IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

Enhanced Approach For Fake Review Detection Using LSTM and RNN

Prof. A. A. Pund[1], Rohit Surve[2], Chetan Sonawane[3], Gauri Sutar[4], Najiya Sayyad[5]
[1]Prof., Department of Information Technology
[2,3,4,5]Students, Department of Information Technology
Dr. Vithalrao Vikhe Patil College of Engineering, Ahmednagar
Ahilyanagar Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract: *In today's digital world, customer decisions are highly influenced by online reviews, but the rise of fake reviews has made it difficult to trust this information. Thispaper aims to develop a system that can accurately identify fraudulent reviews using advanced techniques. Natural Language Processing (NLP) is used to analyze the textual content of reviews, while deep learning models such as RNN(Recurrent Neural Networks) and LSTM help capture patterns and context within the data. The system also considers user behavior, such as rating patterns and review activity, to improve detection accuracy. By combining both textual and behavioral features, the model can effectively distinguish between genuine and fake reviews. The performance of the system is evaluated using standard metrics like accuracy and F1-score. Overall, the proposed solution helps improve the reliability of online platforms and supports better decision-making for users.*

Keywords: LSTM, Recurrent Neural Networks,fake reviews,F1-score,accuracy

I. INTRODUCTION

The rapid expansion of digital commerce has fundamentally altered consumer habits, positioning online testimonials as a critical determinant in the evaluation of product quality and service experiences ("Deep Learning Hybrid Approaches to Detect Fake Reviews and Ratings," 2023; Mohawesh et al., 2021). As these platforms become central to daily life, the integrity of review systems is increasingly compromised by the prevalence of fabricated feedback. Such deceptive content is often strategically deployed—either to artificially inflate a brand's reputation or to maliciously disparage competitors—leading to significant consumer confusion and suboptimal decision-making (Alsubari et al., 2021; Mohawesh et al., 2021).A primary obstacle in identifying these fraudulent entries is their sophisticated mimicry of authentic human sentiment. Unlike traditional spam, modern fake reviews are crafted to mirror the linguistic nuances, emotional depth, and stylistic traits of genuine users, making manual detection highly unreliable (Alsubari et al., 2021; Mohawesh et al., 2021). This creates an urgent requirement for automated detection frameworks capable of processing large-scale datasets to uncover patterns indicative of deception (Mohawesh et al., 2021; Xu et al., 2024).

Recent developments in Natural Language Processing and machine learning have provided the tools necessary for more granular textual analysis. In particular, deep learning architectures such as Recurrent Neural Networks and Long Short-Term Memory networks have proven effective in capturing the sequential dependencies and contextual subtleties inherent in human language (Bhargava et al., 2018; Xu et al., 2024). By training on diverse datasets, these models can identify complex, hidden discrepancies between authentic narratives and fabricated accounts, thereby strengthening the reliability of online review ecosystems (Alsubari et al., 2021; Dai & Wang, 2024).

II. PROBLEM STATEMENT

With the rapid growth of online platforms, user reviews have become an important source of information for customers when making purchasing decisions. However, the increasing number of fake or misleading reviews has reduced the reliability of these platforms. Such reviews are often created intentionally to promote certain products or to damage the

DOI: 10.48175/IJARSCT-36769

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

reputation of competitors, which can mislead users and negatively affect genuine businesses. The main problem lies in accurately identifying these fake reviews, as they are often written in a way that closely resembles real human opinions. Traditional methods based on simple rules or manual checking are not sufficient to handle large volumes of data or to detect complex patterns in review content. Therefore, there is a need to design an efficient and automated system that can analyze review data, extract meaningful features, and classify reviews as genuine or fake with high accuracy.This project aims to address this issue by developing a system that uses Natural Language Processing and deep learning techniques to detect fraudulent reviews. The system focuses on analyzing both the textual content and user behavior to improve detection performance and ensure more reliable results.

III. LITERATURE REVIEW

[1] **Xu et al.** propose a hybridized approach for enhanced fake review detection that utilizes an attention-based long short-term memory (**A-LSTM**) model. The system introduces an attention gate to identify word importance and an attention layer to learn long-term semantic dependencies. By combining these linguistic representations with reviewer behavioral features, the model achieves a detection accuracy of **90.9%** on the Yelp dataset.

[2] **Abd-Alhalem et al.** present a deep learning fusion approach for identifying fraudulent reviews on e-commerce platforms using aspect-based sentiment analysis. The methodology employs an **Aspect Fusion Network** to select significant aspect words and a **Cardinality Fusion Model** to mitigate the effects of stochastic weights in auxiliary models. This framework effectively bridges the gap between genuine and deceptive reviews by focusing on interpreting nuanced human language.

