International Journal of Scientific Research and Engineering Development-– Volume 3 Issues 2 Mar-Apr 2020

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Predicting Fake online Reviews using Machine Learning

Sabira karim*, Dr. Kiruthiga G**

**Department Of Computer Science and Engineering, I.E.S College of Engineering Thrissur India*

karimsabira@gmail.com

***Department of Computer Science and Engineering, I.E.S College of Engineering Thrissur India*

kirthikacsehod@gmail.com

Abstract: Online reviews are very important in decision making of customer whether to purchase a product or service. These are main source of information getting from the past customer experience about the features of that service which we are going to purchase. This paper introduces some machine learning techniques like Naïve-Bayes, Support Vector Machine and Decision Tree for sentiment classification of reviews and to detect fake online reviews using the data set of a Hotel reviews. Sentiment Analysis has become most interesting in analysis of text. Using sentiment analysis we can separate negative and positive reviews as well.

Index Terms **– Spam reviews, machine learning, Naïve-Bayes, Support Vector Machine, Decision Tree algorithm.**

I. INTRODUCTION

A fake review is a misuse of the user review system by fake personalities. Fake reviews are also generated by bots. Fake reviews mislead customers to take decision on wrong product and the customer spends money on the product. The reviews can be either positive or Negative, to increase the promotion and sale or to bring down the competitive company products. Many people look at online reviews before making a decision whether it should be purchase or not. Many companies depend on several applications to detect Fake reviews using machine learning.

In this paper we use Sentiment Analysis to formulate the data.The sentiment is usually formulated as a two-class classification problem, positive and negative. The basis of Sentiment Analysis is detecting the polarity of a give text or document. In this project we are using a set Polarity as negative or positive.

Recent developments in fields like Natural Language Processing (NLP) has paved the way for accurately understanding people's sentiments, emotions, and behavioral patterns. Emotions such as joy, anger, surprise, disgust can be extracted from the reviews. For example If we want to book a hotel then we checks the reviews of that hotel website and gets the past customer experiences. Online reviews have great impact on customers. This application can detect potential fake reviews in order to reduce the misguidance that follows it.

In machine learning based techniques, there are many algorithms can be applied for the classification and prediction. Here we used Naive-Bayes classifier, Support Vector Machine (SVM), Random Forest Classifier and Decision Tree for predicting the reviews. We detect fake positive, fake negative, True positive and True negative reviews. And finally we compare the accuracy of each algorithm. Main objective of this paper is to classify the dataset or reviews

into true and fake reviews using machine learning techniques.

II. RELATED WORKS

A. *Detecting fake reviews through sentiment analysis*

A number of studies conducted and experiments done on several sample data. Many reviews on product are scraped from the webpage of products and conducted studies. In our work we have decided to use the deceptive opinion spam dataset.

We have extracted the data from the dataset and stored in a list. Then we have createdthe data Frame with corresponding labels.Using sentiment analysis all the reviews is analyzed.The polarity is determined as Positive or Negative.Also we have classified the Spamity as True or Deceptive.Later the polarity class and Spamity class converter into 0and 1s. Then we can apply the algorithm. Mainly we used two method Naïve Bayes classification, Support Vector Machine and Decision Tree.

Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems

III. PROPOSED WORKS

We are using the deceptive dataset. The Deceptive opinion spam dataset is a corpus consisting of truthful and deceptive hotel reviews of 20 Chicago hotels. The corpus contains 400 truthful, positive reviews from Trip Advisor, 400 deceptive positive reviews from Mechanical Turk, 400 truthful negative reviews from Expedia, Hotels.com, Orbitz,Priceline, Trip Advisor, & Yelp and 400 deceptive negative reviews from Mechanical Turk. In total we have 1600 reviews.Our task is to classify the truthful and deceptive hotel reviews using a machine learning algorithm.

