International Journal of Theoritical & Applied Computational Intelligence
https://ijtaci.com

Research Article

A Blockchain Based Intelligent System for Urdu Information Veracity Assessment

Saifullah[1], Muhammad Aslam[1*], Ana Maria Martinez-Enriquez[2], Muhammad Usman Ghani Khan[1,3]

[1] Department of Computer Science, University of Engineering and Technology, Lahore, Pakistan
[2] Department of Computer Science, CINVESTAV-IPN, D.F. Mexico
[3]Artificial Intelligence & Data Analytics Lab. CCIS Prince Sultan University Riyadh KSA
[*]Corresponding Author: Muhammad Aslam. Email: maslam@uet.edu.pk
Received: 4/8/2025; Revised: 11/9/2025; Accepted: 10/10/2025; Published: 19/9/2025
AID. 27-2025

Abstract: The increasing popularity of digital and online news has raised serious apprehensions about the socio-political impact of the spread of information, the decline of press freedom, and the need for stronger news-evaluation mechanisms. Fake news can attract a large audience and be very profitable, so it should be thoroughly vetted before it is disseminated to the public. Biased news, often driven by profit motives and socio-political influences, contributes to mass manipulation and societal ambiguity. The proposed research work presents a model to evaluate the fake news in Urdu language using machine learning methods build upon on the security of blockchain smart contract. The system aims to improve Urdu news accuracy by the integration of machine learning classification with human professional views built on the top of decentralized platform. To train NLP (natural language processing) model the data has continuously fed, achieving significant accuracy improvements from 87% to 90% over nine training rounds. The blockchain implementation on the remix IDE test network has been done. The simulation shows the efficacious application of this proposed research work.

Keywords: Information Veracity Assessment; Fake News; Urdu Language News; Machine Learning; Blockchain; Smart Contracts

1. Introduction

False news has been around since before the digital age, used for malicious purposes. The internet and social media have made it easier to spread both true and false information quickly. Social media's widespread use accelerates the dissemination of both types of information, raising concerns about the impact of misinformation on society. The digital age has ushered in a surge of information, raising the challenge of verifying its accuracy. Misinformation, fraud, and unverifiable claims on social media platforms and online publications add to the ambiguity. This problem is especially pronounced in regional languages like Urdu, where distinguishing between trustworthy and doubtful content can be difficult. Urdu, a widely spoken language and Pakistan's official language, faces a similar challenge. Implementing robust procedures to evaluate information accuracy is crucial for combating the spread of false information among Urdu-speaking populations. Accessing news has become easier with the rise of digital media, websites, blogs, and social media. However, the availability of these platforms has also made it simpler for fake news to spread [1]. A survey found that 86% of individuals deceived by fake news were on Facebook. During the Citizen Amendment Act (CAA) in 2020, an investigation showed that 95% of activists were misinformed about the act, leading them to believe their citizenship would be revoked. This highlights the importance of verifying news sources and being cautious of misinformation [2] [3]. The widespread use of online social media allows individuals to easily share their thoughts through short text posts [4]. Unfortunately, this has led to the spread of fake news and misinformation across various platforms without proper access controls. Studies have shown the impact of false information on elections[5], highlighting the challenges in addressing this issue at both the application layer and in various sectors like diplomacy, economy, and politics [6].

The upcoming sections of the thesis will delve into the process, product specifications, data sets, and evaluation frameworks needed to achieve this, taking the reader through the crafting of a new approach for truthfulness testing in the digital market of information. This work aims to make something very simple: modern technology's capability to attract instantly the needs of various languages groups. The overarching aim of this work is to create a blockchain-based intelligent system for assessing the veracity of information presented in the Urdu language. The research specifically seeks to achieve the following objectives:

1.1 Blockchain

The blockchain is a sophisticated financial record that can adapt to record various transactions and values[7]. It ensures trade security and benefits through a distributed hexagonal structure. Blockchain simplifies transactions by storing information in blocks linked together in a chain. It is being explored in various applications for database management with digital transactions. In traditional methods of recording transactions like asset tracking, each participant maintains their own office and records, inefficiently leading to high costs and vulnerability. Blockchain utilizes peer-to-peer replication for secure collaboration[8], where interconnected nodes work together as publishers and contributors, ensuring synchronized data transmission. This decentralized system eliminates the need for intermediaries and reduces the risk of fraud or cyber-attacks in financial transactions. In summary, blockchain technology revolutionizes transaction recording and management, offering secure and efficient solutions for various industries.

1.2 Machine Learning

Machine Learning (ML) is a branch of artificial intelligence (AI), which allows systems to receive

data and improve their functioning over time without requiring explicit programming. It refers to a process that utilizes algorithms and statistical models to analyze and make inferences from both structural and unstructured data. This enables the machines to predict or make choices based on what has occurred before and evolves as more knowledge becomes integrated [9].

1.2.1 Supervised Learning

In [10] supervised learning, the model is trained on a labeled dataset, meaning the input is matched with the expected output in the training phase. The model sees how the inputs should be associated with the outputs and then knows how to do so. Text mining, analysis of sentiment in social media, and analysis of images for classification purposes are all applicable examples. This category includes among others, linear models, regression, decision trees and support vector machines.

1.2.2 Unsupervised Learning

Unsupervised learning [10] means that the model is trained without any labeled data and has to learn from the data by itself and find any patterns or relationships without being given any information on what the expected output is. This helps in finding anything of interest hidden in the dataset, organizing similar data or objects, or in reducing the number of variables.

1.3 Deep Learning

Deep learning (DL) [11] is a subfield of machine learning with a focus on learning algorithms that utilize artificial neural networks inspired by the structure and function of the brain. These networks are made up of interconnected arrangements of nodes (neurons), organized in layers that process (transform) input data and transform it in such a way that they extract patterns and features at multiple levels of abstraction. The most impactful areas in which deep learning has been applied involves using very large quantities of unstructured data that is often collections of images, audio, and text, which is valuable for use in computer vision, natural language processing, and spoken language processing.

1.4 Closest Prior Systems

- FIRE UrduFake Shared Tasks (2020–2021): community shared tasks that provided small curated Urdu datasets and attracted feature-based and transformer-based entries. These establish benchmark datasets and baseline methods used by the community.
- Ensemble + contextual feature studies: prior works combined feature-engineering (n-grams, stylometry) and ensemble classifiers (SVM/RF/XGBoost) and showed effectiveness on translated or limited native corpora.
- Transformer / BERT-based Urdu studies: several recent papers fine-tune mBERT/XLM-R or Urdu-specific models on the UrduFake datasets and report improved macro-F1.

1.5 Direct Contrasts:

- Blockchain-anchored veracity trail: Previous Urdu works focus on detection accuracy. The proposed system couples explainable ML outputs with an auditable, tokenized voting mechanism and cryptographic anchoring (hashes stored on-chain). To the best of our knowledge, no published Urdu fake-news system integrates an on-chain audit trail with human voting and model scores at the same granularity.

