# Fake Account Detection Using Machine Learning

**Chapter** · January 2021

DOI: 10.1007/978-981-15-5258-8_73

CITATIONS
8

READS
776

**4 authors**, including:

Lakshmi Pranathi Yerramreddy
Queen's University Belfast
**5** PUBLICATIONS **13** CITATIONS
SEE PROFILE

Anita Pradhan
Potti Sriramulu Chalavadi Mallikharjunarao College of Engineering & Technology
**12** PUBLICATIONS **249** CITATIONS
SEE PROFILE

Gandharba Swain
K L University
**91** PUBLICATIONS **1,775** CITATIONS
SEE PROFILE

# Fake Account Detection using Machine Learning

Priyanka Kondeti[1,*], Lakshmi Pranathi Yerramreddy [2], Anita Pradhan[3], and Gandharba Swain[4]
[1,2,3,4]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram-522502, Guntur, Andhra Pradesh, India.
Email:[1,*] priyankakondeti1999@gmail.com, [2] pranathi.yr@gmail.com, [3]anita.pradhian15@gmail.com, [4] gswain1234@gmail.com

**Abstract.** Nowadays the usage of digital technology have been increasing exponentially. At the same time the rate of malicious users have been increasing. Online social sites like Facebook and Twitter attract millions of people globally. This interest in online networking has opened to various issues including the risk of exposing false data by creating fake accounts resulting in the spread of malicious content. Fake accounts are a popular way to forward spam, commit fraud, and abuse through online social network. These problems need to be tackled in order to give the user a reliable online social network. In this paper, we are using different ML algorithms like Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and K-Nearest Neighbours (KNN). Along with these algorithms we have used two different normalization techniques such as Z-Score, and Min-Max, to improve accuracy. We have implemented it to detect fake Twitter accounts and bots. Our approach achieved high accuracy and true positive rate for Random Forest and KNN.

**Keywords:** Data mining, Classification, Logistic Regression, Support Vector Machine, K-Nearest Neighbours, Random forest, Normalization.

## 1 Introduction

In this contemporary world, people are being dependent on Online Social Networks (OSNs). As many users are attracted and showing more interest to use OSN in their work-life or for personal uses. This gives an opportunity for the spammers to target people by collecting sensitive information by creating fake accounts. Fake accounts are being created in order to hide their identity and to accomplish their targets [1]. Bucket et al. [2] presented a supervised discretization technique known as Entropy Minimization Discretization (EMD) based on attributes, they have used Naïve Bayes algorithm to evaluate the fake accounts in twitter. This technique can even be applied for all OSNs, for this process they proposed their own dataset with 16 attributes. Finally they presented three different evaluation criteria's before discretization and after that is an increase in accuracy, results from 85.5% to 90.41% and they showed that Naïve Bayes works perfectly for discrete values. To identify fake accounts in OSN Naman et al. [3] proposed a different model in a step by step process firstly they gathered the information and cleaned it. Then created some fabricated accounts. Therefore they validated the data and then injected fake accounts by creating new attributes. At this stage supervised ML techniques are applied and finally results are evaluated. In this process they analysed, identified,

and abolished fictitious bot accounts. This model helped them to identify fake accounts created by humans from a real one.

These activities motivated the researchers to analyse the abnormal activities of Facebook and twitter users by detecting and studying them. Furthermore, in recent years banks and financial providers in the U.S are even analysing Twitter and Facebook accounts before granting the loan [4]. Whereas to create a Robust Fake account detection model Yeh-Chen and Shyhtsun [5] proposed a strategy to analyse user activity by collecting several popular pages which are at an ease to be attacked. They used collector filter accounts to analyse if any malicious activities like spam keywords, extreme promoting a particular company etc. to verify whether they are among the selected group. Then the model is trained. They used three ML algorithms namely Random Forest (RF), C4.5 decision tree algorithm, Adaptive Boosting by giving a cluster of features as input for testing the model and a rank score is produced as output, which produces the probability to be a fake account. In this model, the RF classifier performed the best according to the correction rate and their model performed well in the real-world without any overfit problem.

