ISSN No:-2456-2165

Spam Profile Detection on Facebook and Performance Evaluation of Machine Learning Algorithms

Muhammad Hashim Hameed, Usman Rasheed
Master of Sciences, Department of Computer Science,
University of Agriculture Faisalabad, Pakistan

Akmal Rehan
Lecturer, Department of Computer Science,
University of Agriculture Faisalabad, Pakistan

Abstract:- In recent time, social media have been affected many undesirable threats. Social media provided us an open platform to connect and share our life events with others. Social media also attracted the attentions of the spammers. Spam in social media relates to undesirable, malicious and spontaneous content, shown in different ways including malicious links, massages, fake friends and microblogs, etc. With the expanding of social networks such as Instagram, Facebook, MySpace, Twitter, and Sina Weibo, etc. spammers on them are getting increasingly rampant. Social spammers consistently make a mass of phony records to misdirect the users and lead them to malicious websites and illegal content. This research is highlight features for perceiving spammers on Facebook with the help of different classifiers. Also compare the performance of different Machine Learning Algorithms (MLA) like Support Vector Machine (SVM), Multilayer Perceptron (MLP), K Nearest Neighbor (KNN) and Random Forest (RF) on machine learning tools WEKA and Rapid Miner. We use the primary data collection technique to collect the user profile data of Facebook. Lebel the data "Spam" and "Not Spam" on the basis of Engagement Rate (ER), Duplication Profile Picture and Not Human Name. The outcomes of Support Vector Machine (SVM) from the experiments is better than other algorithms on both Machine Learning Tools (MLT) WEKA and RapidMiner. The results of all algorithms are better using WEKA as compare to RapidMiner. The results will be valuable for researchers who are eager to build machine learning models to recognize spamming exercises on social media networks.

Keywords:- malicious content, spammers, machine learning algorithms, Multilayer Perceptron, Random Forest, K-Nearest Neighbor, Support Vector Machine, RapidMiner, WEKA.

I. INTRODUCTION

In past, interpersonal organizations have rolled out an uncommon improvement in the public activity and it changed the web into "social web" where users and their networks are the habitats for online development, trade, and data sharing. Interpersonal organizations have a one of a kind worth chain which targets distinctive users. To locate an old friend, we used to social media network like Facebook, on the other hand, if it is finding to micro-blogging then we use Twitter. LinkedIn is a social network that is used to maintain and established professional resume and contacts or to discover contact by professional groups. Online Social Networks (OSNs) have developed as modest and effectively open internet-based life, encouraging users around the world for correspondence and data sharing. The users of these social networks are the key components in charge of the content being shared**.** Twitter is an online social media and microblogging, where user can send tweets (i.e., short instant messages restricted to 140 characters). As indicated by an ongoing report, Twitter is the quickest developing social network on the planet. The OSN is a couple of web sites that enable people to generate their personal content on the internet (WWW). The social internet platforms similar to web blogs, online forums, sociable media, and networks sharing sites provide a facility to do have cultural interactions and create public activities. The users in such systems create virtual communities also referred to as the social networks which let users share their comments, opinions, knowledge, and experiences with other users (Adikari and Dutta, 2014).

Facebook is one of the major OSNs. Common users, as well as big names like celebrities, politicians and other individuals, utilize web-based life to spread content and information to other people. Besides, organizations and associations examine social media websites the way of huge scale promotion and goal-oriented publicizing efforts. Facebook is the top social media network that has over 2.2 billion monthly active users and YouTube has 1.9 billion. On the other hand, Instagram and Twitter have 1000 million and 330 million active users respectively. Social media-based life has changed the way of communication and the way we live our lives. From the way we get our news to the manner in which we communicate with our friends and family. Social media life is all over, unavoidable and incredible. One of the thoughtful issues about information procurement in OSNs is the matter of phony user info or even altogether phony profiles. The purpose behind giving fake user information is generally a consequence of safety upgrade procedures because of incompatible safety design and info protection policies brought about by the stage. Though several Facebook users give intermediate phony info in their profiles, a few profiles that don't match with a user who exists in the world. On Facebook, the percentage of enlisted Facebook bogus records is 5% to 6%. Facebook plainly states in lawful terms that users are not tolerable to give false information. This visibly shows that the precision and rightness of Facebook user information is significant for Facebook's business model (Krombholz *et al.*, 2012).

