ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Detection of Fake Online Reviews Using Supervised & Semi Supervised Learning

M NIkitha
21N81A6253
Computer Science and Engineering
(Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
nniki3609@gmail.com

D Prashanthi
21N81A6265
Computer Science and Engineering
(Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
pranshanthidonapati1901@gmail.com

K.Yogi gayathri
21N81A6264
Computer Science and Engineering
(Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
ggayathriyogi70@gmial.com

K Sujith
21N81A685
Computer Science and Engineering
(Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
sugithkoneru44@gmail.com

Mrs. P Sandhya rani
Assistant Professor
Computer Science and Engineering (Cyber Security)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510

ABSTRACT: The authenticity of online reviews is crucial for maintaining trust and credibility in e-commerce platforms. However, the increasing prevalence of fake reviews poses a significant challenge, deceiving consumers and undermining the reliability of review systems. Fake reviews can be generated by individuals or bots, making it difficult for consumers to make informed decisions.

1. INTRODUCTION

In the digital age, online reviews play a crucial role in shaping consumer behaviour and influencing business reputations. Whether it's purchasing a product, booking a hotel, or choosing a restaurant, many people rely heavily on the opinions shared by others on platforms

such as Amazon, Yelp, and TripAdvisor. However, the increasing prevalence of **fake or deceptive reviews**—often generated for fraudulent marketing or defamation—poses a serious threat to the reliability of these platforms. Detecting fake reviews is a challenging task due to the subtle nature of deceptive writing and the limited availability of labelled data. Traditional supervised machine learning techniques require large amounts of labelled data to build accurate classifiers, but obtaining such data is often time-consuming and costly. To overcome this limitation, **semi-supervised learning** offers a promising alternative by leveraging a small set of labelled reviews alongside a large volume of unlabelled ones to improve detection accuracy. This project aims to explore and implement both **supervised and semi-supervised learning approaches** to effectively identify fake online reviews. Supervised methods such as Logistic Regression, Support Vector Machines, and Random Forests are first employed using available labelled datasets. Then, semi-supervised techniques like Self-training, Label Propagation, and Semi-Supervised SVMs are applied to enhance the model's performance by exploiting the patterns in unlabelled data. By comparing the performance of these approaches, the project seeks to identify the most effective strategies for detecting fake reviews with limited labelled data—ultimately contributing to more trustworthy online review ecosystems.

2. LITERATURE REVIEW

[1] **Jindal, N. & Liu, B. (2008).**
Opinion Spam and Analysis.
Proceedings of the 2008 International Conference on Web Search and Data Mining
This seminal paper introduced the problem of opinion spam and proposed supervised classifiers based on textual and behavioral.

[2] **Feng, S., Banerjee, R., & Choi, Y. (2012).**
Syntactic Stylometry for Deception Detection.
Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics

Applies semi-supervised techniques using syntactic features for deception detection.

[3] **Alsmadi, I., & Qablan, B. (2020).**
Detection of Fake Reviews Using Machine Learning Algorithms: A Survey.
IEEE Access.
A comprehensive survey of machine learning approaches including supervised and semi-supervised methods for fake review detection.

3. Existing methods

In the current digital era, a large number of consumers rely on online reviews to make purchasing decisions. However, the presence of fake or deceptive reviews has become a growing concern for both users and online platforms. Existing systems for fake review detection primarily focus on supervised learning techniques. These systems depend on labelled datasets, where reviews are annotated as either fake or genuine. Various classification algorithms such as Support Vector Machines (SVM), Logistic Regression, Random Forest, and Naive Bayes have been widely used. These models utilize features derived from textual content (e.g., n-grams, sentiment scores, linguistic cues), behavioural patterns of users (e.g., review

posting frequency, account age), and metadata (e.g., timestamp, review length). While these approaches have shown good accuracy, they are limited by the availability of large-scale labelled data, which is often expensive and time-consuming to obtain.

PROPOSED SYSTEM

The proposed system aims to enhance the accuracy and robustness of fake online review detection by integrating both supervised and semi-supervised learning approaches. Unlike traditional systems that rely solely on large labelled datasets, this model leverages a small set of labelled data along with a larger corpus of unlabelled reviews to improve detection performance while minimizing the need for extensive manual annotation. The system is designed to automatically classify reviews as genuine or fake based on a combination of textual analysis, behavioural patterns, and metadata features.

