Detecting Suspicious News over Social Media using Ant Colony Optimization

Asha Kumari, Balkishan

Abstract: ***The online social media platforms has become the trending as it provide the convenient and free access to users to share their day to day activities and other information. Despite the significance of these online social media platforms, there are also the people that can mislead the other by posting the fake news. These kinds of news are termed as suspicious news. Such kind of misleading news can badly affect the society. It is the way too hard to completely restrict such people from posting anything on social media. But after the detection of such activities, these posts can be removed from social media and users can be restricted from the respective platform. From the years, researchers are continuously working to detect the suspicious activities using machine learning and data mining techniques. This research work addresses the problem of suspicious news detection using the ant colony optimization based ant miner plus technique. This proposed approach is termed as ACODSN (Ant Colony Optimization for the Detection of Suspicious News). The experimentation is conducted on the dataset of FakenewsNet. The system performance is analyzed in terms of evaluation metrics of recall, precision, and f-measure.***

Keywords: Fake News, Ant Colony Optimization, Ant Miner Plus, Suspicious News, Data Mining, Swarm Intelligence.

I. INTRODUCTION

The invention of internet and smart phones has abruptly changed the life style of human beings. The telephonic conversations changed to online message chats, face to face meeting changes to online face time meetings, public gathering changed to online group discussions and many more such instances are available. The freedom of social media applications has made the human independent to post and share anything publically. In these daily tending posts, some dubious posts or news can be seen which can lead to unethical activity and badly disturb the people. Such types of doubtful news are terms as fake news or suspicious news. The day by day increasing population, increasing online applications, increasing poverty & unemployment, and various other factors are responsible to make the user to create and post unauthentic news content.

Due to lack of awareness, some legitimate people also share such posts on their social media walls which overall create the serious problem for the society. The easy accessibility of internet and smart phones is also responsible to spread more fake news in various forms such as satires, fake political statements, rumors, fake advertisements & reviews for any product, etc.

Fake news can be created or spread by humans or non-humans (cyborgs and bots). Non-human entities can spread and create the content automatically in a quick manner as compared to human. These fake news are created in both the positive and negative manners. There is the popular instance of US elections held in 2016 in which millions of non-human entities were created to support the Clinton or Trump and oppose their respective other candidate [1]. In another case of humans, there are also instances in which human has shared and spread the fake news. The popular instance of human based fake news is related to the death of a FBI agent who was found as a suspect to leak the emails of Hillary [2]. This news was shared by half a million of users on Facebook and more than 15 million people reacted to this post.

The targeting users for the suspicious news depend on the target agenda of suspicious person. The terrorism groups can target the people who have some anti-national views, the fake advertisements or fake reviews can target the online customers, fake health news can target the old age people, fake educational news can target the students and parents, etc. In this way, the fake news directly or indirectly can affect the human life and can lead to suspicious activity. For instance, recently, the fake news of terror attacks were spread on the social media in June, 2019 [3]. The spreaded news photos were actually of the terror attack happened in Paris in the year 2015. Some suspicious entities have restructured the old photos and spreaded over the social media. It was either to create the operational preparedness or to affect the Muslim communities.

The fake news can be created and spread by any individual. The four major components of fake news are social scenario, news content, target users, and fake news creator [4]. As discussed earlier, fake news creator can be either human or non-human entities that can create the fake news. The users who spread the fake news either intentionally or unintentionally are equally responsible to create the intense situation. The target users are the legitimate users that become the target of the fake news creator through news channels or online social media platforms. The target users can be any old age people, voters, students, parents, online customers, etc.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author

Asha Kumari *, Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, India, Email: asha198829@gmail.com

Balkishan, Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, India, Email: balkishan248@gmail.com

The news content involves both the textual content of the news and public sentiments related to that news. The last component is social scenario. The social scenario is the way adapted by fake news creator to spread the news on social media platforms. It depends on the followers of user and his behavior to share posts.

The detection of suspicious news is the procedure to access and detect the truth about some of the pieces of news [5]. The focus of researchers in the past years was mainly on the detection of rumors but the recent research focus has shifted from the detection of rumors to fake news [6]. The recent data statistics of papers published and indexed in the Scopus database related to rumors and fake news is illustrated in fig. 1.

