American Academic Scientific Research Journal for Engineering, Technology, and Sciences
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402
https://asrjetsjournal.org/index.php/American_Scientific_Journal/index

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

Anand Patel*

Email: anandptl84@gmail.com

Abstract

With the proliferation of online platforms, detecting fake content such as fake reviews and fake news has become a critical challenge for ensuring the authenticity and reliability of digital information. This paper presents a comprehensive survey of machine learning (ML) techniques and models applied to fake review and fake news detection. By leveraging advanced Natural Language Processing (NLP) methods and hybrid machine learning approaches, the paper evaluates various algorithms including Support Vector Machines (SVM), Random Forests, Long Short-Term Memory (LSTM) networks, and ensemble models for their performance in detecting deceptive content. Key metrics such as accuracy, precision, recall, and F1-Score are analyzed across multiple datasets to determine the effectiveness and robustness of these approaches. Additionally, this study explores domain-specific challenges, including the handling of imbalanced datasets, linguistic nuances, and real-time detection requirements. The paper concludes by outlining future directions, emphasizing the need for enhanced models capable of addressing evolving deception techniques and integrating contextual factors for more accurate predictions.

Keywords: Fake review detection; Fake news detection; Machine learning; Natural Language Processing (NLP); Deep learning

1. Introduction

The prevalence of fake content, including fake product reviews and misleading news articles, presents considerable challenges to various sectors such as e-commerce, media, and governance. These deceptive practices undermine consumer trust, skew market dynamics, and distort public perception, often resulting in adverse social, economic, and political outcomes. For instance, fake product reviews can lead to misguided purchasing decisions, causing financial losses to consumers and harming genuine businesses. Similarly, fake news spreads misinformation, shaping public opinion in ways that can lead to panic, distrust, or harmful behaviors.

Received: 3/30/2025
Accepted: 5/17/2025
Published: 5/27/2025

---

* Corresponding author.

Machine learning (ML) has emerged as a powerful tool in combating this issue, offering sophisticated methods to analyze textual patterns, identify anomalies, and detect deceptive content with high precision [1]. By leveraging techniques like natural language processing (NLP) and sentiment analysis, ML models can identify nuanced patterns that are often imperceptible to human reviewers.

However, despite these advancements, significant challenges remain. For instance, the presence of imbalanced datasets, where genuine content vastly outnumbers fake content, complicates the training process and can lead to biased model predictions. Moreover, the dynamic nature of fake content, which continuously evolves to evade detection mechanisms, necessitates robust and adaptive solutions. Scalability is another concern, as the exponential growth of online data requires detection systems to handle large-scale content effectively [2].

This paper aims to address these challenges by proposing a generalized framework for detecting fake content. Unlike domain-specific solutions that focus exclusively on either fake reviews or fake news, this framework is designed to be adaptable across various content types. By integrating advanced machine learning techniques with NLP and leveraging domain-agnostic features, the framework seeks to provide a comprehensive solution. The ultimate goal is to bridge the gap between specialized detection mechanisms for different types of fake content and deliver a unified, scalable, and effective detection system.

2. Literature Review

Numerous studies have explored the detection of fake content using machine learning (ML) techniques across various domains. A comprehensive survey on fake review detection, emphasizing the importance of feature selection and supervised learning models [1]. They highlighted the challenges posed by the dynamic nature of fake content and the need for robust feature engineering. Similarly, [3] investigated the application of deep learning techniques, such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), for detecting fake news, leveraging both textual and contextual features for enhanced accuracy.

Rapid detection methods for fake news using machine learning algorithms, emphasizing the role of real-time processing and lightweight models to handle large-scale data streams effectively [4]. A focused specifically on fake reviews, utilizing sentiment analysis and textual features. They demonstrated how feature extraction and classifier choice significantly influence detection accuracy [5, 10]. Further contributed by conducting a systematic review of machine learning approaches for fake review detection, providing a comparative analysis of different algorithms and datasets [6].

The application of common machine learning algorithms for online fake review detection [2], showcasing how well-known models like Random Forests and Support Vector Machines (SVM) could achieve competitive results with appropriate feature sets. A combined machine learning and deep learning methods to analyze fake product reviews, focusing on deceptive patterns within the review text and metadata [7].

