Proceedings of the Sixth International Conference on Inventive Computation Technologies [ICICT 2021]
IEEE Xplore Part Number: CFP21F70-ART; ISBN: 978-1-7281-8501-9

Opinion Mining based Fake Product review Monitoring and Removal System

Syed Mohammed Anas
Dept. of CSE,
M S Ramaiah University of Applied Sciences,
Bangalore, India
syedanas.570@gmail.com

Santoshi Kumari
Dept. of CSE,
M S Ramaiah University of Applied Sciences,
Bangalore, India
syedanas.570@gmail.com

2021 6th International Conference on Inventive Computation Technologies (ICICT) | 978-1-7281-8501-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICICT50816.2021.9358716

Abstract—**Fake review detection and its elimination from the given dataset using different Natural Language Processing (NLP) techniques is important in several aspects. In this article, the fake review dataset is trained by applying two different Machine Learning (ML) models to predict the accuracy of how genuine are the reviews in a given dataset. The rate of fake reviews in E-commerce industry and even other platforms is increasing when depend on product reviews for the item found online on different websites and applications. The products of the company were trusted before making a purchase. So this fake review problem must be addressed so that these large E-commerce industries such as Flipkart, Amazon, etc. can rectify this issue so that the fake reviewers and spammers are eliminated to prevent users from losing trust on online shopping platforms. This model can be used by websites and applications with few thousands of users where it can predict the authenticity of the review based on which the website owners can take necessary action towards them. This model is developed using Naïve Bayes and random forest methods. By applying these models one can know the number of spam reviews on a website or application instantly. To counter such spammers, a sophisticated model is required in which a need to be trained on millions of reviews. In this work ”amazon Yelp dataset” is used to train the models and its very small dataset is used for training on a very small scale and can be scaled to get high accuracy and flexibility.**

Keywords—opinion mining, sentiment analysis, text mining

I. INTRODUCTION

The elegance with 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 unwantedly/wontedly depend on reviews of other buyers this must be made truthful as much as possible so that the buyer is not cheated with fake reviewers or spammers time and again. The problem is simple yet tiring to be accomplished through/read every review to mark it as a fake or ambiguous category this must be done systematically to get to the root of the problem. This problem can be addressed by training an ML model which deals with the review section to flag a particular review as genuine or spam. The interesting thing is spammers who didn't use the product can be caught this way. A spam review or the usage of different customer id can be used to filter review of the product falsely to get a good rating of the product. This can be filtered by checking the use of words like "awesome", "so good", "fantastic" etc. can be flagged. Since they tend to hype the product or they try to emulate genuine reviews with the same words using it again and again to make an impact on the buyer. Hence the issue of spam filtering requires huge data to train and be effective with added domain knowledge such as sarcasm sentences used by users to show their dissent towards the product, sometimes the product is good but not the delivery or the packing which affects the review classification. Here, an NLP technique is used to identify such reviews instead of misclassification to a negative review as in sentiment analysis. To remove unwanted or outdated product reviews those include data pre-processing.

The focus of this research is to create an environment of online E-commerce industry where consumers build trust in a platform where the products they purchase are genuine and feedbacks posted on these websites/applications are true, are checked regularly by the company where the number of users is increasing day by day, henceforth companies like Twitter, WhatsApp, Facebook use sentiment analysis to check fake news, harmful/derogatory posts and banning such users/organizations form using their platforms. Parallel to that E-commerce (Flipkart, Amazon) industries, hotels booking (Trivago), logistics, tourism (Trip Advisor), job search (LinkedIn, Glass door), food (Swiggy, Zomato), etc. use algorithms to tackle fake reviews, spammers to deceive the consumers in buying below average products/ services. And the users need to be alerted of the spammer like “not verified profile” hence users need not worry about such false users. Manual labelling of the reviews is practically time-consuming and less effective. So supervised learning model is used for labelling the reviews and then predicting the label is not feasible. For example, Mukherjee et al. manually labelled 2431 reviews for over 8 weeks, so automated review labelling should be possible to contain time and energy which is difficult and proposed by Sunil Soumya et.al. Some industries pay money to write fake reviews of their products and services where it is not possible to label the review as spam or not spam. Amazon’s “Yelp” dataset has 30% to 40% of spam reviews. Feature selection is an important aspect in selecting and training these models. In this work, the comparison of two models developed to justify the model performance for this “Amazon’s yelp” dataset and their relevance to deploy these models in real-time software. The Random Forests (RF) model performed well compared to the Naïve Bayes algorithm by a large margin in the fake review data analysis. The fake review detection problem is addressed fairly and gives a fair insight into its legality and need. The purpose is to select a suitable algorithm to fulfill the task of fake review detection and its elimination.

The rest of the article is arranged as follows: Section II describes the related works, Section III explains methods and datasets, and Section IV depicts the model outline and working. Section V demonstrates the result and analysis. Finally, Section VI draws the conclusion of the research work along with the future scope of the research.

II. RELATED WORK

The previous analysis is done on the expressed views through text, blogs, reviews, feedbacks, etc. as opinions by users which are unique to compute, study to obtain relevant information, that is nothing but sentiment analysis.

Exiting research used a two-step approach, SVM classifier for classifying tweets [1]; other used emoticons, smileys, and Hashtags to classify labels into multiple sentiments [2]. The other researcher used an SVM classifier for training data using emoticons [3].

