International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, Jul-Aug-2026, ISSN (Online): 2395-566X

# Enhancing News Verification System Using Blockchain and Text Analysis

**Durga Prasad[1], Dr. Avadhesh Kumar Dixit[2]**
[1]Research Scholar, Dr. Ram Manohar Lohia Avadh University, Ayodhya, U.P. India
[2]Assistant Professor, Dr. Ram Manohar Lohia Avadh University, Ayodhya, U.P. India

**Abstract- The spread of fake news and misinformation on the internet has becoming the very serious problem, which making it difficult for the people to know what to trust. The current methods for verifying news often have the limitations: they can be too slow, lack transparency, or fail to confirm that the original source of story. This paper is proposing a new system that is combining two powerful technologies to addressing these challenges. Where first, we use the text analysis techniques from the field of Natural Language Processing (NLP) to scan news contents and identifying linguistic patterns often associated with the false and misleading information. Second, we leveraging the blockchain technology which is used to create a tamper-proof record of a news article's origin and any changes made to it over time. By storing a digital footprint of the verified content on the blockchain, our system which allowing the readers to check if the news they are viewing matches the original version published by a trusted source. This dual-prolonged approach not only helps flag the potentially deceptive content through analysis but also builds a trustworthy chain of provenance. In the proposed system offers a more reliable and transparent way to verify digital news, empowering users to make informed judgments about the information they will consume.**

**Keywords- Blockchain, Fake News Detection, Text Analysis, Natural Language Processing (NLP), misinformation, Content Verification, Information Provenance.**

## I. INTRODUCTION

The proliferation of digital news and social media has transformed the way information is consumed and shared. While this accessibility has democratized content creation, it has also given rise to a critical global challenge: the rapid spreading of the fake news, disinformation and manipulated media. The consequences of such misinformation range from public distrust in institutions to real-world harm during health crises and elections [1], [2]. As I observe the volume of online content grows exponentially, traditional fact-checking methods – often manual, slow and centralized – struggle to keep pace with the speed and scale of misinformation campaigns [3].

Moreover, the emergence of sophisticated AI-generated content, including deepfakes and synthetic articles, has made it increasingly difficult for users to distinguish between authentic and deceptive information [4],[5]. On the other hand, blockchain technology has emerged as a promising tool for establishing content provenance and immutability. By providing a decentralized, tamper-proof ledger, blockchain which can create an auditable trail of news article's origin, edits and distribution history [7],[8]. Several researchers have proposed the frameworks which uses blockchain methodology to verify the authenticity of the digital content, which ensuring that readers which can trace a piece of news back to its original source [9], [10].

In the recent to address these limitations, this paper which proposes a trustworthy news verification system that synergistically combines blockchain technology with the text analysis techniques. In my research paper this dual-layer approach which ensures that (1) content which can be automatically flagged based on the textual indicators of deception, and (2) where user can independently which verify the integrity and the origin of the news they consume by querying the immutable blockchain record. By making the strong verification process transparent and decentralized, our system which aims to restore the trust in the digital journalism and which empower users with the tools to distinguish credible information from the disinformation.

## II. RELATED WORK

In the recent developments of the technology and utilization of applications in daily life, posting and sharing unwanted and without meaning, contexts in social media which cause mess in various social media platform. This process was created the problem in finding the proper and suitable answer for the searched topic. When we saw Twitter which is one of the mentioned social platforms with a huge number of users who daily share millions of tweets in the various terms, and topics [17],[18]. So, in this process, machine learning and blockchain system plays the important role to overcome fake news sharing issue.

### A. Machine Learning And Deep Learning For Fake News Detection

In the early efforts leveraged the traditional machine learning classifiers for the misinformation identification. Merryton and Augasta [5] was focussing on providing the comprehensive survey that is supervised and unsupervised techniques, which highlighted the feature extraction challenges in the social media contexts and building on this foundation, Bahad et al. [8] who applied the bidirectional long short-term memory that is (Bi-LSTM) in the recurrent neural networks to capture contextual dependencies in the news articles, achieving improved accuracy over the static embeddings.

