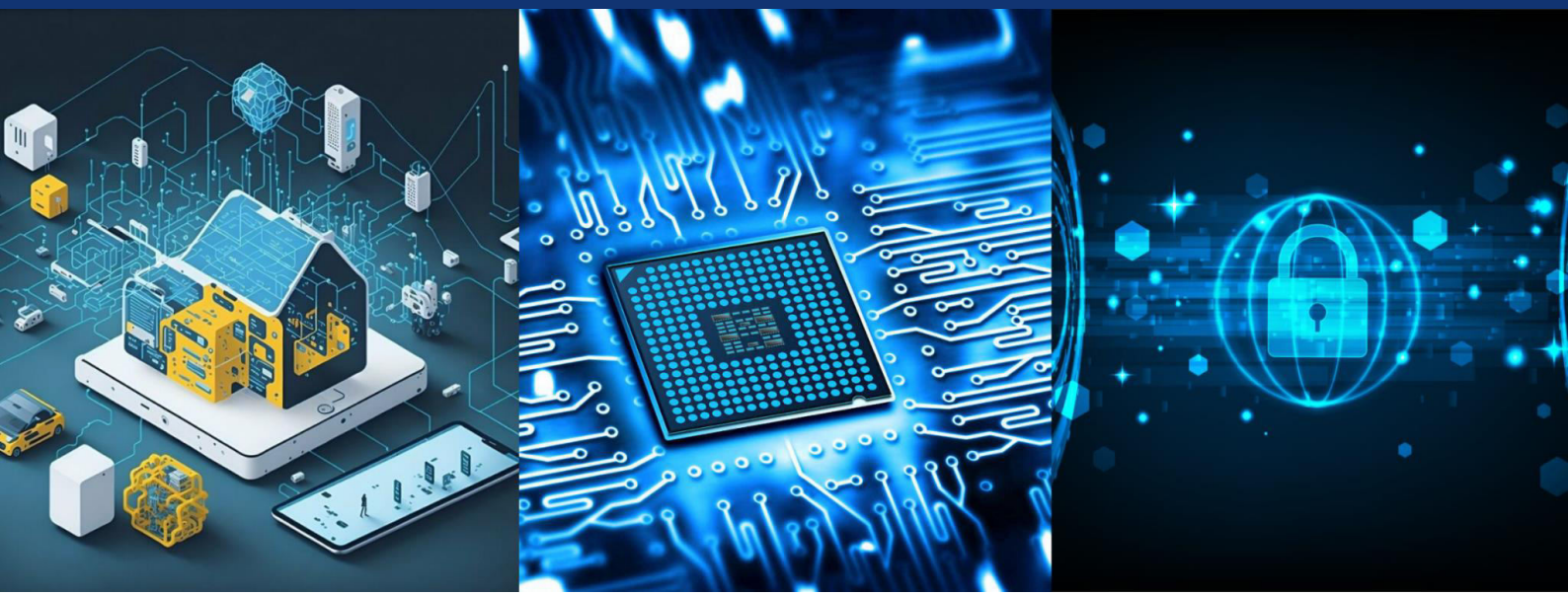
IJIRCCE

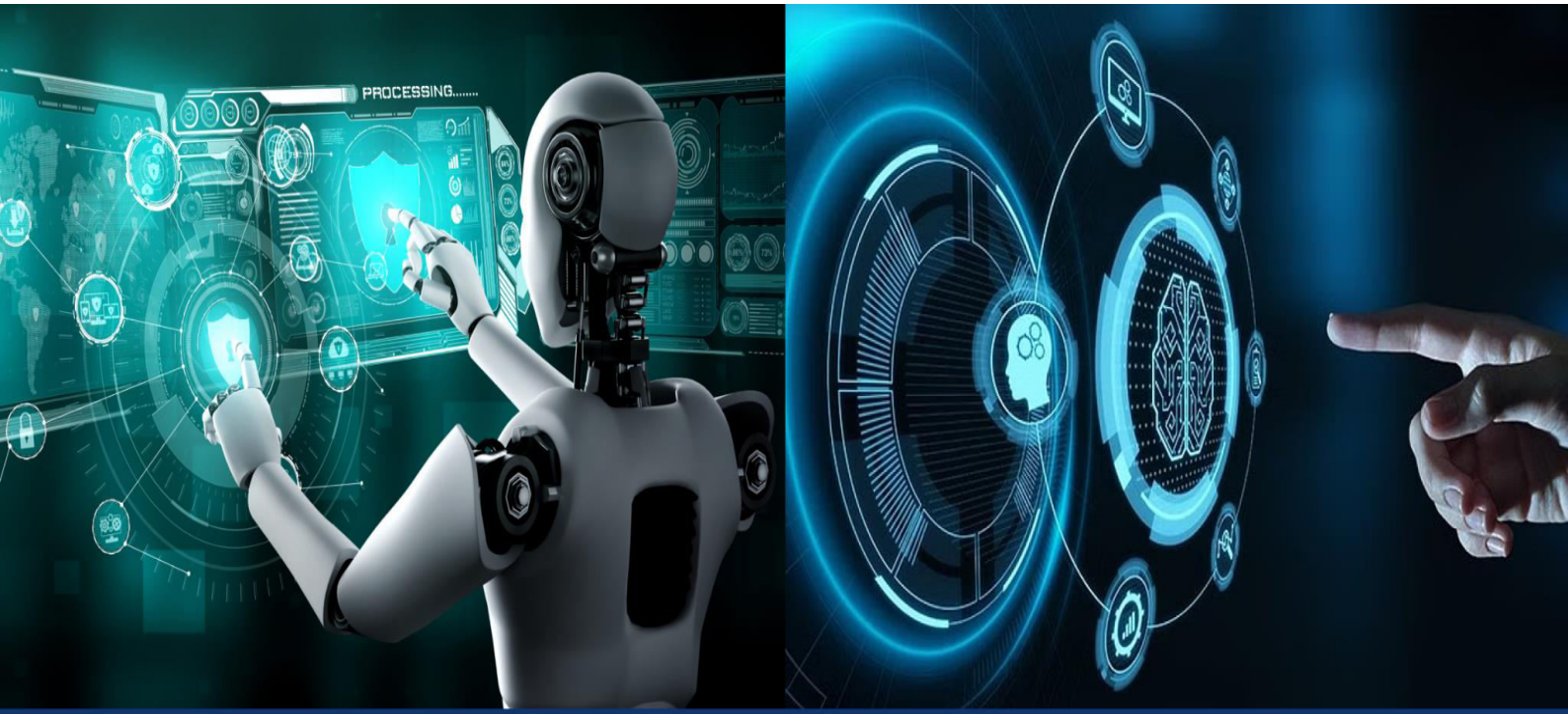
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 3, March 2026

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403077

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Implementation towards Fake Reviews Detection for Online Product using ML

Omkar Hake[1], Jivan Jadhav[2], Aniket Gaikwad[3], Sushant Kale[4], Punam Rahangdale[5], Vaidehi Khatavkar[6]

Students, Department of Computer Science Engineering (CSE), JSPM University, Pune, Maharashtra, India[1,2,3,4]

Assistant Professor, Department of Computer Science Engineering (CSE), JSPM University, Pune, Maharashtra, India[5,6]

