IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 6, 2021, accepted September 9, 2021, date of publication September 14, 2021,
date of current version September 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3112607

Fake Media Detection Based on Natural Language Processing and Blockchain Approaches

ZEINAB SHAHBAZI AND YUNG-CHEOL BYUN
IIST, Department of Computer Engineering, Jeju National University, Jeju-si 63243, South Korea

Corresponding author: Yung-Cheol Byun (ycb@jejunu.ac.kr)

This work was supported by the Ministry of Education and National Research Foundation of Korea through the ''Leaders in INdustry-university Cooperation +'' Project.

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.

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

I. 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

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

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 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.

The rest of the process arranged as follows: Section 2 presents the brief literature review of the recent techniques in this area. Section 3 presents the detailed methodology and data collection process. Section 4 presents the implementation of the developed framework based on an integrated system, and we conclude this research in the conclusion section.

II. RELATED WORK

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

A. MACHINE LEARNING TECHNIQUES FOR FAKE NEWS DETECTION

Machine learning (ML) algorithms and techniques which use the Natural Language Processing (NLP) to determine and underline the linguistic patterns in term of fake or real news [19]. Most of the ML process is related to classifier models, which are able to detect fake information from real information. Vijay *et al.* [20] applied Random Forest and NLP to detect fake news by counting vectors that are used for words. They also applied the RID matrix approach for finding the similarity and copy sources between documents. Their main approach is the usage of machine learning and NLP to classify the documents based on words. Chokshi and Mathew [21] presented a survey related to fake news detection using deep learning and NLP. Various deep learning methods are presented for detection based on a huge amount of data. Similarly, text classification and convolutional neural network (CNN) for image classification were analyzed. Wang *et al.* [22] proposed a fake news detection framework using the combination of three components as: fake news detector, reinforcement learning and annotator. This process have done to extract the weak labels of news and select the high quality samples to identify the fake news. The differences of the related works with the current proposed work is lack of trust in this process which in our case by applying blockchain network the users are supposed to register in network and after identification they have the possibility of sharing news in social media.

B. BLOCKCHAIN FRAMEWORK FOR FAKE NEWS DETECTION

Ochoa *et al.* [23] presented the centralized blockchain platform for fake news detection. This approach consensus algorithm is designed based on data mining to acknowledge the shared information. The presented system can fake news identification, alerting readers and punishing the one who unravels the information and remunerates the one who shared the truth. Shae and Tsai [24], presented the Artificial Intelligence (AI) based blockchain platform. This process contains a real news database for generating the supply chain graph and rank the fake news based on the crowd source mechanism. Paul *et al.* [25] presented news authentication among shared information using decentralized Ethereum smart contract blockchain. Qayyum *et al.* [26] proposed, news publication based on smart contracts to avoid the spread of fake news. In this process, before news publication, publisher authentication is processed, giving the publisher a public key and verification for sharing information. Balouchestani *et al.* [27] proposed, SANUB method for

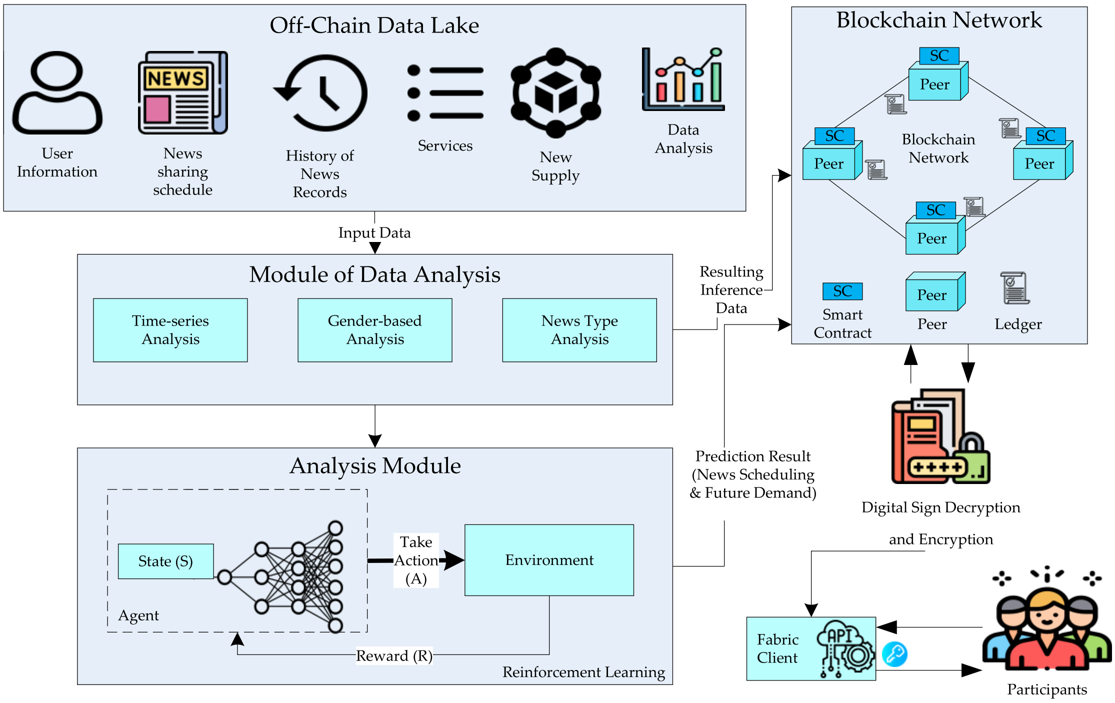

FIGURE 1. Block diagram of the proposed fake news detection approach.

blockchain-based news sharing and analysis. SANUB contains fake news detection, anonymously information sharing, evaluation of news, validation report, and proof of shared information. Islam *et al.* [28] presented news broadcasting as a kind of trade using blockchain by removing the third party from news transmission. Arquam *et al.* [29] proposed a news tracking system using blockchain-based allocating credit to users. These credits are in two types as local trust and global trust. Validating the news using this system causes the user news validation request from the publisher. Table 1 shows the related terms and concept of fake news detection from social media. The mentioned concepts below are not considered as fake news in social media [30], [31].

