© DEC 2025 | IRE Journals | Volume 9 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I6-1712519

Detecting Online Fake Reviews Using Supervised and Semi Supervised Learning.

AKSHAYA D, DANIYA BEGUM[1], NISARGA H N, SYED SAIFULLA[2]
[1, 2]*Department of Computer Science & Ghousia college of Engineering Ramanagara, Karnataka, India*

Abstract- ***Online reviews play a crucial role in influencing consumer decisions, which makes them a target for manipulation through fake or misleading feedback. Detecting such deceptive reviews is challenging because fraudulent content is often written to closely resemble genuine opinions. This project presents a hybrid approach for *detecting online fake reviews using supervised and semisupervised machine learning techniques*. The system 254 abelled both linguistic features and reviewer behavior to classify reviews as genuine or deceptive. Supervised learning models such as Support Vector Machines, Logistic Regression, and Random Forest are trained on 254 abelled datasets to establish a strong baseline. To address the scarcity of high-quality 254abelled data, semisupervised methods—including Self-Training and Label Propagation—are integrated to utilize large amounts of unlabeled reviews. This combination enhances model robustness and improves detection accuracy in real-world scenarios. Experimental results demonstrate that the semi-supervised models significantly improve performance, especially when 254abelled data is limited. The proposed hybrid approach offers an effective and scalable solution for identifying deceptive content, helping e-commerce platforms protect customers and maintain trust.***

I. INTRODUCTION

Online reviews have become a primary source of information for customers before making purchasing decisions. Their influence has led to a rapid rise in deceptive or fake reviews created to artificially promote or defame products and services. These manipulated reviews distort consumer trust and negatively impact e-commerce platforms.

Detecting fake reviews is challenging because deceptive reviews often resemble genuine ones and may be generated by humans or bots. Traditional rule-based detection methods are insufficient due to evolving writing styles and reviewer patterns. This project addresses the problem by developing a system that uses *supervised and semisupervised machine learning techniques* to identify fake reviews with improved precision and scalability

II. LITERATURE REVIEW

Fake review detection has been explored using various computational approaches:

Text-based approaches:
Research shows that deceptive reviews often differ from genuine ones in terms of sentiment intensity, vocabulary richness, emotional tone, and writing style. TF-IDF, ngrams, POS tagging, and sentiment scores are commonly used linguistic features.

Behavioral and metadata-based approaches:
Studies indicate that reviewer patterns—such as account age, review frequency, posting time, and rating deviation—can strongly signal suspicious behavior. Graph-based features such as reviewer–item networks have also been applied.

Supervised learning approaches:
Models like Logistic Regression, SVM, Random Forest, XGBoost, and Neural Networks have shown high performance but require large labeled datasets, which are expensive to obtain.

Semi-supervised learning approaches:
Methods such as Label Propagation, Label Spreading, Self-Training, and Co-Training allow the use of unlabeled data to improve model robustness. These approaches are becoming increasingly important as fake reviews evolve and 254 abelled datasets remain limited.

III. PROPOSED SYSTEM

To solve the major problem faced by online websites due to opinion spamming, this project proposes to identify any such spammed fake reviews by classifying them into fake and genuine. The method attempts to classify the reviews obtained from freely available datasets from various sources and categories including service based, product based, customer feedback, experience based and the crawled

© DEC 2025 | IRE Journals | Volume 9 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I6-1712519

Amazon dataset with a greater accuracy using Naïve Bayes [7], Linear SVC, SVM and Logistic regression algorithms. In order to improve the accuracy, the additional

IV. PROPOSED METHODOLOGY

A classifier is built based on the identified features. And those features are assigned a probability factor or a weight depending on the classified training sets. This is a supervised learning technique applying different Machine learning algorithms to detect the fake or genuine reviews.

V. SYSTEM DESIGN

Design is a meaningful engineering representation of something that is to be built. It is the most crucial phase in the development of a system. Software design is a process through which the requirements are translated into a representation of software. Design is a place where design is fostered in software Engineering. Based on the user requirements and the detailed analysis of the existing system, a new system must be designed. This is the phase of system designing. Design is the perfect way to accurately translate a customer's requirement in the finished software product. The design creates a representation or model, and provides details about software data structure, architecture, interfaces and components that are necessary to implement a system. The logical system design arrived at as a result of systems analysis is converted into physical system design.