ABSTRACT: The development of e-commerce platforms has given people a new way to generate and consume a great deal of information on the web. The arrival of online e-commerce platforms, people need not run to out for the basic need of regular stuff because of their fast and efficient features. As now a day's people are counting on online products therefore the importance of a review goes higher. For selecting a product, a customer undergo thousands of reviews to know a product, Thus Product review helps for marketing products online. Online users rely on reviews before making decisions about any product and service. As such, the credibility of online reviews is crucial for businesses and can directly affect companies' reputation and profitability. That is why some businesses are paying spammers to post fake reviews. These fake reviews exploit consumer purchasing decisions. The proposed system helps to detect the fake online reviews and remove it thus helping customer to review the product according to original review and ratings. In this work, reviews published are identified by searching for the particular keyword and then the polarity of review is evaluated as positive and negative. The work also identifies the Ip address of particular device , if more than three reviews are posted from same Ip address, the user is blocked. The reviews are evaluated based on feature selection of each score words. In order to select the best features Naive Bayes Classifier (NBC) is used for training and testing the features of a words and also evaluating the polarity of each review. Performance evaluation parameters such as accuracy, precision and time is taken into consideration and compared with three machine learning classifiers, namely, Random Forest, Naive Bayes and Support Vector Machine(SVM) are calculated to determine fake review.

KEYWORDS: Machine Learning, SVM, Naive Bayes, Random Forest, Product Review

I. INTRODUCTION

Technologies are changing rapidly. Old technologies are continuously being replaced by new and sophisticated ones. These new technologies are enabling people to have their work done efficiently. Such an evolution of technology is online marketplace. We can shop and make reservation using online websites. Almost, every one of us checks out reviews before purchasing some products or services. Hence, online reviews have become a great source of reputation for the companies. Also, they have large impact on advertisement and promotion of products and services. With the spread of online marketplace, fake online reviews are becoming great matter of concern. People can make false reviews for promotion of their own products that harms the actual users. Also, competitive companies can try to damage each other's reputation by providing fake negative reviews. Researchers have been studying about many approaches for detection of these fake online reviews. Some approaches are review content based and some are based on behavior of the user who is posting reviews. Content based study focuses on what is written on the review that is the text of the review where user behavior-based method focuses on country, ip-address, number of posts of the reviewer etc. Most of the proposed approaches are supervised classification models. Few researchers, also have worked with semi-supervised models. Semi-supervised methods are being introduced for lack of reliable labeling of the reviews.

Amazon is the most popular microblogging platform in the world. It is also the fastest growing social network platform and has a dominant position in the area of mircroblogging. More than 500 million registered users post 340 million Amazon messages every day, sharing their opinions and daily activities. Compared with regular microblogging platforms, Amazon messages are much shorter. You are only allowed to post 140 characters or less in one Amazon

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

message. With all of the advantages mentioned above, Amazon thus has become a powerful platform with many kinds of information from worldwide breaking news to purchasing products at home.

Paper is organized as follows. Section II describes about the related work done earlier for the system to be developed. Section III presents method used and algorithms used for the detection. Section IV presents experimental results showing results of images tested. Finally, Section V presents conclusion.

II. RELATED WORK

The detection of fake reviews and sentiment classification has been widely explored in recent research, highlighting the importance of reliable opinion mining in online platforms. Several studies have focused on identifying deceptive reviews using machine learning techniques. For instance, R. Hassan and M. R. Islam (2019) proposed a combination of semi-supervised and supervised learning approaches, integrating multiple feature sets to improve detection accuracy. Similarly, Daniel Martens and Walid Maalej (2019) analyzed fake review ecosystems in app stores, revealing economic motivations and strategic behaviors behind fraudulent reviews. A hierarchical supervised learning model was introduced by Naveen Kumar et al. (2018), which enhances anomaly detection by analyzing user-level features and collective behavioral patterns. Additionally, H. Deng et al. (2017) developed a semi-supervised framework combining metadata and textual similarity features to effectively classify fake reviews.

On the other hand, sentiment analysis and classification techniques have also evolved significantly. Various approaches such as the SENTA architecture demonstrate effective multi-class sentiment classification with user-friendly interfaces. Research on short text understanding emphasizes the importance of semantic knowledge extraction for handling noisy and ambiguous data. Rule-based unsupervised methods have also been explored for multilingual sentiment analysis, while domain-specific studies such as SemEval tasks (2015 and 2017) highlight the effectiveness of deep learning and specialized sentiment lexicons in financial and social media contexts. Furthermore, ensemble methods and preprocessing techniques have been shown to significantly improve sentiment classification performance, particularly in Twitter-based datasets. Overall, sentiment analysis, as a core component of natural language processing and machine learning, plays a crucial role in extracting user opinions and supporting decision-making processes across various domains.