TABLE 1. Concepts and related term of fake news.

Type	Authentication	Intention	Reported News
Spam	Might be True	Advertise/Bad	No
Rumors	Disclosed	Disclosed	Disclosed
Hoaxes & Scams	False	Acceptable	No
Conspiracy Theories	Disclosed	Disclosed	No
Parodies & Satires	False	Acceptable	No
Misinformation	False	Bad	Disclosed
Disinformation	False	Disclosed	Disclosed
Clickbait	Might be True	Advertise	No

C. NATURAL LANGUAGE PROCESSING FOR FAKE NEWS

Natural Language Processing (NLP) is a famous linguistic computation which process to understand the human language and overcome the practical problems. Based on the manifestation and speech or text form of this process, NLP understand the various form of human daily conversion. Continuously changing of NLP, makes the explicit rules establishing difficult for computers [32]. NLP contains five key stages for analysis and extracting the computational meaning from document. This stages named as: tokenization, semantic analysis, lexical analysis, pragmatic analysis and syntactic analysis. Monteiro *et al.* [33] proposed the automatic detection of fake news in Portuguese language. Their process is based on uncovering the linguistic characteristic by applying automatic detection and machine learning algorithms in the presented system. Table 2 shows the comparison of recent existing works related to the proposed approach in term of advantages and presented work.

TABLE 2. Comparison of recent existing works.

Author	Proposed Approach	Advantages
Somya et al. (2021) [34]	Applying deep learning for analysis of multiple user accounts in Facebook	Adding more information in user profile to easily find the suitable decision
Nida et al. (2021) [35]	Using ensemble based deep learning model in LIAR dataset	Fake news detection based on two deep learning model for textual context
Ashutosh et al. (2020) [36]	Applying blockchain technique	Controlling the fake news propagation in social media framework

III. PROPOSED FAKE NEWS DETECTION SYSTEM

This section presents a detailed explanation of fake news detection with a combination of Natural Language Processing, Reinforcement Learning (RL) and blockchain. There are two main parts in the presented system which the learning steps and securing presented approach.

A. FAKE NEWS DETECTION USING NATURAL LANGUAGE PROCESSING

NLP is a decisive task in data preparation process which carry out with data cleaning, segmentation, stop words, feature extraction, word indexing and embedding. Data preparation has the responsibility of cleaning the content before starting any process on data. In the next step the data goes through feature extraction to convert the data into vectors and save them into database. The received data for feature extraction is taken from information retrieval and send the query once at a time to data source to request the related news from Internet. The feature extraction module has the ability of measuring the similarity between contents of news with the news storage and based on the distance and query sorting them into a list of news. Figure 2 shows the mentioned process of the NLP framework.

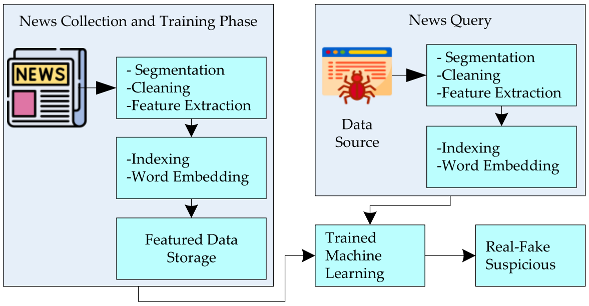

FIGURE 2. NLP framework for fake news detection.

B. FAKE NEWS DETECTION USING REINFORCEMENT LEARNING

Deep reinforcement learning is the incorporation of reinforcement learning and deep learning for decision-making from unstructured data. Most of the shared news in social media needs confirmation to know the shared topic's reality. The main reason for applying this algorithm for the proposed system is the learning-based aspect, which is a positive step in improving the performance of fake news detection. Problem formulation is based on Markov Decision Process (MDP). Every timestamp agent is in a state to take action and catch the award for the next state. The agent determines for learning policy based on increasing the reward returns. MDP contains the 4 tuples as (State E, Action C, Probability of actions taken from state P_e and Reward A). Table 3 shows the notations used in this process.

TABLE 3. Notations used for fake news detection.

Notations	Meaning
E	State
C	Action
A	Reward
Pe	Probability of Action
E	Time
π	Specifying action

Equation 1 presents the probability of the taken action C in the state T based on the time t which is lead to the next state $\hat{T}$ owing to the action C [22].

$$P_c(E, \hat{E}) = Pr(E_{t+1} = \hat{E}|E_t = E, c_t = c) \quad (1)$$

To make the proper decision based on MDP, a suitable policy is needed. Action needs to specify with function to decide the state. When the MDP integrates with policy and the integration result is similar to the Markov chain. Equation 2 presents the Markov transition matrix. π is a function to specify the action [22].

$$\pi(E), Pr\ (E_{t+1} = \hat{E}|E_t = E, c_t = c) \\ -Pr(E_{t+1} = \hat{E}|E_t = E) \quad (2)$$

Figure 3 presents the process of fake news extraction in training and classification steps. This process is based on the DRL architecture. The key solution of DRL is the learning-based of this system which makes the process having the high performance and quality in term of fake news detection. Based on the defined rules for the system and trained model based every step of training learns more and improves the quality of this framework. As it shown in architecture there are two type of news source as fake news and unlabeled news which are the input of trained DRL model which based on the classification process and ensemble with the strong classifier it labels the news.

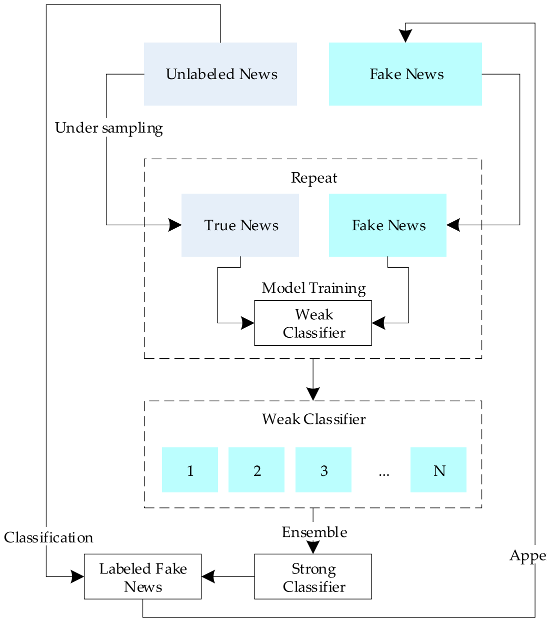

FIGURE 3. Deep reinforcement learning framework.

