

# Detect Fake Identities Using Machine Learning

<sup>1</sup>Ms.K.Nagalakshmi, <sup>2</sup>Ms.P.Nanthini, <sup>3</sup>Ms.A.Saranya, <sup>4</sup>Mrs.B.Revathi  
<sup>1,2,3</sup> Students, CSE dept, Mangayarkarasi College of engineering, Madurai, Tamilnadu.  
AP/ CSE dept, Mangayarkarasi College of engineering, Madurai, Tamilnadu.

## 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 others, 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. In this paper, a new algorithm, SVM-NN, is proposed to provide efficient detection for fake Twitter accounts and bots, 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), and our newly developed algorithm, SVM-NN. The proposed algorithm (SVM-NN) uses less number of features, while still being able to correctly classify about 98% of the accounts of our training dataset*

**Keywords—** Social, Machine Learning, Networks

## I. INTRODUCTION

Identity deception on big data platforms is an increasing problem, due to the continued growth and exponential evolution of these platforms. Social media is one of the preferred means of transmission [1] and has become a target for spammers and scammers alike [2]. Cyberthreats like spamming, which involves the sending of unsolicited emails, are common in email applications. These same threat and more now materialize on social media platforms (SMPs), although in different demonstration.

Much can be learned about people's behaviour and needs through analysing their synergy with one another. Habits and topics of conversations can be

evaluated to deliver a better service or product to customers and ultimately to people at large [1], [3]. The same information can however also be used against people, very often in a ambiguous way. For example, a cluster of people may consequence an opinion [4] when the other participants in the conversation are unaware that the "people" in the cluster are not real Since the detection of fake social engagement is quite challenging [5], this vulnerability is greatly abused [6]. We believe that these fake accounts can be attributed to, among others, the following factors.

The privacy policies of SMPs not expecting persons to reveal their true identity [7]. The authenticity of people is constantly being questioned [1], and this can detrimentally affect [2] those who are falsely accused or misled. An example is the case of cyberbullying [8] where children are bullied online through the spreading of false rumours. Malicious individuals and groups on SMPs striving to spread chaos and pandemonium. A recent example was the spreading of fake news about Hurricane Sandy in the US [9]. False news about the hurricane went viral and became a main source of information for those affected by the storm.

The gratification of sites, with more "likes" or "followers" inadvertently meaning greater popularity and instances of identity deception by humans on SMPs. We found that standard attributes alone, such as the number of friend and followers that are available through application programming interfaces (APIs) and describing accounts in SMPs like Twitter, were not sufficient to successfully detect fake identities created by humans. In this paper we evaluate whether readily available and engineered features that are used for the successful detection, using machine learning models, of fake identities created by bots or computers can be used to detect fake identities created by humans. This is done in the hope that similar features can serve as a catalyst for uncovering identity deception by humans on SMPs.

Higher social ratings [2]. This trend drives people to find new means to artificially or manually [2] stay ahead of their competitors. By analogy, the most popular candidate in a political election usually receives most of the votes [10]. The ease with which false

accounts and actions can be obtained. An example is false accounts being bought online at a marketplace [11] at minimal cost, or delivered through crowdsourcing services [12]. It is even possible to buy Twitter followers and Facebook “likes” online [5].

## **II. RELATED WORK**

If a certain message contains these words or number of words these are addressed to as spam. These rules have also been used in social media platform with a success. Although the main drawback is that the process of developing new words are easy and constant and the use of shortened words are becoming more common on platform for example lol which means laugh out loud. Pattern matching techniques are being used to detect these shortened words on these platforms. For instance, from any account a tweet is published regarding trending information on the social media platform or a new account not more than a day old starts advertising about the trending topics is regarded as fake [5]. Facebook uses algorithms which can identify bots using the number of friends for deceptiveness which could either be related to relationships history or tagging. The above rules specified are successful for identifying bot accounts by have not been successful in identifying fake accounts by humans [11].

Unsupervised ML was effectively connected by "Gu et al. His exploration demonstrated bunching, that is a typical Unsupervised ML technique, could be utilized for distinguishing bots. By Unsupervised ML machine taking in, information are not labelled, information is assembled in view of closeness. Grouping functions admirably to recognizes bots as these bots as a rule share co attributes and has a similar reason.

