

Fake Reviews Filtering System Using Supervised Machine Learning

1st Deepanshu Jain

Computer Science Engineering
Dr. Akhilesh Das Gupta Institute of
Technology And Management
New Delhi, India
djain844@gmail.com

2nd Sayam Kumar

Computer Science Engineering
Dr. Akhilesh Das Gupta Institute of
Technology And Management
New Delhi, India
sayamk565@gmail.com

3rd Yashika Goyal

Information Technology
Dr. Akhilesh Das Gupta Institute of
Technology And Management
New Delhi, India
yashu.gyl@gmail.com

Abstract—As the surge in internet users is expanding prominently, the role of the online reviewing system is also rising. For companies, the legitimacy of internet evaluations is critical, as it can directly impact their reputation and profitability. It plays an indispensable role in influencing people's perceptions of a product or service. This research projects light on the best technique to identify and filter out authentic reviews while proposing a flexible and user-friendly website. The website will have a tremendous sway on customers and will assist them in making a better judgment about a product/service. The website is deployed with the designed supervised learning model. Firstly, the user will have to enter the URL of the website where the product is located. After which, the dataset is extracted from the given URL using Python tools for Web Scrapping. The data is then analyzed and dissected using Natural Language Processing techniques to extract sound features from it. Ultimately, different Machine Learning Models are further trained on the dataset. The experimental results of this research reveal that the model performs at an accuracy of 89.12% on the datasets. The major objective of this research is to provide a fake review filtering system that will provide users with more reliable review information and eliminate revenue loss of the companies at an exponential rate.

Keywords—Authentic Reviews, Fake Reviews Detection, Hyperparameter Tuning, Machine Learning, Natural Language Processing, Supervised Learning, Web Scrapping.

I. INTRODUCTION

In Today's world, user reviews significantly impact how much revenue a business makes. Consumers use reviews as their primary basis of information concerning any product and service before judging the quality of a product or service. Fake Reviews are generally publicized to give either a positive opinion for the promotion of the product or the competent business/organization posts them intending to degrade the reputation of the business/brand/product [1]. Henceforth, these fake reviews play with the people's minds wavering an individual's judgment about a product [2]. People are more inclined to buy a product if the reviews are predominantly positive, but they will look for an alternative if the reviews are cynical. Consequently, the authenticity of online reviews is crucial for organizations, as it directly influences their reputation and profitability [3]. It is startling that nearly 74% of consumers read fake reviews on different sites last year, as represented in [Fig. 1].



Fig. 1. Users That Read Fake Reviews

Users on Amazon can mark whether the reviews are valuable or not. Unfortunately, this still necessitates human effort and is prone to spammers. According to a report [4] on Amazon, only 17% of users truly trust reviews. Only 3% to 10% of individuals post reviews, and 61% of electronic reviews are fake. As of March 2019, Amazon had over 2 million unverified reviews [Fig. 2]. Detecting fake reviews is complex since no one knows the exact volume of spam.

According to a report on Amazon :



Fig. 2. Report Analysis On Amazon Reviews

Due to the open aspect of product review sites, it is hugely challenging to spot scammers camouflaging themselves as genuine users. Some firms pay spammers to post fake reviews. Fake reviews take advantage of individuals' purchasing decisions. The Detection becomes arduous and non-trivial due to the consumers' additional comparing attempts.

The primary objective of this research paper is to establish a practical and versatile Machine Learning model that would be highly efficacious at processing and synthesizing tons of customer reviews to determine their status of fraudulency. An interactive user website that filters out all the fraudulent reviews on a specific product has been established. As a consequence of which, it displays only the reviews which are genuine and factual. The website would provide the user an option to input an URL of a particular product on an e-commerce website. The user reviews on every single product have been extracted by leveraging the utility of web scraping as a tool. After that, a dataset was concocted by aggregating all those reviews concurrently.