C. FAKE NEWS DETECTION USING BLOCKCHAIN

The blockchain concept appeared from the crypto-currency platform, which means any node can be a miner and join blockchain. This can be in terms of Proof-of-Work (PoW), Proof-of-Authority (PoA), and Proof-of-Stake (PoS). PoA is

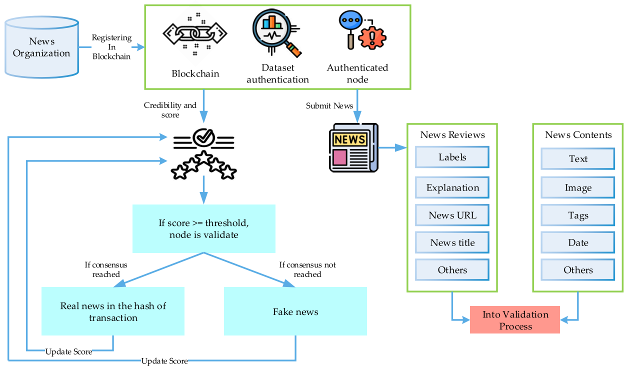

FIGURE 4. Fake news detection based on blockchain framework.

one of the new parts of the Byzantine Fault Tolerant (BFT) algorithm. The mining leader and proposing the new block of the algorithm operating process is the elected party. PoA messaging process is less than BFT, and it shows more acceptable performance. PoA implementation is based on the network assumptions to show the best interest of system preservation. The correspondence identity is related to the identification of validator on the platform, e.g., credibility score. Figure 4 shows the organization of news based on the decentralized manner of the validator. The presented scenario of blockchain for fake news detection is based on PoA with a high transaction rate. The credibility score applied for the consensus. The main components of this process presented as below:

- **Organization of News:** The entities which have enough power for registration are presented and involved in the validation of transaction and publishing. Building the network based on the trust and power of news publication in real life is not contemplated as a news organization. Some entities like BBC and CNN need to submit the application for registering into blockchain.
- **Data Authentication:** News organization validation requires some information e.g. exact data or document. This information will be verified and get license either to operate on TV, newspaper or radio. In the next step, the organization of news registers into blockchain with the smart contract rules to authenticate in nodes. All this process is in the blockchain registration as shown in Figure 4. The organization of news only registers once in the blockchain.
- **Proof-of-Authority (PoA):** The important step of the proposed system is the PoA which is plays the role of consensus algorithm for the fake news detection. First, publish the news, need the node authentication, and the next news organization can request publication. The publication is related to the status of credibility score. In this stage, some nodes become validators to check the truth of news and validate the transaction. In more detail, after submitting the news for validation, the transaction also gets into the process of validation. In this step, the validator pinpoints if the news is fake or real and the quality of fakeness. The transaction hash gives flexibility to blockchain for publishing. If the condition is not met, then news considers fake news.
- **Fake Media:** The content of every shared news contains image, text, tags, etc which are the main attributes of the shared topic that should be saved. It is required to have the related context to the fake news and the important information which can trace the news sources. Similarly, the presented system contains the backlog for fake news detection, and it's important and useful for future detection.

1) BLOCKCHAIN FRAMEWORK ENTITIES

Simple sense validation visualization of entities summarized in Figure 5. The highlighted field means the explanation of entities e.g. field UID means userID.

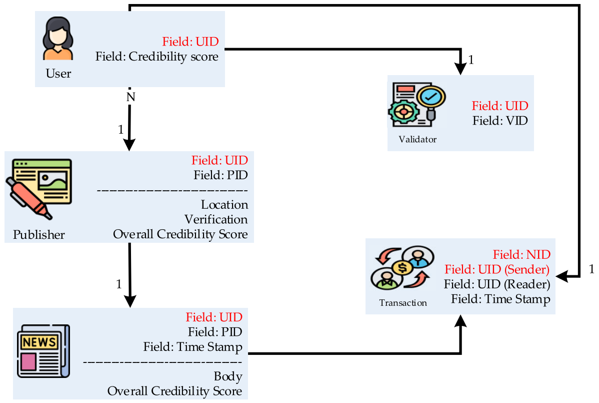

FIGURE 5. Diagram of relationship between entities.

a: USER

Based on the activities in this process, every participant gets a credibility score. Generally, each participant has a credit score and primary key. After registering the user in the blockchain network, the unique ID will be given to the user with the primary credibility score.

b: PUBLISHER

Every user is a publisher if they share content on social media. While proposing, the user has the primary key, which can be the hash value if the number of users gathers together for being a publisher. Each publisher has a foreign key which means verifying themselves and identity with the information of location. One important part is when the validator wants to view the publisher, the news can send with the publisher ID. At the same time, the publisher ID also sends to the credibility system to check the news based on publisher ID and hash value to check their validity and equality. If it's equal, then the credibility system gives the reward report.

c: VALIDATOR

The users which are not publishing any news. Based on the credibility score, it's possible to become a validator. The validator's responsibility is checking the publisher's news.

After the agreement of validators, the pending transactions collect in the news block of the blockchain. The proof of authority is the validator responsibility. Each validator has the validator ID and user ID.

d: TRANSACTION

The transaction records based on the total user activity in news acceptance or rejection records in the system and save into the blockchain. The transactions have the user and validator ID as well as publisher ID. Along with the news ID the timestamp of the news also generated. Blockchain traceability gives future access to complete the credibility testing and be sure about Real or Fake news.

e: NEWS

The process of sending news is from the publisher to validator and sending the news to blockchain. The publisher information doesn't reveal in the validation stage, but the total credibility score will share for the validators judgment. In the proposed system, every piece of news contains a news ID and publisher ID. The news ID is along with the publisher ID from the hash, and the validator can't know about the publisher. The publisher ID in this system is used for double-checking from the credibility system with the user ID. The reason is the publisher ID, and user ID is in the credibility system. In this case the credibility system is able to double-check the news ID based on the available hash value of user and publisher ID.