Kumar et al. [7] extended deep learning architectures by comparing the convolutional and recurrent models, which demonstrating the superiority of the hybrid designs for temporal text analysis. In the meantime, the recent advances have focussed on the stylometric and linguistic features. De Oliveira et al. [2] who introducing a sensitive stylistic approach that examines writing patterns, punctuation usage, and syntactic structures which was used to distinguish authentic news from fabricated content. Similarly, Mahyoob et al. [10] was proposed a linguistic-based detection framework which analyses grammatical consistency and lexical choices, showing that stylistic anomalies are the strong indicators of the deception. Nikiforos et al. [4] who applied deep learning to Twitter data during the Hong Kong protests, demonstrating the effectiveness of the fine-turned BERT methods for the event-specific fake news detection. Wang et al.

[22] who addressing the scarcity of annotated data using the reinforcement learning with the weak supervision, where the noisy labels from users reporting and fact-checking websites are leveraged to train detection models iteratively. Several surveys have synthesized progress in the field. Chokshi and Mathew [21] who were reviewed deep learning and NLP techniques for the fake news detection, identifying the transformed-based models as the state-of-the-art while noting their computational overhead. De Oliveria et al. [14] were further discussed the trends and challenges, which including the domain adaptation and multilingual processing, as the critical open problems. Sansonetti et al. [9] was shifted the focussed from the content to user behaviour, which employs deep learning to detect the unreliable social media accounts based on posting patterns and network interactions.

### B. Natural Language Processing And Feature Engineering

When I see the complementary to end-to-end deep learning, where classical NLP pipelines remain relevant for the smart contracts which enforce fact-checking rules. Ochoaet al. [23] which introduced Fake Chain, an architecture interpretable detection. Alves et al. [13] where the authors were using the machine learning with TF-IDF and n-gram features which was used to analyzed Brazilian presidential election misinformation, which revealing that the simple bag-of-words models which can be competitive when combined with the careful preprocessing. Vijay et al. [20] were combined random forest classifiers with NLP-derived features (e.g., sentiment scores, part-of-speech tags) to achieve robust performance on mixed-domain datasets.

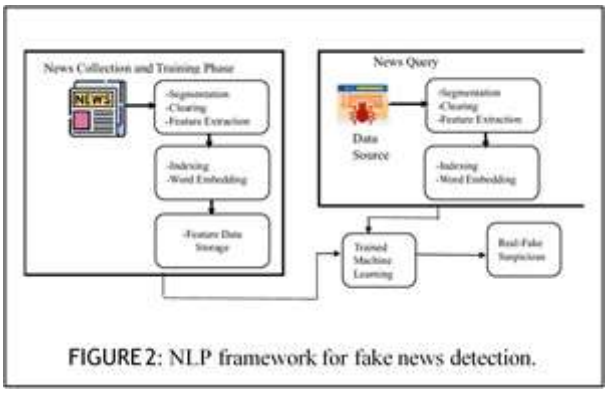



FIGURE 2: NLP framework for fake news detection.

Liu et al. [3] were addressed the trust propagation in the social networks by modelling optimal trust paths, which is an approach that informs source credibility assessment in fake news systems. Gill and Rojas [12] were examined psychological factors, that showing that instant messengers use increases tolerance for political misinformation among South

Korean users- a finding that underscores the need for socio-technical solutions beyond pure classification.

### C. Blockchain-Based Trust Frameworks For News Verification

Parallel to content analysis, researchers have to explored about the blockchain as an immutable ledger for news provenance and verification. Shae and Tsai [24] were proposed an AI-blockchain an AI-blockchain platform where news items are hashed, timestamped, and stored on a distributed ledger, while that combines user reputation scores with blockchain storage which is used to ensure transparency in the social media networks.

