ProBlock: a novel approach for fake news detection

Eishvak Sengupta[1] · Renuka Nagpal[1] · Deepti Mehrotra[1] · Gautam Srivastava[2,3]

Abstract

The world is diving deeper into the digital age, and the sources of first information are moving towards social media and online news portals. The chances of being misinformed increase multifold as our reliance on sources of information are getting ambiguous. Traditional news sources followed strict codes of practice to verify stories, whereas today, users can upload news items on social media and unverified portals without proving their veracity. The absence of any determinants of such news articles' truthfulness on the Internet calls for a novel approach to determine the realness quotient of unverified news items by leveraging technology. This study presents a dynamic model with a secure voting system, where news reviewers can provide feedback on news, and a probabilistic mathematical model is used for predicting the truthfulness of the news item based on the feedback received. A blockchain-based model, ProBlock is proposed; so that correctness of information propagated is ensured.

Keywords Fake news · ProBit model · Blockchain · Probabilistic model

1 Introduction

In recent years, fake news and rumours have been a cause of significant societal losses. Misinformation, in the form of doctored articles, memes, and unverified posts from anonymous users has triggered multiple incidents in the real world that have caused the loss of life and reputation worldwide. The year 2019 was termed the "Year of Fake News" by The Economic Times.[1] The distribution of misinformation on sensitive socio-political issues has caused widespread outrage among citizens, even leading to riots. India, which leads the world social media user charts, is at a greater risk of being affected by the spread of such fake news, hate propaganda, and rumours. Major factors for the spreading of fake news include social factors, cognitive factors, political factors, financial factors, and malicious factors [1]. In 2016, a series of false tweets by Netizens about a pizza joint being part of a pedophile sex ring involving former U.S.A. secretary of state Democratic presidential candidate Hillary Clinton and her campaign members triggered a shooting incident in Washington, D.C., U.S.A.[2] The primary reasons for the prevalence of fake news have been revenue generation [1]. clicks (or clickbait), inculcation of social and political biases in the minds of the audience, and the degradation of social image of individuals, groups, or organizations [2, 3]. As fake news cases rise, solutions must be deployed to identify and stop the spread of this misinformation online.

Expert knowledge and machine learning algorithms are two common approaches used to identify the fictitious articles deliberately fabricated to deceive readers [4].

✉ Gautam Srivastava
srivastavag@brandonu.ca

Eishvak Sengupta
eishvak.sengupta@student.amity.edu

Renuka Nagpal
rnagpal1@amity.edu

Deepti Mehrotra
dmehrotra@amity.edu

[1] Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India

[2] Department of Mathematics and Computer Science, Brandon University, Brandon R7A 6A9, Canada

[3] Research Center for Interneural Computing, China Medical University, Taichung 40402, Taiwan, Republic of China

---

[1] https://economictimes.indiatimes.com/news/politics-and-nation/fake-news-still-a-menace-despite-government-crackdown-fact-checkers/articleshow/72895472.cms.

[2] https://www.marubeni.com/en/insight/potomac/backnumber/19.html.

Computational approaches involve various Natural Language Processing based indexing and Deep Learning algorithms [5]. Fake news and hoaxes are created in all media types, including text, image, audio, or video clips. Researchers have developed efficient computational algorithms for one type of news article, but if one wants to develop a real-time system that can handle all types of fake news, the system itself will be very complex [6]. However, an expert-based system, though involving human input may have reduced complexity and increased efficiency. Furthermore, the date and context of news are equally essential for judging the worth of news pieces. It is often observed that an old news piece or some clipping of an old event or movie scene is recirculated as news. These issues themselves motivate us to look into expert-based detection methods. Our novel approach in this paper involves analyzing feedback given by one or more subject matter experts reviewing news items, one-by-one in a centralized manner, giving votes, labels, or scores to the news items and attempting to determine its degree of truthfulness [7]. The expert may differ in opinion, so while evaluating the score of trust for given news, the expert's votes should be weighted as per their expertise. It has encouraged us to explore a probabilistic model for calculating the score. Although all mentioned techniques show promising results and have been used widely, a centralized voting technique allows experts to give the score one after the other, thus lengthening the review process. Also, there is a threat to the security of new pieces and their corresponding votes caused due to the insecure method of their storage, making them easily editable with the integrity of votes being questionable in such a system. Therefore, to counter this known issue, in this paper we use blockchain technology to design the model.

In this paper, a framework called ProBlock is proposed and implemented as an efficient, secure, and reliable fake news detection technique. ProBlock uses blockchain technology to design a secured framework where the experts' votes are dynamically stored and shared. The blockchain is an immutable ledger to maintain all news pieces' entries and securely store their corresponding votes. The cryptographic encryption of blocks provides for a secure environment for storing the news pieces [8]. The blockchain ensures that the news pieces and votes cannot be changed or modified at any point in time [9]. A news piece, whose truthfulness is to be determined, is voted on by authorized reviewers giving scores based on its degree of genuineness and confidence about it. Next, the votes are weighted based on the credentials of the reviewer such as experience, designation, affiliations. As the reviewers come from diverse backgrounds and work experiences, some votes may be more trustworthy than others; hence a weighted system is vital and allows us to consider a more significant

number of factors to evaluate them in a more trustworthy manner. Profiling of the experts helps to get a better understanding of their credibility. We introduce the ProBit model, a probabilistic mathematical model that analyzes the weighted votes to predict their accuracy. The ProBit model allows us to consider the reviewers' features and the votes to generate a final score for credible deception detection. As the truthfulness label of the variable newsPiece (dependent variable) in this case can only take two possible values (genuine or fake), the ProBit model is the most appropriate model of this analysis [10]. The immutability of data stored on the blockchain is utilized to ensure no modification of the posted newsPieces and the simultaneous casting of votes by the reviewers in a distributed environment taking into account exogenous factors trust in the system.

The proposed framework for ProBlock is given in Fig. 1. The weighted majority voting model is implemented here by calculating a score based on the experts' interpretation of its fakeness. The weight of experts' votes is evaluated using a dynamic scoring approach where experts' career statistics and their confidence about their vote to the newsPiece is considered. An expScore is computed for each review cycle consisting of static and dynamic inputs for every expert. The static component of the expScore consists of a score based on the analysis of the reviewers' experience, organization of affiliation and designation. The dynamic score is computed based on the frequency of the reviewer and the accuracy of each review. The probability of the newsPiece being genuine is computed using the ProBit model where different experts' ratings and their expScore is considered as input. Based on the news rating, the false news is deleted from the blockchain; else continued in the blockchain, and finally, the upgraded rating of the experts is recalculated. The Proof of Trust consensus algorithm is used for the implementation of the model.

In this paper, a private blockchain has been leveraged to ensure the privacy of the reviewer votes at all times. ProBlock involves simplified data handling processes that are not accessible to every block. It offers a faster output with high power efficiency and at the same time ensures sufficient data privacy [11]. The distributed records created by each reviewer vote are transparent and immutable [12–14]. It finds major applications in securing and managing information systems by reducing dependency on other outer applications [15] .