2) DEFINITION OF SMART CONTRACT FOR THE PROPOSED FAKE NEWS DETECTION ENVIRONMENT

The smart contract is the defined section of blockchain which performs the access point of blocks, transactions, and history. The programs of smart contract stores in the distributed database. Smart contract performs the business logic, adding constraints, transaction validation, etc. Information adding, operation performing, and contract existence is based on smart contract. Figure 6 shows the ledger querying and updating based on the smart contract in the proposed system. The reason to make the limitation of transaction query is performing all nodes. To overcome this problems, the transactions supposed to execute into nodes that are defined. This idea implemented in term of proper subsets and installing smart contract on selected subsets only. In this system, we have designed and implemented the Hyperledger Composer in smart contracts in the blockchain framework. The Hyperledger Composer creates security in blockchain, and it's an open-source framework. There are four main components in a smart contract defined as participants, assets, model files, and transactions to present access control rules in terms of data definition, query definition, and control policies. Participants of this network are those who interact network and represent the organization of the business network.

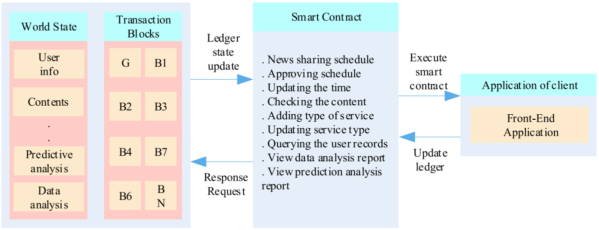

FIGURE 6. Updating and querying ledger based on smart contract.

IV. PREDICTIVE ANALYSIS BASED ON FAKE DATA DETECTION

This section presents the applied predictive analysis in fake news detection to improve system performance and help manage and formulate the business network effectively. The presented predictive model contains the following steps as: data collection, data pre-processing, using data mining techniques to extract the underlying patterns, data normalization, evaluation, data parameter selection, and processing the machine learning algorithms. Figure 7 shows the predictive analysis architecture in detail. The total process of the following prediction analysis shows the collected data information which are User ID, News ID, news prediction schedule, publication day, type of news, user gender, user name and user contact information. The raw data processing goes through pre-processing techniques, computed features and data normalization. After feature selection and splitting data into train and test set, reinforcement learning algorithm used for the fake news detection. RL is the learning-based algorithm to improve the performance of the process and extracting the fake news form social media based on defined rules.

A. DATA COLLECTION AND ANALYSIS

In the proposed fake news detection system, the dataset was collected from online sources and social media networks such as Twitter, Facebook, BBC news, etc. A total number of processed data contains 900.000 records of news from 2015-2019. Table 4 shows the presented features in the collected dataset. There in total nine attributes in the dataset named as: ID, title, text, URL, users, source, publish date, movies, and images. Each of these attributes represents the web page information that shared fake news, the title of shared news, the news contents, the news web address, user

TABLE 4. Feature presentation of dataset.

#	Attributes	Description
1	ID	Represents the webpage of shared fake news
2	Title	Represents the title of shared news
3	Text	Represents the contents of news
4	URL	Represents the web address of the news
5	Users	Represents the information of user
6	Source	Represents the source file of shared news
7	Publish date	Represents the date and time of shared news
8	Movies	Represents if the news contents are movie
9	Images	Represents if the news contents are images

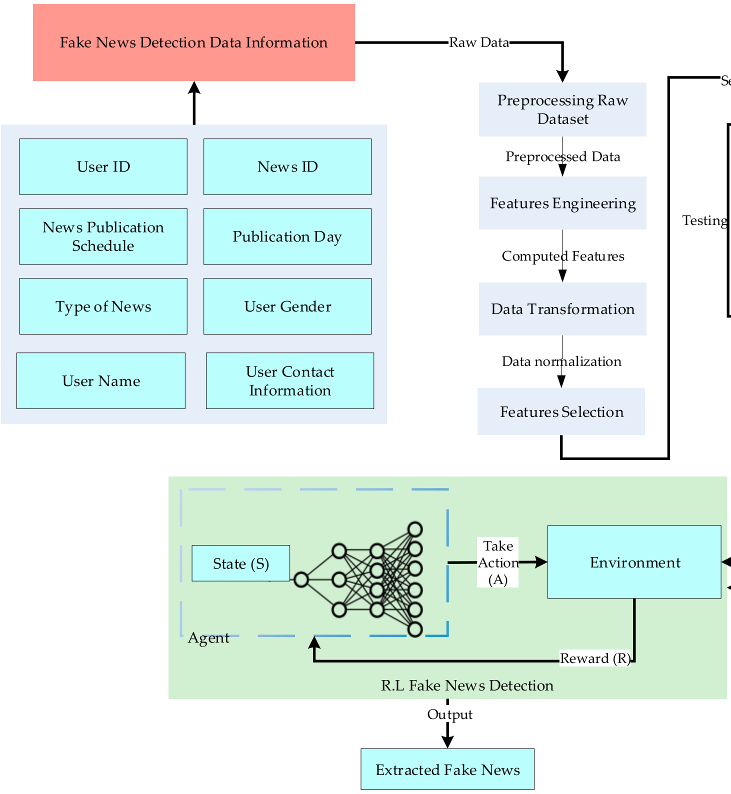

FIGURE 7. Reinforcement learning based predictive analysis for fake news detection.

information, the source file of the shared news, date and time of shared news, and contents of shared information.

Table 5 presents the available fake news datasets which are accessible online. Each of them contains different categories and various records with different topics as: Facebook posts and comments, articles, Wikipedia statements, politics and twitter posts.

TABLE 5. Available fake news dataset.

