# Abnormal User Detection of Malicious Accounts in Online Social Networks using Cookie Based Cross Verification

**Nirmala B, SP.Chokkalingam, G.Sai Neelima**

***Abstract:*** ***Malicious account detecting is a serious problem on the Internet today. Online social media services like Facebook, LinkedIn, and Instagram, these services include good quality service like opinions, comments as well as poor quality services like rumors, spam, and other malicious activity. In this paper, we review the existing research work done on Facebook, Instagram and LinkedIn, study the techniques used to identify and analyze the poor quality content on Facebook, and other social networks, and we proposed a combined technique like dynamic user profile verification and cookie-based cross-verification to detect malicious activity in an online social network by using random forest machine learning algorithm We also attempt to understand the limitations posed by Facebook in terms of availability of data for collection, and analysis, and try to understand if existing techniques can be used to identify and study poor quality content on Facebook and other social networks.***

***Keywords: Machine Learning, Online Social Networks, random forest algorithm***

## I. INTRODUCTION

A special kind of website allows the people having same interests to come together for sharing information, images and videos is social network. People can share information may be a business or personal perspective in social networking. People interact with each other using various forms as audio, video, images and texts. Social networking is an online platform where people can create and build social relationship with other people and share similar personal, career, and real-life interests. The social network is a distributed network where computers, people, knowledge and organisations. The features and format of the social network services are different from one another. The devices involved in social network communication are desktop, laptop, tablet, and smartphones. Various services are interlinked with the social network are individual-centred or group-centred services. In certain situation government, e-business, and educational information are shared.

**Revised Manuscript Received on July 13, 2019.**

**Nirmala B,** Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. Assistant Professor, Department of Computer Science and Engineering, Sri Ramachandra Faculty of Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research, Chennai.

**SP.Chokkalingam,** Professor, Department of Computer Science Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai.

**G.Sai Neelima,** UG Student, Department of Computer Science Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai.

Twitter, Facebook, QZone, Telegram, VKontakte, and Odnoklassniki are some of the social networks function based on web platform. People share activities and interests across economic, geographical and political impacts. Using e-mail, chatting, instant-messaging are created by encouraging the cooperation. Twitter is a network where people share mainly about political, economical and entertainment information. Facebook is a network where people can share entertainment, family, personal and other life-oriented information. Linked-in is also a network, where people can share educational, career-based information.

### Online Social Network Sites (OSNS)

Online Social Network Sites (OSNS) from ancient time man is called as a social animal. From his beginning man has maintained a social relation with nature, animals and with a fellow human being. It was this social relationship that helps him to have a close relationship in the universe with one another. In modern times with increase in population the OSNSs have become an easy and a much efficient platform in maintaining social relationships. Online Social Network sites like Facebook, Twitter, Linkedln, and Sharechat has become popular sites in Internet platform. They have attracted of all ages from technicians to novice users. In the wide area sphere like research, industries, business, working Office, news media, organization, entrepreneurship OSNS have become a daily practice in use. Mostly OSNS have mainly used for information sharing and to express on common interest views example political view.

### Service Providers

Table 1 provides details about online social networks providers, utilization ways and launched date.

**Table. 1 Online Social Network**

| Service Provider Name | Description | Date Launch |
|---|---|---|
| Facebook | General: photos, videos, blogs, apps | February 2004 |
| Instagram | A photo and video sharing site. | October 2010 AS |
| LinkedIn | Business and professional networking | May 2003 |

**Malicious Account**

Duplicate Account: A duplicate account refers to an account maintained by a user in addition to his/her principal account. False Accounts: False accounts are further broken down into two categories user misclassified accounts and undesirable accounts.

User-misclassified accounts: It represents the personal profiles created by users for a business, organization, or non-human entity such as a pet (Facebook's terms of service permits such entities as a Page rather than a personal profile).

Undesirable accounts: These are the user profiles that are intended to be used for purposes that violate Facebook's terms of service, such as spamming. Fake accounts are mainly used to unfairly increase ones power and influence within a target community.

**Literature Survey**

DeepScan method is proposed in [1] for malicious detection based on the user account information especially the dynamic behaviour of the user. DeepScan is specially designed for Location based social networks (LBSN). In our day to day life, the number of social networks users are increasing rapidly with the huge amount of information. Since the volume of data transfer in the LBSN is huge, data analyzation is a crucial task. Earlier machine learning approaches are semi-automatic because the feature extraction is manual. DeepScan also involves Long-short-term-memory neural network for analysing time series user activities. It combines both time series features and conventional features and used supervised machine learning approaches for anomaly detection.