Fake news and its influence have attracted researchers as this misinformation is considered a significant threat to journalism, and freedom of expression. In Sect. 2 an extensive literature review is conducted to understand the researcher's approaches to detect the deception in circulated information. The proposed model verifies the news using a secure weighted majority model implemented using
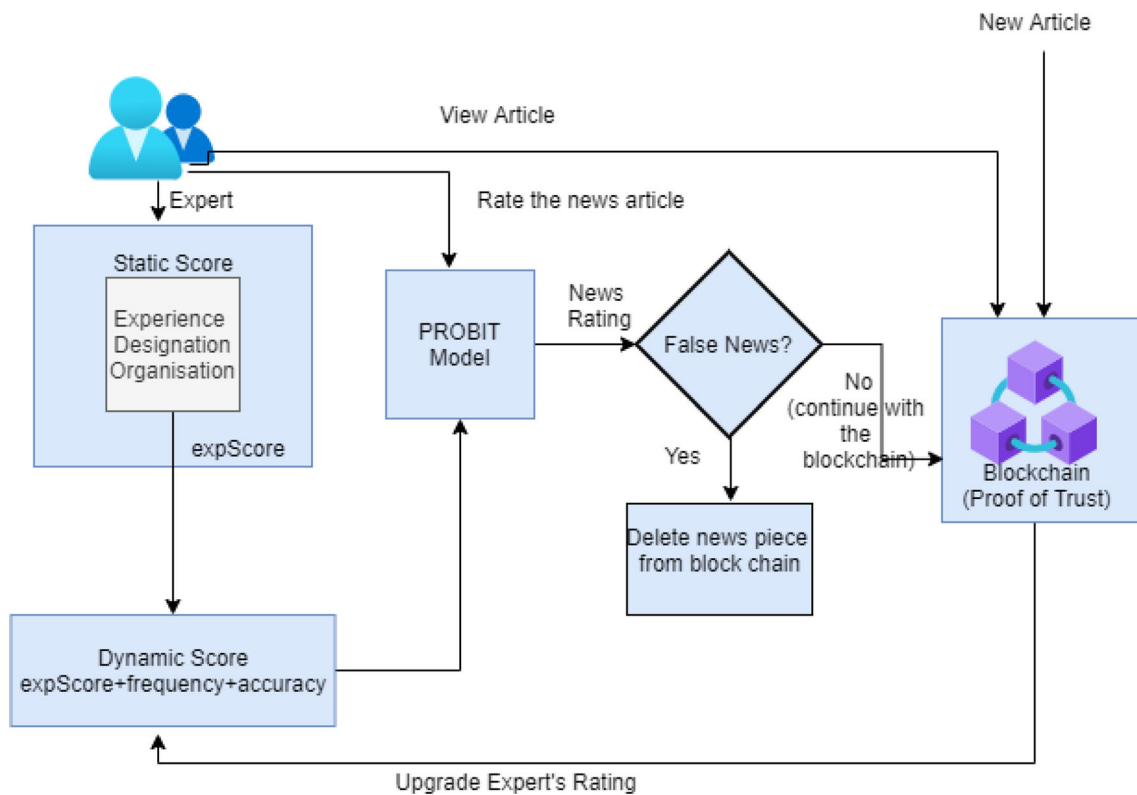
Fig. 1 Framework for fake news detection

blockchain technology. The methodology and concept used in the proposed model are explored in this section as well. The proposed methodology and numerical example, demonstrating an experimental simulation of ProBlock are given in Sect. 3. The algorithm of ProBlock and its implementation using blockchain is discussed in Sect. 4. The performance measure and result are given in Sect. 5, followed by some concluding remarks in Sect. 6.

2 Existing methods of fake news detection

The problem of detection of fake news online has been a very popular topic. The most vital need for the detection of fake content online is that it must be accurate, secure, and timely. Its accountability needs to be high on online social networks. Most of the methods used to address this issue consider it to be a classification problem giving fake or 'not fake' Boolean responses. Previously, Zhang et al. developed a credibility inference model for fake news detection by extracting explicit and latent features from the dataset building deep diffusive networks [16]. Some adopted methodologies try to rate given news on a fakeness scale. Mavroforakis et al. introduced a wide variety of supervised machine learning, deep learning, and data mining techniques that gave promising results [17]. The supervised machine learning techniques of a Support Vector Machine (SVM) include the formation of a hyperplane that divides the two classes of data after training on a set of labelled data points [17]. Kwon et al. implemented a Random Forest classifier trained with news temporal, structural and linguistic features that gave the Twitter graph dataset a precision of 0.90 [18]. Ferreira et al. designed a logistic regression for fake news detection as well but did not give impressive results on Emergent Dataset to a maximum of 0.74 [19]. Liliana et al. designed a novel Probabilistic Graph Model (PGM) that represented the probability distribution among given variables, while classification enables Conditional Random Field (CRF) to infer based on large sets of input features [20]. A CRF can be used in fake news and rumour detection that utilizes its learnings from sequential dynamics of news websites and social media posts with existing systems. It would not have to observe the news pieces or question its position but it would use the context-based features learned [2]. Ciampaglia et al. developed a classifier that gives an edge over existing sequential models by giving an F-measure of 0.6 on the PHMEME dataset [21]. A Tensor modelling method proposed by Seyedmehdi et al. captures latent relations between articles and terms, as well as spatial/contextual relations between terms, towards unlocking the full potential of the content. Furthermore, an ensemble method

that can consolidate and fine-tune the results of multiple tensor decompositions into a single, high-quality, and high-coherence set of article clusters [22].

Yildirim et al. implemented an ensemble-based learning approach where multiple classifiers and regressors models can be utilized to improve model performance and reduce the probability of selecting a wrong response has been implemented to assign a degree of confidence in a news piece [23]. Wang et al. also used a classification technique based on the prediction of an independent constituent is used to determine the veracity of a news piece [24]. The authors went on to use a series of content and context-based features to train the learning algorithms. More than 12,000 entries were manually labeled to make the ensemble achieve an accuracy of 0.77 on the RumourEval test set [25, 26]. A Hidden Markov model is a statistical model that is used to learn the basic information from given sequential data. Datasets in a time series format containing context and content-based features were used to train two such models. In this case, two models were used to keep track of the true and false data(s) respectively. Zou et al. experimented with the probability of outcomes of many models and their best result gave approximately a 0.75 accuracy [27]. An incentive-aware blockchain-based solution has been proposed by Chen et al., highlighting the prevention of fake news propagation by bringing together the benefits of blockchains and smart contracts along with a specially curated consensus algorithm [28].

Fake news detection methods based on knowledge include manual and automatic fact-checking that compares expert knowledge to the news pieces in hand to be verified [29]. Manual fact-checking techniques involve the labelling or voting for the news pieces in hand by a small group of domain experts giving fairly accurate results. PolitiFact is an American political news verification portal that analyses textual data to give labels like True; Mostly true; Half true; Mostly false; False; Pants on fire.[3] It follows a highly centralized architecture where domain specialists vote or one by one label the news items. Other platforms following a similar approach are The Washington Post Fact Checker,[4] FactCheck,[5] Snopes,[6] where the process of verification relies on expert knowledge. All these techniques provide 3–5 predefined labels for selected domain news verification. Crowdsourced fact-checking has also been leveraged on a large scale through individual congregations or websites. A group of regular individuals act as fact-checkers and cast their votes for the news pieces. This technique is highly unreliable as fact-checkers' credibility

is unverified. Their individual biases may lead to ambiguous results. Website-based crowdsourced news verification systems include the users uploading news articles and headlines with appropriate tags, and the sentences are rated to distinguish the content types (e.g., news versus non-news) and determine their veracity (true vs. not-true). The tags are included to study patterns as to where the probability of news being fake is highest. Yadav et al. proposed a secure voting mechanism for private computation using the Schulze voting method over the cloud [30]. The method uses homomorphic encryptions to perform computation over encrypted data and the computed result cannot be decrypted without a private key. However, the method's the model's computational complexity increases drastically by increasing the number of levels in it Existing automatic fact-checking methodologies rely on fact retrieval followed by natural language processing. The facts are extracted from the newsPiece being taken into consideration and processed to make it efficiently interpretable by the NLP analyzer [31]. The analyzer compares the extracted facts from the article to a "knowledge base" to generate an authenticity index.

Though these results display promising levels of accuracy, having a centralized approach makes them susceptible to modification and tampering. Blockchain technology enables a decentralized and distributed environment with no need for a central authority. Transactions are simultaneously secure and trustworthy because of the use of cryptographic principles. A distributed environment allows for immediate feedback from multiple networks which allows reviewers to participate simultaneously. ProBlock aims at implementing a secure voting and news storage environment for the detection of fake news via majority voting. Fake news on the internet is in the form of textual, image, and video-based forms. A machine-based approach would involve the development and implementation of a highly complex system that would be able to process all types of data efficiently. On the other hand, an expert-based system would be relatively less complex, be more efficient, and involve greater accuracy in real-time.

3 Proposed methodology

Section 2 shows the need for a robust fact-checking mechanism that is secure and trustworthy. ProBlock aims to implement a secure voting and news storage environment for detecting fake news via a weighted majority voting.

