ISSN 2278-3091
Volume 10, No.3, May - June 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse561032021.pdf
https://doi.org/10.30534/ijatcse/2021/561032021

Execution Assessment of Machine Learning Algorithms for Spam Profile Detection on Instagram

Usman Rasheed[1], Akmal Rehan[2], Salman Afsar[3], Ahmed Mateen[4], Tayyaba Raza[5], Ayesha Khalid[6], Hira Naeem[7], Javeria Jameel[8]
[1,2,3,4,5,6,7]Department of Computer Science, University of Agriculture Faisalabad, Pakistan, 38000.
[8]National Textile University, Faisalabad
Corresponding Author: ahmedmatin@hotmail.com

ABSTRACT

With every passing second social network community is growing rapidly, because of that, attackers have shown keen interest in these kinds of platforms and want to distribute mischievous contents on these platforms. With the focus on introducing new set of characteristics and features for counteractive measures, a great deal of studies has researched the possibility of lessening the malicious activities on social media networks. This research was to highlight features for identifying spammers on Instagram and additional features were presented to improve the performance of different machine learning algorithms. Performance of different machine learning algorithms namely, Multilayer Perceptron (MLP), Random Forest (RF), K- Nearest Neighbor (KNN) and Support Vector Machine (SVM) were evaluated on machine learning tools named, RapidMiner and WEKA. The results from this research tells us that Random Forest (RF) outperformed all other selected machine learning algorithms on both selected machine learning tools. Overall Random Forest (RF) provided best results on RapidMiner. These results are useful for the researchers who are keen to build machine learning models to find out the spamming activities on social network communities.

Key words: Support-vector machines (SVM), random forest (RF), K-nearest neighbor (KNN), rapid miner, weka, spam

1. INTRODUCTION

Online social networking platforms have appeared as easily accessible, cheap and effective social media that facilitates worldwide users for information sharing and communication. Although the very basic goal of social networking sites is online communication and interaction, nevertheless the patterns of usage and specific goals vary on different services. In the recent years, social networking platforms like Facebook, Instagram and Twitter, have become worldwide sensation and one of the quickest emerging e-services, as stated by [1]. Instagram, the photo sharing app is ranked 6th with over 1 billion active monthly accounts. Key elements that is the base for the components being shared on social mediaare its users [2]. Users are usually recognized by a profile at these social media sites. However, these social media sites do not strictly identify that the one creating the profile and using it, is actually the same personas stated in the profile. Someone else is using somebody else's identity, if that is not the case, this is known as false identity. One can easily create profile with fabricated names and other information that is not associated with any person living in any part of the world. In these kinds of cases the identity is known as faked identity. There are number of mischievous activities executed by the spam profiles, which consists of phishing, following great number of users with little followers, overflowing the social media platforms with fake profiles, random link connection, spreading malware and endanger existing valid accounts etc. In spite of the benefits got from the social relationships, user's profile has turnedout to be one of the focused resource by the spammer's, who among user's, influence the trust relationship to acquire more unfortunate victims [3] . The purposeof this research is to find the different features among the Instagram user's profile data that will help to detect and identify the spam profiles on Instagram. The performance of machine learning algorithms will be checked on machine learning tools, to find which of the machine learning algorithm performs better than the other machine learning algorithms. A new concept of spammers is arising by use of these social media platforms. By using machine learning algorithms many useful models are predicted by the researchers to help detect these spammer's profiles. The assessment of these illegitimate profiles is essential to avoid the valid users to be caught up in their plans [4]. To save the privacy of user's profiles, the researchers set their attention towards identifying the spammers. Day after day unsolicited profiles are increasing that are destroying the rights of the actual user's and increasingtheir focused objective. The aim of detecting the spammers is forcing the researchers to apply the appropriate techniques by using statistical approaches, sentiment analysis, machine learning and deep learning [5] . The objectiveof this study was to find the different features among theuser's profile data that will help detect and identify the spam profiles on Instagram. The performance of machine learning algorithms was also checked on machine learning tools, to findwhich of the machine learning algorithm performed better than the other machine learning algorithms. The accuracy level of the machine learning tools was also measured.

