FAKE NEWS DETECTION OF SOCIAL MEDIA NEWS IN BLOCKCHAIN FRAMEWORK

Mr. Akash Dnyandeo Waghmare
1Research Scholar, Department of Computer Engineering, SSBT’s COET Jalgaon Maharashtra, INDIA
Email_id: aakashjan8@gmail.com

Dr. Girish Kumar Patnaik
2Professor, Department of Computer Engineering, SSBT’s COET Jalgaon Maharashtra, INDIA
Email_id: patnaik.girish@gmail.com

Abstract

Social media news are most important in today’s worlds, it puts positive or negative influence on social views. There is a wide propagation of fake news on social media so it will be difficult to believe on the news. Fake news has negative impacts on individuals as well as on society. Information spreads rapidly over the social media and so there is a need of mechanism which detects and stops the spreading of fake news. Therefore, detection of fake news is the need of time and also a challenging problem. The goal of this proposed research work is to detect fake news and minimize spreading of the fake news. In the proposed research a machine learning approach is used for detection of fake news with blockchain framework. In first section a supervised machine learning techniques is design to identify the trustiness of specific news while blockchain framework revoke the malicious activity of spreading fake news. A blockchain environment is created with mining, smart contract as well as Proof of Work (PoW) of consensus. The current systematic review broadly focuses on the various methods to detect fake news in social media. After partial implementation of system, performance evaluation has done with traditional blockchain framework. It is found that 10% less time for transaction verification by consensus in P2P environment over the existing systems.

Keywords: Fake News, Social Media, Classification, Blockchain, Cryptocurrency, Mining, Hash generation, Consensus.

1 Introduction

The enormous amount of data is produced on social networking in different formats of social media. There have been very large amounts of posts that have increased social media data on the web explosively. Many people discuss it on the web through social networking when some event has occurred. They are searching for or retrieving and discussing news incidents as the daily routine. However, very big volumes of news or posts caused users to face the issue of overloading data during search and retrieving. Unreliable sources of data expose individuals to a dose of false news, hoaxes, rumors, theories of conspiracy and fake news.

Fake news for different political and economic benefits occurs in large numbers following the successful development of online social media platforms. Several analytical experiments were performed using Natural Languages Processing (NLP) analytics to estimate the origins of false information or analyze the news text to characterize news as fake or accurate in different classifications. Current methods have detected false statements using other computer vision models, namely Naïve Bayes, ANN, SVM, etc. While there are several tools and techniques to identify fake news outlets, such as whether a website or person releases fake news, this problem can be discussed as an example of text analysis, utilizing news and information to function based on description, body, and context. Using it helps to identify a false statement by creating deep learning models. Specific parameter estimates will ensure higher predictive accuracy for the best results by tuning and regularly training the model.

Security methods using deep neural networks can recognize fake news and prevent malicious writers from disseminating fake information. In comparison, this paper integrates the advantages of blockchain with machine learning techniques to identify and prevent the spread of false information in widespread communication.

The remaining sections are structured as follows. Section 2 is a review the past work done in the field of blockchain and fake news detection. Section 3 introduces the proposed approach and model. Section 4 discussed about the blockchain algorithm design. Section 5 presents the experiments conducted to evaluate proposed fake news detection model using blockchain framework. Section 6 makes concluding remarks and discusses future work.
2 Review of Literature

Sharma et al., in (IEEE 2019) [18], explored the system of fake news detection from large text data using machine learning. This approach is the collaboration of machine learning and natural language processing techniques. The numerous techniques used by prior studies in the very same field were even included in the scheme. Use several experiments as examples, and the device studied the fundamentals of specific strategies used in detecting fake news. In this study, the future implications are also included together with the obstacles one faces and doing research here in this area. The propaganda of the world, Fake Media refers in laymen's terms to the deliberate or accidental distribution of misleading info also on a public forum.

Fitwi et al., in (IEEE 2019) [8], discussed about a Confidentiality Security Compact (CSC), cryptocurrency framework for edge camera. This helps the VSS to carry out monitoring before violating the privacy of individuals caught in the videos. The Lib-Pri system converts the VSS implemented into a design that acts as a nonfederal bitcoin blockchain consisting of integrity testing, maintenance of obscuring keys, exchanging of functionality, and penalizing of video access. The policy-based implementation of privacy provisions is carried out without burdening the system at the edge of actual video monitoring devices. Concerning security cameras, a variety of claims and opinions are being raised in support of and those against.