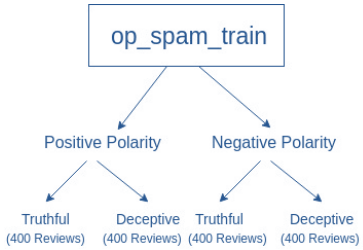

Fig.1 Structure of the reviews set on which we are going to work

A. *Data preprocessing*

We have created the data frame with three columns named reviews, polarity class and spamity class. The column named reviews gives the text or reviews posted by the customerwhereas spamity class shows whether it is deceptive or True. And the polarity class shows that whether the polarity is positive or negative.

The extracted review look like the below table:

Table 1

0	**polarity_class**	**review**	**spamity_class**
0	negative	"My $200.."	t
1	negative	"This was a .."	t
2	negative	"The hotel.."	t
3	negative	"Going to.."	t
...	...	...	.....	.....

[1600 rows x 3 columns]

First of all we need to remove the stopwords from the reviews. For removal of stop words we used **nltk** package from sklearn. Text mining techniques have been applied and the strings are converted into numbers.Extracted parts of speech from reviews which will be fed as a Feature Input to the model.The reviews are stored as array format. The spamity class and Polarity class of dataFrame converted into 0's or 1's instead of True or False then the model can be created. The dataset contains only 1600 rows so that we splitthe data into 80:20ratiosfor train and test, stored as an array.

B. Model selection and Prediction

To fit the model we have used sklearn Of Python programming language provides the needful libraries for the classifiers.We have different classification techniques in machine learning like Naïve-Bayes, Support Vector Machine, Decision Tree and Random Forest classifiers. We have applied different predictions methods to reach the more accurate model.

Random Forest algorithm isan Ensemble model which creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting .algorithm creates decision trees on data samples and then gets the prediction from each tree and finally selects the best solution from that. This produces the highest accuracy. Random Forest can also use for classification as well as Regression analysis.

Naïve-Bayes is popularly used for text categorization to predict the text with word frequencies as the features. Naïve –Bayes typically uses bag-of-words feature from NLP to identify the fake in text categorization.

SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. For SVM classifier we have gamma parameter keeping constant for perfect fit model.

IV. RESULTS AND PERFOMANCE ANALYSIS

A. Experimental Environment

We have applied our experiments on a machine with Processor: Intel core i3 – 2330M and CPU- 2GHz, RAM: 4GB,S system with 64 bit OS, We have used Windows as an operating system. We have used Python as programming language with sklearn, numpy and pandas packages. Spyder 4.1.1 used as IDE.

B. Results

We have used Naïve Bayes classifier, Support Vector Machine (SVM), Decision Tree and Random Forest classifiers to classify the reviews dataset. We have divided the dataset of 1600 rows with 3 columnswith column names reviews, polarity and spamity for each classification process. The data split into train and test in the ratio 80:20.

For Naïve-Bayes classification we applied Multinomial NB. After fitting the model the predicted data NB has given accuracy of 90.31%. SVM has given accuracy of 83.75 % Using Decision Tree algorithm we got the accuracy of 66.56 %. Comparing each of the algorithms, we found that Random Forest is giving highest accuracy, secondly the Naïve-Bayes and Decision Tree classifier has given the least accuracy.

TABLE 2: Accuracy Comparison

NB	SVM	DT	RFC
90.31	84.166	66.56	92.7279

C. Performance Analysis

We can choose Random Forest Classifier as well as Naïve-Bayes as our model since both are giving highest accuracy.By choosing Random Forest Classifier, we could improve the performance accuracy up to 92.7 percent. It is the highest accuracy compared to the other techniques. By importing metric from skleaarn package we can have the confusion metric for the same predictions. Confusion metrics has given the perfect accuracy for each algorithm.

i. The Accuracy graphical representation of each algorithm shown in figure2

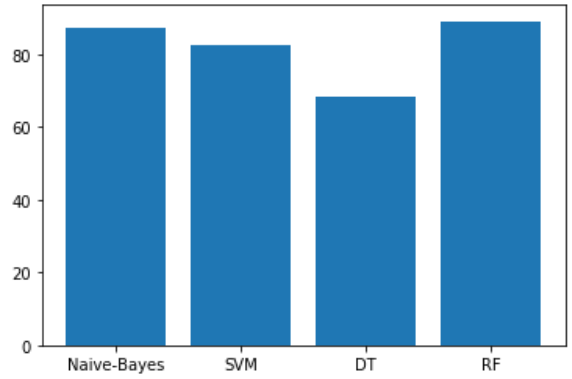

Fig 2

ii. ROC CURVE

The true positive rate is calculated as the number of true positives divided by the sum of the number of true positives and the number of false negatives. It describes how good the model is at predicting the positive class when the actual outcome is positive. The true positive rate is also referred to as sensitivity.