---

[3] https://www.politifact.com/.

[4] https://www.washingtonpost.com/news/fact-checker/.

[5] https://www.factcheck.org/.

[6] https://www.snopes.com/.

3.1 Majority voting

A pure majority voting model is a decision that selects alternatives that have a majority, that is, more than half the votes. Each vote is equal and holds the same value. A majority voting model can be incorporated for the detection of fake news. A fake news identification majority voting model would include a panel of journalists, experts, and reviewers who would study and analyze news items before they are uploaded to the portal and give boolean decisions on their veracity. Every reviewer or expert can pass a vote on the truthfulness of the news after thorough verification of the news item. A pure majority voting model would ideally have only two votes, real or fake, and hence the outcome of the voting process would be deterministic, excluding the case of a draw. Each vote is assumed to be equal, free, and fair. However, the pure majority voting model is precarious. The factor that makes this pure majority voting model successful for detecting fake news is that the votes of all reviewers or experts may not be on the same level. In terms of trustworthiness and experience, some experts outweigh others. Biases may get incorporated into the system based on their organization of affiliation. Their judgment capabilities and accuracy of reviews may vary. Hence, every vote in a majority voting method cannot hold the same value. This paper presents a weighted majority voting system where selected parameters and factors would be determining the weight of each vote for the detection of fake news via blockchain. The weight of each vote will be computed based on a predetermined rule system.

The weighted majority voting model is incorporated into the blockchain via the Proof of Trust (PoT) consensus protocol [32]. In PoT, a digital token is sent to the network users, and a special class of users (experts in this case) are sent a "puzzle" (generally a hash function or a simple integer factorization) which is to be solved, and its solutions are compared. The solution getting the highest number of responses is considered correct, and the block gets placed in the chain. In the given scenario, the "puzzle" is the voting system where the experts give votes (which are weighted as per calculations) and are compared to find the most common solution or range of solutions. If for a certain block, the solution lies in the parts to the completely confident range, the block gets incorporated into the chain. The PoT protocol eliminates the low throughput and resource-intensive problems linked to Proof of Work (PoW), while at the same time addressing the scalability issues known to exist in traditional Byzantine Fault Tolerance (BFT)-based protocols.

3.2 Voting system

In a voting system, multiple individuals come together to analyze the pros and cons of an object, thing, or situation by assigning labels or scores to it based on its characteristics. For ProBlock, a voting system is implemented by creating a class of users consisting of reviewers, subject experts, and journalists who are the voters of the majority voting system. Each newsPiece is reviewed and analyzed by this class of users, and each expert passes a semi-deterministic vote as a judgment on the veracity of the newsPiece. The weighted majority voting model is implemented here by calculating a score based on the experts' interpretation of its fakeness and his career statistics and a score based on their confidence about their vote to the newsPiece. An expScore is computed for each review cycle which is an integer score consisting of a static and dynamic basis for every expert. The static component of the expScore consists of a score based on the analysis of the reviewers' experience, organization, and designation as referred to in Table 1. The relative score is given to each criterion and help create the expert's profile, which helps attach greater accountability and trust with the vote. The profiling of the expert helps to attach greater accountability and trust with the vote. The integer score is calculated from the stated factors and is combined with the dynamic component. The dynamic component of the expScore is re-calculated after each review.

The dynamic component of the expScore is re-calculated after each review cycle is over. It is based on the frequency of the reviewer and the accuracy of each review. The dynamic component is also an integer value. Both the static and dynamic components are added to form the expScore.

Based on the number of correct predictions of reviews for the total number of reviews, the reviewer is given an accuracy score (accScore) which determines the success rate of the reviewer. It acts as a measure of dependability for the system. The consensus algorithm being used is a modified version of the standard PoT consensus protocol. The weighted voting and dynamic component further add multiple layers of trust towards the system by determining the credibility of the voter and his corresponding votes.

Table 1 ExpScore calculation metrics

Experience	Organisation	Designation
0–5 years—1	Freelancer—1	Jr. Journalist—1
5-010 years—2	Regional—2	Journalist—2
10+ years—3	Local—3	Sr. Journalist—3
	National—4	Executive—4
	International—5	Sr. Executive—5
Computes yrsScore	Computes orgScore	Computes desScore

These layers of trust also contribute towards more realistic scenarios of fake news detection as the voting process becomes much more credible. The expScore is calculated by calling the calExpScore() method as shown in Listing 1.

Listing 1 calExpScore()

```
Int calStExpScore(Reviewer r){
    stExpScore = r.yrsScore + r.orgScore + r.desScore;
Return expScore;
}
```

The expScore is calculated by calling the calExpScore() method as shown in Listing 2.

Listing 2 calExpScore()

```
Float calExpScore(Reviewer r){
    expScore = expScore + r.freqScore + r.accScore;
ReturnexpScore;
}
```

A confidenceScore indicates the surety of the vote cast by the reviewer in a range of -1 to $+1$. This confidenceScore is given by the reviewer itself and indicates his/her confidence in his vote. A scale is prepared where:

- 1: **Not confident**
- 2: **Confident**
- 3: **Highly Confident**

The newsVote is given by the reviewer on a scale of -2 to $+2$ indicating the reviewer's analysis of whether the newsPiece is genuine where:

- -2: **Completely False**
- -1: **Partially False**
- $+0$: **Neutral**
- $+1$: **Partially True**
- $+2$: **Completely True**

The confidenceScore is mathematically combined with the newsVote to calculate the cummVote i.e the cumulative vote.

The finalVote is then weighted with the expScore and hence the weighted majority voting model is implemented. The expScore is converted to a decimal value between 0 and 1 for the weighting purpose. The expScore is taken as the weight and the cummVote is taken as the vote in the weighted voting system.

3.3 ProBit model

A probabilistic analysis of the votes is made to determine the probability of the newsPiece taken into consideration being genuine or fake. The probability of the newsPiece falling into the score range of 'real' news is determined using the **ProBit** model [33]. The **ProBit** model is used to calculate the probability of occurance of binary-valued response variable Y in as a function of regressor X [11] as shown in Eq. 1.

$$Pr(Y = 1|X) = \phi(\beta X), \tag{1}$$

where Pr describes the probability of response variable Y taking value 1 and ϕ is the cumulative distribution function for normal distribution given in Eq. 2.

$$\beta X = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n, \tag{2}$$

where β = parameter of maximum likelihood. The parameter of maximum likelihood is defined as a function of regressor x, mean μ and standard deviation σ shown in Eq. 3.

$$\beta(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}, \tag{3}$$

where the Mean denoted by μ is described as shown in Eq. 4.

$$\mu = \frac{\sum_1^n x_i}{N}, \tag{4}$$

and Standard Deviation denoted by σ is defined as in Eq. 5.

$$\sigma = \sqrt{\frac{\sum_{i=0}^{N}(x_i - \mu)^2}{N}} \qquad (5)$$

The cumulative distribution function for normal distribution ϕ is given in Eq. 6.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-x^2}{2}} dx \qquad (6)$$

Apart from the constituents of the block, the votes cast by each expert and their corresponding weights (in the form of expScore) are also cryptographically secure. This encryption is done to ensure no tampering is done to the vote. A method is included which prompts a dialogue box on sensing any change in the hash of the votes.

The total number of reviews made by a reviewer is incremented every time a review is made by that particular reviewer. Furthermore, a monthly frequency is generated out of the reviews made by a reviewer in a particular month. The accuracy of the reviewer is further calculated for updating dynamic components of expScore and is expressed as the probability of a review of being correct based on the value stored in the correctness counter.

3.3.1 Numerical example

An experimental analysis of an online news portal, was created in a fashion so that the news outlet's goal would be to get all news items they have gathered to be verified before they are published on their platform. Table 2 represents a mix of genuine and fake news headlines with their respective details that are used in our analysis. Each news item (row) has information about that speficic story and has an *Actual Veracity* binary value of True or False (Fake). A total of 40 newsPieces are analyzed by 9 reviewers. The number of reviewers voting on the newsPiece keeps incrementally increasing in no particular order.

