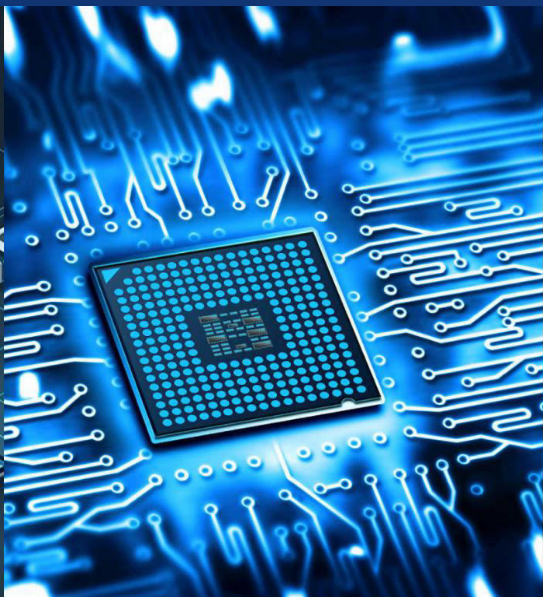
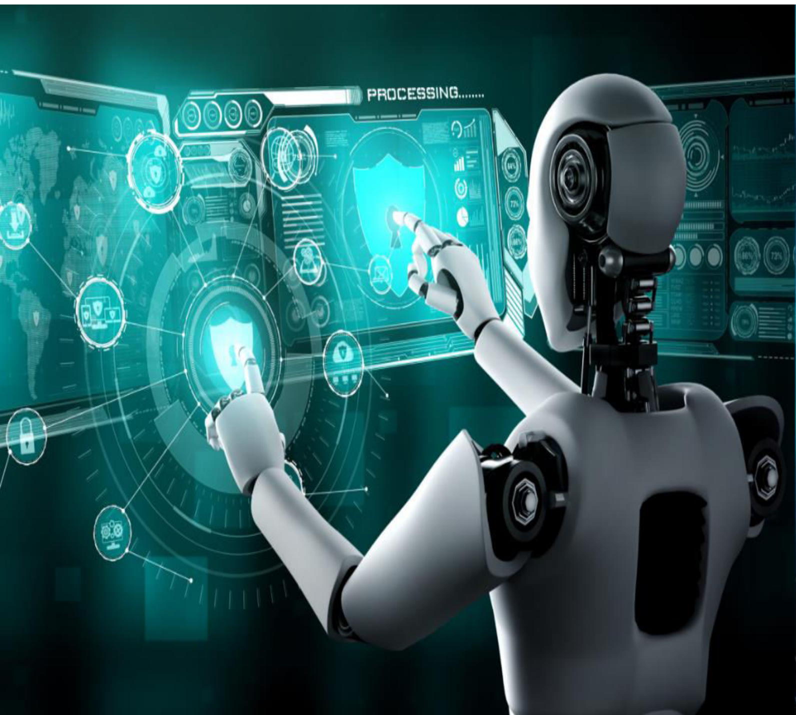
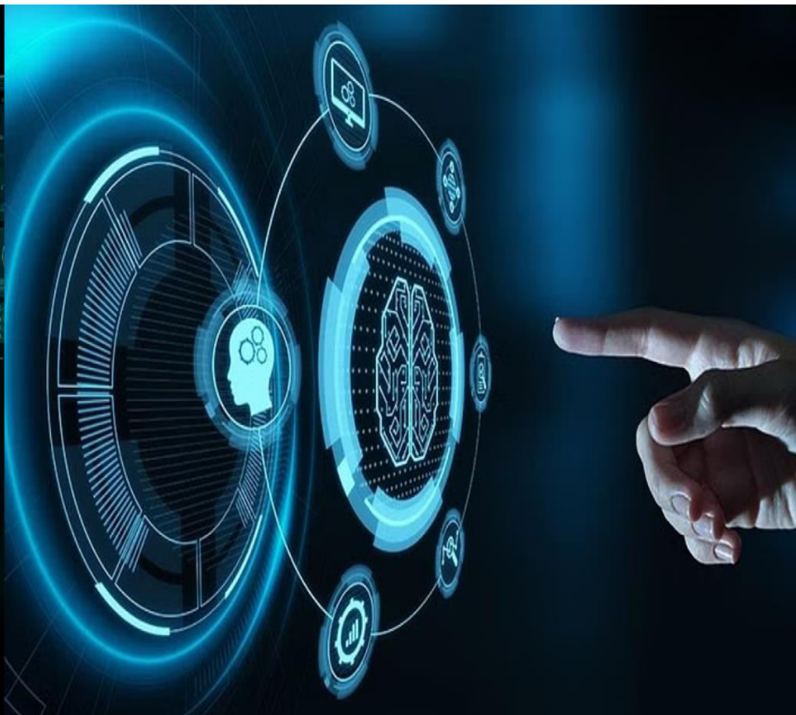
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.927