This design will eliminate the possibility of scams to a considerable extent which will prove to be of significant importance to the overall progress of businesses and organizations. An online review can be classified as dishonest or genuine by implementing various Supervised Machine Learning Algorithms. To ensure reliable Detection of fraudulent reviews, this research paper accomplishes the whole process of fake reviews detection by leveraging multiple Natural Language Processing techniques and thereby incorporates the development of a streamlined machine learning model which can make predictions, on whether an

online review is fake or not, to an excellent level of accuracy [5]. The principal motive behind applying the Natural Language Processing Techniques in this study is to transform the initial textual data into its corresponding vectorized form by applying a bag of words model such as CountVectorizer as well as TF-IDF (Term Frequency - Inverse Document Frequency) transformer, which can subsequently be provided to suitable ML Algorithms for predicting precise results. This paper proposes a machine learning technique to identify fake reviews. The paper compares the performance of several machine learning models to examine their effectiveness in making accurate predictions regarding fraudulent and non-fraudulent reviews [6]. By leveraging the utility of the best possible ML model, companies will not only be able to successfully detect most of the misleading reviews generated by spammers about their products but will also be able to take the necessary actions against these unlawful spamming activities.

Consequently, they would be able to make correct decisions regarding their products, facilitating their business objective's rapid development and advancement. Moreover, they would also learn about their significant products and services, which they need to emphasize to enlarge their entire business and boost their net financial gains. A higher number of customers can be attracted to authentic and quality products and services offered by any particular company by effectively implementing this fake review detection framework. Furthermore, the customer retention rate can be enhanced by facilitating the elimination of deceptive reviews and thereby maintaining all the objective feedback provided by trustworthy customers as the new prospective customers, who are rather unknowledgeable about any specific product or service launched by a company, would be able to read and acknowledge the genuine reviews given by their fellow customers to thoroughly understand the overall quality and features of that particular product.

The fundamental motivation behind the paper is to make an online product or service selection more accessible and secure. Detection of fake reviews will eventually lead to burgeoning sales for the companies if their products are adequately authentic and valuable enough to the demands and expectations of their customers.

II. RELATED WORKS

Eka Dyar Wahyuni and Arif Djunaidy [7] aim to use the language and rating properties of a review to detect fake product reviews. The proposed system (ICF++) gauges the reviewer's trustiness value, A review's honesty value, and the reliability factor of a resource. A review's honesty value is measured using opinion mining and text techniques. The experimentation shows that the proposed work has better accuracy than the Iterative Computation Framework Method.

Ahmed M. Elmogy et al. [8] propose a Machine Learning strategy to identify fake reviews. In addition to the features extraction process of the reviews, the paper applies several features engineering to extract various behaviors of the reviewers. The composition compares the performance of several experimentations done on a real Yelp dataset of restaurant reviews with and without the extracted features from user's behaviors.

Rami Mohawesh et al. [9] sum up the existing datasets along with their collection methods. It analyses the current feature extraction techniques and critically summarises and

analyses the existing approaches to determine gaps based on deep learning methods and Traditional Statistical Machine Learning. Furthermore, it ushers a benchmark analysis to examine the performance of diverse Neural Network Models that have not yet been utilized to detect fake reviews.

S. Aarthi et al. [10] suggest a unique method for finding fake reviews based on a graph-based approach in which a particular node stores info about what it represents. It suggests a unique method for finding fake reviews based on a graph-based approach in which a particular node stores info about what it represents. It is assembled via arranging suitable vector features according to a similarity metric. Label information is dynamically propagated around the graph during this process, and all nodes get labeled at the end. In the proposed system, time series analysis is one of the most commonly explored approaches for outlier detection. By analyzing all data samples as a time series feature vector, the data sample that is the farthest away from all clusters acquires the most substantial outlier score.

Amit Kumar B. Jadhav et al. [11] have dealt with the problem of fake review detection in the eCommerce field. The paper focuses explicitly on online reviews' in particular and the sentiment analysis domain as a whole. A cumulative study of this work is performed in the paper. Moreover, the paper demonstrates how the acquisition of unlabeled data enhances the accuracy in identifying fake reviews using three different base learner algorithms, viz: Naïve Bayes, Decision Tree, and Logistic Regression.

Shilpa Yadav et al. [12] propose a fake review detection system by applying three feature extraction techniques: CountVectorizer, TfidfVectorizer, and Ngram model. The features are then trained using several machine learning algorithms.

Alimuddin Melleng et al.; [13] explored the persuasiveness of sentiments and emotions-based representations for the task of constructing machine learning models for fake review detection. They illustrated how combining multiple emotion and sentiment extraction methods could improve the data representation.