2. RELATED WORK

Reference [6] researched how straightforward it would be for a likely attacker to dispatch the automated crawlingand the fraud ambushes against various well-known social communication websites to access an enormous volume of individual client data. Reference [7] utilized blacklist-based method to decrease spamming activities and to identify spammers on Twitter. Reference [8] examined that to what degree spam has entered the social media. They established the methods to find spammers in social networks and gathered their messages in huge spam operations on the analysis of this behavior. By providing the results of their analysis to Twitter, thousands of spamming accounts were shut down. Reference [2] offered an approach based on Markov Clustering (MCL) for the recognition of spam accounts on OSNs. Reference [9] observed that unluckily, spammers affect the people by posting the spam content. Reference [10] introduced a conventional statistical approach to recognize spam profiles on Online Social Networks. Reference [11] stated that initially, certain researchers paid attention on the improvement of honey pots to distinguish spams. Reference [12] proposed technique, which was analogous to the results, gets by other existing techniques, detecting the fake profile with an accuracy of 84% and 2.44% false negative. Reference [13] exam- ined and differentiated existing detection procedures of Sybil accounts, and a TSD system presented that detect twitter Sybil accounts and notify the users before retrieving or ensuing these accounts by dynamically utilizing the supervised machine learning methods. Reference [14] highlighted how similar methods were used to recognize are composed of specific high-profile accounts. Sai Sundara [15] applied different approaches to recognize or distinguish spams in the social network by removing essential material from web pages. [16] Lee and Kim suggested an original detection patternto filter possible malevolent account groups around the timeof their formation. Reference [4] proposed to construct social activities profiles for particular OSN users that explains their interactive designs. Reference [5] presented a model for spammer identification based on machine learning, forsocial network (Twitter). The author in [17] utilized Random Forest algorithm to detect the spammers. They proposed new time- based features and advanced the design of some existing features that were used before. [18]Zhang and Sun proposed a system relating feature-based technique and supervised learn- ing method to identify spam posts from Instagram. Reference [19] proposed the solutions to reduce the influenceof spammers by utilizing Markov Clustering algorithm to identify the spam profile. The author in [20] focused mainly on debating and estimating machine learning techniques for spam SMS identification. As expected, among conventional classifiers, NB and SVM display great outcomes, extremely near CNN for both datasets. Substantial outcomes have been acquired from this work, relating to which this research canbe taken to genuine application level for the identification of spam SMS. Reference [3] introduced other aspects to increase the classifier performance and proposed the group of attributes for classifying the spammers on Twitter. Reference [21] studied several present methods on spam profile identification in online social platforms. The author in [22] proposed work which introduced the systemthat is capable to capture the spammer's common behavior, and the users who are habitual to share the malicious URLs and patterns presence in spammer's tweets [23]. Aggarwal et al. presented an algorithm in which detailed examination of user's account was performed based on his actions and specific profile statistics, thus based upon the facts of the profile categorizing these accounts.

3. METHODOLOGY

For research purpose, data of social platform Instagram is collected manually by visiting this platform and carefully observing and recording the selected features of the Instagram. Initial dataset contained the values of limited attributes ofInstagram profiles that were publicly available and can be observed clearly by opening any profile on Instagram. The first created dataset contained the attributes Profile Pic, Name, Username, Followers, Following, Posts, Verified, Description, Hashtags, Number of hashtags, Mentions, Number of mentions, External URL, URL, Private, Highlights, Number of highlights and Tagged. The initial dataset can be seen in the figure 1.