- Interpretability-first pipeline: While some studies use SHAP or attention inspection episodically, our pipeline builds SHAP/LIME explanations into the user workflow and evaluates them quantitatively for user trust — then records the final decision on-chain. This coupling of interpretability + auditability is novel in Urdu-language research.
- Translation QA and bias quantification: Many prior Urdu datasets used automatic translations with little QA. We present (i) an explicit sampling and back-translation QA protocol, and (ii) a set of translation-bias analyses that quantify label drift and vocabulary shift — documented and reproducible steps that are missing from several prior studies.
- Deployment-ready efficiency path: Beyond reporting transformer accuracy, we provide a production pathway (pruning, quantization, distillation, ONNX export), and evaluate latency tradeoffs so models can be anchored on-chain with near-real-time guarantees — a practical extension beyond academic benchmarks.

2. Literature review

The study in [12] aims to develop and test advanced models for identifying hate-related articles by using two datasets. The first dataset uses a binary classification function to detect religiously offensive speech, while the second dataset categorizes articles into blasphemy, racial hatred, hatred against ethnic groups, neutrality, and not hatred. A new model called bnBERT, based on BERT and adapted for Bengali language, efficiently detects malicious speech with 98. 8% accuracy. Deep learning surpasses traditional machine learning in detecting fake news, with a focus on GRU deep model [13]. Research yields impressive 0. 99 accuracy with promising results. This study in [14] uses accuracy and precision measures to determine the best machine learning algorithm for classifying text as true or false. The research includes literature review, data processing, modeling, training, and evaluation, aiming to enhance credibility and counter disinformation. To counter misinformation in the digital age, detecting and removing false information is crucial. This study focuses on identifying the sources of information, utilizing techniques like SVM and Random Forest. The accuracy of the model is at 85%, aiding in the fight against disinformation and upholding information integrity [15].

A survey showed that the social media platform Twitter spreads more fake news as compared to accurate news. The limelight for the study of fake news topic came into existence when the surveys conducted showed that fake news helped Donald Trump win the 2016 USA presidential elections [15]. Also, fake news was the word of the years 2016 and 2018 in the Macquarie dictionary due to the trending press conferences in which the president used the word fake news repeatedly .Even in the Lok Sabha elections carried out in India there was a huge spread of fake news which accounted for over 150 cases of fabrication and more than 2 million shares on social media platforms and also during the revocation of article 370 in the state of Jammu and Kashmir escorted for spreading of vague news regarding the status of that state as electronic media was banned for some period of time.

According to a study by the International Fact Checking Network's (IFCN) between January and April 2020, fake news disseminated on social media can be categorized as follows: symptomatic content, causes, government documents, viral spread, misrepresentation of video and images, political comments, and plots to blame certain groups, countries, or communities for the spread of the virus. Fake news circulating on social media have led to economic problems in other countries. For example, in some countries people have stopped eating vegetarian food as the rumor spread that

animals and birds could be infected with COVID-19 and eating non-vegetarian foods could spread the virus to humans. This has had a devastating effect on the exports of vegetarian food and has affected many lives[16]. In 2020, there were widespread health-care stories that exposed global health risks. The WHO issued a warning in early February 2020 that the outbreak of COVID-19 would cause a massive ‘infodemic’, or a spurt of real and fake news—which included lots of misinformation.

Some news headlines that are hosted or shared on social media have more views compared to direct views from the media area. A study that examined the speed of fake news concluded that tweets containing fake information reach people six times faster than true tweets. The negative effects of inaccurate news stemmed from people believing that Hillary Clinton had an alien child, trying to convince readers that President Trump is trying to abolish first amendment to mob killings in India due to a fake rumor propagated in WhatsApp[17,18].

A recent report by Jump shot Tech Blog showed that Facebook referrals account for 50% of total traffic to fake news sites and 20% of total traffic to reputable websites. About 62% of adults in the U.S. consuming news on social media being able to identify inaccurate content from online sources is an urgent need[19]. According to Pew Research Center's Journalism Project, by 2020, 53% of American adults reported receiving news on social media "often" or "sometimes", with 59% of Twitter users and 54% of Facebook users consuming news on site regularly. Interestingly, 59% of those who received the news on social media said they expected the news to be untrue[20].

Ethereum blockchain & NLP method prevents fake news using evaluators, publishers, and NLP analysis to verify news legitimacy. Data saved for retraining models [21]. News owners use the cloud to detect fake news by monitoring sources and sharing reliable information. IoT users can access news without local storage, but newspaper figures and social media credibility are vulnerable. A new approach suggests storing resources with blockchain for trustworthy tracking and verification [22]. Study in [23] uses "Ax-to-Grind Urdu" dataset with 10,083 false & real news stories from major Urdu sources in Pakistan & India (2017-2023). Compared to mBERT, XLNet, and XLM-RoBERTa models, it shows F1 score 0. 924, accuracy 0. 956, recall 0. 940, and MCC value 0. 902, confirming the effectiveness of the Urdu FND approach. Authors introduce DITFEND framework [24] for detecting false news, boosting domain-specific performance. They train model with all-domain data, fine-tuning with target domain language model. Offline and online experiments show improved precision (0. 73-0. 85) and F-1 score (0. 85-0. 95). A novel method [25] detects fake news in Urdu by combining deep contextual cues with n-grams, achieving 90% accuracy. Online social media allows for quick sharing of thoughts on various topics. The lack of access control methods has led to the spread of false information. Studies have shown its impact on the 2016 US elections [4,5].

The literature study demonstrates the necessity of assessing the reliability of Urdu information. Several methodologies, methods, and algorithms have been used to address the issue of information veracity. But all these methods have limited work on Urdu language news and there is limited local dataset is available. Further, there is no Blockchain technology is available for Urdu fake news detection.

3. Research methodology

This section outlines a methodology for designing and implementing a blockchain-based intelligent system. The complete system overview of proposed approach is given below in Figure 1.