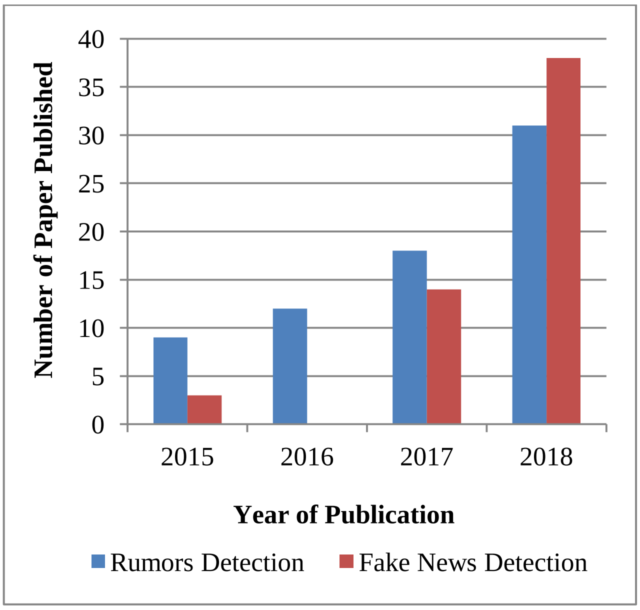

Fig. 1: Statistics of the papers published and indexed in Scopus database based on rumor detection and fake news detection topic [6]

The fig. 1 illustrates the statistics of paper published and indexed in Scopus database from year 2015 to 2018. It also indicates the switching behavior of the researchers from rumor detection to fake news detection. There are various existing methods and concepts related to fake or suspicious news detection. Different researchers have adapted different methods to detect the fake news from different social media platforms with the experimentation on different datasets. In the current research work, an artificial intelligence based concept of ant colony optimization (ACO) is considered. The detection of suspicious news is conducted using the Ant Miner Plus classifier of ACO algorithm. This proposed approach is termed as ACODSN (Ant Colony Optimization for the Detection of Suspicious News). The research experimentation is conducted on the dataset of FakenewsNet having news content from BuzzFeed and PolitiFact. The performance of the ACODSN approach is accessed in terms of recall, precision, and f-measure parameters.

The other sections of the paper are structured as follows. Section 2 illustrated the work related to the suspicious news detection. Section 3 elaborates the proposed research methodology of ACODSN approach. The research database of FakenewsNet is also discussed in this section. Further, Section 4 presents the performance assessment based on the evaluated results and comparison. The last section 5 concludes the paper with the future view points.

II. RELATED WORK

Fake news enormously affects the knowledge and perception of people that leads to distort their awareness and decision making capability regarding any news. The major factor responsible for the enormous growth of fake news is the independence authorized by social media to post any information. Researchers are continuously working on the techniques to detect the fake news available on online social media platforms so that its impact can be diminished. In this research work, the latest and quality contributions of authors for fake news detection are discussed.

Ruchansky et al. [7] have hybridized the characteristics related to the fake news such as the textual information, user profile & response, and news spreading source. There characteristics have been utilized by authors to propose the CSI approach (consists of elements: Capture, Score, & integrate). In the first element 'capture', the neural network was used to determine the pattern based on the textual information and response in respect to the available information. The second element collects the source attributes on the basis of user behavior. The last module was responsible for the classification of news as fake or real. The system performance was accessed for the Twitter and Weibo social media platforms. The authors noted the efficient performance results for the proposed CSI model.

Granik and Mesyura [8] have used the Naïve Bayes classifier for the classification and detection of fake news articles. The authors have conducted the research experimentation as a basic model for the detection of fake news. The research dataset of BuzzFeed news was used for the training and testing of classifier. The database carries the news from the Facebook posts along with the some news pages related to politics. The authors have noted the accuracy results of 75.40% for the considered classifier and suggested to consider the advanced artificial intelligence concept for the result improvement.

Shu et al. [9] have proposed the semi-supervised approach by integrating the attributes of news instance, publisher bias, and user engagement. The proposed approach was termed as TriFN which represents the framework for the detection of fake news with tri-relationship. The authors have also proposed two datasets of PolitiFact and BuzzFeed News for the detection of fake news. The research experimentation conducted on the mentioned dataset using the proposed TriFN approach indicates the remarkable result values for the mentioned approach.

Further, Shu et al. [10] have proposed the FakeNewsTracker to analyze the fake news. The proposed framework has the capability to collect, detect, and visualize the data of fake news. The authors have collected the data from PolitiFact and BuzzFeed News with the information of ground truth value and social engagement of user. The detection of fake news is performed using the social article fusion approach.