Dataset	Topic	Number	Label
Buzzface	Facebook posts and comments	2.263	Four categories
BuzzfeedNews	Facebook posts	2.282	Four categories
Buzzfeed-Webis	Facebook posts	1.687	Four categories
Fakenewsnet	Articles	23.921	Binary
Fever	Wikipedia statements	185.000	Three categories
Fake.br Corpus	Articles	7.200	Binary
Emergent	Titles and statements	300	Binary
Liar	Politics	12.800	Six categories
Credbank	Twitter posts	60.000.000	Five categories
pheme	Twitter posts	330	Binary

B. ANALYSIS AND PATTERN ENGINEERING

One of the important processes of discovering hidden features is patter engineering based on data mining (DM) techniques. The effective way to improve train and test set accuracy is feature extraction based on machine learning techniques. The presented model uses the following features as: shared news date and time, type of news, and user gender for knowledge discovery through collected information based on DM techniques. The time-series information contains helpful content, which is the best option of the prediction model input. In the first step, this model peruses the scheduled day of sharing news based on day, month, and year. Second, analyzing the total number of the news based on daily, monthly, and yearly analysis. Figure 8 presents the daily analysis of shared news based on two male and female gender groups. The daily analysis discloses the average of shared news which the frequency of male users is higher than female users.

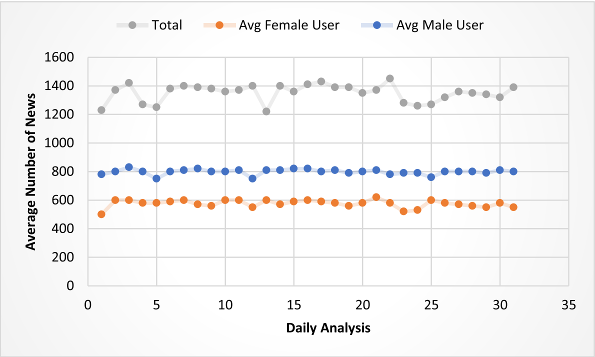

FIGURE 8. Time series analysis of daily news sharing of users.

Figure 9 shows the monthly analysis of shared news that presents the frequency of news sharing from January to December. The monthly analysis shows that from February to December number of shared news is relatively high.

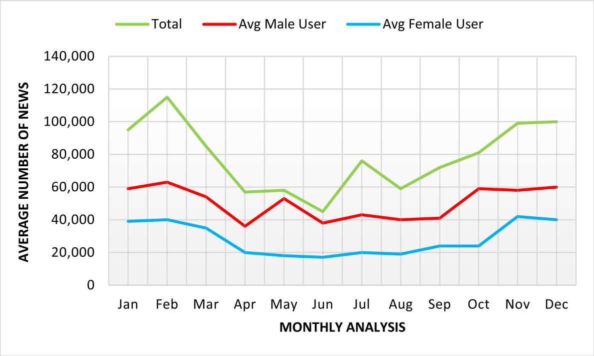

FIGURE 9. Time series analysis of monthly news sharing of users.

Figure 10 presents the shared news on the basis of yearly analysis. This process is according to users' gender. The yearly analysis shows the shared news records range from 120.000 to 160.000 in terms of male and female frequency, which male records are higher.

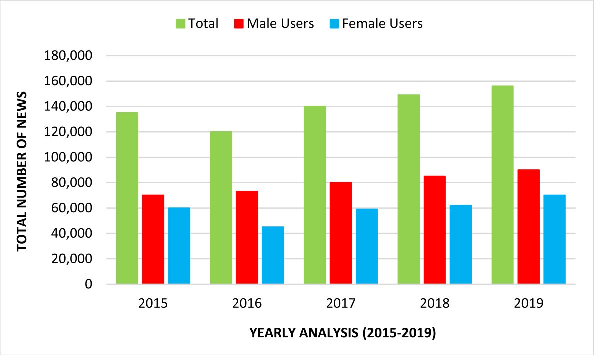

FIGURE 10. Time series analysis of yearly news sharing of users.

C. DATA NORMALIZATION

In the collected dataset, some of the features of data values are skewness, e.g., time series frequency of the shared news. Based on this reason, data normalization is required to avoid skewness and get a uniform range of data values. There are various techniques for data normalization, such as z-score normalization, min-max normalization, and decimal scaling normalization. In this process, the applied normalization technique is min-max normalization [37]. This technique normalizes the attributes based on scaling them between zero and one to get data features in uniform mode. Equation 3 is used to normalize data in a range of zero and one.

$$MinMax = \frac{Y_n - min(B)}{max(B) - min(B)} \tag{3}$$

D. FEATURED DATA EXTRACTION

NLP techniques analysis the retrieved data from data source. Next the text segmented into word tokens. Next step is the segmented tokens cleaning to remove the no meaning characters and words. In this stage the feature extraction, extracts the characteristic of the news content. The main features extracted in this process categorise into five groups as: fake score, real score, real domain, fake domain, sim matched. Fake score is the representation of fake group of words and negative words which appear in news content. The real score shows the right and positive group of words in news contents. The fake and real domain represents the number of social media platforms which shares fake or real news and the sim matched feature represents the similarity between the queries and content of news. Figure 11 shows the joint plot of the applied NLP techniques in feature data.

V. ENVIRONMENTAL IMPLEMENTATION OF THE PROPOSED ML AND BLOCKCHAIN FRAMEWORK

This section presents the environmental implementation of fake news detection based on predictive analysis and blockchain technology to predict the number of shared fake news and reduce the waiting time of the system. There are two categories for implementation of this system as blockchain-based implementation and predictive analysis.

A. BLOCKCHAIN-BASED IMPLEMENTATION FOR THE PROPOSED FAKE NEWS DETECTION

The implementation setup for the developed environment of the proposed fake news detection listed in Table 6. The operating system for the proposed approach is Ubuntu Linux 18.04 LTS, which contains a series of experiments. The Docker environment with the version of 18.06.1 applied for the system configuration and setting the virtual machine. Moreover, we have used the Hyperledger fabric for the run-time development of blockchain, the open-source framework. Composer playground applied for the design and development of the business network. Supporting the

TABLE 6. Development environment.