In the present scenario, researchers are using various techniques to analyse fake accounts on OSN platforms by using various attributes. Some analysists detected fake accounts in OSN using user profile. Some other analysists detected by using both sentiment analysis and user behaviour. Furthermore, some researchers used ego networks to analyse the clusters in the social networks and even their tweets. Some crawling tools are also used to extract the data which is available publicly [6]. Qiang et al. [7] proposed a new OSN user system known as Sybil Rank which is dependent on ranking the users using social graph. In this model, social relationships are bidirectional which helps to detect fake accounts in large scale OSN's. Mauro et al. [8] proposed a new approach according to an empirical analysis and structure of typical social network interactions and their statistics to detect fake accounts created in OSN's. They analysed from dynamic point using social network graphs within the content of confidentiality threats. Kaur and Singh [9] proposed a wide range of approaches like supervised, semi-supervised and unsupervised methods. Besides this to detect anomalies in data mining and social networking domain they even analysed according to cluster based, proximity based and classification based networks. Yazan et al. [10] presented an Integro which is a robust and scalable defence machine by using distinct classification theory. They mainly analysed on real accounts who accepted the fake requests and the process in the Integro system that takes place from user-level activities with the help of supervised machine learning algorithms. Finally Integro uses probability in order to rank user accounts. Tsikerdekis and Zeadally [11] proposed a method using nonverbal behaviour in order to detect and identify deception in online social media. In this paper they used Wikipedia as an experiment and their method achieved a high detection accuracy than other methods. They even demonstrated how developers and designers had overcome these nonverbal data in analysing the deceit by increasing the reliability in online communications. The proliferation of hoaxes, fake news and deceptive online data in Indonesia had casted a tremendous effect on Jakarta election which is already divided [12]. According to the statistics the political advertising spending on Facebook by sponsor category between 2014 and

2018 have been increased. As per the findings, in the measured period, non-profits spent approximately US$ 2.53 millions sponsoring political advertising on Facebook, this demand for online social network giving opportunity for the hacker to spread fake news [13]. Twitter has also become an important outlet for administration and significant for state U.S. communications [14]. In CBC news Facebook shared that 955 million active monthly users in 8.7 percent users could be false accounts or facade. Fake accounts are being mainly used for businesses or for spamming purpose only [15]. Estee and Jan [16] proposed an feature based engineering model to detect bots in the social media platforms (SMPs) by extracting attributes such as follower-count and friend- count and showed an advanced successful score of 41.5% in identification of fake identities fabricated by human using SMPs. However, analysing an account as genuine or fake in any social network is a complicated task and the various attacks like fake profiles, social engineering, fake news and profile compromise are increasing in the online social network. Hence there is a need to improve and implement new techniques. One of the well-known techniques used in recent days is data mining, which was a user friendly in reading reports and significant approach, minimizing the errors and controlling in the standards of data sets.

In this paper we are using a data mining approach which can be used in a wide range of areas including images, business marketing, transactions, banking, hospitals, medicine, insurance, monitoring video, satellites, e-mail messaging and repositories. Data mining is mainly used to extract the data from a collection of data and transforms it to a structure which is understandable for future use like Classification, Clustering, Regression, Association Rules, Prediction, and Sequential patterns. Our approach is based on Classification with normalization, Association Rules, and Prediction by using ML algorithms. There are eclectic algorithms but we have approached supervised learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbour (KNN) and by collecting datasets from Twitter users and online [19]. By using these datasets we performed pre-processing and classification to the data by improving the true positive rate in order to predict the fake accounts on Twitter.

