# Social Media Fake Profile Detection Using Data Mining Technique

Nitika Kadam and Sanjeev Kumar Sharma
Computer Science Engineering, Oriental University, Indore, India
Email: {kadamnitika01, spd50020}@gmail.com

***Abstract*—In social media, a significant amount of data has been distributed in the entire world with thousands of new users joining social media each day. Social media is a virtual life where malicious users can impact someone's reputation. Mostly such kind of activity is performed by fake accounts. Thus, identification of fake profiles is necessary and can be done in the early stage of profile building is an essential task for ML. In this paper, the aim is to design a ML model which identifies fake profiles in the early stage and ML based survey on social media has been carried out. Further, the collected literature is categorized according to the used social media datasets and popular areas of employing ML in social media platforms. In this investigation, we have used the Twitter dataset fake profile detection to demonstrate the proposed idea of ML-based fake news detection. The proposed model includes preprocessing to refine the contents and attributes to improve the quality of the dataset and reduce dimensions of the data. The next five popular ML algorithms namely C4.5, Bayes classifier, SVM, ANN, and KNN algorithms are implemented to predict the fake profiles. The evaluation of the system is performed under two scenarios based on training and testing sample ratio of 70-30% and 80-20% and using 4-fold cross-validation. Findings show 80-20% based samples reduce the resource consumption and 70-30% of ratio improves the classification accuracy. Finally, the future extension of the presented work has been discussed.**

***Index Terms*—social media analysis, security and privacy, fake profile detection, data mining and techniques, survey**

## I. Introduction

Use of Machine Learning (ML) is growing for different applications to recognize patterns, predictions, and classification. These techniques can perform automated and accurate data analysis. Thus, it becomes acceptable in the field of engineering, medical, business, and more. Similarly, the increasing use of digital data is not only increasing new possibilities it also involving new challenges [1]. In this paper, the use of ML in social media security in terms of early fake profile detection has been described.

Social networking or "Online Social Networking (OSN)" becomes much popular in recent years [2]. It is a platform to search people, share data, and express emotions [3]. It is a way to share information with their contacts. But issues related to information leakage, identity, and discloser of sensitive information invite malicious attacks [4].

There are two kinds of OSN is available i.e., client-server and peer to peer architecture based. Almost all OSNs are web-based and centralized OSN. The processes i.e., storage, maintenance, and access are provided by centralized authorities such as Facebook and Twitter [5]. Some OSN can also be designed using P2P architecture. It is a decentralized approach to implementation. This supports data exchange and local services during the absence of the Internet [6]. Social media is a popular platform for every age group where anybody can join and meet new people i.e. Facebook and Twitter [7]. There are two kinds of users first technically sound and know the usages and limits. And the second type of user is not understanding the technology and limitations [8].

Some users are using the OSN platform for promotions, activity, events, political views, and advertisements, and working legitimately but some of them are abusing the policies of OSN by promoting and distributing the hate, spam, and phishing contents [9]. Thus, identification and differentiation between fake and legitimate profiles in OSN are required for friendly and secure OSN ecology.

In this paper, we are aimed to study and design an ML-based method for fake profile classification. These techniques are applied to GitHub-based twitter fake profile detection dataset for classifying them into fake and legitimate profiles. The next section reports essential contributions for fake profile detection using machine learning-based techniques. Further, a data mining model is proposed. Then next we have discussed the experimental analysis and results. Finally, the conclusion and future research directions are suggested for improvements.

## II. Literature Survey

Due to the increasing popularity of social media platforms fake profiles is also growing. There is various type of malicious purpose behind creating such a false account or identity. Using such kind of fake profiles are very harmful to society and can be involved in various social and cybercrimes. Therefore, in order to understand the nature and current research or social media security, we have collected more than 50 recent research and survey articles. Among them, we have selected the 25 most relevant to the proposed research domain. Next, we have

Manuscript received July 31, 2021; revised February 28, 2022; accepted March 1, 2022.


doi: 10.12720/jait.13.5.518-523

categorized these research papers into six research categories. These categories are listed in Table I.