[3] **Ram et al.** develop a fake review detection system using supervised machine learning algorithms, including **SVM, Random Forest, and Naive Bayes**. The system utilizes Natural Language Processing (NLP) to extract text features such as writing style and sentiment, while also considering reviewer behavioural patterns like review time and date. The techniques are designed to optimize performance metrics including accuracy, precision, and F1-score across various datasets.

[4] **Ennaouri and Zellou** explore a hybrid approach that integrates **BERT embeddings, fuzzy logic, and a multicentroid K-means classifier** for robust detection. The system uses fuzzy logic membership functions to highlight extreme linguistic patterns and "extravagant words" that signify deception. By dynamically weighting linguistic cues with deep semantic features via an attention mechanism, the model improves interpretability and classification accuracy.

[5] **Jnoub et al.** introduce a fact-checking reasoning system for fake review detection based on **Answer Set Programming (ASP)**. The framework utilizes two knowledge bases—one for review classification and another for author classification—to analyze both textual (review-centric) and behavioural (user-centric) features. The system models timestamps and source IP addresses to identify suspicious patterns, such as multiple reviews posted within short time intervals.

[6] **Shunxiang et al.** frame fake review detection as a data streaming classification task, proposing a model based on **sentiment intensity and PU learning (SIPUL)**. The system divides review data into subsets based on sentiment intensity and iteratively trains the model to adapt to new data arrival. This continuous improvement cycle helps the model identify hidden features in fake reviews that often lose their concealment in updated contexts.

7] **Liu et al.** propose a model that incorporates **part-of-speech (POS) and first-person pronoun features** into bidirectional long short-term memory (**BiLSTM**) networks. The methodology, referred to as BiLSTMWF, embeds these specific linguistic markers directly into word embeddings to enhance document-level representations. By combining these enriched text features with behavioral data, the system achieves superior results compared to standard weighted neural network models.

[8] **Kennedy et al.** evaluate the effectiveness of transformer-based architectures, specifically **BERT**, for deceptive opinion spam detection. The research demonstrates that fine-tuning pre-trained BERT models on specific deception

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, June 2026 Impact Factor: 8.2

datasets can yield accuracies exceeding **90%**. However, the study also notes challenges related to the high number of parameters and the necessary truncation of long reviews required by the transformer structure.

[9] **Shehnepoor et al.** address the challenge of data scarcity by introducing **ScoreGAN**, a generative adversarial network designed for fake review detection. The system utilizes an LSTM-based generator and a CNN-based discriminator, incorporating reviewer behavioral features into the generation process. This approach not only improves detection performance but also enables data augmentation for better training on smaller datasets.

[10] **Budhi and Chiong** propose a **Multi-type Classifier Ensemble (MICE)** model that integrates multiple base classifiers to adapt to diverse domains. The system leverages the individual strengths of various classifiers to compensate for weaknesses when detecting reviews across hotels, restaurants, and medical services. Experimental results confirm that this ensemble method outperforms single-type classifier approaches in adapting to the varied linguistic styles of different industries.

IV. SYSTEM METHODOLOGY

The framework for the proposed fake review detection system is structured into three primary segments: the User Interface, the Processing & Analysis Module, and the Deep Learning Classification Module. These components collaborate to identify and exclude fraudulent submissions. By integrating Natural Language Processing with deep learning architectures, the system ensures high levels of detection precision and reliability.

1. User Interface

This layer manages the interaction between the system and the external platforms or users providing the data. It facilitates the submission of product reviews or the retrieval of existing datasets from online marketplaces. The input typically consists of text content, numerical ratings, and metadata related to the reviewer. Every entry is captured and channelled into the system for subsequent evaluation.

2. Processing & Analysis Module

Serving as the central data management hub, this module prepares the raw input through a sequence of refinement and extraction phases:

Data Acquisition: Aggregating reviews from digital sources or specific datasets.

Data Sanitization: Eliminating noise, such as redundant entries, special symbols, and non-essential characters.

Textual Preprocessing: Standardizing text through tokenization, the removal of stop-words, and stemming.

Feature Engineering: Distilling critical indicators including TF-IDF vectors, sentiment polarity, and linguistic markers.

Behavioural Assessment: Examining reviewer-specific traits, such as historical rating trends and the frequency of submissions.

The refined data is then transitioned to the classification stage.

3. Deep Learning Classification Module

This component utilizes sophisticated computational models to distinguish between authentic and fabricated reviews through the following workflow:

Embedding Generation: Mapping words into high-dimensional vector spaces (using BERT, GloVe, or Word2Vec) to retain semantic converts context.

