

FAKE MEDIA DETECTION BASED ON NATURAL LANGUAGE PROCESSING AND BLOCKCHAIN APPROACHES

¹MUGALA YESHWANTH REDDY, ²VANGALA SHRUTHI, ³SUNKARA NIKHIL,
⁴Dr. Bagam Laxmaiah

^{1,2,3} B.Tech Student, Department of Computer Science and Engineering, CMR Technical
Campus, Medchal, Hyderabad, Telangana, India

¹207r1a05a1@cmrtc.ac.in, ²207r1a05b7@cmrtc.ac.in, ³207r1a05a6@cmrtc.ac.in

⁴Professor, Department of Computer Science and Engineering, CMR Technical Campus,
Medchal, Hyderabad, Telangana, India, blaxmanphd@gmail.com

ABSTRACT:

Social media network is one of the important parts of human life based on the recent technologies and developments in terms of computer science area. This environment has become a famous platform for sharing information and news on any topics and daily reports, which is the main era for collecting data and data transmission. There are various advantages of this environment, but in another point of view there are lots of fake news and information that mislead the reader and user for the information needed. Lack of trust-able information and real news of social media information is one of the huge problems of this system. To overcome this problem, we have proposed an integrated system for various aspects of blockchain and natural language processing (NLP) to apply machine learning techniques to detect fake news and better predict fake user accounts and posts. The Reinforcement Learning technique is applied for this process. To improve this platform in terms of security, the decentralized blockchain framework applied, which provides the outline of digital contents authority proof. More specifically, the concept of this system is developing a secure platform to predict and identify fake news in social media networks.

Keywords: *Block chain, NLP, fake news.*

1. INTRODUCTION:

Variety of shared information is the realistic part of social media. From 2017, fake news has become a very considerable topic until now, which 365% frequently used online [1]. Struggling with fake news becomes an unsolved problem in social networks in the data and information consumption application layer and

becomes a serious and challenging issue in information advancement that appears in diplomatic, economic, and political sectors. The fake information revelation point to the unnecessary process in the network resources. Moreover, it contains the content totality and validity based on the available service [2].



Therefore, the wrong information sharing relevance the Quality of Trust (QoT) to apply on the news distribution [3].

Machine learning text classification improves the level of security that is needed in social media daily-based networking. Expressing feelings or sharing an opinion through the social networking portal from the non-government organization's survey contains many fake accounts and information circulating the portal based on a suitable channel. In this case, the harmful and unwanted accounts need to pass from the network to give more space to the data center and manage the mess and political problems in the network.

Another related area for information extraction is propaganda which is special for political purposes [4]–[6]. The fake news forging language is very crafty in terms of predesignate to arouse and aggravate the emotion of users for spreading fake information [7]–[9]. Fake news detection is the capability of contents analysis based on truth in the shared information [10]–[12]. Along with the number of noisy and unstructured data, growth of the number of users, and news, there is a need for an automatic solution for extraction of fake news [13]–[15]. These terms become limited based on the recent developments in machine learning, deep learning, and artificial intelligence. Proofing the digital contents authorship is one of the mandatory steps for information sharing. To do this, blockchain is a suitable and promising framework that is the decentralized and secure platform to improve fake information extraction. A blockchain system continuously increases the number of blocks which each block has the previous block cryptographic hash, timestamp and transactions information [16]. The data integrity is guaranty with the blockchain and all

the transnational information store in it. This aspect of blockchain makes it a famous platform in this approach

As a case study, we collected the social media contents from Facebook and Twitter, which are famous information sharing platforms with thousands of users that upload millions of daily news and posts on various topics. This research aims to authorize fake users and information using the blockchain, NLP, and machine learning techniques. More specifically, the proposed system is the preventative approach based on the integrated techniques for the concept of fake data extraction combining with gamification components. Reinforcement learning is the learning-based algorithm that improves the system quality based on the provided information. If the information is wrong, the system prevents using similar information as before to reduce the fake and wrong information rating. The main contribution of this paper is threefold:

Designing the fake news prevention system instead of a detection system and applying the Natural Language Processing (NLP) for the detailed text analysis based on the shared contents.

