Lightweight Chain for Detection of Rumors and Fake News in Social Media

Yazed Alsaawy[1], Ahmad Alkhodre[2], Nour M. Bahbouh[3], Adnan Abi Sen[4], Adnan Nadeem[5]
Department of Information Technology, Faculty of Computer Science and Information System
Islamic University, Medina, KSA[1, 2, 4, 5]
University of Granada Spain[3]

***Abstract*—Social media has become one of the most important sources of news in our lives, but the process of validating news and limiting rumors remains an open research issue. Many researchers have suggested using Blockchain to solve this problem, but it has traditionally failed due to the large volume of data and users in such environments. In this paper, we propose to modify the structure of the Blockchain while preserving its main characteristics. We achieve this by integrating customize blockchain with the Text Mining (TM) algorithm to create a modified Light Weight chain (LWC). LWC will speed up the verification process, which is carried out through proof of good history where the nodes will have the weights according to their previous posts. Moreover, the LWC will be compatible with different applications such as verifying the authenticity of news or legal religious ruling (fatwas). In this research, we have implemented a simple model to simulate the proposed LWC for the detection of fake news and preserving the characteristics and features of the traditional Blockchain. The results on experimental data reflect the effectiveness of the proposed algorithm in establishing the chain.**

Keywords—Component; fake news detection; text mining; blockchain; detection algorithms words

I. INTRODUCTION

Actually, "the production of false rumors or news that might prejudice public system or security, or send or re-send them through social media or any technical means, is an information crime that carries a sentence of five years in prison and a fine of three million riyals, Noting that the sixth article of the system "fighting information crimes" Any person who commits any of the following information crimes; production that violates public order, religious values, public morals, or the sanctity of private life, shall be punished with imprisonment for a period not exceeding five years and a fine not exceeding three million riyals, or one of these two penalties, preparing it, sending it, or storing it through the information network, or one of the computers. "[Okaz Newspaper]. All this poses a big challenge for researchers, to find an effective solution to the problem of information validity and reliability.

When digital content is shared, video or pictures of any event, how can we trust its authenticity and credibility? Considering that, the public is losing their trust in media due to the absence of reliable indication of facts. The 21st century's beginning set the premise for today's disruptive digital economy, wherever creating and presenting digital content has become convenient and easy. Digital content in the form of many multimedia types of blogs is being created and published at large scales today. Fake news misguides the public or just seeking more page views to get more income by dishonestly transforming a new kind of yellow press. Whereas, putting an end to these wrongdoings has become necessary.

An instinctive way to neutralize fake news in social networks is to apply a central regulatory authority to handle the news flow. However, applying such authority requires adjusting the functional and the model of trust in social networks that is tough as it is a public and open network.

Blockchain technology is a technology that can ensure the security and reliability of various forms of information such as news. It can establish the largest distributed record that allows the transfer of ownership or the exchange of trade transactions without the need for a third party. In addition to achieving a high degree of safety and reliability. To date, this technology has only been used in the commercial and financial domains, and its applications have been limited to digital currencies, electronic payment, smart contracts, and some latest applications in medical information confidentiality.

One of the major challenges is to uphold the decentralization of social networks while limiting fake news sharing.

Our study in this paper involves:

- Analyzing and detecting the fake news on social networks.
- Verifying the source integrity of information.
- Ranking the news with an indicator of authenticity.

Due to the anonymity discussed earlier, users can validate the news feed with no pressure. Consequently, their validation cannot be biased to any agendas. After the news is published, the deployment will be in the form of a chain transaction. And then with a certain degree of vitality, validators receive a request to validate the news [2]. The validator will give value to the correctness of this news. Those values will indicate the authenticity of the news. Due to the decentralization and anonymity of the system, their verification is transparent and trustworthy. Once the verification is performed, the news will have a ranking of authenticity. These ranks will be included anywhere the news is distributed.

Our contribution is summarized by first providing a new solution to the problem of documenting and verifying news in social media applications or Islamic electronic fatwas applications. Second, we will propose a modification to the blockchain structure as the Light Weight chain, through integration with text mining algorithms. Third, we will explain how the consensus algorithm is established based on the good reputation to select the master nodes and estimating their importance in the news evaluation process. Fourth, to verify the feasibility of our research idea, we test and implement the proposed model. Fifthly, to automatically verify and create the modified string, we develop and implement the Text-Mining algorithm, and finally, we suggest a set of important future work in the field of news verification.

The rest of the article is organized as follows. In Section 2, the literature review demonstrates previous work in Section 3 and our analysis as discussion in Section 4. Then we will present our proposed framework, algorithm and presents a case for applying our idea in Section 5, and Section 6 depicts the results and evaluation.

II. Literature Review

The use of technology to detect fake news and classify it correctly is a challenging task. Researchers have made efforts to deal with this important issue. This section presents an overview of several of them, discussing their approaches, results, and effectiveness. Many studies have been published to address the problem of detecting fake news in cyberspace, such as "Detecting fake news on social media [3] is one such study. Authors claim a very high accuracy ratio of fake news detection. They used WEKA classifiers which can be described as an algorithm that evaluates the given data and generates the final result after processing [3], but there is still a confidentiality issue. In another example of Detecting Fake News from Multiple Sources and Multiple Classes is proposed in [4], called the MSMC technique for detecting fake news. The authors endeavored to solve the problem by answering two questions [4]:

- How to effectively combine information from multiple sources for fake news detection?
- How to mathematically discriminate between degrees of falsehood?