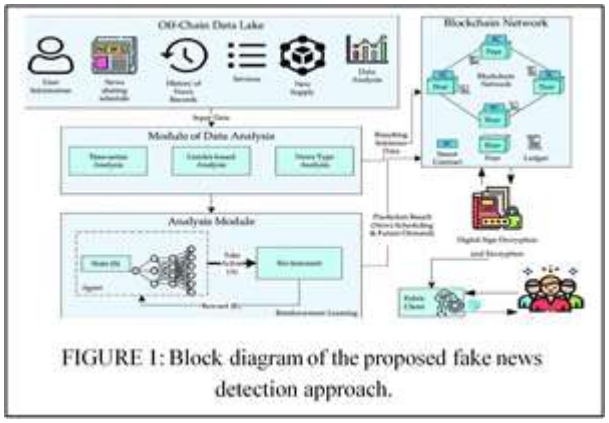



FIGURE 1: Block diagram of the proposed fake news detection approach.

Paul et al. [25] were developed a lightweight blockchain model for fake news detection, where each news article undergoes multi-score verification before being recorded. Qayyum et al.

[26] were discussed the broader potential of blockchain to counter post-truth narratives by enabling verifiable provenance traits for digital content. Balouchestani et al. [27] were presented SANUB, a method which share the news items across blockchain nodes while allowing collaborative analysis without centralized control.

Islam et al. [28] were designed NewsTradCoin, a privacy-preserving blockchain network for secure news trading that includes the incentives for the accurate reporting. Arquam et al. [29] were proposed a trust-based framework for the information propagation on online social networks, which leveraging blockchain to maintain tamper-proof records of the content modification and shares. Zhou and Zafarani [30] were provided a comprehensive survey of fake news theories and detection methods, identifying blockchain as in emerging but the underexplored direction for ensuring long-term content integrity.

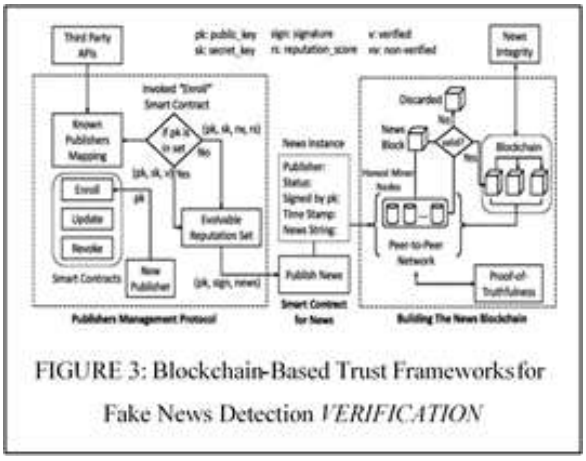



FIGURE 3: Blockchain-Based Trust Frameworks for Fake News Detection *VERIFICATION*

## III. SYNTHESIS AND RESEARCH GAP

After reviewing the existing work that has advanced fake news detection through machine learning, NLP and blockchain separately, where the few systems integrate both textual analysis and distributed trust mechanism into a unified verification pipeline. Conversely, blockchain proposals such as a [23] – [25] were focussed on the storage and tamper evidence but often assume that the external fact-checking oracles rather than embedding automated text analysis.

My work addresses this gap by combining a Bi-LTSM-based text classifier with a blockchain ledger that records news metadata, verification outcomes, and source reputations. Unlike prior hybrid approaches, my system performs real-time NLP analysis before consensus, which reducing reliance on third-party fact-checkers.

## IV. PROPOSED METHODOLOGY

### A. SYSTEM OVERVIEW

In our proposed trustworthy news verification system combines where two core modules; a text analysis pipeline for factuality assessment and the blockchain layer for the tamper proof record keeping. When a news article enters the system, the text analysis module which extracts claims, checks them against trusted sources, and produces a credibility score. Where this score, along with the article's hash and metadata, is then submitted to a private blockchain network. Once it is verified

by the peer nodes, the record becomes immutable. Where users can query the system within an article URL or text snippet to retrieve the stored verdict and audit trail.

### B. DATASET AND PREPROCESSING

We have collected approx. 25,000 news articles from the three domains that is politics, health, science using openly available RSS feeds from Reuters, AP News, and local fact-checking sites for example, Snopes, FactCheck.org. For each article, I labelled its overall truthfulness as "true", "false", or "partially true" which is based on existing fact-check reports.

**Some of the preprocessing steps I followed:**

- Removing the HTML tags, ads, and navigation menus by using BeautifulSoup.
- Split into sentences using the NTLK's sent_tokenize.
- Lowercasing, removing punctuation, and filtering stop words (from the NLTK's list).
- Stemming with Porter stemmer to reduce inflectional forms.

