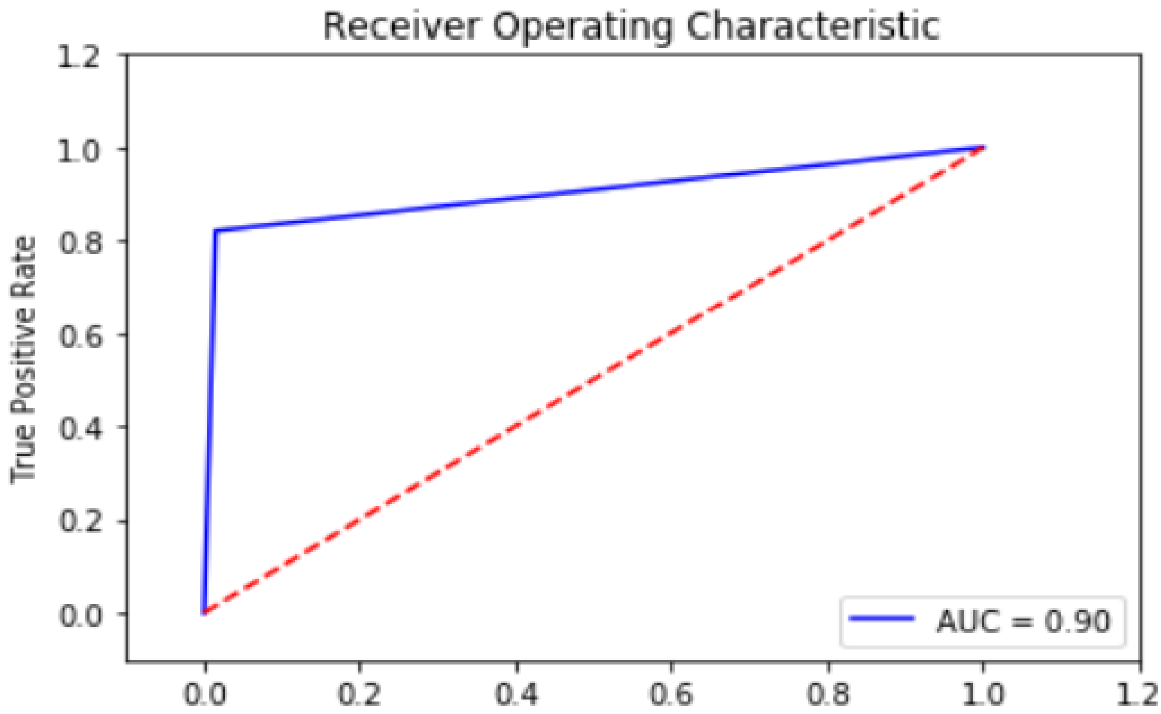
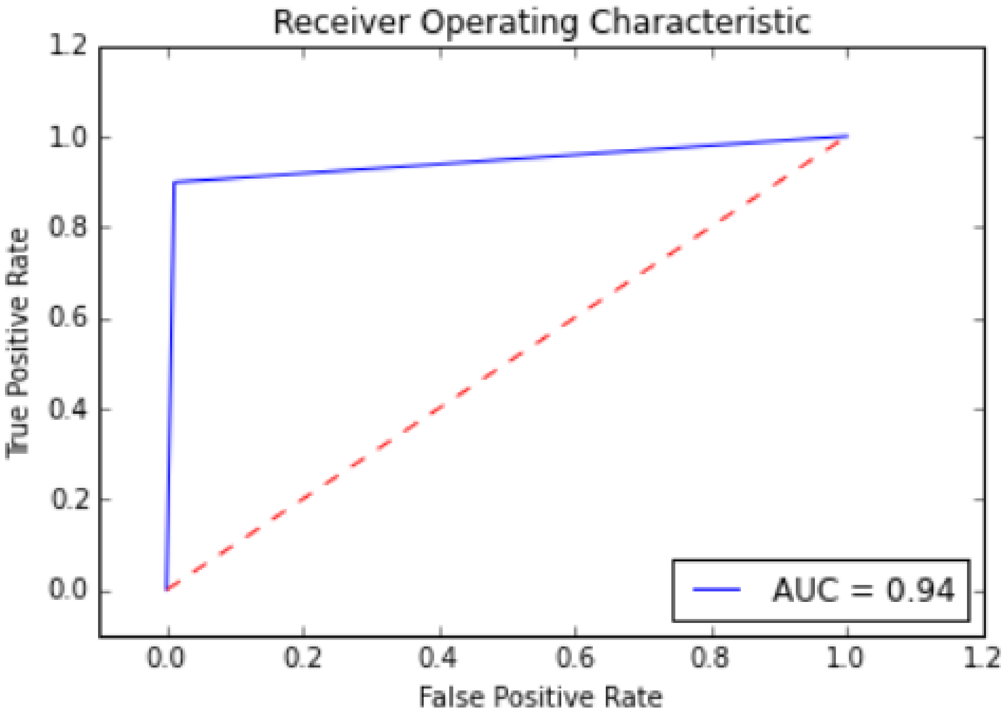