It integrates three elements of coherence in an automated form of end-to-end features: interpretable multi-source fusion, extraction, and falsity discrimination [4]. In the extraction, they use a deep model to extract characteristics from textual sources based on the CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory) [4]. The interpretable multi-source blending component aims to combine features from different sources by applying different formulas and final falsehood discrimination that is responsible for the degree of falsehood. Their model obtained 38.81% accuracy and the most notable point is the discrimination of falsehood, regardless of the minimum precision compared to other results.

The studies we mentioned above discuss the fake news detection part, but our goal is to integrate detection and Blockchain to get the benefit of them. Credibility Test: A Blockchain approach to detect and block fake news on social media [5] is the most relevant to our use case, due to the use of Blockchain in its model. The authors proposed a modification in the scheme of the social media system, to be redesigned with Blockchain technology to provide control over the news shared and back to users on a P2P network. The proposed algorithm is called Proof of Credibility, aimed at people and preventing the exchange of false news on social networks [5].

The concept of this algorithm is to redesign the architecture of social networks to be more decentralized, where users are represented in P2P communication as nodes. Each node contributes to the distributed ledger, which is identified as an immutable and cryptographically protected record of identified fake news. The authors of this article simulate their algorithm on Twitter with two hashtags, and the most interesting finding from this study is that the satisfied value of accuracy is 89% on personal fake news.

In [12] the authors proposed a new method for sharing and analyzing the news to detect fake news using Blockchain, called SANUB. SANUB provides features such as publishing news anonymously, news evaluation, reporter validation, fake news detection and, proof of news ownership. The results of our analysis show that SANUB outperformed the existing methods.

In [13] the authors have briefly explained the discussion of how cryptography and Blockchain technology can be used to detect video fraudulence and the possible way of its implementation is also discussed.

In this paper [14], researchers proposed a new Blockchain system that overcomes current challenges and limits the spread of fake news across the network by analyzing information workflow in social networks and building an optimal detection system that can be deployed effectively with minimal overhead.

Using the advantages of Blockchain's [11] peer-to-peer network concepts, researchers discussed a method for detecting fake news in social media. Their method is based on giving weight for the users to determine the probability of being selected as a validator that will be useful to rating the sources [15].

There are many research studies on detecting fake news on social media, and some research studies have used the Blockchain method for their purposes [16]. In a research study, Gilda explored the application of natural language processing techniques to detect fake news [17].

Another research work used the Naïve Bayes classifier to identify fake news [18]. A research study applied the closest neighbor algorithm to classify the news polarized from the trusted news [19]. Youngkyung Seo, Deokjin Seo, and Chang-Sung Jeong provided a model for the detection of fake news using media reliability [20]. Shivam B. Parikh, Pradeep K. Atrey conducted text analyses to uncover fake news.[21]

In research [23], one speaks of fake news and one-off news as indicated in the vector space model, combining vector

display of the repetition term, the frequency and repetition of the opposite report are reversed with mutual consent 10 superpositions using the computation classifier of vector reinforcement machine [22].

Cai Shuo, Huan Liu, and Suhang Wang have formulated a dataset with accurate information [24]. They gave two news rankings (fake and real) [25-28] for two of the client meetings [29]. One meeting was from an experienced client (they could easily spot fake.

There are many mobile app for fake news detection, along with a comparison of them posted in the affiliation with this subsection. Such as: NewsCop [8], ELISA [9] and Oigetit [10] but all have used AI algorithms without integration with Blockchain in their models.

III. Discussion

We now discuss the main challenges in news validation according to existing literature and our proposed solution.

1) The huge volume of data, which makes it impossible to save an entire series in the social media space.

2) The difficulty of having this chain for every user of communication sites.

3) The difficulty of involving all users in the consensus process, which is why we have a Master Node in our proposed solution with a consensus mechanism based on proof of good history algorithm.

On the other hand, we differ from Researchers in [6, 7, and 14], which discussed ideas close to our own research. But in [6], authors have cooperated with BlockChain with AI to detect and track fake news by detecting DeepFake in videos and flagging suspicious accounts, as it relied on a group of editors to commonly supervise the process. The proposal in [7], relies on a prior definition of reliable news agencies as a source to verify other news and use the BFS algorithm in the search process. Finally, the proposal in [14] is interested in reducing the problem of re-sharing news irregularly or modifying it.

In the current research, we provide a solution to the problem of the huge size of the fake string in the news that all the research suffered from, including [6, 7, and 14]. We achieve this by enabling TM to create a modified light string, and thus at the same time providing a solution to enable a larger number of nodes to participate in the evaluation process as a master node. We suggest a special algorithm for the consensus process that fits the nature of the news and the nature of the blockchain, by linking a historical record to the reputation of the node participating in the evaluation and the final classification of the news as either true, false or true in a certain percentage.

The most important characteristics of blockchain that we have preserved in the proposed modified model are as follows:

1) It does not depend on a limited number of authorities or governments so that the system does not become semi-centralized, and this affects freedom of expression in news applications.

2) Allow readers to share news in the decision process to promote freedom and decentralization.

3) Integration of trust between people (as auditors) and trust between machines such as automated auditing, as per blockchain technology process.

4) Non-denial so that a node cannot deny its ownership of a particular block within the chain.

5) Non-modification and reliability, so that any modification to a block will be revealed at the level of the block itself and at the level of the chain of the node that did the modification process.

6) The strength of the code, or the code is the rule, that is, the node becomes a master or gains weight according to the consensus algorithm used, which is managed automatically through the software codes without administrative intervention. The most important of these algorithms are illustrated in Table I.

