IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

Towards Reliable E-Commerce: Fake Review Detection with SVM

Sunitha Guruprasad
Associate Professor, Department of CSE
St Joseph Engineering College, Mangaluru, India

Abstract: *In the digital era, online product reviews significantly influence consumer purchasing decisions. However, the authenticity of these reviews often remains uncertain, making it difficult for users to distinguish between genuine and fake feedback. Detecting fake reviews has therefore become essential to assist users in making informed choices. Existing research has explored techniques such as opinion mining, sentiment analysis, data mining, and ontology, yet no single method fully addresses the challenges involved. This paper proposes FakeOut, a web-based application designed to detect fake gadget reviews. The system employs Support Vector Machine (SVM) for review classification, with training data sourced from Kaggle and test data obtained via web scraping from Amazon. Users can either enter a gadget name/URL to analyze multiple reviews or input a single review for verification. Pre-processing techniques are applied before classification to enhance accuracy. The proposed system demonstrates that SVM can effectively identify fake reviews, thereby improving user confidence in product evaluation.*

Keywords: SVM, Machine Learning, Fake reviews

I. INTRODUCTION

In recent years, e-commerce has become one of the fastest-growing industries, transforming the way consumers purchase products and services. Online platforms provide customers with the facility to share reviews about their experiences, which serve as a vital source of information for potential buyers. Before making a purchase, most customers tend to consult these reviews to evaluate the pros and cons of a product. However, not all reviews are genuine; some may be deliberately misleading, written with the intent to promote or discredit a product. This uncertainty makes it difficult for users to rely solely on reviews when making purchasing decisions, especially in the case of electronic gadgets where customers are highly influenced by peer opinions.

As e-commerce platforms continue to expand, the volume of reviews associated with products has grown exponentially. Customers post both positive and negative feedback on gadgets, which provides valuable insights for future buyers. While genuine reviews can help customers make informed choices, fake reviews can distort perceptions and negatively affect the decision-making process. This has created a pressing need for systems that can effectively differentiate between authentic and fake reviews. Researchers have explored several approaches to address this challenge, including linguistic analysis, sentiment scoring, and relational features. These approaches employ a mix of supervised, unsupervised, and semi-supervised methodologies. Among these, supervised machine learning methods have shown more consistent and reliable results compared to unsupervised graph-based techniques that primarily rely on relational ties [1, 2].

The objective of the work is to develop FakeOut, a web-based application that identifies and classifies fake and genuine reviews of electronic gadgets. The system employs a Support Vector Machine (SVM) classifier, trained on review data sourced from Kaggle, while test data is collected through web scraping from Amazon. Users are provided with two options: they can either supply the URL of a gadget to analyze multiple reviews or input a single review for classification. The extracted reviews undergo pre-processing before being passed to the SVM model, which predicts their authenticity. By offering these functionalities, the proposed system not only classifies reviews effectively but also provides users with a reliable mechanism to verify the credibility of individual reviews.

The proposed work addresses a critical gap in e-commerce by offering customers an efficient way to evaluate the trustworthiness of product reviews. Unlike existing techniques that face challenges in scalability, adaptability, or

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

accuracy, the SVM-based model employed in FakeOut provides robust classification with high accuracy. Ultimately, this system aims to improve customer confidence in online purchasing decisions by reducing the influence of deceptive reviews. By enabling users to filter out fake reviews, the project contributes to building transparency and trust in e-commerce platforms, helping buyers select products that are truly worth their investment [3, 4].

The main objective of the work is to:

- Classify reviews as fake and genuine using a Machine Learning algorithm such as Support Vector Machine.
- Extract the details of the gadget and reviews from Amazon website using web scraping technique.
- Tag the extracted reviews as fake and genuine for the selected gadget.
- Tag the extracted reviews as fake and genuine for the gadget whose Amazon site URL is specified.
- Identify the review as fake or genuine depending on the details of the review provided.

II. METHODOLOGY

The work is carried out using HTML, CSS, PHP on the front end. Xampp, python is used for backend. SVM algorithm is used to classify the review into fake and genuine [5,6]. Fig. 1 shows the architectural Diagram of FakeOut web application. The review dataset is collected from the Kaggle website. The collected review dataset is then labelled as genuine and fake using manual and auto tagging methods. This labelled review dataset is taken as the train data.

