International Journal of Scientific Research and Engineering Development-– Volume 4 Issue 2,Mar- Apr 2021

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Supervised Machine Learning Based Approach to Detect Fake Online Hotel Reviews

Vijay*, Shubham Kumar**

*(Computer Science and Engineering, Neelkanth Institute of Technology)

** (Computer Science and Engineering, Bharat Institute of Technology)

----------**********----------

Abstract:

Online reviews regarding a product or a service influence the customers. Customers will always prefer to purchase the product which has good online reviews rather than the product having bad online reviews. Customers write online reviews to share their experiences which can be used by other customers for decision making regarding a particular product or service. Online reviews are very helpful in decision making if they are true. Sometimes people write fake reviews either to make their product or service popular or to mar the reputation of their competitors. Customers may regret in future if they make decisions on the basis of fake reviews. There is a need to detect fake online reviews. In this paper we used four machine learning based text classifiers to detect fake online reviews of hotels and discussed their performances. The four classifiers which we used in this paper are: Naïve Bayes, Random Forest, Support Vector Machine (SVM) and Classification and Regression Trees (CART). Among these four algorithms Naïve Bayes gave maximum accuracy.

Keywords **— Text Classification, machine learning, SVM, CART, Naïve Bayes, Random Forest.**

----------**********----------

I. INTRODUCTION

While deciding whether to book a particular hotel or not, customers can take help of online reviews [1]. Many tourists write reviews on social media, forums and other websites to share their experiences. These reviews can be used by other people to make decisions. These online reviews can also be used by business owners to design strategies to make their hotels popular. Tourism industry is greatly influenced by the electronic word of mouth (eWOM) [2]. Online reviews are beneficial in decision making as long as they are true. Making decisions on the basis of false reviews can be harmful. False negative reviews can damage the reputation of a hotel which may result in the decline in number of its customers. People who write fake reviews are called opinion spammers [3].

Our purpose is to detect whether an online review of hotel is true or false. In this paper we used supervised machine learning based text classifiers to classify the online reviews as true or false. In supervised machine learning there is a need of already classified training data. Text classifiers were trained to classify a review as true or false on the basis of its content. Text classification is the process of labelling categories (from an already defined set of categories) to text documents [4]. We trained four different text classifiers by already classified (as true or false) online hotel reviews: Naïve Bayes, Random Forest, Support Vector Machine (SVM) and Classification and Regression Trees (CART). After training these four classifiers were tested on test data to check how well they can predict whether a given text document is a false review or a true review. Their accuracies were compared.

II. DATA DESCRIPTION

The dataset which we used for training and testing was "deceptive-opinion.csv". We downloaded it from kaggle.com. This dataset contains 1600 reviews of 20 hotels of Chicago. 800 reviews in dataset were labelled as truthful and remaining 800 were labelled as deceptive. The dataset also contains the polarity of the review, name of the hotel about which review was written and the source of the review. Our purpose is to classify the review as true or false on the basis of its content. Therefore, we did not include name of the hotel, polarity of the review and source of review for training and testing of classifiers.

III. DATA PREPROCESSING

Quality of a dataset can be improved for Text Classification by preprocessing [6]. Therefore, we first cleaned the dataset and then converted it into the form of document term matrix so that machine learning based algorithms can be easily applied on it. Original dataset had five columns. First we removed the columns specifying the polarity of the review, name of the hotel and the source of review from the dataset "deceptive-opinion.csv". After that dataset had only two columns which were labelled as text and deceptive. Column labelled as text contained all text reviews and in column named deceptive there was one value for each text review i.e. either truthful or deceptive specifying whether the text review is true or false respectively.

A. Data Cleaning

A text document contains many words which are not useful for training a classifier and such useless words are known as stopwords [7]. To clean the data, first we converted the text reviews into lower case then we removed all punctuations, stopwords, numbers and URLs from them. After that we stripped white spaces from text reviews.

B. Document term Matrix

After cleaning the data all text reviews were transformed in a form of document term matrix. Document term matrix is a two dimensional matrix which specifies the frequency of occurrences of terms in various documents. In document term matrix rows specify documents and columns specify terms.

