SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 7 | July 2025 - Peer Reviewed Journal

FAKE REVIEW IDENTIFICATION THROUGH TOPIC MODELING AND SUPERVISED LEARNING

[1]Md Sirajul Huque, [2]Lavanya Kumari.Ch, [3]Gundekar Ashwini
[1, 2, 3] *Assistant Professor, Department of Computer Science and Engineering, GNITC, Hyderabad, Telangana,India*

Article DOI: https://doi.org/10.36713/epra23255
DOI No: 10.36713/epra23255

ABSTRACT

*Online reviews significantly influence consumer purchasing decisions and serve as a vital source of public opinion on products and services. These reviews not only help customers make informed choices but also provide businesses with constructive feedback to improve their offerings. However, the increasing reliance on online reviews has led to the rise of opinion spam—intentionally deceptive reviews created to falsely promote or criticize a product for personal or competitive gain **[1]**. This manipulation undermines trust in digital platforms, making it essential to detect and filter out fake reviews. Since manually identifying deceptive content is both time-consuming and inefficient, automated methods powered by Natural Language Processing (NLP) and machine learning have become necessary **[2][3][4]**. NLP enables the extraction of key linguistic features from review texts, revealing subtle differences in language usage between authentic and fake reviews **[3][5]**. Fraudulent reviewers often use exaggerated or emotionally charged language to influence readers, creating distinguishable patterns in word choice and topic focus **[3][6]**. By leveraging these linguistic and structural differences, machine learning models such as Support Vector Machines, Random Forests, and neural networks can be trained to classify reviews as genuine or fake **[7][8][9][10]**. This study presents a model that utilizes such language-based cues to effectively distinguish deceptive content, ultimately enhancing the accuracy and performance of automated fake review detection systems **[2][4]**.*

KEYWORDS: *Fake Review Detection, Opinion Spam, Deceptive Reviews, Natural Language Processing (NLP), Machine Learning, Text Classification, Consumer Behavior*

1. INTRODUCTION

With the rapid advancement of technology, consumer behavior has significantly shifted toward online shopping. A growing number of individuals now rely heavily on reading and writing online reviews before making purchasing decisions. These reviews have become an influential factor in shaping consumer opinions and, consequently, in driving the success or failure of businesses. As online reviews play a vital role in both market reputation and economic performance, there has been a notable increase in opinion spamming—the practice of posting deceptive reviews to unfairly promote or damage a product or brand [1].

Such deceptive practices can have far-reaching impacts. While positive reviews can elevate a company's image and profitability, negative or fake reviews can lead to significant reputational damage and financial losses. In some cases, businesses even hire individuals to generate fake positive reviews for their products or negative ones targeting competitors [1][2]. This manipulation creates a need for effective fake review detection systems to preserve trust and provide accurate information to consumers.

Natural Language Processing (NLP), a field of artificial intelligence focused on enabling computers to understand and interpret human language, offers promising solutions for addressing this issue. NLP techniques, combined with machine learning (ML), can be employed to analyze review texts, extract meaningful patterns, and identify deceptive content [3][4]. Fake reviewers often use specific language styles or emotional expressions to influence readers, making linguistic and structural features valuable indicators for classification [2][5].

We have identified three common types of fake reviews:

- Untruthful reviews – fabricated opinions about a product.
- Misattributed reviews – content posted under incorrect product or brand categories.
- Non-reviews – irrelevant content such as advertisements or promotional links.

Among these, untruthful reviews pose the greatest threat as they directly compromise the integrity of the review ecosystem [1].

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 7 | July 2025 - Peer Reviewed Journal

Due to the subtle nature of deception, it is challenging to manually distinguish between real and fake reviews. As a result, automated detection systems have become essential. These systems often serve as part of broader decision support systems (DSS), which assist consumers and businesses in navigating dynamic and complex decision-making environments [6][7].

Current ML-based detection techniques can be broadly classified into supervised and unsupervised learning approaches. Additionally, they typically utilize three categories of features: behavioral, linguistic, or a hybrid of both. By analyzing user behavior patterns and textual characteristics, these systems aim to improve the accuracy and reliability of fake review detection in online platforms [1][6][8].