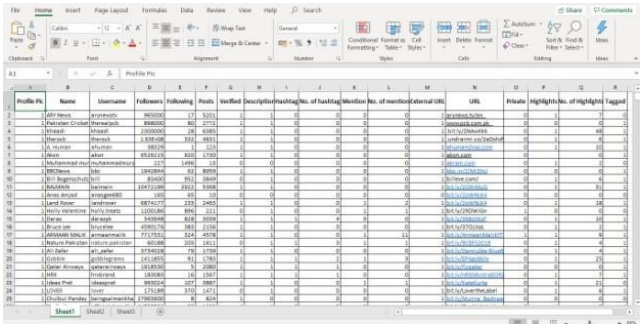


Figure 1: Initial Dataset

A. Creating new features

When the basic dataset was collected, the collected features were used to change the type of current features and to create new features. For example, the name and username of profile was already recorded in the dataset, but we changed it into binary form i.e., in the form of 0 and 1, where 0 represents the absence of name or username and 1 represents the presence of name and username. Likewise, presence of numeric characters in name and username was extracted from the name and username features, also their lengths were recorded.

Furthermore, the URLs that were collected originally were checked from two sites, namely PhishTank and Sitechecker. Phishtank was used to check whether the collected URL is phishing or not and Sitechecker uses Google Safe Browsing and checks the URLs according to its rules. Two new columnswere created that recorded the results from these two sites. The results from PhishTank site were recorded in

Phishing column and the results from Sitechecker were recorded in Blocked column.

In the figure 2, we can see how the PhishTank gives its results. The figure tells us that the URL that was checked from the site was phishing.

While in the figure 3, we can see the results from the Sitechecker. The figure above tells us that the URL checked is found as safe from the Google Safe Browsing.

New dataset that was formed is shown in the figure 2.

Figure 2: Phish Tank Result

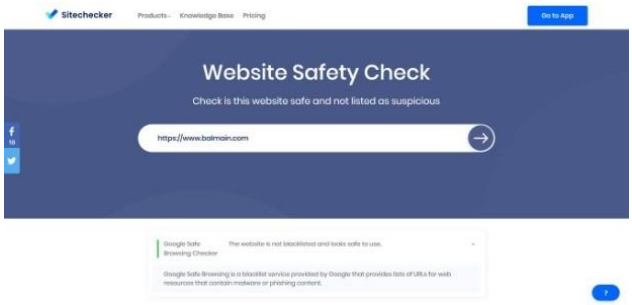


Figure 3: Sitechecker Result

B. Metrics

Three metrics were created after carefully observing the users behavior and used in this research and the results from these metrics were further used to label our target variable as spam and non-spam as shown in Figure 4. Formulas were applied in excel sheet in which the data was recorded in the first place and also the new features that were created were also present in the same file. Results of all three metrics were recorded in binary form i.e., in the form of 0,1. All of these metrics and the features that were used by these metrics are discussed below.

In M1 or metric one, 2 features were used that are Phishing and Blocked. In this metric these 2 were checked and if either of the feature was 1, then its result was recorded as 1.

M1 = IF(OR(Phishing=1,Blocked=1), “1”, “0”)

In M2 or metric two, 2 features were used that are Num Name and Verified. This metric checks that if the Num Name is 1, that means the Name of the user contains a numeric character in it and Verified is 0, which indicates that the user account is not verified by the Instagram, then its result was recorded as 1.

M2 = IF(AND(Num Name=1, Verified=0), “1”, “0”)

In M3 or metric three, 4 features were used that are Profile Pic, Posts, Description and Verified. This metric checks that if

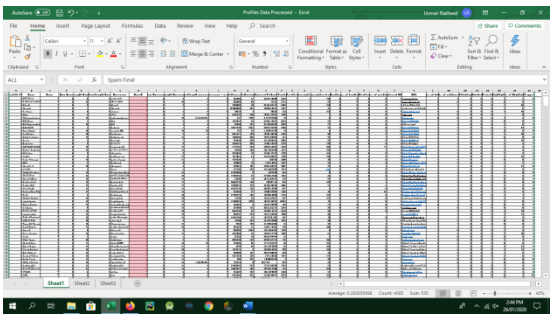
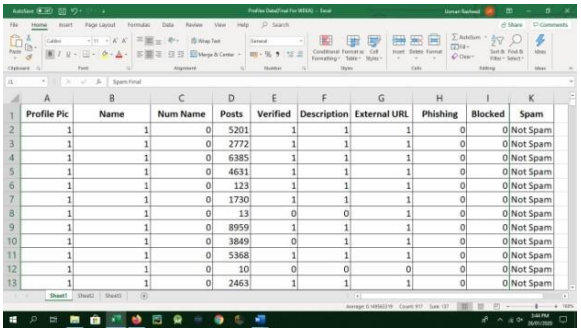