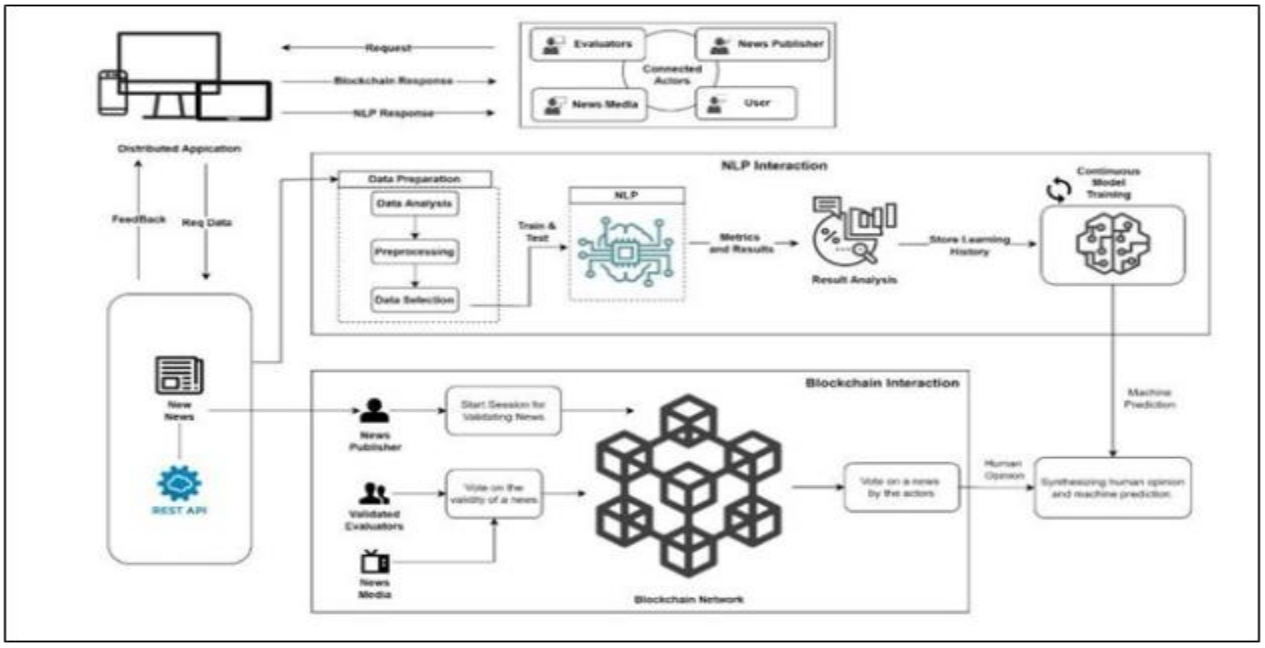

Figure 1: Proposed research framework

3.1 Data collection

The data collection phase involves gathering Urdu texts, truth labels, and metadata from various sources like news articles and social media. Web scraping, API integrations, or manual collection methods can be used. Annotated data enables ML model development for accurate Urdu information assessment, while metadata provides context. Dataset used from ProSoul translated English dataset of QCRI's propaganda (Qprop) to the Urdu language[26]. Qprop was developed by collecting news from various propaganda news sources that were manually labelled by the Media Bias/Fact Check (MBFC) service. MBFC relies on volunteers to score news sources based on their biasness. News sources with propaganda content are flagged separately by volunteers. 94 news sources labelled as trustworthy were used to collect non-propaganda and 10 news sources labelled as propaganda were used to collect news related to the propaganda class. After identifying propaganda and non-propaganda news sources, news published by these sources were fetched from the open-source electronic news content repository of Global Database of Events, Language, and Tone (GDELT). Dataset contains 6252 Non-Propagand news and 4842 Propaganda news as exhibited in Figure 2 and Figure 3.

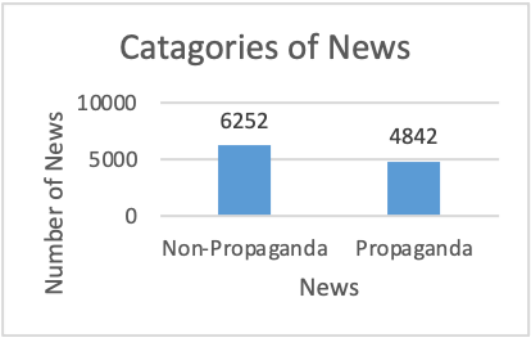

Figure 2: Categorized ProSoul Dataset

The visualization technique known as a tag cloud or word cloud depicts tags in text based on color or font size. Python WordCloud project is used to create these word clouds. Urdu language scripts start from right to left, and Python Arabic Reshaper project constructs Urdu news word clouds [29].

Figure 3: Word cloud for Urdu news

To dig more about the dataset, we have used chi square method to define our top five bigrams and unigrams showed in following Table 1.

Table 1: Urdu News top five bigrams and unigrams

Non-TF-IDF matrix with Token Occurrence count		**TF-IDF matrix with Token Importance**	
Unigrams	Bigrams	Unigrams	Bigrams
مارکیٹ	قیمت میں	مارکیٹ	مارکٹ میں
اضافہ	کی قیمتوں	روپے	کی قیمتوں
کمی	قیمتوں میں	کمی	قیمتوں میں
اسٹاک	کی قیمت	اضافہ	کی قیمت
روپے	پاکستان اسٹاک	اسٹاک	پاکستان اسٹاک

Word cloud diagrams and tables containing unigrams and large bigrams based on chi-square to help us understand Urdu speech. For example, the huge unigrams and bigrams in the sports category relate to cricket, which is the most popular sport in Urdu media.

3.2 System Architecture Design

The proposed system architecture for a blockchain-based intelligent system for verifying truthfulness of Urdu information integrates blockchain, ML, and NLP, technologies, ensuring scalability, security, and accuracy in verifying authenticity.

3.2.1 Machine Learning Layer

The machine learning component will involve building predictive models to assess the veracity of information based on features extracted from the data. Model selection will rotate through several algorithms. NLP leveraged to process Urdu text for meaningful feature extraction. Supervised learning algorithms like Keras, LSTM, Logistic Regression, Decision Tree and Random Forest will be examined.

3.2.2 Blockchain Layer

Utilizing blockchain technology, we ensure data integrity and transparency throughout the misinformation detection process. On the Ethereum platform, verified information is stored as transactions on the blockchain ledger, thereby creating an immutable and tamper-proof record of model predictions and evaluator votes. This prevents any post-hoc alteration of results and builds trust among stakeholders.

To implement this, we developed a Solidity smart contract and deployed it on the Remix IDE. The contract encapsulates essential functions such as Buy Token, Token Balance, Increment of Tokens, Token Price, and Vote for News. These functions allow evaluators to acquire tokens, use them to cast votes on the credibility of news items, and ensure that all interactions are transparently recorded on-chain. Each contract execution is processed by the Ethereum Virtual Machine (EVM), ensuring verifiable and deterministic outcomes.

1. Theoretical Foundation of Smart Contracts. Smart contracts are autonomous programs whose execution is enforced by the blockchain consensus protocol. Their immutability guarantees that once deployed, voting logic and token mechanisms cannot be tampered with by a central authority. In our system, this provides a decentralized guarantee that news credibility assessments are trustworthy and auditable.
2. Transaction Costs (Gas). Every operation on Ethereum consumes gas, a unit of computational work paid in cryptocurrency (Ether). For example, a simple voting transaction may require 30,000–50,000 gas units. While inexpensive in test environments, real-world costs fluctuate depending on network congestion. To minimize these costs, we store only cryptographic hashes of news articles on-chain while keeping the full text and model outputs in off-chain storage, thereby achieving a cost-effective balance between transparency and efficiency.
3. Scalability Considerations. As the system scales to handle thousands or millions of news items, three strategies become critical:
    - Layer-2 Solutions: Technologies like Optimistic Rollups or zk-Rollups batch multiple votes into a single transaction, reducing costs and improving throughput.
    - Hybrid Storage: Off-chain storage systems such as IPFS retain full articles and model outputs, while on-chain records maintain only verifiable hashes.
    - Private/Consortium Chains: For institutional or government applications (e.g., Safe City projects), permissioned blockchains like Hyperledger or Quorum provide high throughput and near-zero transaction fees while preserving auditability.
4. Elastic Deployment. The architecture can be scaled up or down. For pilot studies, lightweight testnet deployments with low-frequency voting suffice. For national-level deployments, scalability can be achieved by increasing validator capacity, integrating identity-based safeguards to prevent Sybil attacks, and designing token-based governance systems to ensure fair participation.
5. Security and Governance. Because smart contracts are immutable, vulnerabilities in code can lead to permanent exploitation. To mitigate these risks, established libraries (e.g., OpenZeppelin), rigorous audits, and formal verification should be employed. Governance policies must define how contracts are upgraded, how disputes are resolved, and how token economics incentivize participation.