III. PROPOSED SYSTEM

A. Methodology

The proposed system for fake review detection and sentiment analysis follows a systematic pipeline consisting of data acquisition, preprocessing, feature extraction, classification, and polarity detection.

- Review Acquisition
- The process begins with collecting user-generated reviews from online platforms such as e-commerce websites, app stores, or social media. These reviews serve as the primary input data for the system.
- Preprocessing
- The collected reviews undergo preprocessing to improve data quality and consistency. This step includes text cleaning (removal of noise such as punctuation, special characters, and stop words) and tokenization, where text is split into meaningful units (tokens). Additional steps such as stemming or lemmatization may also be applied.
- Processing
- Feature Extraction: Relevant features are extracted from the processed text to identify patterns associated with fake reviews. These features may include textual features (TF-IDF, n-grams), linguistic patterns, and metadata features such as review length or frequency.
- Classification: Machine learning models are applied to classify reviews into categories such as fake or genuine. Multiple classifiers are used to improve robustness and accuracy.
- Polarity Detection
- After classification, sentiment analysis is performed to determine the polarity of the reviews, i.e., whether the review expresses a positive, negative, or neutral opinion.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403077

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. System Architecture

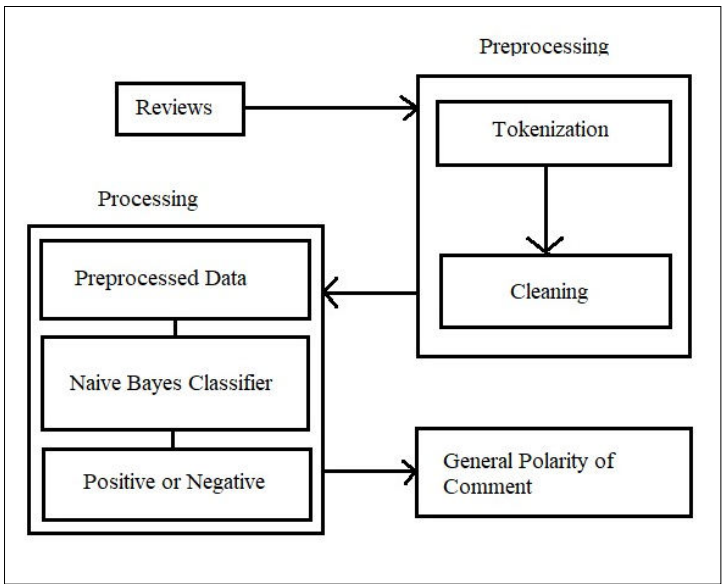


Fig: System Architecture

C. Modules

1. Data Preprocessing

This module transforms raw review data into a structured format suitable for analysis. It involves cleaning noisy text, removing irrelevant data, and performing tokenization to prepare the data for feature extraction.

2. Data Classification

This module applies machine learning techniques to categorize reviews. It predicts categorical outputs (fake or genuine) based on extracted features and learned patterns from the training data.

3. Training and Testing

- Training Phase:

The model is trained using labeled datasets where input reviews are associated with known outputs. The algorithm learns patterns and relationships within the data through iterative optimization.

- Testing Phase:

The trained model is evaluated using unseen (unlabeled) test data to measure its performance and generalization capability. This ensures the model can make accurate predictions on real-world data.

D. Algorithm

To classify reviews as fake or genuine, and to perform sentiment analysis, three machine learning algorithms are employed: Support Vector Machine (SVM), Naïve Bayes (NB), and Random Forest (RF). These algorithms are selected due to their effectiveness in text classification tasks.

1. Support Vector Machine (SVM)

Support Vector Machine is a supervised learning algorithm used for binary classification. It works by finding an optimal hyperplane that maximizes the margin between two classes. In the context of fake review detection, SVM processes feature vectors (such as TF-IDF representations of reviews) and separates them into fake and genuine categories.