IV. TRAINING AND TESTING OF CLASSIFIERS

For training the text classifiers 1200 rows of document term matrix corresponding to the 1200 text reviews were used. The remaining 400 rows of document term matrix corresponding to remaining 400 text reviews were used for testing purpose.

A. Naive Bayes

Naïve Bayes classifier is a probabilistic classifier [8]. It is based on Bayes theorem. According to Bayes theorem the probability of occurrence of X when C has already been occurred is given by the formula:

$$\boldsymbol{P(X|C) = \frac{P(C|X)P(X)}{P(C)}}$$

According to Bayes theorem the probability that document d belongs to category c is given by [9]:

$$\boldsymbol{P(c|d) = \frac{P(d|c)P(c)}{P(d)}}$$

Suppose document d is represented as features x1, x2,……,xn then according to Bayes theorem the probability that document d belongs to class c is given by:

$$\boldsymbol{P(c|x1, x2, \ldots..xn) = \frac{P(x1|c)P(x2|c).....P(xn|c)P(c)}{P(x1)P(x2).....P(xn)}}$$

1200 rows of document term matrix corresponding to 1200 text reviews were used for training the Naïve Bayes text classifier. In these 1200 text reviews 596 reviews were labelled as deceptive and remaining 604 reviews were labelled as truthful. After training the classifier was tested on test data consisting of remaining 400 rows of document term matrix corresponding to remaining 400 text reviews. In test data 196 text reviews were truthful and remaining 204 reviews were deceptive. Classes predicted by the trained Naïve Bayes classifier for test data is compared with the real classes of the test data. The Naïve Bayes classifier correctly predicted 180 deceptive reviews as deceptive and 174 truthful reviews as truthful. It incorrectly predicted 22 truthful reviews as deceptive and 24 deceptive reviews as truthful. The

accuracy of Naïve Bayes classifier was 88.5%. Fig. 1 shows the confusion matrix and statistics of test results for Naive Bayes classifier.

```
Confusion Matrix and Statistics

          Reference
Prediction  deceptive truthful
  deceptive       180       22
  truthful         24      174

               Accuracy : 0.885
                 95% CI : (0.8496, 0.9146)
    No Information Rate : 0.51
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.77

 Mcnemar's Test P-Value : 0.8828

            Sensitivity : 0.8824
            Specificity : 0.8878
         Pos Pred Value : 0.8911
         Neg Pred Value : 0.8788
             Prevalence : 0.5100
         Detection Rate : 0.4500
   Detection Prevalence : 0.5050
      Balanced Accuracy : 0.8851

       'Positive' Class : deceptive
```

Fig. 1 Confusion Matrix and Statistics for Naive Bayes Classifier

B. Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm. It can be used for classification as well as regression [10]. In this algorithm data items are plotted as points in n-dimensional plane and then hyperplane is found to differentiate the classes of points. SVM finds the hyperplane with maximum margin. We used two SVMs in our experiment. One SVM was with linear kernel and another SVM was with radial kernel. Both were trained with 1200 rows of document term matrix corresponding to 1200 text reviews in which 603 were deceptive and 597 were truthful. Both of these trained SVMs were tested on remaining 400 rows of document term matrix with 197 rows corresponding to deceptive text reviews and 203 rows corresponding to truthful. We got 87% accuracy from SVM with radial kernel and 85.5% accuracy from SVM with linear kernel. Fig. 2 shows the confusion matrix and statistics of test results for SVM with radial kernel.

```
Confusion Matrix and Statistics

          Reference
Prediction  deceptive truthful
  deceptive       175       30
  truthful         22      173

               Accuracy : 0.87
                 95% CI : (0.833, 0.9014)
    No Information Rate : 0.5075
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.7401

 Mcnemar's Test P-Value : 0.3317

            Sensitivity : 0.8883
            Specificity : 0.8522
         Pos Pred Value : 0.8537
         Neg Pred Value : 0.8872
             Prevalence : 0.4925
         Detection Rate : 0.4375
   Detection Prevalence : 0.5125
      Balanced Accuracy : 0.8703

       'Positive' Class : deceptive
```