In Table 3, data of the reviewers is taken as input for the members of the reviewer class and the expScore is calculated. The factors on which expScore is evaluated for each reviewer are experience,organization they belong to, designation,number of reviews done per month. The factors on which expScore is evaluated for each reviewer are experience, the organization they belong to, designation, number of reviews done per month. Table 1, Table 3, Listing 1, and Listing 2 are used for calculating expscore.

A probabilistic analysis of newsPiece being genuine is made using the ProBit model. By using the ProBit model. Simultaneously, a probability is also calculated by counting the number of positively, negatively, and neutrally weighted votes for comparison.

3.4 Calculations

From the probabilistic analysis conducted in Table 4 using the ProBit model, we can see the probabilities of news items being genuine. We also conclude that the probability calculation using the ProBit model is much more accurate in classifying the newsPieces as genuine or fake.

For the experimental data, the first newsPiece being considered for review is Tech101 which gets a confScore of 3, 2, and 2 and a newsVote of 1, 2, and 1 from the three reviewers, respectively. This leads to a combined score of 3, 2 and 2. The cumulative scores get weighted by multiplying each of them with the expScores giving the finalScores. Using Eqs. 4 and 5, we get values of 23.33 and 9.84. Taking these values as input for Eq. 6 we get 0.912 indicating a percentage probability of 91.2%. Table 4 represents the results obtained by applying Eqs. 1 to 6 on the votes of the reviewers. We see in Table 4 that each newsPiece has specific metrics that we can calculate for it such as Standard Deviation, Probability of Fakeness, etc. These are all summarized in Table 4.

Figure 2 shows the Gaussian distribution of the probabilities. The area under the curve corresponds to all the news items having a probability of genuineness above the mean.

4 Implementation of model using block chain

4.1 Blockchain approach

A blockchain is a time-stamped series of immutable records of data that is managed by a cluster of computers not owned by any single entity. Each of these blocks of data (i.e. block) is secured and bound to each other using cryptographic principles (i.e. chain). Blockchain architecture of the proposed model consists of a `SHA256` encrypted blockchain. Each block of the proposed model consists of 2 key fields i.e newsPiece and `uploaderName` apart from the hashes and the timestamp. The newsPiece component can be uploaded by anyone. Open access is provided to all users to upload their desired newsPiece. The security that is provided to users by using cryptographic principles can make it widely popular and globally trusted. Public key cryptography is an integral component of blockchain. A public-private pair of keys is generated in the blockchain through high-order cryptographic algorithms. Effective security is maintained by keeping the private key secure whereas the public key may be given for access purposes. Such cryptographically encrypted security is used in the proposed weighted

Table 2 NewsPiece(s) with category, newsCode and actual veracity

S no.	Headline	News code	Tag	Veracity
1	Paytm requires location to log in now	Tech101	Technology	True
2	Richa Chadha shared an old tweet screenshot of Rishi Bagree (BJP IT cell worker) accepting that he spreads propaganda for BJP daily, which he claimed was a fake tweet and threatened her with a legal defamation case. Later it was found that the fake tweet is indeed real and was tweeted by him	Pol209	Politics	True
3	Netflix's new documentary on the Coronavirus includes Arunachal Pradesh AND Taiwan in their map of China	GPol655	Geopolitics	True
4	Coca-Cola Co. announces to to start producing milk bottles	Fnb897	Food	False
5	BJP lawmaker makes fun of PM Narendra Modi's 'clap, bang utensils' appeal	Pol218	Politics	True
6	Spurious Chart on 'Jihad' Leads to 'Insulting Religion' FIR Against Zee News Editor	Pol256	Politics	True
7	Testing increased to 95,000 tests a day	nCov332	Coronavirus	True
8	Indian companies are contributing lavishly to PM-CARES – even amid layoffs and pay cuts	bizFin987	Business and Finance	True
9	Israel names street after Rabindranath Tagore to celebrate his 159th birth anniversary	NPol143	Non Political	True
10	Twitter buys 5.2% stake in Reliance Jio	Tech241	Technology	False
11	Indian education can't go online—only 8% of homes with young members have computer with net link	PolEco892	Policy and Economy	True
12	Ranbir Kappor and Ranveer Singh to be casted in the same movie directed by Karan Johar	Bol456	Bollywood	False
13	Chennai bakery owner arrested for ad saying 'no Muslim staff'	NPol082	Non political	True
14	Azim Premji among the highest private donors for the covid19	nCov478	Coronavirus	True
15	Ahead of key WHO meet, Taiwan seeks India's support	NPol214	Non Political	True
16	AIIMS doctor reported to be infected by corona virus	nCov908	Coronavirus	False
17	Hong Kong Government Will Prioritize Bill to Make Booing China's National Anthem Punishable by Prison	WPol132	World Politics	True
18	Mumbai cop dies of infection as Maharashtra Police department records 786 cases	nCov243	Coronavirus	True
19	New education policy in India will be based on vedic knowledge and modern science	Edu823	Education	True
20	David Icke removed from Facebook	Tech927	Technology	False
21	Hungarian top court confirms Roma students unlawfully segregated at school, awards $310,000 to children's families	WPol802	World Politics	True
22	After 5 days of no new infections, Iceland will open to all tourists from June 15	NCov019	Coronavirus	True
23	Karnataka takes Rs 10,675 crore coronavirus blow	PolEco324	Policy and Economics	True
24	Large meteor headed towards earth may end life forever	Spax903	Space	False
25	Aarogya Setu scored 2 out of 5 stars in MIT's Covid Tracing Tracker	NCov819	Coronavirus	True
26	Britain would not support Israeli West Bank annexation: minister	WPol004	World Politics	True
27	IMF grants India $10bn fund to fight unemployment	Eco615	Economics	False
28	Japan's central bank gets its first woman executive director in 138 years	Eco187	Economics	True
29	India draws up Rs 1 billion Covid-19 medical assistance plan, targets 90 countries - india news	NCov521	Coronavirus	True
30	German chancellor replaced because of economic failure in COVID situation	WPoL90l	World Politics	False
31	In court filing, FBI accidentally reveals name of Saudi official suspected of directing support for 9/11 hijackers	WPol776	World Politics	True
32	Qatar denies permission to Air India India has been misinterpreted	WPol562	World Politics	True
33	Mexico's president orders military on streets to curb rising violence. Mexico's president has ordered the armed forces to tackle security on the streets for another four years, extending a policy he had previously criticized as the government struggles to curb runaway violence	WPol905	World Politics	True
34	India, China locked in tense military stand-off in eastern Ladakh	WPol 812	World Politics	True
35	Actor Tom Hanks in critical health condition post COVID infection	NCov298	Coronavirus	False
36	United States announces complete shutdown to tackle COVID	NCov097	Coronavirus	False
37	Dengue kills too'—Latin America faces two epidemics at once	Health128	Health	True
38	Chennai: Manager at herbal products firm tests corona 'drug' on himself, dies	Health765	Health	True

Table 2 (continued)

S no.	Headline	News code	Tag	Veracity
39	IPL scheduled to start from 15th June	Sports523	Sports	False
40	Support for Brazilian President Jair Bolsonaro has dropped as criticism grows of his handling of the coronavirus pandemic, according to a poll	WPol 124	World Politics	True

Table 3 Reviewer details with experience, designation. frequency

	Experience	Organisation	Designation (years)	Frequency of review /month	Calculated expScore
Reviewer 1 (r1)	13	National	Sr. Journalist	6	13
Reviewer 2 (r2)	3	Free Lancer	Journalist	10	7
Reviewer 2 (r3)	7	International	Journalist	6	12
Reviewer 4 (r4)	9	Regional	Sr. Journalist	7	10
Reviewer 5 (r5)	2	International	Jr. Journalist	9	12
Reviewer 6 (r6)	4	National	Journalist	11	10
Reviewer 7 (r7)	20	Regional	Sr. Journalist	4	10
Reviewer 8 (r8)	6	Freelancer	Sr. Journalist	12	9
Reviewer 9 (r9)	11	Local	Sr. Journalist	7	12

majority voting model. For more information on the security of blockchain technology, we refer readers to some informative survey articles on blockchain security and privacy [34, 35].