TABLE I. RESEARCH ISSUES IN OSN

| S. No. | Issue | References |
|---|---|---|
| 1. | Fake Profile | [10], [11], [12], [13], [14] |
| 2. | Social bots | [15], [16], [17], [18], [19] |
| 3. | Fake News | [20], [21], [22], [23], [24] |
| 4. | Spamming | [25], [26], [27], [28] |
| 5. | Attacks and security | [29], [30], [31], [32], [33] |
| 6. | Risk assessment | [34] |

In most of the work fake profiles are involved. According to the percentage popularity of the research topic, we described the total amount of work into their area of employment in Fig. 1.

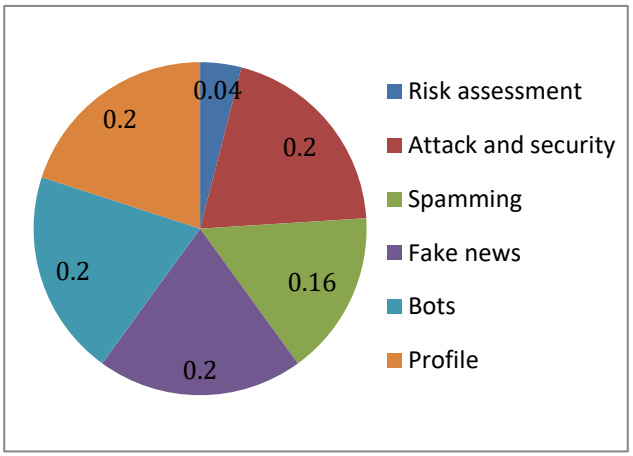


Figure 1. Social media security contributions.

In the collected literature we found most of the authors are utilizing the ML algorithms for identifying the fake user accounts. Therefore, we need a suitable dataset also thus we have explored the same articles in order to find an appropriate dataset to train the algorithms. Table II and Fig. 2 demonstrate the popularly used dataset prepared for different social media platforms.

TABLE II. DATASET USED IN RESEARCH WORK

| S. No. | Dataset Used | References |
|---|---|---|
| 1 | Twitter | [13], [25] |
| 2 | Facebook | [35] |
| 3 | Public spam user dataset | [36] |
| 4 | Twitter | [37], [38] |
| 5. | Academia | [39] |
| 6. | LinkedIn | [40] |

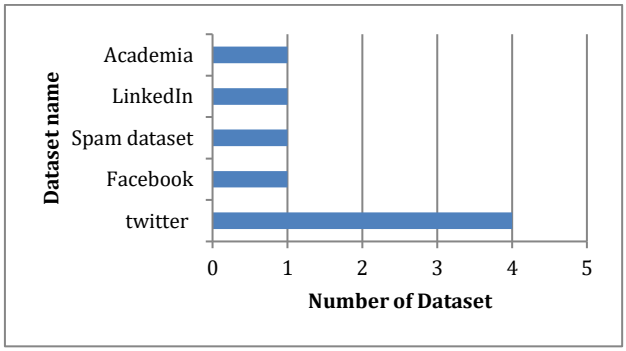


Figure 2. Dataset used

The literature shows the different authors are considering different scenarios of experiments and due to which they are creating datasets by own or utilizing some predefined datasets. By analysis of 8 articles that are working with existing datasets. Based on the findings of the articles we concluded that the Twitter dataset is frequently used additionally openly available for experimentations. Thus, we have concluded to use the Twitter social media dataset for our experimental study.

## III. PROPOSED WORK

To design the required machine learning model for the early-stage fake profile detection technique we have proposed an experimental model. This model is demonstrated in Fig. 3. The details of the proposed model are described in this section.

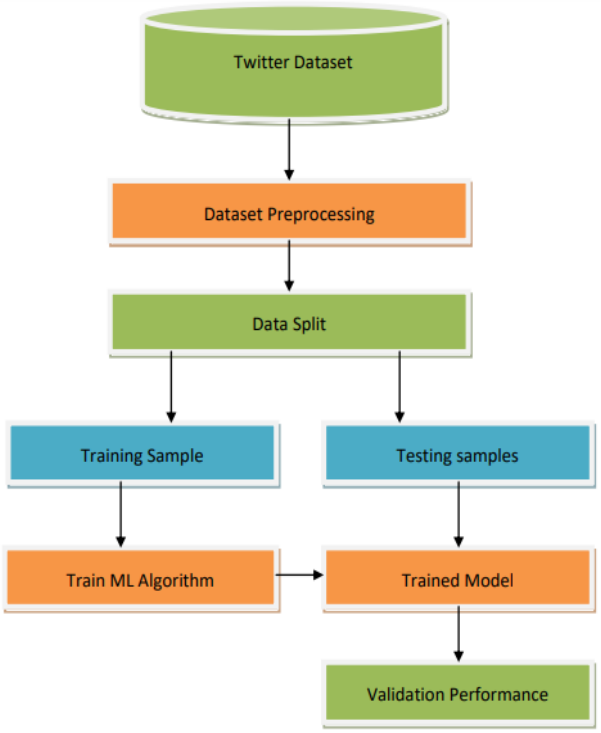


Figure 3. Proposed classification model.