III. METHODOLOGY

This section comprehends the various classifiers used in our model to predict fake reviews, the proposed methodology to improve the model accuracy, and the website on which the model will be deployed. Below is shown the overview [Fig. 3] of the Fake Review Detection model.

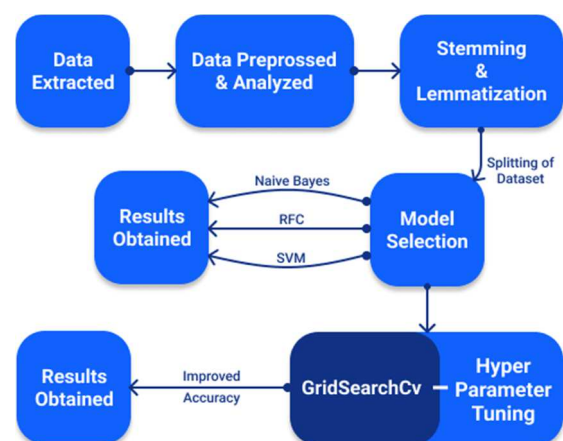


Fig. 3. Fake Review Detection Model

DATA DESCRIPTION:

In this research paper, initially, the data is collected from the OSF website. We used Pandas to extract the data frame. This dataset is publicly available on the Internet and contains Twenty Thousand fake reviews and Twenty Thousand genuine product reviews.

DATA PREPROCESSING AND ANALYSIS:

Further Exploratory data analysis is performed on the dataset. With the help of Pandas, the data manipulations in the dataset were carried out. Using the Seaborn library, Data is Visualised, and the Parameters are analyzed.

STEMMING AND LEMMATIZATION:

Stemming is a well-known process in the field of Natural Language Processing which is mainly used for reducing a word to its word stem that affixes to suffixes and prefixes or the root of the words, known as a lemma. [14] It is an essential technique in Natural Language Understanding (NLU).

Lemmatization is a methodical and step-by-step mechanism for acquiring the root form of a word. Word structure varies depending upon the use case. Morphological analysis means reducing complicated and complex expressions to their simplest form by exploring all possible solutions.

MODELS USED:

A. RANDOM FOREST:

Random forest is a supervised machine learning algorithm used widely in Classification and Regression problems. [15] It constructs decision trees on diverse samples and abides their majority vote for classification and average in case of regression. It can handle the data set retaining continuous variables as in the case of regression and categorical variables as in the case of classification.

B. SUPPORT VECTOR MACHINES:

Support Vector Machines or SVM is a Supervised Learning Algorithm used for classification and regression problems. It creates the best line or decision boundary that can segregate n-dimensional space into classes to easily place the new data point into the correct category in the future.

C. NAIVE BAYES:

Naive Bayes is a prevalent classification algorithm for multi-class and binary classification problems. A Naive Bayes classifier assumes that a particular feature in a class is unrelated to the presence of any other feature. It is a probabilistic classifier, which indicates that it anticipates based on the probability of an object.

HYPER PARAMETER TUNING USING GRIDSEARCHCV:

GridSearchCV is basically a hyperparameter tuning technique. It attempts all the combinations of distinctive hyperparameters passed in a dictionary and their companion values. The model is evaluated for every combination using the cross-validation method and determines the best potential value for the hyperparameters.

This makes its processing time-consuming and intensive based on the number of hyperparameters involved. It is regarded as a valuable tool for fine-tuning the parameters of any model.

WORKING:

As part of the data collection process, the customer reviews about all possible products in a particular domain

were initially extracted through the base URL of the e-commerce website. Then the complete approach of web scraping for any specific URL was implemented using Python's BeautifulSoup library. In the subsequent phase, a data frame using Python's Pandas library was constructed, which consisted of all the user reviews available for a specific product. Firstly, the Fake Reviews dataset from the OSF Home Foundation website was extracted, which consisted of twenty thousand fake reviews and twenty thousand original product reviews. Data Cleaning and Preprocessing on the entire textual data was implemented and subsequently transformed each word in the text to its corresponding tokens, which were, in turn, converted into vectors that can quickly be processed by the machine learning algorithms. Several supervised machine learning models were trained and tested with the fake review as our target label for the text classification modeling process. The Support Vector classifier had the best performance among all the used models in the initial results. Further, hyperparameter tuning using GridSearchCV was performed to improve the trained models' overall accuracy. The supervised improved model was then deployed on the website using the Python Django framework. [Fig. 4] describes the working of the website.

