

# Fake Review Detection in Online Platforms: A Machine Learning and Deep Learning Approach

**Prakash Wagh**

*School of Data Science*

*Symbiosis Skills and Professional University, Pune, Maharashtra, India*

## ABSTRACT

With the rapid growth of e-commerce platforms, online reviews have become a critical factor influencing consumer decision-making. However, the increasing presence of deceptive or fake reviews poses a serious threat to trust and transparency in digital marketplaces. Traditional machine learning techniques often fail to capture the contextual semantics of textual data, limiting their effectiveness in identifying sophisticated fake reviews.

This research proposes a hybrid fake review detection system that integrates Natural Language Processing (NLP), traditional machine learning models, and a lightweight transformer-based model, DistilBERT. The system performs automated data collection, preprocessing, contextual embedding, and classification to distinguish between genuine and fake reviews. Additionally, real-time analysis is achieved through a Flask-based deployment architecture.

Experimental insights suggest that transformer-based models significantly improve detection accuracy while maintaining computational efficiency. The proposed system provides a scalable and practical solution for enhancing the credibility of online review systems.

**Keywords:** *Fake Reviews, Fake Reviews, NLP, Machine Learning, Deep Learning, DistilBERT, Sentiment Analysis, E-commerce, Text Classification*

## 1. INTRODUCTION

The rapid expansion of digital marketplaces and e-commerce platforms has fundamentally transformed the way consumers make purchasing decisions. Online reviews have become one of the most influential sources of information for customers, often replacing traditional word-of-mouth recommendations. Platforms such as **Amazon**, **Flipkart**, and **eBay** rely heavily on user-generated reviews to build credibility and guide potential buyers. Studies indicate that a significant percentage of consumers read online reviews before purchasing a product, making review authenticity a critical factor in digital commerce.

However, the growing dependency on online reviews has also led to the proliferation of fake or deceptive reviews. These reviews are intentionally written to mislead customers by either artificially boosting product ratings (promotional spam) or damaging competitors' reputations (defamatory spam). Fake reviews may be generated by individuals hired for promotional purposes or by automated bots using advanced text-generation tools. Such practices distort product rankings, manipulate consumer perception, and negatively impact fair market competition.

Detecting fake reviews presents several technical challenges. Deceptive reviews are often written in a manner that closely resembles genuine opinions, making manual detection inefficient and unreliable. Traditional rule-based filtering systems and keyword-based approaches lack the ability to understand deeper contextual and semantic meanings within textual data. As a result, more advanced computational techniques are required to effectively identify deceptive patterns.

Machine Learning (ML) approaches, such as Logistic Regression, Support Vector Machines (SVM), and Random Forest classifiers, have been widely used for text classification tasks. These models typically rely on feature extraction techniques like Bag-of-Words and TF-IDF. While these approaches can capture frequency-based information, they struggle to understand contextual dependencies and semantic relationships between words.

