

# Classification of Positive and Negative Fake Online Reviews using Machine Learning Techniques

**Anurag Goel**

Department of Computer Science and Engineering, Delhi Technological University, New Delhi-110042  
Email: anurag@dtu.ac.in

**Tanmay Nigam**

Department of CSE & IT, Jaypee Institute of Information Technology, Noida-201309  
Email: tanmayanigam@gmail.com

**Tarundeep Singh**

Department of CSE & IT, Jaypee Institute of Information Technology, Noida-201309  
Email: tarundeepsingh198@gmail.com

**Harsh Agrawal**

Department of CSE & IT, Jaypee Institute of Information Technology, Noida-201309  
Email: harshagrawal685@gmail.com

---

## ABSTRACT

E-Commerce is one of the most flourishing businesses in today's world. A large part of the population, especially in urban areas, is switching towards e-commerce websites to fulfill all of their shopping requirements, whether groceries, electrical appliances, clothing, etc. In an online purchase, product review is considered a significant factor in deciding the right choice of product. Therefore, e-commerce businesses are primarily dependent on product reviews. Due to the lack of authenticity of the reviewer information, while posting a review of any product or service online, the presence of fake reviews is increasing day by day. The presence of these fake reviews of various products or services impacts the customers and the sellers. The customers might choose the wrong brand of a product or service, while the sellers might face low sales of their high-quality products because of these fake reviews. This paper used different machine learning approaches to detect fake reviews of services on e-commerce sites. We have further categorized the fake reviews into positive and negative based on the reviewer's rating.

**Keywords - Logistic regression, Support Vector Machine, Decision Tree, Random Forest, Neural Networks**

---

Date of Submission: May 25, 2021

Date of Acceptance: June 22, 2021

---

## I. INTRODUCTION

Product reviews help a buyer in the most significant way to determine whether to buy a product or not. For most buyers, it is the only way to assess the quality of the product. The positive or the negative character of the reviews makes the final impression on the buyers' minds. If the reviews, for instance, turn out to be negative in the majority, the buyer might go for some other brand that has better reviews. Positive reviews play an essential role for the sellers. If the reviews are positive, the profits might increase for the sellers, and if they turn out to be negative, the consequences may not be as expected by the seller. For their benefit, fraud sellers sometimes post fake reviews to increase the sale of their products. The buyers, reading positive reviews, buy that product and get conned.

Similarly, rival companies may spam negative reviews about each other to decrease their sales [5]. It can even affect the decisions associated with the product's design or manufacturing or services that are needed to be given to the actual consumers. Therefore, the detection of fake reviews is crucial in today's world.

## II. RELATED WORK

Previous work on fake review detection was using the filtered data taken from yelp.com containing the reviews from 85 hotels and 130 restaurants in Chicago. The reviews used were having an equal number of real and fake reviews. POS-Tagging was also implemented in their work [2].

Fake reviews have high similarity among themselves. In [3], the authors have used a dataset labeled by random people according to the similarity property of fake reviews and supervised learning techniques applied to detect fake reviews in the labeled dataset. Later, text classification techniques were used to differentiate between fake and genuine reviews.

[4] focused on detection of fake online hotel reviews. [6] and [9] used the supervised learning techniques to detect the spam in product reviews while [7], [8], [10] proposed the approaches of opinion spam detection based on reviewer-centric and review-centric features.

### III. MACHINE LEARNING APPROACH

In this paper, the various supervised machine learning algorithms are applied to test the accuracy of the prediction of fake and genuine reviews:

- Logistic regression predicts the probability of a dichotomy. It is based on several features.
- Support Vector Machine finds the hyperplane that puts the margin between the two classes as distant as possible. Support vectors are the ones that determine that specific hyperplane.
- Naïve Bayes classifier is based on Bayes' theorem. It assumes all features are independent of each other. It consists of several learning algorithms that utilize the independence of statistics.
- Decision Tree utilizes a tree structure for classification. The dataset is broken into small-sized sets while developing a decision tree simultaneously in a gradual way.
- Random Forest algorithm generates a "forest" that contains many trees, similar to a tree created in decision tree algorithm. The more the number of trees, the better is the probability of achieving higher accuracy.
- Kernel SVM algorithms utilize a group of mathematical functions identified as the Kernel. Its role is to take input data and convert it into the structure that is needed. These kernels or functions may be of different kinds.