As shown in Table I, the comparison shows the performance of each consensus protocol where "+" means a good performance of protocol and "++" high performance where "o" means the performance is average. The characteristics of our proposed framework illustrate its strength in performance and completion.

TABLE I. A COMPARISON BETWEEN CENSUSES ALGORITHMS

Algorithm name	Principle	Performance	DLT environment	Completion	Example of use
Proof of Work (PoW)	Difficulty finding solutions with ease of verification	-	Public	probabilistic	Bitcoin, Ethereum, Litecoin
Proof of Stake (PoS)	Whoever has a greater share of validators has the potential to create blocks	+	Private & Public	probabilistic	Ethereum, NXT, Tezos
Delegated Proof of Stake (DPoS)	Producing new blocks for a small, fixed number of elected validators authorized by the participants. It has high competition, and it is very profitable.	+	Private & Public	probabilistic	EOS, BitShares
Proof of Activity (PoA)	It is a mixture between the first two protocols in terms of share and arithmetic power, providing a balance between the miners and the ordinary participants	-	public	probabilistic	Decred
Proof-of-Location (PoL)	To verify a node, its position is marked by GPS and a temporary stamp is given to it	o	public	immediate	FOAM, Platin
Proof-of-Importance (PoI)	Similar to POS algorithm with other important features	+	public	probabilistic	NEM
Proof-of-Elapsed-Time (PoET)	The blocks are created in a secure environment, which is similar to POW, except that it is from a reliable environment and has less electricity consumption	O	Private	probabilistic	Intel
Proof of authority (PoA)	Similar to PoS and DPoS. Blocks are only created when a consensus is reached among the validators. In this network there is something of a decentralization where the validators are pre-selected.	+	Public, private or consortium	probabilistic	ovan, Rinkeby, Giveth, TomoChain, Rublix, Swarm City, Colony, Go Chain.
Proof of Burn (PoB)	Coins owned by a node and obtained from a previous protocol such as Bitcoin are burned to gain the privileges of creating new blocks in that network.	o	public	Immediate	Slimcoin and Counterparty
Proof of Capacity (PoC) or Proof of Space (PoS)	It is the closest to PoW with differences. Instead of making the effort to check, each block the effort is made in advance and is called plotting. The results are used to verify the future block.	++	public	probabilistic	Burstcoin and Bitcoin Ore
Proof-of-Stake-Time (PoST)	The longer a node is in the network, the more probable to participate in the block mining	+	public	probabilistic	VeriCoin Blockchain Explorer
Proof-of-Brain (PoB)	This protocol enables intelligent people to enter the network through their valuable social participation		public	probabilistic	Steemit
Proof-of-Physical-Address (PoPA)/ Proof-of-Bank-Account (PoBA)	Rely Blockchain with the physical address	+	Private	Immediate	ConsenSys and POA Network
Proof-of-concept (PoC)	It is a copy of the concept of proof of concept in Blockchain. An example is the feasibility of check of insured vehicles.	o	Private	Immediate	Shared KYC Platform, FinTech
Our proposed framework	The protocol enables adding block based on the reputation of node	++	Semi-public	Immediate for master node/probabilistic for new nods	-

IV. Proposed Framework

A level of trust among parties is necessary for most forms of transactions. However, if an immutable ledger of data existed, we could remove this dependency on others. A Blockchain-based news consortium aims to do just that by using a decentralized network. With no central authority, it is exceedingly hard to alter the data inside a ledger. An agency would have to control over half of the platform to create a consensus. This fundamental principle leads to a secure system. As a result, certain industries are moving their processes over to this new technology, particularly those that require a greater degree of trust. Now we briefly present core functionalities of our proposed framework including fake news detection technique, node accreditation process, and news verification scenarios to illustrate the working of our proposed framework.

A. Detection Framework

Blockchain is a decentralized system in which there is a requirement of consensus between nodes to perform a secure transaction. We classify nodes as Master or untrusted nodes in our proposed customized blockchain network as shown in Fig. 1. Master nodes are selected based on their good past reputation. We suggest that the news issuance and verification process starts from Master Nodes in the private network of the Blockchain, and each node has its own contracts and database that contains the correct news. The master node is reliable nodes, such as the nodes in Twitter, which have a blue checkmark, or like Facebook.

For the news that is not issued from a reliable node, such as P1, P2, or P3, an automatic and intelligent mechanism will apply to verify the validity of the news. On the other hand, each node of the Blockchain contains complete information, so searching for it will be somewhat slow, especially in the case of many blocks in the chain. Therefore, we propose to develop a news indexing process through keywords that are stored in the nodes themselves, which will facilitate and speed up the verification process.

This research proposes an improved framework that integrates Blockchain and text-mining algorithms. Our proposed mechanism will verify the validity of the news received or the Islamic Fatwa from the Semi-Trust environment. For the proof of concept, we presented two scenarios for our proposed framework (Fig. 1).

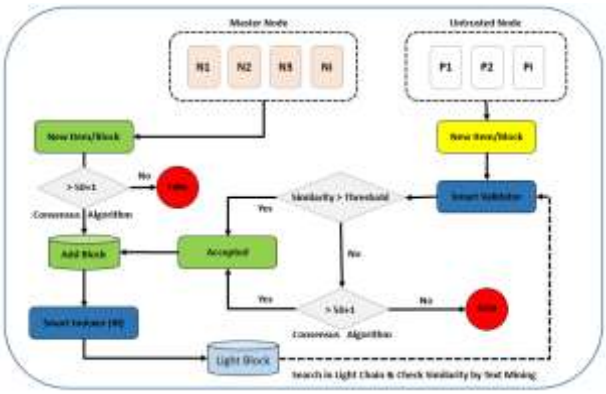

Fig. 1. Proposed Framework – S2.