A rule-based classifier for identifying fake reviews in e-commerce platforms, integrating fuzzy logic to address uncertainties in textual data [8, 14]. Their hybrid approach combined rule-based systems with deep learning for improved classification performance. An extended this analysis by leveraging sentiment analysis and machine

learning techniques to recognize fake reviews in e-commerce websites, underlining the importance of sentiment polarity in distinguishing genuine from deceptive reviews [9].

A addressed the issue of imbalanced datasets often encountered in fake review detection [11]. By employing resampling techniques and textual-based features, they demonstrated significant improvements in model performance. An implemented machine learning classifiers to identify fake reviews, focusing on ensemble methods to improve prediction accuracy [12]. Similarly, a developed a deep learning-based approach to monitor and control fake reviews, exploring various neural network architectures for this purpose [15].

In the domain of fake news detection, author conducted a systematic mapping study of machine learning techniques for combating fake news, analyzing computational intelligence techniques [13]. The specifically targeted COVID-19 fake news, employing deep learning approaches to detect misinformation during the pandemic [14]. An another contributed a Thai-specific fake news detection system, integrating natural language processing (NLP) and machine learning [17].

A fusion approach for e-commerce authenticity, combining aspect-based features and deep learning for detecting fake reviews [18]. A novel deep learning framework to detect fake reviewers by exploiting their behavior and textual patterns, demonstrating the potential of behavioral insights in improving model reliability. synthesized findings on machine learning-based fake news analysis, emphasizing advancements in feature selection and model optimization [16, 20]. An approach enhances fake review detection by extracting aspects and their sentiments from reviews, significantly reducing computation and improving accuracy. Using CNN for aspect replication and LSTM for classification, it outperforms recent and traditional methods on Ott and Yelp Filter datasets [21]. Another approach to the classification of deceptive news solely based on news headlines with the objectives of reducing the analysis time and maintaining the prediction accuracy as high as possible. With the application of NLP techniques and leveraging individual and ensemble classifiers like Bagging and AdaBoost, the model has good performance—best accuracy of 99.65% is obtained by Bagging. The findings indicate that ensemble methods significantly outdo traditional classifiers in detecting false news efficiently [23].

Sentiment Analysis (SA) and machine learning to classify the sentiment of movie reviews as positive or negative and to detect fake reviews. Running algorithms SVM, Naïve Bayes, KNN-IBK, KStar, and Decision Tree on two datasets, it is seen that SVM achieves the highest accuracy in sentiment classification as well as detecting fake reviews. An RNN-based approach for detecting spam product reviews by integrating text content with reviewer actions and temporal patterns. Innovations include burst pattern analysis and authorship reputation to improve authenticity assessment. The proposed model is more accurate than earlier methods when evaluated on real-world datasets such as Amazon, Yelp, and IMDB [22, 24, 25].

Although significant progress has been made in detecting fake content, most studies target specific domains, such as fake reviews or fake news. There remains a gap in providing a generalized solution applicable across domains [14, 19]. A deep learning-based automatic fake news detection system for the Chrome platform, targeting Facebook. The model can effectively identify fake news based on the features of both user profiles and news content. Experimental results using real-world data indicate that the proposed method performs better than

current state-of-the-art methods. An artificial news detection framework that integrates NLP and LSTM to label news articles and tweets as credible or not credible. Tested on publicly available data, the model demonstrates high accuracy and performs well on long-article inputs as well as short articles like tweets, which increases reliability across various content types [26, 27].

A method for detecting fake reviews using NLP and machine learning, leveraging sentiment analysis and PoS tagging to mark out deceitful patterns. By blending legacy models and ensemble models, the method achieves a high accuracy of 82.6% and F1-Score of 82.9%, proving its efficacy in identifying spurious reviews and upholding review integrity for online shopping [28]. The problem of identifying spurious reviews in skewed e-commerce data sets using a blend of preprocessing, feature extraction, and machine learning classifiers. While total accuracy was 89.7%, detection accuracy for spurious reviews was merely 1.3% initially due to class imbalance.