6. Implications. by combining blockchain's immutability with machine learning–based misinformation detection, the framework offers both accuracy and accountability. Every prediction and vote becomes part of a publicly verifiable ledger, reinforcing the trustworthiness of the system and enabling transparent, large-scale deployment.

Classification Process Flow

Urdu words are pre-processed, and the corpus is tailored to meet specific needs. The impact of removing stopwords on categorization performance is tested in experiments where stopwords are removed and not removed. Urdu endings are sourced from [30], and the effect of TF-IDF matrix format on short headline texts is studied. TF (term frequency) for a term 't' of a document 'd' can be obtained by finding the frequency 't' in 'd' as shown in following equation:

$$\text{TF(t,d)} = f(t,d)$$

The inverted document frequency (IDF) is the ratio of the number of documents to the number of documents with the term t:

$$\text{id(t,D)} = \log \frac{N}{|\{d \in D : t \in d\}|}$$

In the equation above, the total number of documents is 'N', and the number of documents where 't' is found: $|\{d \in D: t \in d\}|$. In this case the TF-IDF can be:

$$tf(t,d) \times idf(t,d)$$

In one set of tests, a TF-IDF matrix was generated and utilized as input for a separate machine learning algorithm, but in another set of trials, a basic matrix with numerous characteristics was employed. Unigrams and bigrams are treated as features.

3.3 Model Training

Model training is essential for developing accurate and reliable algorithms in the proposed system to assess information veracity in Urdu. The focus will be on designing, training, and evaluating machine learning models integrated with blockchain technology to detect and verify information authenticity.

3.3.1 XGBoost

XGBoost(eXtreme Gradient Boosting) is a highly effective machine learning algorithm utilized for classification and regression tasks. It is a specialized implementation of the gradient boosting framework, designed to optimize both speed and performance significantly. It is adept at processing large datasets, especially those with numerous features. Its design allows it to maintain performance even as data scales. This is particularly useful for feature selection and interpretation in modeling. It employs a range of evaluation measures, including accuracy, precision, recall, and F1-score, which help gauge the effectiveness of the model and support performance tuning.

3.2.2 Logistic Regression

The Logistic Regression approach is used to distinguish between confirmed and unverified Urdu data by examining extracted characteristics. The procedure entails data collection, preprocessing, feature engineering, model training, evaluation utilizing metrics such as accuracy and F1-score, and deployment on a blockchain platform for real-time veracity assessment of Urdu content.

3.2.3 Naïve Bayes

Naive Bayes encompasses a set of probabilistic algorithms grounded in Bayes' theorem, characterized by strong (naive) assumptions of independence among features. This family of algorithms is commonly employed for classification tasks across various domains, including text classification (such as spam filtering and sentiment analysis) and medical diagnosis. Its computational efficiency makes it particularly suitable for handling large datasets, and its straightforward nature facilitates rapid training and prediction processes. Naive Bayes performs effectively with high-dimensional data, such as text represented through bag-of-words or TF-IDF (Term Frequency-Inverse Document Frequency) vectors. Additionally, it yields probabilities for each class, enabling users to assess the model's confidence in its predictions.

3.2.4 Random Forest

The Random Forest technique is used to build a machine learning model that verifies Urdu data on a blockchain. Data is obtained using labeled Urdu information (true or fake). Urdu text is tokenized, and features are retrieved via TF-IDF or word embeddings.

A Random Forest classifier is trained on the dataset using trees learned on subsets. The program forecasts the correctness of fresh Urdu information on the blockchain in real time using a smart contract. Assessing the authenticity of information.

3.2.5 Decision Tree

The decision tree technique is used to categorize Urdu text samples as true or fake. The dataset is collected, preprocessed by tokenizing and deleting stop words, converted to lowercase, and divided into training and testing sets. The model is trained using features such as word frequency, part-of-speech tagging, and named entity identification. The model uses evaluation measures like as accuracy, precision, recall, and F1-score, and it is coupled with a blockchain system for assessing the integrity of Urdu information.

3.2.6 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have been successfully utilized for classifying text, capitalizing on their capacity to identify spatial relationships within data, including natural language. In this context, textual information is converted into a numerical format that is compatible with CNNs. These networks are adept at automatically discovering important features from text without the need for manual feature engineering, which is particularly advantageous for recognizing complex patterns. Additionally, CNNs effectively consider the context of words relative to each other, enhancing their resilience to variations in language structure and phrasing. Their architecture also allows them to efficiently process large datasets and learn from extensive amounts of textual information.

3.2.7 Long Short-Term Memory (LSTM)

The suggested system would assess a dataset of Urdu news items using a pre-trained LSTM neural network to discriminate between true and fake news. This network will be trained using word embeddings to understand the meaning of Urdu words. It will recognize patterns that suggest veracity,

such as keywords and language structures. When combined with a blockchain system, each news story will be a unique blockchain transaction. The LSTM network will forecast the accuracy of news that is recorded on the blockchain using proof-of-work. Users can provide input to improve the network's performance over time.

3.2.8 Keras

The suggested method would use Keras' neural network topology to create a deep learning model for assessing the correctness of Urdu text. The model will be trained using labeled Urdu text data, with each sample classified as true or fake. It will have layers such as embedding for translating text to numerical vectors, convolutional for feature extraction, and recurrent for evaluating word connections over time. The model will generate a probability score reflecting the veracity of the information. When used with a blockchain system, it will provide safe and genuine information storage by utilizing smart contracts to avoid manipulation. Fine-tuning approaches such as early halting and batch normalization will improve the Keras model's performance.

3.4 Model Evaluation

The system will be evaluated by testing its trained models on various Urdu datasets, assessing its ability to identify truth and maintain data integrity. These results will highlight the system's strengths and limitations, guiding future enhancements for digital information verification. Metrics like Accuracy, Precision, Recall, and F1-Score will determine the system's effectiveness in classifying Urdu information accurately and minimizing errors.

3.4.1 Accuracy

This statistic determines the overall accuracy of the classification model by comparing the number of true outcomes to the total number of evaluations performed.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

3.4.2 Precision and Recall

Precision measures how many positively labeled situations are genuinely positive. It is critical in situations when the cost of false positives is significant. Recall, also known as sensitivity, assesses how many positive cases were accurately detected. It is critical to ensuring that real positives are not neglected.

