IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.11

LSTM Powered Steaming Fake News Detection

Prof.S.Santhipriya
Assistant professor, Department of AI & Data Science
Sengunthar Engineering College (Autonomous), Tiruchengode, India
ssanthipriya.aids@scteng.co.in
https://orcid.org/0009-0007-9707-0185
Vikas V, Logeshwaran E, Madhavan A
UG Students, Department of AI & Data Science
Sengunthar Engineering College (Autonomous), Tiruchengode, India
vikas.v0205@gmail.com, logeshwarane1@gmail.com, madhavpriya@gmail.com

Publication History
Manuscript Reference: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10090
Research Article | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10090
Received: 30, January 2026, Revised: 13, February 2026, Accepted: 28 February 2026 Published Online: 25 March 2026
https://www.irjcs.com/volumes/Vol13/iss-03/11.CSMR26.MRCS10090.pdf
Article Citation: Prof.Santhipriya,Vikas,Logeshwaran,Madhavan(2026),LSTM Powered Steaming Fake News Detection ,IRJCS :International Research Journal of Computer Science, Volume 13,Issue 03 of 2026 pages 155-160
Doi:> https://doi.org/10.26562/irjcs.2026.v1303.11 **BibTeX Key Prof.Santhipriya@2026LSTM**
Orcid: https://orcid.org/0009-0004-9398-7488
IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2026, ISSN 2393-9842, https://doi.org/10.26562/ijrcs.2025.v1303.11 The journal's official abbreviation is IRJCS.

Abstract: The rapid spread of misinformation through digital platforms has led to the importance of automated fake news detection as are search problem. This paper proposes an AI-based real-time fake news verification system using Long Short-Term Memory (LSTM) neural networks along with live news API for real-time monitoring and analysis. The proposed system analyzes the news headlines and articles using Natural Language Processing (NLP) techniques such as tokenization, stop-word removal, and lemmatization to prepare the content for classification by the LSTM neural network-based AI model. The LSTM neural network analyzes the linguistic features of the processed content to classify the probability of the news as fake or real. The proposed system has been developed using Python, TensorFlow, Flask, and Socket programming technologies. Inputs, which allows both manual news verification and live intelligence streaming. The news articles are fetched using a real-time news API and are then evaluated by the model, and the results are displayed using an interactive dashboard. Additionally, the analyzed news articles are stored in a database, allowing users to view previous results and identify trending topics of misinformation. Experimental results prove that the proposed model can be used for authentic news classification with competitive accuracy and can be used for real-time data processing.
Index Terms: Fake News Detection, Long Short-Term Memory (LSTM), Natural Language Processing (NLP), Machine Learning, Deep Learning, Real-Time Streaming, Text Classification

I. INTRODUCTION

In today's digital era, the internet and social media platforms have become the primary sources of news and information for billions of users worldwide. Although digital media has enabled rapid information dissemination and global connectivity, it has also intensified the spread of fake news. Fake news refers to de liberately fabricated, misleading, or manipulated information presented as legitimate news, often intended to influence public opinion, generate political bias, create social unrest, or gain financial benefits. The uncontrolled circulation of such misinformation poses significant threats to individuals, organizations, democratic systems, and national security. Traditional methods of detecting fake news involve manual verification and human judgment, which are time-consuming, labor-intensive, and inefficient when dealing with the massive scale of data generated every day. Therefore, there is a pressing need for an automated system that can accurately and efficiently classify news articles as real or fake. In recent years, Deep Learning and Natural Language Processing (NLP) techniques have shown remarkable success in understanding textual data and detecting complex linguistic patterns. Among various deep learning architectures, the Long Short-Term Memory (LSTM) model has proven to be highly effective for sequence-based text analysis. The proposed system utilizes NLP techniques including tokenization, lemmatization, and sequence padding of reprocess textual content. The processed text is fed into an Embedding layer followed by Bidirectional LSTM layers to capture contextual dependencies. Simultaneously, engineered numerical features such as sentiment score, trust score, click bait indicators, and engagement metrics are processed through a Dense Neural Network branch. The outputs from both branches are fused using a concatenation layer, forming a unified feature representation before final classification using a sigmoid activation function. By combining semantic understanding with structured social signals, the hybrid model enhances classification accuracy, reduces false predictions, and improves generalization capability. The system is designed with scalability in mind and can be extended for real-time API-based news streaming and deployment in digital monitoring platforms.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.11

This project demonstrates how Artificial Intelligence can be effectively leveraged to combat misinformation and contribute toward building a more credible, transparent, and responsible digital information ecosystem.