2. IDENTIFYING THE CHALLENGE AND AIM OF THE STUDY

With the rapid growth of online shopping and digital services, competition among businesses has intensified. In an effort to promote their own products or undermine their competitors, some businesses engage in unethical practices—such as hiring individuals or using fake accounts to post deceptive reviews [1][2]. These fabricated opinions can significantly influence consumer behavior and distort market dynamics.

From a financial perspective, the impact of such fake reviews can be substantial. While genuine, positive reviews help trustworthy companies gain recognition and boost sales, fake negative reviews can mislead potential buyers, causing them to avoid quality products. This can result in serious financial setbacks for businesses that actually offer superior services or products [2].

To protect the credibility of online reviews and ensure fair competition, it is essential to implement effective detection mechanisms that can identify and filter out fake opinions [3][4]. One such approach involves using machine learning techniques to analyze review content and detect deception [1][5].

Our proposed solution focuses specifically on the textual content of customer reviews. Individuals who write fake reviews often use emotionally charged language or specific keywords aimed at manipulating potential buyers [6][2]. By examining these unique patterns in word usage and topic selection, machine learning models can be trained to differentiate between authentic and deceptive reviews [1][3][5].

Accurate detection of fake reviews can lead to their automatic removal, ensuring that only trustworthy and factual opinions are visible to consumers. This not only strengthens consumer confidence but also supports ethical business practices, ultimately benefiting the entire digital marketplace [4].

The objective of this study is to develop a method that enhances the efficiency and automation of fake review detection by capturing the distinct characteristics found in genuine and deceptive reviews. The proposed approach emphasizes the analysis of textual content within user reviews, based on the assumption that individuals writing fake reviews often use unnatural or exaggerated language to sway readers [2][6].

These linguistic patterns, which tend to differ significantly from those found in truthful reviews, can be identified and grouped into unique thematic categories. To achieve this, the method utilizes the generative framework of Latent Dirichlet Allocation (LDA) to uncover hidden topics within the text [7]. These extracted topics are then used as key features to differentiate between authentic and fake reviews, providing a meaningful basis for classification and improving the overall accuracy of the detection process [1][5][7].

3. LITERATURE REVIEW: FAKE REVIEW DETECTION USING MACHINE LEARNING AND NLP

The proliferation of deceptive reviews on online platforms has created significant challenges in maintaining the trust and integrity of digital ecosystems. As a result, researchers have employed Machine Learning (ML) and Natural Language Processing (NLP) techniques to automatically detect such fake or misleading content. The literature on this topic can broadly be categorized into content-based, behavior-based, and hybrid approaches.

i. Content-Based Detection using NLP

Content-based approaches rely on the textual content of reviews, leveraging linguistic and semantic features to distinguish between genuine and deceptive texts.Shojaee et al. demonstrated the use of lexical and syntactic features for detecting deceptive reviews and found that combining them yielded better accuracy [1]. Yuan et al. reinforced this by showing improved results when merging multiple textual attributes [2].Zhang et al. introduced DCWord, a deep learning-based method using Word2Vec embeddings to capture contextual word semantics, enhancing the model's ability to distinguish fake reviews [3].Jain et al. explored pretrained deep neural networks (DNNs) using Google News corpus embeddings and achieved strong performance even in noisy datasets [4].Traditional methods such

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 7 | July 2025 - Peer Reviewed Journal

as TF-IDF, n-grams, and skip-grams also remain effective for extracting semantic relevance and word co-occurrence patterns, as shown in the work by Huque & Kumar [5].

ii. Behavior-Based Detection

Behavioral detection focuses on the metadata associated with reviewers, such as their posting patterns, emotional tone, and review volume.Lim et al. proposed detecting fake reviews by monitoring rating consistency, review frequency, and extreme score tendencies, helping identify unusual reviewer behavior [6].Liu et al. introduced temporal feature-based detection, where timestamps of reviews are analyzed to catch sudden bursts typical of fraudulent campaigns [7].The model by Huque & Kumar employed a multi-dimensional behavior scoring framework involving reviewer length, activity, rating variance, and emotional index to assess credibility [5].

iii. Hybrid and Ensemble-Based Models

Recent studies have explored combining text and behavior features to enhance accuracy.Huque & Kumar [5] integrated textual, behavioral, and anomaly-based features using PCA, Isolation Forest, and LOF for detecting deceptive reviews. They implemented ensemble feature selection methods like Information Gain, Chi-square, and XGBoost importance, demonstrating that such hybridization improves model robustness on imbalanced datasets.Their experiments with RUS and Borderline-SMOTE resampling techniques significantly enhanced classification metrics (AUC > 0.85, F1-Score > 0.89), outperforming traditional classifiers such as KNN, SVM, and Logistic Regression [5].