The list of trusted nodes is formed by the rate of participation of the node in verifying the news and its authenticity as trusted news. In case the news is verified is authentic then the verifying node gets a score of +1 and a score of -1 is assign in case of wrongly verifying news. After reaching the desired score, the node is upgraded as the master node and gains a higher power to vote and verify the news.

A group of unreliable public contracts representing individuals or entities that are not accredited. Here we will have different scenarios to work according to the nature of the system and the level of sensitivity of the information as well as the type of node created for the news. We will discuss all these scenarios in the following paragraphs.

For adding a new block, the voting is done by the trusted nodes in the proposed model as shown in Fig. 2. A new block consists of the contents of the block's header. Then the proposed TM algorithm finds the keywords from the news text and stores them instead of the whole text to reduce the size significantly by more than 80%.

The voting process will take place in the first stage through the TM algorithm to measure the degree of similarity with other reliable news, and this will be completely automated. The transition to the traditional voting model is carried out so that each node (as an editor of the news) votes on the news through manual verification (here only trusted nodes) in the event that the achieved percentage is not higher than the threshold. This threshold will be determined based on experience and testing, which can be adapted to the nature of the news, for example, technical, sports, political.

B. News Verification Scenarios

1) Scenario 1 - A trusted node added a news (Algorithm 2): When a trusted node adds news, there is no need for voting or get a consensus on the news. A new block will be created that contains in its header the hash code of the previous node, as well as the date of its creation, block creator, a random number, and the news itself. This is all to ensure that it is connected to the previous block in the chain, and to ensure that its content is not modified.

At the same time, important keywords will be found in the news via the algorithm (3) based on the principles of information retrieval algorithms and the processing of natural texts. A parallel series of news will then be updated, but with much less information. Therefore, we called it a light chain.

Header	- Hash of previous block (HA_P) - Time Stamp (DT) - Owner Name (N1) - Nonce (n) Unique Random Value - Hash (Data + HA_P + N1 + DT + n) - Number of Block (ID) (Last ID+1)
Content	Text Data
Signature	Encryption of (Hash) by Public Key of N1

Fig. 2. Block Format in our System with Text News.

The importance of previous light chain and index steps will be illustrated in the following untrusted contract scenarios. Importantly Blockchain is used in the process to ensure that the data is not modified and saved in a multi-copy version of all trusted contracts, for future verification. However, accepting the news directly by the trusted node and canceling the consensus process will be linked to the sensitivity of the information or application. Therefore, a constraint in the consensus process is added before the news or information is added to the blockchain. For example, a consensus score of 10% of a trusted contract can be accepted as an alternative to the 50+1 score used in public chains. So here, we have reduced the load and accelerated the performance by overcoming some unnecessary steps.

Algorithm 1 : Trusted_node_Validation method

Input: Node N, Message Msg
Output: Accept/ reject, Update (Lightweight.Chain, CA.Index)
Begin
Accept (Msg)
Nonce ← **Generate**_Random_ID()
P_Hash← **Get**_Hash_of_Last_Block_in_the Chain()
N_Hash←**Hash**_Function(Msg, N, Date, P_Hash, Nonce)
Header ← New Header (N, Date, Nonce, N_Hash)
NBlock←**Create**_New_Block(Header, Msg)
Broadcast (List_of_Trusted_Block[], NBlock)
K ← **Find**.Keywords (Msg) // Details of this function will explain later
Update(Lightweight.Chain(K))
Update (CA.Index (K))
End

2) Scenario 2 - An untrusted node added a news (Algorithm 3): In case that the news is coming from a member of un-trusted nodes, then an intelligent principle is applied to verify the consensus through three successive phases of verification. It stops when a threshold is achieved.

Phase one:

A proposed similarity measurement algorithm (Algorithm 4) will be used to verify the node. This algorithm will search within the Certified Authority (CA) example master node index to find block numbers that contain similar content in the series. Then the search will be in the light chain within only the specified block numbers, and the proposed similarity algorithm (Algorithm 4) will be used to measure the similarity ratio with each block, and if a similarity is found greater than a threshold, it is considered as a positive vote for the new block. Similarly, one or more votes are required depending on the nature of the application. It is also possible to check more than one series (i.e. at more than one reliable node) to check for the presence of a fake or hacked chain.

Phase two:

If no similarity is found with a previous block, at this stage the similarity algorithm will be applied to all unsupported blocks that have been submitted by the untrusted nodes. If the similarity ratio with a certain number of nodes is greater than the specified threshold, the new block will be trusted.

Phase three:

If the verification is not achieved in the previous phase, then the third phase is initiated, which is the traditional vote request from the trusted nodes. The news is accepted only if it is voted (approved) by a certain number of trusted nodes, for example, 10%, or the traditional principle of Blockchain can be relied upon and a vote of 50+1 of all nodes.

Here, the reputation of the nodes that participated must then be modified based on the comparison between the type of voting and the final result of the block.

