The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375

Identification of Source of Misleading Information and Stop the Dissemination through Blocking the User

Dr. Girishkumar K. Patnaik
Professor, Department of Computer Engineering, SSBT, COET, Jalgaon,
patnaik.girish@gmail.com

Mr. Akash D. Waghmare
Research Scholar, Department of Computer Engineering, SSBT, COET, Jalgaon,
aakashjan8@gmail.com

Mr. Dinesh D. Puri
Research Scholar, Department of Computer Engineering, SSBT, COET, Jalgaon,
ddpuri@gmail.com

Abstract

Introduction: At present, people are more dependent on Internet sources for any sort of information or news. So, the news/information needs to be preserved and should not be modified by any user. Providing security for the news data is a major concern. The decentralized approach of a chain of blocks is used in order to strengthen the security of the news. The existing blockchain framework that offers openness, tamper-proofing, privacy, controlling information, and monitoring is inherited in the proposed work. Precisely, the idea is to build a safe platform that can detect bogus news on social media platforms. Even if the environment is fragile, the chain of blocks-based decentralised peer-to-peer environment provides security to the published information.

Objectives: As a result of recent innovations and advancements in the field of computer technology, social media networks have emerged as one of the most crucial aspects of contemporary human existence. Social media has developed into a well-known platform for information dissemination and news, as well as for daily reports. There are a variety of benefits associated with social media; but, on the converse, there is a great deal of misleading news and data that can mislead the reader. One of the major issues with social media is that there is a dearth of information that can be relied on as well as real world news. Because of misleading news on social media, users are misled. So, to build a trustful environment, early detection of misleading news is necessary. Innovative machine learning methodologies are useful to identify and recognize misleading news more accurately.

Methods: Misleading news is more viral than real news. People instantly believe on the false information. So, there is a need to reduce the dissemination of misleading information on social media. In order to minimize the spread of misleading news, the source of the news needs to be traced. In overall, proposed system utilizes the chain of blocks and applies proposed machine learning methodologies in order to identify misleading news and thereafter reduce the propagation of misleading information by blocking the fake user.

Results: An experimental analysis reveals that the proposed classification algorithm obtains a better accuracy rate. In order to produce useful training rules and evaluate the test classifier, a number of features are extracted like TF-IDF, N-Gram features, and dependency-oriented NLP features from the data input.

Conclusions: The proposed method analyzes every user's uploaded information, identify fraudulent users and reduce the propagation of false information by blocking the user.

Keywords: Misleading News, Chain of block, Classification, Machine Learning, Peer-to-peer, NLP, TF-IDF, N-gram

1. Introduction

Many activities, such as looking for new data or news, have been redirected to the various online social media platforms in favour of the conventional information sources, such as newspapers

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375
and radio. This is because lives are becoming increasingly centered on internet communication. It has been demonstrated that over the past few years, there has been a significant growth in the number of people who utilize social media as the medium to collect news. In 2012, only 49 percent of individuals were believed to be seeing news on social networking sites; however, currently approximately 70 % of the population is concerned with social media [11]. Although there are numerous benefits offered by social media platforms in contrast to conventional media, the quality of the material must be sacrificed in order to use these platforms [12]. Because it is so simple to share information online, there is a growing problem of people spreading false information for a variety of motives, including political and financial gain. There are various challenges toward recognition of misleading information [14, 15, 16]. This tough subject is being worked on by a large number of scientists and major organizations such as Facebook and Google in an effort to create solutions that would halt the dissemination of harmful data to the general public.

The paper is organized as follows. Section II presents the literature review related to recognition of misleading information and reducing the propagation of misleading information. Section III depicts the framework of proposed system. Section IV demonstrates the proposed method to stop the propagation of misleading information and recognition of misleading information. Section V states the result and discussion. Finally, Section VI concludes the paper and provides future direction.

2. Related Work

This paper focuses on recognition of misleading information and preventing it from spreading. This chapter provides a brief review of the existing research on recognition of misleading information on various social platforms and existing ways to stop the spreading of misleading information. Table 1 shows the literature survey to identify the false information.