Figure 4: Processed Dataset

all of these 4 features are 0, then its result was recorded as 1. This means that there is no profile picture, no short description is given in profile, no post is shared, and the account is alsonot verified by Instagram.

M3 = IF(AND(Profile Pic=0,Post=0,Description=0, Verified=0), “1”, “0”)

Target variable was created by using the results of the three metrics that are discussed before. If the result of any metric is recorded as 1, then it means this profile is behaving differently from the legitimate users. So, we applied the formula in the excel sheet which uses the results from these three metrics and if any of them was recorded as 1 then it labelled the target variable Spam as spam and if all of them were recorded as 0 then it labelled the target variable as not spam. All of the extra features that were not used in this research were removed from the file and a clean dataset that carries the target variable was produced. The final data file is shown in figure 5.



	A	B	C	D	E	F	G	H	I	K
1	Profile Pic	Name	Num Name	Posts	Verified	Description	External URL	Phishing	Blocked	Spam
2	1	1	0	5201	1	1	1	0	0	Not Spam
3	1	1	0	2772	1	1	1	0	0	Not Spam
4	1	1	0	6385	1	1	1	0	0	Not Spam
5	1	1	0	4631	1	1	1	0	0	Not Spam
6	1	1	0	123	1	1	1	0	0	Not Spam
7	1	1	0	1730	1	1	1	0	0	Not Spam
8	1	1	0	13	0	0	1	0	0	Not Spam
9	1	1	0	8959	1	1	1	0	0	Not Spam
10	1	1	0	3849	0	1	1	0	0	Not Spam
11	1	1	0	5368	1	1	1	0	0	Not Spam
12	1	1	0	10	0	0	0	0	0	Not Spam
13	1	1	0	2463	1	1	1	0	0	Not Spam

Figure 5: Final Dataset

C. Support-Vector Machines (SVM)

1) Applying SVM in WEKA: After importing the data in WEKA, I applied SVM on dataset by using SMO classifier.

2) Applying SVM in RapidMiner: Before executing SVM in RapidMiner, the model port of SVM was connected withthe model port of Apply model and original port of Select attributes operator with the unlabeled data

port of the Apply model. At the end, performance port of the Performance operator was connected with the result port.

D. Multilayer Perceptron (MLP)

3) Applying MLP in WEKA: After importing the data in WEKA, MLP was applied on dataset by choosing Multilayer- Perceptron which implemented the classifier.

4) Applying MLP in RapidMiner: MLP was executed in RapidMiner by connecting the performance port of the Performance operator with the result port.

E. Random Forest (RF)

5) Applying RF in WEKA: After importing the data in WEKA, RF was applied by choosing RandomForest which implemented the classifier.

6) Applying RF in RapidMiner: To apply RF in Rapid-Miner, the performance port of the Performance operator was connected with the result port.

F. K-Nearest Neighbor (KNN)

7) Applying KNN in WEKA: After importing the data in WEKA, from the lazy list that is present in classifier, IBk was selected which implemented the k-nearest neighbors classifier.

8) Applying KNN in RapidMiner: To apply KNN in Rapid- Miner, the performance port of the Performance operator was connected with the result port.

4. RESULT

This study uses four algorithms on two different tools to test the proposed model and check the accuracy of these algorithms. First the result of each algorithm is discussed one by one, and then at the end comparative analysis is made basedon gathered results.