Initially, the system applies data preprocessing techniques such as tokenization, stop-word removal, and stemming to clean and normalize the review text. Feature extraction is then performed using both linguistic and statistical methods, including TF-IDF vectors, sentiment scores, and review-specific attributes like word count and punctuation usage. Additionally, user behavioural features such as the frequency of reviews, account age, and rating distribution are incorporated to provide a more comprehensive understanding of potential fraudulent behaviour.

4. IMPLEMENTATION

The implementation of the fake review detection system involves a series of well-defined steps, starting from data acquisition and preprocessing, followed by feature extraction, model training using both supervised and semi-supervised learning, and finally classification and evaluation. Below is a detailed explanation of each step:

4.1 Data Collection

The first step involves gathering online review data from platforms such as Amazon, Yelp, or publicly available datasets. Each review entry typically contains the text of the review, user information, star rating, and a label indicating whether the review is fake or genuine. In the absence of a labelled dataset, a small portion of the data is manually labelled, and the rest is treated as unlabelled for semi-supervised learning.

4.2 Data Preprocessing

The raw text data often contains noise, which must be cleaned before further processing. Preprocessing steps include:

- Converting all text to lowercase.
- Removing punctuation, special characters, and digits.

- Eliminating stop-words (common words that carry little meaning, e.g., “is,” “the,” “and”).
- Tokenizing the text into words.
- Applying stemming or lemmatization to reduce words to their base forms.

This step ensures uniform and clean data, which helps improve model performance.

4.3 Feature Extraction

After preprocessing, the textual data is transformed into numerical features suitable for machine learning algorithms. This is accomplished using:

- **TF-IDF (Term Frequency-Inverse Document Frequency):** Measures the importance of a word in a document relative to a collection of documents.
- **Behavioural Features:** If available, features like review length, user activity, and review timestamps can also be added.

The final feature set is represented as vectors that describe each review numerically.

4.4. Supervised Learning

A small portion of the dataset with labelled reviews is used to train a supervised learning model. Algorithms like Logistic Regression, Random Forest, or Support Vector Machines (SVM) are commonly used. The model learns to distinguish between fake and genuine reviews based on the extracted features.

4.5 Semi-Supervised Learning

To leverage the large amount of unlabelled data, a semi-supervised approach is employed. The system uses a method such as **Self-Training**, where:

- The supervised model is first trained on the labelled data.
- It then predicts labels for the unlabelled data.
- The predictions with high confidence are treated as pseudo-labels.
- These pseudo-labelled reviews are added to the training set.
- The model is retrained with the expanded dataset to improve generalization.

This process helps improve accuracy without requiring extensive manual labelling.

4.6 Classification and Prediction

After training, the model can classify new/unseen reviews as either fake or genuine. When a user inputs a review, it is pre-processed, converted to a feature vector, and passed through the trained model, which outputs a prediction.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