Name	Components	Description
Blockchain Network	CPU	Intel(R) Core(TM) i7-8700 @3.20 GHz
	Operating System	Ubuntu Linux 1804 LTS
	Docker Engine	V 18.06.1-ce
	Docker Composer	V 1.13.0
	IDE	Composer Playground
	Programming Language	Node.js
	Hyperledger Fabric	V 1.2
	Node	V 8.11.4
	Database	Couch DB
	Memory	12GB
	CLI Tool	Composer REST server
Machine Learning	Operating System	Windows 10
	Browser	Chrome, IE, Firefox
	Programming Language	Python, IDE

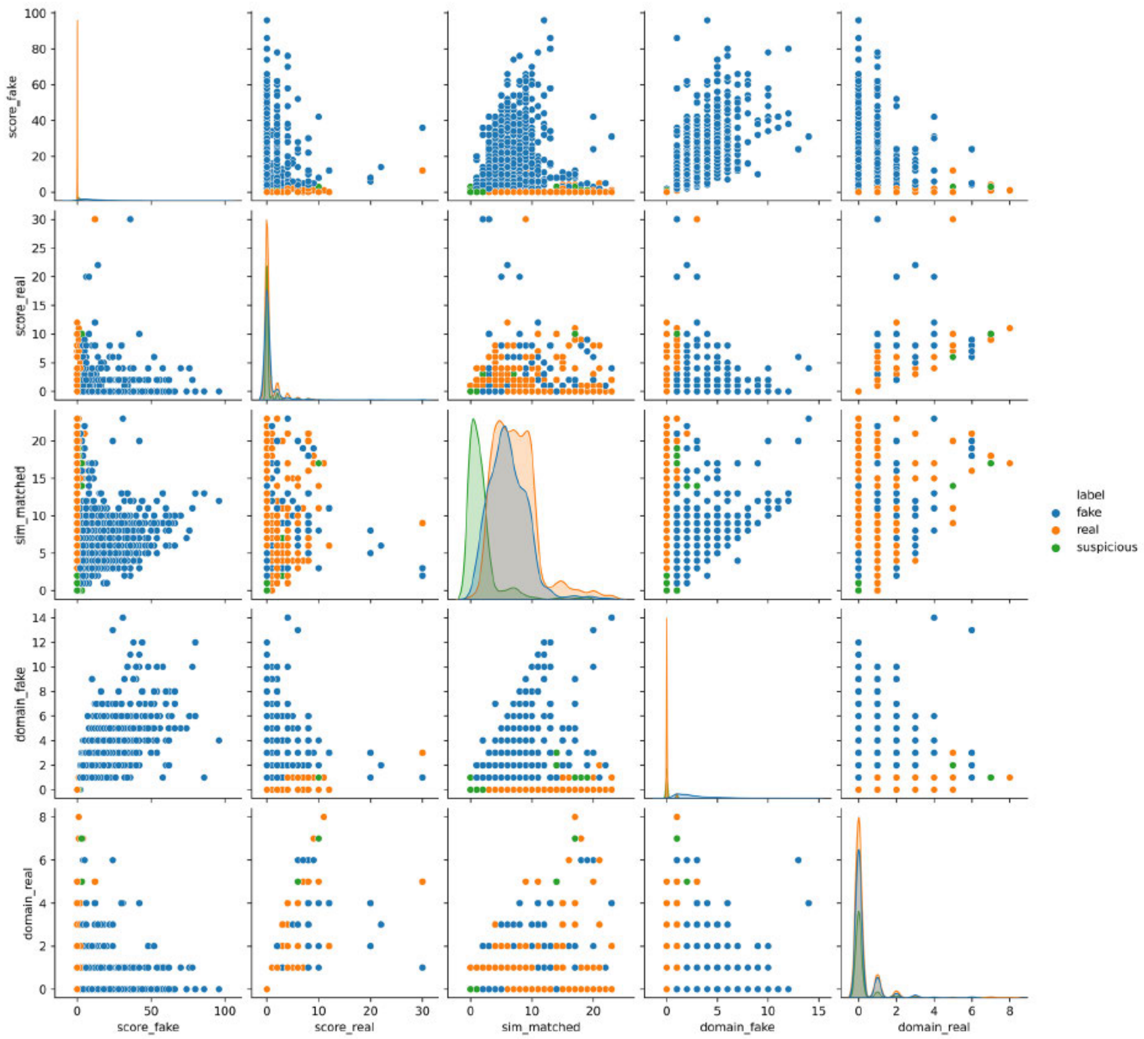

FIGURE 11. Joint plot of the featured data using NLP techniques.

blockchain in term of local-host and web, the composer web-playground provided. Fetching data based on blockchain technology needs the configuration of Composer REST server using CLI Tool.

Table 7 presents the list of RESTFul API generated by composer-rest-server to make the relationship between blockchain and client. The REST server can perform the CRUD operation for manipulating distributed ledger state. There are three main parts in the composer-rest-server define as: resource, verb, and action. These resources present the path of data in terms of request and response and specify the given resources based on required action in terms of POST, DELETE, GET, and PUT. The request for POST means creating or updating the new participants or assets, and GET means recapturing the data from a data source.

TABLE 7. Proposed blockchain system based on RESTFul API.

Resource	Verb	Action
/api/userProfile	POST, DETETE, PUT, GET	User management
/api/newsSharingSchedule	POST, DELETE, PUT, GET	Schedule of sharing
/api/updateSharingTime	POST	Update time of sharing
/api/updateSharingStatus	POST	approve or reject
/api/service	POST, DELETE, PUT, GET	Management of service
/api/updatePastRecords	POST	Update user history data
/api/news/supply	POST, DELETE, PUT, GET	News supply management
/api/dataAnalysis	GET	Repost of data analysis management

B. PREDICTIVE ANALYSIS IMPLEMENTATION AND EXPERIMENTS

The implementation setup of the predictive model is summarized in Table 6 Machine learning section. The presented model is based on the reinforcement learning (RL) algorithm, which is learning-based and fits this environment because of the decision-making strategy. The positive aspect of the learning-based model is an improvement of the system due to training based on defined rules for extracting fake news.

TABLE 8. Summary of discovered and base features.