- Applying the proof of authority protocol and designing financial roots. This process is the strong aspect of this system to find fake user information and accounts.
- Applying the Reinforcement Learning technique for predicting the learning rate of the system and extracting fake accounts. Finding the relationship between contents, extracting the similar meaning and structure of the shared information to avoid sharing fake news.



Figure 1 shows the overview of the proposed approach in fake news detection based on the integrated method. The applied blockchain system is permission network that every participants are supposed to register and required authentication to make them qualified to join to blockchain network. In permission blockchain only authenticate users have allowance of joining to network which this process is the responsibility of user identification manager. This process also required the authentication certificate and enrollment for the valid participants. The aim of the proposed system is to store the news data in the distributed ledger which is reliable and secure platform.

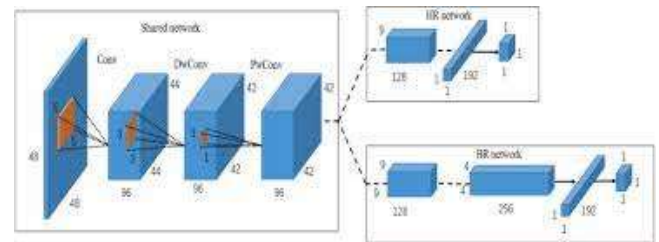
2. EXICITING SYSTEM:

Variety of shared information is the realistic part of social media. From 2017, fake news has become a very considerable topic until now, which 365% frequently used online [1]. Struggling with fake news becomes an unsolved problem in social networks in the data and information consumption application layer and becomes a serious and challenging issue in information advancement that appears in diplomatic, economic, and political sectors. The fake information revelation point to the unnecessary process in the network resources. Moreover, it contains the content totality and validity based on the available service [2]. Therefore, the wrong information sharing relevance the Quality of Trust (QoT) to apply on the news distribution [3].

PROPOSED SYSTEM:

The Overview Of The Proposed Approach In Fake News Detection Based On The Integrated Method. The Applied Blockchain

System Is Permissioned Network That Every Participants Are Supposed To Register And Required Authentication To Make Them Qualified To Join To Blockchain Network. In Permissioned Blockchain Only Authenticate Users Have Allowance Of Joining To Network Which This Process Is The Responsibility Of User Identification Manager. This Process Also Required The Authentication Certificate And Enrollment For The Valid Participants. The Aim Of The Proposed System Is To Store The News Data In The Distributed Ledger Which Is Reliable And Secure Platform.



3. METHODOLOGY

Upload credit crad dataset : use this module to get Upload credit crad dataset

Normalize & dataset split : use this button to get on Normalize & dataset split.

Train auto encoder & decoder model : to encode and decode use this button to get Train auto encoder & decoder model

Extract encoder & decoder prediction : to extract encoder & decoder use this module to get Extract encoder & decoder prediction.

MAE histogram on fraud transaction: use this button to get MAE histogram on fraud transaction



IJARST

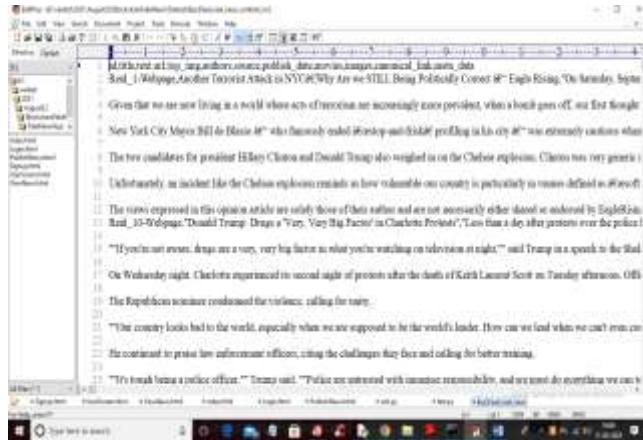
International Journal For Advanced Research In Science & Technology

A peer reviewed international journal

ISSN: 2457-0362

www.ijarst.in

To implement this project author has used BUZZ NEWS and many more dataset but we are using BUZZ news dataset only and below is the dataset screen shots