The comparative study on fake review detection techniques in summarized in the below Table 1

Comparative Study on Fake Review Detection Techniques

S. No.	Authors	Topic	Year	Website Covered	Tools & Technologies Used	Type of Fake Reviews Detected
1	Nitin Jindal & Bing Liu	Opinion Spamming & Analysis	2008	Amazon	- NLP (Logistic Regression) - Data Mining Techniques	- Untruthful Opinions - Brand Reviews - Non-reviews
2	Arjun Mukherjee, Bing Liu, Natalic Glance	Identification of Fake Review Groups in Customer Reviews	2012	Amazon	Behavioural & Relational Models	Spammer Groups
3	Andrej, Gregoe, Karosak	Enhancing Detection of Opinion Spam Groups	2013	Facebook & Twitter	Quantitative Psycholinguistic Text Analytics Tool	Spam Groups
4	Yuqing Lu, Zhang, Yudong Xiao, Yangguang	Detection of Fake Reviews and Reviewer Groups Simultaneously	2013	Amazon	Factor Graph Model	Fake Reviews & Spammers
5	Kuldeep Sharma, King-Ip Lin	Using Rating Consistency Check for Detection of Fake Reviews	2013	Amazon	- PHP Insight - Alchemy API	Fake Reviews
6	Snehashish Banerjee & Alton Y. K. Chua	Analysis of Authentic & Manipulative Reviews	2014	Amazon	Binomial Logistic Regression	Authentic & Manipulative Reviews

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 7 | July 2025 - Peer Reviewed Journal

7	Aakas Zhiyuli, Xun Liang, Yige Wang	Recognizing Deceptive Reviews	2015	Amazon	Sentiment Analysis on Product Attributes	Fake Reviews
8	Rupesh Kumar Dewang, A.K. Singh	Using Lexical & Syntactic Features to Identify Fake Reviews	2015	Amazon	- Naive Bayes - Decision Tree - WEKA Toolkit	Fake Reviews
9	Kyungyup Daniel Lee, Kyungah Han, SungHyon Myaeng	Capturing Word Choice Patterns of Fake and Truthful Reviews	2016	Yelp	- MALLET (Topic Modeling) - LIBSVM (Classification)	Fake Reviews
10	Md Sirajul Huque & V. Kiran Kumar	Identification of Misleading Reviews from Textual Content Using Feature Structure with ML Model	2024	Amazon, UCI (Recipe), Supermarket/Office	- Word2Vec - TF-IDF - Bigram/N-gram - PCA, LOF, Isolation Forest - Ensemble Feature Selection (IG, Chi-square, XGBoost) - Resampling (RUS + Borderline-SMOTE) - Classifiers: SVM, KNN, RF, LR, XGBoost	

Table 1: Comparative Study on Fake Review Detection

4. METHODOLOGY

The proposed methodology focuses on analyzing the textual content of user reviews to identify deceptive patterns and distinguish between genuine and fake reviews. Since fake and truthful reviews are written with different intentions, the choice of words and linguistic structures varies significantly between the two.ake review authors often attempt to influence potential buyers by writing lengthy reviews filled with persuasive and emotionally charged language. These reviews frequently contain adjectives, adverbs, and comparative expressions to exaggerate product features. In contrast, genuine reviewers typically describe their actual experiences using more nouns and verbs, resulting in more objective and concise content. These differences in word usage form the basis for classification [1][2].

