IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

Fake Product Review Detection and Removal System using NLP

Prof. Prathibha R[1], Sahana H S[2], Yashashwitha R[3]
Assistant Professor, Department of Information Science and Engineering[1]
Students, Department of Information Science and Engineering[2,3]
S J C Institute of Technology, Chikkaballapur, Karnataka, India

Abstract: *Fake review detection and its elimination from the given data set using different natural language processing techniques is important in several aspects fake review dataset is trained by applying two different machine learning models to predict the accuracy of how genuine are the reviews in a given data set. The fake review problem must be addressed so that these large ecommerce industries such as amazon, Flipkart, etc.*

Keywords: Sentiment Analysis, Text Mining

I. INTRODUCTION

The elegance of online review posting has grown at a faster rate and people buying almost everything online that gets delivered at their doorsteps. hence, People are not subject to physically inspect the product when buying online so they drastically unwanted/wanted to depend on reviews of other buyers these must be made truth full as much as possible so that buyers is not cheated by fake reviewers time and again. these can be filtered by checking the use of words" awesome fantastic etc". Since they tend to hype the product or try to emulate genuine reviews with the same words using it again and again to make an impact on the buyer.

II. PROBLEM IDENTICATION

Fake reviews that have been created people for various purposes. They write fake reviews to mislead readers detection system by promoting or demoting target product to promote them. The focus of this research is to create and environment of online ecommerce industries where ecommerce build trust in a flat form there the product they purchase at genuine and feedback posted on these websites at true, are checked regularly by the company number of users is increasing day by day, Hence forth companies like twitter WhatsApp Facebook ,use sentiment analysies to check fake news.

III. METHODOLOGY

3.1 Steps Involved to Generate the Abstraction-Based Summary

- **Data Set: I**t is used in amazon academic reviews which contain reviews, ratings, user id, and many other attributes. the useful parameters are retrieved for feature engineering and it contains thousands of original and fake reviews mixed to easily assess.
- **Preprocessing-**it is the first step in analyzing any data set which includes removing unnecessary attributes, stop words, missing words etc. to clean the data set for training purposes. This ensures proper training of the model.
- **Feature Engineering:-**Function involves all the methods to remove unwanted information from the data set is called data cleaning. This set is very necessary to find the gaps and relationships between different attributes and use them to draw valid conclusions.
- **Sampling of Data:-**Huge number of reviews are used in data set the data is subjected to sampling before even fed to the classifier. Sampling is done lower to wait on the classifier that loads the data in chunks .here, different are used to authenticate the fake reviews and then concatenate two columns after labeling return the labeling.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

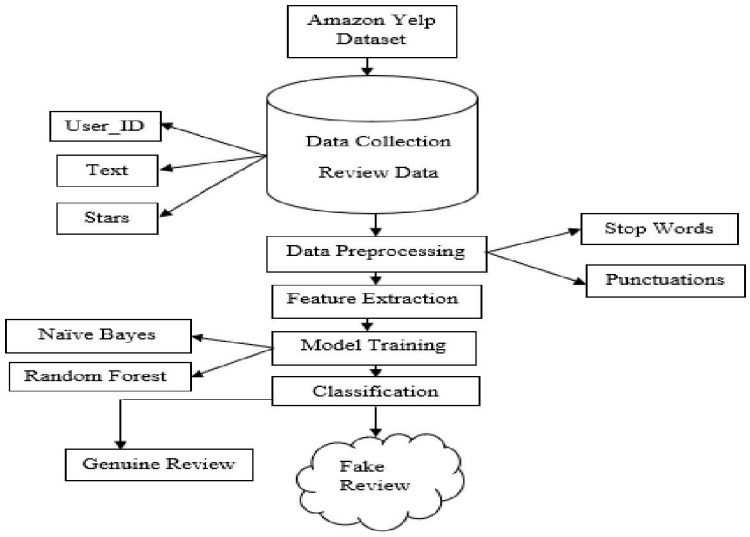

Figure 1: Summary Generation Process

IV. MODELS OUTLINE AND WORKING

4.1 Naïve Bayes Algorithm

A Naive Bayes calculation was utilized to assemble a double arrangement model that would anticipate if the survey's conclusion was positive or negative. A Naive Bayes classifier expects that the estimation of a specific component is free of the estimation of some other element. the information. It expects that all the highlights in the dataset are autonomous and similarly significant. The equations (1), (2), and (3) shown below are the standard form of any Naïve Bayes constituted problem, these are used to compute the probabilities for predicting values that are in the range (0, 1). Where p is a probability, a, b, x_i, y, y_i are values of which probability is calculated, σ is the standard deviation and μ is the mean of the attributes [13]