Model Training: Utilizing a hybrid Recurrent Neural Network and Long Short-Term Memory architecture to process both immediate and long-range dependencies within the text.

Contextual Analysis: Deciphering sequential data patterns and the relationships between words in a review.

Classification: Assigning a "Real" or "Fake" label to each review based on the identified patterns.

Evaluation: Assessing the system's performance through metrics such as precision, recall, accuracy, and F1-score.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

4. Output Generation

The final phase produces the results by labelling reviews according to their verified authenticity. By distinguishing legitimate feedback from deceptive content, the system assists users in making better-informed choices while protecting the integrity of online review ecosystems.

V. SYSTEM METHODOLOGY

The system for detecting fake reviews follows a structured path starting with data collection from logs, followed by filtration and noise removal to ensure the text is clean and relevant. This processed data is then split into training and testing sets using cross-validation to optimize overall model performance. During the training phase, the system extracts key features and them into numerical embeddings to capture the semantic depth and context of the reviews. These embeddings are fed into a hybrid RNN and LSTM model, which learns both short-term and long-term textual dependencies while an optimizer minimizes errors to refine accuracy. Finally, unseen data is passed through identical preprocessing steps and analysed by the trained classifier to predict whether a review is real or fake. This streamlined approach ensures precise detection and strengthens the reliability of online review platforms.

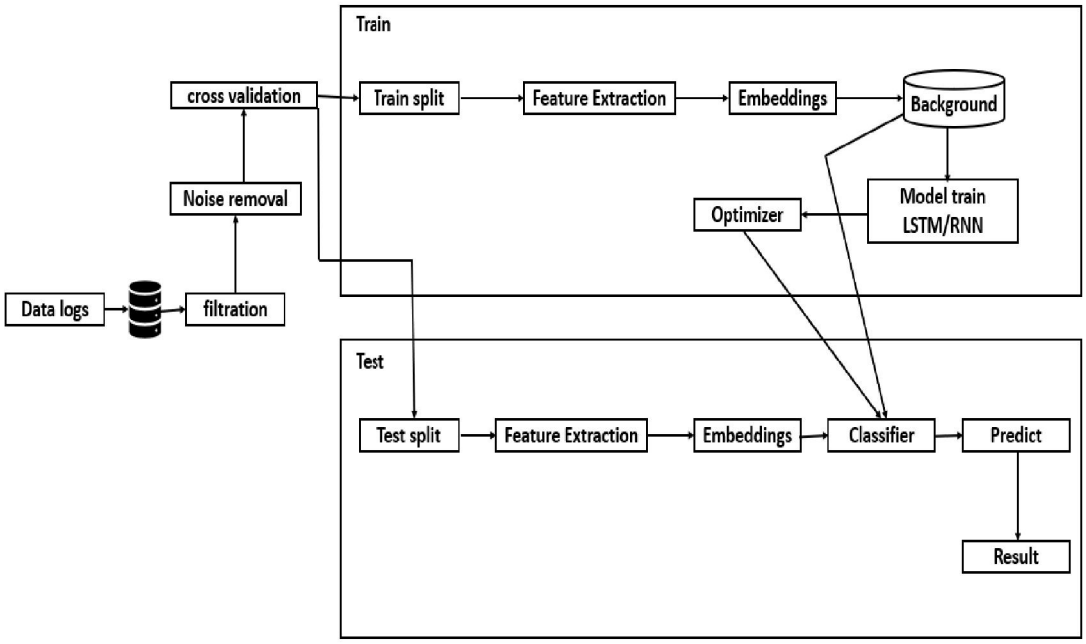

Fig.1: System Architecture Design

VI. RESULT AND ANALYSIS

The results and discussion section presents a comprehensive evaluation of the proposed fake review detection system developed using machine learning and deep learning techniques. The system aims to identify deceptive online reviews by analyzing textual patterns, contextual information, and sequential dependencies present in review content. To assess the effectiveness of the proposed approach, multiple classification models including Naive Bayes, Support Vector Machine (SVM), Random Forest, Recurrent Neural Network (RNN), and the proposed RNN-LSTM model were evaluated and compared.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

Model	Accuracy (%)	Error Rate (%)	Performance Analysis
Naive Bayes	81.5	18.5	Limited contextual understanding
SVM	87.3	12.7	Improved classification capability
Random Forest	89.1	10.9	Enhanced prediction stability
RNN	91.4	8.6	Sequential dependency learning
Proposed RNN-LSTM	94.2	5.8	Superior long-term contextual modeling

Table 1. Comprehensive Performance Analysis of Fake Review Detection Models

Table I presents the comparative performance analysis of all evaluated models. The results indicate that the proposed RNN-LSTM model achieved the highest classification accuracy of 94.2% while maintaining the lowest error rate of 5.8%. In comparison, Naïve Bayes achieved an accuracy of 81.5%, SVM achieved 87.3%, Random Forest achieved 89.1%, and the conventional RNN achieved 90.4%. These findings demonstrate the superiority of the proposed architecture in accurately distinguishing between genuine and deceptive reviews.