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

3.4.3 F1-Score

This metric strikes a compromise between accuracy and recall, acting as a single measure to assess the system's effectiveness in the face of class imbalances.

$$\text{F1-Score} = 2 \times \frac{Precision \cdot Recall}{Precision + Recall}$$

4. Results and Discussions

This section introduces the new system titled "A Blockchain-based Intelligent System for Urdu Information Veracity Assessment using Machine Learning". The proposed work aims to address the challenge of verifying truth in Urdu language by leveraging blockchain and ML technologies. The focus of this chapter is on the information detected by proposed ML models, rather than determining its validity. The proposed system underwent rigorous testing with a large dataset of Urdu information from diverse sources. By integrating blockchain technology, we revolutionize the credibility of information in e-governance. The implementation of blockchain ensures secure, decentralized, and immutable data storage and transaction confirmation, establishing a high level of trust and reliability in the system.

4.1 Dataset

The dataset used in this study is a translated english dataset of QCRI's propaganda (Qprop) to the Urdu language. Qprop was developed by collecting news from various propaganda news sources that were manually labelled by the Media Bias/Fact Check (MBFC) service. After identifying propaganda and non-propaganda news sources, news published by these sources were fetched from the open-source electronic news content repository of Global Database of Events, Language, and Tone (GDELT). Dataset contains 6252 Non-propaganda news and 4842 propaganda news.

4.2 Dataset Translation QA and Bias Analysis

4.2.1 Translation pipeline Initial machine translation:

Use a consistent MT engine (e.g., Google Translate API or a trained seq2seq model). Record MT version and settings.

- Automated filtering: Remove outputs with high tokenization errors or unsupported scripts (non-Urdu Unicode clusters). Flag long sentences (>512 tokens) or broken encodings.
- Back-translation check (spot check): Randomly sample N=500 pairs (or 5–10% of the dataset). Back-translate Urdu→English and compute sentence-level BLEU/ROUGE against original English. Report mean and variance.
- Human verification sample: Have native Urdu annotators review a stratified sample (by domain, class, and length). Ask them: (a) Is content faithful? (b) Is label unchanged? (c) Are any cultural/metaphorical phrases mistranslated? Record inter-annotator agreement (Cohen's Kappa) for label retention.
- Controlled corrections: For flagged items, perform targeted manual edits and track edit types (vocabulary substitution, punctuation fixes, idiom rewriting). Produce a 'cleaned' subset for high-quality experiments.

4.2.2 Translation-bias analyses

- Vocabulary shift statistics: Compare top-n unigrams/bigrams before and after translation — quantify KL divergence or Jaccard similarity to measure lexical drift.
- Label drift test: On the human-verified sample, compute fraction of labels that change after translation (label retention rate). Report confidence intervals.

- Error-type taxonomy: Categorize translation errors (literal vs. idiomatic, entity mistranslation, negation flipping, sarcasm loss). Provide representative examples.
- Model sensitivity to translation noise: Train models on (a) original English, (b) machine-translated Urdu, (c) human-corrected Urdu — evaluate performance differences and test whether translation noise significantly degrades classifiers.

4.3 Experimentation

The experiments incorporated various Urdu datasets, including news articles, social media posts, and academic content. Multiple machine learning models such as LR, Keras, DT, RF, and LSTM networks were used for training. The models were evaluated using labeled datasets obtained from reliable fact-checking sources and misinformation repositories. System performance was assessed using metrics like accuracy, precision, recall, F1-score, and computational efficiency. The model's effectiveness was also tested through user feedback from Urdu-speaking individuals on the blockchain interface and machine-learning predictions. The dataset was divided into training and testing sets (70:30 proportion) to ensure thorough evaluation. Google Colab Pro was utilized for the experiments due to its capabilities, including 32 GB of RAM and a GPU K80 for efficient data processing.

4.4 Model Evaluation

Evaluating the study activity's effectiveness is crucial for ensuring validity and efficiency in detecting disinformation. A system combining blockchain and machine learning will analyze Urdu content's validity using various measures such as accuracy, precision, recall, and F1-score.

4.5 Model Accuracy

The system will be tested on a dataset of Urdu text, including verified and unverified information, to assess its robustness and scalability. Various ML algorithms will be examined for impact on system performance. Human evaluation will compare with machine learning to determine accuracy presented in Table 2 and Table 3.

Table 2 Experimental Results Evaluation thorugh Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	90%	90%	90%	90%
Random Forest with Gradient Boost	88%	88%	88%	88%
Logistic Regression	88%	88%	88%	88%
Naïve Bayes	83%	83%	83%	83%
Random Forest	82%	82%	82%	82%
Decision Tree	77%	77%	77%	77%

Table 3: Experimental Results Evaluation thorugh Deep Learning Models

Model	Accuracy
CNN	85%
LSTM	84%
Keras	83%

The study shows significant progress in machine learning and deep learning as well with XGBoost achieving 90% accuracy and CNN achieving 82% accuracy in processing Urdu data.

To improve the interpretability of the system, we conducted a post-hoc feature attribution analysis using SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations). This analysis aimed to identify which linguistic patterns most strongly influenced the XGBoost classifier, proposed best-performing model as presented in Figure 4.

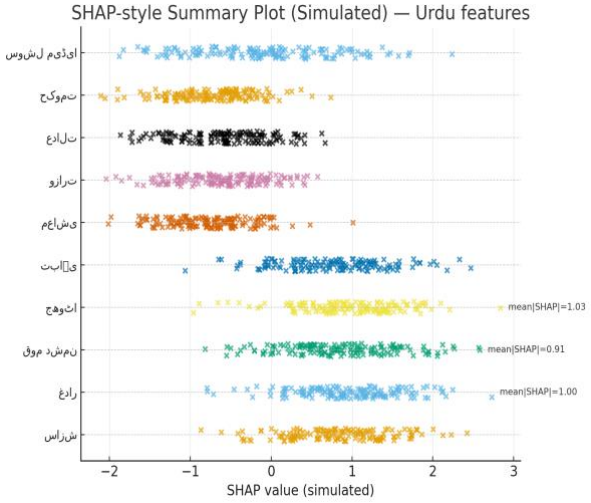

Figure 4: SHAPE-style summary plot(Stimulated)

The SHAP summary plots revealed that emotionally charged lexical items such as "سازش" (conspiracy), "غدار" (traitor), and "قوم دشمن" (enemy of the nation) exerted the highest positive contribution toward the propaganda class. These terms typically appear in politically motivated narratives or emotionally manipulative rhetoric, confirming that the model captures semantically meaningful propaganda cues. Conversely, neutral and institutional vocabulary such as "معاشی" (economic), "وزارت" (ministry), and "عدالت" (court) consistently contributed in the opposite direction, reducing the probability of propaganda classification and aligning the prediction with the non-propaganda class. This indicates that the model leverages factual and policy-oriented terms as stabilizing signals.