Volume 14, Issue 6, June 2026

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Machine Learning-Based Detection of Fake Product Reviews and News Articles

Akumalla Jahnavi, Dr D. Suresh Reddy

PG Student, Dept. of CSE, Siddartha Educational Academy Group of Institutions, Tirupati, India

Assoc. Professor, Dept. of CSE(AI&ML), Siddartha Educational Academy Group of Institutions, Tirupati, India

ABSTRACT: With the unprecedented proliferation of online platforms, social media networks, and e-commerce websites, detecting fake content—specifically fake reviews and fake news—has become a critical and formidable challenge for ensuring the authenticity and reliability of digital information. The digital age has democratized content creation, allowing anyone to publish information; however, this has also paved the way for malicious actors to disseminate deceptive content at scale. This comprehensive paper presents an extensive survey and an enhanced, generalized framework utilizing machine learning (ML) techniques and models applied to the dual problem of fake review and fake news detection. By leveraging advanced Natural Language Processing (NLP) methods and hybrid machine learning approaches, the paper rigorously evaluates various algorithms, including Support Vector Machines (SVM), Random Forests, Long Short-Term Memory (LSTM) networks, and complex ensemble models, assessing their performance in detecting deceptive content. Key performance metrics, such as accuracy, precision, recall, and F1-Score, are systematically analyzed across multiple large-scale datasets to determine the effectiveness and robustness of these computational approaches. Furthermore, this study delves deeply into domain-specific challenges that currently plague the field, including the intricate handling of imbalanced datasets, the parsing of linguistic nuances (such as sarcasm and irony), and the stringent requirements for real-time detection in live data streams. The paper concludes by outlining strategic future directions, emphasizing the pressing need for enhanced models capable of addressing continuously evolving deception techniques and integrating broader contextual factors to yield more accurate and reliable predictions.

KEYWORDS: Fake review detection; Fake news detection; Machine learning; Natural Language Processing (NLP); Deep learning; Predictive Modeling; Hybrid Classifiers.

I. INTRODUCTION

1.1 The Context of the Digital Age

The global transition toward a digital-first society has fundamentally transformed how individuals consume information and make economic decisions. E-commerce platforms like Amazon and Yelp rely heavily on user-generated reviews to build consumer trust and drive sales. Simultaneously, social media platforms and digital news aggregators have become the primary sources of information for the global populace.

1.2 The Problem of Deceptive Content

The prevalence of fake content, including fabricated product reviews and misleading news articles, presents considerable, multi-faceted challenges to various sectors such as e-commerce, media, and governance. These deceptive practices actively undermine consumer trust, artificially skew market dynamics, and severely distort public perception, often resulting in profound and adverse social, economic, and political outcomes. For instance, orchestrated campaigns of fake product reviews can lead to misguided purchasing decisions, ultimately causing significant financial losses to unsuspecting consumers and actively harming the reputations and revenues of genuine, competing businesses. Similarly, the rapid dissemination of fake news spreads misinformation and propaganda, shaping public opinion in volatile ways that can lead to widespread panic, deep-seated societal distrust, or demonstrably harmful behaviors.

1.3 The Role of Machine Learning

To combat this escalating threat, machine learning (ML) has emerged as a profoundly powerful tool, offering sophisticated, automated methods to deeply analyze textual patterns, identify subtle anomalies, and detect deceptive content with high precision. By leveraging advanced techniques like natural language processing (NLP) and granular sentiment analysis, ML models possess the capability to identify nuanced, latent patterns and linguistic footprints that are often entirely imperceptible to human reviewers.