Table: 1: Review of Machine Learning in recognition of Misleading Information

Author	Approaches	Feature /Method	Findings
Sheng et al. [9]	KNN	TF-IDF	The optimal neural network system developed for this project can actually manage to reach 90.3 percent accuracy.
Purnendu et al. [8]	Textual entailment recognizer	Tokenization	The suggested technique detects bogus news with an efficiency of much more than 90 percent.
Nihel et al. [7]	SVM	TF-IDF	The top features for detecting fake news stories are text, publisher, origin, period, and emotion.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375

Ethar et al. [5]	Stance based	Google's pre-trained word embedding model	Bidirectional LSTM (BiLSTM) concatenated system has an accuracy rate of 85 percent, followed by the Multi-head LSTM system with an accuracy of around 83 percent.
Fatemeh et al. [3]	CNN	Publisher and User Features	The hybrid algorithm has the potential to improve the efficiency of fake news identification when compared to existing strategies.
Ankit et al. [1]	KNN	User Features	The maximum classification performance attained with this method was 79 percent.
Mayank et al. [6]	RF, SVM, LR, Naive Bayes and K nearest neighbour	Stylometric features set and Word vector features	RF, SVM, LR, Naive Bayes and KNN work well on stylometric variables. Naive Bayes, RF, and LR have worked well on word representations.
Xishuang et al. [10]	Lexical CNN	TF-IDF	A new organizations of two-way deep semi supervised learning method is intended to enable timely identification of bogus information in social networking sites, with one path for supervised learning and another for unsupervised classification

The literature survey to stop the spreading of misleading information shows in Table 2.

Table: 2: Review to halt the Dissemination of Misleading Information

Title	Author	Method/ Algorithm	Findings
Design and implementation of transaction privacy by virtue of ownership and traceability in blockchain based supply chain	Mohit et. al. [18]	Supply chain with encrypted ledger	The suggested solution permits product movement and allows owners to trace the product.
Credulous Users and Fake News: a Real Case Study on the Propagation in Twitter	Alessandro et al. [4]	NLP and Text mining	Explored the behavior as well as extent of participation of gullible users in the spread of misinformation
Tracing the Source of Fake News using a Scalable	Ashutosh et al. [2]	Watermarking -based social	A naive blockchain as well as a watermarking-based social

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375

Blockchain Distributed Network		networking system	networking system are used to limit the spread of bogus news
Tracing the Source of News Based on Blockchain	Wenqian et al. [17]	Blockchain-based	Applying blockchain technology to the process of news tracing has laid a solid foundation for cracking down on fake news.
Smart Contract-Based Product Traceability System in the Supply Chain Scenario	Shangping et al. [19]	Smart Contract-Based	Propose a blockchain-based product traceability system which uses smart contracts to continuously record the movement history of all products on a distributed ledger, forming a chain of blocks that can trace the source of the product.
Fighting fake news spread in online social networks: Actual trends and future research directions	Alina et al. [13]	Human checked and Algorithm checked	User education is very important to increase the success of these algorithms. Users need to know how to spot fake news and not be complicit in spreading it.

3. Objectives

The various perspectives described in the field of recognition of misleading information are reviewed. Several authors have presented numerous machine learning and deep learning approaches to handle the challenge of recognition of misleading information. An approach to detect bogus news with more accuracy and precision is expected. Method is required to recognise the misleading information at an early stage before it spreads.

Misleading information spreads rapidly compared to real news, so a mechanism is required to reduce the spread of misleading information. Once misleading information spreads all over, the real challenge of social media is to identify the source of misleading information. At present, research on user traceability is in the development stage. In some literature, users can verify the authenticity of the source themselves.

4. Methods

In the beginning, an NLP-based method of feature extraction is used for the evaluation of data, and then detect whether the information is real or misleading. The chain of blocks is designed for the decentralized storage of news data.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375