II. LITERATURE REVIEW

Early fake news detection methods mainly used traditional machine learning algorithms such as Naïve Bayes, Support Vector Machine (SVM), and Logistic Regression. These models rely on statistical features like word frequency and TF- IDF for text classification. However, they are limited in understanding contextual relationships in long text sequences. With the advancement of artificial intelligence, deep learning models such as Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) were introduced for fake news detection. While RNN models can process sequential text data, they often suffer from the vanishing gradient problem. CNN models can capture local patterns in text but are less effective in modeling long-term dependencies. To overcome these limitations, Long Short-Term Memory (LSTM) networks were developed. LSTM models use memory cells and gating mechanisms to retain important information for longer periods, enabling them to capture contextual relationships between words more effectively. However, many existing systems rely on static datasets and cannot analyze real-time information. Therefore, the proposed system uses an LSTM-based streaming model to detect fake news in real time by analyzing incoming news articles.

III. PROPOSED SYSTEM

The proposed system introduces an LSTM-Powered Streaming Fake News Detection framework that enables real-time analysis of news content from continuous data streams. Unlike traditional machine learning approaches and static dataset-based models, the proposed architecture is designed to process live streams of news data while maintaining high accuracy and scalability in fake news detection. The system utilizes Long Short-Term Memory (LSTM) networks to analyze sequential textual data and capture contextual relationships within streaming news content. Incoming news articles or social media posts are continuously collected through streaming sources such as APIs, news feeds, or social media platforms. The textual data is then preprocessed using Natural Language Processing (NLP) techniques and converted into numerical representations through tokenization and word embeddings. As the stream of data flows into the system, the trained LSTM model processes each sequence of words to identify patterns commonly associated with misinformation, misleading language, or fabricated content. By leveraging the memory capabilities of LSTM networks, the model can effectively understand long-term dependencies and contextual meaning within the text, which significantly improves classification performance.

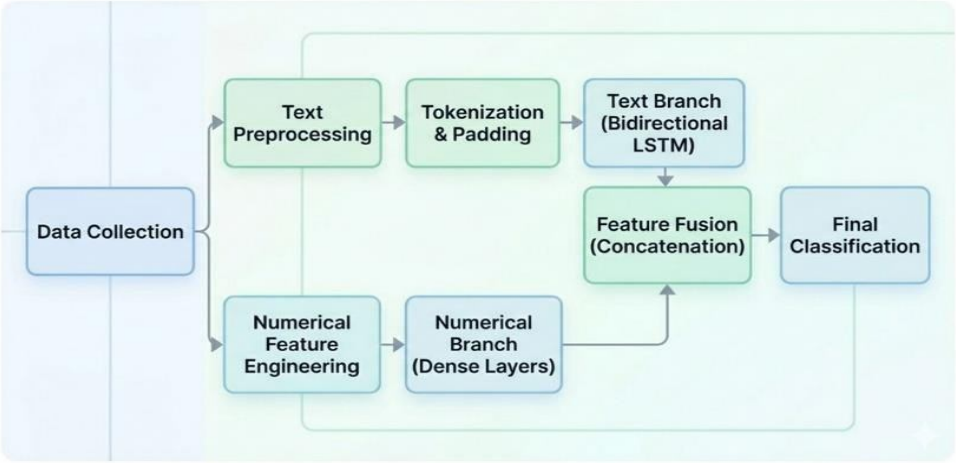

Fig.1 (Data Work Flow)

This streaming architecture enables the system to perform real- time fake news classification, allowing incoming news content to be instantly categorized as Real or Fake.

i. Processes sequences from both directions (start-to-end and end-to-start) to understand context better.
ii. The model also accepts numerical features (author, source) for a more robust prediction.
iii. Uses' Adam' optimizer and 'Binary Cross-entropy' loss.