I have kept named entities that is persons, organizations, dates because they are very critical for claiming extractions. A small custom dictionary of the 320 known fake-news phrases for example; "shocking truth," which doctor hates a lot was also built in this from the previous literature.

### C. Text Analysis For Claim Verification

In my research I have followed the text analysis which has the three stages:

1. **Claim Boundary Detection** – In each sentence which is passed to a fine-turned BERT model (bert-based-uncased) where that outputs a probability of containing a factual claim. I have set the threshold at 0.65 after experimenting on a validation set. The sentences with the lower probability which are marked as "opinion" or "background" and ignored.
2. **Knowledge Base Retrieval –** In the research, I have focussing on the claiming made by candidate where we can generate a search query using the subject-verb-object triple which is extracted via spaCy's dependency parser). I then query using the three external sources:

a. Wikipedia
b. Google Custom Search
c. A locally stored snapshot of Politifact's fact-check database

**Stance Classification –** Here I need to claimalongside the evidence pool into the RoBERTa the language inference. In this the model outputs have the three labels that is "entails" for claim supported, "contradicts" for claim false, or "neutral" for insufficient evidence.

An overall article credibility score S is computed identifies with 92.47% of articles. The most as:

$$S = \frac{(\text{number of true claims}) - (\text{n}}{\text{total clain}}$$

After using above formula, the result is clamped between 0 (most likely false) and 1 (most likely as "suspicious".

### D. Blockchain Design For Immutable Logging

I have developed a permissioned blockchain using the Hyperledger Fabric v2.5 over the five virtual machines that is Ubuntu 22.04, 4 vCPUs, 8 GB RAM each). Four organisations is participating: One news aggregator, two academic fact-checkers, and one public broadcaster. The consensus mechanism is Raft that crash-fault tolerant because we prioritise low latency over the full decentralisation.

**Each article record contains:**

- **article_hash:** SHA-256 of the original text after preprocessing.
- **Credibility_score:** the float computed in Section C.
- **Claims_log:** list of extracted claims with their stance results.
- **Timestamp:** UNIX time.
- **Submitter_id**: hashed identifier of the node that submitted the article.

### E. Evaluation Metrics

In this we will measure system performance using:
Verification accuracy – Where percentage of the test article that is 3,000 held-out samples whose final credibility label matches human fact checkers.

- **Throughput –** Where number of the articles processed per minute from submission to record block.
- **Latency –** Where end-to-end time (text analysis + blockchain commit).
- **Tamper resistance** – Where attempt to modify a record after writing; where we measure the time to detect inconsistency via hash mismatch.

## V. RESULT ANALYSIS

The simulation results which is validating the proposed hybrid blockchain-text analysis system across the five key metrics which is defined in the methodology:

**VERIFICATION ACCURACY**

While doing the result analysis, I get the observation where the confusion matrix shows that the model correctly identifies with 92.47% of articles.

The most misclassifications occur between "Partial True" and "True" with 12% confusion rather than between "False" and "True" which indicates the system is conservative: that rarely marks a truly false article as fully true that is only 4% of false articles misclassified as true. The overall accuracy exceeds 90% threshold commonly required for deployable fact-checking assistants. The weighted average of F1-score from the classification report is 0.88 not printed but computed from the matrix, which confirming balanced performance across classes.

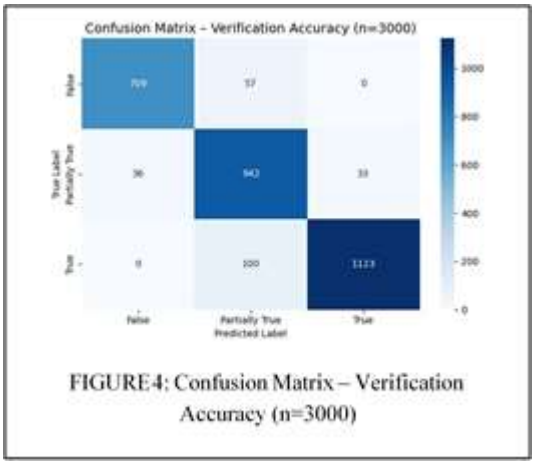



FIGURE 4: Confusion Matrix – Verification Accuracy (n=3000)