## 2 Related Work

Cao et al. [1] created a system which used pipelining to group accounts into clusters; they implemented a cluster level model to detect fake accounts by giving cluster level features for ML algorithms as input. They proposed a generic pattern encoding algorithm to compute statistical features. Pipelining is based on three levels; firstly they implemented a Cluster Builder for cluster account score, secondly profile features were utilized for feature extraction and evaluate ML, thirdly account scorer is generated after training and evaluating ML models on new data. They used three algorithms like SVM, RF and LR. Among those three, RF gave the best result for all metrics 95% and even for the out-of-sample testing data 97%.

Using an ML algorithm Sarah et al. [4] implemented a new classification algorithm SVM-NN by combining SVM and NN (Neural Networks) along with the Management Information Base (MIB) baseline dataset to improve the efficiency for detecting bots and fake twitter accounts, they used dimension reduction and attribute selection methods. This new technique used a very tiny fraction of attributes even though they are able to classify correctly about 98% of the twitter accounts in their dataset.

Apart from traditional methods Myo and Nyein [6] proposed a new method by creating their own blacklist by data collection, pre-processing, extracting topic and keywords from Honeypot dataset. In order to show difference between fake accounts and legitimate accounts, first they have done feature extraction. They have extracted features from user which is available publicly. Then they classified the data using decorate which is a specially designed artificial training examples for meta-learner to build a diverse ensemble classifier. They have added a classified data to this ensemble to increase the rate of accuracy. This method is repeated until they have reached the maximum iteration and achieved their desired committee size. Twitter APIs are collected and this blacklist is used in the attribute extraction process. Finally classification is done as real or fake. Their method has reached an accuracy of 95.4 per cent, while the true positive value is 0.95.

In the previous papers, researchers used cluster-based features, network-based features, keyword and Profile analysis, classification algorithms due to crawling problems, network-based and profile features are difficult to extract. In this paper we are implementing an account level detection by applying supervised learning methods to a twitter dataset beside this Z-score and Min-Max normalization is mainly used to improve the accuracy for SVM, LR, RF and KNN to detect the fake accounts.

## 3 Proposed Work

The proposed work comprises of four steps, (i) data classification, (ii) data pre-processing, (iii) data reduction or transformation, and (iv) Algorithm selection. These steps are described below.

### 3.1 Data Classification

Data classification is characterized as the process for deciding the appropriate type, origin of data and appropriate resources for collecting data. In the data classification step the data is selected from various Twitter accounts. We collected twitter datasets for analysis and to test our model, dataset consists of different attributes such as name, status-count, friend-count and followers-count.

We selected these columns as feature attributes status-count, friends-count, followers-count, sex-code, favourites-count, Lang-code.

### 3.2 Data Pre-processing

We used machine learning algorithms in this process to convert and analyse the available raw data into feasible data. It is mainly used for better and accurate results. For instance, some algorithms like Random Forest do not support null values or they need particular format. In such a case data pre-processing is a necessary step. Then we extracted two Comma-Separated Value (CSV) files fake and genuine users, we combined both files by sampling noise in the data then feature labels were added as 0/1 to distinguish fake or real.

We selected particular columns as feature attributes status-count, friends-count, followers-count, listed-count, sex-code, favourites-count, Lang-code and then we have removed columns having more null values sex-code, Lang-code, and normalized the data for better accuracy.

In this method, we distributed the information for training and testing purpose at 80:20. To represent the values in confusion matrix, the model is trained with x- train and y-train data, and then it is trained with test data for precision and recall values. Graphs are represented as Area under ROC Curve (AUC) and Receiver operating characteristic curve (ROC).

### 3.3 Data transformation

It is a method of converting information from one format or structure into another format. For tasks such as data integration and management, data transformation is a crucial step to improve the accuracy. In our model we used two normalization techniques for data transformation.

#### Normalization in Data Mining

We are using two data normalization techniques such as Z-Score, Min-Max, it is mainly used when dealing with multiple attributes on different scale and to scale the information into smaller range, it is commonly applied for classification algorithms to improve the performance rate, so the attributes are normalized to bring on the same scale.