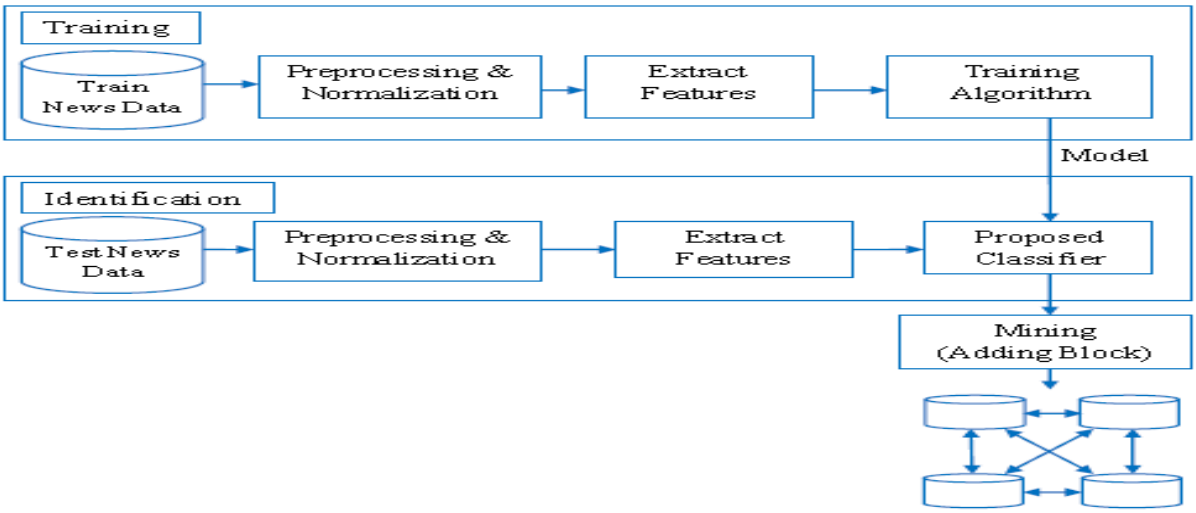

Figure 1 Framework for Recognition of Misleading Information

The framework of recognition of misleading information is shown in Figure 1, including a phase-by-phase breakdown. Initially, during the training phase, news dataset is pre-processed and normalized. Essential features are extracted from news data which helps in the identification of misleading information. Patterns are observed and a classifier model is created using extensive training. In the identification phase, initially, news instances are pre-processed and normalized. Essential features are extracted from news data. Then news instances are fed to the proposed classifier (method) to detect whether the news is real or fake. After recognition of news/information as misleading or real, blocks are created to store news data and are linked together to form a chain of blocks on every node.
Methods are proposed for recognition of misleading information and halt the dissemination of misleading information.

a. Method for Recognition of Misleading Information

The objective of proposed method is to determine whether the information/news is fraudulent or authentic. During the first step of feature selection, every feature is given a weight, and the feature that has the highest weight is selected to move on to the next step. The remaining features within each category are chosen in incremental order in the second step, with correlation between the features taken into account. Weight is applied to choose important features, and correlation is utilized to avoid data redundancy.
Steps during training phase are as follows:
Step 1: Preprocessing & Normalization of Train News Dataset
Step 2: Feature Extraction from Train News Dataset

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375
Step 3: Selection of unique features set
Step 4: Create vectors as per class labels.

The extracted features are used for the training. The proposed model selects unique contents in order to construct the vectors for each class label. Once the model is built, optimization is carried out. In the identification phase of the proposed classification algorithm, features are extracted from the testing instance. Compute the fitness of testing news instance with a training dataset. At every iteration, the computation for fitness includes a comparison of the fitness of one class of vectors to that of another class of vectors. Update the weight of the news instance by determining its maximum fitness. The major objective is to find out the maximum fitness of the news test instance.

b. Method to identify the Source and Stop the Dissemination of Misleading Information

Following news classification, the proposed method read all news from database with all the class labels and User_id. Each user's trust score is generated using the trust calculation function. If the trust_score or misleading information percentage is greater than particular threshold, then block the user and form a group of such users. If trust score/ misleading information percentage is below a particular threshold, then warn or alert the user about the activity. In this algorithm, $\sum$ is the number of news uploaded by the user. Each information instance is classified as misleading or other class label. If information receives a misleading class label, the number of penalties is increased.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375

Algorithm: Algorithm to Halt the Dissemination of Misleading Information
Input: News_info[k], User_id[k], define threshold Th, Legitimate_User[], Misleading _User[]
Output: Updated Legitimate_User[], Misleading_User[] list according to current trust weight of each user.
Step 1: Read all news data including class labels and user id

$$News_Data[] = \sum_{n=1}^{data.length} \left(News[i], User_{id[i]} \dots \dots News[n], User_id[n]\right)$$

Q_p =Number of penalty counts
Q_r =Number of reward counts
$\sum$ = total uploaded news by any user
Step 2: If Information is Real then

$$Trust_{weight} = \frac{\text{Qr (Reward_count)}}{(total_{no}.ofnews\ uploaded\ by\ user)}$$