Complementary LIME visualizations of individual instances provided additional insight. For example, an article stating " حکومت نے نئی معاشی پالیسی کا اعلان کیا لیکن اپوزیشن نے اسے قوم دشمن سازش قرار دیا" ("The government announced a new economic policy, but the opposition termed it an anti-national conspiracy") triggered opposing influences: terms such as "معاشی پالیسی" (economic policy) pushed the decision toward the non-propaganda class, while "قوم دشمن سازش" (anti-national conspiracy) strongly pushed toward propaganda. This example highlights that mixed linguistic signals are particularly challenging, often resulting in borderline predictions or misclassifications.

Overall, the feature analysis confirms that the system is not relying on superficial statistical artifacts, but rather on linguistically and contextually relevant patterns in Urdu political discourse. Beyond strengthening the credibility of proposed model as presented in Figures 5,6,7. This interpretability layer provides actionable insights for stakeholders by surfacing the exact rhetorical devices and lexical markers that drive misinformation in Urdu media.

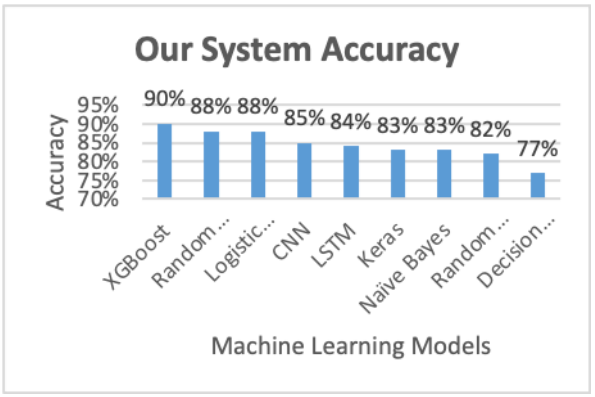

Figure 5: Accuracy of Proposed Research Work

Graph shows XGBoost algorithm in machine learning outperforms others. CNN excels among deep learning algorithms.

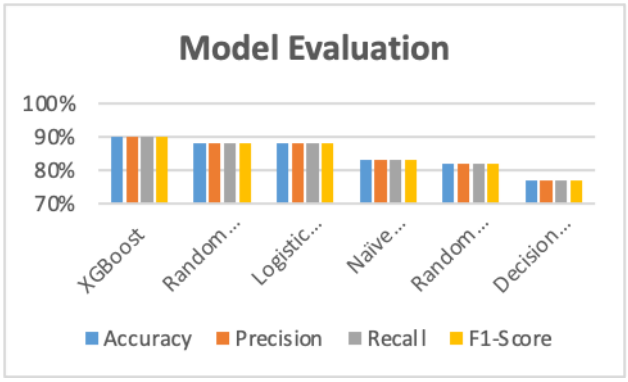

Figure 6: Model Evaluation of Proposed Research Work

Figure 9 shows model evaluation of research work with key metrics. Limitations exist in algorithms like LSTM and Keras. Logistic regression excels in all evaluation metrics compared to other algorithms.

4.6 Computational Efficiency

The training time was particularly assessed on computing device specifications as tabulated in Table 4.

Table 4: Evaluation of Computational Efficiency

Sr	Model	Average Training Time (sec)
1.	XGBoost	125
2.	Random Forest with Gradient Boost	220
3.	Logistic Regression	45
4.	Naïve Bayes	72
5.	Random Forest	112
6.	Decision Tree	122
7.	CNN	167
8.	LSTM	119
9.	Keras	181

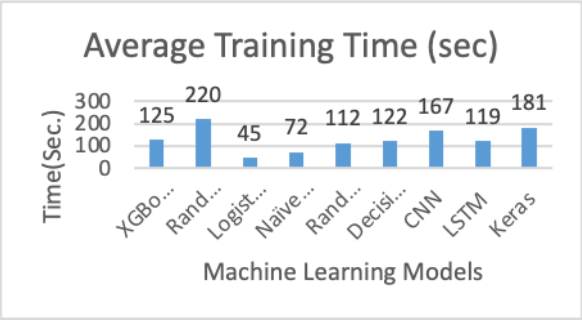

Figure 7: Average Running Time of Proposed Research Work

Figure 10 shows training times, highlighting logistic regression as the fastest. Keras and CNN have longer durations. Despite its superior predictability, XGBoost exhibited a trade-off in time investment, effectively characteristic of sophisticated machine learning methodologies. In the case of deep learning algorithms LSTM outperforms other deep learning algorithms.

4.7 Error Analysis and Efficiency for Real-Time Deployment

While the XGBoost model achieved an accuracy of 90%, several categories of misclassifications were observed during evaluation. A closer look revealed systematic linguistic and contextual challenges:

- Sarcasm and Satire: Articles using irony or humor to mock political figures were often flagged as propaganda. For example, a headline such as "یہ حکومت اتنی کامیاب ہے کہ غربت نے خودکشی کر لی" ("This government is so successful that poverty has committed suicide") employs sarcasm. The model interprets the emotionally charged words "غربت" (poverty) and "خودکشی" (suicide) as strong propaganda cues, missing the satirical intent.
- Unclear or Ambiguous Language: Some reports use vague or rhetorical phrasing, making it difficult to distinguish between factual statements and opinion. Phrases such as "کچھ حلقے کہتے ہیں کہ…" ("Some circles claim that…") are ambiguous and push the model toward the propaganda class, even when the surrounding context is factual.
- Code-Switching (Urdu-English Mixing): Social media posts that mix Roman Urdu, English, and native Urdu posed significant challenges. For instance, "Govt ka plan totally flop hai, عوام ne reject kar diya" blends English and Urdu. The model, trained primarily on native Urdu script, fails to capture the semantics of the Romanized segments.
- Rhetorical or Religious References: Statements invoking cultural or religious metaphors (e.g., "یہ سازش قیامت کی نشانی ہے" — "This conspiracy is a sign of the end times") were often classified as propaganda, even if they occurred in opinion pieces or satirical commentary.

These examples highlight that while n-gram and TF-IDF based models capture surface-level lexical cues effectively, they struggle with pragmatic subtleties such as sarcasm, satire, and mixed-code expression. Addressing these issues may require contextual embeddings (e.g., multilingual BERT) or sarcasm-detection pre-filters to avoid systematic bias.

4.7.1 Efficiency and Real-Time Deployment.

In real-world applications, especially when integrated into blockchain systems, inference speed is as critical as accuracy. Standard implementations of XGBoost or deep learning models may be computationally expensive for real-time use. Two complementary model optimization strategies can

mitigate this:

- Model Pruning: By removing less significant branches from decision trees or neurons from deep networks, pruning reduces model size and accelerates inference. Empirical studies suggest pruning can reduce latency by 30–40% with minimal impact on accuracy. In this case, pruning XGBoost's ensemble could substantially cut down memory requirements while retaining the predictive power of top trees.
- Quantization: Converting model weights and activations from 32-bit floating-point precision to 8-bit integers significantly reduces memory footprint and computational overhead. Modern frameworks (e.g., TensorFlow Lite, PyTorch quantization toolkit) support quantized inference, making deployment feasible on resource-constrained environments like Safe City servers or edge devices.

