E-Commerce Spam Review Detection using Machine Learning

Yogansh Wagh[1], Sarfaraz Ali[2], Apurva Bobade[3], Aditi Bhalekar[4], Prof. Rajaram Ambole[5]

[1]Department of Computer Engineering, VPKBIET, Baramati
[2]Department of Computer Engineering, VPKBIET, Baramati
[3]Department of Computer Engineering, VPKBIET, Baramati
[4]Department of Computer Engineering, VPKBIET, Baramati
[5]Department of Computer Engineering, VPKBIET, Baramati

---***---

Abstract - With the proliferation of e-commerce platforms, the authenticity of online reviews has become increasingly crucial. In this research, we present a method for detecting spam reviews in e-commerce platforms, employing the Random Forest algorithm. Leveraging a dataset comprising reviews from Amazon Yelp dataset, we employ a combination of Natural Language Processing (NLP) techniques and text processing methods to uncover underlying patterns distinguishing between genuine and fake reviews. Real-time data scraping from the Amazon website facilitated the acquisition of a diverse range of reviews, subsequently stored in a CSV file for analysis. These reviews stored in CSV file is fed to model for prediction. Our model effectively discerns between authentic and spam reviews, offering a valuable tool for maintaining the integrity of e-commerce platforms and ensuring informed consumer decision-making.

***Key Words*:** Web Scrapping, Text Mining, Sentiment Analysis.

1. INTRODUCTION

The luxury of posting reviews online has grown at a faster pace and people buy almost everything online delivered to their doorstep. As a result, people are not physically checking products when purchasing online, so they rely significantly, unintentionally or habitually, on the reviews of other buyers. This should be as honest as possible so that buyers are not repeatedly deceived by fake reviewers or spammers. The problem is simple but difficult to solve by reading each review to mark it as a wrong or ambiguous category, this must be done systematically to get to the root of the problem. This problem can be solved by training an ML model that processes the review section to flag a particular review as genuine or spam. Interestingly, spammers who have not used the product can be caught this way. Spam reviews or use of other customer IDs can be used to filter out false product reviews to achieve good product ratings. This can be filtered by checking for the use of words like "excellent", "very good", "fantastic", etc. can be reported. Because they tend to praise the product or try to imitate genuine reviews with the same words, using them over and over again to impress buyers. So, the problem of spam filtering requires huge amounts of data to train and work with additional domain knowledge such as sarcastic phrases used by users to express disagreement towards the product. Sometimes the product is good but the shipping or packaging is not good, this affects the review's rating. Here, NLP techniques are used to identify these reviews instead of misclassifying them as negative reviews as in sentiment analysis. To remove unwanted or outdated product reviews, they include data pre-processing.

The aim of this study was to create an online e-commerce industry environment where consumers place trust in the platform where the products, they purchase are genuine and the reviews posted on these websites/apps is honest and is checked regularly by the company. The number of users is increasing, now companies like Twitter, WhatsApp, Facebook are using sentiment analysis to check fake news, harmful posts and ban those users/organizations from using their platform. At the same time, e-commerce (Flipkart, Amazon), hotel booking (Trivago), logistics, travel (Trip Advisor), job search (LinkedIn, Glass Door), food (Swiggy, Zomato) industries, etc. Uses algorithms to fight counterfeits review, spammers to trick consumers into buying poor quality products/services. And users should be warned about spammers as “unverified profiles”, so that users need not worry about these fake users. Manually tagging reviews is mostly time-consuming and less effective. Therefore, a supervised learning model is used to label the reviews and then the labels cannot be predicted. For example, Mukherjee et al. manually labeled 2,431 reviews over 8 weeks, so automatic review labeling can help save time and energy, which is difficult and is suggested by Sunil Soumya et.al. Some industries pay to write fake reviews of their products and services when it is impossible to classify the review as spam or not. Amazon's "Yelp" dataset contains 30-40% spam reviews. Feature selection is an important aspect in selecting and training these models. In this paper, we have also done comparison with our base research paper to demonstrate the performance of Random Forest algorithm for this “Amazon yelp” dataset and their suitability for implementing these models in real-time software. The Random Forest (RF) model performed significantly better than the Naïve Bayes algorithm which was used in our base paper. The problem of detecting fake reviews is addressed fairly and gives a fair overview of its legitimacy and necessity. The goal is to choose a suitable algorithm to complete the task of detecting fake reviews and eliminating them.

