

Research Paper

Fake Media Detection Based on Natural Language Processing and Blockchain Approaches

¹P. Sandhya Rani ¹MTech, ²Dr. V. Anantha Krishna, ²Professor ³Dr. U. Srilakshmi ³Professor & HOD

Department of Computer Science and Engineering,

Sridevi Women's Engineering College, Gopandpally, Hyderabad, Telangana, India¹²³⁴

Email- sandhyaraniparmangari@gmail.com

Abstract: Social media networks are critical in the conveyance of data and information in today's digital world. Despite their many benefits, there is one big issue with fake and misinformation news: it exists. Reliable content is vital for the credibility of social media platforms and is a serious threat to social media users who seek factual information. To tackle this problem we present an integrated system using blockchain technology and NLP - enhanced with ML techniques - to successfully detect and fight bogus news. We use Reinforcement Learning to enhance the system for detecting fake user accounts and fake posts. This is more effective for safeguarding the integrity of information distributed through social media networks. We additionally deploy the decentralised blockchain architecture in our approach, increasing the safety of the system and generating undeniable proof of the right of the digital contents. The main goal of this system is to build a safe platform for forecasting and recognising bogus news in order to create a more safe and dependable social media ecosystem. We aim to integrate these technologies in a novel manner which might help alleviate the dissemination of misinformation and help people get accurate, reliable information.

“INDEX TERMS: Natural language processing, blockchain, fake media, reinforcement learning.”

1. INTRODUCTION

Right now, the amount of information shared is crucial to our social media landscape. However, in the midst of all this content, the emergence of false news is a big issue and a persistent one. The use of fake news has increased exponentially since 2017, as seen in the increased amount of content online – 365% more [1]. The proliferation of misinformation not only threatens the integrity of the social media platforms but also has major consequences in diplomatic, economic and political arenas.

The issue of fake news propagation is a multidimensional problem in the social networks' fabric, affecting the consumption and propagation of data and information. This problem not only consumes the network resources but also affect the integrity and authenticity of offered services [2]. Thus, the concept of QoT is gaining significance in the assessment of the credibility of news providers: [3].

A sophisticated technological solution is needed to combat the issue of fake news, while ML text classification is an essential instrument in boosting security in social media networks. Social networking portals abound, and misinformation and bogus accounts are breeding grounds, typically from non-governmental organisations with an agenda to sway public opinion [4].

One of the major areas affected by fake news is political propaganda. Misinformation is intentionally created to elicit emotional responses, and to shape public dialogue [5]. Fake news articles are designed to evoke a specific emotional response, and to provide false information [6]. To detect fake news, complex algorithms are needed to be able to distinguish between truth and fabrication in the content analysis [7].

With an exponential growth in the amount of unstructured data and the number of users and news items, the need for automated solutions in fake news detection [8] becomes even more pressing. This problem can be solved by recent advancement of ML, DL and AI [9]. However, if these solutions are going to be successful, one of the key challenges will be determining the authenticity of digital content – and that's where blockchain technology comes in.

The decentralised and secure framework offered by blockchain technology can be useful to make the content shared on social media platforms more trustworthy. The integrity and authenticity of the data are maintained on the blockchain by sequentially appending a chain of blocks with cryptographic hashes, time stamps, and transactional data [10]. This intrinsic characteristic of blockchain makes it a

suitable platform to battle misinformation and create trust in online information [11].

To sum up, fake news is an important issue in social media networks, which has influenced several aspects of the information consumption and sharing process. With the help of cutting-edge machine intelligence, DL and blockchain technology, this threat can be mitigated and trust can be restored in digital content. By working together with researchers, industry and regulators, the tide can be turned towards a more secure and trustworthy social media environment.

2. LITERATURE SURVEY

In recent years, fake news has gained a lot of attention, as it is common on social media and has a significant impact on public opinion and discussion. Several methods and techniques have been explored in a battle against this problem. In this literature analysis, we summarize recent studies and latest developments within the field of fake news detection, focusing on ML methods, stylistic methods and social media's role in the spread of fake news.

Miletskiy et al. [1] discuss the changes in professional political communication in the digital society, with a special emphasis on communication strategy of fake news. The findings of this study highlight the transformation of political communication and the problems resulting from the use of fake news.

In this paper, [2] present a sensitive stylistic technique that they have developed to detect bogus news on social networks. The aim of the research is to give a solid basis for the identification and prevention of misinformation propagation through language aspects and stylistic differences.

Nikiforos et al. [4] are interested in identifying fake news about the Hong Kong events from the Twitter stream. The study investigates the potential of utilising ML algorithms for the detection and classification of fake news especially in the area of socio-political issues.