Though the efficiency of DeepScan is lesser for variety of data. Though DeepScan has various limitations such as, DeepScan is an idea used for exploring the way of identifying malicious account in different LBSNs. Nevertheless, it used only the dataset for evaluating the design of DeepScan. Though it is not performed well on the spatio-temporal dataset, because the temporal data is not fully open to the public. Hence, from the existing DeepScan, it is motivated that any new approaches can be designed for large-scale and reliable LBSN dataset to examine the behavioural patterns.

Information Security is of growing interest of policy makers as society become more dependent on secure communication. Andreson and Moore [1] in their research work have briefly explained about the security concern in economic perspective how these malicious content have impact the economic issue.

Detecting Clusters of Fake Accounts in Online Social Networks Fake accounts are a preferred means for malicious users of online social networks to send spam, commit fraud, or otherwise abuse the system. A single malicious actor may create dozens to thousands of fake accounts in order to scale their operation to reach the maximum number of legitimate members. Detecting and taking action on these accounts as quickly as possible is imperative in order to protect legitimate members and maintain the trustworthiness of the network. However, any individual fake account may appear to be legitimate on first inspection, for example by having areal-sounding name or a believable profile.

Mohammadreza Mohammadrezaei et al proposed [5] "Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms" It is a new model which is based on similarity between the users' friends' networks was proposed in order to discover fake accounts in social networks. Similarity measures such as common friends, cosine, Jaccard, L1-measure, and weight similarity were calculated from the adjacency matrix of the corresponding graph of the social network. To evaluate the proposed model, all steps were implemented on the Twitter dataset. It was found that the Medium Gaussian SVM algorithm predicts fake accounts with high area under the curve=1 and low false positive rate=0.02.

Deep Learning-Based Malicious Account Detection in the Momo Social Network research work proposed by Jiaqi Wang et al, It explore the malicious account detection problem by introducing a deep learning-based framework. By using the long short-term memory (LSTM) neural network, we are able to build a classifier to achieve the binary classification. By using the real data collected from Momo, a widely used LBSN which has more than 180 million users around the world, they evaluated their framework and the results show great promise for malicious account detection tasks, Figure 1 shows Architecture of malicious account detection in deep learning based malicious account.

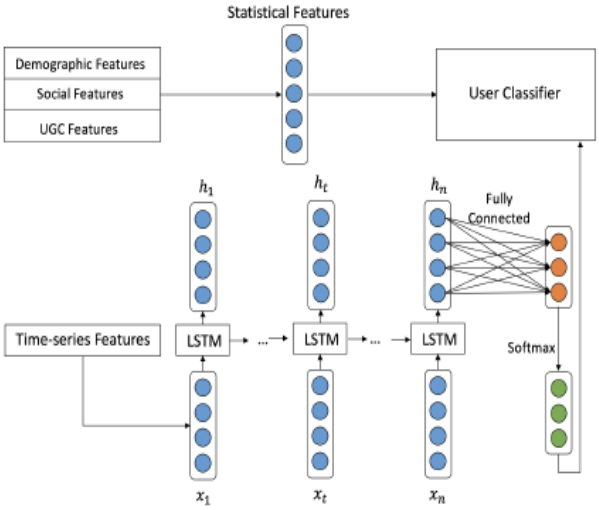


**Fig. 1 Architecture of malicious account detection system**

Beyond Blacklist: learning to detect malicious web sites from suspicious URL's. [6] Authors: - J. Ma, L. K. Saul, S. Savage, and G. M. Voelker. Description: In this paper we describe an approach to this problem based on automated URL classification, using statistical methods. The resulting classifier obtain 95-99% accuracy, detecting large number of malicious web sites from their URL's, with only modest false positive.

**Abnormality Identification**

Analyzing the text, image, audio, and video they are sharing through the social network. Contraction among the question and answer (comments, reply), current location and registered location.

## II. METHODS AND TECHNIQUES

### Text Analysis

An extensive keyword set based on the textual content and URL features to identify malicious content on Facebook at zero time. The intent is to catch malicious or vulgar content that is currently evading Facebook's detection mechanisms using text analysis. Steps involved in text analysis.

- Text Data Collection
- Corpus Processing
- Classification process using Similarity Index

Malicious activity can done by using image also but it can't detect by text analysis detection mechanism.

### Graph

The works based on graph mainly use the location and social relationship of spammers, and they need to build a huge social graph, which leads to much computing cost.

### Cookie-Based Information Cross Verification

A Meta data is generated for each user during user registration, user entry, and data processing, each Meta data provides the hidden information about the ISN, DNS, IP, URL, and URI. These information can be compared during critical investigation about the user as genuine or not.

## III. PROBLEM STATEMENT

People communicating and binding new friendship, business partnership, marketing their products and skill in social media. Social websites are one of the social media where unknown people can share their information, making friendship, etc. Comparing with the advantages, there are some disadvantages occur in the social network, where it spoils the next generation of human. To protect the culture, behavior, and attitudes it is necessary to detect and eliminate the abnormal people from the social network.