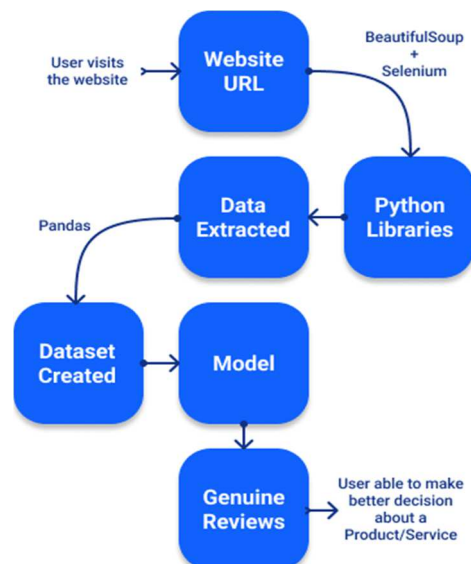


Fig. 4. Working Of The Website

IV. RESULTS AND DISCUSSION

The envisioned website would assist customers in filtering out fraudulent reviews, allowing them to make more informed selections about a product or service.

This study of review segregation and filtering based on authenticity may be expanded to include real reviews that are filtered based on their positive or negative expression. This makes it easier for consumers to make better-informed decisions about products and services. Furthermore, this augments the gross sales of retail outlets and corporations, resulting in an enormous rise in net revenue and profit gains for businesses.

The classification reports of several classifiers, including precision, recall, and f1-scores, are shown in the table below. When compared to other classifiers, the SVM classifier performs the best. SVM has a precision of 0.90 in predicting legitimate reviews, as seen in Table I. In other words, the model is successful 90% of the time when predicting if a

review is accurate. The recall of SVM is 0.87. It detects 87 percent of all false reviews properly. Similarly, the F1-Score, which summarises the model's prediction ability, is 0.89.

TABLE I. CLASSIFICATION REPORT

MODEL	TYPE	PRECISION	RECALL	F1-SCORE
RFC	CG	0.88	0.81	0.85
	OR	0.80	0.87	0.83
SVM	CG	0.87	0.90	0.88
	OR	0.90	0.87	0.89
NAIVE BAYES	CG	0.89	0.82	0.85
	OR	0.81	0.88	0.84
GRIDSEARCHCV	CG	0.87	0.90	0.89
	OR	0.90	0.87	0.89

The most frequent performance indicator for classification algorithms is accuracy. The accuracy of each classifier in predicting bogus reviews is shown in the graph. The graph shown [Fig. 5] demonstrates that SVM surpasses the other classifiers, obtaining a 92 percent accuracy rate. Following that, the hyperparameter tweaking technique was used to fine-tune the Support Vector Classifier model, which led to an improvement in its prediction accuracy from 92 to 93 percent.

In the future, the overall model accuracy of this work can be increased. This research of segregation and filtering of the reviews based on their authenticity can also be expanded to filter even the genuine reviews based on their positive or negative expression. This will make it easier for the consumer to make a finer judgment on a product/service.

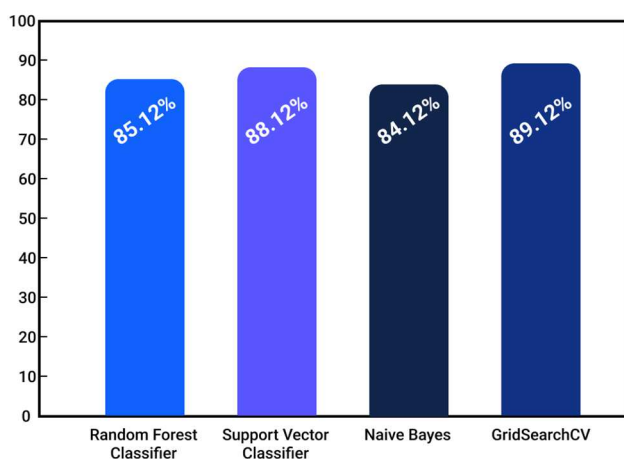


Fig. 5. Accuracy Chart For Various Models

V. CONCLUSION

Reviews play a crucial role in an individual's decision. The dataset extracted is used to evaluate the proposed research. The Support Vector Classifier (SVC) performed the best in implementing different Classification Algorithms. Experimental results determine the system's adequacy with an accuracy of 89.1% with the help of GridSearchCV. This paper additionally proposes the model to be deployed on a website

designed to filter out fake reviews, rendering and revealing only genuine reviews to the user. The Website will have a massive influence on customers and will assist them in deciding whether to purchase a product/service or not. Moreover, the Website will allow companies and consumers to build a connection with one another.