If a certain message contains these words or number of words these are addressed to as spam. These rules have also been used in social media platform with a success. Although the main drawback is that the process of developing new words are easy and constant and the use of shortened words are becoming more common on platform for example lol which means laugh out loud. Pattern matching techniques are being used to detect these shortened words on these platforms. For instance, from any account a tweet is published regarding trending information on the social media platform or a new account not more than a day old starts advertising about the trending topics is regarded as fake [5]. Facebook uses algorithms which can identify bots using the number of friends for deceptiveness which could either be related to relationships history or tagging. The above rules specified are successful for identifying bot accounts by have not been successful in identifying fake accounts by humans [11].

All RR interval series were also checked by visual inspection analysis. Subjects with sustained tachyarrhythmia were excluded from the study. Baseline clinical characteristics of the patients enrolled in this study are shown in Table I. The study was approved by the ethics committee of Fujita Health University and conformed to the principles outlined in the Declaration of Helsinki. All patients provided written informed consent

## **III. CLASSIFICATION**

Classification consists of predicting a certain result based on a given input. In order to predict the result, the algorithm processes a training set contain a set of attributes and the individual outcome, usually called prediction attribute.

Data classification is the process of formulating data into categories for its most potent and adequate use. There is some algorithm in classification which helps to analyze our work are random forest, SVM, ANN.

### **A. Recursive**

A recursive algorithm is one that calls itself repeatedly until a certain condition matches. It is a method common to functional programming. Iterative algorithms use repetitive constructs like loops. Some problems are better suited for one implementation or the other

### **B. Logical**

An algorithm may be viewed as controlled logical deduction. A logic component expresses the axioms which may be used in the computation and a control component determines the way in which deduction is applied to the axioms.

### **C. Serial**

Algorithms are usually discussed with the assumption that computers execute one guidance of an algorithm at a time. This is a serial algorithm, as antithetical to parallel algorithms, which take advantage of computer architectures to process several instructions at once.

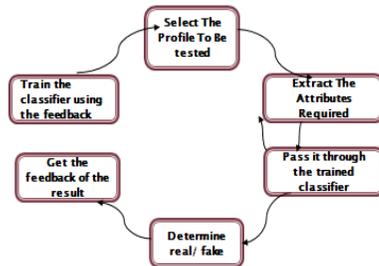
### **D. Deterministic**

Deterministic algorithms solve the problem with a predefined process whereas non-deterministic algorithm must perform guesses of best solution at each step through the use of heuristics.

## **IV. PROPOSED SYSTEM**

Classification starts from the selection of profile that needs to be classified. Once the profile is selected, the useful features are extracted for the purpose of classification.

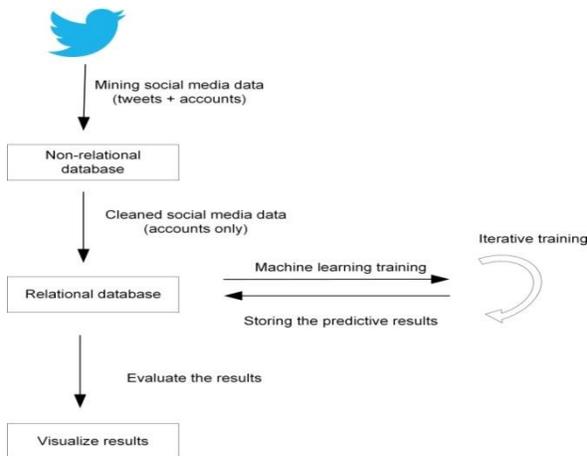
The extracted features are then fed to trained classifier. Classifier is trained regularly as new data is fed into the classifier. Classifier then determines whether the profile is genuine or fake .



The result of classification algorithm is then verified and feedback is fed back into the classifier. As the number of training data increases the classifier becomes more and more accurate in predicting the fake profiles.

The social networking sites are making our social lives better but nevertheless there are a lot of issues with using these social networking sites.

These are done mostly by using fake profiles. we came up with a framework through which we can detect a fake profile using machine learning algorithms so that the social life of people become secured.