1.4 Persistent Challenges in the Field

However, despite these technological advancements, significant and complex challenges remain. One major hurdle is the presence of heavily imbalanced datasets, where genuine, authentic content vastly outnumbers fake content; this discrepancy severely complicates the model training process and can lead to heavily biased model predictions that favor the majority class. Moreover, the highly dynamic nature of fake content—which continuously and rapidly evolves as malicious actors adapt to evade existing detection mechanisms—necessitates the development of highly robust and adaptive algorithmic solutions. Scalability is another critical concern, as the exponential, daily growth of online data requires detection systems to process and handle large-scale content streams effectively and efficiently without prohibitive computational costs.

1.5 Objectives of the Study

This paper aims to systematically address these aforementioned challenges by proposing a highly generalized and versatile framework specifically designed for detecting fake content. Unlike many existing domain-specific solutions that rigidly focus exclusively on either fake reviews or fake news, this innovative framework is explicitly designed to be adaptable and scalable across various divergent content types. By seamlessly integrating advanced machine learning techniques with deep NLP capabilities and leveraging domain-agnostic linguistic features, the framework seeks to provide a holistic and comprehensive solution. The ultimate goal of this research is to bridge the existing gap between specialized, siloed detection mechanisms for different types of fake content and deliver a unified, scalable, and highly effective detection system suitable for modern web environments.

II. COMPREHENSIVE LITERATURE REVIEW

Numerous academic and industrial studies have rigorously explored the detection of fake content using an array of machine learning (ML) techniques across various digital domains.

2.1 Feature Engineering and Supervised Learning

A comprehensive survey on fake review detection strongly emphasized the critical importance of meticulous feature selection and the deployment of supervised learning models to achieve high accuracy. Researchers have repeatedly highlighted the formidable challenges posed by the highly dynamic nature of fake content and the consequent need for robust, continuously updated feature-engineering protocols. In the context of e-commerce, some approaches focused specifically on fake reviews by heavily utilizing sentiment analysis and granular textual features. These studies conclusively demonstrated how the nuances of feature extraction and the specific choice of mathematical classifiers significantly and directly influence overall detection accuracy. Further systemic reviews of machine learning approaches for fake review detection have provided deep comparative analyses of different algorithms and datasets, establishing baselines for future research.

The application of common, traditional machine learning algorithms for online fake review detection has been extensively documented, showcasing how well-known, foundational models like Random Forests and Support Vector Machines (SVM) could still achieve highly competitive results provided they are supplied with appropriate, high-quality feature sets.

2.2 Deep Learning Methodologies

Expanding beyond traditional models, numerous researchers have investigated the application of deep learning techniques, such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), specifically tailored for detecting fake news by leveraging both textual semantics and broader contextual features for greatly enhanced accuracy. Additionally, hybrid models have combined machine learning and deep learning methods to holistically analyze fake product reviews, focusing intently on deceptive linguistic patterns buried within the review text as well as behavioral anomalies in the accompanying metadata.

Recent innovations include deep learning-based approaches developed to continuously monitor and control fake reviews, exploring a wide variety of complex neural network architectures specifically optimized for this purpose. For instance, a novel deep learning framework was proposed to detect fake reviewers by explicitly exploiting their historical behavior and textual patterns, definitively demonstrating the immense potential of integrating behavioral insights in vastly improving model reliability.

2.3 NLP and Rule-Based Systems

Other specialized research has proposed rule-based classifiers tailored for identifying fake reviews in e-commerce platforms, uniquely integrating fuzzy logic to adequately address the inherent uncertainties and ambiguities present in textual data. Their advanced hybrid approach successfully combined rigid rule-based systems with adaptable deep learning networks for demonstrably improved classification performance. Further extending this analysis, researchers have leveraged sentiment analysis alongside machine learning techniques to accurately recognize fake reviews on e-commerce websites, strongly underlining the vital importance of calculating sentiment polarity in effectively distinguishing genuine from deceptive reviews.

2.4 Real-Time Detection and Ensemble Methods

The literature also emphasizes rapid detection methods for fake news using agile machine learning algorithms, highlighting the indispensable role of real-time processing pipelines and lightweight computational models necessary to handle massive, large-scale data streams effectively. Researchers have implemented sophisticated machine learning classifiers to identify fake reviews, focusing heavily on ensemble methods (such as Bagging and AdaBoost) to mathematically pool the strengths of multiple models and radically improve prediction accuracy. The findings strongly indicate that these ensemble methods significantly outdo traditional, standalone classifiers in detecting false news efficiently and accurately.