Feature	Description	Feature Type
Frequency of daily scheduled news	It shows the total frequency of news based on daily schedule	Discovered
The distribution of news schedule per month	It shows the average frequency of news based on monthly schedule from January to December	Discovered
Distribution based on user gender	It shows the news distribution based on the users gender	Discovered
Type of news	It shows the topic and type of shared news	Base

The total process of the presented RL-based approach is divided into two main parts as train and test set. There are various techniques used for data splitting, such as: holdout and k-fold, etc. The sklearn library of python language divides the data into two subsets as 70 to 30 split ratio in this system. Equation 4 shows the training and testing split ratio of news data.

$$Split - Ratio - Data = \begin{pmatrix} 70, Train \\ 30, Test \end{pmatrix} \tag{4}$$

Table 8 shows the discovered features of time-series based on existing knowledge domains.

VI. BLOCKCHAIN-BASED EXPERIMENTAL RESULT FOR TRANSPARENT AND SECURE ENVIRONMENT

This section provides the performance evaluation and experimental result of the presented blockchain framework.

A. EXECUTION RESULTS

The execution results of the presented blockchain approach are summarized in Figure 12. This dashboard is based on the Hyperledger Fabric framework. The core functionality of the proposed system contains the user profile, news sharing schedule, news record history, services, news supply, and data analysis. The user profile represents the personal data of the user. It gives the possibility of creating, editing, and deleting the profile. The news scheduling module presents the news information related to the user trying to share it and the news status. Users can check the status and information of the news to save secure records in the blockchain. Furthermore, the admin can check the operation and news records history in the distributed ledger. The data analysis module gives useful information related to the hidden data of news, which is important for future strategies.

B. PERFORMANCE EVALUATION OF BLOCKCHAIN FRAMEWORK

The performance evaluation of blockchain network presented in this part. We have used the PBFT and RAFT [38] consensus algorithms for the performance evaluation of this system. The PBFT follows the Byzantine fault tolerance (Bft) with the low latency and high transaction for the bounded type of nodes [39]. Poor scalability is the limitation of the PBFT algorithm, along with high transmitting records for consensus messages. The applied blockchain network is a permissioned blockchain. The message-based aspect of PBFT is the main reason for using it for operating in the trusted environment.

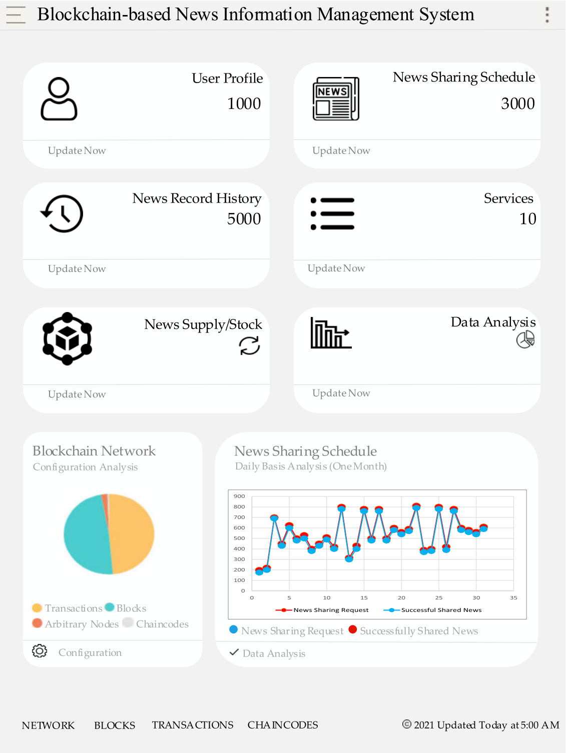

FIGURE 12. Hyperledger fabric front-end user interface for the proposed model.

Similarly, this algorithm due to proof of Authority (PoA) widely used in permissioned blockchain in terms of latency and throughput [40]. The RAFT algorithm usage is related to providing low TL and high TT. To assure data consistency, it uses log synchronization. Achieving data consistency based on RAFT algorithms contains four main steps: security assurance, leader node selection, log synchronization, and member change. Figure 13 presents the performance evaluation of the blockchain network based on transaction latency. There are four user groups defined to evaluate the network efficiency such as: 250,500,1000, and 1500. The transaction latency average record for 250 users is 1300 milliseconds (ms), for 500 users increased to 1497 ms. This process continues increasing by increasing the number of users in this process.

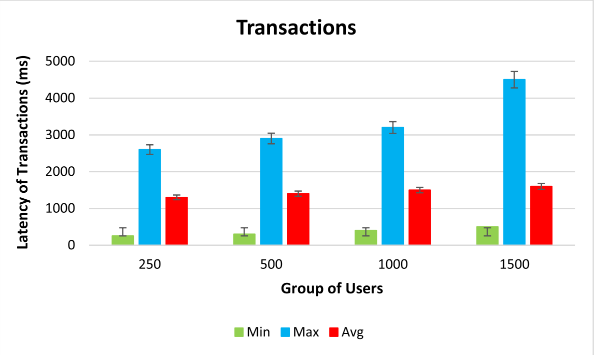

FIGURE 13. Performance evaluation of transactions based on number of users.

Figure 14 presents the performance evaluation of the RAFT algorithm in terms of transaction latency which the sending rate is between 25 to 200 TPS. Based on the presented process, the solo raft and raft have the higher transaction latency comparing with adding solo ordering based on transport layer security (TLS), the cause the improvement in transparency and security of blockchain networks among peers.

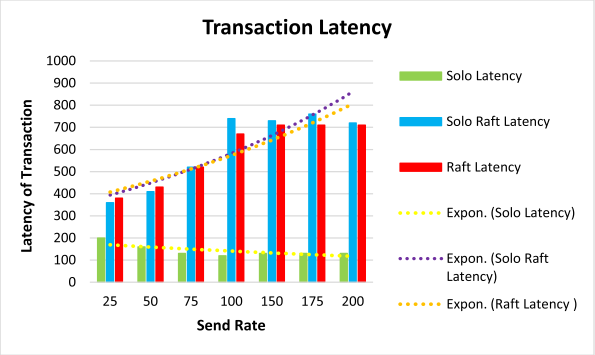

FIGURE 14. Performance evaluation in term of latency for ordering service.