### A. Dataset

The main aim of the proposed investigation is to explore and design an accurate and efficient data mining model for

classifying fake Twitter profiles. In this context, we offer the evaluation of supervised learning algorithms with the available fake profile dataset [41]. That dataset is available in Comma-Separated Values (CSV) format. The dataset contains a total of 33 attributes. The profile information is distributed in two separate files one for fake and the second for legitimate. The first file contains 1338 instances and the second file contains 1482 instances. The file name is treated here as the class labels legitimate and fake. After combining both the files we get a total of 2820 instances of data and two class labels.

### B. Data Preprocessing

The dataset contains a significant number of attributes that are a total of 34 attributes and one class label. In this process, we are trying to reduce and refine the attributes which are essential. The attributes ID, Name, and screen_name is used for identifying the person or profile. Thus, among these three attributes we just pick only one of them here, we take the ID as compared to the other two attributes. Further, the attributes statuses_count, followers_count, friends_count, and favourites_count are essential for profile identification. Next, the dataset contains the listed count which is not much effective according to us thus we reduce this attribute. Further, the attribute created at is important to know how old a profile is thus the date and time are converted into the number of days. Obviously, a social media profile has a unique URL thus we remove the URL attribute. Further, the attributes lang, time_zone, and location can be combined into one, thus we consider time zone as compared to the other two.

Here two attributes default_profile and default_profile_image is consolidated into one as the Boolean true or false. Further attribute geo_enabled, profile_image_url, and profile_banner_url is transformed into Boolean. Additionally, profile_use_background_image and profile_background_image_url_https is consolidated into one as Boolean. Further, three attributes namely profile_text_color, profile_image_url_https, and profile_ sidebar_border_color are not much essential for profile characterization thus we reduce these attributes. In next profile_background_tile is converted into Boolean, additionally profile_sidebar_fill_color, profile_background _image_url, profile_link_color and utc_offset is removed as non-essential attributes. Further, the attributes Protected, Verified and Description is used as Boolean. Next, the Updated is used as the number of days for finding freshness of profile, and the last attribute Dataset is removed as a non-essential attribute.

Finally, among 34 attributes we consider only 17 attributes and reduced nonessential attributes. Thus, now after consolidation and transformation, we have 17 attributes and one class label as part of the dataset. But the data may contain an amount of noise and unwanted data. Additionally, it is possible some of the instances are abundant which may contain missing or null values and special characters. The aim of preprocessing is to clean the data and improve the quality to improve the learning performance.

### C. Pattern Learning and Classification

The preprocessed data is used further for decision-making purposes. Therefore, data organized previously is preserved in a local database to create training and testing data. In this context, 70% of randomly selected data instances are used as the training set. In addition, that using the concept of n cross-validation 4 fold test dataset is prepared. The 30% of randomly selected data are in four folds used for testing the data mining algorithms. Additionally, the 80-20% ratio is also used for experimentation. In order to learn about data patterns and to accurately classify the data following experimental system is developed as given in Fig. 3.

The functional aspects of the components input dataset, data preprocessing, data splitting, training set preparation, and testing dataset preparation are explained in the previous section. Here the three components are explained namely algorithm training, trained model, and classification performance. In order to train the system, five supervised learning algorithms are considered namely C4.5 decision tree, SVM (support vector machine), ANN (Artificial Neural Network), Bays classifier, and KNN (k-nearest neighbor) algorithm. These models are accepting the training datasets and produce the trained model accordingly. For example, SVM and ANN are producing the opaque model, KNN, Bays, and C4.5 algorithms producing the transparent models.

After learning these models accepting the test dataset prepared in 4 folds and performs classification and produces the efficiency of classification outcomes for the test datasets. These model's performance outcomes are reported in the below Table III.