A blacklist service that is provided by the Google, is known as Google Safe Browsing. It provides the lists of the URLs that contain phishing content or malware, for web resources. Lists provided by the Google Safe Browsing service are used by Safari, Google Chrome, GNOME Web, Vivaldi and Firefox, to check the pages for potential threats. A public API for this service is also provided by the Google. By sending alerts to operators via email, regarding threats that are being hosted on their networks, information is provided by Google to internet service providers. Almost over 3 billion devices, according to Google are protected by this particular service as of September 2017. Safe Browsing Lookup API is maintained by the Google with a privacy drawback. Firefox, Safari and Chrome browsers use the Safe Browsing Update API. A compulsory preferences cookie is also stored by the Safe Browsing on the computer. Client-side checks are conduscted by Google Safe Browsing. Websites that has no malware itself but carry the infected ads, might also be blacklisted by the Google Safe Browsing (Aggarwal *et al.*, 2012).

Anti-phishing site PhishTank was in October 2006 launched as a side-shoot by the entrepreneur named David Ulevitch. A phish verification system which is based on a community where different users submit the phish suspects and voting is done by the other users, that either it is a phish or not, is offered the company. Opera, Yahoo! Mail, Kaspersky, Mcafee, WOT, CMU, ST Benard, Firetrust, Mozilla, Officer Blue, Message Level, Finra, Sanesecurity for Clam AV, SURBL, Site Truth, Career Builder, PhishTank Sitechecker, C-Sirt and Avira use PhishTank. PhishTank data can be downloaded for free or it can be accessed through an API call, for commercial use, underneath a restraining license. It was announced in 2018, that with new functionality and features, the website will be rebuilt by the PhishTank (Saleh *et al.*, 2019).

II. LITERATURE REVIEW

Grier *et al.*, (2010) stated that email spam has a broad collection of research investigating how to recognize, characterize, and block spam. Normal methods to filter email spam contain IP and URL blacklisting with clarifying email content. Twitter has developed millions of followings in the last few years, so celebrities and politicians attract of following, spammers have quick attention on their operations to target Twitter with phishing and malware attacks. The guessing of weak password with brute force guessing technique is one of the notable attacks on Twitter.

Chu *et al.*, (2012) expressed that traditional spam strategies contain delivering spam emails and making spam web content. A previous couple of years have seen the fast rise of OSNs. One key component of such a framework is the dependence on the content contributed by people. Shockingly, the system or framework openness combined with huge user papulation has made OSNs a perfect objective of social spammers. By misusing social trust among users, social spam may make a huge success or progress with compare to traditional spam strategies.

Ahmed and Abulaish, (2012) articulated that there has been some study for the identification and avoidance of spam on OSNs. The researchers offered a real-time URL-spam discovery plot for Twitter. They logged browser action as a URL stack in the program and observed much information including redirects, domains in the time of constructing a page, pop-up messages and HTML content, HTTP headers and Java content to recognize spam links.

Aggarwal *et al.*, (2012) expressed that Phishing is an online erroneous procedure to get an individual's information of internet users. In 2011, it is determined that there were 520 million dollars lost worldwide in the aftereffect of phishing assaults. As a rule, email clients focused by these phishing assaults. With the sensational impact in omnipresence of (OSN) like Facebook, Twitter and Youtube, foes have started using these media to spread spam and phishing stunts. In 2010, 1% of the hard and fast Facebook customers have been setbacks of phishing attacks, which means 5 million Facebook customers.