**1. Literature Review**

Year	Publisher	Author(s)	Name of the Paper	Objective	Methodology	Gap
2019	IRJET	Ms. Raishri P. Kashti, Dr. S. Prasad	Enhancing Techniques for Fake News Detection	NLP Improve fake detection	Advanced NLP techniques for linguistic analysis	Complex language variations and need for labeled data
2020	MCES Press (IJIEEB)	Naznin Sultana, Prof. Sellappan Palaniappan	Deceptive Detection Machine Learning Techniques	Opinion Using Machine Learning Detect deceptive opinions in online reviews	Applied machine learning classification algorithms	Accuracy may decrease across domains and languages
2021	IEEE	M.F. Mridha, Ashfia Jannat Keya	A Comprehensive Review on Fake News Detection with Deep Learning	Evaluate learning techniques detection	deep Dataset preprocessing and for deep learning models	Requires large datasets and high computational resources
2021	IEEE	Rami Mohawesh, Shuxiang Xu, Son N. Tran	Fake Reviews Detection: A Survey	Survey review detection approaches	Analysis of fake datasets, features, and neural networks	Depends on dataset quality and may not address emerging spam
2021	IJACSA	Ahmed M. Elmogy, Usman Tariq, Atef Ibrahim	Fake Reviews Detection using Supervised Machine Learning	Develop using supervised Machine model for review detection	Applied ML algorithms to textual features	ML depends on quantity and quality of labeled data
2022	Springer	K. Pooja, Pallavi Upadhyaya	What Makes an Online Review Credible? A Systematic Review	Identify factors affecting credibility	Systematic literature review	Limited by platform differences and cultural context
2023	MDPI	Junwen Lu, Xintao Zhan, Guanfeng Liu, Xinrong Zhan, Xiaolong Deng	BSTC: Fake News Detection Based on Pre-Trained Language Model and CNN	Review Model Combine transformer model with CNN	Pre-trained language model integrated with CNN	Does not cover all fake review detection techniques
2023	Springer	Rahul Kumar, Shubadeep Mukherjee, Nirpendra Rana	Exploring Characteristics of Fake Reviews	Latent Analyze hidden characteristics of fake reviews	Text analysis to identify latent patterns	Hard to generalize across multiple platforms
2023	IJERT	Abhijeet Rathore, Gayatri L. Bhadane, Ankita D. Jadhav, Kishor H. Dhale	Fake Reviews Detection Using Model and Neural Network	Detect reviews using NLP and neural networks	fake NLP feature extraction with neural network classifier	Requires large datasets and risk of overfitting
2024	IEEE	Dr. Atika Qazi, Dr.	Machine Learning- Analyze	Analyze	ML Categorization	Dependent on

Year	Publisher	Author(s)	Name of the Paper	Objective	Methodology	Gap
		Najmul Hasan, Rui Mao	Dr. Based Opinion Spam Detection: Systematic Literature Review	Spam techniques A spam detection Literature Review	for and evaluation of dataset ML models	quality and new spam techniques

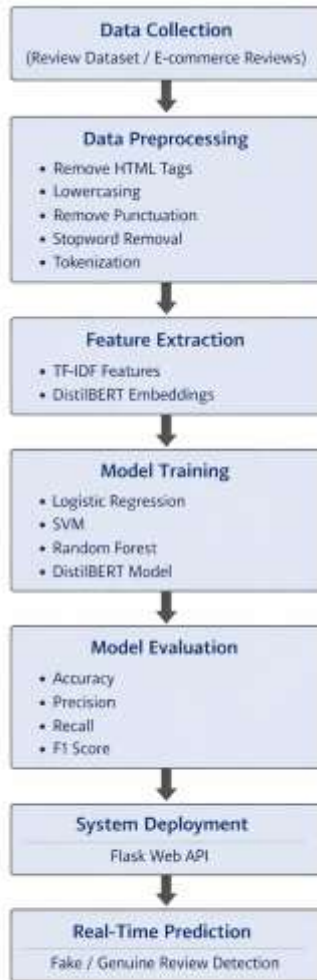


Figure: Architecture of Fake Review Detection System

## 1.PROJECT ARCHITECCTURE

### 1 Data Collection

The process begins with the collection of review data from reliable sources such as publicly available datasets or e-commerce platforms. These datasets typically contain user-generated reviews along with labels indicating whether the review is genuine or fake. The quality and diversity of the dataset play a crucial role in improving the robustness and generalization capability of the model.

### 2 Data Preprocessing

Raw textual data collected from online platforms is often noisy and unstructured. Therefore, preprocessing is a critical step to enhance data quality before feeding it into the models. The following preprocessing techniques are applied:

- **Removal of HTML Tags:** Eliminates unnecessary web elements embedded in the text
- **Lowercasing:** Converts all text to a uniform format for consistency
- **Punctuation Removal:** Removes symbols that do not contribute to semantic meaning
- **Stopword Removal:** Filters out commonly used words (e.g., “is”, “the”) that add little value

- **Tokenization:** Splits text into smaller units (tokens) for further analysis

These steps ensure that the input data is clean, standardized, and suitable for feature extraction.