Algorithm 2 : UnTrusted_node_Validation method

```
Input: Node N, Message Msg
Output: Accept/ reject, Update (Lightweight.Chain, CA.Index)
Begin
        Consensus ← 0
        Block_IDs[] ←Search_in_CA_Index(Find.Keywords(Msg))
        Msgs[] ← Get_Light_Block (Block_IDs[])
        foreach Msg~ in Msgs do
                Sim ← Check_Similarity ( Msg~, Msg )
                If Sim > Threshold2 Then
                Consensus ← Consensus+1
                end if
        end for
        if Consensus > Threshold2 then
         Accept ( Msg )
                Nonce ← Generate_Random_ID()
         P_Hash ← Get_Hash_of_Last_Block_in_Chain()
                N_Hash ← Hash_Function(Msg, N, Date, P_Hash, Nonce)
                Header ← New Header(N, Date, Nonce, N_Hash)
                NBlock ← Create_New_Block( Header, Msg)
                Broadcast ( List_of_Trusted_Block[], NBlock )
                K ← Find.Keywords ( Msg )
                Update ( Lightweight.Chain(K) )
                Update ( CA.Index (K) )
        else
                foreach Msg~ in Un_confirmed_Msgs
                        Sim ← Check_Similarity ( Msg~, Msg )
                        if Sim > Threshold1 then
                                Consensus ← Consensus+1
         end for
                if Consensus > Threshold2 then
                Accept ( Msg )
                Nonce ← Generate_Random_ID()
         P_Hash← Get_Hash_of_Last_Block_in_Chain()
                N_Hash←Hash_Function(Msg, N, Date, P_Hash, Nonce)
                Header ← New Header(N, Date, Nonce, N_Hash)
                NBlock←Create_New_Block( Header, Msg)
                Broadcast ( List_of_Trusted_Block[], NBlock )
                K ← Find.Keywords ( Msg )
                Update ( Lightweight.Chain(K) )
                Update ( CA.Index (K) )
                else
                        Consensus ← Traditional_Consensus( Msg )
                        if Consensus > Threshold2 then
                                Accept ( Msg )
                        else
                                Reject (Msg )
          end if
                end if
        end if
End algorithm
```

Extracting keywords from a message:

Algorithm 3 : Find Keywords (Msg) method // extracting keywords from a message

```
Input : message
Output : keywords
Begin
CMsg ← Clean_Text (Msg) // Replace Special Characters by " " by using Replace function or Regex
NMsg ← Normalize ( CMsg ) // Convert all characters to small letters by using convtolower
Terms [] ← Tokenizing ( NMsg ) // Convert sentences to vector of terms by using Split (" ")
Terms [] ← Removing_StopWords (Terms[], English_Stopwords_List)

// Removing un important terms as " the a an in on they I he … "
foreach term in Terms[]
        if English_Stopwords_List.Contain(term) then
            Terms.Remove (Term)
            Terms [] ← Porter_Stemmer ( Terms [] )
            HashTable HTerms ← Find_Freq ( Terms )
        end if
end for
foreach term in Terms
        if HTems.Has (term) then HTerms[term] ←+1
            else HTerms[term] ← +1
        end if
end for
// Find weight | importance of each term by using IF-IDF
WTerms ←Calculate_IDF (HTerms)
foreach term in HTerm
         W ← ((Freq_Term_in_Doc) / (Max Freq in Doc)) * Log (Num of Docs have Term / Num of Docs)
        RTerms ← Remove_unimportant_term ( WTerms, Threshold )

// Remove who importance less than Threshold
foreach term in RTerms
        if term.W < Threshold then RTerms.Remove(term)
        end if
end for
foreach Term in RTerms
        Keywords [] ← New Keyword ( Term, Freq_in_Doc )
end for
return Keywords[]
```

The algorithm 4 calculates the similarity ratio between two messages. We relied on WordNet [1] anthropology to calculate the similarity between two messages by calculating the distance to their nearest root within the WordNet dictionary tree.

Algorithm 4 : Find Similarity between two terms

```
Input : list of message
Output : the percentage of similarity

Max ←0
Result ← 0
Keywords ← Find_Keywords (Msg)
Keywords~ ← Find_Keywords (~Msg)
for Key1 in Keywords
        for Key2 in Keywords~
                Sim ← Calculate_Similarity ( Key1, Key2)
                if Sim > Max then Max ← Sim end if
        end for
        Result ← Result + Max
        Max ← 0
end for
return Result.
```

To sum up, all algorithms integrate with each other to complete the news verification process. The traditional first voting algorithm of the master nodes is to accept a piece of news, and at the end of the algorithm, the fourth algorithm is invoked to form the modified light chain. Here, those who vote incorrectly will lose a point of reputation credit. After a while, if it continues, they will return as a normal node. In order to verify a piece of news, the fifth algorithm is used by measuring the percentage of similarity. As the participant wins a point in reputation and in the future he may become a master node.

V. Result and Discussion

To test our proposed algorithms, we divide the testing process in five main stages:

- Testing the algorithm that finds the most important words in each news to form the Lightweight chain.
- Testing the Similarity Algorithm to determine the percentage of similarity in the news.
- For the BlockChain, we built our own software to prove the applicability of the proposed model. A new block is generated and a chain is formed. We also tested the modification process on a block in addition to the modification process on one of the chains (that is, the news of the basic operations included in any BlockChain platform).
- Finally, we evaluate the performance of the proposed protocol according to the traditional consensus algorithm in the BlockChain (Table I).
- Our proposal was tested on the Ethereum environment, in addition to an online BlockChain simulator [28] and distributed to users.