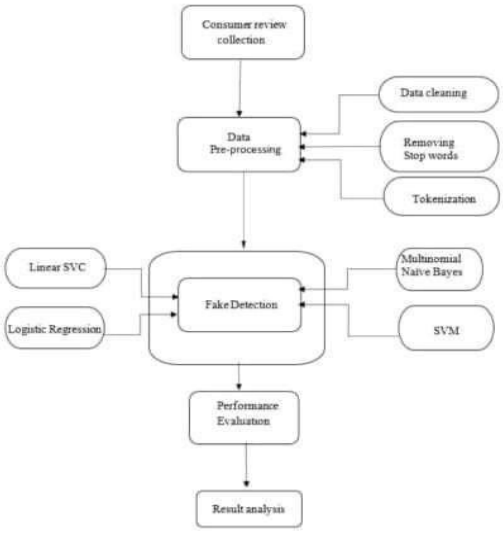


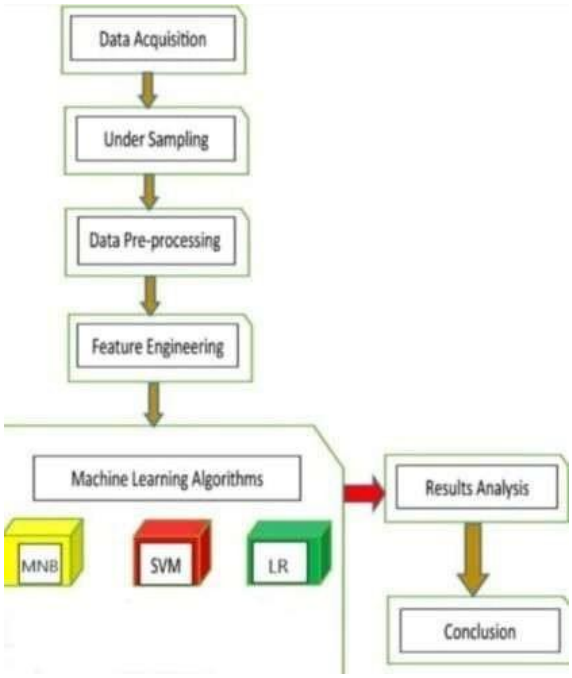
Figure 5.1. Flow Chart

5.1 DATA FLOW

Data flow models are an intuitive way showing how data is processed by a system level, they should be used to model the way in which data is processed in the existing system.

5.2 HIGH LEVEL DESIGN

In the High-Level Design, the proposed functional requirements of the software are studied. Overall solution architecture of the solution is developed which can handle those needs.



VI.METHODOLOGY

To solve the major problem faced by online websites due to opinion spamming, this project proposes to identify any such spammed fake reviews by classifying them into fake and genuine. The method attempts to classify the reviews obtained from freely available datasets from various sources and categories including service based, product based, customer feedback, experience based and the crawled Amazon dataset with a greater accuracy using Naïve Bayes [7], Linear SVC, SVM, Logistic Regression algorithms. In order to improve the accuracy, the additional features like comparison of the sentiment of the review, verified purchases, ratings, product category with the overall score are used in addition to the review details.

© DEC 2025 | IRE Journals | Volume 9 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I6-1712519

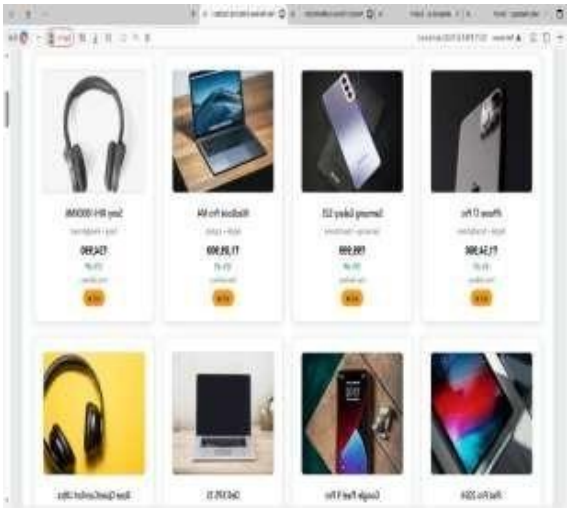
Multinomial Naïve Bayes Classifier:
Naive Bayes [7] is a statistical classification technique based on Bayes Theorem. It is one of the simplest supervised learning algorithms. Naive Bayes classifier is the fast, accurate and reliable algorithm. Naive Bayes classifiers have high accuracy and speed on large datasets. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.