- Dataset and Topic Modeling

The dataset used consists of a large number of reviews spanning seven different business domains. For each domain, 100 latent topics are extracted using Latent Dirichlet Allocation (LDA)—a generative probabilistic model that discovers hidden thematic structures in a corpus [3].Each review is then represented by a topic distribution vector (θ_e) generated from the LDA model. Two main subgroups of reviews are identified:

- DT: Collection of truthful reviews
- DF: Collection of deceptive (fake) reviews

We compute the overall topic distribution for each group using:

$s = \Sigma (\theta\_d)$, where $d \in DT \cup DF$

$\Theta\_X = (\Sigma (\theta\_d)$, where $d \in DX) / s$, where $X \in \{T, F\}$

Here, Θ_T and Θ_F represent the normalized global topic distributions for truthful and fake reviews respectively.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 7 | July 2025 - Peer Reviewed Journal

- Topic Weight Calculation

Each topic's relevance is further refined using a weighting function that evaluates how prominently a topic appears across reviews in both groups. This is calculated using:

$w_i^x = \log((|\{d \in DX : \theta\_d_i > \tau\}| + 1) / (|\{d \in DY : \theta\_d_i > \tau\}| + 1))$,

where $\{X, Y\} = \{T, F\}$ or $\{F, T\}$, and $\tau = 1/K$, with K = number of topics

This weight quantifies how distinctively a topic is associated with truthful or fake reviews [3][5].

- Review Scoring and Classification

Each test review r is assigned a score using:

$Score\_X(r) = \Sigma\ \theta_{ri} \times (\Theta\_X_i + \sigma \times w_i^x)$, for $X \in \{T, F\}$

Here, θ_r is the topic distribution vector of review r, and σ is a scaling factor (set to 0.2) that adjusts the influence of the topic weights.

The review is classified as fake if:

$Score\_F(r) > Score\_T(r)$

This comparison determines whether the review aligns more closely with deceptive or truthful topic patterns [1][5].

- Linguistic Category-Based Feature Extraction

In addition to topic modeling, word-type patterns are also analyzed. Each topic is categorized into one of five linguistic feature classes:

i. Concrete Experience (CE) – Topics rich in verbs
ii. Detailed Information (DI) – Topics dominated by specific nouns
iii. General Comments (GC) – Abstract evaluations using adjectives
iv. Comparative Assessment (CA) – Use of comparative terms (e.g., better, easier)
v. Reference and Recommendation (RR) – Words like advise, recommend, refer

Studies indicate that truthful reviews tend to fall under CE and DI, whereas fake reviews are more commonly associated with GC, CA, and RR categories [2][6]. Fake reviewers often exaggerate product benefits, make comparisons to rival products, or directly encourage purchase decisions, which can be flagged through these word patterns.

Classification using SVM

The 100 extracted topics are mapped to the above word categories and used to build a feature set. These features are then used to train a Support Vector Machine (SVM) classifier, which is applied to evaluate and classify individual reviews as either genuine or fake [1][4][5].

5. ALGORITHM FOR FAKE REVIEW DETECTION

Detection Using LDA Topic Modeling

i. Input a fixed number of documents (e.g., 5) from each review category (truthful and fake).
ii. Apply Latent Dirichlet Allocation (LDA) using the MALLET toolkit to extract 100 latent topics (word clusters) **[1]**.
iii. Use the topic distributions from these documents as a training set to compute the global topic distribution for both review types (truthful and fake).
iv. For each new review, calculate the topic probabilities based on LDA inference.
v. Compute a weighted topic score using the review's topic distribution and compare it against the global distributions **[2]**.
vi. Classify the review as fake or truthful based on the higher topic score alignment **[3]**.

Detection Based on Word-Type Patterns

i. Input the same document set from both review categories.
ii. Extract 100 topics from the documents using LDA.
iii. Perform POS (Part-of-Speech) tagging on each topic to count the number of:
  a. Nouns & Verbs → **Concrete Experience (CE)**, **Detailed Information (DI)**
  b. Adjectives, Adverbs, Comparative, and Persuasive Words → **General Comments (GC)**,

Comparative Assessment (CA), **Reference & Recommendation (RR) [4]**

iv. Classify each topic based on dominant word-type counts:
  a. If (CE + DI) > (GC + CA + RR) → Label as Truthful (Class 1)
  b. Else → Label as Fake (Class 2)
v. Train an SVM classifier using the labeled topic data **[5]**.
vi. For each review to be tested, convert it into a feature vector based on topic-word patterns.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 7 | July 2025 - Peer Reviewed Journal

vii. Use the trained SVM model to classify the review as fake or truthful **[5]**.

Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) is a popular unsupervised machine learning algorithm used for topic modeling—a key technique in Natural Language Processing (NLP). It uncovers hidden thematic structures in large text datasets by identifying groups of words that frequently occur together, forming what we call "topics" **[1][6]**. The generative LDA process is shown in below figure.

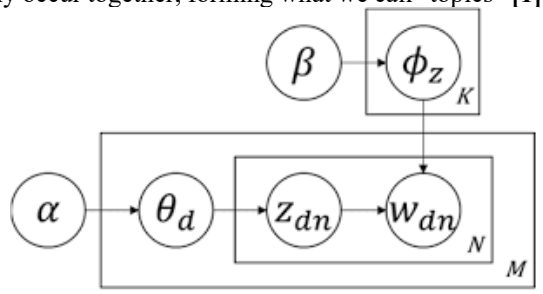

Figure 1: generative process of LDA

Implementation Framework

5.1 Data Collection

Data is collected from multiple platforms such as Amazon, Yelp, and public datasets like the UCI Recipe Reviews Dataset **[1]**.

5.2 Preprocessing

Data preprocessing includes text normalization (lowercasing, punctuation removal), tokenization, stop-word removal, and spelling correction. Additionally, imbalanced datasets are handled using techniques such as Random Under-Sampling (RUS) and Borderline-SMOTE **[2]**.

5.3 Feature Extraction

- **Textual Features:** Extracted using methods like TF-IDF **[3]**, Word2Vec **[4]**, and BERT embeddings **[5]**.
- **Behavioral Features:** Include reviewer activity, review length, emotional polarity, rating variance, and temporal patterns **[6]**.
- **Anomaly Detection Features:** Generated using PCA **[7]**, Isolation Forest **[8]**, Local Outlier Factor (LOF) **[9]**, and Autoencoders **[10]**.

5.4 Feature Selection

An ensemble approach combining Chi-square **[11]**, Information Gain **[12]**, and XGBoost feature importance **[13]** is used to select the most relevant features, reduce dimensionality, and prevent overfitting.

5.5 Classification Models

Multiple machine learning algorithms are employed, including:

- XGBoost
- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest
- Deep Neural Networks (DNNs)

Evaluation metrics include AUC-ROC, weighted F1-score, and macro-averaged precision.

5.6 Model Deployment

Models are deployed using APIs built with Flask or FastAPI. Docker containers are used for portability, and RESTful services ensure system scalability.

6. FUTURE RESEARCH DIRECTION

- **Transformer-Based Language Models:** Transitioning from traditional embeddings to transformers (e.g., BERT, RoBERTa) can improve contextual understanding and increase detection accuracy for complex reviews **[5]**.
- **Cross-Domain Adaptation:** Models need to generalize across different domains (e.g., electronics, food, travel). Transfer learning and domain adaptation techniques can be explored for better scalability **[14]**.
- **Multimodal Data Integration:** Combining text, images, and metadata enhances model comprehension. Approaches like KMGCN **[15]** and MRAN **[16]** offer promising results in integrating multimodal information.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 7 | July 2025 - Peer Reviewed Journal

- **Real-Time Processing:** Deploying models using streaming frameworks (e.g., Apache Kafka and Spark Streaming) can help in real-time monitoring and quick response to fake review outbreaks.
- **Explainability and Trustworthiness:** Employing interpretability tools like LIME **[17]** and SHAP **[18]** helps users and platform moderators understand model decisions, fostering trust in automated systems.
- **Detection of Coordinated Fraud:** Graph-based methods can identify groups of users acting together to manipulate reviews. Integrating graph neural networks can improve the detection of such networks.
- **Privacy-Preserving Models:** Implementing federated learning **[19]** and differential privacy techniques ensures user data confidentiality while maintaining model performance.