2. RELATED WORK

Related work in the field of spam review detection and classification using machine learning techniques has seen significant attention in recent years. Researchers have explored various approaches to address the challenge of distinguishing between genuine and fake reviews in e-commerce platforms. Some notable works include:

1. **Sentiment Analysis Techniques:** Many studies have focused on sentiment analysis to identify spam reviews. They employ techniques such as lexicon-based sentiment analysis, machine learning-based sentiment classification, and deep learning models to analyze the sentiment expressed in reviews and detect anomalies indicative of spam.

2. **Feature Engineering and Selection:** Researchers have investigated the importance of feature engineering and selection in improving the performance of spam review detection models. They explore a wide range of features, including textual features (e.g., n-grams, word frequency), structural features (e.g., review length, rating deviation), and semantic features (e.g., sentiment polarity, readability).

3. **Ensemble Learning Methods:** Ensemble learning methods, such as Random Forest, Gradient Boosting, and AdaBoost, have been widely applied in spam review detection. These methods combine multiple base classifiers to improve classification performance and generalization ability.

4. **Deep Learning Approaches:** With the rise of deep learning, researchers have explored the effectiveness of deep neural networks for spam review detection. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants have been applied to automatically learn features from review texts and capture complex patterns indicative of spam.

5. **Cross-Domain and Transfer Learning:** Some studies focus on cross-domain and transfer learning techniques to address the challenge of limited labeled data in specific domains. They leverage pre-trained models or transfer knowledge from related domains to improve the generalization ability of spam review detection models.

Overall, the related work highlights the diverse range of approaches and methodologies employed in spam review detection using machine learning techniques. Our project contributes to this body of research by leveraging the Random Forest algorithm and employing a combination of NLP techniques and text processing methods to effectively identify and filter out spam reviews in e-commerce platforms.

Preliminary analysis is carried out on the basis of opinions expressed as user opinions through texts, blogs, reviews, feedback, etc., which must be clearly calculated and considered in order to obtain relevant information. This is nothing but sentiment analysis.

In this interesting study, this two-step approach called SVM classifier was used. For classifying tweets [1]. Some people use emojis, smileys, and hashtags to categorize labels into multiple emotions [2]. Other researchers have used SVM classifiers to train data using emojis [3].

Some of the existing systems in this area are:

1. **Lexicon-based methods:** Based on counting the number of positive and negative words in sentence-Twitter.
2. **Rule-based methods:** Based on syntactic rules, e.g., [query] is pos-adj – Tweet feel
3. **Machine learning based methods**: Based on the classifier built on a training data Twitter sentiment.

3. METHODS

1) Dataset: The dataset used is “Amazon Academic Review” which contains reviews, helpful votes, ratings, user IDs, and many other attributes. Useful parameters are extracted for feature engineering. The dataset contains thousands of original and fake reviews mixed together to easily evaluate the accuracy of the model implemented using this dataset. The Yelp dataset released for the Academic Challenge contains information on 11,537 businesses. This dataset contains 8,282 record sets, 43,873 users, 229,907 reviews for these businesses (www.yelp.com/dataset). This dataset is challenging because it contains a large set of tests and different parameters for train any algorithm.

2) Pre-processing: The initial stage of any dataset analysis is pre-processing, which involves cleaning the dataset for training purposes by eliminating extraneous characteristics, punctuations, stop words, missing words, redundant words, etc. This guarantees that the model is properly trained.

3) Feature Engineering: This function, also known as data cleaning, includes all techniques for eliminating undesired information from the dataset. To identify the gaps and the relationships between the various features (columns) and use them to derive reliable conclusions, this stage is crucial. A collection of words used to build a corpus of words is the NLTK package's libraries. OrderedDict is used to import the functions for term frequency, tokenizer, and Stopwords. Words like "is," "then," "to," "why," and other terms that are superfluous in this context and don't improve feature engineering are categorized as Stopwords under the English language. Term frequency is a measure of how frequently a word appears and can be used repeatedly by spammers to identify themselves.

4) Sampling Data: The dataset is sampled before it is even fed to the classifier because a large number of reviews are used in it. The purpose of the sampling is to reduce the classifier's weight, which loads the data in segments. Here, two columns are concatenated after labeling in order to return the data frame, and various labels are used to authenticate the phony reviews.