Saini, (2013) explained that certain researchers focused on the advancement of honey pots to distinguish spams. To recognize spams, researchers managed the programmed gathering of misleading spam profiles in social media networks dependent on unknown behavior of user by utilizing social honey pots. This made rare user profiles with individual data like age, sex, date of birth and geographic location like locality and sent in MySpace (social network) people group. Spammer follows one of the systems and sending friend request for a long timeframe. The honey profile examines the spammer's action by allocating bots. When the spammers send a friend request, the bots store spammer's profile and slithers through the web pages to recognize the objective page where advertisements arise.

Cresci *et al.*, (2015) briefed that A few organizations had practical experience in social network analysis offer online administrations to evaluate how much a Twitter account is original with respect to its followers. However, the standards utilized for the investigation are not openly disclosed. One form of deception is considered as a fake account on Twitter, and deception in content like personal information. The Twitter accounts created and sold out to costumers, these accounts are we consider fake accounts and fake followers.

Zhang *et al.*, (2016) have performed a study on the impact of spammer recognition in traditional stages like email and web, some effort has been dedicated to identifying spammers in different social sites, for example, Twitter, Facebook, and Sina Weibo. A lot of researchers extract a wide range of user features, for example, profile feature, network, and content feature, etc. and afterward choose various classification algorithms to identify spammers.

Sedes, (2016) said that spammers are objective-oriented and beneficiary people expecting to accomplish unethical objectives, and in this way, they influence their knowledge to achieve the spam tasks in an active manner.

They launch bots of accounts in a small period to increase their profit, which is spam accounts. There are APIs provided by online social media, spammers utilize these APIs and create automated spamming tasks to get information in a systematic way.

Liu and Hu, (2017) said that spams in social communities are of distinct forms and they may change after some time, so a few standards must be made so as to filter them out. However, designing one by one and established these rules consuming more time and error-prone. With the help of machine learning techniques, we can construct classifiers that work consequently to expose spammers and prevent unwanted, malicious and spamming activities on social media websites for the safety of online social media users.

Egele *et al.*, (2017) performed a study that compromising interpersonal organization accounts has grown to be a profitable game-plan for cybercriminals now days. Through seizing (hijacking) control of a well-known social media or industrial account, aggressors were appropriate their malignant contents (messages) or phony data spread through an enormous client base. The effects of these occurrences go from a discolored notoriety to multi-billion-dollar money related misfortunes on financial markets. In this research area, it was introduced how it may use the similar strategies to discover these compromises of users with high profile accounts.

Sohrabi and Karimi, (2018) stated the increasing trend of social media, these systems have turned into a noteworthy instrument for criminal through sending spam. Numerous criminal operations, for example, taking significant data, selling malware, false purposeful publicity and different tasks are done by spammers. There are numerous spammers who divert their users to those pages that spammer wants and spread around various places. Because of the spam information, manual analysis is very difficult. The analyst chipped away at Defensio software on Facebook. This product qualifies the content of remarks that order the intelligent SVM, and furthermore qualifies analysts for revelation among its credits.

Hazim *et al.*, (2018) have performed a study on graph-based technique, the OSN is displayed as a system of nodes (users) and edges (associations). The associations among nodes are investigated so as to identify nodes with unusual attributes. This strategy has demonstrated appropriate for isolating spam profiles from those profiles which are original and authentic ones. For example, Markov clustering (MCL) algorithm applied on set of profiles to differentiate spam and non-spam.

Masood *et al.*, (2019) stated spam in the OSNs (online social networks) was a fundamental difficulty or issue which forces a danger to these services regarding the discouragement (undermining) their incentive (value) to publicists and potential financial specialists, just as harmfully influencing clients' commitment. The investigation analysis or examination incorporates more than 100 million messages gathered through Twitter within 1month. This investigation demonstrates that there were two typically particular classifications of spammers and that they utilize distinctive spamming procedures. At that point, this analysis represents how clients in these two classes exhibit distinctive individual properties just as social collaboration designs.

