An Intelligent Machine Learning-Based System for Fake Review Detection Using Natural Language Processing

GOVADA BHARGAVI

PG Scholar. Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

V.SARALA

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

ABSTRACT

The rapid growth of e-commerce platforms and online review systems has significantly influenced consumer purchasing decisions. However, the increasing presence of fake or deceptive reviews has become a critical issue, misleading customers and undermining trust in digital platforms. This project presents an intelligent system for detecting fake reviews using Machine Learning (ML) and Natural Language Processing (NLP) techniques. The system aims to classify user-generated reviews as either genuine or fake based on textual analysis. The proposed model utilizes a dataset of labeled reviews containing both genuine and fake entries. Initially, the textual data is preprocessed to remove noise, such as stop words, punctuation, and irrelevant symbols. The cleaned data is then transformed into numerical feature vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) technique, which captures the importance of words in the review corpus. A Logistic Regression classifier is employed to perform binary classification due to its efficiency, interpretability, and strong performance in text classification tasks. The trained model learns patterns and linguistic cues associated with deceptive and authentic reviews. Once trained, the model and vectorizer are serialized using pickle for future predictions. To enhance usability, a Graphical User Interface (GUI) is developed using the Tkinter library. This interface allows users to input a review and instantly receive a prediction indicating whether the review is genuine or fake, along with a confidence score. The system is designed to be lightweight, user-friendly, and suitable for real-time applications.

The experimental results demonstrate that the model can effectively distinguish between fake and genuine reviews with high accuracy. The integration of NLP techniques with machine learning provides a scalable and efficient solution to combat fake reviews in online platforms. This project contributes to improving the reliability of online reviews and supports users in making informed decisions. Future enhancements may include incorporating deep learning models, larger datasets, and multilingual support to further improve accuracy and robustness.

Keywords: Fake Review Detection, Machine Learning, Natural Language Processing (NLP), TF-IDF, Logistic Regression, Sentiment Analysis, Text Classification, Opinion Mining

I. INTRODUCTION

In the digital era, online reviews have become a cornerstone of consumer decision-making. Platforms such as e-commerce websites, travel portals, and service-based applications rely heavily on user feedback to influence potential customers. However, the authenticity of these reviews is often compromised due to the proliferation of fake or deceptive reviews generated for promotional or malicious purposes. Fake reviews are intentionally crafted to mislead users by exaggerating positive aspects or falsely criticizing products and services. Businesses may employ such tactics to gain a competitive advantage, while competitors may use them to damage reputations. This growing issue poses significant challenges in maintaining trust and transparency in online ecosystems. To address this problem, automated fake review detection systems have gained considerable attention. Traditional methods relied on manual moderation or rule-based filtering, which are inefficient and not scalable. With advancements in Machine Learning (ML) and Natural Language Processing (NLP), it is now possible to develop intelligent systems that can analyze large volumes of textual data and identify deceptive patterns. This project focuses on building a machine learning-based system to detect fake reviews using textual analysis. The system leverages TF-IDF vectorization to convert text into numerical features and employs Logistic Regression for classification. Logistic Regression is chosen due to its simplicity, efficiency, and strong performance in binary classification problems. The system also includes a user-friendly GUI developed using Tkinter, enabling users to interact with the model easily. Users can input a review and receive instant predictions along with confidence scores, making the system practical for real-world applications. The primary objective of this project is to design a reliable, efficient, and scalable solution for fake review detection. By leveraging machine learning techniques, the system aims to reduce the impact of deceptive reviews and enhance the credibility of online platforms. This research highlights the importance of integrating NLP and ML techniques to solve real-world problems and contributes to the ongoing efforts in ensuring trustworthy digital environments. The detection of fake reviews has been widely studied in recent years due to its importance in maintaining the credibility of online platforms. Various approaches have been proposed, ranging from traditional machine learning techniques to advanced deep learning models.