## IV. PROPOSED WORK

As online social networking sites raised in popularity, cyber-malefactor also began to utilize these sites to propagate malware and to carry out frauds, it is important to detect malicious activity in social networks. It is important to detect the malicious account in the social network; one of the recommend approaches to detect malicious account is a combination of dynamic user profile verification and cookie-based cross verifications.

Steps involved in Proposed Work:

- Corpus Processing
- User Profile Verification
- Cookie Based Cross Verifications
- Classification of legitimate and Malicious Account

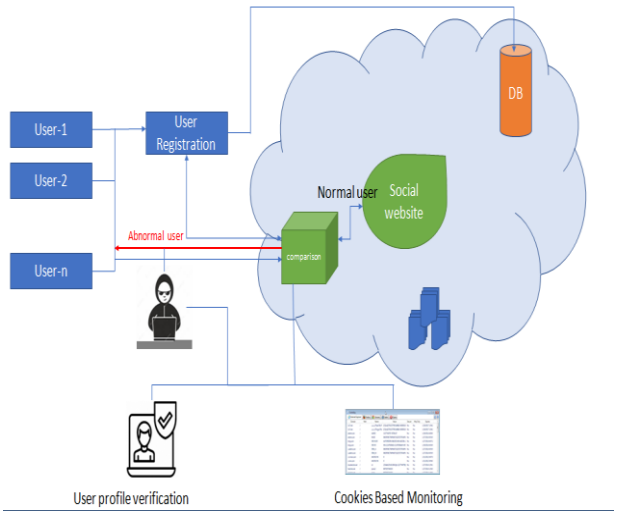


**Fig. 2 Work Flow**

## V. PERFORMANCE EVALUATION

The System Performance to detect malicious account detection in social networks is evaluated by using Random Forest machine learning algorithm.

- True positive is correctly identifying the malicious users as malicious users. (TP)
- False positive is incorrectly identifying the normal user as malicious user. (FP)
- True negative is correctly identifying normal user as the normal user. (TN)
- False negative is incorrectly identifying the abnormal user as normal user. (FN)

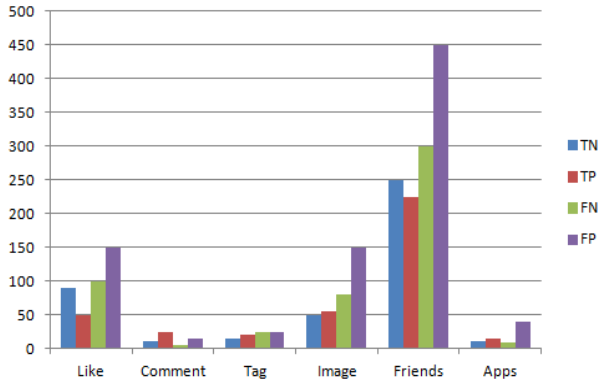


**Fig. 3 Attribute set Types**

## VI. CONCLUSION

In this paper, we investigate the malicious account detection problem in the online social network. By designing a machine learning-based system, we are able to model the statistical pattern and dynamic pattern of a user. Using the real data collected from online social networks, our system can achieve a good performance in malicious account detection. Note that our system is based on publicly-accessible information.

## REFERENCE


1. Brian K Tanner, Gary Warner, Henry Stern, and Scott Olechowski. Koobface: The evolution of the social botnet. In eCrime Researchers Summit (eCrime), 2010, pages 1–10. IEEE, 2010
2. Qingyuan Gong, Yang Chen, Xinlei He, Zhou Zhuang, Tianyi Wang, Hong Huang, Xin Wang, and Xiaoming Fu, (2018), "DeepScan: Exploiting Deep Learning for Malicious Account Detection in Location-Based Social Networks", IEEE Communications Magazine, DOI: 10.1109/MCOM.2018.1700575.
3. Daniele Quercia, MansourehBodaghi, and Jon Crowcroft. 2012. Loosing "friends" on facebook. In Proceedings of the ACM Web Science Conference (WebSci'12).
4. Mike Thelwall. 2008. Social networks, gender and friending: An analysis of MySpace member profiles. Journal of the American Society for Information Science and Technology 59, 8 (2008), 1321–1330.
5. Mohammadreza Mohammadrezaei et all(August 2018), "Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms", Security and Communication Networks Volume 2018, Article ID 5923156.
6. J. Ma, L. K. Saul, S. Savage, and G. M. Volker, ―Beyond blacklists: Learning to detect malicious Web sites from suspicious URLs,‖ in Proc. KDD, 2009, pp. 1245–1254.