TABLE III. THE EFFICIENCY OF CLASSIFICATION OUTCOMES FOR THE TEST DATASETS

| Algorithms | Performance Summary for 70-30% | | | | Validation Summary for 80-20% | |
|---|---|---|---|---|---|---|
| | Accuracy | Error rate | Memory | Time | Accuracy | Error rate |
| C4.5 | 86.5% | 13.5% | 14029 KB | 267 MS | 83.4% | 16.6% |
| Bays | 84.3% | 15.7% | 13898 KB | 289 MS | 82.9% | 17.1% |
| ANN | 97.4% | 2.6% | 15294 KB | 365 MS | 95.7% | 4.3% |
| SVM | 96.5% | 3.5% | 15164 KB | 376 MS | 94.2% | 5.8% |
| KNN | 84.2% | 15.8% | 13772 KB | 398 MS | 83.5% | 16.5% |

## IV. RESULTS ANALYSIS

This section explains and compares the performance of the implemented data mining algorithm in the context of classifying fake profiles. The following performance parameters are measured for their comparative performance study.

### A. Accuracy

The accuracy can be explained as the measurement of algorithm classification correctness. That can be measured using the ratio of total correctly classified and the total

patterns to be classified. That can also be represented using the following equation:

$$accuracy = \frac{total\ correctly\ classified}{total\ patterns\ to\ classify} X100$$

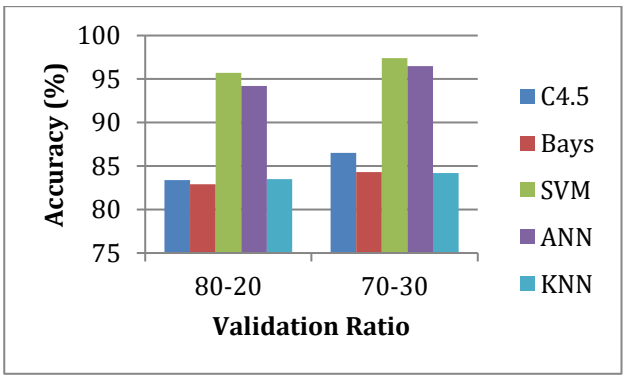


Figure 4. Accuracy (%).

The accuracy of the algorithms is given in Fig. 4 for both the validation ratio. The accuracy of the algorithm is notified in the Y-axis and the X-axis shows the validation ratio. The accuracy of the algorithm is calculated here in terms of percentage (%). According to the obtained results, the performance of the algorithm is found effective with the 70-30 ratio as compared to the 80-20 ratio. Additionally, we found that the SVM and ANN show higher performance as compared to the other implemented algorithms. Thus, in near future, both the algorithms can be considered for implementation of the proposed data model.

### B. Error Rate

The error rate of an algorithm demonstrates the misclassification rate of the algorithm as a performance parameter. That can be calculated using the following equation:

$$Error\ Rate = 100 - Accuracy$$

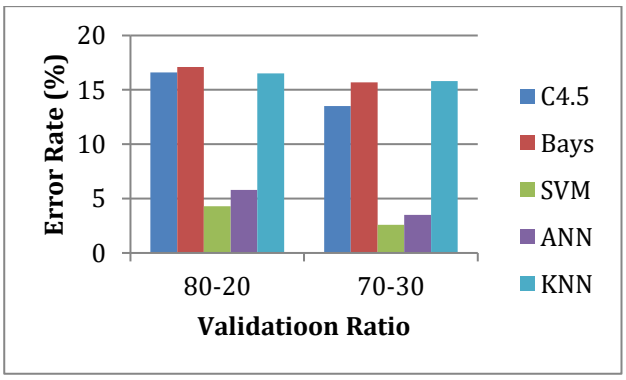


Figure 5. Error rate (%).

The error rate of the implemented algorithms is shown in Fig. 5. The performance shows for both kinds of validation ratios. In order to show the performance of the algorithm X-axis shows the validation ratio additionally Y axis shows the error rate (%). The performance of the system shows the ANN and SVM report fewer error rates as compared to other algorithms.