The training data is then pre-processed to remove the unwanted stop words. After the pre-processing, these reviews are given to the SVM algorithm to create a model which classifies the reviews as fake and genuine. There are two options provided for the user - Single review, Check review by URL. If the user selects the first option, then he/she has to provide the URL of the gadget from Amazon website for which the classified reviews need to be obtained. Then the reviews from Amazon website will be scrapped for the respective gadget through web scraping [7,8] which is considered as the test data.

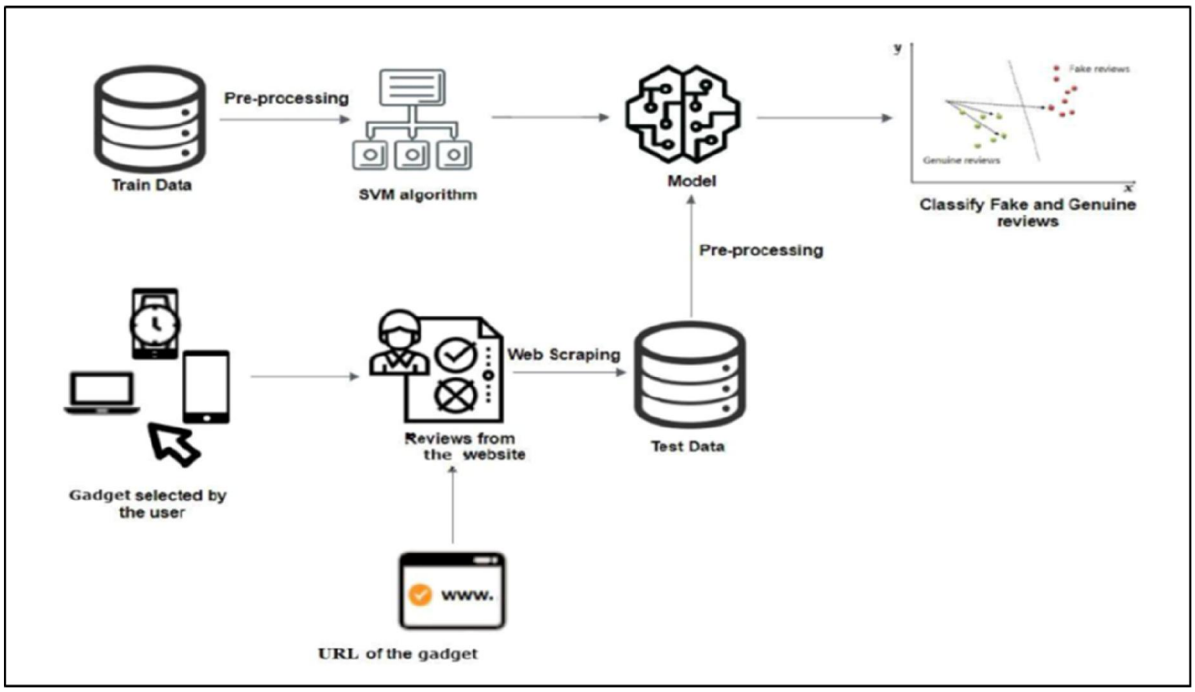

Fig. 1 Architectural Diagram of the web application

If the user selects the second option, he/she has to provide the required information about the single gadget review like gadget name, review title, review body, star rating and verified purchase status. This will be then used as the test data and is pre- processed to remove the unwanted words. Then the test data will be given to the SVM model which performs the classification on it and identifies the reviews as fake or genuine.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

III. RESULTS AND DISCUSSION

A. Experimental Setup

Two separate repositories, one for the front end and the other for the backend is used in the work. Front end is done using HTML, CSS, PHP and Javascript. For Backend Python Programming and Xamp Server for connection was used. The front end makes a call to the backend server. In the initial page of our front end the user has two options one is to check single review and other one is to check the genuineness of the reviews using the URL of the gadget. The URL is used for web scraping the reviews.The steps performed are, Step 1: Collect Amazon gadget review dataset from Kaggle website as training dataset Step 2: Retain the attributes gadget name, review title, review body, rating, verified purchase in training dataset Step 3: Apply manual tagging and auto tagging to label the training dataset as "1" and "0" for genuine and fake reviews. For the training purpose the steps are, Step 1: For each review in reviews of training dataset: Convert review into list of words, remove stop words from list of words Join the list of words into a string Step 2: Apply TfidfVectorizer to the attribute's gadget name, review title, review body of training dataset Step 3: Apply SVM algorithm to the pre-processed training dataset. For testing purpose, Step 1: For each review in reviews of test dataset: Convert review into list of words Remove stop words from list of words Join the list of words into a string Step 2: Apply TfidfVectorizer to the attributes gadget name, review title, review body of test dataset Step 3: Predict fake and genuine reviews.