III. RESEARCH OBJECTIVE

- Identify Machine Learning Algorithms (MLA).
- Identify Machine Learning Tools (MLT).
- Collect user profile data of Facebook social media platform.
- Select features for labeling the data “Spam” or “Not Spam”.
- Convert the data file for Machine Leaning Tools (MLT).
- Get results of selected Machine Learning Algorithms (MLA) on Machine Learning Tools (MLT).
- Evaluate the performance of both Machine Learning Algorithms (MLA) and Machine Learning Tools (MLT).

IV. RESEARCH METHODOLOGY

A. Data Collection and Processing

To collect data, we use the primary data collection technique. Manually data collection technique is time-consuming, but this technique gives better results because the researcher collects those attributes which are required. In this technique, we visit one by user profile on the Facebook social media platform and collect public data. The tool, which is used for saving the data is Microsoft Excel. First of all, we visit the user profile and observe the number of friends because this attribute is very important data processing. Fig. IV.1 shows the view of a dataset in the Excel sheet.

	A	B	C	D	E	F	G	H	I
1	No. of Friends	Name	Profile Url	Gender	No. of Profile Pics	No. of Cover Photos	Profile Pic Likes	Profile Pic Comments	Profile Pic Url
2	3813	Umair Irshad	https://www.fa	M	75	60	111	17	https://scontent.
3	1555	Abubakar Jutt	https://www.fa	M	5	1	53	16	https://scontent.
4	412	Fari Jutt	https://www.fa	M	55	33	66	32	https://scontent-
5	501	Mariner Waheed	https://www.fa	M	5	1	81	21	https://scontent.
6	465	Usman Rasheed	https://www.fa	M	79	47	124	42	https://scontent.
7	195	Sana Rajpoot	https://www.fa	F	2	1	5	11	https://scontent.
8	438	Komal Khan	https://www.fa	M	3	0	3	3	https://scontent.
9	732	Hassan Mahmood	https://www.fa	M	16	7	44	4	https://scontent-
10	586	Periya Periya	https://www.fa	F	1	0	13	7	https://scontent.
11	709	Iqra Iqra	https://www.fa	F	1	0	20	9	https://scontent.
12	411	Hair Oil	https://www.fa	F	2	1	5	0	https://scontent.
13	62	Sajal Ali	https://www.fa	M	4	1	1	0	https://scontent.
14	1609	Kazim Rajpoot	https://www.fa	M	5	4	38	28	https://scontent-
15	4990	Khawar A Li	https://www.fa	M	10	8	30	0	https://scontent.
16	282	Akintade Romoke	https://www.fa	F	1	0	40	14	https://scontent.
17	253	Habab Cheema	https://www.fa	M	11	7	8	0	https://scontent.
18	155	Roya Roya	https://www.fa	F	1	0	5	1	https://scontent-
19	1485	Umer Hayat	https://www.fa	M	7	5	138	24	https://www.face
20	770	Mohammad Tayyab	https://www.fa	M	3	1	77	17	https://scontent-
21	123	Tariq Bashir	https://www.fa	M	1	0	45	56	https://scontent.

Fig. 1: Dataset in the Excel sheet

ISSN No:-2456-2165

The details of our dataset are as follows:

- **No. of Friends:** First of all we observe this attribute of the Facebook user profile, which is very important in the data labeling section.
- **No. of Profile Pics:** is also an important attribute of the Facebook user profile, which is used in the data labeling section. In this, we observe No. of profile pics in the photos section of the user profile.
- **Profile Pic Likes:** is a third important attribute that is used in the data labeling section. In this, we observe user profile pic likes.
- **Profile Pic Comments:** is forth important attribute that is used in the data labeling section.
- **Profile Pic Address:** is also collected from the user profile.
- **Name:** is one of the attributes of the user profile, which are observed and save in the Excel sheet and used in Data Labeling.
- **Profile Url:** is also collected from the user profile.
- **Gender:** is observe from the user profile.
- **No. of Cover Photos:** is an attribute that is observed in the Photos section of the user profile.