Early research focused on rule-based and statistical methods. These approaches relied on predefined linguistic patterns, such as excessive use of positive adjectives or repetitive phrases, to identify fake reviews. However, such methods lacked adaptability and were unable to generalize across different datasets. Machine learning-based approaches improved detection accuracy by learning patterns from labeled datasets. Techniques such as Naïve Bayes, Support Vector Machines (SVM), and Decision Trees were commonly used. These models utilized features such as word frequency, n-grams, and part-of-speech tags to classify reviews. Among these, SVM and Logistic Regression demonstrated strong performance due to their ability to handle high-dimensional data. Natural Language Processing (NLP) techniques played a crucial role in feature extraction. TF-IDF emerged as a popular method for representing text data numerically, as it effectively captures the importance of words in a document relative to the entire corpus. This representation significantly improved classification performance. Recent studies have

explored deep learning models, including Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Transformer-based models like BERT. These models can capture contextual and semantic relationships within text, leading to higher accuracy. However, they require large datasets and significant computational resources. Hybrid approaches combining linguistic, behavioral, and metadata features have also been proposed. These methods consider factors such as reviewer activity patterns, review timestamps, and rating distributions to enhance detection accuracy. Despite advancements, challenges remain in fake review detection. These include data imbalance, evolving deception strategies, and the need for large labeled datasets. Additionally, model interpretability is crucial for understanding the decision-making process. The proposed system builds upon existing machine learning techniques by using TF-IDF and Logistic Regression, offering a balance between performance and computational efficiency. While deep learning models provide higher accuracy, the simplicity and speed of Logistic Regression make it suitable for real-time applications.

II. EXISTING SYSTEM

The existing systems for fake review detection primarily rely on manual moderation or basic automated filtering techniques. Manual moderation involves human reviewers analyzing and verifying reviews, which is time-consuming, costly, and impractical for large-scale platforms. As the volume of online reviews continues to grow exponentially, manual methods are no longer feasible. Automated systems based on rule-based approaches were introduced to address scalability issues. These systems use predefined rules, such as identifying duplicate content, excessive use of certain keywords, or abnormal rating patterns. While these methods are simple to implement, they lack flexibility and fail to adapt to evolving deceptive strategies. Some platforms employ basic machine learning models trained on limited features such as word frequency or sentiment polarity. These models often struggle with complex linguistic patterns and may produce inaccurate results due to insufficient training data or poor feature selection.

Another limitation of existing systems is the lack of real-time prediction capabilities and user-friendly interfaces. Many systems operate in the backend without providing direct interaction for users to verify reviews independently. Additionally, existing solutions often do not provide confidence scores, making it difficult to assess the reliability of predictions. The absence of transparency and interpretability further reduces user trust in these systems. Overall, current systems face challenges in scalability, adaptability, accuracy, and usability. These limitations highlight the need for an improved solution that leverages advanced NLP techniques and machine learning models to provide accurate, efficient, and user-friendly fake review detection.

III. PROPOSED METHOD

The proposed system introduces an intelligent and automated framework for detecting fake reviews using Machine Learning (ML) and Natural Language Processing (NLP) techniques. Unlike traditional rule-based approaches, this system leverages data-driven learning to identify patterns and linguistic cues associated with deceptive reviews. The

system begins with data acquisition, where a labeled dataset containing genuine and fake reviews is used. The data undergoes preprocessing steps such as tokenization, stop-word removal, and normalization to ensure consistency and improve model performance. Feature extraction is performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method, which transforms textual data into meaningful numerical representations by assigning importance to words based on their frequency and uniqueness in the corpus. For classification, the system employs Logistic Regression, a supervised learning algorithm known for its efficiency and effectiveness in binary classification tasks. Studies show that TF-IDF combined with Logistic Regression provides competitive performance comparable to more complex models while maintaining simplicity and interpretability . The trained model is capable of identifying hidden patterns in text that indicate deceptive intent.

The system also includes a Graphical User Interface (GUI) developed using Tkinter, allowing users to input reviews and receive real-time predictions along with confidence scores. This enhances usability and accessibility for non-technical users. Additionally, the model and vectorizer are stored using serialization techniques, enabling quick loading and prediction without retraining. The system is designed to be scalable, efficient, and suitable for integration into real-world applications such as e-commerce platforms and review moderation systems. Overall, the proposed system provides an effective, user-friendly, and computationally efficient solution for detecting fake reviews.