### IV. IMPLEMENTATION

#### 4.1 Datasets Used

The dataset has been taken from the website "yelp.com" which Rayana and Akoglu [1] use. It includes 608,598 reviews of restaurants in the cities of New York and Chicago, U.S. The product id, reviewer id, rating, date of the review, and the review text have been used as the input parameters.

In the first experiment, 16,000 reviews are selected which are later increased to 100,000 reviews in experiments. These reviews contained an equal number of real and fake reviews. The selection of features used to predict the genuineness of the reviews is one of the most crucial tasks. We have used two types of features, namely, review-centric features and reviewer-centric features.

Review-centric features include POS-tagging percentages, 100 unigram features from the reviews, and 100 bigram features, product id. Reviewer-centric features include reviewer id, rating given by the reviewer, deviation of a particular reviewer's rating from the average rating of the product, and review date. Further, we have also classified reviews as positive and negative based on the ratings provided by the reviewer.

#### 4.2 Data Pre-processing

During implementation, text wrangling and pre-processing the review text in our dataset was the first step we implemented using NLP techniques. The various steps performed in data pre-processing are as follows:

- Removing HTML text.
- Removing accented characters
- Expanding contractions of the text data of the review
- Removing special characters
- Stemming of the review text
- Removing Stop-words
- Creating Unigrams Matrix
- Creating Bigrams Matrix
- Percentage of POS Tagging.
- Adding features like reviewer id, product id, date of review as independent variables along with the cleaned review sparse matrix.

#### 4.3 Experiments

We added the deviation of the customer's rating from the average product rating as an independent variable. The number of positive and negative reviews has also been incorporated as an independent variable. We have included all the features into one dataset and chose them randomly. The dataset is divided into the training set and testing set.

LDA dimensionality reduction technique is applied to find a linear combination of features that characterizes or separates the two classes of objects or events. The other classification models, including Logistic Regression, Support Vector Machine (SVM), Naïve Bayes classifier, Decision Tree, Random Forest, Kernel SVM are also applied for prediction of real and fake reviews. Then, we applied 5-fold cross-validation technique after each classification model to avoid overfitting or underfitting on the dataset.

We have also applied Neural Networks using three hidden layers, activation function "relu" for hidden layers and activation function "sigmoid" for the output layer. The number of neurons in the hidden layers is 120. The optimizer used is "Adam" and loss considered is Binary cross entropy loss. Grid Search has been applied which uses different hyperparameters to increase the accuracies. Hyperparameter optimization is usually done using grid search. Grid search is guided by some performance metric, measured by cross-validation on the training set. As some of the parameters may include real or unbounded value spaces, we have set boundaries and did discretization before applying grid search.

The experiments are performed on two datasets of different sizes. The first dataset contains 16,000 reviews, while the second dataset contains 100,000 reviews. The reviews are randomly selected from the main dataset containing more than 600,000 reviews in both datasets. The accuracies of the different machine learning models are compared and analyzed in the next section.

We also analyzed the accuracies of the different classification models for finding the fake positive reviews and fake negative reviews to determine whether the fake reviews tend to be more positive or more negative. The classification of positive and negative reviews solely depends on the reviewer.

The positive or negative character of a review is determined using the rating provided by the reviewer corresponding to that review. The reviews having a rating of three or above have been considered positive reviews while the reviews having rating below three have been considered negative reviews.

## V. RESULTS AND DISCUSSION

Fig. 1, Fig. 2 shows the results of detecting the fake reviews by the various machine learning models on a dataset of size 16000 reviews. Fig. 1 shows the accuracies of the various models used without applying 5-cross validation, while Fig. 2 shows the results after applying 5-cross validation on the different machine learning models. It can be clearly seen from the results that SVM outperforms all the other models while the Naïve Bayes classifier performs the worst.

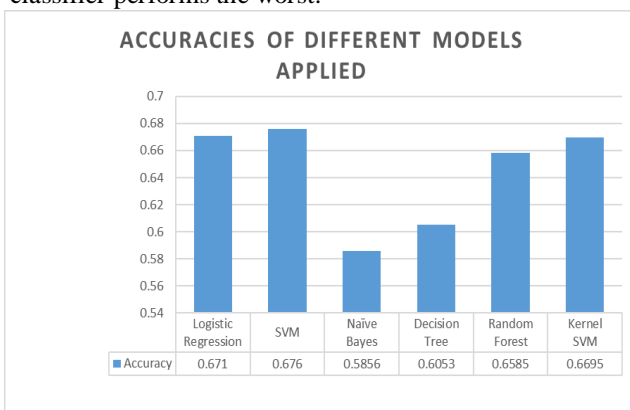


Figure 1: Accuracies of Models on dataset of 16,000 reviews (Without 5-fold Cross Validation)