As the newsPiece gets uploaded with the uploaderName and timestamp, the cryptographic hash for the block is generated and the blockchain is formed where each block is linked to the next by the reference of the hash of the previous block. The newsPiece, uploaderName, hashes, and timestamp determine the generated hash. These hashes are immutable and are regenerative. Any change to the input data changes the output. Hence, any change to the input data means a complete change in the flow of the blocks of the chain. The hashes of all the block change from the block in which the change was made onwards.

Proof of Trust integrates trust components and is widely adopted in the service industry, and crowdsourcing environments as it can address the unfaithful behaviour of members of the public service network [36–38].Proof of Trust is a consensus algorithm that selects validators based on the predefined criterion and Shamir's secret sharing algorithms. The Proof of Trust protocol avoids the low throughput and resource-intensive pitfalls associated with Bitcoin's "Proof-of-Work" (PoW) mining while addressing the scalability issue associated with the traditional Paxos-based and Byzantine Fault Tolerance (BFT)-based

algorithms [32]. Proof of Trust is an extension of the Proof of Work consensus algorithm, a trustless leader "election mechanism" based on demonstration of computational power. Proof of Work provides blockchain security in trust-less Peer-to-Peer (P2P) environments but comes at the expense of wasting huge amounts of energy. A Proof-of-Trust blockchain where peer trust is evaluated in the network based on a trust graph that emerges in a decentralized fashion and that is encoded in and managed by the blockchain itself [39]. Efficient consensus BFT algorithms like RCanopus are utilized to make the process of extraction of queries and transactions faster and providing dedicated fast peer server channels [30]. As the newsPieces that are taken into consideration, the only data items in the public domain and the votes of reviews and scores associated with them are passed privately, generic smart contracts have been used in this scenario.

4.2 Algorithm of ProBlock

The newsPieces is put in the blockchain in the following manner with the hashes calculated using SHA256 hashing algorithm as shown in Table 5. Table 5 contains a full representation of the hashes of all the newsPieces on the blockchain. The consensus algorithm utilized by ProBlock securely collects the votes of the reviews and the votes are

Table 4 ProBit analysis of newsPieces

News code	Standard deviation	Cummulative distribution function	Parameter of maximum likelihood	Probability of fakeness	Percentage probability of the news being GENUINE is	Genuineness verdict based on ProBit model	Probability of genuine-ness WITHOUT ProBit model(mean method) (%)
Tech101	23.33	9.84	1.7494	0.912	91.20	Genuine	100
Pol209	15.33	10.87	1.314	0.839	83.90	Genuine	66.67
GPol655	48	12.96	19.116	0.5	50	Non-determinable	50
Fnb897	− 44.66	22.05	− 1.547	0.083	8.30	Fake	0
Pol218	19	35.86	1.3937	0.887	88.70	Genuine	66.67
Pol256	32	24.65	1.051	0.835	83.50	Genuine	100
nCov332	2.33	0.47	45.61	0.5004	50.04	Genuine	50
NPol143	31.3	9.56	2.661	0.7411	74.11	Genuine	100
Tech241	− 36	8.485	− 0.044	0.2941	29.41	Non-determinable	33.33
PolEco892	39.33	14.63	1.818	0.9092	90.92	Non-determinable	100
Bol456	− 30.33	7.76	− 3.019	0.3476	34.76	Genuine	33.33
NPol082	20	16.32	0.921	0.82	82.00	Non-determinable	66.67
NPol214	28	10.70	2.099	0.8726	87.26	Genuine	100
nCov908	− 28	27.195	− 1.0159	0.1574	15.74	Fake	0
WPol132	40	16.33	2.8558	0.7433	74.33	Genuine	100
Edu823	50.4	21.66	2.238	0.8369	83.69	Genuine	100
Tech927	− 30.4	11.82	− 1.202	0.1201	12.01	Fake	0
WPol802	43.2	19.16	2.51	0.7982	79.82	Genuine	100
NCov019	33.2	20.99	1.735	0.9199	91.99	Genuine	100
PolEco324	36.4	25.56	1.789	0.9174	91.74	Genuine	100
NCov819	31	23.32	1.64	0.9177	91.77	Genuine	100
WPol004	22.25	15.10	1.7727	0.9143	91.43	Genuine	100
Eco615	− 36.25	12.55	− 3.134	− 0.3162	31.63	Non-determinable	33.33
Eco187	49	8.33	3.892	0.5384	53.84	Non-determinable	50
NCov521	34.5	8.64	5.346	0.5006	50.06	Non-determinable	50
WPoL90l	− 25.25	22.21	− 0.941	0.1782	17.82	Fake	0
WPol776	30.33	21.67	1.18	0.5031	50.31	Non-determinable	50
WPol562	43.33	20.22	1.618	0.9198	91.98	Genuine	100
WPolx905	30	16.16	3.002	0.7063	70.63	Genuine	83.33
WPol 812	34	23.4	0.952	0.8281	82.81	Genuine	83.33
NCov298	− 26.83	22.45	− 1.604	0.0785	7.85	Fake	0
NCov097	− 43	19.38	− 3.583	0.1285	12.85	Fake	0
Health128	36.57	16.72	3.25	0.939	93.90	Genuine	100
Health765	33.33	27.02	1.763	0.9626	96.26	Genuine	100
Sports523	− 58	13.85	− 4.164	0.3775	37.75	Non-determinable	33.33
WPol 124	34.28	18.28	3.589	0.7175	71.75	Genuine	100

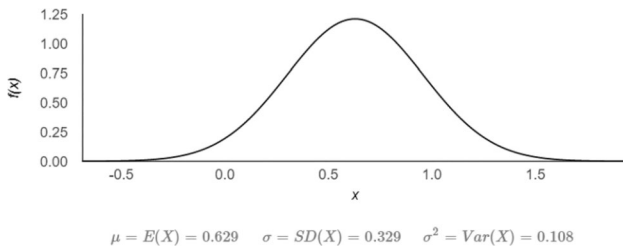$\mu = E(X) = 0.629 \quad \sigma = SD(X) = 0.329 \quad \sigma^2 = Var(X) = 0.108$

Fig. 2 Gaussian distribution with mean and standard deviation

weighted based on factors like organizations of affiliation, years of experience, designation, the frequency of their reviews and its accuracy.

The ProBlock algorithm is shown in Algorithm 1. To implement Algorithm 1, a blockchain approach is used. The flowchart given in Fig. 3 describes how the block for each newsPiece is created and validated by evaluating the

score given by the experts. The final prediction of the genuineness of the newsPiece is done using the ProBit model. The feedback mechanism is incorporated so that experts trust score is updated after each judgment of the newsPiece. In comparison to contemporary models of fake news detection, ProBlock offers a highly credible trust-based voting system in which multi-media of any format including images, videos, texts, sounds can be analyzed by verified reviewers, and scores can be generated to determine their veracity. The fakeness scale used for the determination of the fakeness of news in [17] makes use of the machine and deep learning techniques while ProBlock works on human-based expert systems to ensure realistic data analysis. It utilizes a similar decentralized blockchain authority system through consensus algorithms like [40], but the consensus algorithm used is a modified version of Proof of Trust as mentioned earlier, improving throughput and scalability by separating metadata from data items of