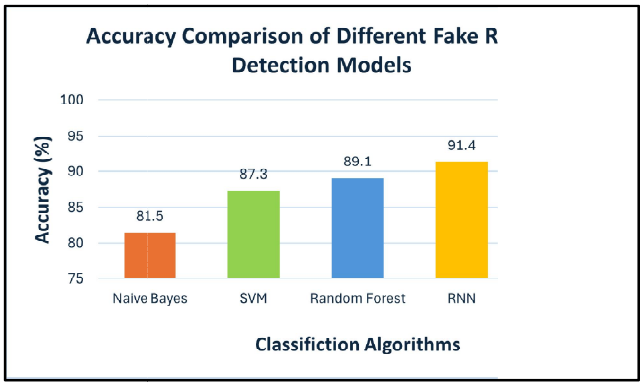

Figure 2: Comparative accuracy analysis of machine learning and deep learning models for fake review detection

Figure 2 illustrates the accuracy comparison among different fake review detection models. It can be observed that deep learning approaches outperform traditional machine learning techniques, with the proposed RNN-LSTM model achieving the best overall performance.

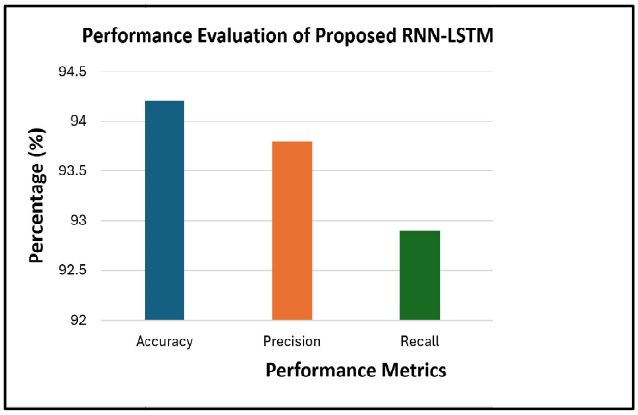

Figure 3: Performance evaluation of the proposed RNN-LSTM model using accuracy, precision, recall, and F1-score

Figure 3 presents the performance evaluation of the proposed RNN-LSTM model using these classification metrics. The model achieved an accuracy of 94.2%, precision of 93.8%, recall of 92.9%, and F1-score of 93.3%. The close distribution of these values indicates balanced classification behaviors without significant bias toward any particular

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, June 2026 Impact Factor: 8.2

class. High precision reduces false positive predictions, while high recall ensures effective detection of deceptive reviews. The strong F1-score further confirms the reliability and robustness of the proposed approach.

VII. CONCLUSION

The proposed fake review detection system helps in identifying misleading and deceptive reviews on online platforms. By using Natural Language Processing (NLP) along with RNN-LSTM models, the system is able to analyse review text and distinguish between genuine and fake reviews more effectively. The results obtained show that the proposed model performs better than traditional machine learning techniques in terms of classification accuracy. The system can reduce manual effort in reviewing large amounts of online feedback and can assist users in making more reliable decisions based on reviews. Overall, the proposed approach provides an effective solution for improving the quality and authenticity of online review systems. Future work may include testing the model on larger datasets and incorporating advanced deep learning techniques to further enhance detection performance.

REFERENCES

[1]. S. Xu, H. Cuan, Z. Yin, and C. Yin, "A Hybridized Approach for Enhanced Fake Review Detection," IEEE Transactions on Computational Social Systems, vol. 11, no. 6, pp. 7448–7466, Dec. 2024, doi: 10.1109/TCSS.2024.3411635.

[2]. S. M. Abd-Alhalem, H. A. Ali, N. F. Soliman, A. D. Algarni, and H. S. Marie, "Advancing E-Commerce Authenticity: A Novel Fusion Approach Based on Deep Learning and Aspect Features for Detecting False Reviews," IEEE Access, vol. 12, pp. 116055–116070, 2024, doi: 10.1109/access.2024.3435916.