Bhoir et al., in (IEEE 2020) [6], discussed about the strength of the Internet which has grown massively over the past decades, thereby giving paws to disinformation, which is a rising issue as it is the misinformation that makes it all too harder to discern the truth in material. Therefore, the panacea is to provide a well thought and total proof device that tracks the various trends in the news that really can help us decide if it is true or not. With the assistance of computer vision, the proposed approach aims to recognize fake news leveraging hybrid models. This design significantly decreased learning time and improved SVM classifiers. This hybrid model performed well in comparison to the Random Forests (RF) algorithm and Vector Support Operator when designed separately.

Xu et al., in (IEEE 2019) [22], present a Cryptocurrency Decentralized Infrastructure for Smart Community Safety Kubernetes. A protection framework based on micro-services is implemented inside a licensed public blockchain to protect data access control in an SPS framework. Intelligence services technology is decoupled into independent container-based microprograms that are designed using a consensus protocol and implemented on nodes of edges and cloud technology. A comprehensive observational design confirmed that the theoretical BlendMAS is capable of offering a distributed IoT-based SPS framework with hierarchical, scalable, and stable sharing data and intrusion detection. Smart National Security systems store multichannel audio from the edge of surveillance and use multiple embedded technologies. Cloud services deployed on the private Cryptocurrency blockchain are converted into digital form to the hash index user authentication experienced numerous. Smart monitoring and intelligence services functionality is decoupled into the distinct container-based micro-service architecture and distributed on decentralized edge or cloudlet nodes.

Traylor et al., in (IEEE 2019) [19], discussed about a software defect prediction in which records the output of a Russian propaganda classifier. To create a new Russian propaganda tracker that uses referenced acknowledgement in a Probabilistic machine learning method as a defining aspect for estimate the probability that a news piece is fake, the Text blob, Artificial Intelligence, and SciPy Development tools were used. The resulting method accuracy is 63.33% successful in determining the probability that an essay with quotations is fake. This approach is called digging pressure, but this new method is portrayed as a methodology that can be used to detect false news and sometimes even misinformation. The qualitative research, advanced analytics, technical anthropology work, and reliability and results of classifiers are discussed in this paper. The pieces end with a summary of how the existing system can grow into an impact computing system.

Wang et al., in (IEEE 2019) [20], developed two dimensions. First, to help SaaS infrastructure, the centralized SaaS service platform focused on public cloud needs to recognize the functionality of virtualized environments, automated implementation and distribution, centralization of structured data, exchanging of knowledge, and efficient user and strategic efficiency. The centralized support network for the service is a way for computing. Third, our systems utilize P2P networks for data distribution in part in response to the cryptocurrency system. The findings are appropriate. Bitcoin has been the most rudimentary and essential implementation of the network, as an authenticated cryptocurrency. Every node reproduces and updates the position of all cryptocurrency transactions in the Bitcoin protocol. Still, each time an order is created; all processors migrate to the next new state which records all purchases throughout the upgrade process.

Kang et al., in (IEEE 2020) [12], illustrates how accurate the reporting has done in existing systems. MCE thus learns the latent space such that of the importance of the correct vectors of information becomes more incredible than with the false vectors of news. Because when element vectors are consistent and when its constituent vectors are well matched, a news vector becomes larger throughout the training. Through doing so, MCE will capture the dynamic relationships between the factors as well as each device's performance. Our detailed studies on two benchmark functions demonstrate that MCE outperforms all the timelines. A research method of the subspace is also provided by the device to validate its capacity to complete the specifications. Fake news may deceive and manipulate viewers, creating negative consequences or even exploiting large meetings.
Ai et al., in (IEEE 2020) [2], discussed about a power exchange decentralized resolution system for smart grid blockchain-based that aimed at the challenges of knowledge disunity. A trust framework problematic to build, power variability waste, and value progress caused by electrical pre-sale. The results of the tests show that through a rational set storage system that could fulfill the criteria of real-world applications, this device obtains impressive results. Smart grid services apply to distributed populations and storage facilities linked to the transmission network to provide finished with versatile loads.

Qawasmeh et al., in (IEEE 2019) [17], explored Automatic recognition of fake news via social networking platforms. Besides, the framework proposes an automated detection of false information utilizing modern methods of machine learning. Asymmetric LSTM convolutional model, which is implemented with 85.3% accuracy efficiency on the FNC-1 data is the standard method. Digital mass media have dramatically transformed the transmitting media of news through printed paper, telephone, radio, and TV to the Internet. The evolution of modern technological trends has contributed to the popularity of the Internet.