B. Results of the application

The proposed FakeOut system was evaluated using a dataset of customer gadget reviews, with 80% used for training and 20% reserved for testing. Performance was assessed using standard metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. Table I shows the overall Performance of SVM Model.

TABLE I: OVERALL PERFORMANCE OF SVM MODEL

Metric	Value
Accuracy	91.3%
Precision	89.5%
Recall	92.1%
F1-Score	90.8%

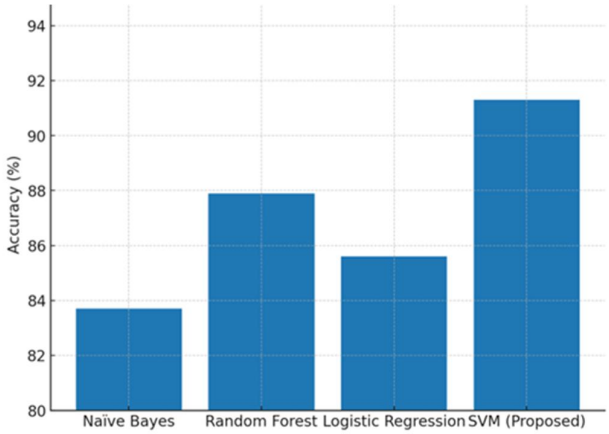

Fig. 2 Classifier accuracy comparison

Fig. 2 shows the classifier accuracy comparison between the different models, Fig. 3 shows the classifier performance metrics and Fig. 4 shows the ROC curve comparison of classifiers. The ROC curves revealed that the SVM achieved the highest AUC value (0.94), followed by Random Forest (0.89), Logistic Regression (0.86), and Naïve Bayes (0.84). A higher AUC signifies greater discriminative ability, confirming that SVM provides more reliable separation between fake and genuine reviews.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