B. Data Labeling

Data Labeling technique used to label the data that the user profile is a spam profile or not a spam profile. There are three ways adopted for data labeling which are Engagement Rate (ER), Profile Pic Duplication Check on TinEye and Not Human Name. These three ways are explained below.

Sr.	Features
1	Engagement Rate (ER)
2	Profile Pic Duplication Check on TinEye
3	Not Human Name

Table **Error! No text of specified style in document.**1: Features

a) *Engagement Rate (ER)*

An Engagement Rate (ER) is a metric that estimates the degree of commitment that a bit of made substance is accepting from a crowd of people. It shows how much individuals communicate with the substance. Elements that impact engagement that is user' comments, share, likes, No. of friends and more. This is a significant measurement to watch out for on the grounds that higher buyer engagement is an indication of extraordinary substance. In this Engagement Rate (ER) four attributes of the user profile are used, these four attributes are No. of Friends, No. of Profile Pics, Profile Pic Likes and Profile Pic Comments. The Engagement Rate (ER) metric is given below.

$$ER = \frac{((\text{Profile Pic Likes} + \text{Profile Pic Comments}) * \text{No. of Profile Pics})}{\text{No. of Friends}} * 100$$

Calculate the Engagement Rate (ER) of each user profile and save it in the Microsoft Excel sheet. For labeling the data for the profile spam or not spam, the minimum value of the Engagement Rate (ER) is set which is 0.01%. If the value of Engagement Rate (ER) is less than 0.01% then the user profile label with "Spam", on the other hand, the user profile label with "Not Spam" if the value of Engagement Rate (ER) is higher than 0.01%.

b) Profile Pic Duplication Check

TinEye is the main site to ever utilize for picture identification innovation and to this date is as yet one of the most well-known and broadly utilized reverse search engines. It's extraordinary for proficient picture takers or creatives who have worked on the web and need to check whether any of it has been taken or altered and reused. At the time, TinEye flaunted 38 billion indexed pictures.

c) Not Human Name

Facebook is quite clear on what considers a genuine name. There are likewise several different rules that your name must avoid some words, which are given below.

- Images, numbers, strange capitalization, and such things are not included in your name.
- A mixture of characters from various languages is not included in your name.
- A title like Doctor or Father is also not included in your name.
- Words that aren't your name; for example, I was unable to have "Magnificent Harry Guinness" as mine, regardless of the amount I needed it.
- Offensive words are not included in your name.

V. RESULTS AND DISCUSSIONS

The outcomes of Machine Learning Algorithms (MLA) is obtained from Machine Learning Tools (MLT) WEKA and RapidMiner.The following table shows the results comparison of SVM, MLP, KNN and RF in WEKA.

Algorithms	Accuracy %	Error Rate %
SVM	99.331	0.668
MLP	99.665	0.334
KNN	100	0.000
RF	98.996	1.003

Table 2: Results Comparison in WEKA

When we compare the results of all these algorithms, the algorithm K Nearest Neighbor (KNN) gives us the highest accuracy which is 100 % While the algorithm Random Forest (RF) gives us the lowest accuracy which is 98.99% . In this comparison the algorithm Multilayer Perceptron (MLP) is on the 2nd position which gives us 99.66 % accuracy and Support Vector Machine (SVM) is on the 3rd postion which gives us 99.33 % accuracy.