Desai et al., in (IEEE 2019) [7], explained a new hybrid blockchain framework that incorporates private with public cryptocurrencies to enable a blockchain to unlock confidential bids so that only the seller then no one can learn that bids. At the same time, the protocol leverages the sample average to account for the identification of the auction winner and payments. Besides, the framework demonstrates how to promote honest conduct among the auction respondents using digital currencies deployed on the bitcoin ledger. Compared to that of pure public cryptocurrency auction deployments, our comprehensive empirical outcomes show that such a model is more effective in terms of run memory and time cost.

Ahuja et al., in (IEEE 2020) [1], created model S-HAN to identify the fake news. This model uses an improved version of Hierarchical Attention Networks using stacked bidirectional Gated Recurrent Units (GRU) to identify the important words and sentences in the news which assists in the identification of misinformation. By operating at a level of the hierarchy, the stacking of GRU lets the model learn different complex text representations. In order, the HAN algorithm itself works by concentrating first on terms and then on phrases to interpret the most meaningful collection in the entire text.

Li et al., in (IEEE 2019) [15], developed a trust model which focus on the decentralization of blockchain technology, an agglomeration strategy for node group clustering. The framework introduces a mechanism for multi-chain space. These experiments show that the suggested solution decreases time and processing power for data synchronization, increases device efficiency by allowing successful parallel processing. The Virtual System (CPS) is a smart system that combines measurement, communication and power, and unifies computer but physical processes. A confidence model of a spanning tree that is arbitrary. In this model, the device in an
undirected graph is abstracted by the system as a node. The edge power defines that same communication relationship between these two network devices, and the degree of communication tightness is the connectivity or between node but the other joint points, which is a metric to calculate the degree of confidence across devices.

Wang et al., in (IEEE 2019) [21], discussed about decentralized and quantifiable services, a blockchain-based IoT architecture. The needed storage space, however, poses a grant challenge for hydrocarbon IoT infrastructures when introducing blockchain to the IoT Framework. The authors propose a centralized blockchain storage framework, Chain-Splitter, for the storage problem. In particular, the proposed architecture has a hierarchical system of storage where much of the database is placed in the clouds. In contrast, the latest blocks become stored and in the overlay.

To build the decentralized network space, this work simultaneously links local IoT networks, the blockchain overlay network, and the cloud infrastructure together via two connectors, the database connector as well as the cloud connectors. In the entire network, the ledger connector creates blocks from information made in IoT networks and general network, and the cloud connectors solve the problems of blockchain integration between the whole network as well as the clouds. A case study to demonstrate the feasibility of the proposed hierarchical database storage in a functional Industrial Automation case is also supported by the system.

Latifi et al., in (IEEE 2019) [14], discussed about blockchain jobs in the RE market and to represent the facilities it can offer to the RE market. The research to date leads to the following conclusions: the classical problems that RE faces can be solved by blockchain technology and smart contracts and they provide far more meaningful resources for a stable-priced market in game theory.

Zhu et al., in (IEEE 2019) [25], discussed about a novel hybrid blockchain crowdsourcing network to achieve decentralization and protection of privacy. To ensure safe contact between the applicants and the staff, our platform integrates with a hybrid blockchain system, dual-ledgers, and dual consensus algorithms. Besides, the smart contract and zero-knowledge evidence are used to guarantee the automated execution of the tasks and the privacy of the security users, respectively. Finally, by contrasting it with the state-of-the-art, the device performs simulations to validate the performance of the implemented smart contract on our network. Bitcoin is the first generation of digital currency blockchain that embraces the consensus of Confirmation.

Hasavari et al., in (IEEE 2019) [9], implemented a mix of secure file transfer procedures and blockchain as a tool for capturing hospital critical medical data as patients walk from one pharmacy facility to the another, establishing a consistent patient footprint as a protected and decentralized source of data. Thus, emergency crews can access it and use it to provide before the care of high quality. In this strategy, all health record communication and access problems, such as authentication, safety, protection, interoperability and acoustic impedance, confidentiality, were taken into account.

Yaqing Wang et al., in (ACM 2018) [23], proposed the Event Adversarial Neural Network (EANN) end-to-end framework, which can derive event-invariant features and thus benefit the detection of fake news on newly arrived events. It comprises of three primary parts: the extractor of multimodal features, the detector of fake news, and the discriminator of events. It is the multi-modal extractor that extracts textual and visual characteristics from messages. It cooperates with the fake news detector to learn how to discriminate in detecting fake news. The function of event classifier is to remove event-specific characteristics and to maintain shared characteristics between events. Extensive experiments are carried out on Weibo and Twitter gathered multimedia datasets. The experimental findings indicate that our suggested model of EANN can outperform state-of -the-art techniques and learn to represent transferable features.