In this regard, Merryton and Augusta [5] present the latest advances of ML approaches for Fake News detection. The paper surveys the latest algorithms and approaches for false news detection and mitigation in various digital platforms.

By Dong et al. [6] a deep two-path semi-supervised learning approach to identifying fake news is proposed. The aim of this paper is to enhance the performance and efficiency of fake news detection systems, particularly in the scenario of scarce labeled data, using DL techniques.

To tackle the problem of fake news identification, Kumar et al. [7] came up with a novel approach using DL models. In this work, we try to leverage the power of deep neural networks to deliver strong and scalable solutions to fake news identification and mitigation on online platforms.

Bahad et al. [8] study a bi-directional LSTM-recurrent neural network to detect bogus news. The research has shown the potential of recurrent neural networks in the learning of the temporal dependency and the relation between semantic words in text, so as to increase the accuracy of false news detection systems.

Sansonetti et al. [9] examine the detection of unreliable users in social media with the help of DL methods. The study will aim to develop automated tools that can detect suspicious users and accounts

who are spreading fake news based on user interaction and behavioural patterns.

A review of the literature also shows the different strategies and methodologies employed to identify false news, including linguistic and stylistic methods, DL and ML. These advances will be employed to build strong and scalable solutions to prevent the spread of disinformation in social media networks.

3. METHODOLOGY

a) Proposed Work:

The proposed system aims to address the challenges of current false news detection algorithms [21] through the use of advanced technologies and novel approaches.

The technology will use cutting-edge ML models trained on multiple datasets to improve the detection of fake news. To detect misinformation more accurately, the models will use NLP [19] techniques to explore the linguistic characteristics of the social media content.

The system will use Blockchain [16] technology to guarantee the content origin and authenticity with an immutable record, to foster transparency and accountability.

The solution we propose will enable users to validate & verify content via community driven procedures.

Privacy will be a key focus of the proposed system with privacy-preserving design principles and minimal gathering of personally identifying information.

The proposed approach takes these factors into account and presents a comprehensive solution to solve the problem of identifying and fighting fake news in social networks. By leveraging the power of blockchain technology, sophisticated ML algorithms, and community-driven validation procedures, the system aims to promote a more reliable and robust digital landscape, where truth triumphs over falsehoods.

b) System Architecture:

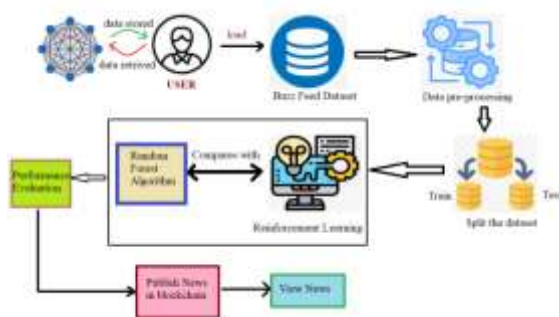


Fig1 Proposed Architecture

The suggested system architecture consists of many important components to tackle the problem of fake news identification and propagation in social media networks.

First, the architecture starts with loading the Buzz feed dataset which is used to train and test the models for fake news identification. Next, the data pre-processing techniques are used to remove data from the dataset and get it ready for analysis.

This is then divided into training and test sets to use for training and testing ML models. Reinforcement learning algorithms are used and compared with RF [20] algorithms to boost the accuracy and effectiveness of fake news identification.

The performance evaluation indicators are used to assess the trained models in terms of their accuracy in detecting and categorizing false news material. Then, certified news stories are published on a Blockchain [16] network to assure transparency, immutability and traceability of the information.

The published news articles are stored on the blockchain, allowing users to trust and verify the information provided. Information saved and retrieved in the blockchain network assures the quality and integrity of published info, reduces the spread of false news [22] and boosts transparency of social media networks. This all-encompassing system architecture is intended to improve online information dissemination's credibility and address the problem of false information spreading.

c) New User Signup Here:

The new module Signup allows for the streamlined sign up and registration process. Users enter their information, and these contracts, recognised for their transparency and immutability, safely preserve it on the blockchain. The strategy establishes confidence and reliability by safeguarding users' information on the platform and ensuring its security and reliability.

d) User Login Here:

The "User Login" module is meant to provide users registered to the system with their login. After logging in, users are able to access their personal accounts as well as access the services provided by the platform. This allows people to interact with the system and participate in tasks such as recognising bogus information. The module ensures a smooth experience for users entering, allowing for them to

participate and collaborate in combating false information.

1. Load Buzz Feed Dataset:

The "Load Buzz Feed Data Set" module loads any data set, including the Buzz Feed data set. This is a crucial step in training and test the false news detection models as it delivers a set of relevant training examples which facilitates the precision of the trained models and encourages the use of DL algorithms.