To visualize the data, the cloud approach is adapted with the geo-location and social network of user. To test the system performance, the dataset was tested with methods of naïve bayes, logistic regression, and support vector machine. The system efficacy was noted with considerable results.

Atodiresei et al. [11] have presented a real time application for the detection of fake news. The authors have used the dataset entities from Twitter and detected the fake news & their respective suspicious fake users. The research work was conducted with the consideration of natural language processing modules, naïve bayes classifier, support vector machine, and maximum entropy approach. The authors have attained the considerable detection reports but with the limitation of method to use in English language and sometimes system lacks in case of novel fake news from trustworthy channel.

Ko et al. [12] have adapted the backtracking approach for the classification of fake news and true news. The research experimentation considers the cognitive system based approach for the considered case studies. The methods of backtracking and dummy news detection are applied by generating some of the dummy news to detect it with reverse tracking approach with their prediction to be real or fake. The constructive results been noted with the system drawback in case of subjective news as subjective news can be generated with the journalist's philosophy instead of any facts.

Vishwakarma et al. [13] have proposed rule based system to test the veracity of news and information available on the social media platforms. The authors conducted the research work by extracting the textual data from the images and the veracity of news was analyzed by comparing the results with the top 15 web links available on the Google. The authors explained the research methodology in the four modules of extraction of textual data from image, extraction of entity, web based processing, and process unit. The research experimentation also conducted by considering the dataset of PHEME, FakeNewsNet, and BuzzFeed Election. The authors have noted the remarkable results but the system lacks in case of local news due to availability of lesser information on web for local news.

III. RESEARCH METHODOLOGY

The research methodology of ACODSN is discussed in section. There are four major steps of proposed ACODSN approach. These steps are Data Collection and Consideration, Pre-processing, feature extraction and selection, and classification of news. The overall working procedure of proposed ACODSN approach is illustrated in fig. 2.

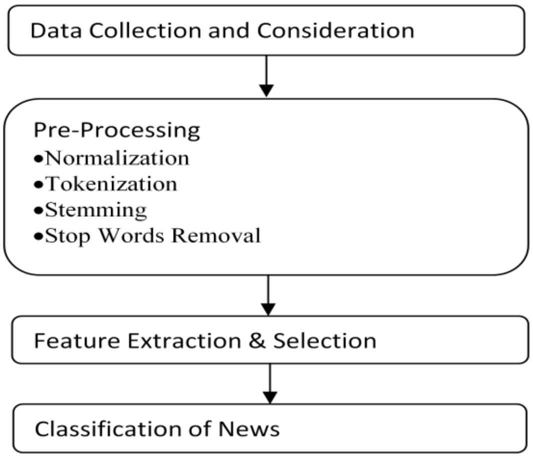

Fig. 2: The Working Procedure of Proposed ACODSN Approach

A. Data Collection and Consideration

In this research work, the dataset of FakenewsNet is utilized. This dataset is prepared by Shu et al. [9] by taking the news content, the information related to user, and the publisher information. In this dataset, there are two categories as mentioned: BuzzFeed news and PolitiFact news. The ground truth for both the categories is labeled from buzzfeed.com and politifact.com respectively. In the BuzzFeed category, there are a total of 182 news. In another category of PolitiFact, there are a total of 240 news. The statistics of FakenewsNet is illustrated in table I. Here, fake news are referred as the suspicious news.

Table I: Statistics of FakenewsNet Dataset

Feature/Category	**PolitiFact**	**BuzzFeed**
Suspicious News	120	91
True News	120	91
Total News	240	182
Publisher	91	9

This research dataset is considered for the experimentation and pre-processing is applied on this dataset in the next step.

B. Pre-processing of dataset

This is an important module of proposed ACODSN approach as raw data can't be directly considered for classification. The pre-processing steps are significant as it helps to reduce the search space and makes the data worthy to extract the features. In this research work, the pre-processing steps of normalization, tokenization, stemming, and stop word removal is considered. These modules are illustrated in fig. 3.