ISSN No:-2456-2165

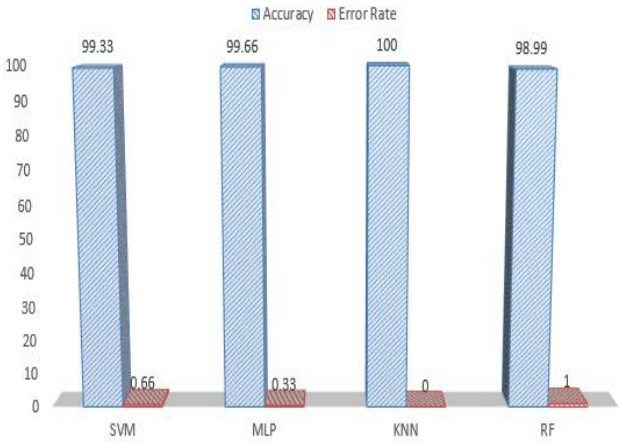

Fig. 2: Results Comparison in WEKA

The following table shows the results comparison of SVM, MLP, KNN AND RF in RapidMiner.

Algorithms	Accuracy %	Error Rate %
SVM	100	0
MLP	86.62	13.38
KNN	86.62	13.38
RF	100	0

Table 3: Results Comparison in RapidMiner

When we compare the results of all these algorithms, the algorithm Support Vector Machine (SVM) and Random Forest (RF) gives us brilliant results with100 % accuracy. While the algorithm Multilayer Perceptron (MLP) and K Nearest Neighbor (KNN) gives us the same results with 86.62 % accuracy and 13.38 % error rate.

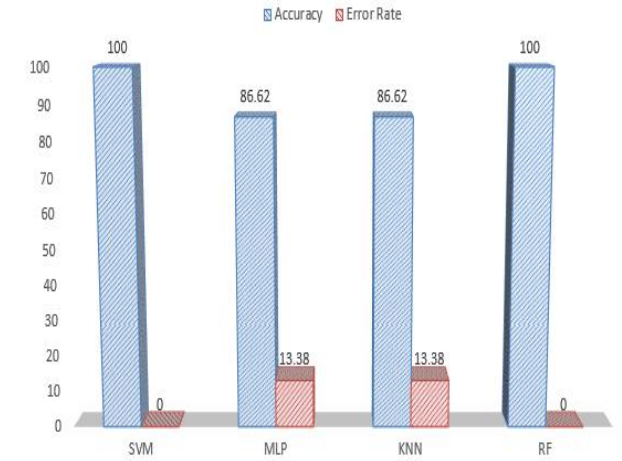

Fig. 3: Results Comparison in RapidMiner

The following table shows the accuracy comparison of SVM, MLP, KNN AND RF in WEKA and RapidMiner.

Algorithms	WEKA	Rapid Miner
SVM	99.331	100
MLP	99.665	86.62
KNN	100	86.62
RF	98.996	100

Table 4: Accuracy Comparison between WEKA and RapidMiner

In this comparison, Support Vector Machine(SVM) and Random Forest (RF) algorithms gives us the best results with minimum difference But there is a big difference in the case of other algorithms like Multilayer Perceptron (MLP) and K Nearest Neighbor (KNN).When we apply Multilayer Perceptron (MLP) in WEKA and RapidMiner the results are 99.66 % accuracy and 86.62 % accuracy. Similarly, when we apply K Nearest Neighbor (KNN) in WEKA and RapidMiner the results are 100 % accuracy and 86.62 % accuracy.