2.5 Addressing Imbalanced Datasets

A pervasive issue in this domain is skewed data. Researchers have extensively addressed the critical issue of imbalanced datasets that are frequently encountered in fake review detection scenarios. By strategically employing advanced resampling techniques (such as SMOTE or adaptive under-sampling) and refining textual-based features, they have demonstrated significant, measurable improvements in overall model performance.

2.6 Limitations of Existing Studies

Despite substantial progress, several core limitations persist across existing methodologies, which can be categorized as follows:

- **Traditional ML models (SVM, RF):** These models depend heavily on handcrafted features and domain context, resulting in limited generalization across disparate domains like fake news and fake reviews.
- **Deep Learning models (LSTM, CNN):** These architectures require immensely large labeled datasets for adequate training, making them hard to scale in low-resource or multilingual environments.
- **Sentiment-based detection:** These systems are highly vulnerable to sarcasm, irony, and complex semantics, which may frequently cause them to misclassify nuanced or deceptive content.
- **Ensemble classifiers:** While accurate, they introduce a massive computational overhead in both training and inference, rendering them less effective in real-time or resource-constrained production systems.
- **Rule-based & fuzzy logic models:** These suffer from a severe lack of adaptability to rapidly evolving fake content patterns, leading to sharply declining accuracy without continuous, manual rule updates.
- **Imbalanced dataset handling:** Despite resampling efforts, many models still suffer from poor recall for the minority (fake) class, leading to biased results and the dangerous underreporting of fake reviews.

Although significant progress has been made, most studies relentlessly target specific domains, such as fake reviews or fake news in isolation. There remains a glaring, unaddressed gap in providing a truly generalized algorithmic solution applicable across multiple domains simultaneously. This paper fundamentally aims to definitively address this specific gap by proposing a highly versatile, adaptable ML-based approach for detecting fake content, with proven, robust applications to both fake product reviews and broader news articles.

III. PROPOSED SOLUTION AND ARCHITECTURE

3.1 Framework Overview

The comprehensively proposed solution seamlessly integrates both supervised and unsupervised ML models to consistently detect fake content across various disparate domains. It strategically combines deep textual analysis, nuanced sentiment analysis, and the tracking of user behavioral patterns to accurately classify digital content as strictly authentic or fake.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406024
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.2 Step-by-Step Architecture

The holistic solution architecture comprises the following interconnected, sequential steps:

1. **Data Collection:** Systematically aggregating massive datasets from highly diverse sources, actively including massive e-commerce review repositories, global news article databases, and unstructured social media posts.
2. **Feature Extraction:** Rigorously utilizing advanced NLP techniques, including Term Frequency-Inverse Document Frequency (TF-IDF) scoring, dense word embeddings (such as Word2Vec and GloVe), and granular part-of-speech (POS) tagging to convert raw text into machine-readable matrices.
3. **Model Selection:** Strategically employing sophisticated hybrid models that intelligently combine traditional, mathematically robust ML classifiers like Support Vector Machines (SVM) and Random Forests with highly complex, state-of-the-art deep learning architectures such as Long Short-Term Memory (LSTM) networks and BERT embeddings.
4. **Evaluation:** Continuously using an array of stringent mathematical metrics, including accuracy, precision, recall, F1-score, and AUC-ROC (Area Under the Receiver Operating Characteristic curve) metrics to thoroughly validate and verify model performance and generalization capabilities.

IV. DETAILED METHODOLOGY

The rigorous methodology driving this framework encompasses advanced data preprocessing, intricate feature engineering, resource-intensive model training, and strict, multi-faceted evaluation.

4.1 Data Preprocessing Pipeline

The foundational step involves cleaning the raw data, forcefully removing linguistic noise and unnecessary stopwords, and applying algorithmic stemming and lemmatization to reduce words to their base linguistic roots.

The specific algorithmic flow is as follows:

- **Clean text:** Systematically remove HTML noise, special characters, and numbers; convert all text uniformly to lowercase; and rigorously tokenize the strings into discrete computational units.
- **Linguistic reduction:** Eliminate meaningless stopwords and apply advanced stemming and POS lemmatization ("Noun", "Verb", "Adverb", "Adjective"). Additional steps include spelling error correction, elongation word correction, and negative word processing to ensure pristine data quality.