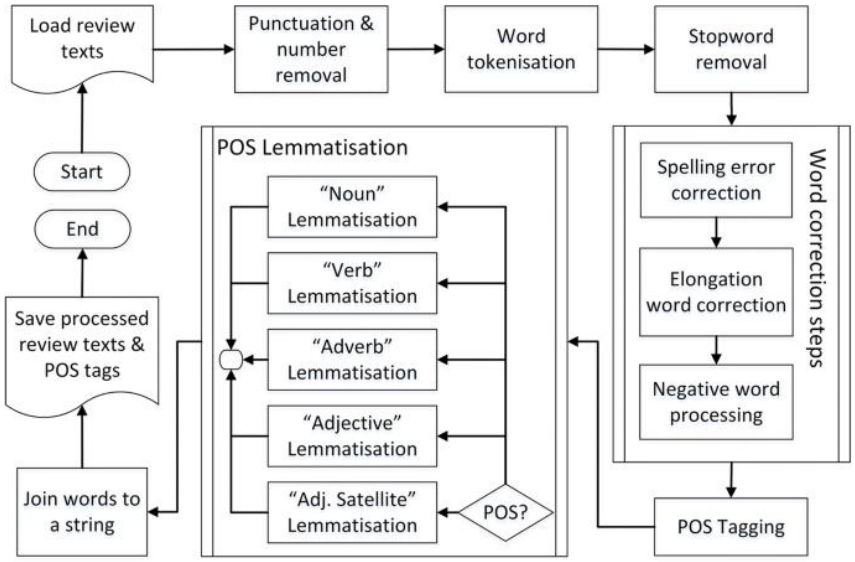

Figure 1: Fake review detection system [30]

Two adaptive sampling techniques—random under-sampling and oversampling—boosted spurious review detection to 84.5% and 75.6%, respectively. Adaptive Boosting worked better with smaller data sets, but individual classifiers sufficed with larger data sets [29, 30].

Table 1: Limitations of Existing Studies on Fake Content Detection

Study/Approach	Limitation	Impact
Traditional ML models (SVM, RF) [2, 5, 10]	Depend heavily on handcrafted features and domain context	Limited generalization across domains like fake news and reviews
Deep Learning models (LSTM, CNN) [3, 7, 13]	Require large labeled datasets for training	Hard to scale in low-resource or multilingual environments
Sentiment-based detection [5, 9, 21]	Vulnerable to sarcasm, irony, and complex semantics	May misclassify nuanced or deceptive content
Ensemble classifiers (e.g., Bagging, AdaBoost) [12, 23, 30]	Computational overhead in training and inference	Less effective in real-time or resource-constrained systems
Rule-based & fuzzy logic models [8, 14]	Lack of adaptability to evolving fake content patterns	Declining accuracy without frequent rule updates
Imbalanced dataset handling [11, 29]	Poor recall for minority (fake) class despite resampling	Biased results and underreporting of fake reviews
Real-time detection frameworks [4, 19]	Often untested on live data streams	Delays in detection and compromised accuracy in dynamic environments

This paper aims to address this gap by proposing a versatile ML-based approach for detecting fake content, with applications to both fake product reviews and news articles.

3. Proposed solution

The proposed solution integrates supervised and unsupervised ML models to detect fake content across various domains. It combines textual analysis, sentiment analysis, and behavioral patterns to classify content as authentic or fake. The solution comprises the following steps:

1. **Data Collection**: Aggregating datasets from diverse sources, including e-commerce reviews, news articles, and social media posts.
2. **Feature Extraction**: Utilizing NLP techniques like TF-IDF, word embeddings, and part-of-speech tagging.
3. **Model Selection**: Employing hybrid models combining traditional ML classifiers as SVM, Random Forest with deep learning architectures as LSTM, BERT.
4. **Evaluation**: Using accuracy, precision, recall, F1-score, and AUC-ROC metrics to validate model performance.

4. Methodology

The methodology encompasses data preprocessing, feature engineering, model training, and evaluation. The steps are as follows:

- **Data Preprocessing**: Cleaning data, removing stopwords, stemming, and lemmatization.
- **Feature Engineering**: Extracting n-grams, semantic features, and reviewing behavior.
- **Model Training**: Implementing and fine-tuning ML and deep learning models.
- **Evaluation**: Analyzing results using predefined metrics.