Kai Shu et al., in (IEEE 2018) [11], Build real-world data sets that measure user confidence in fake news and select representative groups of both skilled consumers who can acknowledge fake news items as fake and "naive" users who are more likely to think fake news items. Build real-world data sets that measure users’ level of trust in fake news and select representative groups of both experienced users who can recognize fake news items as fake and "naive" users who are more likely to believe fake news.

Arjun Mukherjee et al., in [5], studied cooperative spam detection, i.e. discovering false reviewer groups. To discover a set of candidate organizations, their suggested technique first utilizes a frequent item set mining technique. It then utilizes many cognitive models extracted from the collusion concept between fake reviewers and relationship model based on the relationships between individual reviewers, groups and products that they examined to detect fake reviewer groups. They also build a labeled dataset of fake groups of reviewers. While it is very difficult to label individual fake reviews and reviewers, it is much easier to label fake reviewer groups to our surprise. This technique was départs from the traditional supervised learning approach for spam detection because of the inherent nature of our problem which makes the classic supervised learning approach less effective.

Mykhailo Granik et al., in (IEEE 2017) [16], showed a easy strategy to detecting fake news using the naive classifier Bayes. This strategy has been introduced as a software system and tested against such a collection of news articles from Facebook. On the test set, which is a good outcome given the comparative simplicity of the model, they had accomplished classification precision of roughly 74 %. In several respects that are also outlined in the article, these outcomes can be enhanced. Received outcomes indicate that artificial intelligence techniques can address the issue of fake news detection.
Youngkyung Seo et al., in (IEEE 2018) [24], presented a new fake news detection model, FaNDER (Fake News Detection Model Using Media Reliability) that can effectively classify the level of reality for news in a question answering scheme based on altered CNN profound learning model. Their model represents the reliability of different media by training with the input data set containing the truthfulness of each media and the proposal. Their model is intended in terms of information increase, batch size control and model alteration for greater precision with media dataset.

Akshay Jain et al., in (IEEE 2018) [3], proposed system for detecting fake news. They had a classification model for Naive Bayes to predict if a post on Facebook will be marked as true or false. They had contrasted the four received AUC scores and identified the best model and parameters suitable for better model implementation.

The above literature survey describes how system deals with fake news detection approach using various supervised machine learning classification algorithms. Numerous researchers also have done lot of research on implementation of blockchain technology for transaction verification purpose with large data. But still most system facing data leakage and corrosion attack problems, it needs to improve some immutable technology and provide security to data.

3 Proposed System Design

The model-based on supervised Machine Learning (ML) algorithms to detect the fake news generated by reporter as well as generated by malicious bots. From a natural language processing viewpoint, it solves the issue of identifying false information. The ML-based monitoring systems classify articles through two incorrect or actual forms due to the previously posted online source in the network. As reporter' identifiers are not revealed before the news is written, the machine learning model doesn't recognize reporters' addresses for news identification.

As in the proposed model, the news can be released by anybody, and everyone can verify the validity of the news. Consequently, blockchain offers a decentralized, trustworthy, storage facilities to store any user activity is used as evidence, further used only to identify false information in the ML model. The above Figure 1 describes the social media blockchain framework for fake news detection using machine learning approach. The system consists three important modules that is describes in below.

**Data Publisher:** In data publication module any anonymous user can upload the news, when data upload successfully done, it automatically converts as block and stored in blockchain. The blockchain framework has followed by system when data publisher publishes any news on portal. Four basic methodologies have been followed by system during news uploading like Hash Generation, Mining of transaction, Smart contract generation and consensus for P2P verification respectively.

**Analytical Model:** This model contains the various supervised classification algorithm that deals with NLP for feature selection and extraction. Tokenization, stop words removal, lemmatization, dependency based feature
extraction etc. has done in first phase. Then it collaborates the supportive data for validation of current news, which is extracted from blockchain. The supportive data used as training rules for validation model and according to trust factors classifiers decide publisher’s news real or fake.

**Data Provider:** Basically, this module holds the real news dataset that is basically used for proof verification of reporter’s news. In other words, this repository considers as training dataset by classification algorithms. The basic assumption for this repository, it is trustworthy according to current events or news.