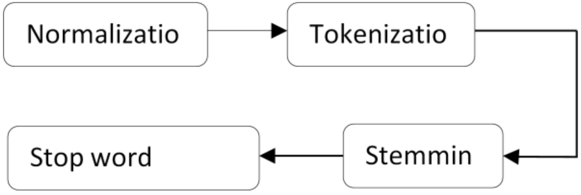

Fig. 3: Pre-processing Steps of Proposed ACODSN Approach

The first module of pre-processing is normalization. Normalization is considered to detect and normalize the data available in the other than English language. Here, Google translator is exploited to translate the other language based news into English language. This normalized data is processed for the tokenization. Here, unigram tokenization is applied and each word in the dataset is separated based on the 'space' separator. The separated words are considered as tokens. These token are further processed for the stemming process. Stemming process is performed using the porter stemmer approach that removes the adjoining affix & suffix of the word and converts the work into the basic root word. For instance, there are various feasible words for the word 'laugh' such as 'laughter', 'laughing', 'laughable', 'laughs', 'laughingly', 'laughably'. The stemming processing removes the suffix of 'er', 'ing', 'able', 's', 'ingly', 'ably' respectively from the mentioned words and converts these words into root word 'laugh'. These stemmed tokens are accessed to check and remove the stops words. Stop words removal process removes the insignificant words such as 'in', 'an', 'are', 'you', and 'who', 'that', etc. The removal of stop words process reduces the search space and enhances the system performance.

C. Feature Extraction & Selection

The features are the only component that decides the category of news as suspicious or true. To analyze and decide the behavior of news as suspicious or not, we have considered the features based on the user behavior, textual content, and URL information. The considered list of features is illustrated in table II.

Table II: List of Features

Authenticity of Source URL
Authenticity of Authors
Publisher Details
Length of news subject/title
Availability of special symbols in subject/title
Availability of Suspicious Words in news subject/title
Frequency of Suspicious Words in body
Relevance score of news title with textual body content

Table II illustrates the list of features based on the URL information, textual content, and user authenticity. For instance, if the news belongs to reliable source, having trustworthy authors & publishers, having great relevancy among the news title & textual body, then the news article will be considered as the authentic (true). In the opposite scenario, it will be counted as suspicious.

D. Classification of News

The classification of news as suspicious or true is the final step of the proposed ACODSN approach. The classification of process is performed using the ant colony optimization (ACO) based Ant Miner Plus (AM+) algorithm [14]. The classification rules for the AM+ approach were defined by Martens et al. [14]. These rules are based on the algorithm of ACO and work on the scenario of <antecedent> <consequent> prediction.

The process of AM+ begins by initialization of parameters in the scenario. Consider an undirected graph *G = (V, E)* with n number of vertices ($V = \{v_1, v_2, \dots, v_n\}$) and m number of edges ($E = \{e_1, e_2, \dots, e_n\}$).

The pheromone value is also initialized to the maximum possible extent to make the exploitation of ants. Further the heuristic function value is determined. Further, the probability for the detection of edge location is determined.

Further, the pheromone is updated for the ants. The algorithm stops after meeting the stopping criteria. The stopping criteria are the achievement of best solution and completion of iterations. The final decision of news as suspicious or not is finalized based on the feature matching as suspicious or not. In this way, news can be determined as true or suspicious.

IV. RESULTS AND DISCUSSION

The results of the proposed ACODSN approach are determined for the FakenewsNet dataset. The categories of FakenewNet dataset are BuzzFeed and PolitiFact. For testing, 60% utilized for training and 40% utilized for testing. The results are also determined as a testing section which makes to consider 40% dataset. This remains the 72 news (36 suspicious and 36 True) in the BuzzFeed Category and 96 news (48 suspicious and 48 true) in the category of PolitiFact. The performance of the system is determined in terms of evaluation metrics of recall, precision, and f-measure. The formulations of these terms are mentioned in the table III.

Table III: Evaluation Measures

Evaluation Measure	Formulation
Recall	$\frac{TP}{TP+FN}$
Precision	$\frac{TP}{TP+FP}$
F-Measure	$2 * \frac{Precision\ * Recall}{Precision + Recall}$

The evaluations illustrated in table III indicate that there is need to define the TP, TN, FP, and FN for the determination of evaluation measures. Here, TP (True positive) indicate that the news is predicted as suspicious by proposed ACODSN approach is same as the ground truth values of that news. FP (False Positive) identifies the news that is predicted as suspicious by the proposed ACODSN approach but the ground truth value of that news is true. In a similar manner, TN (True Negative) identifies the news that is predicted as true by proposed ACODSN approach and ground truth value is also true. FN (False Negative) identifies the news that is predicted as true by the proposed ACODSN approach but the ground truth value is suspicious. The values of these measures are evaluated in table IV.