**Z-Score:** Z-Scores are mainly based on mean value and standardized score these scores are linearly transformed data value with a mean of 0 and the scores have been given a common standard. It helps to understand the rate of a score as per the normal distribution of the data.

**Min-Max:** In this method linear transformation is performed on original data, according to this minimum values of that feature is transformed as 0 whereas maximum values are transformed into 1 and remaining values have been changed into decimals between 0 and 1.

### 3.4 Algorithm Selection

**K-Nearest Neighbour**

K-Nearest Neighbour (KNN) is the type of supervised learning method. It is used for both pattern recognition along with classification. In KNN, a specific test tuple set is compared to the training data set already available which is identical to the test data set. It calculates the distance between the training data and the testing data using the Euclidean distance function. Class membership is the output of the KNN classification. Therefore KNN classification has two stages, the first is the determination of the nearest neighbours and the next step is the determination of classes using the neighbours.
The working process of KNN classifier is defined below
1. Distance between the attributes is calculated from testing and training data sets.
2. Training data is sorted according to the distance values.
3. Obtain the neighbours (k) which are approximately close to the testing data.
4. Majority class of training data is added to the testing data.
The Euclidean distance between the training data set and testing data set is estimated using Eqn. (1) where $p_i$ stands for an element of training dataset and
$q_i$ stands for an element of testing dataset. This D(p,q) derives the smallest distance k and locate the corresponding data.

$$D(p,q) = \sqrt{\sum_{i=0}^{n}(p_i - q_i)^2} \qquad (1)$$

**Random Forest Algorithm**

The Random Forest consists of a large number of individual trees which functions as an ensemble individual tree divides a random forest class prediction, and the class with the most votes is our model's prediction.

Random forest (RF) is similar to bootstrapping algorithm along with Regression tree Classification and Decision Trees(CART).We have 1000 observations of 10 parameters in the entire population. RF attempts to create several CART models with different initial variables with samples. For example, it chooses randomly from the sample of 100 observations and randomly 5 are selected from initial variables to design a CART model. This procedure is repeated 10 times and each observation is analysed finally. A final prediction is also a function of each prediction and this prediction is simply a mean value.

**Support Vector Machines (SVM)**

Support Vector Machine (SVM) is also a type of computer algorithm that can be trained to assign labels to the objects. It is a powerful tool for solving both classification and regression problems. It is one of the supervised learning methods and one of the best-known classification methods. SVMs are based on statistical

learning theory, which is used to solve two-class binary problems without the loss of generality.

The main objective of SVM classifiers is to locate the decision boundaries like hyper planes, which produces the separation of classes optimally is mainly used to resolve the linear problems besides this it can be extended to handle nonlinear decision problems. SVM's features is a good generalization performance, lack of local minima and sparse solution distribution. It is based on the principal of Structural Risk Minimization (SRM) that minimizes the generalization error.

**Logistic Regression**

Logistic Regression (LR) it is a type of linear algorithm, which is the method of relating dependent and independent variables using a logistic distribution functional form. The regression model can be mathematically designed by describing the likelihood of certain events Eqn. (2). It obtains a linear relationship between the output and input. LR measures the probability of class inclusion for one of the data set's different categories. This is used for modelling the binary response data. If the response is in binary, it takes the form of the success and indicates failure. Consider data, weights and class label 1/0.

$$P(c = \pm 1 | d, a) = \frac{1}{1+\exp\left(-c\left(a^T d+b\right)\right)} \qquad (2)$$

The proposed technique has been represented in the form of a flowchart in Fig.1.