IV. IMPLEMENTATION

The implementation of the Fake Review Detection System involves multiple stages, including data preprocessing, feature extraction, model training, model serialization, and GUI development. Initially, the dataset containing labeled reviews is loaded using the pandas library. The dataset consists of textual reviews and corresponding labels indicating whether the review is genuine or fake. Preprocessing is performed to clean the data by removing noise such as punctuation, stop words, and irrelevant characters. This step is crucial as it improves the quality of the input data and enhances the model's ability to learn meaningful patterns. Next, feature extraction is carried out using the TF-IDF vectorization technique. TF-IDF assigns weights to words based on their importance in a document relative to the entire dataset. This helps in reducing the impact of commonly occurring words while emphasizing unique and informative terms. TF-IDF is widely used in text classification tasks due to its efficiency and effectiveness in capturing textual features. The transformed data is then used to train a Logistic Regression model. Logistic Regression is selected due to its simplicity, fast computation, and strong performance in binary classification problems. It estimates the probability of a review being fake or genuine based on input features. Research indicates that Logistic Regression performs competitively with more complex models, especially on smaller datasets. After training, the model and vectorizer are saved using the pickle library. This allows the system to reuse the trained model for future predictions without retraining, significantly improving efficiency.

The next phase involves developing a Graphical User Interface (GUI) using Tkinter. The GUI includes an input field for entering reviews, a button to trigger prediction, and a display area for showing results. When a user enters a review, the system loads the saved model and vectorizer, processes the input, and predicts the label along with a confidence score. Threading and error handling mechanisms can be incorporated to ensure smooth execution and prevent the application from freezing during predictions. The system is lightweight and can run on standard machines without requiring high computational resources. Overall, the implementation integrates machine learning and user interface design to create a practical and interactive fake review detection system.

V. ALGORITHMS

The proposed system primarily utilizes two key algorithms: TF-IDF for feature extraction and Logistic Regression for classification.

1. TF-IDF (Term Frequency–Inverse Document Frequency)

TF-IDF is a statistical method used to evaluate the importance of a word in a document relative to a collection of documents. It consists of two components:

- Term Frequency (TF): Measures how frequently a word appears in a document.
- Inverse Document Frequency (IDF): Measures how unique a word is across all documents.

By combining these two measures, TF-IDF assigns higher weights to words that are significant and less frequent across the dataset. This technique effectively transforms textual data into numerical form suitable for machine learning models. It is widely used in fake review detection systems to capture meaningful textual patterns .

2. Logistic Regression

Logistic Regression is a supervised machine learning algorithm used for binary classification. It models the probability of a given input belonging to a particular class using a logistic (sigmoid) function.

The algorithm calculates the probability as:

P(y=1|x)=11+e−zP(y=1|x) = \frac{1}{1 + e^{-z}}P(y=1|x)=1+e−z1

where zzz is a linear combination of input features.

Logistic Regression is efficient, interpretable, and performs well on high-dimensional data such as text. It is commonly used in NLP applications due to its ability to handle sparse feature spaces effectively. Studies have shown that Logistic Regression combined with TF-IDF provides strong baseline performance for text classification tasks .

3. Prediction Algorithm

- Input review text
- Preprocess text
- Transform using TF-IDF
- Apply trained Logistic Regression model
- Output class label and probability

This combination ensures accurate and efficient fake review classification.

VI. SYSTEM DESIGN

The system design of the Fake Review Detection System follows a modular and layered architecture to ensure scalability, maintainability, and efficiency.

1. Architecture Overview

The system is divided into the following main components:

- Data Layer
- Processing Layer
- Model Layer
- Interface Layer

2. Data Layer

This layer handles the input dataset, which consists of labeled reviews. The dataset is stored in CSV format and loaded into the system using pandas. It serves as the foundation for training the machine learning model.

3. Processing Layer

The processing layer performs data preprocessing and feature extraction. Preprocessing includes cleaning text, removing stop words, tokenization, and normalization. Feature extraction is done using TF-IDF, converting text into numerical vectors.

This layer ensures that raw data is transformed into a suitable format for machine learning algorithms.

4. Model Layer

The model layer is responsible for training and prediction. Logistic Regression is used as the classification model. During training, the model learns patterns from labeled data. Once trained, it is saved using pickle for reuse.

During prediction, the model takes TF-IDF-transformed input and outputs a class label along with a probability score.