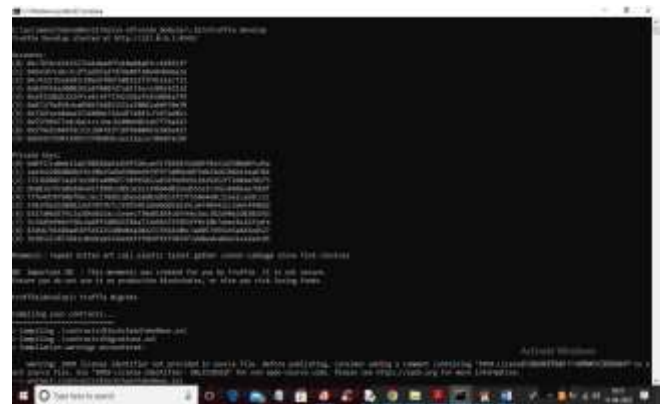
In above screen first row contains dataset column names and remaining rows contains dataset values and by using above dataset we will train Reinforcement algorithm.

To store or access data from Blockchain we need to develop SMART CONTRACT which will contains functions to STORE and READ data and below is the SMART CONTRACT for Fake News application.

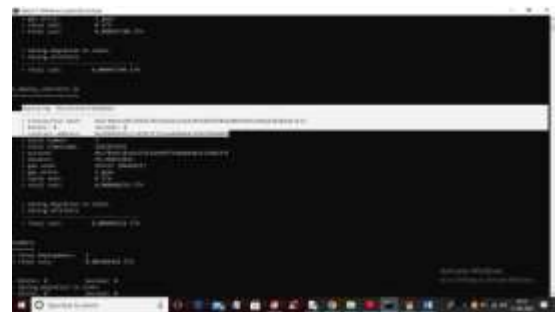


In above screen we have define smart contract functions to store USER & NEWS details and we need to deploy this contract in Blockchain server and for deployment we need to follow below steps

- 1) First go inside 'hello-eth/node_modules/.bin' folder and then double click on 'runBlockchain.bat' file to get below screen



In above screen Blockchain generated Private keys and default account address and now type command as 'truffle migrate' and then press enter key to deploy contract and get below output



In above screen in white colour text we can see Blockchain Fake News contract deployed 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 read data from Blockchain.



```

1 # Importing the libraries
2 import numpy as np
3 import pandas as pd
4 import tensorflow as tf
5 from tensorflow.keras.preprocessing.text import Tokenizer
6 from tensorflow.keras.preprocessing.sequence import pad_sequences
7 from tensorflow.keras.models import Sequential
8 from tensorflow.keras.layers import Dense, LSTM, Embedding
9
10 # Loading the dataset
11 data = pd.read_csv('data.csv')
12
13 # Tokenizing the text
14 tokenizer = Tokenizer()
15 tokenizer.fit_on_texts(data['text'])
16
17 # Converting the text to sequences
18 sequences = tokenizer.texts_to_sequences(data['text'])
19
20 # Padding the sequences
21 max_length = 100
22 padded_sequences = pad_sequences(sequences, maxlen=max_length)
23
24 # Building the model
25 model = Sequential()
26 model.add(Embedding(max_length, 100))
27 model.add(LSTM(100))
28 model.add(Dense(100))
29 model.add(Dense(1))
30
31 # Compiling the model
32 model.compile(optimizer='adam', loss='binary_crossentropy')
33
34 # Training the model
35 model.fit(padded_sequences, data['label'], epochs=100)
36
37 # Saving the model
38 model.save('model.h5')
39
40 # Loading the model
41 model = tf.keras.models.load_model('model.h5')
42
43 # Predicting the class
44 predictions = model.predict(padded_sequences)
45
46 # Converting the predictions to binary
47 binary_predictions = (predictions > 0.5).astype(int)
48
49 # Printing the predictions
50 print(binary_predictions)

```

In above screen read red colour comments to know about how to call Blockchain function to store and read data using Python program.

Now after deployment double click on 'Start_IPFS.bat' file to start IPFS server to store image of application and to get below screen



In above screen IPFS server started and now double click on 'runServer.bat' file to start python server and get below output



In above screen python server started and now open browser and enter URL as

'http://127.0.0.1:8000/index.html' and press enter key to get below home page.