Table 5 Respresentation of initial newsPieces with hashes of the blockchain

NewsPiece	PrevHash	CurHash	timeStamp	uploaderId
Paytm requires location to log in now	Nil	c9d01ed4cb0cafa 49e50f3dc2ca5ec 326d1bdf036e8ad 919f389c0b04a 963a5c	10/04/ 2020 7:29:46	AfzalOzil360
Richa Chadha shared an old tweet screenshot of Rishi Bagree (BJP IT cell worker) accepting that he spreads propaganda for BJP daily, which he claimed was a fake tweet and threatened her with a legal defamation case. Later it was found that the fake tweet is indeed real and was tweeted by him	c9d01ed4cb0cafa 49e50f3dc2ca5ec3 26d1bdf036e8ad9 19f389c0b04a963 a5c	2ca9d5fcea6c369 35a8823905ad044 513bbf7c761cbf6 7ad2d8691c19cd8 5a60	12/04/ 2020 6:44:16	sarazond
Netflix's new documentary on the Coronavirus includes Arunachal Pradesh AND Taiwan in their map of China	2ca9d5fcea6c369 35a8823905ad044 513bbf7c761cbf6 7ad2d8691c19c d85a60	5cc9156cf0ef6bc 786221a7ace7924 b83f32820ed63 83415c2 c0edff13688321	11/04/ 2020 21:19:46	Nogoldsplease
Coca-Cola Co. announces to to start producing milk bottles	5cc9156cf0ef6bc 786221a7ace7924 b83f32820ed63 83415c2c0edff1 3688321	1abc7daa7a00158 22ddd605898cd e76e15995e1c73e 4e62ea724059217 2ea9e2	10/04/ 2020 2:13:23	hammyhammad
BJP lawmaker makes fun of PM Narendra Modi's 'clap, bang utensils' appeal	1abc7daa7a0015 822ddd605898c de76e15995e1c7 3e4e62ea724059 2172ea9e2	8e5e8d0082 2e8573b5921cc4 fdda3566653721 0b1b07d4e555bb a6387aa9c8f3	12/04/ 2020 14:39:43	sudden_dust
Spurious Chart on 'Jihad' Leads to 'Insulting Religion' FIR Against Zee News Editor	8e5e8d00822e8 573b5921cc4fdda 35666537210b1 b07d4e555bba6 387aa9c8f3	b9ed9fcab5b5a29 fa70711ac58b0 fda584dffc26dee 8f25dbde64bd3ce 8d0966	09/05/ 2020 22:31:43	ilali307
Testing increased to 95,000 tests a day	b9ed9fcab5b5a29 fa70711ac58b0f da584dffc26dee 8f25dbde64b d3ce8d0966	481b404b78d285 9b5fe625db2a60 119276d669b2c1 9e0892381efdba e786c524	09/05/ 2020 22:16:28	maybevaibhav

---

Algorithm 1: ProBlock

1 **Function** `blockchain()`:
2   Constructor to initialize data members of Blockchain class
3 Create ArrayList of objects of Blockchain class.
4 **Function** `reviewer()`:
5   Constructor to initialize data members of Reviewer class
6 Create ArrayList of objects of Reviewer class.
7 **for** *all objects of Reviewer ArrayList* **do**
8   **for** *all objects of BlockchainArrayList* **do**
9   reviewer[i].vote(blockchain[j]);

10 **Function** `calStExpScore`:
11   Calculate *expScore* using *yrsScore, orgScore, desScore*;
12 **Function** `expScore`:
13   Update *expScore* on the basis of *freqScore, accScore*;
14 **Function** `updateExpScore`(*frequency*):
15   Update *expScore* based on yrsScore(Frequency of review)
16 **Function** `weightedVotes`(*blockchain[], expScore*):
17   *weightedVotes[]* calculated by multiplying *cummVote* with *expScore*
18 **Function** `probit`(*weightedVotes[]*):
19   Calculate *mean*;
20   Calculate *stdDev*;
21   Calculate *beta*;
22   Calculate *phi*;
23   **return** phi;(probability of being genuine.)

---

the blockchain [28]. Each newsPiece is referenced through its newsCode, and the newsCode is passed to the blockchain. This goes on to reduces the computational time and improves throughput, and scalability. The complexity of the model is independent of the number of reviewers of the block. The passing of the meta-data through the blockchain reduces the complexity significantly.

4.2.1 Addition of nodes

This data is taken as the input of the blockchain class where newsPieces are entered with their news code on an array list and the hashes are generated using the SHA256 hashing algorithm. Algorithm 1 gives all details of how the function is called in the ProBlock Model.

4.2.2 Removal of nodes

A removal mechanism to remove nodes on the blockchain with a great probability of fakeness is also included in the model. A node for which $P(X) \leq 0.70$ is removed from the chain. The removal procedure can be carried out by a Pragmatic Approach to Erasure using the functionality-preserving local erasure (FPLE) as proposed by Martin Florian et al. [41].

5 Results and performance measure

The average size of a block generated by the network averages to 760 bytes depending on the length of the newsPiece and the corresponding hash generated. The blockchain network is successfully able to furnish the 500 test blocks in 4.87 ms. The SHA256 hashing algorithm operates at 100 megabytes per second, For every generated block, each reviewer takes to cast their respective votes depending on the complexity of the item under consideration and the level of confidence they hold in their vote. For sake of simplicity, the test of voting was conducted on
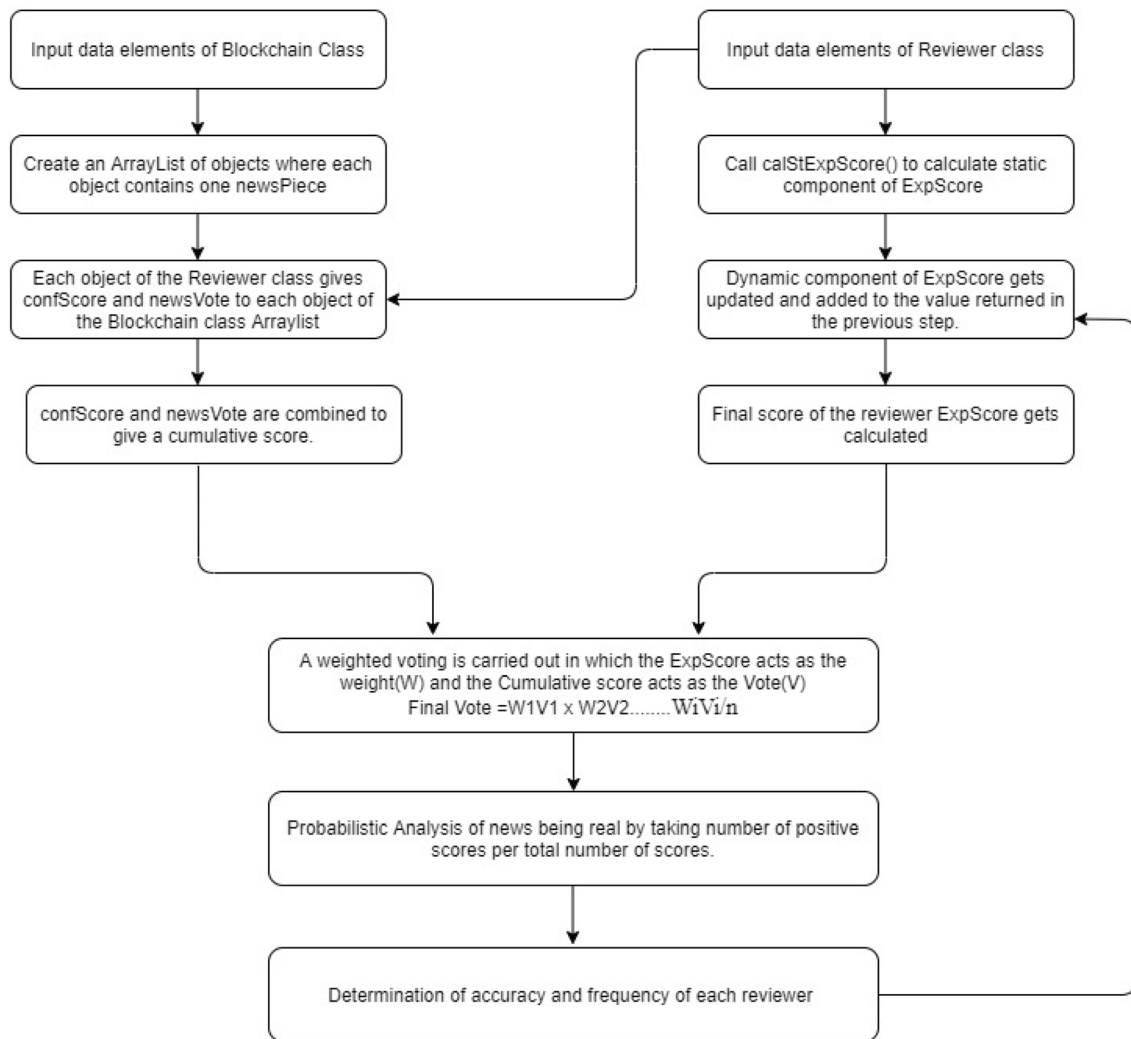
Fig. 3 Flow diagram of **ProBlock**

100 of the 500 test blocks generated. Taking the average of the time taken by reviewers to cast their vote over 40 newsPieces and the time taken to validate those votes to approve the newsPiece is roughly 108 s per block.