A. User Profiles

In the final processed dataset, a total of 9 features are present and one target variable. Data of 916 users profiles from Instagram is recorded manually. Among these profiles there are 779 Non-Spam profiles and 137 Spam profiles labelled after applying the three metrics.

B. Results in WEKA

1) KNN in WEKA: In the previous research, the KNN provided an accuracy of 95.12% with an error rate of 4.88% inWEKA. Here 10 folds cross validation is done. It achieved an accuracy of 99.78% with an error rate of 0.22%. It correctly classified all the non-spam data while it classified 2 instances of spam data incorrectly [3].

2) SVM in WEKA: In the previous research, the SVM provided an accuracy of 95.18% with an error rate of 4.82% in WEKA. Here 10 folds cross validation is done. It achieved an accuracy of 98.8% with an error rate of 1.2%. It incorrectlyclassified 11 instances from the total data [3].

3) MLP in WEKA: In the previous research the KNN provided with an accuracy of 95.42% with an error rate of 4.58% in WEKA. Here 10 folds cross validation is done. It achieved an accuracy of 99.45% with an error rate of 0.55%. It incorrectly classified 5 instances from the total data [3].

4) RF in WEKA: In the previous research the RF provided with an accuracy of 94.51% with an error rate of 5.49% in WEKA. Here 10 folds cross validation is done. It achieved an accuracy of 99.89% with an error rate of 0.11%. It correctly classified all the non-spam data while it classified 1 instanceof spam data incorrectly [3].

5) Comparison of results from WEKA: From the gathered results of four selected algorithms that we used in WEKA, Random Forest(RF) gave the best accuracy and after that KNN. MLP gave the third best results and SVM at the last with lowest accuracy among the four.

Comparative results including accuracy and error rate of all four algorithms in WEKA are shown in the form of chart in the figure 6 and in table1.

Table 1: Comparison of results in WEKA

Algorithm	Accuracy	Error Rate
KNN	98.78	0.22
MLP	99.45	0.55
SVM	98.8	1.2
RF	99.89	0.11

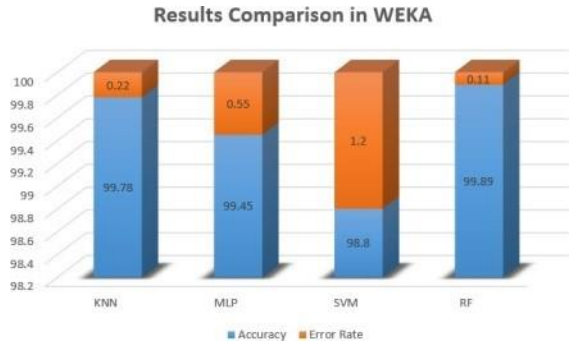


Figure 6: Comparison of results in WEKA

C. Results in RapidMiner

Detail comparison of these Algorithmic techniques are explained below [3].

i. *KNN in RapidMiner:* In the previous research, the KNN provided with an accuracy of 93.31% with an error rate of 6.69% in RapidMiner. It achieved an accuracy of 98.47% withan error rate of 1.53%.

ii. *SVM in RapidMiner:* In the previous research, the SVM provided with an accuracy of 89.35% with an error rate of 10.65% in RapidMiner. It achieved an accuracy of 98.91% with an error rate of 1.09%.

iii. MLP in RapidMiner: In the previous research, the MLP provided with an accuracy of 95.34% with an error rate of 4.66% in RapidMiner. It achieved an accuracy of 97.38% withan error rate of 2.62% .

iv. RF in RapidMiner: In the previous research, the RF provided with an accuracy of 95.44% with an error rate of 4.56% in RapidMiner. It achieved an accuracy of 100% .

v. Comparison of results from RapidMiner: From the gathered results of four selected algorithms that we used in RapidMiner, Random Forest (RF) gave the best accuracy and after that SVM. KNN gave the third best results and MLP at the last with lowest accuracy among the four as shown in Table 2 and Table 3.