In above screen click on 'New User Signup Here' link to get below signup screen



In above screen user is entering signup details and then click on 'Submit' button to get below output



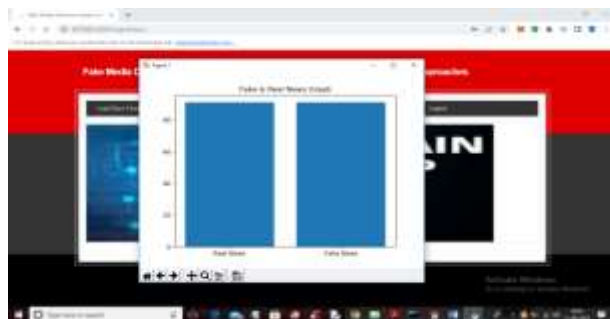
In above screen in red colour text we can see user details saved in Blockchain and now click on 'User Login Here' to get below login screen



In above screen user is login and after login will get below screen



In above screen user can click on 'Load Buzz Feed Dataset' button to load dataset and get below output



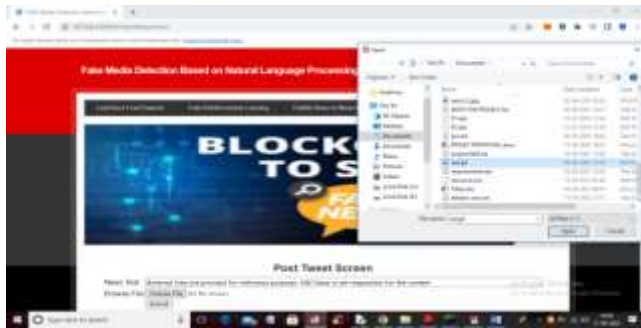
In above screen dataset loaded and we can see number of Fake and Real NEWS graph and in above graph x-axis represents NEWS TYPE and y-axis contains count of that news type and now close above graph to get below output



In above screen we can see News loaded from dataset and now click on 'Train Reinforcement Learning' link to train Reinforcement algorithm and get below output



In above screen we have trained dataset with existing Random Forest and propose Reinforcement algorithm and then we got RMSE, MAE, reward and penalty for both algorithms. Existing Random forest do not have support for REWARD and Penalty and we can see Propose reinforcement got less RMSE and MAE compare to existing algorithm. The lower the RMSE and MAE the better is the algorithm. Now algorithm model is trained and ready and now click on 'Publish News in Blockchain' link to publish news and then Reinforcement will classify weather news is Fake or Real.



In above screen I entered some NEWS and then uploading picture 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



In above screen we can see News is stored in Blockchain and we got hashcode of news storage and Transaction storage DELAY and now click on 'View News' link to view LIST of all news published by all users



In above screen we can see names of USERS who publish news and then we can see detection

output as FAKE or REAL. Similarly you can upload and test other news.

CONCLUSION

Fake news sharing is one of the popular research problems in recent technology based on lack of security and trust in terms of the truth of shared news in social media. In this article, we have presented the combination of blockchain and machine learning techniques to provide solutions and design a trust-based architecture toward shared news online. We have applied the reinforcement learning technique, a learningbased algorithm, to make a strong decision-making architecture and combine it with blockchain framework, smart contract, and customized consensus algorithm, which is well fit for the Proof-of-Authority protocol. Social media plays a key role in this process. The shared information platform contains fake news, and its a beneficial challenge to enhance and investigate the Proof-of-Authority protocol and user validation.

ACKNOWLEDGEMENT

We thank CMR Technical Campus for supporting this paper titled **"FAKE MEDIA DETECTION BASED ON NATURAL LANGUAGE PROCESSING AND BLOCKCHAIN APPROACHES"**, which provided good facilities and support to accomplish our work. Sincerely thank our Chairman, Director, Deans, Head Of the Department, Department Of Computer Science and Engineering, Guide and Teaching and Non- Teaching faculty members for giving valuable suggestions and guidance in every aspect of our work

REFERENCES

- [1] V. P. Miletskiy, D. N. Cherezov, and E. V. Strogetskaia, "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>