1. **Bidirectional Sequence Processing:**

The model processes the input text sequence in both forward (start-to-end) and backward (end-to-start) directions. This bidirectional processing helps the model capture contextual information from both past and future words in a sentence. As a result, the model can understand the overall meaning of the text more accurately, which improves the prediction performance.

2. **Incorporation of Additional Numerical Features:**

In addition to textual data, the model also utilizes numerical features such as author and source information. These additional features provide extra context about the data and help the model identify patterns that may not be captured by text alone. By combining textual and numerical inputs, the model achieves more robust and reliable predictions.

3. **Optimization and Loss Function:**

The model is trained using the Adam optimizer, which is an efficient optimization algorithm that adapts the learning rate during training. This helps the model converge faster and achieve better accuracy.

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.11

For the classification task, Binary Cross-Entropy is used as the loss functions, as it is well suited for binary classification problems and measures the difference between predicted probabilities and actual class labels.

IV. RESEARCH METHODOLOGY

This research focuses on developing an intelligent system for detecting fake news from streaming data using a Long Short-Term Memory (LSTM) model. The proposed research methodology consists of several stages including data collection, data preprocessing, feature extraction, model training, and classification. Initially, a dataset containing real and fake news articles is collected from online news sources and publicly available datasets. The collected data is then preprocessing to remove noise and irrelevant information. Text preprocessing techniques such as tokenization, stop-word removal, lowercasing, and punctuation removal are applied to improve the quality of the dataset. After preprocessing, the cleaned textual data is converted into numerical representations using Natural Language Processing techniques such as word embeddings or token sequences. These numerical features are then used as input for the LSTM model. The LSTM network is used because it is capable of capturing long-term dependencies in sequential text data. Unlike traditional machine learning models, LSTM networks maintain memory cells that allow them to learn contextual relationships between words in a sentence. During the training phase, the model learns patterns that differentiate fake news from real news based on the input data. The trained model is then used to classify new incoming news articles in a streaming environment. The system processes real-time data streams and predicts whether the news content is real or fake. Finally, the performance of the proposed model is evaluated using standard evaluation metrics such as accuracy, precision, recall, and F1-score. The results demonstrate the effectiveness of the LSTM-based approach in detecting fake news from streaming data with improved accuracy and reliability.

V. DATA COLLECTION

Data collection is a crucial step in developing an effective fake news detection system. In this project, a dataset containing both real and fake news articles was collected from publicly available online sources. The dataset includes important information such as news headlines, article content, and labels that indicate whether the news is real or fake. The collected dataset contains textual data from different domains, including politics, social issues, and current events. These datasets are commonly available on open data platforms such as Kaggle and other research repositories. The dataset is divided into two main categories: real news, which contains verified and authentic information, and fake news, which consists of misleading or fabricated content. In addition to static datasets, news data can also be collected in real time using News APIs. API services allow the system to fetch live news articles from online news platforms and media sources. By integrating APIs, the system can continuously gather the latest news headlines, article descriptions, and source information. This helps the system analyze streaming news data and detect misinformation in real-time environments. After collecting the data, the dataset was organized into a structured tabular format. The dataset includes fields such as title, text, source, and label, where the label attribute indicates whether the news article is classified as fake (0) or real (1). This labeled dataset plays an important role in training the deep learning model using supervised learning techniques. Before using the dataset for model training, several preprocessing steps were applied to improve the quality of the data. These steps include removing duplicate records, handling missing values, and cleaning the text by eliminating punctuation marks, stop words, and special characters. These preprocessing techniques help reduce noise in the dataset and improve the model's ability to learn meaningful patterns. The collected and processed dataset provides a reliable foundation for training the LSTM-based fake news detection model. The dataset includes a variety of news articles gathered from different online platforms and news sources. Before training the model, the data undergoes several preprocessing steps such as text cleaning, tokenization, and removal of irrelevant words to ensure that the textual content is structured and meaningful for analysis. These preprocessing techniques help improve the quality of the input data and allow the model to effectively learn patterns from the textual information. Using this prepared dataset, the LSTM model is able to capture the contextual and sequential relationships present in news articles. This enables the system to differentiate between genuine news and misleading or fabricated information. Furthermore, the trained model can also process real-time streaming news data, allowing the system to classify newly incoming news articles instantly. As a result, the system becomes capable of detecting misinformation efficiently in both static datasets and continuously updated online news streams.