Table 2: Comparison of results in RapidMiner

Algorithm	Accuracy	Error Rate
KNN	98.47	1.53
MLP	97.38	2.62
SVM	98.91	1.09
RF	100	0

Comparative results including accuracy and error rate of all four algorithms in RapidMiner are shown in the form of chart in the figure 7 and Figure 8.

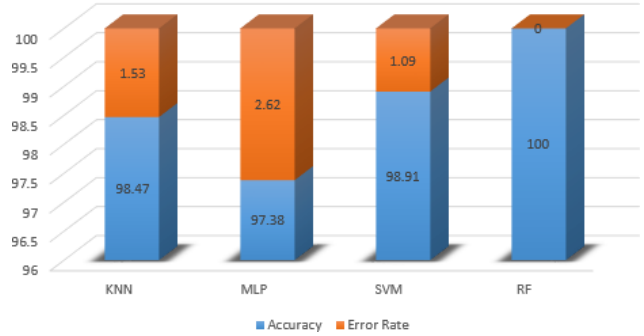


Figure 7: Comparison of results in RapidMiner

Table 3: Comparison WEKA vs RapidMiner

Algorithm	WEKA	RapidMiner
KNN	98.78	98.47
MLP	99.45	97.38
SVM	98.8	98.91
RF	99.89	100

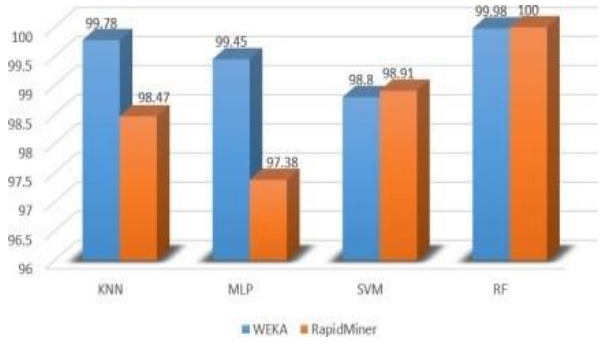


Figure 8: Accuracy comparison of WEKA vs RapidMiner

5. Conclusion

Mostly WEKA provided better results. RF gave best results in both tools. RF was on top with 99.89% accuracy in WEKA and 100% accuracy in Rapid Miner. Overall RF gave the best results in RapidMiner with 100% accuracy. KNN gave second best results in WEKA and SVM gave second best results in RapidMiner. MLP gave the third best results in WEKA and KNN gave the third best results in RapidMiner. Least good results are given by SVM in WEKA and MLP in RapidMiner. Overall RF proved to be the best amongst the four selected Machine Learning algorithms in both selected Machine Learning tools. In figure 8 accuracy of all four selected Machine Learning algorithms in both of the selected Machine Learning tools i.e., WEKA and RapidMiner, is shown in the form of graph. We can see clearly that RF proved to be the best amongst the four selected algorithms.

REFERENCES

1. N.B. Elliso. **Social Network Sites : Definition, History, and Scholarship.** Journal of Computer-Mediated Communication. Vol. 13, pp. 210-230, 2008.
2. F. Ahmed and M. Abulaish. **An MCL-based approach for spam profile detection in online social networks.** Proceedings of the 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom-2012 - 11th IEEE International Conference on Ubiquitous Computing and Communications, IUCC-2012. pp. 602-608, 2012.
3. M.H.M, Hanif, K.S. Adewole, N.B. Anuar and A. Kamsin. **Performance Evaluation of Machine Learning Algorithms for Spam Profile Detection on Twitter Using WEKA and RapidMiner.** Advanced Science Letters. Vol. 24, pp. 1043-1046, 2018.
4. X. Ruan, Z. Wu, H. Wang and S. Jajodia. **Profiling Online Social Behaviors for Compromised Account Detection.** IEEE Transactions on Information Forensics and Security. Vol. 11, pp. 176-187, 2016.
5. G. Khurana. **Efficient Spam Detection on Social Network.** International Journal for Research in Applied Science & Engineering Technology (IJRASET). Vol. 4, pp. 43-50, 2018.
6. L. Bilge, T. Strufe, D. Balzarotti and E. Kirda. **All your contacts are belong to us: Automated identity theft attacks on social networks.** WWW'09 - Proceedings of the 18th International World Wide Web Conference. pp. 551-560, 2009.
7. G. Chris, K. Thomas, P. Vern, M.Z. **@spam: The Underground on 140 Characters or Less.** Proceedings of the ACM Conference on Computer and Communications Security, pp. 1-11, 2010.
8. G. Stringhini, C. Kruegel and G. Vigna. **Detecting spammers on social networks.** Proceedings - Annual Computer Security Applications Conference, ACSAC. pp. 1-9, 2010.