4 Algorithm Design

The major blockchain algorithm and single machine learning classification algorithm has used for proposed system implementation. The below algorithms have used for blockchain implementation, those algorithms have described in below:

1. **SHA 256 based Hash Generation**
   **Input:** initial transactional or input data Data[
   **Output:** Generation hash using SHA256 algorithm
   **Step 1:** Input data Data[
   **Step 2:** Perform SHA 256 from SHA suitable algorithms
   **Step 3:** NewHash = SHA256(data[])
   **Step 4:** Return String (NewHash)

2. **Protocol for peer to Peer node verification**
   **Input:** User input query, Current Node blockchainCurrent_Node[chain], Additional Outstanding Nodes blockchainNodesChain[Node_id] [Node_chain],
   **Output:** Automatic recover blockchain if nodes fails
   **Step 1:** Transactional data or any event data for input to blockchain
   **Step 2:** Extract kth server’s blockchain of time[t]
     Current_chain = Current_node[Chain]
   **Step 3:** foreach 
     NodesChain [Nodeid, Chain] ∑
   **Step 4:** foreach (i into Node_Chain)
     If (!.equalsNodeChain[i] to (Current_chain))
     FlagVal 1
   otherwise continue; commit query into the nodes
   **Step 5:** if (FlagVal == 1)
     CCount = SimilarityyNodesBlockchian()
   **Step 6:** Determine majority of all server
     Recover unacceptable blockchain from precise node
   **Step 7:** end if; end for; end for

3. **Transaction Mining Algorithm to generate valid hash**
   **Input:** Smart contract SC[], Transactions current hash CH[], Previous hash or genesis block PH[]
   **Output:** Generate of valid hash and nonce
   **Step 1:** Generate the hash_Value for kth transaction consuming Algorithm no. 1
   **Step 2:** if (hash_Value.valid to smart_contract [])
     Current hash is validate
     FlagVal =1
   Else
     FlagVal=0
     Continue; Mine the current hash again for next iteration
   **Step 3:** Return validated_hash[] till when flagVal = 1

5 Results and discussion

Evaluate the performance of blockchain execution as per the requirement. The device runs with an INTEL 2.8 GHz i3 machine and 4 GB RAM with a distributed manner on the Java 3-tier analytics platform. Liar dataset has used for partial implementation of that is taken from www.kaggle.com. The below Figure 2 shows the time
required for a consensus with Proof of Work (PoW) to authenticate the blockchain in minimum 4 nodes. First experiment analysis was given on blockchain implementation for the validation of results.

![Figure 2: Time required for transaction with no. of transactions with blockchain](image)

In next experiment, evaluation of the system with smart contract validation for defined consensus algorithm in different number of data nodes in peer to peer environment is carried out.

![Figure 3: Time required for smart contract validation with different no. of P2P network in blockchain](image)

In the third test case, the sum of variability obtained from the proposed SHA value by the model is analyzed. Effectively, this was done to determine whether or not the proposed hash sequence is valid and according to defined mining policy. The computing policy is never fulfilled in many cases when the system generates SHA code for provided transactional data. According to the mining circumstance, to comply with the mineral extraction policy to produce multiple variations on the text sequence. The time needed to build the correct SHA string for particular transactions is shown in Figure 4 below.

![Figure 4: Time required for generation of valid hash for number of transactions](image)

The above figures show valid hash generation time for various transactions, based on define smart contracts. The valid hash generation sometimes depends on difficulty of algorithm and processing environment. When more difficulties have mention in smart contract for hash generation it takes maximum nonce for creation of valid hash.

### 6 Conclusion

In this research, a novel idea of the machine Learning-based Model to Detect Fake News using Blockchain is presented. There are several methods experimented within the existing systems but the proposed system deals
with custom blockchain and machine learning classification algorithms. In experiment, it is found that minimum computation time taken by custom blockchain during the module execution. At present, the system framework is partly deployed. We plan to introduce a fully working real-time framework to identify and eliminate fake news as part of our future research.

References


Authors Profile

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

Dr. Girish Kumar Patnaik has completed PhD degree in Computer Science & Engineering from Motilal Nehru National Institute of Technology Allahabad. Currently he is working as Professor & Head, Department of Computer Engineering, SSBT’s College of Engineering & Technology, Jalgaon. He is recognized PhD Guide in Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. He has published 28 research papers in reputed peer reviewed journals in addition to 10 papers in International Conferences to his credit. He is Senior Member in IEEE, Professional Member in ACM, Life member of ISTE and CSI. His research interests are Wireless Sensor Networks and Security, Machine Learning, Blockchain and Natural Language Processing.