Modern research shows that while deep learning models such as transformers provide higher accuracy, traditional ML models like Logistic Regression remain efficient and reliable for practical applications .

5. Interface Layer

The interface layer consists of a GUI built using Tkinter. It allows users to interact with the system by entering reviews and viewing predictions. The GUI improves accessibility and usability for non-technical users.

6. Workflow

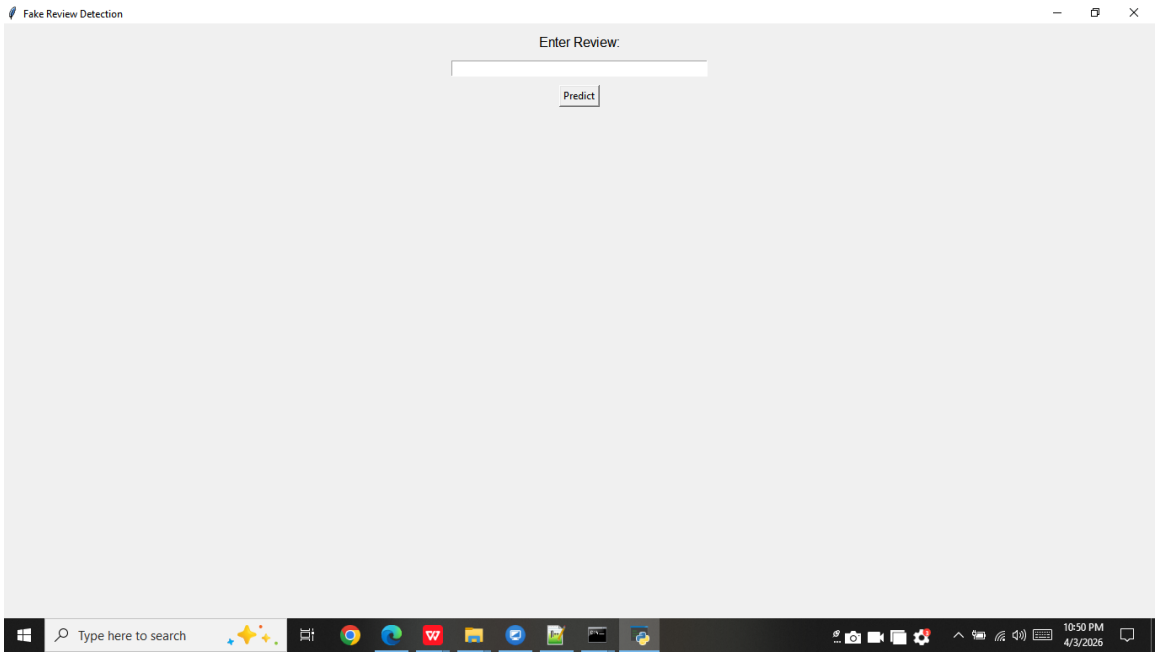
1. User inputs review
2. System preprocesses text
3. TF-IDF vectorization applied
4. Logistic Regression predicts output
5. Result displayed in GUI