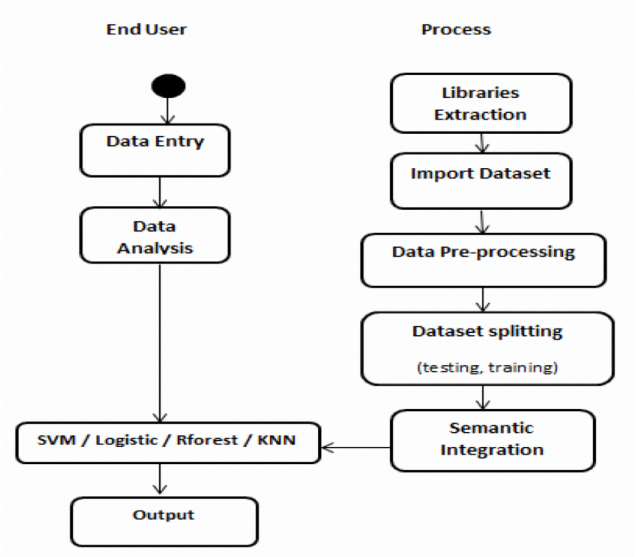



**Figure 1: Fake account detection strategy using data mining techniques.**

## 4 Results and Experimentation

The hardware used to implement this program is a laptop with i5 processor, 500 GB Memory and 4 GB Ram. The Software used for this implementation is Anaconda, and the language used is Python. The dataset used in this program is taken from twitter online repository. The result of this technique is obtained and measured using the following metrics.

### Confusion Matrix

In the area of machine learning, confusion matrix solves the statistical classification problem. Confusion Matrix is often used to discuss about the functioning of the classification system of an algorithm on a collection of testing data that is defined for the true positive values. It is also called as an error matrix.
This allows ambiguity between groups to identify easily, e.g. one class is often mislabelled as another. Maximum performance measurements are obtained from the confusion matrix.

**Table 1: Confusion Matrix**

| | Class 1 Predicted | Class 2 Predicted |
|---|---|---|
| Class 1 Actual | TP | FN |
| Class 2 Actual | FP | TN |

Here,
• Class 1: Positive
• Class 2: Negative

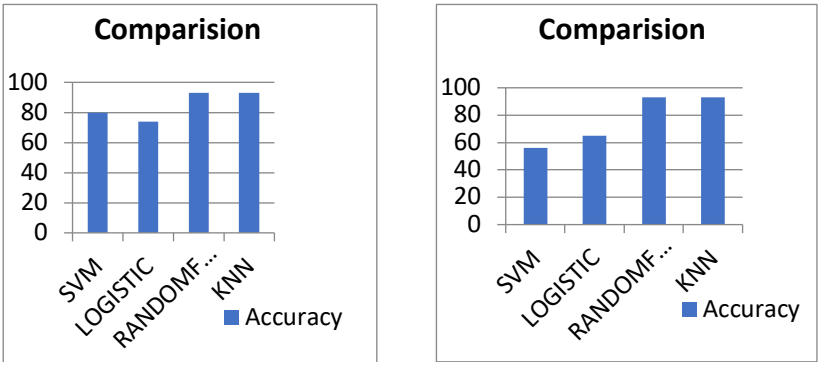
### Description of Terms

• Positive (P): positive values(for instance: is a ball).
• Negative (N): If the values are not positive (for example: is not a ball).
• True Positive (TP): If the prediction is positive, but the observed values are even positive.
• False Negative (FN): Examined as positive, and the predicted value is negative.
• True Negative (TN): If the examined values are negative, but predicted as negative.
• False Positive (FP): Examined values are negative, and predicted as positive.
Accuracy for detecting fake accounts can be obtained by using TP, TN, FP and FN from Eqn. (3).

$$\text{Accuracy Prediction} = \frac{TP + TN}{TP+TN+FP+FN} \qquad (3)$$

AUC, ROC curve is an efficiency calculation at various threshold rates for the classification issues. ROC is a curve of likelihood and AUC is the degree of separability factor. This shows the design value that can differentiate between classes. The model predicts accurately as 0's and 1's if the AUC is higher and it distinguishes better between accounts as fake or real.