Fig. 2 Confusion Matrix and Statistics for SVM with radial kernel

C. Classification and Regression Tree (CART)

Classification and regression trees (CART) was proposed in 1984 by Breiman [11]. It is a decision tree algorithm which can be used for classification as well as regression. It is an algorithm which constructs decision trees. A decision tree consists of internal nodes which are decision nodes and leaf nodes which represent the results. CART generates strictly binary decision trees. CART is a supervised machine learning algorithm and needs to be trained by already classified data. The same rows of document term matrix which were used for training and testing of SVM are also used for training and testing of CART. We got 72.25% accuracy from CART. Fig. 3 shows confusion matrix and statistics of test results for CART.


```
Confusion Matrix and Statistics

          Reference
Prediction  deceptive truthful
  deceptive       131       45
  truthful         66      158

               Accuracy : 0.7225
                 95% CI : (0.6758, 0.7658)
    No Information Rate : 0.5075
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.444

 Mcnemar's Test P-Value : 0.05765

            Sensitivity : 0.6650
            Specificity : 0.7783
         Pos Pred Value : 0.7443
         Neg Pred Value : 0.7054
             Prevalence : 0.4925
         Detection Rate : 0.3275
   Detection Prevalence : 0.4400
      Balanced Accuracy : 0.7216

       'Positive' Class : deceptive
```

Fig. 3 Confusion Matrix and Statistics for CART

D. Random Forest

It is a supervised machine learning algorithm. It can be used for both classification and regression. It is an ensemble algorithm. It builds a number of decision trees. For predicting the class of the test objects Random Forest classifier considers the prediction of all decision trees. Its prediction is based on the majority of votes. The Random Forest which we used built 20 decision trees. The same rows of document term matrix which were used for training and testing of SVM and CART are also used for training and testing of Random Forest. We got 79.5% accuracy from Random Forest. Fig. 4 shows the confusion matrix of test results for Random Forest.

```
            Actual
Prediction  deceptive truthful
  deceptive       167       52
  truthful         30      151
> |
```

Fig. 4 Confusion Matrix for Random Forest

V. CONCLUSION

In our experiment we used four supervised machine learning algorithms for detecting fake reviews. We got maximum accuracy from Naïve Bayes classifier and minimum accuracy from CART.

REFERENCES

[1] Casalo LV, Flavian C, Guinaliu M, Ekinci Y (2015) Do online hotel rating schemes influence booking behaviors? Int J Hosp Manag 49:28–36

[2] Bompotas A., Ilias A., Kanavos A., Makris C., Rompolas G., Savvopoulos A. (2020) A Sentiment-Based Hotel Review Summarization Using Machine Learning Techniques. In: Maglogiannis I., Iliadis L., Pimenidis E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2020 IFIP WG 12.5 International Workshops. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 585. Springer, Cham. https://doi.org/10.1007/978-3-030-49190-1_14

[3] Algur, S., Hiremath, E., Patil, A. and Shivashankar, S., "Spam Detection of Customer Reviews from Web Pages." In Proceedings of the 2ndInternational Conference on IT and Business Intelligence.2010.

[4] Sebastiani, F., Text categorization. In Zanasi, A., editor, Text Mining and itsApplications, pages 109–129, Southampton, UK: WIT Press, 2005.

[5] https://www.kaggle.com/rtatman/deceptive-opinion-spam-corpus

[6] HaCohen-Kerner Y, Miller D, Yigal Y (2020) The influence of preprocessing on text classification using a bag-of-words representation. PLoS ONE 15(5): e0232525. https://doi.org/10.1371/journal.pone.0232525

[7] Ikonomakis, Emmanouil & Kotsiantis, Sotiris & Tampakas, V.. (2005). Text classification: a recent overview. 125.

[8] Meena, M. J., & Chandran, K. R. (2009). Naïve Bayes text classification with positive features selected by statistical method. 2009 First International Conference on Advanced Computing. doi:10.1109/icadvc.2009.5378273

[9] Dey Sarkar, S., Goswami, S., Agarwal, A., & Aktar, J. (2014). *A Novel Feature Selection Technique for Text Classification Using Naïve Bayes. International Scholarly Research Notices, 2014, 1–10.* doi:10.1155/2014/717092

[10] V. Vapnik. The Nature of Statistical Learning Theory. NY: Springer-Verlag. 1995.

[11] BREIMAN L, FRIEDMAN J H, OLSHEN R A, et al. “Classification and Regression Tree”. Wadsworth International Group, U.S.A, 1984.