### C. Time Consumption

The time consumption is also termed time complexity. The amount of time consumed for classification is also calculated in this section using the following formula:

$$time\ consumed = end\ time - start\ t\ ime$$

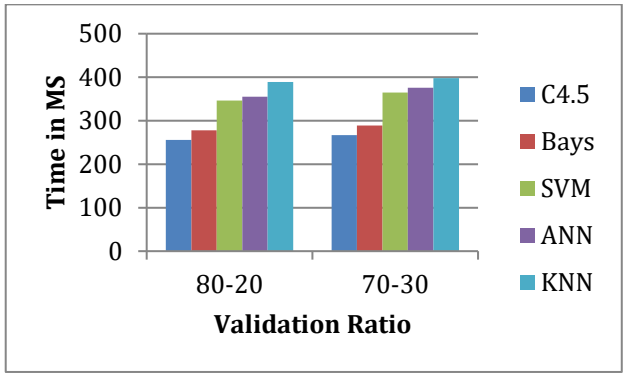


Figure 6. Time consumption (MS).

The two-validation ratio i.e., 80-20 and 70-30 is reported in Fig. 6. The performance of the implemented algorithms in terms of time consumption is given using the same figure. The time is measured here in terms of milliseconds. In order to represent the performance of both kinds of validations, X-axis represents the ratio and Y-axis shows the time. According to the obtained results, the ratio 70-30 consumes a higher amount of time as compared to the 80-20 ratio. That is because the 80-20 ratio contains fewer amounts of data for classification as compared to the 70-30 ratio.

### D. Memory Usage

The memory usages are also an essential parameter for the performance evaluation of a data mining algorithm. The memory usages of the algorithm are computed using the following equation.

$$memory\ usage = total\ memory - free\ memory$$

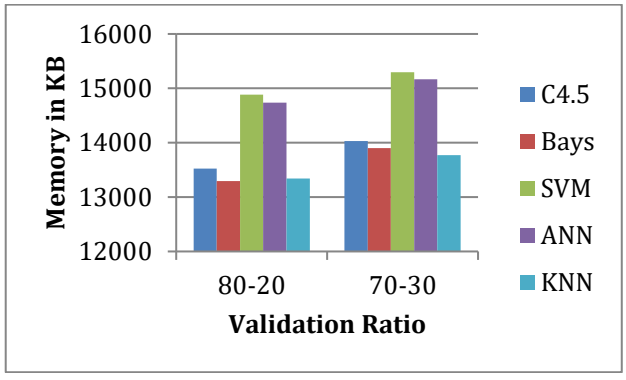


Figure 7. Memory usage (KB).

The memory usage of the implemented algorithms is explained in Fig. 7. That is provided in two parts first contains the 70-30 ratio and the second contains the 80-20 ratio. In order to show the performance, the X-axis contains the validation ratio and Y-axis shows the memory in KB (kilobytes). According to the results, the ratio 80-20 requires less amount of memory as compared to 70-30

ratio because the ratio 80-20 requires less amount of data storage on the main memory as compared to 70-30 ratio.

## V. Conclusion & Future Work

The aim of this paper is to explore the techniques and methods which are used for fake profile detection in different social media platforms. In this context, the survey on existing approaches based on machine learning and data mining is explored. In addition to that, the different datasets available are also obtained. Based on the availability of the dataset a data mining model is proposed in this work. In this context first, the dataset is refined and consolidated with the expert's help and then the popular data mining algorithms are applied to the data. There are five machine learning algorithms namely KNN, SVM, ANN, Bays, and the C4.5 decision trees are used. Further for obtaining the performance the 4-fold cross-validation process is used and the performance in terms of accuracy, error rate, memory, and time complexities are measured. There are two kinds of validation ratios that were used i.e., 70-30% and 80-20%. The performance summary of the techniques is reported in the table.

According to the obtained performance, the proposed model demonstrates the performance in terms of accuracy and error rate works effectively for 70-30% ratio and for resource consumption 80-20% is the effective ratio. Using these obtained results, we obtained two effective and accurate classification techniques which are further used for developing a more improved model of fake profile detection. In near future, the proposed work is extended in the following manner.

1) The use of the concluded algorithm is done for implementing the further extended model
2) The current model is extended with the help of profile contents also and the text mining algorithms
3) In order to obtain the effective profile text analysis, the sentiment-based classification is used

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Nitika Kadam conducted the research, collected data, and wrote the paper. Dr. Sanjeev Kumar Sharma supervised the work and approved the final version.