The analysis of the proposed system is done by increasing the number of blocks and by increasing the number of reviewers. As the voting is taking place in a distributed environment i.e all the reviews can concurrently vote, increasing the number of reviewers for a newsPiece only shows a variation in the time taken for validation and computation as demonstrated clearly in Fig. 4. On varying the number of reviewers voting on the same newsPiece, we obtain the following rate of increase in time taken to validate. The time is taken increases from 107.22 to 136.45 s.

In the second analysis, we introduce variation in throughput of the system with an increase in the number of blocks. The number of reviewers is kept at 5 and the number of newsPieces is increased, thereby increasing the length of the chain. The results are depicted in Fig. 5 where the increase in time taken for validation increases from 111.81 to 112.66 s. 2 reference points, 70% (0.70) and 30% (0.30) are taken for measuring the Accuracy of the model.
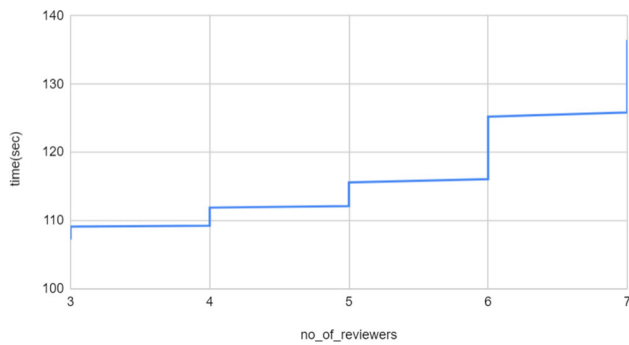
Fig. 4 Variation of time taken to give response with increase in number of reviewers
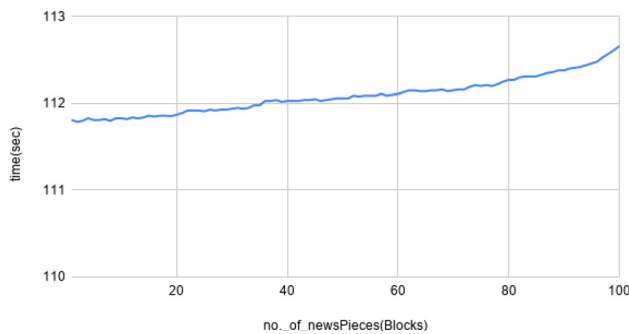

Fig. 5 Variation of time taken to give response with increase in number of newsPieces

All newsPieces with probability percentages greater than 70% are considered genuine. The usage of the ProBit model enables us to improve the accuracy of the model because, in the ProBit model, the inverse standard normal distribution of the probability is modelled as a linear combination of the predictors. ProBlock shows an accuracy of 82.79% for detecting genuine newsPieces which is a significant improvement to the Non-ProBit based methods

of probability calculation. These results are all depicted in Table 6. We can see clearly in Table 6 the strong results garnered by ProBlock

6 Conclusion and future work

This paper aims at providing a comprehensive model for fake news detection. The proposed model is advantageous over existing approaches in many dimensions. First, it can handle any type of news piece whether it is text, image, video, or audio format. The authenticity of the model is high as it considers expert knowledge for testing news pieces, and a dynamic weight voting approach is used that considers the credibility of reviewers. News pieces are further classified as fake or genuine using the ProBit model. The considered approach has outperformed the simple weighted approach. For a secured and faster implementation of the model, we consider the use of a distributed ledger. The whole model is implemented using blockchain technology which allows simultaneous voting and immediate feedback to reduce scalability issues. By analyzing both qualitative and quantitative results provided in the paper, we show that by taking into consideration factors regarding the credibility of reviewers through the ProBit model over a distributed platform like a blockchain, the probabilistic analysis of the veracity of news items can be done with higher accuracy (Table 6). In terms of future directions, the given model of fake news detection through a probabilistic analysis over a blockchain can be made more efficient by deploying a greater number of servers in the network. The model can be implemented on the back-end of a front-end web application to be used more widely by a greater number of people.

Table 6 Results of ProBit model vs. no ProBit model

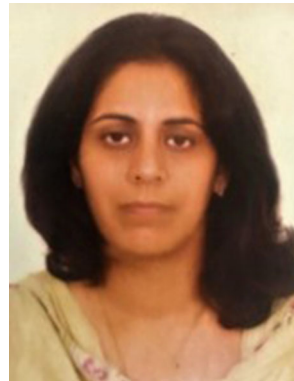| Condition	Count	Total number of correct newsPieces	%Accuracy of model
Number of newsPieces with %probability $\geq$ 70% WITH PROBIT MODEL	24	29	82.79
Nnumber of newsPieces with %probability $\geq$ 30% WITH PROBIT MODEL	7	11	63.36
Number of newsPieces with %probability $\geq$ 70% WITHOUT PROBIT MODEL	16	29	55.17
Number of newsPieces with %probability $\geq$ 30% WITHOUT PROBIT MODEL	6	11	54.55

References

1. Celliers, M., Hattingh, M.: A systematic review on fake news themes reported in literature. In: Proceedings of the Conference on e-Business, e-Services and e-Society, pp. 223–234. Springer (2020)

2. López-López, P.C., Onate, P., Rocha, A.: Social media mining, debate and feelings: digital public opinion's reaction in five presidential elections in Latin America. Clust. Comput. pp. 1–12 (2020)

3. Qi, E., Yang, X., Wang, Z.: Data mining and visualization of data-driven news in the era of big data. Clust. Comput. **22**(4), 10333–10346 (2019)

4. Bondielli, A., Marcelloni, F.: A survey on fake news and rumour detection techniques. Inf. Sci. **497**, 38–55 (2019)

5. Gurav, S., Sase, S., Shinde, S., Wabale, P., Hirve, S.: Survey on automated system for fake news detection using NLP & machine learning approach. Int. Res. J. Eng. Technol. (IRJET) **6**(01), 308–309 (2019)

6. Shrivastava, G., Kumar, P., Ojha, R.P., Srivastava, P.K., Mohan, S., Srivastava, G.: Defensive modeling of fake news through online social networks. IEEE Trans. Comput. Soc. Syst. **7**(5), 1159–1167 (2020)

7. Zhou, X., Zafarani, R., Shu, K., Liu, H.: Fake news: Fundamental theories, detection strategies and challenges. In: Proceedings of the twelfth ACM international conference on web search and data mining, pp. 836–837 (2019)

8. Lange, M., Leiter, S.C., Alt, R.: Defining and delimitating distributed ledger technology: results of a structured literature analysis. In: Proceedings of the International Conference on Business Process Management, pp. 43–54. Springer (2019)

9. Gaurav, A.B., Kumar, P., Kumar, V., Thakur, R.S.: Conceptual insights in blockchain technology: Security and applications. In: Proceedings of the Transforming Businesses With Bitcoin Mining and Blockchain Applications, pp. 221–233. IGI Global (2020)

10. Shao, X., Ma, X., Chen, F., Song, M., Pan, X., You, K.: A random parameters ordered probit analysis of injury severity in truck involved rear-end collisions. Int. J. Environ. Res. Public Health **17**(2), 395 (2020)

11. Kwak, K.H., Kong, J.T., Cho, S.I., Phuong, H.T., Gim, G.Y.: A study on the design of efficient private blockchain. In: Proceedings of the International Conference on Computational Science/ Intelligence & Applied Informatics, pp. 93–121. Springer (2018)