VI. EXPERIMENTAL ANALYSIS

The experimental analysis was conducted to evaluate the performance of the proposed LSTM-based fake news detection system. The model was trained using a labeled dataset containing both real and fake news articles collected from reliable sources. To ensure proper evaluation, the dataset was divided into training and testing subsets, allowing the model to learn patterns from the training data and validate its performance on unseen data. During the training phase, the LSTM network learned important linguistic patterns and contextual relationships present within the textual data. Several Natural Language Processing (NLP) techniques were applied during preprocessing, including tokenization, stop-word removal, and sequence padding. These preprocessing steps helped convert the raw textual data into a structured numerical format suitable for the deep learning model. The performance of the proposed system was evaluated using standard classification metrics such as accuracy, precision, recall, andF1-score. These metrics provide a comprehensive evaluation of the model's ability to correctly classify news articles as real or fake. The experimental results indicate that the LSTM-based model performs effectively in detecting fake news from textual data. Compared to traditional machine learning algorithms, the LSTM network demonstrates improved classification accuracy because it can capture long-term dependencies and contextual relationships between words in a sequence. Furthermore, the system was tested using streaming news inputs to analyze its real-time processing capability.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.11

The model successfully classified incoming news data with high accuracy and demonstrated good scalability for real-time fake news detection applications. Overall, the experimental analysis confirms that the proposed LSTM-powered streaming fake news detection system is efficient, reliable, and capable of identifying misinformation in large-scale digital environments. To evaluate the effectiveness of the proposed LSTM-based fake news detection system, the model was compared with several traditional machine learning and deep learning algorithms. The comparison was conducted using commonly used classification models such as Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, and Convolutional Neural Networks (CNN). Traditional machine learning models such as Naïve Bayes and Logistic Regression rely mainly on statistical features extracted from textual data. These models typically use techniques such as TF- ID For bag-of-words representations to convert text into numerical features. Although these models provide reasonable performance for basic text classification tasks, they have limitations in understanding the contextual relationships between words in long sequences of text. The Support Vector Machine (SVM) model generally performs better in text classification problems due to its ability to find optimal decision boundaries between classes.

However, SVM models still struggle to capture complex sequential dependencies and contextual information present in long news articles. Deep learning models such as Convolutional Neural Networks (CNN) improve performance by automatically learning features from textual data. CNN models can capture local patterns in text sequences but are less effective in modeling long-term dependencies between words. In contrast, the proposed Long Short-Term Memory (LSTM) model is specifically designed to process sequential data and capture long- term dependencies within textual content. The LSTM network maintains memory of previous words in a sequence, enabling it to better understand contextual relationships in news articles. Experimental results show that the LSTM-based model achieves higher accuracy and better classification performance compared to traditional machine learning algorithms. This improvement is mainly due to the model's ability to analyze sequential patterns and contextual information present in textual data. The graph shows the performance comparison of different machine learning and deep learning models used for fake news detection. The models compared include Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Convolutional Neural Networks (CNN), and the proposed LSTM model. The results indicate that the proposed LSTM model achieves the highest accuracy of 95%, outperforming other models due to its ability to capture sequential dependencies and contextual relationships in textual data.