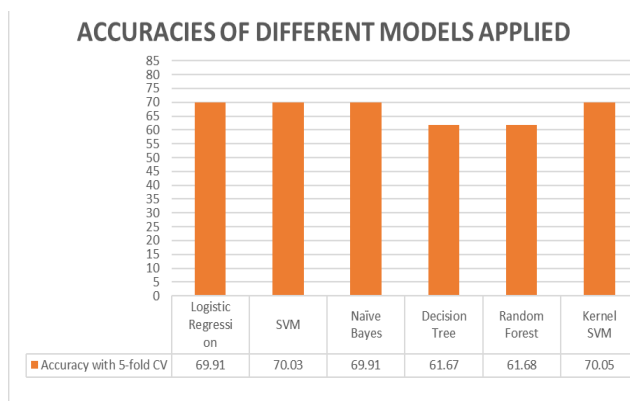


Figure 2: Accuracies of Models on dataset of 16,000 reviews (With 5-fold Cross Validation)

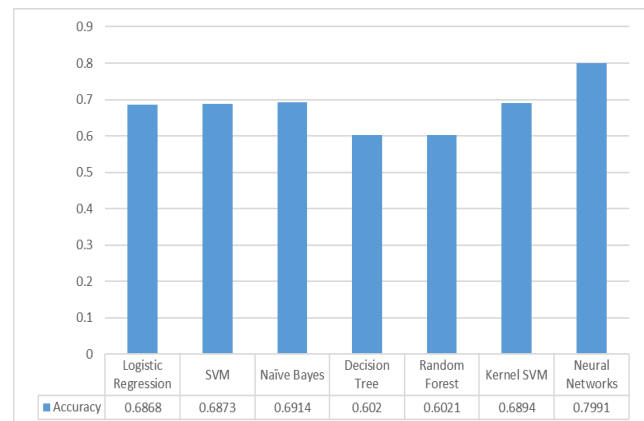


Figure 3: Accuracies of Models on dataset of 100,000 reviews (Without 5-fold Cross Validation)

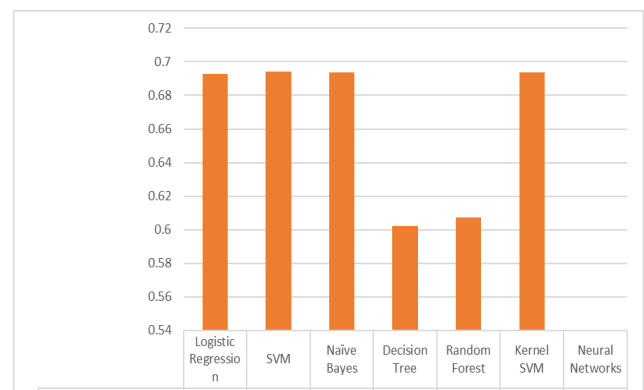


Figure 4: Accuracies of Models on dataset of 100,000 reviews (With 5-fold Cross Validation)

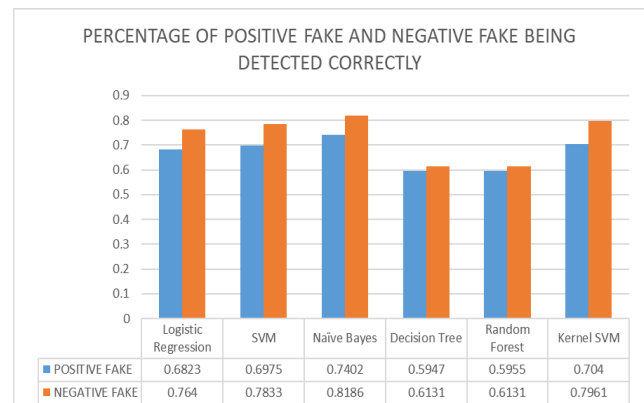


Figure 5: Class-wise Accuracies of positive and negative fake reviews on dataset of 16,000 reviews

Fig. 3, Fig. 4 shows the results of detecting the fake reviews by the various machine learning models on a dataset of size 100000 reviews. Fig. 3 shows the accuracies of the various models without applying 5-cross validation, while Fig. 4 shows the results after applying 5-cross validation. Although the results for this dataset with 5-fold cross-validation could not be obtained for the neural network due to the limited memory of our machine used for the experiments but Fig. 3. clearly depicts that the neural network outperforms all the other models.