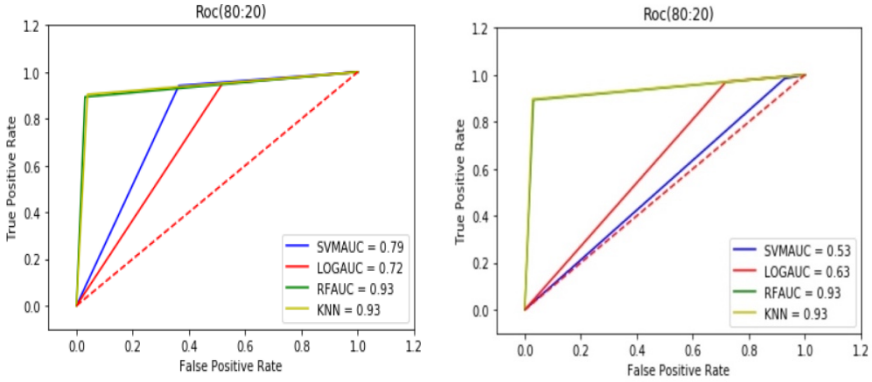
Based on the techniques described before, we assessed our strategies with ML classifiers, which includes Random Forest, Logistic Regression, KNN, and SVM. In this experiment for testing the model we selected the datasets from twitter, these datasets are pre-processed and split into 80:20 cross-validation process this split data is used for testing and training the algorithms. We applied our technique to the testing data as per the analysis Random Forest and KNN performed well with high accuracy (98%). By applying Z-Score Normalization SVM and Logistic regression, accuracy rate had been increased as shown in Figure 2.




**(a) Z-Score Normalization** **(b) Min-Max Normalization**

**Figure 2: Accuracy prediction using machine learning algorithms by account level.**

To calculate the performance of the classifier, we generated a ROC curve, based on true positive and false Positive Rate. The rank allocated to each account is based on the results of a particular classifier, so we tested the reliability by concentrating on a different range of ranks to see if our method works better by the rank created by our classification system. We used the ROC curve which shows the cut-off for a test, whereas the best cut off has the highest true positive rate with a low false-positive rate, as shown in Figure 3. ROC curve based on 80:20 split and account-level classification worked well with RF and KNN by giving a 93% True Positive rate with both type of the normalization techniques, however, SVM and Logistic performance are very low in Min-Max normalization when compared to Z-Score. On the other hand, we even focused on AUC. It can provide an integrated probability of performance with comparison to all possible classification thresholds. AUC is the rate at which the model is ranked random positive more than a random negative.




**(a) Z-Score Normalization** **(b) Min-Max Normalization**

**Figure 3: Comparison of Supervised algorithms using ROC curve learning.**

Table 2 shows the AUC precision for four algorithms at the account level, this shows that the identification of each algorithm is more accurate for every single account. When compared to the other algorithms we can observe that Random Forest and KNN are giving the highest accuracy.

**Table 2: 80:20 split account level testing performance**

| Z-Score Normalization | | Min-Max Normalization | |
|---|---|---|---|
| **Algorithm** | **AUC** | **Algorithm** | **AUC** |
| KNN | 0.932 | KNN | 0.934 |
| Random Forest | 0.931 | Random Forest | 0.931 |
| Logistic Regression | 0.715 | Logistic Regression | 0.626 |
| SVM | 0.788 | SVM | 0.528 |

## 5. Conclusions and Future Work

Finally, we conclude that our research has been done to analyse, detect and remove the fake accounts created on Twitter, but our method can be applied to other datasets from OSN platforms such as Facebook and LinkedIn. Due to the ease of ML algorithms, fake accounts can be analysed using different ML classifiers. In this paper we have used SVM, KNN, Random forest, logistic algorithms along with Z-Score and Min-Max normalization techniques to predict fake users. By using these techniques we have improved the accuracy to 98%. As future work, this can be extended by implementing feature selection for this method or to cluster the accounts into groups and develop an efficient model by analysing more data to improve the accuracy for predicting the fake account in OSN.