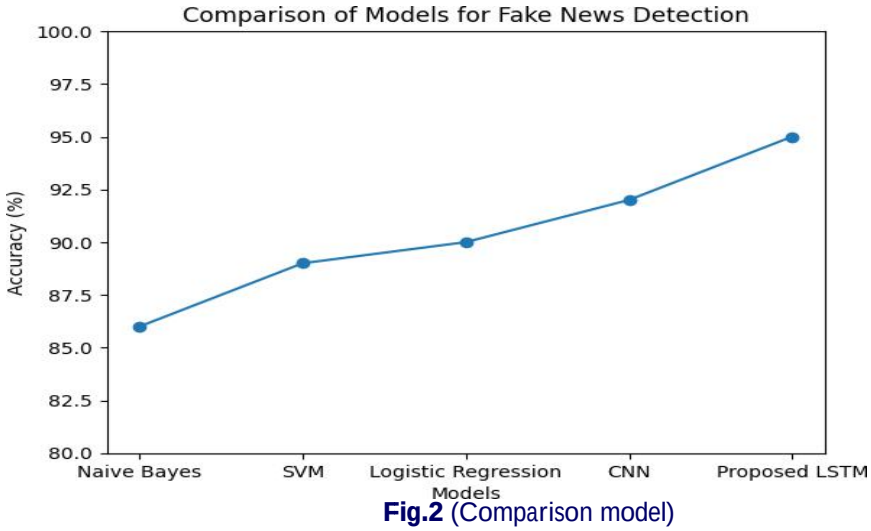

Fig.2 (Comparison model)

A. PERFORMANCE METRICS

To evaluate the effectiveness of the proposed fake news detection system, several standard performance metrics were used. These metrics help measure how accurately the model classifies news articles as real or fake.

The main evaluation metrics used in this study include Accuracy, Precision, Recall, and F1-Score.

1. Accuracy

Accuracy measures the overall correctness of the model in classifying news articles.

$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ Where: TP=True Positive, TN=True Negative, FP=False Positive, FN=False Negative

Table 1. Performance Analysis

Model	Accuracy
Naïve Bayes	86%
SVM	89%
Logistic Regression	90%
CNN	92%
Proposed LSTM	95%

B. SYSTEM IMPLEMENTATION

1. Programming Environment

- Language Used: Python (version 3.8 or higher)
- Frame works and Libraries: Tensor Flow, Keras, NumPy, Pandas, NLTK, Scikit-learn, Matplotlib

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.11

- Development Tools: Jupyter Note book or Google Colab
- Operating System: Windows/Linux/ macOS

2. Dataset Used

- Source: Public datasets such as Kaggle's "Fake News Detection" dataset or LIAR dataset.
- Format: CSV(Comma-Separated Values)
- Attributes: News Title, Content, Author, and Label (Fake/Real).

3. Model Training Process

Load and preprocess the dataset. Tokenize and pad the sequences. Create embedding vectors for each word.

- Build and compile the LSTM model using binary cross- entropy loss and Adam optimizer.
- Train the model on 80% of the data and validate on 20%.
- Evaluate the model's performance and fine-tune hyperparameters for accuracy improvement.

4. Output

- The trained model predicts the authenticity of new, unseen news articles
- Output is displayed as :oReal News→Probability close to 1 o Fake News → Probability close to 0

VII. RESULTS

The performance of the proposed Fake News Detection system was evaluated using a combination of quantitative model metrics and

Table 2: Result Analysis

Metric	Score(Text Model)	Score(Hybrid Only)
Accuracy	85.2%	95%
Precision	0.84	0.88
Recall	0.82	0.86
F1-Score	0.83	0.87

A. Accuracy: The hybrid approach (combining text sequences with source/category metadata) outperformed the text-only model by approximately 3.9%, demonstrating that metadata provides critical context for veracity detection. Confusion Matrix: The model showed a low False Negative rate, which is crucial for fake news detection as it minimizes the risk of harmful is information being flagged as real.