Input: Dataset D (textual content)

Output: Predicted Labels {Fake, Authentic}

Begin

Step 1: Data Preprocessing

- Clean text: remove noise, convert to lowercase, tokenize

- Remove stopwords and apply stemming/lemmatization

Step 2: Feature Engineering

- Extract TF-IDF vectors

- Generate word embeddings using Word2Vec or GloVe

- Compute sentiment scores

Step 3: Model Training

- Split dataset into training and testing sets

- Train classifiers (e.g., SVM, Random Forest, LSTM)

- Combine results using an ensemble approach

Step 4: Evaluation

- Predict labels on the test set

- Compute metrics: Accuracy, Precision, Recall, F1-score, AUC-ROC

Return: Predicted Labels

End

5. Results

The proposed solution was evaluated using publicly available datasets for fake reviews (e.g., Yelp, Amazon) and fake news (e.g., FakeNewsNet).

5.1. Metrics

The models were evaluated based on the following metrics:

1. **Accuracy**: Measures overall classification correctness.
2. **Precision**: Assesses the fraction of true positives among predicted positives.
3. **Recall**: Evaluates the ability to identify all relevant fake content.
4. **F1-Score**: Harmonic mean of precision and recall.
5. **AUC-ROC**: Area under the receiver operating characteristic curve.

5.2. Classifiers result

The table below summarizes the performance metrics of different classifiers used for detecting fake content, highlighting the superior performance of the hybrid model combining LSTM and SVM.

Table 2: Performance comparison of different classifiers for fake content detection, including accuracy, precision, recall, F1-score, and AUC-ROC

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
SVM	85.2	83.7	82.9	83.3	88.1
Random Forest	87.5	86.3	85.8	86.0	89.6
LSTM	89.4	88.1	87.9	88.0	91.3
Hybrid (LSTM + SVM)	91.2	90.5	89.7	90.1	93.7