2. Train Reinforcement Learning:

The "Train Reinforcement Learning" module is used to train the RL algorithm with the loaded data. The system analyses news articles to see if they are genuine or bogus and enhances categorisation results. This training process is essential to enhance the accuracy of the algorithm and ensure uniformity of the detection results.

3. Publish News in Blockchain:

After verification by the reinforcement learning system, they can publish news to the blockchain by using the "Publish news in blockchain" module. The process of distribution is made secure, transparent and immutable with smart contracts to store each item. An integral part of the battle against fake news.

4. View News:

The "View News" module allows users to access and browse the list of all news that have been posted on the blockchain. To promote transparency and verification of authenticity of content stored in the blockchain, the users can access the articles stored in

the blockchain, as well as the classification (authentic or not).

e) Blockchain Integration:

The blockchain acts as a decentralised and secure storage mechanism, with each news article being stored as a block or transaction. This will stop unauthorized edits and ensure that news data can't be changed. Decentralisation of the blockchain boosts security and transparency, which is crucial in combating fake news.

User details & news details are stored securely using smart contracts that are deployed on the Ethereum blockchain. These contracts specify functions for storing and retrieving information in a secure manner. The solution uses Ethereum's blockchain for a trustless and transparent method to manage user and news data making it more secure and reliable.

Every block or transaction in the blockchain has a hashcode that is used to validate and guarantee the validity and integrity of the news data recorded in the block. The hashcodes are computed and verified before new blocks are added. This is to ensure accuracy of data and to avoid any unauthorised alteration of data. The technique adds an additional layer of security and confidence in the content of news stored.

4. EXPERIMENTAL RESULTS

To save or access data from Blockchain, we need to build SMART CONTRACT which will include function to save and READ data, below is the SMART CONTRACT for Fake News application.



Let's publish this contract in Blockchain server, for this purpose follow the below steps, above screen defined functions of smart contract for USER & NEWS so now we have to publish this contract in Blockchain server.

1) Inside the 'hello-eth/node_modules/.bin' folder, just double click the 'runBlockchain.bat' file, you will get the following screen.



After Blockchain generated Private keys and default account address, type the command 'truffle migrate' and press enter key then it will deploy the contract and give below output.



In above screen in white colour text we can see Blockchain Fake News contract launched and we

got contract address also and this address we need to specify this address in python program to access above contract to store and get data from Blockchain.

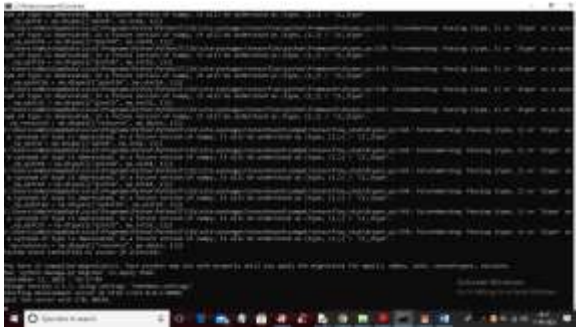


For below screen read, comments are in red colour, which shows as to how to call Blockchain function to store and read data using python program.

Once deployed, double click on 'Start_IPFS.bat' to start IPFS server to store the image of the program and get the following screen:



Now, above mentioned IPFS server is started, double click on 'runServer.bat' file and start python server to get following output:



You will see that above screen python server has been started and the browser is opened and you type url as 'http://127.0.0.1:8000/index.html' and press enter key to get below home page.



In above screen in red colour text we can see user details saved in Blockchain and now click on 'User Login Here' to reach below login screen Next step is to login



The above screen displays 'New User Signup Here' which will get you the following screen for signing up.



Above Screen user will be login, after login will be given below screen.



The user enters the signup details in the above rectangle and then clicks the 'Submit' button and the output below will appear:



If screen user clicks on 'Load Buzz Feed Dataset' button, then he/she will be able to load the dataset and get below result.



Now close above graph to get below output in which x-axis represents NEWS TYPE and y-axis contains number of Fake or Real NEWS type.

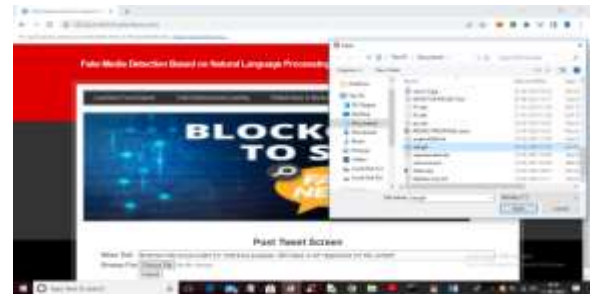


We can see that above that we have imported the news from the data set and now click on the link 'Train Reinforcement Learning' and we get below output after training the algorithm of Reinforcement.