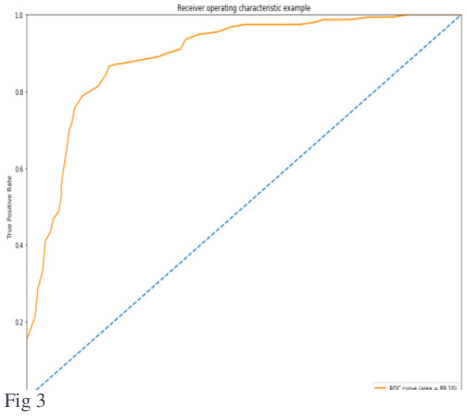

Fig 3

The ROC curve is a graph with: The x-axis showing 1 – specificity (= false positive fraction = FP/(FP+TN)) The y-axis showing sensitivity (= true positive fraction = TP/(TP+FN))

In a Receiver Operating Characteristic (ROC) curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points.

Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. Area under curve (AUC) is a summary measure of the accuracy of a quantitative diagnostic test. It is 89.01

V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed several methods to analyze a dataset of hotel reviews. We also presented sentiment classification algorithms to apply a supervised learning of the hotel reviews dataset.

For future work, we would like to extend this study to use other datasets such as Amazon dataset or eBay dataset and use different feature selection methods. Furthermore, we may apply sentiment classification algorithms to detect fake reviews using various tools such as Python and R or R studio, Statistical Analysis System (SAS), and Stata; then we will evaluate the performance of our work with some of these tools.

ACKNOWLEDGEMENT

This research was supported by Technical University of Kerala. We are thankful to our colleagues who provided expertise that greatly assisted the research, although they may not agree with all of the interpretations provided in this paper.

REFERENCES

[1] Rakibul Hassan and Md. Rabiul Islam "*Detection of fake online reviews using semi-supervised and supervised learning*" 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)

[2] Chengai Sun, Qiaolin Du and Gang Tian, "*Exploiting Product RelatedReview Features for Fake Review Detection*," Mathematical Problems inEngineering, 2016.

[3] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, "*Detection ofreview spam: a survey*", Expert Systems with Applications, vol. 42, no.7, pp. 3634–3642, 2015.

[4] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "*Finding deceptiveopinion spam by any stretch of the imagination*," in Proceedings ofthe 49th Annual Meeting of the Association for Computational Linguistics:Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319,Association for Computational Linguistics, Portland, Ore, USA, June
2011.

[5] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "*Linguistic Inquiryand Word Count: Liwc,*" vol. 71, 2001.

[6] S. Feng, R. Banerjee, and Y. Choi, "*Syntactic stylometry for deceptiondetection*," in Proceedings of the 50th Annual Meeting of the Associationfor Computational Linguistics: Short Papers, Vol. 2, 2012.

[7] J. Li, M. Ott, C. Cardie, and E. Hovy, "*Towards a general rule foridentifying deceptive opinion spam*," in Proceedings of the 52nd AnnualMeeting of the Association for Computational Linguistics (ACL), 2014.

[8] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "*Detectingproduct review spammers using rating behaviors,*" in Proceedings ofthe 19th ACM International Conference on Information and KnowledgeManagement (CIKM), 2010.

[9] J. K. Rout, A. Dalmia, and K.-K. R. Choo, "*Revisiting semi-supervisedlearning for online deceptive review detection,*" IEEE Access, Vol. 5,pp. 1319–1327, 2017.

[10] J. Karimpour, A. A. Noroozi, and S. Alizadeh, "*Web spam detection bylearning from small labeled samples*," nternational Journal of ComputerApplications, vol. 50, no. 21, pp. 1–5, July