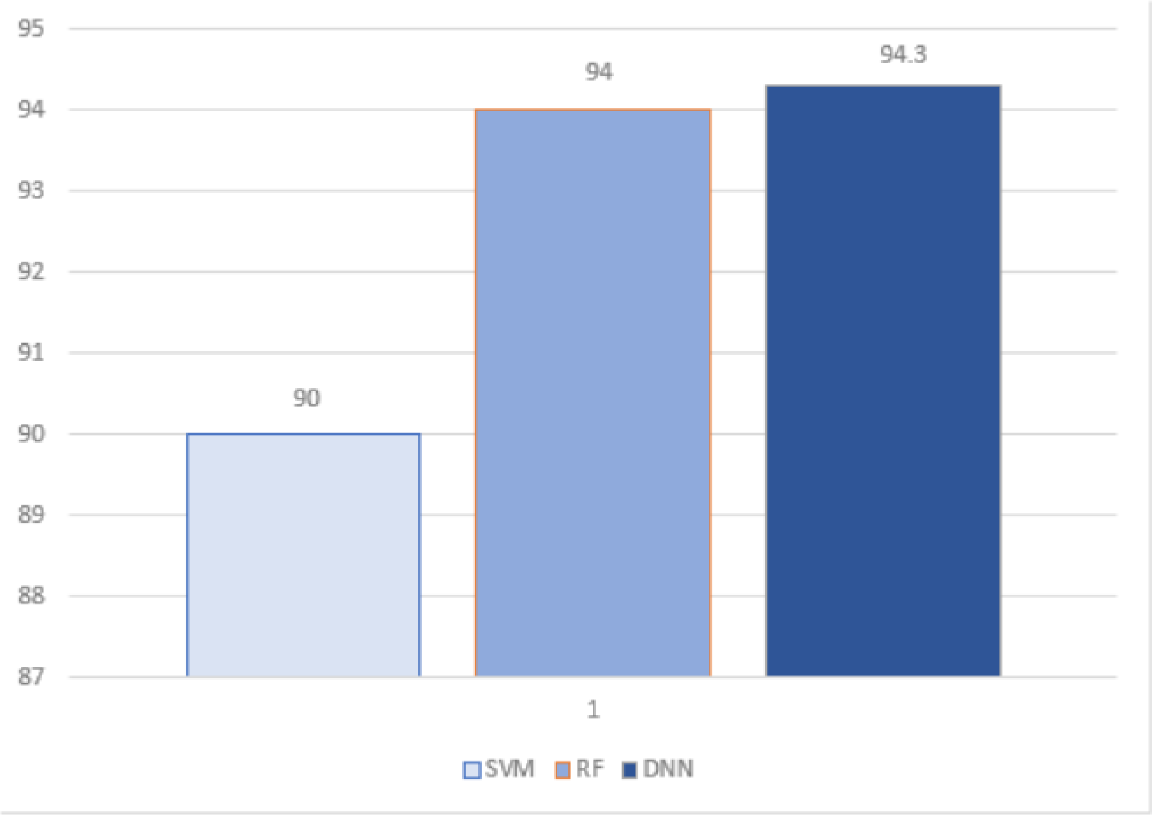