$$p\left(\frac{a}{b}\right) = \frac{p\left(\frac{b}{a}\right)p(a)}{p(b)} \quad (1)$$

$$Posterior = \frac{prior*livelihood}{excellence} \quad (2)$$

$$p\left(\frac{x_i}{y}\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2{}_y}}\right)\exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma^2{}_y}\right)$$

4.2 Random Forest Classifier

It is a supervised learning algorithm used to train and test machine learning models. The "forest" means an ensemble of decision trees trained with the "bagging" method. Here decision trees are combined to increase the performance and learning of the model to get good overall results. It basically merges multiple decision trees to amplify the performance of the random forest and get a more accurate prediction [13].

1. Accuracy= TP+TN/ FP+FN+TN
2. Precision= TP/TP+FP
3. Recall (sensitivity) = TP/TP+FN
4. F1_score= 2*(Recall*Precision)/ (Recall + Precision)

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

V. TESTING

Sl.NO	Test Name	Test Description	Input	Excepted output	Actual Output
1	User input format	To test user input values	Product review As input	The review should read And display in console	The review read and Displayed in console
2	User input format	To test user input values	Product review as null	Show alert messages please full the fields	Shown alert messages please select the fields
3	Preprocess	Check for data cleaning	Dataset.csv file	Remove the null fields	Removed the null fields
4	Prediction	To test whether its prediction the review is fake or Real	Review text	Predict the review is fake are real based on the historical data using the machine learning model.	Predict the review is fake are real based on the historical data using the machine learning model.

VI. RESULTS

Algorithms	Accuracy Score	Precision Score	Recall Score	F1 Score
Naïve Bayes (in %)	79.007	70.224	99.099	82.169
Random Forest (in %)	89.487	85.577	94.389	89.768

From Table 1, It can infer that the two models performed fairly well except that the random forests classifier is better when compared. Hence random forests have got better accuracy, precision score, and F1 score. It is concluded, a random forest classifier can be used for the fake product review monitoring and removal approach. When compared to the models for diverse applications, they perform well in certain fields and are incompatible in some areas, hence their application needs some experience.

REFERENCES

[1]. Barbosa, Luciano & Feng, Junlan. (2010). Robust Sentiment Detection on Twitter from Biased and Noisy Data. Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference. 2. 36-44.

[2]. Enhanced Sentiment Learning Using Twitter Hashtags and Smileys Dmitry Davidov, Oren Tsur, ICNC / 2, Institute of Computer Science The Hebrew University 2010.

[3]. Go, Alec & Bhayani, Richa & Huang, Lei. (2009). Twitter sentiment classification using distant supervision. Processing. 150.

[4]. Fake review detection using opinion mining" by Dhairya Patel, Aishwerya Kapoor and Sameet Sonawane, International Research journal of Engineering and technology (IRJET) , volume 5, issue 12,Dec 2018.

[5]. Ravi, k. Ravi. 2015. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge based systems, 89.14-46.

[6]. Khan, K. et al., "Mining opinion components from unstructured reviews: A review". Journal of King Saud University – Computer and Information Sciences (2014), http://dx.doi.org/10.1016/j.jksuci.2014.03.009.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 9, June 2022
Impact Factor: 6.252

[7]. "Fake review detection from product review using modified method of iterative computation framework", by EkaDyarWahyuni&ArifDjunaidy, MATEC web conferences 58.03003(2016) BISSTECH 2015.

[8]. Saumya, S., Singh, J.P. Detection of spam reviews: a sentiment analysis approach. CSIT 6, 137– 148 (2018). https://doi.org/10.1007/s40012-018- 0193-0.