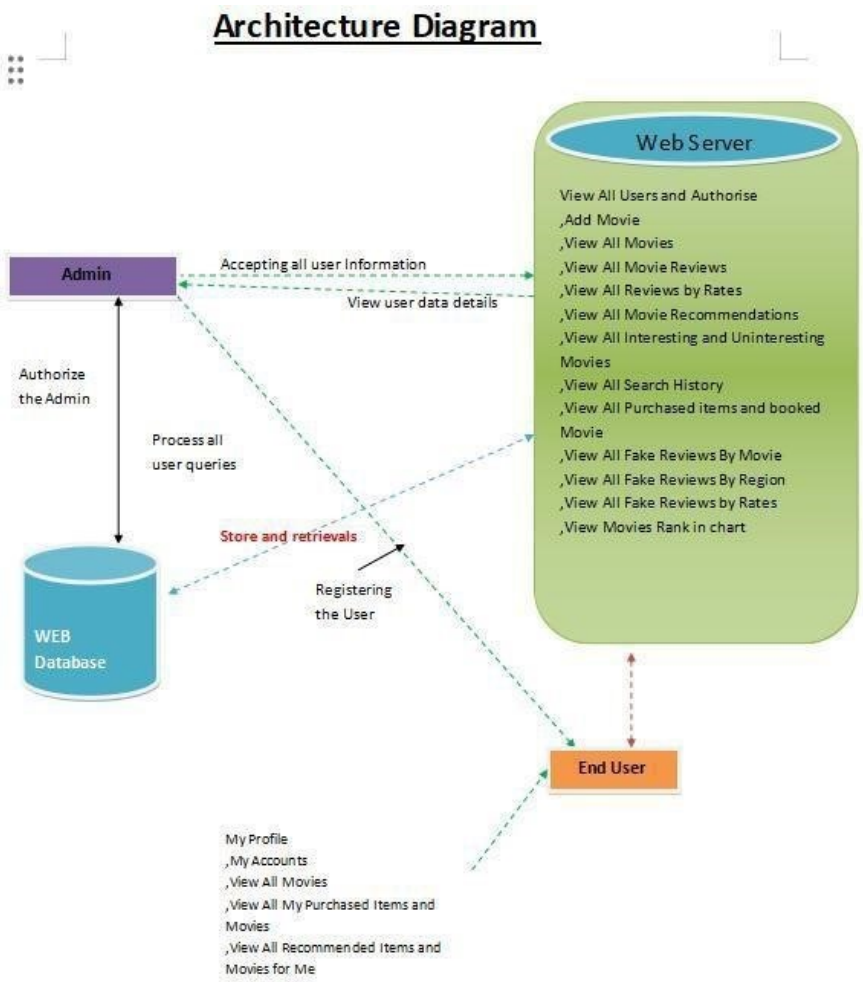
4.7 Evaluation

The model's performance is evaluated using standard classification metrics:

- **Accuracy:** Overall correctness.
- **Precision:** How many predicted fake reviews are actually fake.
- **Recall:** How many actual fake reviews were correctly identified.
- **F1-Score:** Harmonic mean of precision and recall.
- **Confusion Matrix:** Shows true vs. predicted labels.

These metrics help assess the effectiveness of the system and guide any necessary refinements.