A non-relational database is a database that does not incorporate the table/key model that relational database management systems (RDBMS) promote.

These kinds of databases require data manipulation techniques and processes designed to

provide solutions to big data problems that big companies face

A relational database is a set of formally described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables.

The standard user and application programming interface (API) of a relational database is the Structured Query Language SQL statements are used both for interactive queries for information from a relational database and for gathering data for reports

### V. SUPPORT VECTOR MACHINE

More formally, a support-vector machine erect a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification,backsliding, or other tasks like deviation detection. Naturally, a good estrangement is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class, since in general the larger the margin, the lower the generalization error of the classifier

Support Vector Machine is an elegant and robust technique for classification on a large data set not unlike the data sets of Social Network with several millions of profiles It is a binary classification algorithm that finds the maximum separation hyper plane between two classes. It is a supervised learning algorithm that given enough training examples, divides two classes fairly well and classifies new examples.maximum margin Classification - Given a weight vector  $w$  and bias weight  $b$ , we formulize the classification methodology as:

$$wTx + b > 0 \Rightarrow \text{positive class}$$

$$wTx + b < 0 \Rightarrow \text{negative class}$$

This equation gives a separator and it is intuitive that depending upon the choice of the above-mentioned parameters; we can have several separators for the same dataset.

### VI. EXPERIMENTAL RESULTS

The engineered features created during step 5 of the research were explored to understand the corpus and it was noted that most accounts had few friends and followers. The distribution of friends.

The data exploration looked at the profile descriptions of these accounts. The exploration showed that not all accounts had a profile description and that

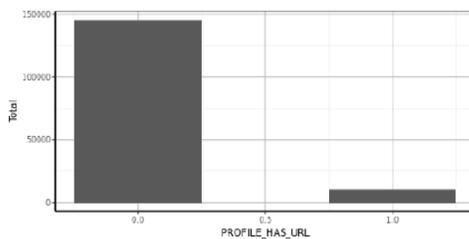
some profile descriptions were shared among accounts, a few profile descriptions also contained URLs.

The results of profiles having an URL as part of their profile is where 0 D no and 1 D yes. These exploratory results showed that even though we are dealing with human accounts only, they still show characteristics known to bots, such as having a URL in their profile description. This further affirmed that research previously conducted to detect fake bot accounts on SMPs could well be applicable to detect fake human identities too.

Confusion Matrix is a technique for epitomizing the act of a classification algorithm. Calculating a confusion matrix can give you a better idea of what your classification model is getting right and what types of lapse it is making.

**True Positive Rate (TPR) =  $TP / TP + FN$**   
**False Positive Rate (FPR) =  $FP / FP + TN$**   
**True Negative Rate (TNR) =  $TN / FP + TN$**   
**False Negative Rate (FNR) =  $1 - TPR$**

Features	svmRadial	rf
ACC_AGE_MONTHS	58.70	34.64
DUP_PROFILE	89.65	79.02
FF_RATIO	0.57	1.24
FOLLOWERS_COUNT	1.76	21.67
FRIENDS_COUNT	41.79	20.07
GEO_ENABLED	28.10	21.23
HAS_IMAGE	0.00	0.00
HAS_NAME	100.00	100.00
HAS_PROFILE	9.03	11.93
LISTED_COUNT	14.05	9.68
PROFILE_HAS_URL	9.87	20.24
STATUS_COUNT	43.05	51.89
USERNAME_LENGTH	58.70	34.64



**Supervised machine learning results.**

Model	Accuracy	F1 score	PR-AUC
svmLinear	68.05%	32.16%	27.76%
rf	87.11%	49.75%	49.90%
Adaboost	85.91%	47.54%	49.53%

Furthermore, a tailed distribution seemed to occur regarding the length of user names chosen for accounts. Any outliers on this distribution could indicate potential deception, as supervised machine learning models will be able to detect this type of anomaly.

Features are selected to apply classification algorithms. The classification algorithm is discussed further.

Attributes are selected as features if they are not dependent on other attributes and they increase efficiency of the classification. The features that we have chosen are discussed further.