If Information is Misleading then

$$Trust_{weight} = \frac{\text{Qp(Penalty_count)}}{(total_{no}.ofnews\ uploaded\ by\ user)}$$

Step 3: If ($Trust_{weight}$>= Th)
$Block\ User_{[id]}$
Form Group User[id]
Else
$Warn\ or\ Alert\ User_{[id]}$
End

5.Results

A considerable amount of news is being examined for classification for the purposes of conducting an experiment. With 10-fold cross-validation, the data is split for training and testing, with 70% data selected for training and the remaining 30% for testing. Experimentation for the identification of false information is conducted 10 times and accuracy is observed. An average of ten experiments is considered for the calculation of the accuracy of the proposed method to detect fake news.

The proposed model extracts a number of features, such as TF, TF-IDF, Lemmas, and TF features with co-relational coefficient characteristics, but the Lemmas characteristic seems to provide us the best results in terms of accuracy. The suggested technique is capable of correctly classifying the news as real or misleading. As a result, the proposed approach has a identification rate of 95.20 percent.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375

Figure 2 shows the percentage of misleading information recognition using the proposed method based on Features. The proposed method extracts the features and accordingly, the accuracy is calculated. For a feature like TF (Term Frequency) the accuracy is 89.70, for TF-IDF (Term Frequency–Inverse Document Frequency) the accuracy is 90.25, for TF-CC (TF with co-relational coefficient) the accuracy is 92.60 and for Lemmas the accuracy is 95.20. So the use of Lemma as a feature gives better accuracy. The proposed method uses Lemma as a feature for the recognition of misleading information since it provides the highest level of accuracy.

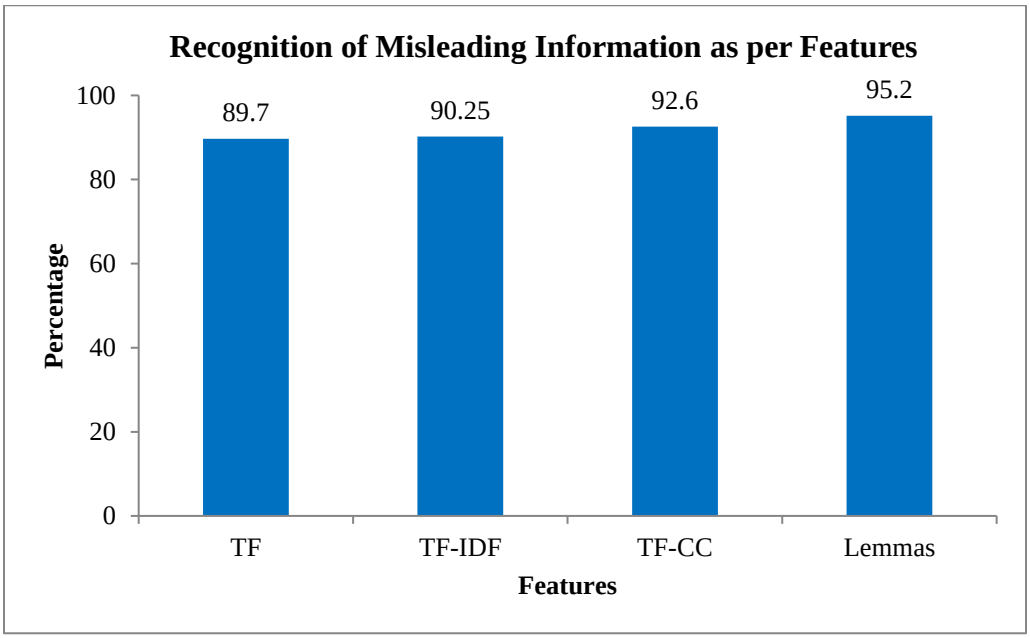

Figure 2 : Recognition of Misleading Information Percentage according to Features

Table 3 shows an example of trust calculation of a user using proposed algorithm with the Threshold considered to block or alert the user is 60%. $Trust_{weight}$ of a user at the time of posting News id 3, 4, 5, 6, and 13 is more than or equal to the threshold. If a user posts misleading news, $Trust_{weight}$ (trustworthiness) is calculated by dividing the number of penalty counts by total number of news items uploaded by the user. The penalty for posting misleading news is one, and the reward for posting true news is zero.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375