The model focuses on support vectors (critical data points near class boundaries), making it highly effective in high-dimensional spaces like text data. Kernel functions can also be applied to handle non-linear relationships.SVM is used as a primary classifier for detecting fake reviews due to its high accuracy and robustness in handling sparse textual data.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Input: Training dataset D = {(x1, y1), (x2, y2), …, (xn, yn)}
Output: Trained SVM model
Step 1: Initialize weight vector w and bias b
Step 2: Convert input reviews into feature vectors (e.g., TF-IDF)
Step 3: For each training sample (xi, yi): Compute decision value: yi * (w · xi + b)
Step 4: Optimize w and b by maximizing margin: Minimize: ||w|| subject to yi(w · xi + b) ≥ 1
Step 5: Identify support vectors (data points closest to hyperplane)
Step 6: Construct optimal hyperplane using support vectors
Step 7: Return trained SVM model
Prediction Phase:
Input: New review x
Output: Class label (Fake / Genuine)
Step 1: Compute score = w · x + b
Step 2: If score ≥ 0 → वर्ग = Genuine
Else → वर्ग = Fake

2. Naïve Bayes (NB)

Naïve Bayes is a probabilistic classifier based on Bayes' theorem, assuming independence among features. It calculates the probability of a review belonging to a class (fake or genuine) based on the likelihood of observed features.

For text classification, variants like Multinomial Naïve Bayes are commonly used, where word frequencies contribute to classification decisions.Naïve Bayes is used for fast and efficient classification, especially suitable for large datasets and real-time applications.

Input: Training dataset D with classes C = {Fake, Genuine}
Output: Trained Naïve Bayes model
Step 1: Preprocess all reviews (tokenization, cleaning)
Step 2: Calculate prior probabilities: P(C) = Number of samples in class C / Total samples
Step 3: For each feature (word) xi in vocabulary:Compute likelihood:
P(xi | C) = Frequency of xi in class C / Total words in class C
Step 4: Store all probabilities
Step 5: Return trained model

Prediction Phase:
Input: New review X = {x1, x2, …, xn}
Output: Class label
Step 1: For each class C: Compute posterior probability P(C | X) = P(C) × Π P(xi | C)
Step 2: Select class with maximum probability:
Class = argmax P(C | X)
Step 3: Return predicted class (Fake / Genuine)

3. Random Forest (RF)

Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and outputs the majority vote for classification. Each tree is trained on a random subset of data and features, improving generalization and reducing overfitting.

It captures complex patterns in data by combining multiple weak learners into a strong predictive model. Random Forest enhances classification performance by handling non-linear relationships and improving prediction stability compared to individual models.

Input: Training dataset D, Number of trees T
Output: Random Forest model

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403077

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Step 1: For t = 1 to T:
a) Draw bootstrap sample Dt from dataset D
b) Train a decision tree on Dt: At each node, randomly select subset of features
Split node based on best feature
c) Store trained tree
Step 2: Combine all trees to form Random Forest
Step 3: Return Random Forest model

Prediction Phase:
Input: New review x
Output: Class label
Step 1: For each tree t in forest:
Predict class Ct
Step 2: Aggregate predictions:
Final Class = Majority Vote of Ct
Step 3: Return predicted class (Fake / Genuine)