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

**VII.CONCLUSION**

In the end, we conclude that the research work have been done to detect, identify and eliminate fake bot accounts created and cyborgs cannot be used for differentiating fake account created by human beings.

As machine learning has evolved in recent days. We can differentiate fake accounts easily by applying a data set with fake accounts and marking them as fake and real accounts marking them as real . So, after the model knows which account fake and which account is real, the model will be successfully able to differentiate a fake account created by human from a real one when the actual data set will be given to it.

The Findings indicate that engineered features that were previously used to detect fake accounts generated by bots, at best predicted fake accounts generated by humans with an F1 score of 49.75%.

This can be attributed to the fact that humans have different characteristics and behaviours than bots which cannot be modelled similarly we investigated whether the results from past studies to detect bot accounts could be applied successfully to detect fake human accounts.

A corpus of human accounts was enriched with engineered features that had previously been used to successfully detect fake accounts created by bots.

### REFERENCES

- [1] S. Gurajala, J. S. White, B. Hudson, B. R. Voter, and J. N. Matthews, "Profile characteristics of fake Twitter accounts," *Big Data Soc.*, vol. 3, no. 2, p. 2053951716674236, 2016.
- [2] C. Xiao, D. M. Freeman, and T. Hwa, "Detecting clusters of fake accounts in online social networks," in *Proc. 8th ACM Workshop Artif. Intell. Secur.* 2015, pp. 91\_101.
- [3] S. Mainwaring, *We First: How Brands and Consumers Use Social Media to Build a Better World*. New York, NY, USA: Macmillan, 2011.
- [4] V. S. Subrahmanian et al. (2016). "The DARPA Twitter bot challenge." [Online]. Available: <https://arxiv.org/abs/1601.05140>
- [5] Y. Li, O. Martinez, X. Chen, Y. Li, and J. E. Hopcroft, "In a world that counts: Clustering and detecting fake social engagement at scale," in *Proc. 25th Int. Conf. World Wide Web*, 2016, pp. 111\_120.
- [6] D. M. Freeman, "Detecting clusters of fake accounts in online social networks", 8th ACM Workshop on Artificial Intelligence and Security, pp. 91–101.
- [7] B. Hudson, B. R. Voter, "Profile characteristics of fake twitter accounts", *Big Data & Society*, 2016.
- [8] S. Durst, L. Zhu, "The darpa twitter bot challenge," *arXiv*, 2016. 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE-2018) Paris, France 22-23 June 2018 C. Beileites, K. Geiger, M. Kirsch, S. B. Sobottka, G. Schackert, and R. Salzer, "Raman spectroscopic grading of astrocytoma tissues: Using soft reference information," *Anal. Bioanal. Chem.*, vol. 400, no. 9, p. 2801, 2011.
- [9] G. Gu, R. Perdisci, J. Zhang, and W. Lee, "BotMiner: Clustering analysis of network traffic for protocol-and structure-independent botnet detection," in *Proc. USENIX Secur. Symp.*, vol. 5. 2008, pp. 139\_154.
- [10] W. Wu, J. Alvarez, C. Liu, and H.-M. Sun, "Bot detection using unsupervised machine learning," *Microsyst. Technol.*, vol. 24, no. 1, pp. 209\_217, 2018.
- [11] M. Yahyazadeh and M. Abadi, "BotOnus: An online unsupervised method for botnet detection," *ISC Int. J. Inf. Secur.*, vol. 4, no. 1, pp. 51\_62, 2012.
- [12] S. Venkatesan, M. Albanese, A. Shah, R. Ganesan, and S. Jajodia, "Detecting stealthy botnets in a resource-constrained environment using reinforcement learning," in *Proc. Workshop Moving Target Defense*, 2017, pp. 75\_85.
- [13] M. H. Arif, J. Li, M. Iqbal, and K. Liu, "Sentiment analysis and spam detection in short informal text using learning classifier systems," in *Soft Computing*. Berlin, Germany: Springer, 2017, pp. 1\_11.
- [14] D. Bogdanova, P. Rosso, and T. Solorio, "Exploring high-level features for detecting cyberpedophilia," *Comput. Speech Lang.*, vol. 28, no. 1, pp. 108\_120, 2014.