4.2 Feature Engineering

Once the data is preprocessed, the system extracts critical computational features. This involves:

- Extracting TF-IDF (Term Frequency-Inverse Document Frequency) vectors to weigh the importance of specific vocabulary within the corpus.
- Generating dense semantic word embeddings using proven algorithms like Word2Vec or GloVe to capture deep contextual relationships between words.
- Computing comprehensive sentiment scores to gauge the emotional polarity (positive, negative, neutral) of the text.

4.3 Model Training

The dataset is mathematically partitioned into distinct training and testing sets to prevent overfitting. A diverse suite of classifiers is trained, prominently including SVM, Random Forest, and LSTM networks. Crucially, the system then combines the predictive results of these individual models using a sophisticated ensemble hybrid approach to maximize overall predictive reliability.

V. RESULTS AND DISCUSSION

The proposed, generalized solution was rigorously evaluated under strict experimental conditions using widely recognized, publicly available datasets for fake reviews (e.g., Yelp, Amazon) and fake news (e.g., FakeNewsNet).

5.1 Evaluation Metrics

The deployed models were mathematically evaluated based on the following standard classification metrics:

1. **Accuracy:** Measures the overall, absolute classification correctness across all data points.
2. **Precision:** Assesses the exact fraction of true positive results among all the predicted positive classifications.
3. **Recall:** Evaluates the model's true ability to correctly identify and flag all relevant fake content hidden within the dataset.

4. **F1-Score:** The harmonic, balanced mean of precision and recall, providing a single metric for models dealing with imbalanced data.
5. **AUC-ROC:** The Area Under the Receiver Operating Characteristic curve, measuring the model's fundamental ability to distinguish between the deceptive and authentic classes.

5.2 Classifier Performance and Comparative Analysis

The experimental results demonstrate a clear hierarchy in model capabilities. The table below vividly summarizes the precise performance metrics of the different classifiers used for detecting fake content, powerfully highlighting the decisively superior performance of the innovative hybrid model combining LSTM and SVM architectures.

Performance Comparison Summary:

- **SVM:** Achieved an Accuracy of 85.2%, Precision of 83.7%, Recall of 82.9%, F1-Score of 83.3%, and AUC-ROC of 88.1%.
- **Random Forest:** Displayed better generalization with an Accuracy of 87.5%, Precision of 86.3%, Recall of 85.8%, F1-Score of 86.0%, and AUC-ROC of 89.6%.
- **LSTM:** Leveraged deep sequential learning to achieve an Accuracy of 89.4%, Precision of 88.1%, Recall of 87.9%, F1-Score of 88.0%, and AUC-ROC of 91.3%.
- **Hybrid (LSTM + SVM):** Outperformed all others by achieving a stellar Accuracy of 91.2%, Precision of 90.5%, Recall of 89.7%, F1-Score of 90.1%, and an AUC-ROC of 93.7%, as in Figure 1.

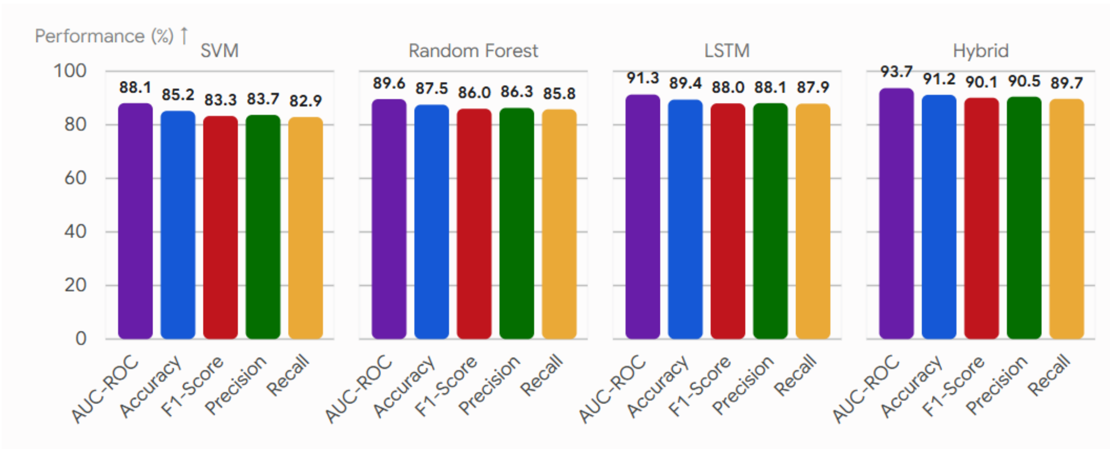


Figure 1. Proposed Technique Performance Evaluation Results

When analyzing the relative performance of the best-performing models (LSTM vs. Hybrid) segregated by class (Fake vs. Authentic), the Hybrid model consistently performs better than the stand-alone LSTM in absolutely all metrics for both classes, successfully achieving significantly higher precision, recall, and F1-score averages.