A. First - an Algorithm to Discover Important Words

We implemented the proposed algorithm using Visual Studio .NET platform and tested it on a sample of news selected from Twitter as experiential dataset by using TF-IDF as a criterion for importance, which is used in most search engines. Then we tested 12 queries (12 new news items) and based on the number of important words returned from each query, the values TP, TN, FN, FP were calculated according to their respective equations 1,2 & 3. Finally, the main criteria to measure accuracy is calculated such as Accuracy, Precision, and F1-Score. The results show reasonably high percentage of all four criteria of accuracy in Fig. 3, with an average accuracy of approximately 91%. However, these results are from a small sample and in future we will optimize our algorithm to works with huge datasets with higher accuracy.

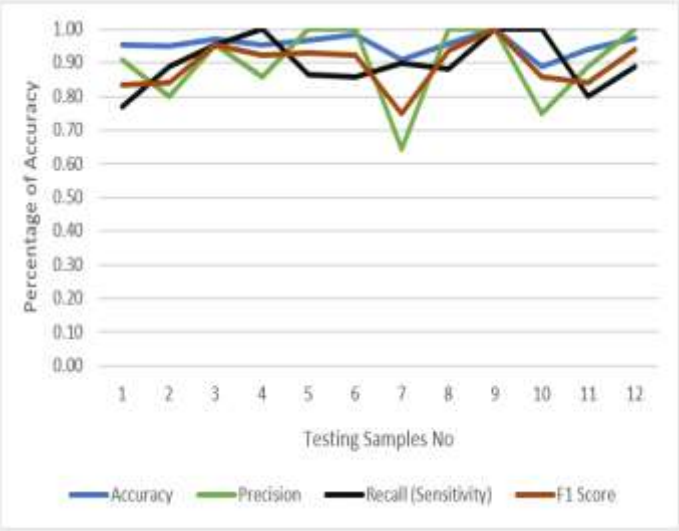

Fig. 3. Features of Proposed Algorithm.

	TP	FP
	FN	TN

$$Accuracy = \frac{TP + TN}{All} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

- TP denotes that the two texts are alike and the result is correct.
- TN denotes that the two texts are not the same and the result is correct.
- FP denotes that the two texts are the same but the result is wrong (here the traditional consensus algorithm will be applied and the error effect will be less than the following case).
- IFN: Indicates that the two texts are not the same, but the result is wrong. That is, the information will be accepted without the need for consensus, and this is a state of great danger.

B. Second - Similarity Measurement Algorithm

Initially the algorithm was programmed using Visual Studio .NET, and then we tested it on a group of news articles, chosen from various topics from the BBC platform. Then we calculated the similarity ratio on each pair of articles using the proposed algorithm, and then compared the result with the average results that were measured by three similarity ratio arbitrators.

Finally, we estimate the MSE (mean square error) using equation (4) [16], where its value is 6.9. This low value of MSE represents very good accuracy of the proposed algorithm. The results of the comparison are shown in Fig. 4. The average of Manual similarity is calculated based of the result of similarity of real news against news widespread on some social media sites. A group of linguistic expert does this calculation.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}[(y_i - \hat{y_i})^2] \quad (4)$$

C. Three - BlockChain

The Visual Studio.Net platform was also used to build our own simulator to achieve the basic operations in the BlockChain: creating a new block and adding it to a chain, in addition to testing a modification process on a specific block at the end or middle of the chain. Finally, the process of synchronization is between several chains.

Fig. 5 shows our simulator. The main goal of using the simulator with the previous two algorithms is to ensure the testing of the complete proposed platform because it is difficult to modify the workings of existing BlockChain platforms. Also, in this research, we proposed a different method of management compatible with the new and proposed use in verifying the validity of news. This all makes it difficult to compare the proposed platform with another existing platform, due to the different purposes of the application and the mechanism of work.

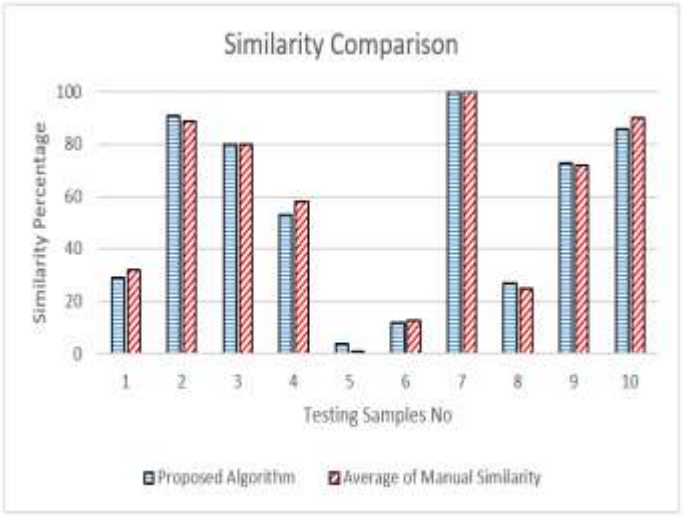

Fig. 4. Comparison of Similarity Check.

Fig. 5. Error appeared in Block 3 and the following after Modifying the Data in this Block during the Test.

Finally, in order to integrate all the ideas together (Blockchain with similarity algorithm) we built our own simulator for all Block-Chain functions (where a formed block containing: a unique number, a random number, the date of construction, the hash code of the previous block, in addition to the hash code of the block itself). When a block is added, the similarity algorithm will be applied, and in case of an acceptable similarity percentage, it will be added in green directly; otherwise, normal voting will take place.