As shown above, the data-set is trained with existing Random Forest and proposed Reinforcement algorithm and then we received RMSE and MAE, reward and penalty for both algorithms. Existing random forest don't support REWARD and Penalty.

We can observe Propose reinforcement obtained smaller RMSE and MAE compare with existing algorithm. Minimizing the value of both RMSE and MAE will be better algorithm. After training, algorithm will be ready to be used, click the 'Publish News in Blockchain' link to publish news then Reinforcement will classify it as Fake or Real.



In above code, I wrote some NEWS, and Uploaded Photo and then click on 'Open' and 'Submit' button to detect news as Fake or Real and then store in Blockchain and will get below output



As seen in above page, News is stored in Blockchain and we have retrieved the hashcode of News Storage and Transaction storage DELAY and now click on 'View News' link, to see the LIST of all news published by all users.



On the above screen, the names of USERS who submit news will be displayed and then the output of detection will be displayed as FAKE or REAL. If you want to upload or test other news, you can also do so.

Note: If there is any TEST news data then you can copy some lines from 'Test_News.csv file' or else you can copy few lines from the provided TEST News data.

5. CONCLUSION

To sum up, the NLP approach such as NLTK and reinforcement learning can be a powerful tool for detecting real and fake news. The combination is able to give a more precise automatic solution, essential in the dissemination of false information. The use of NLP, which provides the contextual knowledge, and reinforcement learning, which provides learning skills, enable the system to be able to effectively distinguish between the real and fake news articles.

Moreover, blockchain technology offers a high degree of confidentiality and integrity to store news data. Blockchain [16] assures the tamper-proof storage by giving each transaction a distinctive block hashcodes [11], which improves data security and reliability and prevents from illegal modifications. This is important to maintain news material validity and gain user trust on the information provided.

Moreover, the comparison between the reinforcement learning model and previous models like the RF indicates the superiority of reinforcement learning model in terms of reduction of RMSE and MAE. This shows the strength of reinforcement learning in building more powerful and accurate fake news detection models, and gives hope for the future of fake news detection by automation. There are, however, modern solutions and technologies, such as blockchain or reinforcement learning, that enable us to decrease the negative effects of misinformation on our societies and pave the way for a more trustworthy and accurate news environment.

6. FUTURE SCOPE

Looking ahead, potential developments in false news identification might involve the incorporation of multimodal analysis algorithms which would go beyond just textual information to also incorporate photographs, videos and audio samples available on social media sites. This can help the system to grasp disinformation in a more complete way, which can lead to a higher accuracy of detection.

Besides, new technologies such as DL, reinforcement learning and federated learning hold the promise to improve the fake news detection [23] capacities. The future system recommended in this document could leverage these technologies to improve detection accuracy, as well as scalability and adaptation to new threats. Incorporating these developments allows the system to stay ahead of the curve and effectively address the spread of disinformation in a more complicated digital context.

REFERENCES

- [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. 121–124.
- [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. 1250–1254, 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. 1391–1398.
- [4] M. N. Nikiforos, S. Vergis, A. Styliadou, 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. 177–186.
- [5] A. R. Merryton and G. Augasta, "A survey on recent advances in machine learning techniques for fake news detection," Test Eng. Manag., vol. 83, pp. 11572–11582, 2020.
- [6] X. Dong, U. Victor, S. Chowdhury, and L. Qian, "Deep two-path semisupervised 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 bidirectional LSTM-recurrent neural network," Procedia Comput. Sci., vol. 165, pp. 74–82, 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. 213154–213167, 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 classification using deep learning," Tech. Rep., 2020. [Online]. Available: <https://www.cs.ait.ac.th/xmlui/handle/123456789/981>
- [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. 470–493, 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. 72–84.
- [14] N. R. de Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. de 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] D. 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] A. 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. 1289–1300, 2018.
- [18] B. Liu, Q. Zhao, Y. Jin, J. Shen, and C. Li, "Application of combined model of stepwise regression analysis and artificial neural network in data calibration of miniature air quality detector," *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, Art. no. 3247.
- [19] A. Karbowski, "A note on patents and leniency," *Gospodarka Narodowa*, vol. 301, no. 1, pp. 97–108, 2020.
- [20] J. A. Vijay, H. A. Basha, and J. A. Nehru, "A dynamic approach for detecting the fake news using random forest classifier and NLP," in *Computational Methods and Data Engineering*. Springer, 2021, pp. 331–341.
- [21] A. 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. 516–523.
- [23] I. 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. 105–118.
- [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. 1610–1619.
- [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. 1–5.
- [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. 16–24, 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. 139–143.