Table 3 Trust Calculation of User using Proposed Algorithm

News id	Classified label	Penalty	$Trust_{weight}$
1	Mislead	1	1/1 =1
2	True	0	1/2=0.50
3	Mislead	1	**2/3=0.66**
4	Mislead	1	**3/4=0.75**
5	True	0	**3/5=0.60**
6	Mislead	1	**4/6 =0.66**
7	True	0	4/7=0.57
8	True	0	4/8=0.50
9	Mislead	1	5/9=0.55
10	true	0	5/10=0.50
11	Mislead	1	6/11=0.54
12	Mislead	1	7/12=0.58
13	Mislead	1	**8/13 =0.61**

Table 4 shows the percentage of misleading information spread by Users.

Table 4 Percentage Misleading Information Spreading

User_ID	Trust Weight (Percentage)
User 1	66
User 4	75
User 6	30
User 9	66
User 10	57
User 20	80
User 21	40
User 35	54

Table 4 shows the extent to which each user contributes to the dissemination of misleading information. Users 1, 4, 9, and 20 are responsible for the dissemination of the most misleading news, so they must be prohibited by blocking them. Threshold considered to block the user is 60%. If users are blocked, they are not able to post or spread any information. Additionally, the

proposed method has the ability to block any user or user group that engages in the dissemination of misleading information.

6. Conclusions

Misleading information is a serious problem in today's world that has an impact on individuals or society. So early recognition is required but recognition of misleading information is a major challenge. The proposed method achieves a higher level of precision and accuracy when compared to conventional classification strategies for machine learning. The extraction of various features from textual information improved the performance of the proposed method.

Once news is determined to be phoney, the source of the misleading information must be determined before that information get spread. The developed model emphasizes the stoppage of the dissemination of misleading information by tracing the source of the news/misleading information. A block data structure is used to provide security to the news data and the proposed approach is used to detect and trace the source of false information.

The focus of future research is to develop a classifier that detects misleading information in all regional languages. Because all of the news in the dataset is published in English, the researchers may develop framework for non-English datasets.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Ankit Kesarwani, Sudakar Singh Chauhan and Anil Ramachandran Nair, “Fake News Detection on Social Media using K-Nearest Neighbor Classifier”, International Conference on Advances in Computing and Communication Engineering (ICACCE), IEEE, pp. 1–4, 2020.

[2] Ashutosh Dhar Dwivedi, Rajani Singh, Sakshi Dhall, Gautam Srivastava and Saibal K. Pal, “Tracing the Source of Fake News using a Scalable Blockchain Distributed Network”, 2020, 17th International Conference on Mobile Ad Hoc and Sensor Systems, IEEE, pp. 38–43, 2020. DOI: 10.1109/MASS50613.2020.00015

[3] Fatemeh Torgheh, Mohammad Reza Keyvanpour, Behrooz Masoumi and Seyed Vahab Shojaedini, “A Novel Method for Detecting Fake news: Deep Learning Based on Propagation Path Concept”, 26th International Computer Conference, Computer Society of Iran, Tehran, Iran, IEEE, pp. 1-5, 2021.

[4] Alessandro Balestrucci and Rocco De Nicola. “Credulous Users and Fake News: a Real Case Study on the Propagation in Twitter”, Conference on Evolving and Adaptive Intelligent Systems (EAIS), IEEE, pp. 1-8, 2020.