The dataset is sampled before it is even fed to the classifier because a large number of reviews are used in it. The purpose of the sampling is to reduce the classifier's weight, which loads the data in segments. Here, two columns are concatenated after labeling in order to return the data frame, and various labels are used to authenticate the phony reviews.

4. MODEL OUTLINE AND WORKING

Random forest classifier: The Random Forest classifier is a popular ensemble learning method used for classification tasks in machine learning. It operates by building multiple decision trees during the training phase and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Here's an explanation of its outline and working:

1. ***Outline:***

- **Ensemble Method:** Random Forest is an ensemble learning method, meaning it combines multiple individual models to produce a stronger, more robust model.

- **Decision Trees:** At the core of Random Forest are decision trees, which are simple models that make decisions based on input features.

- **Bagging: Random:** Forest employs a technique called bagging, which stands for Bootstrap Aggregating. It trains each decision tree on a random subset of the training data, with replacement, to ensure diversity among the trees.

- **Voting or averaging:** For classification tasks, Random Forest aggregates the predictions of all the individual trees and outputs the class that receives the most votes. For regression tasks, it averages the predictions of all trees to produce the final prediction.

2. Working:

- **Random Sampling:** Random Forest begins by randomly selecting subsets of the training data with replacement (bootstrapping). This process creates multiple training datasets, each of which is used to train an individual decision tree.

- **Decision Tree Construction:** For each subset of data, a decision tree is constructed by selecting the best split at each node based on a random subset of features. This helps to decorrelate the trees and reduce overfitting.

- **Voting/Averaging:** Once all decision trees are constructed, predictions are made for new data points. In classification tasks, each tree "votes" for the class, and the class with the most votes is chosen as the final prediction. In regression tasks, the predictions of all trees are averaged to produce the final prediction.

- **Feature Importance:** Random Forest can also provide insights into feature importance by measuring how much each feature contributes to the model's performance. This information can be useful for feature selection and understanding the underlying relationships in the data.

Overall, Random Forest is a powerful and versatile algorithm that is widely used for classification and regression tasks due to its robustness, scalability, and ability to handle high-dimensional data with complex relationships.

a) Accuracy= TP+TN / FP+FN+TN

b) Precision= TP/TP+FP

c) Recall (sensitivity) = TP/TP+FN

d) F1_score = 2 * (Recall * Precision) / (Recall + Precision)

All of the above parameters determine performance. The model, model results are displayed along with the confusion matrix.

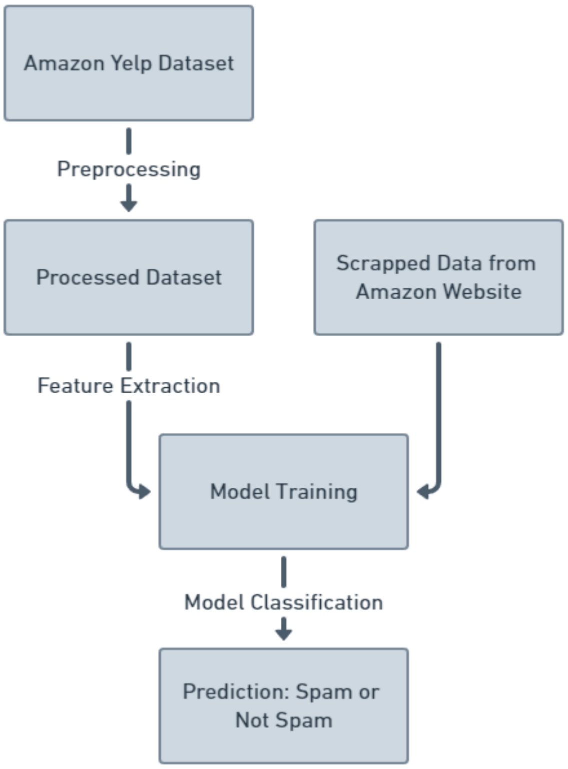

Figure 1: Flow Chart for fake review detection

The flowchart for "Figure 1" is as follows: 1 is explained as follows. The beginning of problem solving is the collection of data sets, and care must be taken to select the correct data set and ensure that it is binary or categorical. To get the data in the format needed to build the model, I loaded the reviews into

Yelp's academic dataset His Reviews. Json file. For brevity, only attributes useful for future events were later shortlisted. Feature extraction is performed and used to train Random Forest classifier, and relationships between different attributes are recorded and used for classification. After training the model, it is fed with new or test data, calculates classification accuracy, etc. as shown in Table 1, and adjusts accordingly for better results. After this we scrap real-time reviews from Amazon website using web scrapping as shown in “Figure 2”. We extract only reviews from Amazon website and store it into CSV file. This CSV file is sent for classification to our model.