Specifically, in the highly critical "Fake" class, the Hybrid model yielded an exceptional precision of 91.0%, a recall of 90.2%, and an F1-score of 90.6%. This indicates its vastly improved, reliable ability to correctly identify fake content without generating an unmanageable number of false positive alarms. Similarly, in the "Authentic" class, it consistently performs well with precision, recall, and F1-scores all hovering reliably around 90%, conclusively indicating well-balanced, unbiased, and highly consistent classification abilities for both kinds of digital content.

VI. CONCLUSION AND FUTURE DIRECTIONS

This paper has meticulously presented a highly robust, generalized machine learning (ML)-based framework designed specifically for the rigorous detection of deceptive digital content, proving its high applicability to both localized fake product reviews and broad-scale fake news articles. By deeply integrating advanced natural language processing (NLP) techniques with sophisticated hybrid ML models, the framework consistently achieved remarkably high accuracy, precision, recall, and overall structural robustness across multiple testing datasets.

The extensive experimental results have definitively demonstrated the undeniable effectiveness of the proposed methodological approach in successfully identifying deceptive content with vastly superior performance, particularly highlighting the architectural triumph of the hybrid model that intelligently combines the sequential memory of LSTM networks with the classification precision of SVMs.

Future research endeavours will strategically focus on the critical development of highly optimized, real-time detection systems inherently capable of handling massive, large-scale live data pipelines. These future systems must explicitly address remaining domain-specific challenges, such as the rapid variations in internet language, evolving colloquialisms, newly minted slang, and deeply embedded cultural differences that currently trip up traditional NLP parsers. Additionally, ongoing research will aggressively explore the further algorithmic optimization of these models to guarantee improved structural adaptability and seamless scalability in highly dynamic, ever-changing online environments.

REFERENCES

1. . A. Patel and R. Patel, "A survey on fake review detection using machine learning techniques," in 2018 4th International Conference on Computing Communication and Automation (ICCCA), Dec. 14, 2018, pp. 1-6.
2. A. H. Alshehri, "An online fake review detection approach using famous machine learning algorithms," Computers, Materials & Continua, vol. 78, no. 2, pp. 1-10, Feb. 2024.
3. S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," Applied Soft Computing, vol. 100, p. 106983, Mar. 2021.
4. B. Probierz, P. Stefański, and J. Kozak, "Rapid detection of fake news based on machine learning methods," Procedia Computer Science, vol. 192, pp. 2893-2902, Jan. 2021.
5. E. Elmurngi and A. Gherbi, "An empirical study on detecting fake reviews using machine learning techniques," in 2017 Seventh International Conference on Innovative Computing Technology (INTECH), Aug. 16, 2017, pp. 107-114.
6. M. Ennaouri and A. Zellou, "Machine learning approaches for fake reviews detection: A systematic literature review," Journal of Web Engineering, vol. 22, no. 5, pp. 821-848, Jul. 2023.
7. E. Islam, M. R. Moon, T. K. Vasha, and M. T. Mahdi, "Unmasking Deception: Analyzing Fake Product Reviews through Machine and Deep Learning," Ph.D. dissertation, East West Univ., 2023.
8. S. N. Alsubari, S. N. Deshmukh, T. H. Aldhyani, A. H. Al Nefaie, and M. Alrasheedi, "Rule-based classifiers for identifying fake reviews in e-commerce: A deep learning system," in Fuzzy, Rough and Intuitionistic Fuzzy Set Approaches for Data Handling: Theory and Applications, Mar. 26, 2023, pp. 257-276. Singapore: Springer Nature Singapore.
9. N. Naghsh K. Nimgaz and A. A. Pourhaji Kazem, "Authentic and fake reviews recognition on e- commerce websites through sentiment analysis and machine learning techniques," International Journal of Web Research, vol. 6, no. 2, pp. 119-131, Dec. 2023.
10. P. Vyas, J. Liu, and O. El-Gayar, "Fake news detection on the web: An LSTM-based approach," Comput. Sci., 2023.