[5] Ethar Qawasmeh, Mais Tawalbeh and Malak Abdullah, “Automatic Identification of Fake News Using Deep Learning”, Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), IEEE, pp. 383-388, 2019.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 1364 - 1375
[6] Mayank Kumar Jain, Dinesh Gopalani, Yogesh Kumar Meena and Rajesh Kumar, “Machine Learning based Fake News Detection using linguistic features and word vector features”, 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), IEEE, pp. 1-6, 2020.
[7] Ghinadya and Suyanto Suyanto, “Synonyms-Based Augmentation to Improve Fake News Detection using Bidirectional LSTM”, 8th International Conference on Information and Communication Technology (ICoICT), IEEE, pp. 1-5, 2020.
[8] Purnendu Kumar Rath and Rohini Basak, “Automatic Detection of Fake News Using Textual Entailment Recognition”, 17th India Council International Conference (INDICON), IEEE, pp. 1-6, 2020.
[9] Sheng How Kong, Li Mei Tan, Keng Hoon Gan and Nur Hana Samsudin, “Fake News Detection using Deep Learning”, 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE), IEEE, pp. 102-107, 2020.
[10] Kumar, S. A. S., Naveen, R., Dhabliya, D., Shankar, B. M., & Rajesh, B. N. (2020). Electronic currency note sterilizer machine. Materials Today: Proceedings, 37(Part 2), 1442-1444. doi:10.1016/j.matpr.2020.07.064
[11] Xishuang Dong, Uboho Victor and Lijun Qian, “Two-Path Deep Semisupervised Learning for Timely Fake News Detection”, IEEE Transactions On Computational Social Systems, Volume 7, Number 6, pp. 1386–1398, 2020.
[12] Mandal, D., Shukla, A., Ghosh, A., Gupta, A., & Dhabliya, D. (2022). Molecular dynamics simulation for serial and parallel computation using leaf frog algorithm. Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 552-557. doi:10.1109/PDGC56933.2022.10053161 Retrieved from www.scopus.com
[13] H. Allcott and M. Gentzkow, “Social Media and Fake News in the 2016 Election”, Journal of Economic Perspectives, Volume 31, Number 2, pp. 211–236, May 2017.
[14] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake News Detection on Social Media: A Data Mining Perspective”, ACM SIGKDD Explorations Newsletter, Volume 19 Issue 1, pp. 22–36, Aug. 2017.
A. Campan, A. Cuzzocrea, and T. M. Truta, “Fighting fake news spread in online social networks: Actual trends and future research directions”, International Conference on Big Data, IEEE, pp. 4453–4457, 2017.
[15] Nirav Shah, M., Ganatra, A., “A systematic literature review and existing challenges toward fake news detection models”, Social Network Analysis and Mining: Journal, Volume 12, Article Number 168, pp. 1-21, 2022. https://doi.org/10.1007/s13278-022-00995-5
[16] Swapna H. and Soniya B., "A Review on News-Content Based Fake News Detection Approaches," *International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*, Kochi, India, pp. 1-6, 2022, doi: 10.1109/IC3SIS54991.2022.9885447.
[17] Rastogi S. and Bansal D., “A review on fake news detection 3T’s: typology, time of detection, taxonomies”, International Journal Information Security, pp. 177–212, 2022. https://doi.org/10.1007/s10207-022-00625-3

[18] Wenqian Shang, Mengyu Liu, Weiguo Lin, Minzheng Jia,**"**Tracing the Source of News Based on Blockchain", 17th International Conference on Computer and Information Science (ICIS), IEEE, pp. 377- 381, 2018.
[19] Mohit Mohit, Sanmeet Kaur, Maninder Sing, "Design and implementation of transaction privacy by virtue of ownership and traceability in blockchain based supply chain", Springer Science+Business Media, LLC, part of Springer Nature, pp. 2223–2240, 2021.
[20] Shangping Wang, Dongyi Li, Yaling Zhang, and Juanjuan Chen, "Smart Contract-Based Product Traceability System in the Supply Chain Scenario", IEEE Access: Open Access Journal, Volume 4, pp. 1–12, 2016.

Professor & Principal of SSBT's College of Engineering & Technology, Jalgaon. He is recognized PhD Guide in Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. Professional Member in ACM, Life member of ISTE and CSI. His research interests are Wireless Sensor Networks and Security, Machine Learning, Blockchain and Natural Language Processing. He has published 30 research papers in reputed peer reviewed journals in addition to 12 papers in International Conferences to his credit. He is Senior Member in IEEE.

Mr. Akash Dnyandeo Waghmare, completed B.E. and M.E. in Computer Science & Engineering. He is Working as Assistant Professor in Department of Computer Engineering, SSBT's College of Engineering and Technology since 2013. He is pursuing his PhD in Computer Science & Engineering in Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. He is a Life member of ISTE. His areas of interest are Machine Learning, Sentiment Analysis, Data Analytics and Blockchain.

Mr. Dinesh Dagadu Puri, completed B.E from Walchand Engineering College, Sangli. and M.Tech. from DBATU, Lonere in Computer Science & Engineering. He is working as Assistant Professor in SSBT's College of Engineering and Technology since 2012. He is pursuing PhD in Computer Science & Engineering from Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. His areas of interest are Machine Learning, Data Analytics and Natural Language Processing.