Fig. 5 shows the class-level accuracies of various machine learning models in detecting the positive fake and negative fake reviews on a dataset of size 16000 reviews.

Naïve Bayes classifier performs better than the other machine learning models in detecting the positive fake and negative fake reviews.

## VI. CONCLUSION

In this work, six different machine learning models, namely Logistic Regression, Support Vector Machine (SVM), Naïve Bayes Classifier, Decision Tree, Random Forest and Kernel SVM, and Neural Networks, have been used for detecting the fake product reviews. We have also applied Neural Networks. Among the machine learning models used, SVM gave the best results in terms of accuracy. Furthermore, some models like Naïve Bayes gave the least accuracy as compared to other models. The accuracy achieved may have been affected by different factors like feature selection, random data selection, data pre-processing techniques used, size of the dataset, etc. Neural Networks outperformed all the machine learning models and gave the best results with an accuracy of 79.91%. On the other hand, Naïve Bayes classifier gave the best accuracy of 74.02% and 81.86% to find positive fake reviews and negative fake reviews.

As future work, we plan to incorporate some more reviewer-centric features like geographical locations, IP addresses, MAC addresses to achieve better accuracy.

## REFERENCES

- [1] S. Rayana and L. Akoglu, "Collective Opinion Spam Detection", *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 15*, 2015.
- [2] Mukherjee Arjun, Vivek Venkataraman, Bing Liu and Natalie Glance, "Fake Review Detection: Classification and analysis of real and pseudo reviews", *UIC-CS-03-2013, Technical Report* (2013).
- [3] Jindal, Nitin, and Bing Liu. "Analyzing and detecting review spam." In *Seventh IEEE international conference on data mining (ICDM 2007)*, pp. 547-552. IEEE, 2007.
- [4] Sandifer, Anna V., Casey Wilson, and Aspen Olmsted. "Detection of fake online hotel reviews." In *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*, pp. 501-502. IEEE, 2017.
- [5] Crawford, Michael, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter, and Hamzah Al Najada. "Survey of review spam detection using machine learning techniques." *Journal of Big Data* 2, no. 1 (2015): 1-24.
- [6] Mevada, Daya, and Viraj Daxini. "An opinion spam analyzer for product Reviews using supervised machine Learning method." *Journal of Information, Knowledge And Research In Computer Engineering* (2015).
- [7] Dominic, Dilsha, and V. P. Lijo. "Opinion spam detection using review and reviewer centric features." In *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, pp. 1154-1159. IEEE, 2017.
- [8] Lin, Yuming, Tao Zhu, Hao Wu, Jingwei Zhang, Xiaoling Wang, and Aoying Zhou. "Towards online anti-opinion spam: Spotting fake reviews from the review sequence." In *2014 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2014)*, pp. 261-264. IEEE, 2014.
- [9] Badresiya, Mr Ashok, and Jay Teraiya. "Performance analysis of supervised techniques for review spam detection." *International Journal of Advanced Networking and Applications* 8, no. SP (2014): 21-24.
- [10] Wahyuni, Eka Dyar, and Arif Djunaidy. "Fake review detection from a product review using modified method of iterative computation framework." In *MATEC web of conferences*, vol. 58, p. 03003. EDP Sciences, 2016.

## Biographies and Photographs



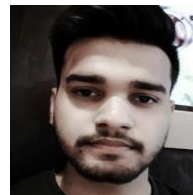
**Anurag Goel** is currently working as an Assistant Professor in the Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India. His areas of interest include Machine Learning, Deep Learning, Deep Clustering, Graph Algorithms, Data Mining and Social Network Analysis.



**Tanmay Nigam** has completed Bachelors in Technology in Computer Science and Engineering from Jaypee Institute of Information Technology, Noida, India. He is currently working as Business Operations Associate in ZS.



**Tarundeep Singh** has completed Bachelors in Technology in Computer Science and Engineering from Jaypee Institute of Information Technology, Noida, India. He is currently working as Associate Software Developer in Kuliza.



**Harsh Agrawal** has completed Bachelors in Technology in Computer Science and Engineering from Jaypee Institute of Information Technology, Noida, India. He is currently working as Software Engineer in Dew Solutions Pvt Ltd.