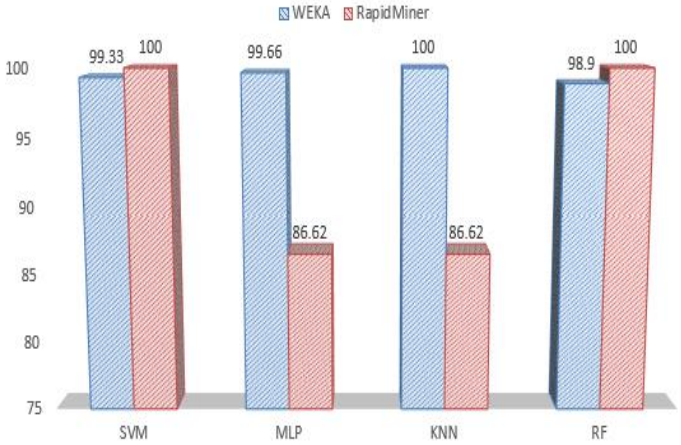

Fig. 4: Accuracy Comparison between WEKA and Rapid Miner

Support Vector Machine (SVM) gave results from the experiments is better than other algorithms on both Machine Learning Tools (MLT) WEKA and RapidMiner. The results of all algorithms are better using WEKA as compared to RapidMiner.

REFERENCES

[1.] Adikari, S. and K. Dutta. 2014. Identifying fake profiles in Linkedin. Pacific Asia Conference on Information Systems. 278:1.

[2.] Aggarwal, A., A. Rajadesingan and P. Kumaraguru. 2012. PhishAri : Automatic realtime phishing detection on Twitter. eCrime Researchers Summit, Las Croabas. 1: 1–12.

[3.] Ahmed, F. and M. Abulaish. 2012. An MCL-Based Approach for spam profile detection in online social networks. IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications.: 602–608.

[4.] Chu, Z., I. Widjaja, H. Wang and B. Laboratories. 2012. Detecting Social Spam Campaigns on Twitter. International Conference on Applied Cryptography and Network Security. 7341: 455–472.

[5.] Cresci, S., R. Di Pietro, M. Petrocchi, A. Spognardi and M. Tesconi. 2015. Fame for sale: Efficient detection of fake Twitter followers. Decis. Support Syst. 80:56–71.

[6.] Egele, M., G. Stringhini, C. Kruegel and G. Vigna. 2017. on Social Networks. 14:447–460.

[7.] Grier, C., K. Thomas, V. Paxson and M. Zhang. 2010. @ spam : The underground on 140 characters or less.

17th ACM conference on Computer and communications security.: 27-37.

[8.] Hazim, M., N.B. Anuar and A. Kamsin. 2018. Performance evaluation of machine learning algorithms for spam profile detection on Twitter using Weka and Rapidminer. Advanced Science Letters. 24(2):1043–1046.

[9.] Krombholz, K., D. Merkl and E. Weippl. 2012. Fake identities in social media : A case study on the sustainability of the Facebook Business Model. Journal of Service Science Research. 4(2):175–212.

[10.] Liu, N. and X. Hu. 2017. Spam detection on social networks. Encyclopedia of Social Network Analysis and Mining.: 1–9.

[11.] Masood, F., G. Ammad, A. Almogren, A. Abbas, H.A. Khattak, I. Ud Din, M. Guizani and M. Zuair. 2019. Spammer Detection and Fake User Identification on Social Networks. IEEE Access. 7:68140–68152.

[12.] Saini, J. S. 2013. A study of spam detection algorithm on social media networks. Computational Intelligence, Cyber Security and Computational Models. 246:195–202.

[13.] Saleh, A.J., A. Karim, B. Shanmugam, S. Azam, K. Kannoorpatti, M. Jonkman and F. De Boer. 2019. An intelligent spam detection model based on artificial immune system. Inf. 10.

[14.] Sedes, F. 2016. Leveraging time for spammers detection on Twitter. Proceedings of the 8th International Conference on Management of Digital Ecosystems. 8: 109–116.

[15.] Sohrabi, M.K. and F. Karimi. 2018. A Feature Selection Approach to Detect Spam in the Facebook Social Network. Arab. J. Sci. Eng. 43:949–958.

[16.] Zhang, X., H. Bai and W.L. B. 2016. A social spam detection framework via semi-supervised learning. Pacific-Asia Conference on Knowledge Discovery and Data Mining. 9794:214–226.