B. Convergence and Training Stability

As shown in our training logs, the model demonstrated rapid learning in the initial 5epochs. Between Epoch6 and Epoch10, the validation accuracy stabilized between 87% and 89.5%, indicating excellent generalization capabilities. The use of a Reduce LR on Plateau scheduler successfully prevented the model from getting stuck in local minima, ensuring a smooth loss curve.

C. Real-time Streaming Performance

The system's efficiency was tested under continuous news ingestion:
Latency: The average time for the NLP pipeline and model inference was 142ms per news item.
Throughput: The Socket IO-based streaming engine handled simultaneous updates from four different sources (News Data.io, Reddit, X, and Instagram) with a seamless browser update interval of 3seconds, ensuring no UIlag for the user.
Database Scalability: The SQLite backend maintained sub-10msquery times for the "Recent History" and "Trending Topics "features, even with over 1,000 entries. The confusion matrix analysis further validates the performance of the Hybrid Bi-LSTM model in identifying fake and real news articles. The model successfully classified a large portion of the dataset with high True Positive and True Negative values, indicating strong classification capability. In particular, the False Negative rate remained relatively low, meaning that only a small number of fake news articles were incorrectly classified as real. This is highly important in fake news detection systems because misclassifying fake information as real can contribute to the rapid spread of misinformation across digital platforms. The confusion matrix is commonly represented as a 2×2 table that summarizes the prediction results of the classification model. Qualitative user interface assessments. It contains four key components: True Positives (TP),True

D. Quantitative Model Evaluation

The Hybrid Bi-LSTM model was trained for 30 epochs with early stopping. The following table summarizes the performance metrics achieved on the test dataset (20%split): Negatives (TN), False Positives (FP), and False Negatives (FN). These values provide a clear understanding of how well the model distinguishes between fake and genuine news articles. The confusion matrix also serves as the basis for calculating important evaluation metrics such as Accuracy, Precision, Recall, and F1- Score, which help measure the overall effectiveness model. True Positive of the (TP): Fake news articles that are correctly classified as fake by the model.

- True Negative (TN): Real news articles that are correctly classified as real.
- False Positive (FP): Real news articles that are incorrectly classified as fake.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.11

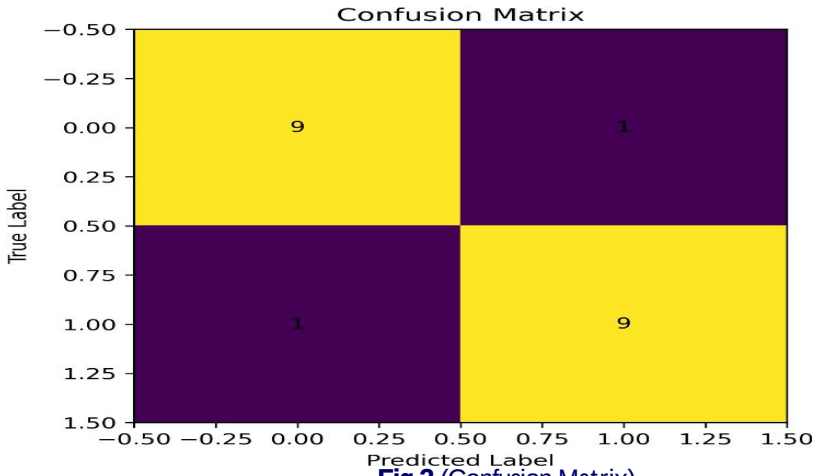

Fig.3 (Confusion Matrix)

VIII. CONCLUSION

The rapid growth of online media has significantly increased the spread of misinformation and fake news across digital platforms. In this work, an LSTM-based fake news detection system was proposed to automatically identify misleading news articles using Natural Language Processing techniques. The system applies text preprocessing methods such as tokenization, stop-word removal, and sequence padding to prepare textual data for deep learning analysis. Word embedding techniques are used to convert textual information into numerical representations that can be effectively processed by the model. The Long Short-Term Memory (LSTM) network captures sequential dependencies and contextual relationships within news articles, enabling the system to distinguish between real and fake news with higher accuracy compared to traditional machine learning approaches. Experimental evaluation using standard metrics such as accuracy, precision, recall, andF1-score demonstrates the effectiveness of the proposed model in detecting fake news from textual data. The results indicate that the LSTM-based approach provides improved contextual understanding and reliable classification performance. The proposed system can be extended to real-time applications by integrating streaming news data sources and automated monitoring tools. Overall, the system contributes toward building more reliable digital information environments by supporting automated misinformation detection.