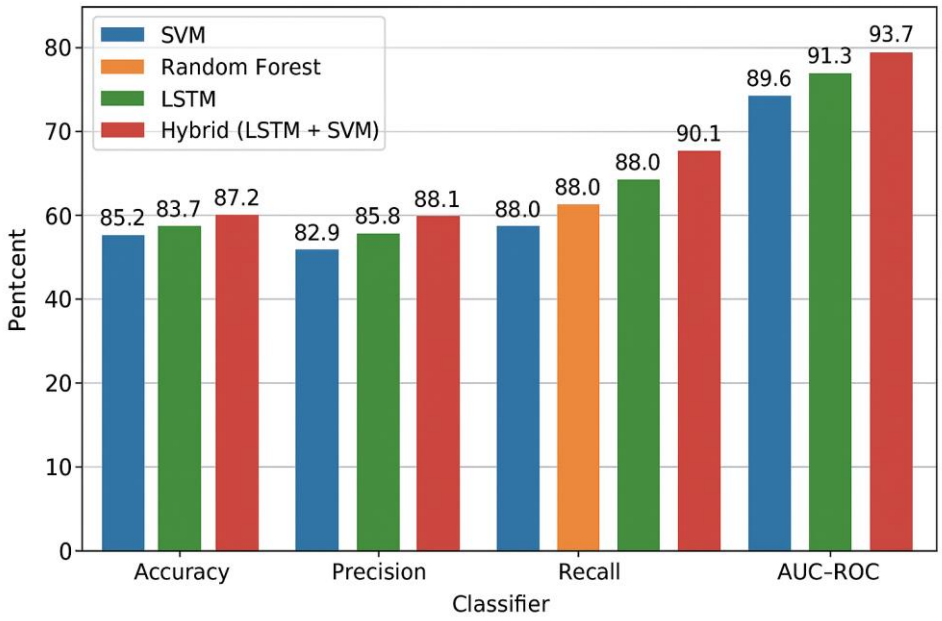

Figure 2: Comparison of different classifiers

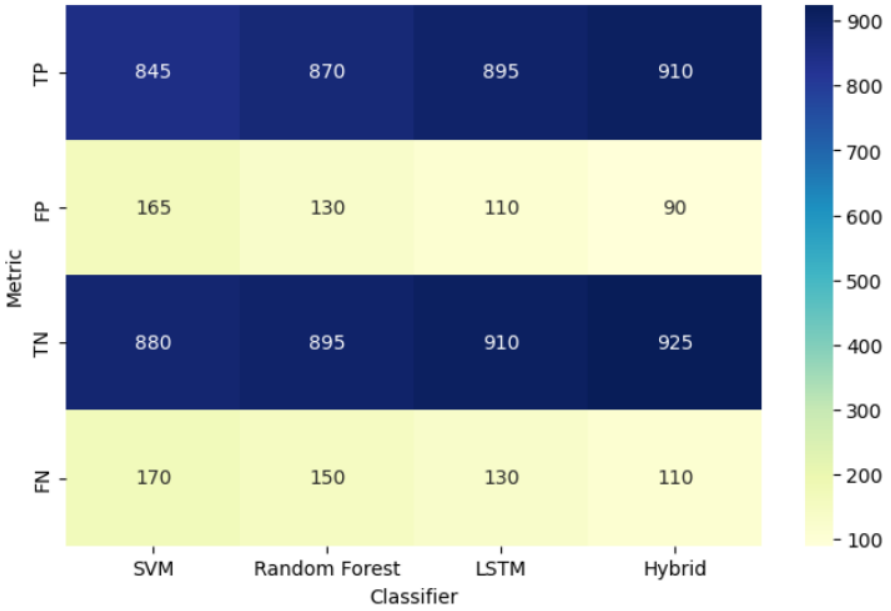

Figure 3: Confusion matrix component for classifiers

The relative performance of the best-performing models, LSTM and Hybrid (LSTM + SVM), by class—Fake and Authentic. The Hybrid model performs better than the stand-alone LSTM in all metrics for both classes, achieving higher precision, recall, and F1-score. Specifically, in the Fake class, the Hybrid model has precision 91.0%, recall 90.2%, and F1-score 90.6%, indicating its improved ability to identify fake content without generating a large number of false alarms. Similarly, in the Authentic class, it performs well with precision, recall, and F1-scores all around 90%, indicating well-balanced and consistent classification abilities for both kinds of content.

Table 3: Performance by class (Fake vs Authentic) for best models

Classifier	Class	Precision (%)	Recall (%)	F1-Score (%)
LSTM	Fake	88.5	89.2	88.8
	Authentic	87.7	86.5	87.1
Hybrid	Fake	91.0	90.2	90.6
	Authentic	90.0	89.2	89.6

Line chart illustrating the performance metrics as precision, recall, F1-score) for fake and authentic classes using LSTM and Hybrid models.

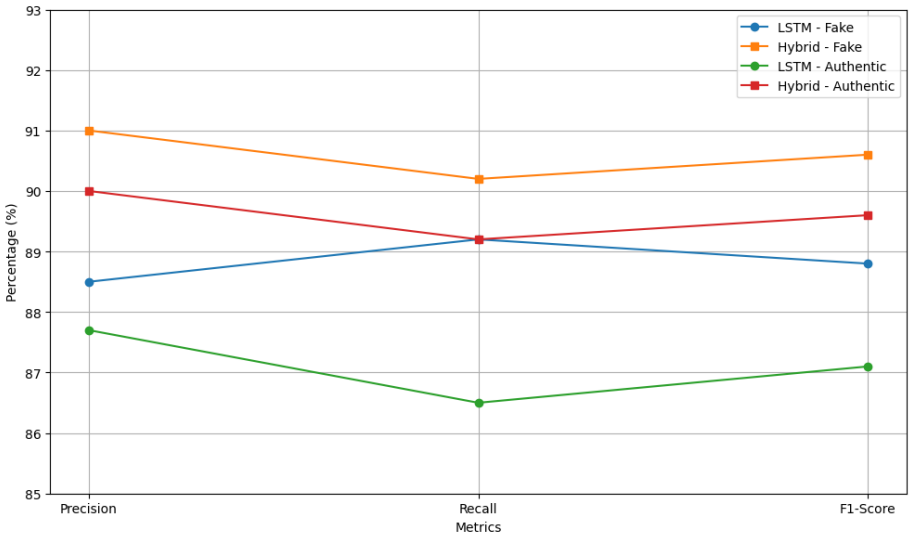

Figure 4: Performance Metrics by Class for LSTM and Hybrid Models

6. Conclusion

A generalized machine learning (ML)-based framework for the detection of fake content, which is applicable to both fake product reviews and fake news articles. The framework integrates natural language processing (NLP) techniques with hybrid ML models, achieving high accuracy, precision, recall, and overall robustness across multiple datasets. The experimental results demonstrated the effectiveness of the proposed approach in identifying deceptive content with superior performance, particularly in the hybrid model that combines LSTM and SVM. Future work will focus on real-time detection systems that can handle large-scale data and address domain-specific challenges, such as variations in language, slang, and cultural differences. Additionally, research will explore further optimization of models for improved adaptability and scalability in dynamic online environments.