**Fig.: Accuracy Using Support Vector Machine Algorithm**

### c. Performance of model using Neural Networks Algorithm:

Neural networks can be defined as "The algorithms in machine learning are implemented by using the structure of neural networks. These neural networks model the data using artificial neurons. Neural networks thus mimic the functioning of the brain." The 'thinking' or processing that a brain carries out is the result of these neural networks in action. The Neural networks algorithm tries to improve the performance of the model by using smart computational methods to create new and better performing types of prediction and detection model.





**Results**

# Conclusion and Future Scope

In this project, we proposed machine learning algorithms for detecting fake accounts in social media sites in this project. Our system classifies clusters of fake accounts to identify whether they were generated by the same person, rather than making a prediction for each individual account. A new classification algorithm was proposed to improve fake account detection on social networks, where the SVM trained model decision values were used to train a NN model, and SVM testing decision values were used to test the NN model. To reach our goal we used dataset from kaggle and run it into pre-processing phase where different feature reduction techniques were used to reduce the feature vector. In the classification phase learning algorithms were used. The results of the analyses showed that ”SVM-NN” has archived better accuracy results with all feature sets comparing with the other two classifiers, with classification accuracy around 98%. It was noticed that the NN algorithm has the lowest classification accuracy compared with SVM, and SVM-NN. This occurred because the SVM algorithm reaches the global minimum of the optimized function, while the NN used the gradient descent technique, and may reach the local minimum, not global minimum like SVM . It was also noticed that using the feature set provided by PCA, encountered a very low classification accuracy, while the correlation feature set achieves high classification accuracy. This happened because PCA performs dimension reduction and generate a new features based on linear combination of original features. But the correlation approach, and other feature selection techniques select the best set of original features, not linear combination of all features. In other words, feature selection selects the most effective original features, but PCA performs a linear combination of the original features event they are not effective. The correlation feature set records a remarkable accuracy among the other feature selection technique sets, because correlation technique not only select the best features, but also removes the redundancy.

From a modelling viewpoint, one significant direction for future study is to apply feature sets used in other spam detection models, thereby achieving multi-model ensemble prediction. Another direction is to make the system robust against adversarial attacks, such as a botnet that diversifies all features, or an attacker that learns from failures.

# References

[1] S. Khaled, N. El-Tazi and H. M. O. Mokhtar, "Detecting Fake Accounts on Social Media," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 3672-3681.

[2] Rao, K. Sreenivasa, N. Swapna, and P. Praveen Kumar. "Educational data mining for student placement prediction using machine learning algorithms." Int. J. Eng. Technol. Sci 7.1.2 (2018): 43-46.

[3] D. M. Freeman, "Detecting clusters of fake accounts in online social networks", 8th ACM Workshop on Artificial Intelligence and Security, pp. 91-101.

[4] R. Kaur and S. Singh, “A survey of data mining and social network analysis based anomaly detection techniques,” Egyptian informatics journal, vol. 17, no. 2, pp. 199–216, 2016.

[5] L. M. Potgieter and R. Naidoo, “Factors explaining user loyalty in a social media based brand community,” South African Journal of Information Management, vol. 19, no. 1, pp. 1–9, 2017.

[6] Y. Boshmaf, D. Logothetis, G. Siganos, J. Ler´ıa, J. Lorenzo, M. Ripeanu, K. Beznosov, and H. Halawa, “´Integro: Leveraging victim prediction for robust fake account detection in large scale osns,” Computers & Security, vol. 61, pp. 142–168, 2016.

[7] E. V. D. WALT and J. ELOFF, “Using Machine Learning to Detect Fake Identities: Bots vs Humans”, 2169-3536, 2018 IEEE, VOLUME 6, 2018.

[8] M. Tsikerdekis and S. Zeadally, “Multiple account identity deception detection in social media using nonverbal behavior,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 8, pp. 1311– 1321, 2014.

[9] K. Thomas, C. Grier, D. Song, and V. Paxson, “Suspended accounts in retrospect: an analysis of twitter spam,” in Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011, pp. 243–258.

[10] Guille, H. Hacid, C. Favre, D. A. Zighed, “Information Diffusion in Online Social Networks: A Survey”, ACM, 2013,

[11] Whiting, D. Williams, “Why people use social media: a uses and gratifications approach”, Qualitative Market Research: An International Journal Vol. 16 No. 4, 2013 pp. 362-369 q Emerald Group Publishing Limited 1352-2752

[12] Maier, S. Laumer, A. Eckhardt and T. Weitzel, “Giving too much social support: social overload on social networking sites”, European Journal of Information Systems (2014), 1–18 © 2014 Operational Research Society Ltd.