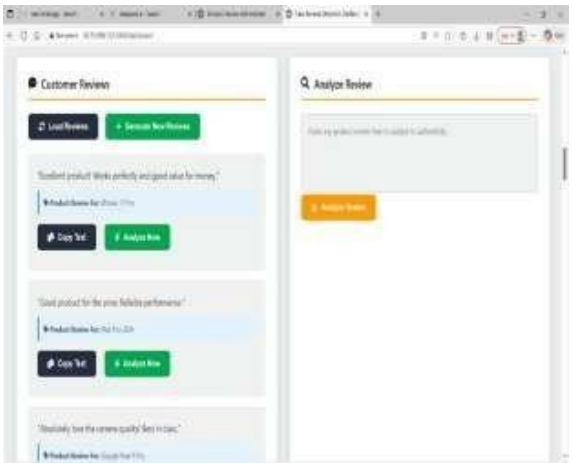
VII. RESULTS

Verified purchases column is the target variable for this project. From the count plot above, it can be seen that there are near equal parts of true VP and false VP (56% and 44% respectively).Amazon has provided their solution to combat fake reviews by implementing this column, Verified Purchases, where the reviewer has to go through a series of verification steps to ensure that the review that they are placing has indeed been bought from the site. This is Amazon's answer to combating fake reviews, and thus provides security on the truthfulness of the reviews since the review has been placed after purchasing the products.

bifurcations, and detection of microaneurysms from images.
Machine learning techniques (Random Forest, SVM, or deep learning models) are applied to the extracted features. The models are trained on labeled datasets to classify individuals according to predicted risk of heart disease based on retinal phenotypes .Validation and Evaluation.

Within the dataset, there are currently only two columns. Out of the two, review_text is going to be assigned as the input variable, and verified_purchases as the target variable. The data is then going to be split accordingly.





COUNT VECTORIZER AND MODELING:
word vectorization maps words or phrases from a lexicon to a matching vector of real numbers, which may then be used to determine word predictions and semantics, and this is done due to the fact that models only understand numerical data. We are going to be utilizing two of the vectorization methods, the first one being count vectorizer. We just count the number of times a word appears in the document in CountVectorizer, which results in a bias in favour of the most common terms.

VIII. CONCLUSION

The fake review detection is designed for filtering the fake reviews. In this project work Logistic Regression classification provided a better accuracy of classifying than the SVM for testing dataset. On the other hand, the Logistic Regression has performed better than other algorithms on the training data. Revealing that it can generalize better and predict the fake reviews efficiently. This method can be applied over other sampled instances of the dataset. The data visualization helped in exploring the dataset and the features identified contributed to the accuracy of the classification. The various algorithms used, and their accuracies show how each of them have performed based on their accuracy factors. Also, the approach provides the user with a functionality to recommend the most truthful reviews to enable the purchaser to make decisions about the product. Various factors such as adding new vectors like ratings, verified purchase have affected the accuracy of classifying the data better. After applying the above model, we have come to theconclusion that, obtaining false reviews requires both linguistic and behavioral features.

In our research work we have worked on just user reviews. In future, user behaviors can be combined with texts to construct a better model for classification. Advanced preprocessing tools for tokenization can be used to make the dataset more precise. Evaluation of the effectiveness of the proposed methodology can be done for a larger data set. This research work is being done only for English reviews. It can be done for Bangla and several other languages.

REFERENCES

[1] Rakibul Hassan, Md. Rabiul Islam “Detection of fake online reviews using semisupervised and supervised learning” 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019 .

[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] J. Li, M. Ott, C. Cardie and E. Hovy, “Towards a General Rule for Identifying Deceptive Opinion Spam,” in Proceedings of 52nd Annual Meeting of the Association for Baltimore, MD, USA, vol. 1, no. 11, pp.

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

[5] J. C.S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, “Supervised Learning for Fake News Detection,” IEEE Intelligent Systems, vol. 34, no. 2, pp. 76-81, May 2019.

[6] B. Wagh, J. V. Shinde and P. A. Kale, Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques,” International Journal of Emerging Research in Management and Technology, vol. 6, no. 12, pp. 37-44, December 2017.

[7] E. I. Elmurngi and A.Gherbi, “Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques,” Journal of Computer Science, vol. 14, no. 5, pp. 714– 26, June 2018.

[8] J. K. Rout, A. Dalmia, and K.-K. R. Choo, “Revisiting semi-supervised learning for online deceptive review detection,” IEEE Access, Vol.5, pp. 1319–1327, 2017

[9] N. O’Brien, “Machine Learning for Detection of Fake News,”[Online].Availablehttps://dspace.mit.edu/bitstream/handle/1721.1/119727/1078649610-MIT.pdf [Accessed: November 2018].

[10] N. Jindal and B. Liu., “Opinion spam and analysis”, Proceedings of the international conference on Web search and web data mining - WSDM 08 (2008), ACM, pp. 219 –230,2008.