Table IV: Evaluation Results of FP, FN, TP, and TN

	PolitiFact	BuzzFeed
TP	44	33
FN	04	03
FP	05	05
TN	43	31

The table IV results indicate that TP has attained the value of 44 and 33 for the PolitiFact and BuzzFeed categories respectively. Furthermore, TN has attained the value of 43 and 31 for the PolitiFact and BuzzFeed categories respectively.

By considering into account the mentioned results illustrated in table IV, the values of recall, precision, and f-measure are determined. The results for the PolitiFact and BuzzFeed categories are illustrated in table V and table VI respectively along with the comparison with existing concepts. The comparison of proposed ACODSN approach is conducted with the results evaluated by Shu et al. [9] and Vishwakarma et al. [13].

Table V: Evaluated Results of PolitiFact Category

	Recall (%)	Precision (%)	F-Measure (%)
Proposed ACODSN	91.97	89.80	90.87
Shu et al. [9]	89.3	86.7	88.0
Vishwakarma et al. [13]	88.8	87.9	88.34

Table VI: Evaluated Results of BuzzFeed Category

	Recall (%)	Precision (%)	F-Measure (%)
Proposed ACODSN	91.67	86.84	89.19
Shu et al. [9]	89.3	84.9	87.0
Vishwakarma et al. [13]	88.4	85.2	86.77

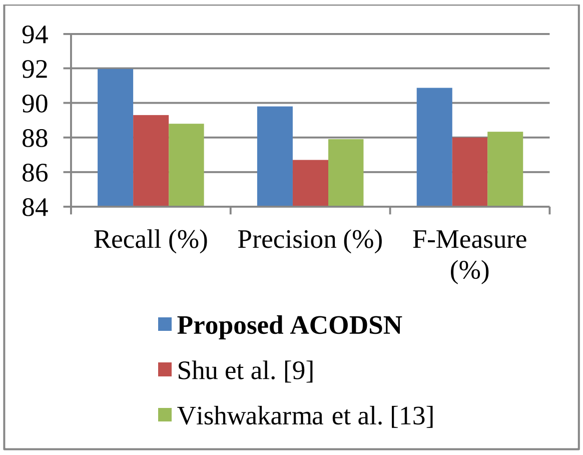

Fig. 4: Comparison of Results based on PolitiFact Category

The evaluated results and comparison in terms of recall, precision, and f-measure are illustrated in table V and table VI for the categories of PolitiFact and BuzzFeed respectively. In case of PolitiFact category, the proposed ACODSN approach has achieved the recall, precision, and f-measure score of 91.97%, 89.80%, and 90.87% respectively. In another category of BuzzFeed, the proposed ACODSN approach has achieved the recall, precision, and f-measure values of 91.67%, 86.64%, and 89.19% respectively. In the other comparison algorithms, the methods discussed by Shu et al. [9] and Vishwakarma et al. [13] lacks than the proposed ACODSN approach. To clearly identify the superiority of proposed ACODSN approach, the comparison is illustrated in the form of graphical representation in fig. 4 and fig. 5 for the categories of PolitiFact and BuzzFeed respectively.

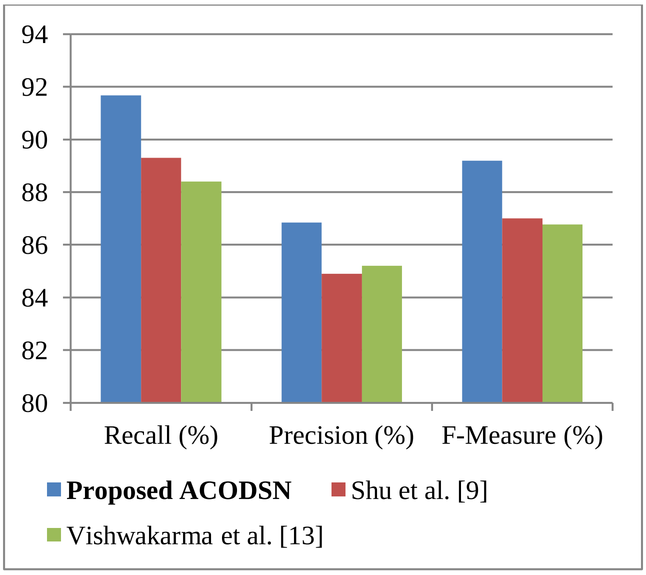

Fig. 5: Comparison of Results based on BuzzFeed Category

The comparison graphs illustrated in fig. 4 and fig.5 clearly indicate the outperformed results of proposed ACODSN approach as compared to considered other concepts. The other concepts lacks in terms of all the evaluation parameters of recall, precision, and f-measure. This indicates the achievement of remarkable result values.