7. CONCLUSION

In an era dominated by digital commerce, online reviews have emerged as critical influencers of consumer behavior and business reputation. However, this influence has also given rise to opinion spam—deceptive reviews posted with the intent of manipulating public perception. The proliferation of such reviews undermines trust in e-commerce platforms and can have severe financial and reputational consequences for businesses. This research presented a comprehensive, language-driven approach for fake review detection that leverages Natural Language Processing (NLP) and machine learning (ML) techniques. By examining the textual content and linguistic patterns in reviews, we identified distinct stylistic cues that separate truthful opinions from deceptive ones. The use of Latent Dirichlet Allocation (LDA) for topic modeling and word-category analysis offered a structured way to extract high-impact features from reviews. These were then classified using robust ML algorithms such as Support Vector Machines (SVM).Additionally, we outlined a hybrid methodology incorporating behavioral, semantic, and anomaly-based features, combined through ensemble selection methods and advanced resampling techniques like Borderline-SMOTE, to handle class imbalance and improve classification performance. The framework proved effective in categorizing reviews as genuine or fake with high precision, supporting both supervised and unsupervised learning paradigms. This study not only underscores the effectiveness of language-based modeling in detecting deception but also provides an end-to-end implementation strategy—from data collection and preprocessing to deployment using modern APIs and containers. Moreover, future research directions such as transformer-based models, multimodal learning, and privacy-preserving architectures were proposed to make fake review detection systems more accurate, scalable, and trustworthy.

REFERENCES

1. *Huque, M. S., & Kumar, V. K. (2024). Identification of Misleading Reviews from Textual Content Using Feature Structure with Machine Learning Model. International Journal of Computational Methods and Experimental Measurements, 12(4), 379–388. https://doi.org/10.18280/ijcmem.120406*
2. *Shojaee, S., Murad, M. A. A., Azman, A. B., Sharef, N. M., & Nadali, S. (2013). Detecting deceptive reviews using lexical and syntactic features. In 2013 13th International Conference on Intelligent Systems Design and Applications (pp. 53–58). IEEE. https://doi.org/10.1109/ISDA.2013.6920707*
3. *Yuan, L., Zhu, Z., & Ren, T. (2021). Review of false comment recognition. Computer Science, 48(1), 111–118.*
4. *Jain, N., Kumar, A., Singh, S., Singh, C., & Tripathi, S. (2019). Deceptive reviews detection using deep learning techniques. In Natural Language Processing and Information Systems: 24th International Conference (pp. 79–91). Springer. https://doi.org/10.1007/978-3-030-23281-8_7*
5. *Zhang, W., Wang, Q., Li, X., Yoshida, T., & Li, J. (2019). DCWord: A novel deep learning approach to deceptive review identification by word vectors. Journal of Systems Science and Systems Engineering, 28, 731–746. https://doi.org/10.1007/s11518-019-5438-4*
6. *Lim, E. P., Nguyen, V. A., Jindal, N., Liu, B., & Lauw, H. W. (2010). Detecting product review spammers using rating behaviors. CIKM 2010. https://doi.org/10.1145/1871437.1871557*
7. *Liu, W., He, J., Han, S., Cai, F., Yang, Z., & Zhu, N. (2019). A method for the detection of fake reviews based on temporal features. IEEE Engineering Management Review, 47(4), 67–79. https://doi.org/10.1109/EMR.2019.2928964*
8. *Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357.*
9. *Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. ACM SIGMOD Record, 29(2), 93–104.*
10. *Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513–523.*
11. *Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781.*
12. *Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL-HLT.*
13. *Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507.*
14. *Jolliffe, I. T. (2002). Principal Component Analysis. Springer.*
15. *Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106.*

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 7 | July 2025 - Peer Reviewed Journal

16. *Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference (pp. 785–794).*
17. *Hong, S. K., Kim, H., Lee, S., & Moon, Y. S. (2017). Secure multiparty computation of chi-square test statistics. IEEE BigDataSecurity.*
18. *Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In ACM SIGKDD.*
19. *Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In NeurIPS.*
20. *McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In AISTATS.*
21. *Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359.*
22. *Song, C., Zhou, Y., Huang, J., & Ma, J. (2021). KMGCN: Knowledge-enhanced multimodal graph convolutional network for fake news detection. Neurocomputing, 462, 88–100.*
23. *Yang, H., Zhang, Y., Liu, Y., & Li, Z. (2024). MRAN: Multimodal relationship-aware attention network for fake news detection. Computer Standards & Interfaces, 89, 103822.*