SYSTEM ARCHITECTURE



INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

5. RESULT ANALYSIS

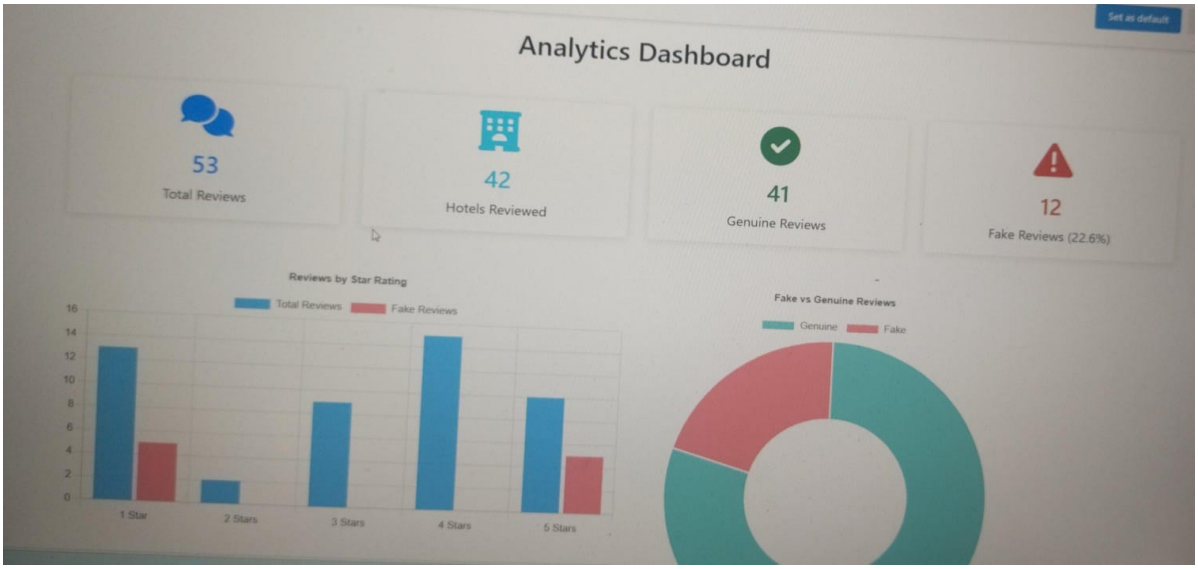


Fig:1 Analytics Dashboard of Reviews

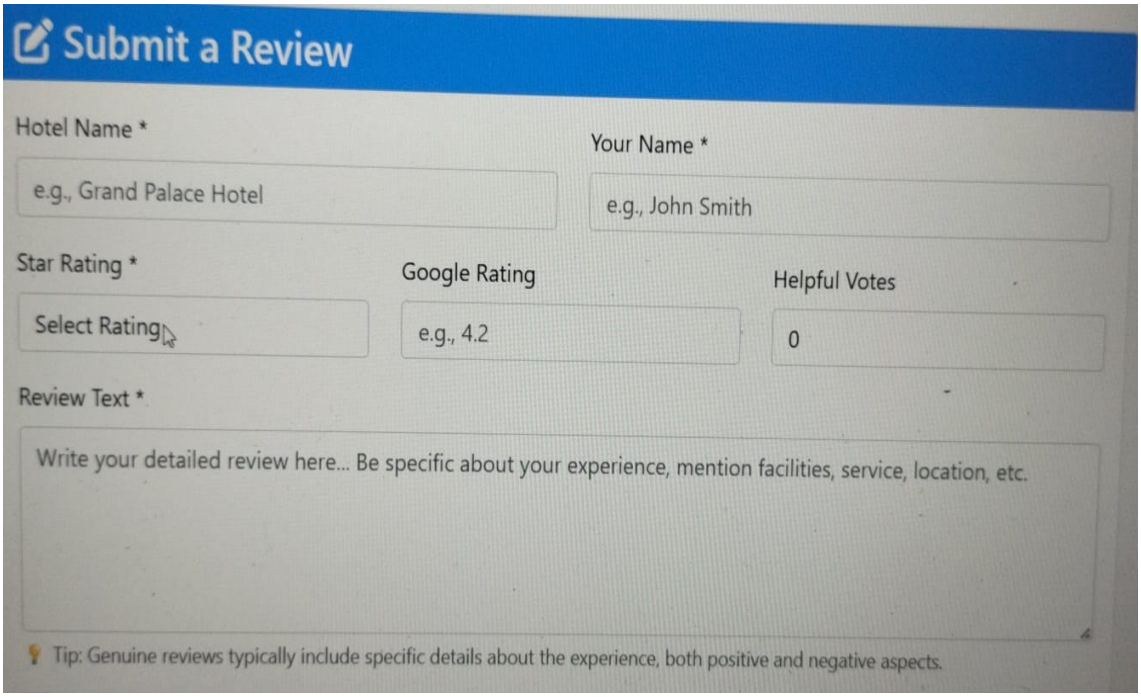


Fig:2 Reviews Textbox

6. CONCLUSION

In this project, we developed a machine learning-based system for detecting fake hotel reviews using both supervised and semi-supervised learning techniques. The system aims to address the growing issue of deceptive online reviews, which can mislead potential travellers and harm the credibility of hotel platforms. The implementation involved several critical stages including

data collection, text preprocessing, feature extraction, and model training. Supervised learning provided a solid foundation using a limited set of labelled reviews. To enhance performance and scalability, semi-supervised learning was incorporated—allowing the model to learn effectively from a large volume of unlabelled data through self-training. The system was evaluated using key performance metrics such as accuracy, precision, recall, and F1-score. Results showed that the semi-supervised model outperformed the supervised model by achieving higher detection accuracy and better generalization. This demonstrates the strength of leveraging unlabelled data in real-world applications where manually labelling large datasets is impractical. Overall, the project successfully shows that combining supervised and semi-supervised learning is a powerful approach to fake review detection. The developed system can assist online platforms in automatically filtering suspicious content, enhancing the trustworthiness of hotel reviews, and improving the decision-making experience for users.

7. REFERENCES

1 Chengai Sun, Qiaolin Du and Gang Tian, “Exploiting Product Related Review Features for Fake Review Detection,” Mathematical Problems in Engineering, 2016.

2 A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, ”Detection of review spam: a survey”, Expert Systems with Applications, vol. 42, no. 7, pp. 3634– 3642, 2015.

3 M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319,

Association for Computational Linguistics, Portland,Ore, USA, June 2011. 4 J. W. Pennebaker, M. E. Francis, and R. J. Booth, ”Linguistic Inquiry and Word Count: Liwc,” vol. 71, 2001.

5 S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deception detection,” in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2, 2012.