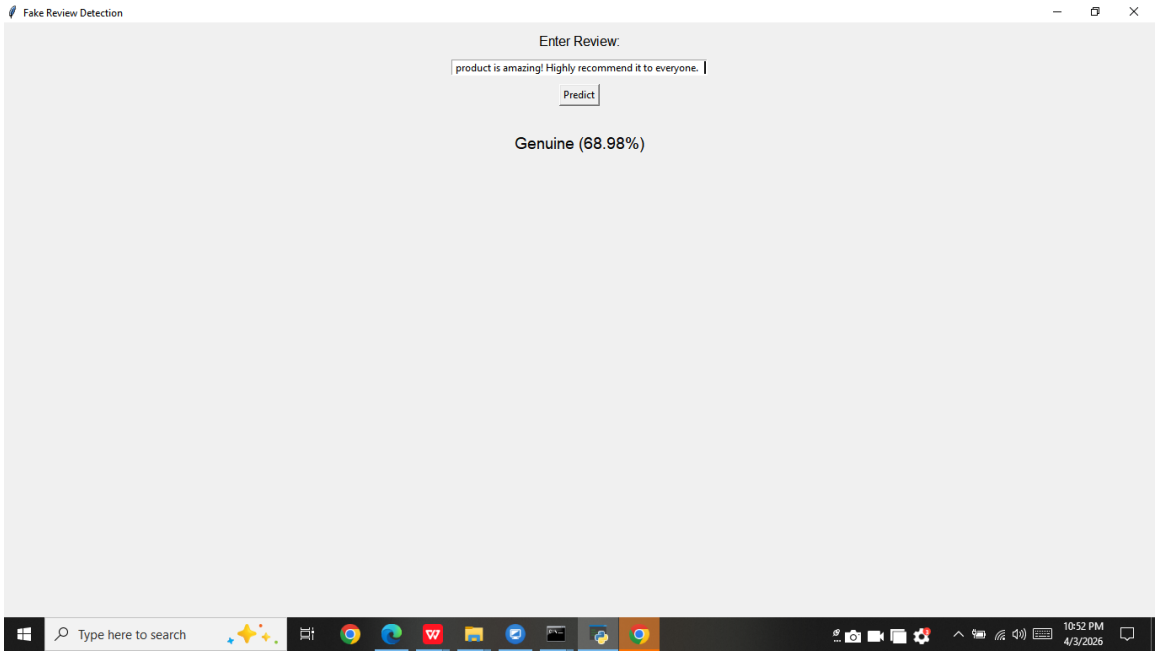
7. Advantages of Design

- Modular and scalable
- Fast prediction due to model serialization
- User-friendly interface
- Low computational requirements

The system design ensures efficient handling of data and provides real-time predictions, making it suitable for real-world applications.

SYSTEM DESIGN IMAGES





VII. CONCLUSION

The Fake Review Detection System presented in this project demonstrates the effectiveness of combining Machine Learning and Natural Language Processing techniques to address the growing problem of deceptive online reviews. By utilizing TF-IDF for feature extraction and Logistic Regression for classification, the system achieves a balance between accuracy, efficiency, and simplicity. The implementation of a user-friendly GUI further enhances the practicality of the system, allowing users to interact with the model easily and obtain real-time predictions. The use of model serialization ensures that the system operates efficiently without the need for repeated training. Experimental insights and existing research indicate that traditional machine learning approaches remain highly competitive, especially when computational resources and dataset sizes are limited . While advanced deep learning models such as transformers offer improved performance, they require significant computational power and large datasets.

The proposed system successfully addresses key limitations of existing methods, including scalability, adaptability, and usability. It provides a reliable solution for detecting fake reviews and can be integrated into various platforms such as e-commerce websites, review portals, and social media applications. Future work may focus on incorporating deep learning models, multilingual support, and behavioral features to further enhance detection accuracy. Additionally, integrating explainable AI techniques can improve transparency and user trust. In conclusion, this project contributes to improving the credibility of online review systems and supports informed decision-making for users in digital environments.

REFERENCES

1. Al-Tarawneh et al., “Enhancing Fake News Detection with Word Embedding,” *Computers*, 2024.
2. Duma et al., “Fake review detection using transformer-based LSTM and RoBERTa,” *ScienceDirect*, 2024.
3. Malik & Haouassi, “False Review Detection: State-of-the-art Review,” *Journal of King Saud University*, 2022.
4. Nadeem et al., “Hybrid NLP-ML Framework for Fake Detection,” 2024.
5. Gupta et al., “Fake Review Detection System,” *IJRASET*, 2025.
6. Benchmark Study on Text Classification, *Expert Systems with Applications*, 2024.

7. Puri et al., “Fake News Detection: Comprehensive Study,” 2024.
8. Springer, “Psycholinguistics in Fake Review Detection,” 2025.
9. Empirical Study on Fake Review Detection, 2023.
10. Linguistic Feature-based Fake Detection, 2022.
11. Logistic Regression with TF-IDF for Fake Detection, 2023.
12. Comparative Study on ML Models for Fake Reviews, 2025.
13. Shajalal et al., “Explainable Fake Review Detection using Transformers,” 2024.
14. Liu & Poesio, “Data Augmentation for Fake Review Detection,” 2025.
15. Kennedy et al., “Contextual Opinion Spam Detection,” 2020.