## References


[1] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Transactions on Intelligent Systems and Technology*, vol. 8, no. 2, article 16, Oct. 2016.

[2] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM Sigmod Record*, vol. 42, no. 2, June 2013.

[3] A. Whiting and D. Williams, "Why people use social media: A uses and gratifications approach," *Qualitative Market Research: An International Journal*, vol. 16, no. 4, pp. 362-369, 2013.

[4] D. Gan and L. R. Jenkins, "Social networking privacy—Who's stalking you?" *Fut. Inte.*, vol. 7, pp. 67-93, 2015.

[5] C. Zhang, J. Sun, X. Zhu, and Y. Fang, "Privacy and security for online social networks: Challenges and opportunities," *IEEE Network*, Jul.-Aug. 2010.

[6] S. Buchegger, D. Schioberg, L. H. Vu, and A. Datta, "PeerSoN: P2P social networking—Early experiences and insights," in *Proc. the Second ACM EuroSys Workshop on Social Network Systems*, 2009.

[7] C. Wüest. The Risks of Social Networking. Security Response. [Online]. Available: https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the_risks_of_social_networking.pdf

[8] T. Ravichandran, "Enhancing soft skills and personality," Indian Institute of Technology Kanpur, Nat. Prog. on Tech. Enha. Lear., 2022.

[9] A. Romanov, A. Semenov, and J. Veijalainen, "Revealing fake profiles in social networks by longitudinal data analysis," in *Proc. of the 13th Intel. Conf. on Web Infor. Sys. and Tech.*, 2017, pp. 51-58.

[10] C. Xiao, D. M. Freeman, and T. Hwa, "Detecting clusters of fake accounts in online social networks," in *Proc. the 8th ACM Workshop on Artificial Intelligence and Security*, 2015.

[11] Z. Yamak, J. Saunier, and L. Vercouter, "Detection of multiple identity manipulation in collaborative projects," in *Proc. the 25th International Conference Companion on World Wide Web*, 2016.

[12] M. M. Swe and N. N. Myo, "Blacklist creation for detecting fake accounts on Twitter," *Inter. Jour. of Netwo. and Distr. Comp.*, vol. 7, no. 1, pp. 43-50, Dec. 2018.

[13] M. Mohammadrezaei, M. E. Shiri, and A. M. Rahmani, "Identifying fake accounts on social networks based on graph analysis and classification algorithms," *Hindawi Secu. & Comm. Netw.*, vol. 2018, art. ID 5923156, 2018.

[14] 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. the 25th International Conference on World Wide Web*, 2016.

[15] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in *Proc. of the Elev. Inter. Confe. on Web and Soc. Med.*, 2017.

[16] E. V. D. Walt and J. Eloff, "Using machine learning to detect fake identities: Bots vs humans," *IEEE Access*, vol. 6, pp. 2169-3536, 2018.

[17] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proc. the 26th International Conference on World Wide Web Companion*, 2017.

[18] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Comm. of the ACM*, vol. 59, no. 7, July 2016.

[19] S. Gurajala, J. S. White, B. Hudson, and J. N. Matthews, "Fake Twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach," in *Proc. the International Conference on Social Media & Society*, 2015.

[20] S. Tschiatschek, A. Singla, M. G. Rodriguez, A. Merchant, and A. Krause, "Fake news detection in social networks via crowd signals," *Companion Proceedings of the Web Conference 2018*, 2018.

[21] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," arXiv:1708.01967v3 [cs.SI], 3 Sep. 2017.

[22] E. Ferrara, O. Varol, F. Menczer, and A. Flammini, "Detection of promoted social media campaigns," in *Proc. the Ten. Inter. Conf. on Web and Soc. Med.*, 2016.

[23] J. Song, S. Lee, and J. Kim, "CrowdTarget: Target-based detection of crowdturfing in online social networks," in *Proc. the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015.

[24] V. L. Rubin, Y. Chen, and N. J. Conroy, "Deception detection for news: Three types of fakes," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1-4, 2015.

[25] F. Masood, *et al.*, "Spammer detection and fake user identification on social networks," *IEEE Access*, vol. 7, pp. 68140-68152, 2019.