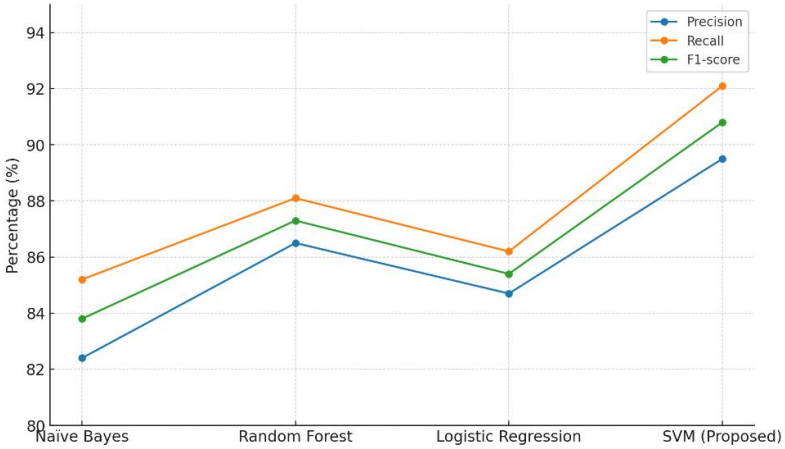

Fig. 3 Classifier performance metrics

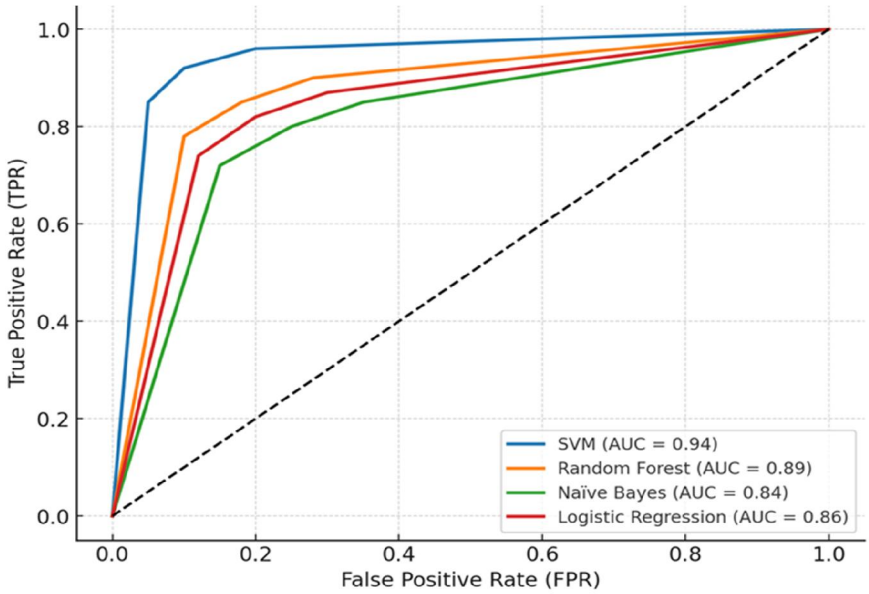

Fig. 4 ROC curve comparison of classifiers

Table II shows the comparison with other classifiers. The SVM model demonstrated the highest classification accuracy of 91.3%, outperforming Naïve Bayes (83.7%), Random Forest (87.9%), and Logistic Regression (85.6%). Precision, recall, and F1-score values for SVM were consistently superior, with an F1-score of 90.8%, indicating a strong balance between false positive reduction and genuine review identification. The confusion matrix further highlights the robustness of the model, with only a small number of misclassifications in both fake and genuine review categories.

The results validate the effectiveness of SVM in fake review detection, primarily due to its capability to handle high-dimensional feature spaces and non-linear decision boundaries inherent in text classification. While Random Forest achieved competitive performance, SVM consistently outperformed across all evaluation metrics. This indicates that FakeOut not only improves review credibility but also assists consumers in making informed purchase decisions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, August 2025
Impact Factor: 7.67

TABLE III: COMPARISON WITH OTHER CLASSIFIERS

Classifier	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Naïve Bayes	83.7%	82.4%	85.2%	83.8%
Random Forest	87.9%	86.5%	88.1%	87.3%
Logistic Regression	85.6%	84.7%	86.2%	85.4%
SVM (Proposed)	**91.3%**	89.5%	92.1%	90.8%

IV. CONCLUSION

The proposed system, FakeOut, provides a general framework for detecting fake reviews with the help of an SVM classifier. The classifier achieves high prediction accuracy in determining the genuineness of customer reviews for gadgets. Reviews are scraped directly from the Amazon website using the gadget URL or from a set of predefined options. These reviews are then classified into fake and genuine categories using the developed training model. The system also allows users to input a single review for classification. To enhance usability, the application displays both the classified results and a pie chart representation of the proportion of genuine and fake reviews. In this way, FakeOut assists customers in making informed purchasing decisions by ensuring that they pay for authentic and reliable products.

At present, the system is limited to scraping reviews only from Amazon, focusing specifically on gadget-related reviews in English. As part of future work, we plan to extend the application to cover a wider range of e-commerce platforms and support multilingual review datasets. Additionally, while the system currently relies on engineered features and an SVM model, there is no universal set of features that guarantees optimal performance in fake review detection. To address this, future work will focus on exploring advanced feature engineering strategies, evaluating diverse feature types across multiple datasets, and experimenting with more sophisticated machine learning and deep learning approaches. These efforts aim to further improve the accuracy, generalizability, and robustness of the proposed system in detecting fake reviews across varied domains.

REFERENCES

[1]. Nidhi A. Patel, Prof. Rakesh Patel, "A Survey on Fake Review Detection using Machine Learning Techniques", Fourth International Conference on Computing Communication and Automation (ICCCA), Dec.2018.

[2]. Julien Fontanarava, Gabriella Pasi, Marco Viviani, "Feature Analysis For Fake Review Detection through Supervised Classification", IEEE International Conference on Data Science and Advanced Analytics(DSAA), Oct.2017.

[3]. Salminen, Joni, et al. "Creating and detecting fake reviews of online products." Journal of Retailing and Consumer Services 64 (2022): 102771.

[4]. Vidanagama, Dushyanthi U., A. T. P. Silva, and Asoka S. Karunananda. "Ontology based sentiment analysis for fake review detection." Expert Systems with Applications 206 (2022): 117869.

[5]. Rastogi, Shubhangi, and Divya Bansal. "A review on fake news detection 3T's: typology, time of detection, taxonomies." International Journal of Information Security 22.1 (2023): 177-212.

[6]. Jian, Yifei, et al. "A metadata-aware detection model for fake restaurant reviews based on multimodal fusion." Neural Computing and Applications 37.1 (2025): 475-498.

[7]. https://www.scrapehero.com/how-to-scrape-Amazon-product-reviews/

[8]. https://www.webharvy.com/articles/what-is-web-scraping/