12. Srivastava, G., Parizi, R.M., Dehghantanha, A.: The future of blockchain technology in healthcare internet of things security, pp. 161–184. Blockchain Cybersecurity, Trust and Privacy (2020)

13. Srivastava, G., Dhar, S., Dwivedi, A.D., Crichigno, J.: Blockchain education. In: Proceedings of the 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 1–5. IEEE (2019)

14. Ch, R., Srivastava, G., Gadekallu, T.R., Maddikunta, P.K.R., Bhattacharya, S.: Security and privacy of UAV data using blockchain technology. J. Inf. Security Appl. **55**, 102670 (2020)

15. Chen, Q., Srivastava, G., Parizi, R.M., Aloqaily, M., Al Ridhawi, I.: An incentive-aware blockchain-based solution for internet of fake media things. Inf. Process. Manage. **57**(6), 102370 (2020)

16. Zhang, J., Dong, B., Philip, S.Y.: Fakedetector: Effective fake news detection with deep diffusive neural network. In: Proceedings of the 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 1826–1829. IEEE (2020)

17. Mavroforakis, M.E., Theodoridis, S.: Support vector machine (svm) classification through geometry. In: Proceedings of the 2005 13th European Signal Processing Conference, pp. 1–4. IEEE (2005)

18. Kwon, S., Cha, M., Jung, K., Chen, W., Wang, Y.: Prominent features of rumor propagation in online social media. In: Proceedings of the 2013 IEEE 13th international conference on data mining, pp. 1103–1108. IEEE (2013)

19. Ferreira, W., Vlachos, A.: Emergent: a novel data-set for stance classification. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pp. 1163–1168 (2016)

20. Liliana, D.Y., Basaruddin, C.: A review on conditional random fields as a sequential classifier in machine learning. In: Proceedings of the 2017 International Conference on Electrical Engineering and Computer Science (ICECOS), pp. 143–148. IEEE (2017)

21. Ciampaglia, G.L., Mashhadi, A., Yasseri, T.: Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13–15, 2017, Proceedings, Part I, vol. 10539. Springer (2017)

22. Pinzón, C., Rocha, C., Finke, J.: Algorithmic analysis of blockchain efficiency with communication delay. In: Proceedings of the FASE, pp. 400–419 (2020)

23. Yildirim, P., Birant, K.U., Radevski, V., Kut, A., Birant, D.: Comparative analysis of ensemble learning methods for signal classification. In: Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU), pp. 1–4. IEEE (2018)

24. Wang, F., Lan, M., Wu, Y.: Ecnu at semeval-2017 task 8: Rumour evaluation using effective features and supervised ensemble models. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pp. 491–496 (2017)

25. Franzese, M., Iuliano, A.: Descriptive statistics. ABC of Bioinformatics, Encyclopedia of Bioinformatics and Computational Biology (2018)

26. Zhou, Z., Guan, H., Bhat, M.M., Hsu, J.: Fake news detection via NLP is vulnerable to adversarial attacks. arXiv preprint arXiv: 1901.09657 (2019)

27. Zou, J., Ye, B., Qu, L., Wang, Y., Orgun, M.A., Li, L.: A proof-of-trust consensus protocol for enhancing accountability in crowdsourcing services. IEEE Trans. Serv. Comput. **12**(3), 429–445 (2018)

28. Gorenflo, C., Lee, S., Golab, L., Keshav, S.: Fastfabric: scaling hyperledger fabric to 20000 transactions per second. Int. J. Netw. Manage. **30**(5), e2099 (2020)

29. Zhou, X., Zafarani, R.: Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315 **2** (2018)

30. Yadav, V.K., Anand, A., Verma, S., Venkatesan, S.: Private computation of the Schulze voting method over the cloud. Clust. Comput. pp. 1–15 (2019)

31. Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E.: A review of relational machine learning for knowledge graphs. Proc. IEEE **104**(1), 11–33 (2015)

32. Bahri, L., Girdzijauskas, S.: When trust saves energy: a reference framework for proof of trust (pot) blockchains. In: Proceedings of the Companion Proceedings of the The Web Conference, 2018, pp. 1165–1169 (2018)

33. Egelhofer, J.L., Lecheler, S.: Fake news as a two-dimensional phenomenon: a framework and research agenda. Ann. Int. Commun. Assoc. **43**(2), 97–116 (2019)

34. Li, X., Jiang, P., Chen, T., Luo, X., Wen, Q.: A survey on the security of blockchain systems. Future Gen. Comput. Syst. **107**, 841–853 (2020)

35. Lin, I.C., Liao, T.C.: A survey of blockchain security issues and challenges. IJ Netw. Security **19**(5), 653–659 (2017)

36. Alfandi, O., Otoum, S., Jararweh, Y.: Blockchain solution for iot-based critical infrastructures: Byzantine fault tolerance. In: Proceedings of the NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium, pp. 1–4. IEEE (2020)

37. Berdik, D., Otoum, S., Schmidt, N., Porter, D., Jararweh, Y.: A survey on blockchain for information systems management and security. Inf. Process. Manage. **58**(1), 102397 (2021)

38. Al Ridhawi, I., Otoum, S., Aloqaily, M., Boukerche, A.: Generalizing ai: Challenges and opportunities for plug and play ai solutions. IEEE Network (2020)

39. Florian, M., Henningsen, S., Beaucamp, S., Scheuermann, B.: Erasing data from blockchain nodes. In: Proceedings of the 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 367–376. IEEE (2019)

40. Leal, F., Chis, A.E., González-Vélez, H.: Multi-service model for blockchain networks. Inf. Process. Manage. **58**(3), 102525 (2021)

41. Lechner, M., Okasa, G.: Random forest estimation of the ordered choice model. arXiv preprint arXiv:1907.02436 (2019)

Eishvak Sengupta is a final year student pursuing Bachelors of Technology in Computer Science and Engineering from Amity School of Engineering and Technology, Amity University, Noida. He is keenly interested in analysis and development of algorithms for statistical learning techniques.

Renuka Nagpal working as Associate Professor in Department of Computer Science, Amity School of Engineering and Technology, Amity University Uttar Pradesh. She completed her PhD in the field of Computer Science and Engineering from Dr. A.P.J. Abdul Kalam Technical University, Lucknow. She has more than 16 + years of experience in teaching and research. She had published more than 30 papers in national, international referred journals and conference proceedings.

Deepti Mehrotra is gold medallist of Lucknow University and completed her Ph. D from Lucknow University. Currently she is working as Professor in Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida. She has more than 23 year of research, teaching and content writing experience. She had published more than 180 papers in international refereed Journals and conference Proceedings. She member of various research committees and board of studies. 10 students have completed Ph.D. degree in her guidance. Her keen research interest area includes evolutionary algorithm, intelligent systems and Machine learning.

Gautam Srivastava was awarded his B.Sc. degree from Briar Cliff University in U.S.A. in the year 2004, followed by his M.Sc. and Ph.D. degrees from the University of Victoria in Victoria, British Columbia, Canada in the years 2006 and 2011, respectively. He then taught for 3 years at the University of Victoria in the Department of Computer Science, where he was regarded as one of the top undergraduate professors in the Computer Science Course Instruction at the University. From there in the year 2014, he joined a tenure-track position at Brandon University in Brandon, Manitoba, Canada, where he currently is active in various professional and scholarly activities. He was promoted to the rank Associate Professor in January 2018. Dr. G, as he is popularly known, is active in research in the field of Data Mining and Big Data. In his 8-year academic career, he has published a total of 143 papers in high-impact conferences in many countries and in high-status journals (SCI, SCIE) and has also delivered invited guest lectures on Big Data, Cloud Computing, Internet of Things, and Cryptography at many Taiwanese and Czech universities. He is an Editor of several international scientific research journals. He currently has active research projects with other academics in Taiwan, Singapore, Canada, Czech Republic, Poland and U.S.A. He is constantly looking for collaboration opportunities with foreign professors and students. Assoc. Prof. Gautam Srivastava received *Best Oral Presenter Award* in FSDM 2017 which was held at the National Dong Hwa University (NDHU) in Shoufeng (Hualien County) in Taiwan (Republic of China) on November 24-27, 2017.