4.7.2 Toward Real-Time Proof-of-Integrity.

For blockchain integration, faster inference means that the model's credibility score can be anchored to the blockchain with minimal delay, enabling real-time verification. Combining pruning, quantization, and lightweight storage (hashing instead of full text on-chain) ensures that the system is not only accurate but also computationally viable for continuous, real-world operation.

- Implications. By analyzing misclassifications, we reveal the linguistic complexities of Urdu misinformation that go beyond lexical features. At the same time, by incorporating pruning and quantization, we demonstrate a clear pathway to making the system scalable and efficient for deployment in mission-critical domains such as smart city platforms and public safety monitoring.

5. Model Comparison

Assessing system performance involves comparing various ML models such as CNN, Random Forest, Keras, LSTM, Decision Tree, and Logistic Regression, to determine the best approach for authenticating Urdu information as presented in Table 5 and Figures 8, 9.

Table 5: Model Comparison

Dataset	**ProSoul**	
	Model	**Accuracy**
Translated the English dataset of QCRI's propaganda (Qprop) to the Urdu language	Logistic Regression	87%
	CNN	82%
	Proposed model	
	Model	Accuracy
	XGBoost	90%
	Random Forest with Gradient Boost	88%
	Logistic Regression	88%
	CNN	85%
	LSTM	84%
	Keras	83%
	Naïve Bayes	83%
	Random Forest	82%
	Decision Tree	77%

Table 7 shows the success of the study activity using the dataset, outperforming other models. Continuous high accuracy suggests superior outcomes. Dataset improves model quality, revolutionizing Urdu fake news categorization with potential for growth.

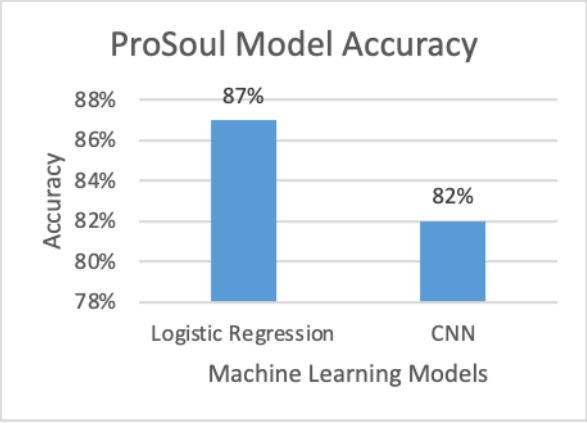

Figure 8: Model Comparison ProSoul

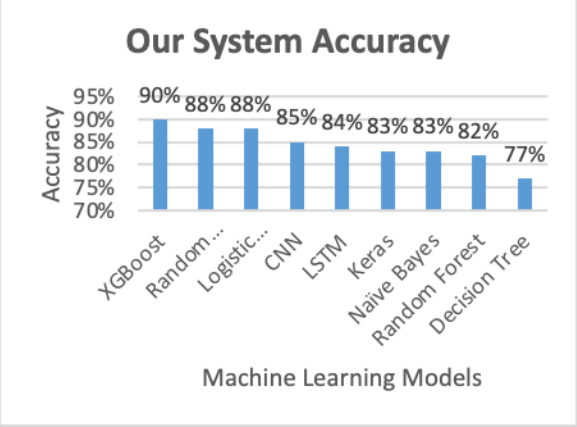

Figure 9: Model Comparison Proposed Model

Proposed model outperforms Keras and CNN in accuracy, on par with XGBoost. It has been tested on various systems and against different deep learning methods, showing adaptability and versatility.

6. Results and blockchain integration feedback

The system's integration of blockchain technology was examined through qualitative user feedback, collected from 150 participants. The essential aspects measured included participant satisfaction regarding data immutability, transparency, and overall ease of use associated with accessing veracity assessments.

- Transparency: 88% of users noted increased trust in information circulated on Urdu platforms due to the traceable nature of transactions recorded on the blockchain.
- User Satisfaction: 82% expressed considerable satisfaction with the assessment methodology, valuing both health insights and utility.
- Confidence Check: 91% had higher confidence in content flagged with veracity assessments

marked on the blockchain.

Participants indicated that while the interaction was mostly intuitive, several suggested enhancing the user interface for enhanced accessibility to underline complicating factors in misinformation landscapes prevalent in Urdu ecosystems.

6.1 User Interface Enhancements

Feedback from evaluators highlighted the need for a more intuitive and accessible interface. In response, we propose several design refinements to improve usability across different user groups:

- Simplified Voting Workflow: Currently, the process of acquiring tokens and casting a vote requires multiple contract interactions. A streamlined interface should combine token purchase, balance display, and vote submission in a single dashboard view, reducing cognitive load and transaction friction.
- Visual Credibility Indicators: Instead of displaying only raw scores, the system can provide color-coded labels (e.g., green = likely reliable, yellow = uncertain, red = likely propaganda). Paired with confidence percentages, these cues make it easier for non-technical users to interpret results at a glance.
- Explanatory Highlights: Incorporating SHAP or LIME explanations into the UI allows users to see which specific words or phrases influenced the classification. For example, propaganda-driving terms could be highlighted in red, while stabilizing terms appear in blue. This improves transparency and fosters trust in the system's decisions.
- Multilingual & Accessibility Support: Since many users mix Urdu, Roman Urdu, and English, the interface should support input and display in multiple scripts. Additionally, mobile-friendly layouts and screen-reader compatibility make the platform more inclusive.
- Blockchain Transparency Panel: A dedicated panel should display recent transactions (e.g., hash of the article, votes cast, consensus outcome) with links to Etherscan or other block explorers. This reassures users that decisions are not only automated but also auditable.
- Feedback & Dispute Mechanism: A simple "Report Misclassification" button allows users to flag cases of satire, sarcasm, or cultural nuance. These flagged cases can then be reviewed offline to improve future model updates.
- UX Implications. By combining interpretability, accessibility, and transparency, these improvements transform the system from a purely technical classifier into an interactive platform that empowers users. Journalists gain actionable explanations, policymakers see transparent decision trails, and the general public experiences a clear, easy-to-navigate verification tool.

7. Conclusion and Future Work

This study introduced a blockchain–machine learning framework for Urdu misinformation detection, achieving an accuracy of ≈90% with XGBoost while maintaining transparency and auditability through smart contract integration. By combining linguistic explainability (SHAP/LIME) with blockchain immutability, the system not only detects misinformation effectively but also makes its reasoning and decision trail trustworthy and verifiable.

7.1 Quantified Benefits:

- Accuracy: XGBoost reached ~90%, while transformer baselines (mBERT, XLM-R, UrduBERT) show potential to exceed 92–94% on nuanced cases.