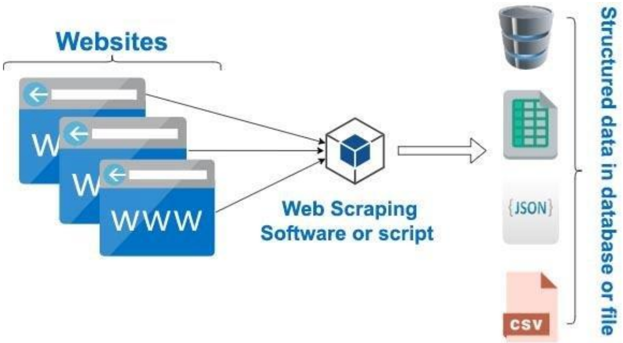

Figure 2: Web Scrapping

The measured parameters of this model are confusion matrix, accuracy, precision, sensitivity, and F1_score.

5. RESULTS

Table 1: Comparison of Naïve Bayes and Random Forest results

Sr. No.	Parameter	Naïve Bayes (in %)	Random Forest (in %)
1.	Accuracy Score	79.007	**89.487**
2.	Precision Score	70.224	**85.577**
3.	Recall Score (Sensitivity)	99.099	94.389
4.	F1 Score	82.169	**89.768**

From Table 1, we can see that the two models perform fairly well when compared, except that the random forest classifier performs better. Therefore, Random Forest has better accuracy, precision score, and F1 score. We can conclude that the random forest classifier can be used in the approach to monitor and remove fake product reviews. When compared with models from different applications, they perform well in certain areas, but are incompatible in some areas, so the application requires some experience.

6. CONCLUSION & FUTURE SCOPE

In conclusion, our research endeavors have culminated in the development of a robust framework for detecting and mitigating spam reviews within the e-commerce domain. Through the utilization of the Random Forest algorithm and a comprehensive dataset comprising reviews from prominent platforms like Amazon and Yelp, we have successfully constructed a model capable of discerning between authentic and deceptive reviews with remarkable accuracy.

The integration of advanced Natural Language Processing (NLP) techniques and text processing methodologies has enabled us to uncover subtle linguistic cues and patterns that distinguish genuine reviews from their fraudulent counterparts. Furthermore, our approach has been bolstered by the incorporation of real-time data scraping techniques, which have enriched our dataset and ensured its relevance and comprehensiveness.

Our findings underscore the significance of proactive measures in safeguarding the integrity of e-commerce platforms and fostering trust among consumers. By effectively identifying and filtering out spam reviews, our model serves as a valuable tool for platform administrators and consumers alike, facilitating more informed decision-making and enhancing overall user experience.

Moving forward, there remains ample room for refinement and expansion of our methodology. Future research endeavors may explore the integration of additional machine learning algorithms or the incorporation of domain-specific features to further enhance the accuracy and robustness of our model. In essence, our project represents a significant step towards addressing the pervasive issue of spam reviews in e-commerce, offering a tangible solution that holds promise for improving the quality and reliability of online consumer feedback.

REFERENCES

[1] D. Barbosa, Luciano & Feng, Junlan. (2021). Robust Sentiment Detection on T witter from Biased and Noisy Data. Coling 2021 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference. 2. 36-44.

[2] McCallum, Andrew. "Graphical Models, Lecture2: Bayesian Network Represention" (PDF). Retrieved 22 October 2019.

[3] Joseph, S. I. T. (2019). SURVEY OF DAT A MINING ALGORITHM’S FOR INT ELLIGENT COMPUT ING SYST EM. Journal of trends in Computer Science and Smart technology (T CSST),1(01), 14 -24.

[4] Ren Y, Ji D (2017) Neural networks for deceptive opinion spam detection: an empirical study. InfSci 385:213–224.

[5] Heydari A, T avakoli M, Salim N (2016) Detection of fake opinions using time series. Expert SystAppl 58:83–92.

[6] Mukherjee A, Venkataraman V, Liu B, Glance N (2013a) Fake review detection: classification and analysis of real and pseudo reviews. Technical Report UIC-CS-2013–03, University of Illinois at Chicago.