We used SH1 as a Hashing algorithm; we built the chain and giving the power to modify a particular block for experimentation purposes (Fig. 6). Just modifying any block from the chain will be re-evaluated, the modified block color will change, and thus all the block that follows in the chain. Therefore, this chain at a certain node will be different from the rest of the chain, and any new block that sends from this node will be rejected immediately because the hash code will be different from the rest when calculated for the verification process.

For the Blockchain, several methods have been adopted to implement the proposed system and emphasize its effectiveness and the possibility of its implementation. Initially, we used a ready platform like Ethereum to build the system. But we faced difficulty in adjusting the consensus algorithm (Fig. 6 and 7). Therefore, the second solution of the simulation platform [28] is used to clarify the form of a chain in the proposed system. In addition, the solution shows the advantages of using Block-Chain use in terms of ensuring Non- Repudiation, with the error detected immediately and spread to all elements of the following chains in case of a modification in any previous adopted news [30].

D. Fourth - Discussing the Performance of the Proposed Model

We point out that the proposed algorithm will achieve a tangible improvement in performance, but that is conditional on the nature of the use as well (for example, verifying the validity of news) in which the similarity ratio can be assessed. This means that the news in which there may be repetition or similarity in the contents of its blocks, but with a different wording.

Here the proposed method can provide a verification mechanism instead of the traditional consensus mechanism, which can be hacked with a 50+1 attack, although current platforms have tried to solve this attack by switching to establish consensus based on proof of good past history. However, it must be noted that it is only suitable for certain text-based applications.

Performance cannot be truly assessed until we fully apply the platform to real data with large numbers of users. Nevertheless, the similarity checking process is a new test factor that limits the effects of the previous attack and saving time in the event of immediate adoption of the news.

However, if the news is completely new, this will affect the performance somewhat for the worse, as the time will be (T1 + T2). Where T1 represents the time taken to apply the similarity algorithm, and T2 represents the time required for the consensus algorithm. Hence, the performance of the proposed model will be as follows

$$performance = T_1 * R + T_2 * (1 - R)) \quad (4)$$

where, R is the rate of finding the proof based on similarity.

If T1<T2 is taken for granted due to the dependence on only part of the chain (trusted users) then it can be said that performance will be better when R>0.5.

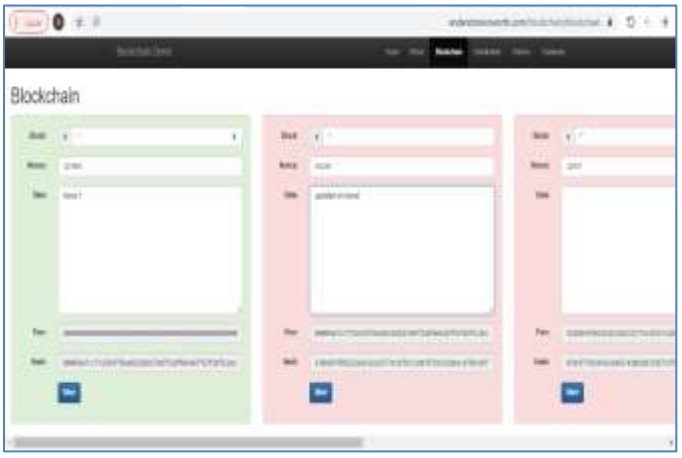

Fig. 6. How the Error Spreads once the Data is Modified.

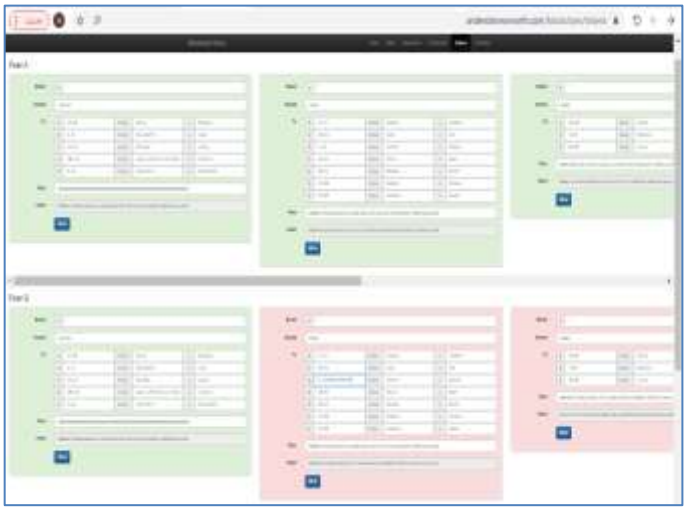

Fig. 7. An Example of more than One Peer of How One of Them has a different Block.

VI. Conclusion

Preventing the dissemination of fake news on various online platforms is one of the major research challenges for the research community. Considering the importance of detecting fake news to stop its dissimilation, we propose a blockchain-based platform to detect fake news. We slightly customize the structure of BlockChain to form a lightweight chain. Then establishes the process including algorithm and technical details to detect fake news. Text mining plays an important role to determine the authenticity of the tested news and its similarity index with existing news. Testing results of the basic simulation scenarios suggests the effectiveness of the proposed solution as a proof of concept. However, in the future, we are planning to perform comprehensive testing in various scenarios. We also plan to test the proposed solution for other similar applications such as verifying the authenticity of published religious rulings (fatwas).

Also, through this research, we started by launching a small initiative of a new consensus algorithm that will be based on the reputation of the participating nodes. This algorithm will be developed and all necessary tests will be performed to become a new consensus algorithm. This last will be effective in such type of applications.

References

[1] http://www.wordnet-online.com/anthropology.shtml.