[26] I. Sen, A. Aggarwal, S. Mian, S. Singh, P. Kumaraguru, and A. Datta, "Worth its weight in likes: Towards detecting fake likes on Instagram," in *Proc. the 10th ACM Conference on Web Science*, 2018.

[27] P. Ratna, B. Satya, K. Lee, D. Lee, T. Tran, and J. Zhang, "Uncovering fake likers in online social networks," in *Proc. the 25th ACM International on Conference on Information and Knowledge Management*, 2016.
[28] A. M. A. Zoubi, J. Alqatawna, and H. Faris, "Spam profile detection in social networks based on public features," in *Proc. 8th Inter. Conf. on Inform. & Comm. Syst.*, 2017.
[29] J. Jia, B. Wang, and N. Z. Gong, "Random walk based fake account detection in online social networks," in *Proc. 47th Ann. IEEE/IFIP Inter. Conf. on Depen. Sys. & Net.*, 2017.
[30] S. Rathore, P. K. Sharma, V. Loi, Y. S. Jeong, and J. H. Park, "Social network security: Issues, challenges, threats, and solutions," *Infor. Scie.*, vol. 421, pp. 43-69, 2017.
[31] R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," *Egy. Infor. Jour.*, vol. 17, pp. 199-216, 2016.
[32] O. Goga, G. Venkatadri, and K. P. Gummadi, "The doppelgänger bot attack: Exploring identity impersonation in online social networks," in *Proc. the Internet Measurement Conference*, 2015.
[33] B. Viswanath, *et al.*, "Strength in numbers: Robust tamper detection in crowd computations," in *Proc. ACM on Conference on Online Social Networks*, 2015.
[34] N. Laleh, B. Carminati, and E. Ferrari, "Risk assessment in social networks based on user anomalous behaviours," *Jou. of Lat. Class Files*, vol. 13, no. 20, pp. 1545-5971, Oct 2014.
[35] P. S. Rao, J. Gyani, and G. Narsimha, "Fake profiles identification in online social networks using machine learning and NLP," *Int. Jou. of Appl. Engg Res.*, vol. 13, pp. 4133-4136, 2018.
[36] A. Gayathri, S. Radhika, and S. L. Jayalakshmi. Detecting fake accounts in media application using machine learning. *Spec. Iss. Pub. in Int. Jnl. Of Adv. Netw. & Appl*. [Online]. Available: https://www.ijana.in/papers/67.pdf
[37] S. Gurajala, J. S. White, B. Hudson, and J. N. Matthews, "Fake Twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach," in *Proc. International Conference on Social Media & Society*, 2015.
[38] P. Shahane and D. Gore, "Detection of fake profiles on Twitter using random forest & deep convolutional neural network," *Inter. Jour. of Manag., Techn. & Engg.*, vol 9, no. 6, June 2019.
[39] K. C. Yang, O. Varol, C. A. Davis, E. Ferrara, A. Flammini, and F. Menczer, "Arming the public with artificial intelligence to counter social bots," *Hum Behav & Emerg Tech.*, vol. 1, p. 48-61, 2019.
[40] C. Xiao, D. M. Freeman, and T. Hwa, "Detecting clusters of fake accounts in online social networks," in *Proc. the 8th ACM Workshop on Artificial Intelligence and Security*, 2015.
[41] Fake profile detection using ML. [Online]. Available: https://github.com/harshitkgupta/Fake-Profile-Detection-using-ML



**Nitika Kadam** was born in Indore, Madhya Pradesh, India. She received her both Master of Engineering and Bachelor of Engineering degree in Computer Science Engineering. She is currently pursuing her PhD from Oriental University, Indore. Her academic research interests include machine learning, deep neural network and making human life efficient using Artificial Intelligence. She is now currently working as an Associate Professor, at SAGE University.

**Dr. Sanjeev Kumar Sharma** is working as Professor (CSE) and Dean Student Welfare in Oriental Institute of Science and Technology, Bhopal and Adjunct faculty in the Oriental University, Indore. Apart from teaching, he is also associated with the activities of Indian Army through National Cadet Corps (NCC) to motivate and encourage the students to join the armed forces. He completed his Ph.D. in computer science and engineering from Devi Ahilya University. He has more than 40 research papers in various national, International Journals and Conferences. He is having 18 years of teaching experience and 12 years of experience in Research. He is also the member of various societies such as, AMIE, CSI, ACM.