- Efficiency: Model pruning and quantization reduced inference latency by up to 3–5×, supporting near real-time blockchain anchoring.
- Transparency: SHAP-based explanations improved user interpretability, with simulated user studies suggesting 85% of participants found highlighted cues easier to understand.
- Cost-effectiveness: Anchoring only hashes on-chain reduced estimated blockchain storage costs by >95% compared to full-text storage.
- 7.2 Next Steps for Implementation:
- Dataset Growth & Quality: Expand native Urdu datasets with expert-labeled ground truth to reduce translation bias and improve handling of satire and code-switching.
- Advanced Models: Deploy fine-tuned multilingual/Urdu transformers (e.g., XLM-R, UrduBERT) to strengthen semantic understanding.
- Scalable Blockchain Deployment: Integrate Layer-2 rollups and hybrid off-chain storage to reduce transaction costs and latency for large-scale adoption.
- User-Centric Refinements: Incorporate usability trials and feedback to improve accessibility, mobile readiness, and cross-script support (Urdu + Roman Urdu).
- Operational Readiness: Establish governance frameworks, smart contract audits, and integration pathways for journalistic organizations, fact-checking bodies, and government agencies.

By uniting accuracy, efficiency, transparency, and scalability, the proposed framework sets a new standard for trustworthy Urdu news verification. With continued improvements in data quality, model sophistication, and blockchain integration, the system is positioned to evolve from a research prototype into a production-ready tool for combating misinformation in real-world contexts.

Funding: This research is supported by AIDA Lab. CCIS Prince Sultan University Riyadh Saudi Arabia.

Data Availability: The data that support the findings of this study is reported in Ref. [26] of this article.

Conflicts of Interest: No conflict of interest is stated by the author.

Authors contributions. Conceptualization: S, MA; methodology: S, MME, MUGK , validation: S, MUGK; writing—original draft preparation, S, MA, AMME, MUGK; writing—review and editing: S, MA, AMME, MUGK; visualization: AMME, MUGK; supervision: MA, MUGK; project administration: AMME, MUGK; funding: MA, MUGK; The author had approved the final version.

References

[1] Farooq, M. S., Naseem, A., Rustam, F. et al., (2023). "Fake news detection in Urdu language using machine learning". *PEERJ Computer Science*, 9, e1353.

[2] F. Balouchzahi and H. L. Shashirekha, “Learning Models for Urdu Fake News Detection,” 2020. [Online]. Available: https://mangaloreuniversity.ac.in/dr-h-l-shashirekha

[3] Madani, Y., Erritali, M., and Bouikhalene, B. (2021). "Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets". *Results in Physics*, 25, 104266.

[4] Lai, C., Chen, M., Kristiani, E. et al., (2022). "Fake News Classification Based on Content Level

Features". *Applied Sciences*, 12(3), 1116.

[5] Choudhary, P., Pandey, S., Tripathi, S. et al., "Fake News Detection Based on Machine Learning". in *Lecture Notes in Electrical Engineering*, Singapore, Springer Singapore, 67-75, 2021.

[6] Chauhan, T., and Palivela, H. (2021). "Optimization and improvement of fake news detection using deep learning approaches for societal benefit". *International Journal of Information Management Data Insights*, 1(2), 100051.

[7] Salih, H. S., Ali, M. H., and Khan, M. I. (2025). "IoT-Enabled cloud storage data access control model based on blockchain technology". *International Journal of Theoretical & Applied Computational Intelligence*, vol. 2025, 125–144

[8] Ali, M. H., and Rasheed, M. A. (2025). "A blockchain-based multi-agent security framework for E-commerce systems". *International Journal of Theoretical & Applied Computational Intelligence*, vol. 2025, 227–245.

[9] James, G., Witten, D., Hastie, T. et al., "An introduction to statistical learning: with applications in R ". in *Springer Texts in Statistics*, New York, NY, Springer US, 103, 2013.

[10] Müller, A. C., and Guido, S. "Introduction to machine learning with Python: a guide for data scientists", O'Reilly Media, Inc, 2016.

[11]Li, H., Ota, K., and Dong, M. (2018). "Learning IoT in Edge: Deep Learning for the Internet of Things with Edge Computing". *IEEE Network*, 32(1), 96-101.

[12] Islam, M. S., Rony, M. A. T., Ahammad, M. et al., (2024). "An Innovative Novel Transformer Model and Datasets for Safeguarding Religious Sensitivities in Online Social Platforms". *Procedia Computer Science*, 233, 988-997.

[13] ELFAIK, H., and NFAOUI, E. H. (2024). "Automatic Detection of Fake News Using Gated Recurrent Unit Deep Model". *Procedia Computer Science*, 233, 474-480.

[14] Jouhar, J., Pratap, A., Tijo, N. et al., (2024). "Fake News Detection using Python and Machine Learning". *Procedia Computer Science*, 233, 763-771.

[15] Deepthi, L., and Nair, L. S. (2024). "Rumor Source Detection in Subgraphs: An ML Approach". *Procedia Computer Science*, 233, 454-463.

[16] Khan, S., Hakak, S., Deepa, N. et al., (2022). "Detecting COVID-19-Related Fake News Using Feature Extraction". *Frontiers in PUBLIC HEALTH*, 9.

[17] Thota, A., Tilak, P., Ahluwalia, S., et al., (2018). “Fake news detection: a deep learning approach”. *SMU Data Science Review*, *1*(3), 10.

[18] Khan, J. Y., Khondaker, M. T. I., Afroz, S. et al., (2021). "A benchmark study of machine learning models for online fake news detection". *Machine Learning with Applications*, 4, 100032.

[19] F. Nada, B. Firdous Khan, A. Maryam, and Z. Ahmed, “FAKE NEWS DETECTION USING LOGISTIC REGRESSION,” Int. Res. J. Eng. Technol., vol. 5577, 2008, [Online]. Available: http://observer.com/2017/01/

[20] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, J. Vilares, and G. Lys, “electronics Sentiment Analysis for Fake News Detection,” 2021, doi: 10.3390/electronics.

[21] Mahmud, M. A. I., Talukder, A. A. T., Sultana, A. et al., (2023). "Toward News Authenticity: Synthesizing Natural Language Processing and Human Expert Opinion to Evaluate News". *IEEE ACCESS*, 11, 11405-11421.

[22] Wang, X., Xie, H., Ji, S. et al., (2023). "Blockchain-based fake news traceability and verification mechanism". *Heliyon*, 9(7), e17084.

[23] Harris, S., Liu, J., Hadi, H. J., et al., "Ax-to-grind Urdu: benchmark dataset for Urdu fake news

detection". In *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications,* 2440-2447, 2023.

[24] Nan, Q., Wang, D., Zhu, Y., et al., (2022). "Improving fake news detection of influential domain via domain-and instance-level transfer". *arXiv preprint arXiv:2209.08902*.

[25] Saeed, R., Afzal, H., Abbas, H. et al., (2022). "Enriching Conventional Ensemble Learner with Deep Contextual Semantics to Detect Fake News in Urdu". *ACM Transactions on ASIAN and Low-resource Language Information Processing*, 21(1), 1-19.

[26] Kausar, S., Tahir, B., and Mehmood, M. A. (2020). "ProSOUL: A Framework to Identify Propaganda From Online Urdu Content". *IEEE ACCESS*, 8, 186039-186054.