[2] S. W. H. L. K. Shu, Understanding User Profiles on Social Media for Fake News Detection, Semanticscholar, 2019.

[3] A. A. Monther Aldwairi, Detecting Fake News in Social Media Networks, Abu Dhabi: ScienceDirect, 2018.

[4] P. C. R. S. S.-S. a. J. T. Hamid Karimi, "Multi-Source Multi-Class Fake News Detection," in Internatonal Conference on Computational Lingusitics, Michigan State, 2020.

[5] E. N. S. Mohamed Torky, "Proof of Credibility: A Blockchain Approach for Detecting and Blocking Fake News in Social Networks," International Journal of Advanced Computer Science and Applications, vol. 10, 2019.

[6] J. J. T. Zonyin Shae, "AI Blockchain Platform for Trusting News," in IEEE 39th International Conference on Distributed Computing Systems (ICDCS) , 2019.

[7] J. I. J. S. S. H. S. S. A. ,. K. D. S. Paul, " Fake News Detection in Social Media using Blockchain," in 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, 2019.

[8] "NewsCop | Fake News Detector," Devkey, [Online]. Available: https://play.google.com/store/apps/details?id=com.surajgiri.newsage&hl=en_US. [Accessed 10th March 2020].

[9] "ELISA - Fake News Detector," RoboMx, [Online]. Available: https://play.google.com/store/apps/details?id=tech.robomx.elisa&hl=en_US. [Accessed 10th March 2020].

[10] "Oigetit Fake News Filter," Oigetit, Inc., [Online]. Available: https://play.google.com/store/apps/details?id=io.scal.oigetit&hl=en_US. [Accessed 10th March 2020].

[11] Ahmad Alkhodre, Toqeer Ali, Salman Jan, Yazed Alsaawy, Shah Khusro and Muhammad Yasar, "A Blockchain-based Value Added Tax (VAT) System: Saudi Arabia as a Use-Case" International Journal of Advanced Computer Science and Applications(IJACSA), 10(5), 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100588.

[12] Balouchestani, Arian & Mahdavi, Mojtaba & Hallaj, Yeganeh & Javdani, Delaram. (2019). SANUB: A new method for Sharing and Analyzing News Using Blockchain. 139-143. 10.1109/ISCISC48546.2019.8985152.

[13] A. Dhiran, D. Kumar, Abhishek and A. Arora, "Video Fraud etection using Blockchain," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 102-107, doi: 10.1109/ICIRCA48905.2020.9182963.

[14] M. Saad, A. Ahmad and A. Mohaisen, "Fighting Fake News Propagation with Blockchains," 2019 IEEE Conference on Communications and Network Security (CNS), Washington DC, DC, USA, 2019, pp. 1-4, doi: 10.1109/CNS.2019.8802670.

[15] Yi Lin,A note on margin-based loss functions in classification, Statistics & Probability Letters, Volume 68, Issue 1,2004, Pages 73-82, ISSN 0167-7152,

[16] S. Gilda, "Evaluating machine learning algorithms for fake news detection - IEEE Conference Publication," Ieeexplore.ieee.org, 2019.

[17] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier - IEEE Conference Publication," Ieeexplore.ieee.org, 2019.

[18] A. Dey, R. Rafi, S. Hasan, and S. Kundu, "Fake news pattern recognition using linguistic analysis," Dspace.bracu.ac.bd, 2019.

[19] A. K. Das, A. Ashrafi and M. Ahmmad, "Joint Cognition of Both Human and Machine for Predicting Criminal Punishment in Judicial System," 2019 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019.

[20] FaNDeR: Fake News Detection Model Using Media Reliability- IEEE Conference Publication.

[21] S. Parikh and P. Atrey, "Media-Rich Fake News Detection: A Survey," Semanticscholar.org, 2019.

[22] A. K. Das, T. Adhikary, M. A. Razzaque, E. J. Cho and C. S. Hong, "A QoS and profit aware cloud confederation model for IaaS service providers," ICUIMC '14 Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, Siem Ream, Cambodia, 2014.

[23] H. Al-Ash and W. Wibowo, "Fake News Identification Characteristics Using Named Entity Recognition and Phrase Detection," Semanticscholar.org, 2019.

[24] F. T. Zohora, M. R. R. Khan, M. F. R. Bhuiyan and A. K. Das, "Enhancing the capabilities of IoT based fog and cloud infrastructures for time sensitive events," 2017 International Conference onElectrical Engineering and Computer Science (ICECOS), Palembang, 2017 , pp.224-230.

[25] T. Adhikary, A. K. Das, M. A. Razzaque, M. E. H. Chowdhury and S. Parvin, "Test implementation of a sensor device for measuring soil macronutrients," 2015 International Conference on Networking Systems and Security (NSysS), Dhaka, 2015, pp. 1-8.

[26] K. Shu, S. Wang, and H. Liu, "Understanding User Profiles on Social Media for Fake News Detection," Semanticscholar.org, 2019.

[27] M. Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro, and L. de Alfaro, "Automatic Online Fake News Detection Combining Content

and Social Signals," Dl.acm.org, 2019. The 7th International Conference on Smart Computing & Communications (ICSCC).

[28] https://andersbrownworth.com/blockchain/blockchain.

[29] S. Shabani and M. Sokhn, "Hybrid Machine-Crowd Approach for Fake News Detection," Semanticscholar.org, 2019.

[30] S. Yu, K. Lv, and Z. Shao, "A High-Performance Blockchain Platform for Intelligent Devices - IEEE Conference Publication", Ieeexplore.ieee.org,201.