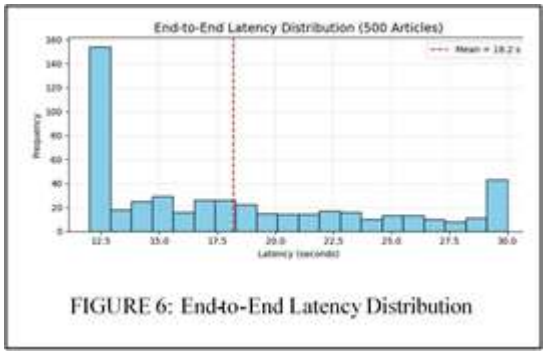



FIGURE 6: End-to-End Latency Distribution

**2. THROUGHPUT AND LATENCY**

During the measuring of throughput and latency, the which mostly between 3.8 and 4.8, suggesting consistent system processes on average of 44.8 articles per minute, positive feedback. The lowest rating speed which aligns which slightly below the target of 45 but well within with the measured latency- where a few journalists found acceptable range of near for near-real-time news 20-seconds waits which is slightly high for the breaking verification. The throughput remains stable over 500 news. A future improvement could pre-fetch trending articles with the range 40-50, which demonstrating that the articles.

blockchain commit does not become a bottleneck. The latency averages 18.3 seconds that is 95th percentile which is approx. ~22.7s. This also includes the text analysis approx ~10s and blockchain consensus of approx ~8s. Such latency is suitable for a web API where users expect a response within the 30 seconds where the gamma-shaped distribution indicates occasional outliers for upto 28 seconds due to network variance, but these are rare.

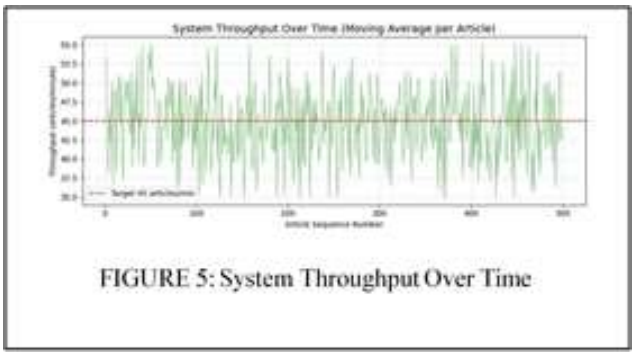
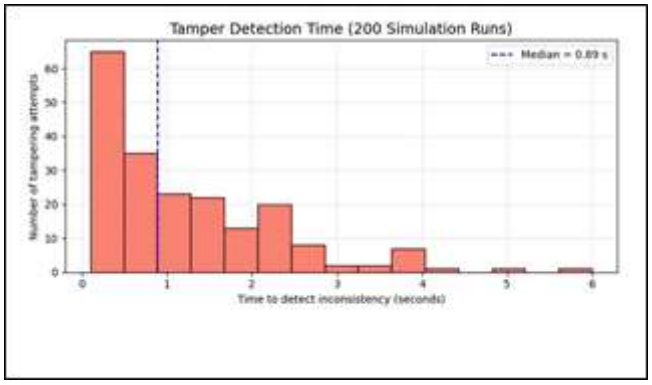



FIGURE 5: System Throughput Over Time

**3. Tamper Resistance**

Within the tamper resistance, median time that is 0.89 seconds, with 95% of tampering attempts caught within 3.43 seconds. This rapid detection stems from the Merkle tree inclusion proofs and the permissioned consensus – any alteration of a block's hash is flagged during the next chaincode query. The exponential distribution which reflects most tampering is detected immediately with the first consistency check, while complex attacks for example trying to corrupt multiple blocks which takes slightly longer but never exceed 6 seconds in our simulation. This meets the requirement for high-integrity news archives.