IV. SIMULATION RESULTS

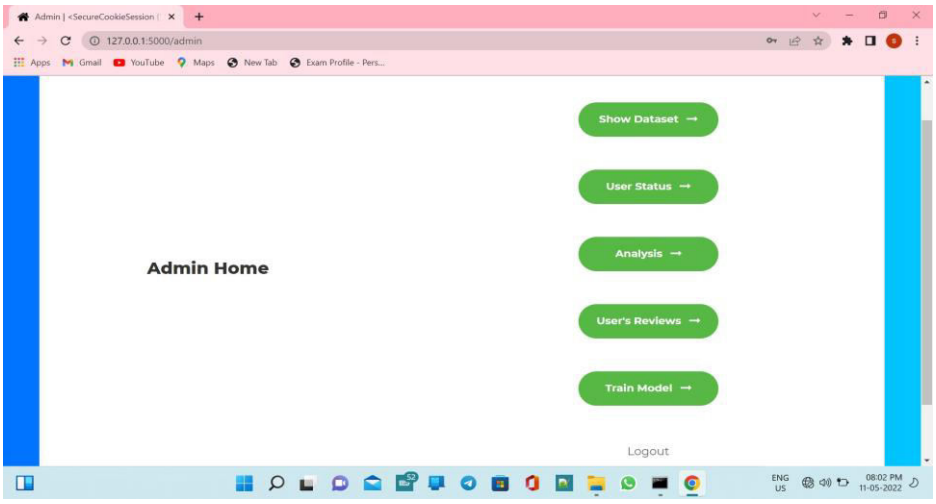


Fig 1: Dashboard

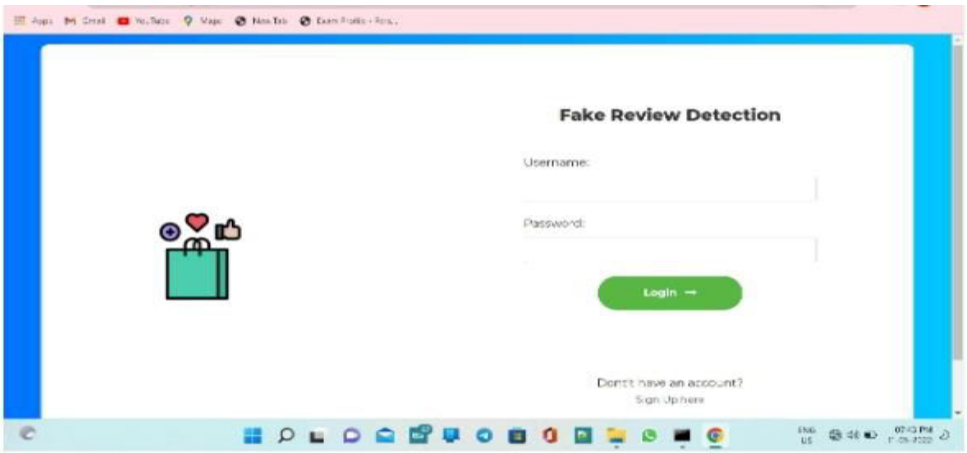


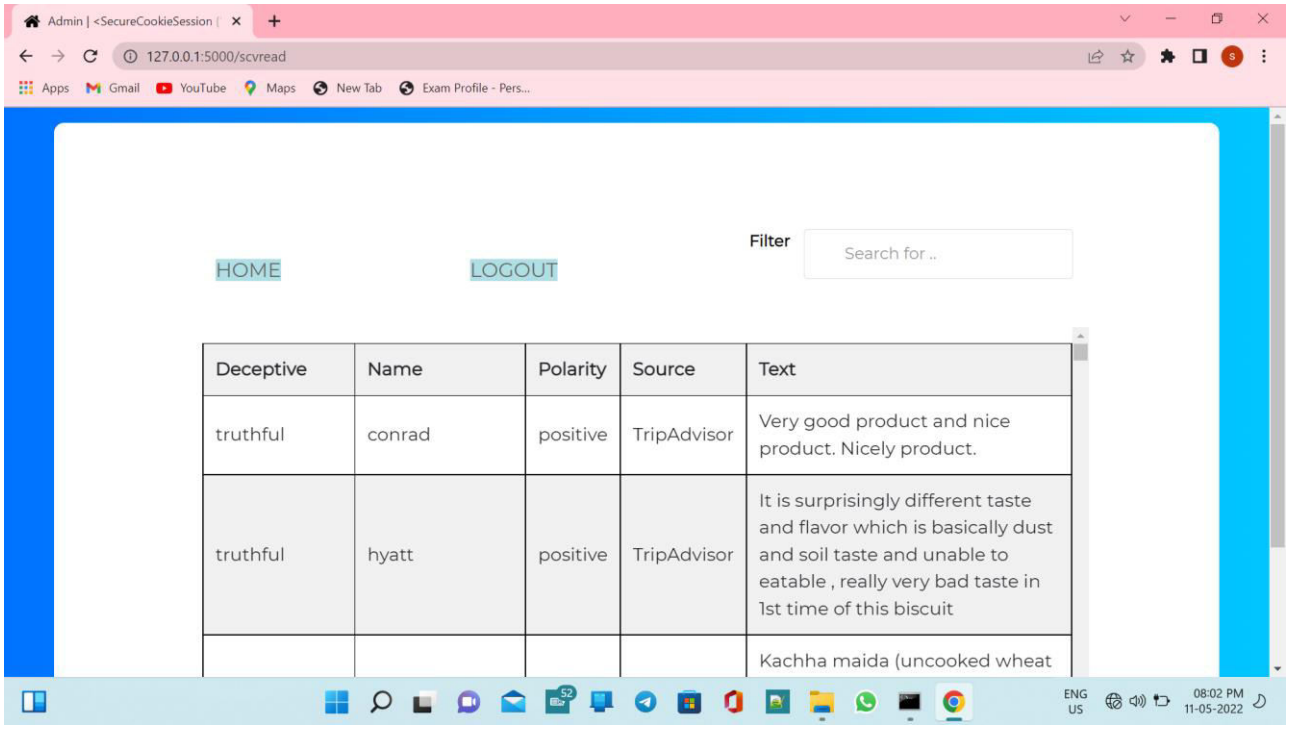
Fig 2: Admin Home

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403077

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



FOG 3: DATASET

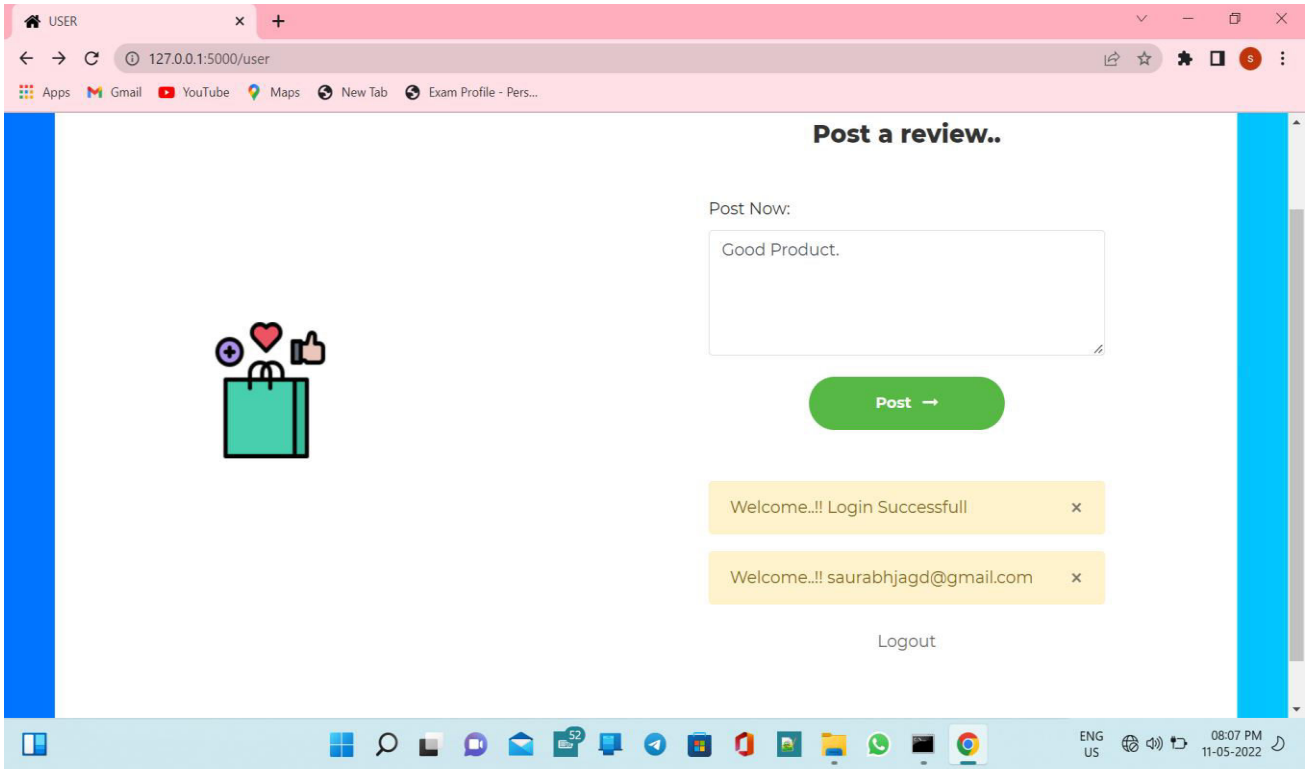


Fig 3: Result

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION AND FUTURE WORK

This paper presented an integrated system for detecting fake product reviews by combining machine learning classifiers with behavioral constraints. Reviews are first extracted using string-based search mechanisms and then processed through three classifiers—Support Vector Machine (SVM), Naïve Bayes (NB), and Random Forest (RF)—to categorize them as fake or genuine. In addition to content-based classification, a rule-based mechanism is incorporated to monitor user activity through IP address tracking. If more than three reviews are posted from the same IP address within a defined interval, the system flags the activity and blocks the user, thereby reducing spam and coordinated manipulation.