Figure 15 presents the performance evaluation of transaction throughput in terms of min, max, and average. It can be noticed that solo ordering has the highest transaction throughput comparing with others because there is no need for additional TLS service.

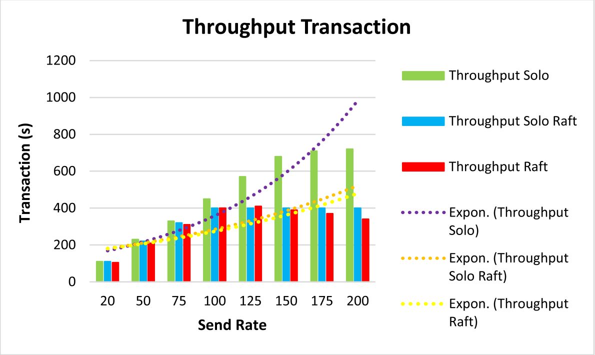

FIGURE 15. Performance evaluation in term of throughput for ordering service.

C. PERFORMANCE EVALUATION OF MACHINE LEARNING TECHNIQUE

There are various measures for the performance evaluation of machine learning techniques. In this process, we have used the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and R2 score. The following equations 5 to 7 used to evaluate the applied process.

- Mean Absolute Error (MAE): Measuring the prediction model performance evaluates based on MAE to find the differences between the actual and predicted value. Equation 5 presents the evaluation process.

$$MAE = \frac{\sum_{n=1}^{x} |Z_1 - \hat{Z}_i|}{x} \tag{5}$$

- Root Mean Square Error (RMSE): presents the MAE square error. This process is used for the overall evaluation of prediction model performance. RMSE range starts from 0 to ∞. Equation 6 presents the evaluation process.

$$RMSE = \sqrt{\frac{\sum_{n=1}^{x} |Z_1 - \hat{Z}_i|^2}{x}} \tag{6}$$

- R^2 Score: To show the variance proportion between two variables, this statistical measurement was applied. R^2 score is between 0 to 1.

$$R^2 Score = \frac{\sum_{n=1}^{x} |Z_1 - \hat{Z}_i|^2}{\sum_{n=1}^{x} |Z_1 - \hat{Z}_i|^2} \tag{7}$$

Table 9 illustrate the results and performance evaluation of the proposed system with the other machine learning models such as XGBoost, Random Forest, RNN, and LSTM. As shown, the presented system has the most significant and promising outputs compare with other models.

TABLE 9. Comparison and performance evaluation of the proposed system.

Model	RMSE	MAE	R2	MAPE
XGBoost	894.2	952.52	0.62	9.82
Random Forest	1175.34	1439.34	0.56	15.88
RNN	678.696	533.388	0.958	3.988
LSTM	629.96	488.356	0.957	3.722
Proposed System	318.036	173.515	0.968	1.87

VII. 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 learning-based 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.

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. 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. 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 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. 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.

[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. 21–32.

[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. 1–40, 2020.

[31] J. Golbeck, B. Auxier, and V. Kori, "Fake news vs satire: A dataset and analysis," in *Proc. 10th ACM Conf. Web Sci.*, 2018, pp. 17–21.

[32] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 604–624, Feb. 2021.

[33] R. A. Monteiro, R. L. Santos, T. A. Pardo, T. A. D. Almeida, E. E. Ruiz, and O. A. Vale, "Contributions to the study of fake news in portuguese: New corpus and automatic detection results," in *Proc. Int. Conf. Comput. Process. Portuguese Lang.* Evora, Portugal: Springer, 2018, pp. 324–334.

[34] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106983.

[35] N. Aslam, I. Ullah Khan, F. S. Alotaibi, L. A. Aldaej, and A. K. Aldubaikil, "Fake detect: A deep learning ensemble model for fake news detection," *Complexity*, vol. 2021, Apr. 2021, Art. no. 5557784.

[36] A. D. Dwivedi, R. Singh, S. Dhall, G. Srivastava, and S. K. Pal, "Tracing the source of fake news using a scalable blockchain distributed network," in *Proc. IEEE 17th Int. Conf. Mobile Ad Hoc Sensor Syst. (MASS)*, Dec. 2020, pp. 38–43.

[37] L. Al Shalabi, Z. Shaaban, and B. Kasasbeh, "Data mining: A preprocessing engine," *J. Comput. Sci.*, vol. 2, no. 9, pp. 735–739, 2006.

[38] W. Fu, X. Wei, and S. Tong, "An improved blockchain consensus algorithm based on raft," *Arabian J. Sci. Eng.*, vol. 46, pp. 8137–8149, Feb. 2021.

[39] D. Lu, Y.-L. Li, S.-J. Wang, and X.-Q. Ding, "Acta automatica sinica," *ACTA Automatica Sinica*, vol. 37, no. 1, 2011.

[40] M. Castro and B. Liskov, "Practical byzantine fault tolerance and proactive recovery," *ACM Trans. Comput. Syst.*, vol. 20, no. 4, pp. 398–461, 2002.

ZEINAB SHAHBAZI received the B.S. degree in software engineering from Pooyesh University, Iran, and the M.S. degree from the Internet Laboratory, Chonbuk National University (CBNU), in 2018. In March 2017, she moved to Republic of Korea for her M.S. studies. In March 2019, she moved to Jeju-do and started working as a Ph.D. Research Fellow with the Machine Learning Laboratory (MLL), Jeju National University. Her research interests include artificial intelligence and machine learning, natural language processing, deep learning, and data mining.

YUNG-CHEOL BYUN received the B.S. degree from Jeju National University, in 1993, and the M.S. and Ph.D. degrees from Yonsei University, in 1995 and 2001, respectively. He worked as a Special Lecturer with SAMSUNG Electronics, in 2000 and 2001. From 2001 to 2003, he was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI). He was an Assistant Professor with Jeju National University, in 2003. He is currently an Associate Professor with the Computer Engineering Department, Jeju National University. His research interests include AI and machine learning, pattern recognition, blockchain and deep learning-based applications, big data and knowledge discovery, time series data analysis and prediction, image processing and medical applications, and recommendation systems.