9. Z. Chu, I. Widjaja and H. Wang. **Detecting social spam campaigns on Twitter.** Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 7341 LNCS, pp. 455-472, 2012.
10. F. Ahmed and M. Abulaish. **A generic statistical approach for spam detection in Online Social Networks.** Computer Communications. Vol. 36 pp. 1120-1129, 2013.
11. J.S. Saini. **A Study of Spam Detection Algorithm on Social Media Networks.** Springer. pp. 195-202.
12. S. Adikari and K. Dutta. **Identifying fake profiles in linkedin.** Proceedings - Pacific Asia Conference on Information Systems, PACIS, pp. 1-16, 2014.
13. M., A. Alsaleh, Al-Salman, M. Alfayez and A. Almuhaysin. **TSD: Detecting sybil accounts in twitter.** Proceedings - 2014 13th International Conference on Machine Learning and Applications, ICMLA, pp. 463-469, 2014
14. E. Mumford. **Social Networks.** IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING. Vol. 14, pp. 447-460.
15. S.S. Krishnan, G., R. Anitha, R.S. Lekshmi, M. Senthil Kumar, A. Bonato and M. Graña. **Computational intelligence, cyber security and computational models:** Proceedings of ICC3, Advances in Intelligent Systems and Computing. Vol. 246, pp. 195-202, 2014.
16. S. Lee and J. Kim. **Early filtering of ephemeral malicious accounts on Twitter.** Computer Communications. Vol. 54, pp. 48-57, 2014.
17. F. Sedes. **Leveraging Time for Spammers Detection on Twitter.** 8th International Conference on Management of Digital EcoSystems (MEDES), pp. 109-116, 2016.
18. W. Zhang and H.M. Sun. **Instagram spam detection.** Proceedings of IEEE Pacific Rim International Symposium on Dependable Computing, PRDC. pp. 227-228, 2017.
19. E.I. Setiawan, C.P. Susanto, J. Santoso, S. Sumpeno and M.H. Purnomo. **Preliminary study of spam profile detection for social media using Markov clustering: Case study on Javanese people.** 20th International Computer Science and Engineering Conference: Smart Ubiquitos Computing and Knowledge, ICSEC, pp. 1-4, 2017.
20. M. Gupta, A. Bakliwal, S. Agarwal and P. Mehndiratta. **A Comparative Study of Spam SMS Detection Using Machine Learning Classifiers.** 11th International Conference on Contemporary Computing, IC3, pp. 1-7, 2018.
21. R. Krithiga, and E. Ilavarasan. **A survey of recent techniques for detection of spam profiles in online social networks.** Journal of Advanced Research in Dynamical and Control Systems. Vol. 11, pp. 1256-1262.
22. F. Concone, G. Lo Re, M. Morana and C. Ruocco. **Twitter spam account detection by effective labeling.** CEUR Workshop Proceedings, pp. 1-22, 2019.
23. A. Aggarwal, A. Rajadesingan and P. Kumaraguru. PhishAri: **Automatic realtime phishing detection on twitter.** eCrime Researchers Summit, eCrime, pp. 1-12, 2012