---

### 3 Feature Extraction

To enable machine learning models to process textual data, it must be transformed into numerical representations. This system employs a hybrid feature extraction approach:

- **TF-IDF (Term Frequency–Inverse Document Frequency):**

Captures the importance of words based on their frequency and uniqueness across documents. It is effective for traditional machine learning models.

- **DistilBERT Embeddings:**

Generates dense, context-aware vector representations of text using a transformer-based model. Unlike TF-IDF, these embeddings capture semantic relationships and contextual meaning within sentences.

This combination allows the system to leverage both statistical and contextual features for improved performance.

---

### 4 Model Training

The processed and transformed data is then used to train multiple classification models. The system incorporates both traditional and deep learning approaches:

- **Logistic Regression:** A baseline model for binary classification
- **Support Vector Machine (SVM):** Effective for high-dimensional text data
- **Random Forest:** Ensemble method that improves prediction stability
- **DistilBERT Model:** A lightweight transformer model that captures deep contextual semantics

By comparing these models, the system identifies the most efficient approach in terms of accuracy and computational cost.

---

### 5 Model Evaluation

To assess the effectiveness of the trained models, several evaluation metrics are used:

- **Accuracy:** Measures overall correctness of predictions
- **Precision:** Indicates how many predicted fake reviews are actually fake
- **Recall:** Measures the ability to correctly identify all fake reviews
- **F1-Score:** Harmonic mean of precision and recall, providing a balanced evaluation

These metrics ensure a comprehensive performance analysis and help in selecting the best-performing model.

---

### 6 System Deployment

Once the optimal model is selected, it is deployed using a **Flask-based web API**. This enables seamless integration with user interfaces and allows the system to process incoming review data dynamically. The deployment layer acts as a bridge between the trained model and real-world applications.

---

### 7 Real-Time Prediction

The final stage of the system involves real-time classification of reviews. Users can input review text, and the system processes it through the trained pipeline to determine whether the review is **fake or genuine**. This capability enhances the practical usability of the system in live environments such as e-commerce platforms.

## 5. CONCLUSION

Fake review detection is essential for maintaining trust and transparency in online marketplaces. This study demonstrates that transformer-based models such as DistilBERT significantly outperform traditional machine learning approaches by capturing contextual semantics within review text.

The proposed system provides an efficient, scalable, and real-time solution for identifying deceptive reviews. Future improvements may include ensemble models, multilingual support, behavioral feature integration, and continual learning mechanisms to adapt to evolving fake review strategies

## 6. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who contributed to the successful completion of this research work on *Fake Review Detection using Machine Learning and Deep Learning techniques*.

First and foremost, I am deeply thankful to my project supervisors, **Prof. Peeyush Pareek** and **Prof. Vishnu Potdar**, for their constant guidance, valuable insights, and continuous encouragement throughout the research process. Their expertise and support played a crucial role in shaping this work.

I would also like to extend my gratitude to the faculty members of the **School of Data Science, Symbiosis Skills and Professional University, Pune**, for providing the necessary academic environment and resources required for this study.

## 7. REFERENCES

- Qazi, A., Hasan, N., & Mao, R. (2024). Machine Learning-Based Opinion Spam Detection: A Systematic Literature Review.
- Mridha, M. F., & Keya, A. J. (2021). A Comprehensive Review on Fake News Detection with Deep Learning.
- Mohawesh, R., Xu, S., & Tran, S. N. (2021). Fake Reviews Detection: A Survey.
- Lu, J., Zhan, X., Liu, G., & Deng, X. (2023). BSTC: A Fake Review Detection Model Based on Pre-Trained Language Model and CNN.
- Kumar, R., Mukherjee, S., & Rana, N. P. (2023). Exploring Latent Characteristics of Fake Reviews and Their Intermediary Role in Persuading Buying Decisions.
- Rathore, A. A., et al. (2023). Fake Reviews Detection Using NLP Model and Neural Network Model.
- Elmogy, A. M., Tariq, U., & Ibrahim, A. (2021). Fake Reviews Detection using Supervised Machine Learning.