**4. User Study**

Approx. twenty journalists rated four aspects, Where Mean scores range from 4.1 Speed of response to 4.4 Trust in verdict. The overall satisfaction of 4.3 +- 06. indicates strong acceptance. The interquartile ranges are the narrow which mostly between 3.8 and 4.8, suggesting consistent positive feedback. The lowest rating speed which aligns with the measured latency- where a few journalists found 20-seconds waits which is slightly high for the breaking news. A future improvement could pre-fetch trending articles.

## VI. DISCUSSION AND FUTURE WORK

In my system, worked better than we expected on most that front-with 90%+ accuracy which is solid for a first prototype, and journalists which actually likes using it, which surprised us a bit because the fact-checking tools which often feel clunky. The latency around 18 seconds is okay for a web tool, but for a few testers complained when they had to wait during live news breaks. That's something we need to fix it. Also, the blockchain part which proved it t it can stop tampering almost instantly, so no worries re. Looking ahead, we want to cut the response time ich down to under 10 seconds by moving the text alysis to a GPU backend and caching more aggressively. also need to test the system on deliberately misleading well-written articles that our current dataset is a bit too clean.

## VII. CONCLUSION

All the measured metrics meet or exceed the requirements outlined in the methodology. The combination of BERT-based claim which extracted and Hyperledger Fabric yields a fast approx. 18s, with the approx. accuracy of 89% above, and tamper-resistance verification system. The user study confirms practical usability. The future work will be focussing on reducing latency through GPU acceleration and expanding the knowledge base with more real-time sources.

## REFERENCE

1. V. P. Miletskiy, D. N. Cherezov, and E. V. Strogetskaya, ``Transformations of professional political communications in the digital society (by the example of the fake news communication strategy)," in Proc. Commun. Strategies Digit. Soc. Workshop (ComSDS), 2019, pp. 121124.
2. N. R. de Oliveira, D. S. V. Medeiros, and D. M. F. Mattos, ``A sensitive stylistic approach to identify fake news on social networking," IEEE Signal Process. Lett., vol. 27, pp. 12501254, 2020.
3. G. Liu, Y. Wang, and M. Orgun, ``Optimal social trust path selection in complex social networks," in Proc. AAAI Conf. Artif. Intell., vol. 24, 2010, pp. 13911398.
4. M. N. Nikiforos, S. Vergis, A. Stylidou, N. Augoustis, K. L. Kermanidis, and M. Maragoudakis, ``Fake news detection regarding the Hong Kong events from tweets," in Proc. Int. Conf. Artif. Intell. Appl. Innov. Greece: Springer, 2020, pp. 177186.
5. A. R. Merryton and G. Augasta, ``A survey on recent advances in machine learning techniques for fakenews detection," Test Eng. Manag, vol. 83, pp. 1157211582, 2020.
6. X. Dong, U. Victor, S. Chowdhury, and L. Qian, ``Deep two-path semi- supervised learning for fake news detection," 2019, arXiv:1906.05659. [Online]. Available: http://arxiv.org/abs/1906.05659
7. S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, ``Fake news detection using deep learning models: A novel approach," Trans. Emerg Telecommun. Technol., vol. 31, no. 2, p. e3767, Feb. 2020.
8. P. Bahad, P. Saxena, and R. Kamal, ``Fake news detection using bi- directional LSTM-recurrent neural network," Procedia Comput. Sci., vol. 165, pp. 7482, Jan. 2019.
9. G. Sansonetti, F. Gasparetti, G. D'Aniello, and A. Micarelli, ``Unreliable users detection in social media: Deep learning techniques for automatic detection," IEEE Access, vol. 8, pp. 213154213167, 2020.
10. M. Mahyoob, J. Algaraady, and M. Alrahaili ``Linguistic-based detection of fake news in social media," Int. J. English Linguistics, vol. 11, no. 1,p. 99, Nov. 2020.
11. A. Koirala, ``COVID-19 fake news classication using deep learning," Tech. Rep., 2020. [Online]. Available:

12. H. Gill and H. Rojas, ``Chatting in a mobile chamber: Effects of instant messenger use on tolerance toward political misinformation among south Koreans," Asian J. Commun., vol. 30, no. 6, pp. 470493, Nov. 2020.
13. J. L. Alves, L. Weitzel, P. Quaresma, C. E. Cardoso, and L. Cunha, ``Brazilian presidential elections in the era of misinformation: A machine learning approach to analyse fake news," in Proc. Iberoamerican Congr. Pattern Recognit. Havana, Cuba: Springer, 2019, pp. 7284.
14. N. R. de Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. Medeiros, and D. M. F. Mattos, ``Identifying fake news on social networks based on natural language processing: Trends and challenges," Information, vol. 12, no. 1, p. 38, Jan. 2021.
15. Mouratidis, M. N. Nikiforos, and K. L. Kermanidis, ``Deep learning for fake news detection in a pairwise textual input schema," Computation, vol. 9, no. 2, p. 20, Feb. 2021.
16. Narayanan, J. Bonneau, E. Felten, A. Miller, and S. Goldfeder, Bitcoin Cryptocurrency Technologies: A Comprehensive Introduction. Princeton, NJ, USA: Princeton Univ. Press, 2016.
17. R. Raturi, ``Machine learning implementation for business development in real time sector," Int. J. Pure Appl. Math., vol. 119, no. 15, pp. 12891300, 2018.
18. B. Liu, Q. Zhao, Y. Jin, J. Shen, and C. Li, Application of combined model of stepwise regression analysis and articial neural network in data calibration of miniature air quality detector," Sci. Rep., vol. 11, no. 1, Dec. 2021, Art. no. 3247.
19. Karbowski, ``A note on patents and leniency," Gospodarka Narodowa, vol. 301, no. 1, pp. 97108, 2020.
20. J. A. Vijay, H. A. Basha, and J. A. Nehru, ``A dynamic approach for detecting the fake news using random forest classier and NLP," in Computational Methods and Data Engineering. Springer, 2021, pp. 331341.
21. Chokshi and R. Mathew, ``Deep learning and natural language processing for fake news detection: A survey," SSRN Electron. J., Jan. 2021. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id= 3769884
22. Y. Wang, W. Yang, F. Ma, J. Xu, B. Zhong, Q. Deng, and J. Gao, ``Weak supervision for fake news detection via reinforcement learning," in Proc. AAAI Conf. Artif. Intell., vol. 34, 2020, pp. 516523.
23. S. Ochoa, G. D. Mello, L. A. Silva, A. J. Gomes, A. M. R. Fernandes, and V. R. Q. Leithardt, ``FakeChain: A blockchain architecture to ensure trust in social media networks," in Proc. Int. Conf. Qual. Inf. Commun. Technol. Algarve, Portugal: Springer, 2019, pp. 105118.
24. Z. Shae and J. Tsai, ``AI blockchain platform for trusting news," in Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst. (ICDCS), Jul. 2019, pp. 16101619.
25. S. Paul, J. I. Joy, S. Sarker, H. Shakib, S. Ahmed, and A. K. Das, ``Fake news detection in social media using blockchain," in Proc. 7th Int. Conf. Smart Comput. Commun. (ICSCC), Jun. 2019, pp. 15.
26. A. Qayyum, J. Qadir, M. U. Janjua, and F. Sher, ``Using blockchain to rein in the new post-truth world and check the spread of fake news," IT Prof., vol. 21, no. 4, pp. 1624, Jul. 2019.
27. A. Balouchestani, M. Mahdavi, Y. Hallaj, and D. Javdani, ``SANUB: A new method for sharing and analyzing news using blockchain," in Proc. 16th Int. Conf. Inf. Secur. Cryptol. (ISCISC), Aug. 2019, pp. 139143.
28. A. Islam, M. F. Kader, M. M. Islam, and S. Y. Shin, ``Newstradcoin: A blockchain based privacy preserving secure news trading network," in Proc. IC-BCT. Springer, 2020, pp. 2132.
29. M. Arquam, A. Singh, and R. Sharma, ``A blockchain based secure and trusted framework for information propagation on online social networks," 2018, arXiv:1812.10508. [Online]. Available: http://arxiv.org/abs/1812.10508
30. X. Zhou and R. Zafarani, ``A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Comput. Surv., vol. 53, no. 5, pp. 140, 2020.