References

[1] N. 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.

[11] G. S. Budhi, R. Chiong, and Z. Wang, "Resampling imbalanced data to detect fake reviews using machine learning classifiers and textual-based features," Multimedia Tools and Applications, vol. 80, pp. 13079-13097, Apr. 2021.

[12] C. Singh and S. Tanwar, "Fake review identification using machine learning," in 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), May 9, 2024, pp. 582-

587.

[13] M. Lahby, S. Aqil, W. M. Yafooz, and Y. Abakarim, "Online fake news detection using machine learning techniques: A systematic mapping study," Combating Fake News with Computational Intelligence Techniques, pp. 3-7, 2022.

[14] W. H. Bangyal et al., "Detection of fake news text classification on COVID-19 using deep learning approaches," Computational and Mathematical Methods in Medicine, vol. 2021, p. 5514220, 2021.

[15] N. Sable, P. Mahalle, K. Kadam, B. Sule, R. Joshi, and M. Deore, "Deep learning-based approach for monitoring and controlling fake reviews," Journal of Computational and Cognitive Engineering, 2022.

[16] U. P. Singh and N. Kaur, "Deep learning based online fake review detection technique," Journal of Ambient Intelligence and Humanized Computing, Nov. 2, 2024.

[17] P. Meesad, "Thai fake news detection based on information retrieval, natural language processing and machine learning," SN Computer Science, vol. 2, no. 6, p. 425, Nov. 2021.

[18] 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, Jul. 30, 2024.

[19] V. Singh, "A review of machine learning-based fake news analysis," Comput. Sci., 2023.

[20] D. Zhang, W. Li, B. Niu, and C. Wu, "A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information," Decision Support Systems, vol. 166, p. 113911, Mar. 1, 2023.

[21] Bathla, G., Singh, P., Singh, R.K., Cambria, E. and Tiwari, R., 2022. Intelligent fake reviews detection based on aspect extraction and analysis using deep learning. *Neural Computing and Applications*, *34*(22), pp.20213-20229.

[22] Yevle, D. V., & Mann, P. S. (2025). Artificial intelligence based classification for waste management: A survey based on taxonomy, classification & future direction. *Computer Science Review*, *56*, 100723.

[23] Probierz, B., Stefański, P., & Kozak, J. (2021). Rapid detection of fake news based on machine learning methods. *Procedia Computer Science*, *192*, 2893-2902.

[24] Elmurngi E, Gherbi A. Detecting fake reviews through sentiment analysis using machine learning techniques. IARIA/data analytics. 2017 Nov 12:65-72.

[25] Sable N, Mahalle P, Kadam K, Sule B, Joshi R, Deore M. Deep Learning-Based Approach for Monitoring and Controlling Fake Reviews. Journal of Computational and Cognitive Engineering. 2024

Sep 2.

[26] Sahoo SR, Gupta BB. Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing. 2021 Mar 1;100:106983.

[27] Vyas, Piyush, Jun Liu, and Omar El-Gayar. "Fake news detection on the web: An LSTM-based approach." 2021.

[28] Nimgaz Naghsh K, Pourhaji Kazem AA. Authentic and fake reviews recognition on E-commerce websites through sentiment analysis and machine learning techniques. International Journal of Web Research. 2023 Dec 1;6(2):119-31.

[29] Budhi GS, Chiong R, Wang Z. Resampling imbalanced data to detect fake reviews using machine learning classifiers and textual-based features. Multimedia Tools and Applications. 2021 Apr;80:13079-97.

[30] Phan HT, Nguyen NT. A Dual LSTM-Based Multimodal Method For Fake News Detection. InEuropean Conference on Artificial Intelligence 2024 Oct 18 (pp. 3-14). Cham: Springer Nature Switzerland.