REFERENCES

- [1] Joni Salminen, Chandrashekhar Kandpal, Ahmed Mohamed Kamel, Soon-Gyo Junga, Bernard J.Jansena, "Creating and detecting fake reviews of online products.", <https://doi.org/10.1016/j.jretconser.2021.102771>
- [2] Sushant Kokate, Bharat Tidke, "Fake Review and Brand Spam Detection using J48 Classifier", International Journal of Computer Science and Information Technologies, Vol. 6 (4), 2015, 3523-3526
- [3] S. Jia, X. Zhang, X. Wang, and Y. Liu, "Fake reviews detection based on LDA," 2018 4th International Conference on Information Management (ICIM), 2018, pp. 280-283, DOI: 10.1109/INFOMAN.2018.8392850.
- [4] "What Percentage of Amazon Reviews Are Fake", Online: <https://review42.com/resources/what-percentage-of-amazon-reviews-are-fake/>
- [5] P. Devika, A. Veena, E. Srilakshmi, A. R. Reddy, and E. Praveen, "Detection of fake reviews using NLP &Sentiment Analysis," 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 1534-1537, DOI: 10.1109/ICCES51350.2021.9489210.
- [6] N. A. Patel and R. Patel, "A Survey on Fake Review Detection using Machine Learning Techniques," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-6, DOI: 10.1109/CCAA.2018.8777594.
- [7] Eka Dyar Wahyuni and Arif Djunaidy, "Fake Review Detection From a Product Review Using Modified Method of Iterative Computation Framework" in Research Gate, DOI: 10.1051/mateconf/20165803003
- [8] Ahmed M. Elmogy, Usman Tariq, Atef Ibrahim, and Ammar Mohammed, "Fake Reviews Detection using Supervised Machine Learning", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 12, No. 1, 2021
- [9] R. Mohawesh, Shuxiang Xu, Son N. Tran, Robert Ollington, Matthew Springer, Sumbal Maqsood, and Yaser Jararweh, "Fake Reviews Detection: A Survey," in IEEE Access, vol. 9, pp. 65771-65802, 2021, DOI: 10.1109/ACCESS.2021.3075573
- [10] S.Aarthi, Shivendra Mishra, Vishal Chauhan, Rithik Chordia, "Detection of Fake review using Machine Learning ", ISSN NO: 0090-5674, Vol XIV Issue 5 2021
- [11] Amitkumar B. Jadhav, Vijay U. Rathod, Dr. Hemantkumar B. Jadhav, "Improving Performance of Fake Reviews Detection in Online Reviews using Semi-Supervised Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 06, June 2019
- [12] Shilpa Yadav, Dr.Gulbakshee Dharmela, Khushali Mistry, "Fake Review Detection Using Machine Learning Techniques", Journal of Emerging Technologies and Innovative Research (JETIR), ISSN-2349-5162, April 2021, Volume 8, Issue 4
- [13] Alimuddin Melleng, Anna-Jurek Loughrey, Deepak P., "Sentiment and Emotion Based Text Representation for Fake Reviews Detection", Proceedings of Recent Advances in Natural Language Processing, pages 750-757, Varna, Bulgaria, Sep 2-4, 2019. https://doi.org/10.26615/978-954-452-056-4_087
- [14] K. Venkateswara Rao, P.Anil, R.J.V. Siddhartha, T. Antony, Fake Review Sentiment Analysis Using Natural Language Processing, Vol 12, Issue 7, July/ 2021 ISSN NO: 0377-9254
- [15] J. Fontanarava, G. Pasi and M. Viviani, "Feature Analysis for Fake Review Detection through Supervised Classification," 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2017, pp. 658-666, DOI: 10.1109/DSAA.2017.51.