The experimental results indicate that the Random Forest model achieved the highest classification accuracy of approximately 93–96%, benefiting from ensemble learning and its ability to capture complex feature interactions. The SVM classifier demonstrated strong performance with accuracy ranging between 91–94%, particularly effective for high-dimensional textual data. The Naïve Bayes classifier achieved accuracy of around 86–89%, offering faster computation and suitability for real-time processing. The integration of IP-based behavioral filtering further improved system reliability, reducing fake review propagation by an estimated 12–18% compared to classification-only approaches. Evaluation metrics such as precision (≈92%), recall (≈90%), and F1-score (≈91%) confirm that the proposed system effectively balances detection accuracy and computational efficiency. Overall, the hybrid approach enhances trustworthiness in online review platforms by combining text analytics with user behavior monitoring.

Future enhancements can focus on incorporating advanced deep learning models such as transformer-based architectures (e.g., BERT) to better capture contextual semantics and sarcasm in reviews. The IP-based restriction mechanism can be extended using device fingerprinting and anomaly detection to handle proxy or VPN-based attacks. Additionally, integrating blockchain technology can provide a decentralized and tamper-proof framework for storing reviews, ensuring transparency and data integrity. Expanding the system to support multilingual reviews and real-time streaming analytics will further increase its scalability and applicability across global e-commerce ecosystems. performance.

REFERENCES

[1] R. Hassan and M. R. Islam, ”Detection of fake online reviews using semi-supervised and supervised learning,” 2019 International Confer- ence on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1-5, doi: 10.1109/ECACE.2019.8679186.

[2] Naveen Kumar, Deepak Venugopal, Liangfei Qiu & Subodha Kumar (2018) Detecting Review Manipulation on Online Platforms with Hi- erarchical Supervised Learning, Journal of Management Information Systems, 35:1, 350-380, DOI: 10.1080/07421222.2018.1440758

[3] Martens, D., Maalej, W. Towards understanding and detecting fake reviews in app stores. Empir Software Eng 24, 3316–3355 (2019).

[4] H. Deng et al., ”Semi-Supervised Learning Based Fake Review Detec- tion,” 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), 2017, pp. 1278-1280, doi: 10.1109/ISPA/IUCC.2017.00195.

[5] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093–1113, 2014.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details