## References:


1. Cao, X., David, MF., Theodore, H.: Detecting Clusters of Fake Accounts in Online Social Networks. In: 8th ACM Workshop on Artificial Intelligence and Security, pp. 91-101 (2015).
2. Buket, E., Ozlem, A., Deniz, K., Cyhun, A.: Twitter Fake Account Detection. In: IEEE 2nd International Conference on Computer Science and Engineering, pp. 388-392 (2017).
3. Naman, S., Tushar, S., Abha, T., Tanupriya, C.: Detection of Fake Profile in Online Social Networks Using Machine Learning. In: IEEE International Conference on Advances in Computing and Communicaton Engineering. pp. 231-234 (2018).
4. Sarah, K., Neamat, E., Hoda, M .O .M.: Detecting Fake Accounts on Social Media. In: IEEE Intenational Conference on Big Data. pp. 3672-3681 (2018).
5. Yeh-Cheng, C., Shyhtsun, F .W.: FakeBuster: A Robust Fake Account Detection by Activiy Analysis. In: IEEE 9th International Symposium on Parallel Architectures, Algorithms and Programming. pp. 108-110 (2018).
6. Myo, MS., Nyein, NM.: Fake Accounts Detection on Twitter using Blacklist. In: IEEE 17th International Conference on Computer and Information and Information Science. pp. 562-566 (2018).
7. Qiang, C., Michael, S., Xiaowei, Y., Tiago P.: Aiding the Detection of Fake Accounts in Large Scale Social Online Services. In: 9th USENIX Conference on Networked Systems Design and Implementation. pp. 1-14 (2012).
8. Mauro, C., Radha, P., Macro, S.: Fakebook: Detecting Fake Profiles in Online Social Networks. In: IEEE International Conference on Advances in Social Networks Analysis and Mining. pp. 1071-1078 (2012).
9. Kaur, R., and Singh, S.: A survey of data mining and social network analysis based anomaly detection techniques. In: Egyptian informatics journal. pp.199–216 (2016).
10. Yazan, B., Dionysios, L., Georgos, S., Jorge, L., Jose, L., Matei, R., Konstatin, B., and Hassan, H.: Integro: Leveraging victim prediction for robust fake account detection in large scale osns ,Computers & Security. pp. 142–168 (2016).
11. Tsikerdekis, M., Zeadally, S.: Multiple Account Identity Deception Detection in Social Media using Non Verbal Behaviour. In: IEEE Transactions on Information Forensics and Security. **9**(8), 1311-1321 (2014).
12. How fake news and hoaxes have tried to derail jakarta's election. Internet draft(online).
13. Political advertising spending on facebook between 2014 and 2018 Internet draft. [Online].Available:https://www.statista.com/statistics/891327/politicaladvertisingspending- facebook by-sponsor-category/.2018.

14. Statista.twitter: number of monthly active users 2010-2018. Internet draft. [Online].Available:https://www.statista.com/statistics/282087/number-of-monthlyactive- twitter-users/.2018.
15. Cbc.facebook shares drop on news of fake accounts. Internet draft. [Online].Available:http://www.cbc.ca/news/technology/facebook-shares-drop-onnews-of-fake-accounts-1.1177067.2012.
16. Estee, VDW., Jan, E.: Using Machine Learning to Detect Fake Identies: Bots vs Humans. In: IEEE Access, pp. 6540-6549 (2018).
17. The list of email spam trigger words. http://blog.hubspot.com/blog/tabid/6307/bid/30684/The-Ultimate-List- of-Email-SPAM-Trigger-Words.aspx.
18. Bayuk, J. (ed.).: "Cyber Forensics: Understanding Information Security Investigations," Springer's Forensic Laboratory Science Series, Humana Press. pp. 59- 101 (2010).
19. https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/user-object .