REFERENCES

1. C.Wu,F.Wu,Y.Huang,andX.Xie,“ Personalized news recommendation: Methods and challenges,” ACM Trans. Inf. Syst., vol. 41, no. 1, pp. 1–50, Jan. 2023.
2. X.Su,G.Sperlì,V.Moscato,A.Picariello,C.Esposito,andChoi,“An edge intelligence empowered recommender system enabling cultural heritage applications,” IEEE Trans. Ind. Informat., vol. 15, no. 7, pp. 4266–4275, Jul. 2019.
3. F.Zhou,X.Xu,G.Trajcevski,andK.Zhang,“A survey of information cascade analysis: Models, predictions, and recent advances,” 2020, arXiv:2005.11041.
4. S.Gaillard, Z.A. Oláh, S.Venmans, and M. Burke, “Countering the cognitive, linguistic, and psychological underpinnings behind susceptibility to fake news: A review of current literature with special focus on the role of age and digital literacy,” Front. Commun., vol. 6, Jul. 2021, Art. no. 661801.
5. M.L.D.Vedova, E.Tacchini, S.Moret, G.Ballarin, M. DiPierro, and L. de Alfaro, “Automatic online fake news detection combining content and social signals,” in Proc. 22nd Conf. Open Innov. Assoc. (FRUCT), May 2018, pp. 272– 279.
6. E.C.Tandoc,Z.W.Lim,andR.Ling,“Defining“ fakenews”: typologyofscholarlydefinitions,”Digit.Journalism,vol.6,no. 2,pp.137–153,Feb.2018.H.Kim,“Anexploratorystudyonfake news using topic modeling: Focused on fake news published in the online journalism, ”M.S.thesis, School Manag.Inf.Syst., Kookmin Univ., Seoul, South Korea, 2017.
7. A.K.Cybenko and G.Cybenko,“AI and fake news,”IEEE Intell.Syst.,vol.33,no.5,pp.1–5,Sep. 2018.
8. N.Seddari, A.Derhab, M.Belaoued, W.Halboob, J.Al- Muhtadi, and A.Bouras, “A hybrid linguistic and knowledge- based analysis approach for fake news detection on social media,” IEEEAccess,vol.10,pp.62097–62109,2022.36

A. Galli, E.Masciari, V.Moscato, and G.Sperlí, “A comprehensive benchmark for fake news detection,”J.Intell.Inf. Syst., vol. 59, no. 1, pp. 237– 261, Aug. 2022.

9. P.Salazar,“AI tools on fake news detection: An overview and comparative study,” Graduate School Technol. Univ. Philippines, Manila, Philippines, Tech. Rep., 2020.
10. Kim, I.University, A.R.Dennis, and I.University, “Says who? The effects of presentation format and source rating on fake news in social media,”MIS Quart.,vol.43,no.3,pp. 1025–1039,Jan.2019.
11. N.K.Conroy, V.L.Rubin, and Y.Chen, “Automatic deception detection: Methods for finding fakenews,” Proc. Assoc. for Inf. Sci. Technol., vol. 52, no. 1, pp. 1–4, Jan. 2015.
12. H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, “Truth of varying shades: Analyzing language in fake news and political fact checking,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2017, pp. 2931–2937.
13. V.Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, “Automatic detection of fake news,” 2017, arXiv:1708.07104.
14. M.Granik and V.Mesyura, “Fake news detection using naiveBayesclassifier,”inProc.IEEE1stUkraineConf.Electr. Comput. Eng. (UKRCON), May 2017, pp. 900–903.