[3]. M. Ennaouri and A. Zellou, "Enhancing Fake Review Detection Using Linguistic Exaggeration, BERT Embeddings, and Fuzzy Logic," IEEE Access, vol. 13, pp. 135956–135970, 2025, doi: 10.1109/access.2025.10820123.

[4]. N. Jnoub, A. Brankovic, and W. Klas, "Fact-Checking Reasoning System for Fake Review Detection Using Answer Set Programming," Algorithms, vol. 14, no. 7, p. 190, Jul. 2021, doi: 10.3390/a14070190.

[5]. N. C. S. Ram et al., "Twitter Data Pre-Processing and Detection of Fake Reviews," Seybold Report Journal, vol. 19, no. 1, 2024.

[6]. G. S. Budhi and R. Chiong, "A Multi-Type Classifier Ensemble for Detecting Fake Reviews through Textual-Based Feature Extraction," ACM Transactions on Internet Technology, vol. 23, no. 1, pp. 1–24, Apr. 2023, doi: 10.1145/3568676.

[7]. S. Shehnepoor, R. Togneri, W. Liu, and M. Bennamoun, "ScoreGAN: A Fraud Review Detector Based on Regulated GAN with Data Augmentation," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 280–291, 2022, doi: 10.1109/TIFS.2022.3142247.

[8]. W. Liu, W. Jing, and Y. Li, "Incorporating Feature Representation into BILSTM for Deceptive Review Detection," Computing, vol. 102, no. 3, pp. 701–715, 2020, doi: 10.1007/s00607-019-00763-w.

[9]. Z. Shunxiang, Z. Aoqiang, Z. Guangli, W. Zhongliang, and L. Kuan-Ching, "Building Fake Review Detection Model Based on Sentiment Intensity and PU Learning," IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 10, pp. 6926–6939, Oct. 2023, doi: 10.1109/TNNLS.2021.3135593.

[10]. M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding Deceptive Opinion Spam by Any Stretch of the Imagination," in Proc. 49th Annual Meeting of the Assoc. for Computational Linguistics: Human Language Technologies, 2011, pp. 309–319.

[11]. N. Jindal and B. Liu, "Opinion Spam and Analysis," in Proc. First ACM Int. Conf. Web Search and Data Mining, 2008, pp. 219–230

[12]. S. Shehnepoor, M. Salehi, R. Farahbakhsh, and N. Crespi, "NetSpam: A Network-Based Spam Detection Framework for Reviews in Online Social Media," IEEE Transactions on Information Forensics and Security, vol. 12, no. 7, pp. 1585–1595, Jul. 2017, doi: 10.1109/TIFS.2017.2675361.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, June 2026
Impact Factor: 8.2

[13]. P. Hajek, A. Barushka, and M. Munk, "Fake Consumer Review Detection Using Deep Neural Networks Integrating Word Embeddings and Emotion Mining," Neural Computing and Applications, vol. 32, no. 23, pp. 17259–17274, 2020.
[14]. S. Kennedy, N. Walsh, K. Sloka, J. Foster, and A. McCarren, "Fact or Factitious? Contextualized Opinion Spam Detection," arXiv preprint, arXiv:2010.15296, 2020.
[15]. L. Ilias, F. Soldner, and B. Kleinberg, "Explainable Verbal Deceptive Detection Using Transformers," arXiv preprint, arXiv:2210.03080, 2022.
[16]. Y. Ren and D. Ji, "Neural Networks for Deceptive Opinion Spam Detection: An Empirical Study," Information Sciences, vol. 385, pp. 213–224, Apr. 2017.
[17]. J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a General Rule for Identifying Deceptive Opinion Spam," in Proc. 52nd Annual Meeting of the Assoc. for Computational Linguistics, vol. 1, 2014, pp. 1566–1576.
[18]. S. Feng, R. Banerjee, and Y. Choi, "Syntactic Stylometry for Deception Detection," in Proc. 50th Annual Meeting of the Assoc. for Computational Linguistics, vol. 2, 2012, pp. 171–175
[19]. X. Wang, K. Liu, S. He, and J. Zhao, "Learning to Represent Review with Tensor Decomposition for Spam Detection," in Proc. Conf. Empirical Methods in Natural Language Processing, 2016, pp. 866–875.
[20]. S. Rayana and L. Akoglu, "Collective Opinion Spam Detection: Bridging Review Networks and Metadata," in Proc. 21st ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2015, pp. 985–994.matics (ICCCI), Coimbatore, India, 2023, pp. 1-7, doi: 10.1109/ICCCI56745.2023.10128545