V. CONCLUSION

The growing usage of social media for sharing posts and news has makes it important to analyze and check the authenticity of news. The freedom provided by social media platforms to post anything have also arisen the count of fake news. These kinds of fake news need to be analyze and detect as the more dissemination of such news can lead to any suspicious activity. In this research work, an ant colony optimization based framework for the detection of suspicious news (ACODSN) is discussed. The classification is conducted by exploiting the ACO based ant miner plus classifier. The research dataset of FakenewsNet is utilized for the experimentation which consists of PolitiFact and BuzzFeed categories.

These dataset categories are named after the availability of their ground truth value from the politifact.com and buzzfeed.com respectively. The results of the proposed ACODSN approach are accessed in terms of recall, precision, and f-measure. The evaluated results further compared with existing concepts. These results indicate the outperformed performance of the proposed concept as compared to existing concept.

REFERENCES

1. H. Allcott, and M. Gentzkow, "Social media and fake news in the 2016 election." *Journal of economic perspectives*, Vol. 31, no. 2, 2017, pp. 211-36.
2. Denver Guardian, 2016. Online Available at: https://en.wikipedia.org/wiki/Denver_Guardian. Last Accessed: 06 September, 2019.
3. A. Hern, 2019, “ 'Fishwrap' fake news campaign recycles old news of terror attacks”. Online Available at: https://www.theguardian.com/media/2019/jun/12/fishwrap-fake-news-campaign-recycles-old-news-of-terror-attacks. Last Accessed: 07 September, 2019.
4. X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion.", *Information Processing & Management*, 2019, Online Available at: https://www.sciencedirect.com/science/article/abs/pii/S0306457318306794
5. A. Vlachos and S. Riedel, "Fact checking: Task definition and dataset construction." In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, 2014, pp. 18-22.
6. A. Bondielli, and F. Marcelloni, "A survey on fake news and rumour detection techniques." *Information Sciences*, Vol. 497, 2019, pp. 38-55.
7. N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection." In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ACM, 2017, pp. 797-806.
8. M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier." In *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, IEEE, 2017, pp. 900-903.
9. K. Shu, S. Wang, and H. Liu, "Exploiting tri-relationship for fake news detection." *arXiv preprint arXiv:1712.07709*, 2017.
10. K. Shu, D. Mahudeswaran, and H. Liu, "FakeNewsTracker: a tool for fake news collection, detection, and visualization." *Computational and Mathematical Organization Theory*, Vol. 25, no. 1, 2019, pp. 60-71.
11. C.S. Atodiresei, A. Tănăselea, and A. Iftene, "Identifying Fake News and Fake Users on Twitter." *Procedia Computer Science*, Vol. 126, 2018, pp. 451-461.
12. H. Ko, J. Y. Hong, S. Kim, L. Mesicek, and I. S. Na, "Human-machine interaction: A case study on fake news detection using a backtracking based on a cognitive system." *Cognitive Systems Research*, Vol. 55, 2019, pp. 77-81.
13. D. K. Vishwakarma, D. Varshney, and A. Yadav, "Detection and veracity analysis of fake news via scrapping and authenticating the web search." *Cognitive Systems Research*, Vol. 58, 2019, pp. 217-229.
14. D. Martens, D. Backer, M., Haesen, R., Vanthienen, J., Snoeck, M. and Baesens, B., “Classification with ant colony optimization”, *IEEE Transactions on Evolutionary Computation*, Vol. 11 No. 5, pp.651-665, 2007.

AUTHORS PROFILE

Asha Kumari received her M.tech degree in Computer Science from Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, India. She is currently working as an Assistant Professor in Department of Computer Science, Bhaskaracharya College of Applied Sciences, New Delhi, India. Her research interests include Data mining, machine learning, Software Engineering.

Balkishan received his PhD degree in Computer Science from Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, India. He is currently working as an Assistant Professor in Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, India. His research interests include Soft computing, Artificial Intelligence, Data Mining, Software Engineering.