2.1 Existing systems:

2.1.1 Lexicon-based methods: - Based on counting the number of positive and negative words in a sentence- Twitter.

2.1.2 Rule-based methods: – Based on syntactic rules, e.g., [query] is pos-adj – Tweet feel

2.1.3 Machine learning based methods: – Based on the classifier built on a training data Twitter sentiment [14].

The fake review detection problem the researchers have developed techniques to address the problem of fake review detection. New models like the ICF++ model that uses honesty value, their research influenced the accuracy and increased by 49% [7]. VADER and Polarity based approach was used to flag the reviews as true, false, suspected categories and assign polarity +1, -1, and 0 to identify/classify fake reviews and eliminate them using this technique [4]. The review of all the other methods and techniques used by researchers of the past decade is a great collection of vast literature compiled for sentiment analysis and gives in-depth information on methods related to fake review detection [5]. Spam review detection based on comments made on the reviews to help in sensing the review reliability and its truthfulness and the method achieved 91% of F1_score for this model [8]. The detection of spam reviews in singleton reviews using singleton spam review correlated temporal pattern was followed [9]. The algorithm in practice KL divergence essentially used to differentiate fake reviews from the original due to its asymmetric property is an issue of pseudo fake review [10]. The sequence of reviews used to filter spam ones used feature extraction up to six times to classify fake reviews and genuine reviews [6]. Another researcher proposed spam review detection using new concept time series prediction method which uses pattern recognition to know the suspicious time-intervals in which spam review was posted [11]. The author's utilized activeness, context similarity, the behavior of ratings of review to compute spam score explored deep neural networks to know models behavior in detecting spam opinions, where recurrent and convolution networks were also monitored to convert raw text to vectors which in turn used as features to locate spam reviews [12].

III. METHODS

1) Dataset

Dataset used is "amazon academic review" which contains reviews, useful votes, ratings, user id, and many other attributes. The useful parameters are retrieved for feature engineering. The dataset contains thousands of original and fake reviews mixed to easily assess the accuracy of the model being implemented using this dataset. The Yelp dataset released for the academic challenge contains information for 11,537 businesses. This dataset has 8,282 check-in sets, 43,873 users, 229,907 reviews for these businesses (www.yelp.com/dataset). The dataset is challenging since it contains a large set of varied reviews and parameters for training any algorithm.

2) Pre-processing

Pre-processing is the first step in analyzing any dataset which includes removing unnecessary attributes, punctuations, stop words, missing words, redundant words, etc. to clean the dataset for training purposes. This ensures proper training of the model.

3) Feature Engineering

This function involves all the methods to remove unwanted information from the dataset it is also called data cleaning. This step is very necessary to find the gaps and the relationship between the different attributes (columns) and use them to draw valid conclusions. The libraries from the NLTK package is a bag of words used to construct a corpus of words. Term frequency, tokenizer, Stopwords functions are imported from OrderedDict. Stop words are removed and unwanted words like is, then, to, why, etc. which are not required in this context and do not add value to feature engineering are grouped under Stopwords coming under the English language. Term frequency counts the number of times a particular word has occurred and that can be used by spammers again and again to identify the spammer.

4) Sampling of data

Since a huge number of reviews are used in the dataset the data is subjected to sampling before even fed to the classifier. The sampling is done to lower to weight on the classifier that loads the data in chunks. Here, different labels are used to authentic the fake reviews and then concatenate two columns after labelling and return the data frame.

IV. MODEL OUTLINE AND WORKING

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, given the class variable. It utilizes the preparation information to compute the likelihood of every result dependent on the highlights. One significant trait of the Naive Bayes calculation is that it makes suspicions about

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].

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

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

$$p\left(\frac{x_i}{y}\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2{}_y}}\right)\exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma^2{}_y}\right) \quad (3)$$

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].

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 the above parameters determine the performance of the model, the results of models are shown along with the confusion matrix.

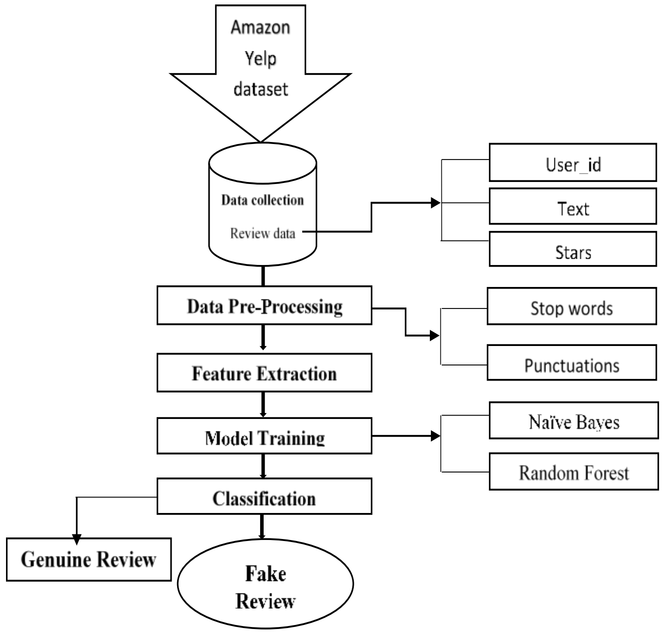

Figure 1: Model diagram for fake review detection

The flowchart in "Fig 1" is explained as follows, the start of the problem solving is done with dataset collection which requires the attention to select the correct dataset to check whether it is binary or categorical. To get the data in the required format for building the model, I loaded the reviews in the Yelp academic dataset review. Json file. Later for brevity, only those attributes were shortlisted which is useful in future events. Feature extraction is done and used to train Random forests and Naïve Bayes, where it records the relationships between different attributes and then uses them for classification. After training the model, the model is fed with the new data or test data where its classification accuracy, etc. as illustrated in table 1 are computed and tuned accordingly for better results. The measurement parameters of this model are confusion matrix, accuracy, precision, sensitivity, F1_score.

V. Results and Discussion

Table 1: Compiled Results of Both models

S. No	Parameter	Naïve Bayes (in %)	Random Forests (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, 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 incompatible in some areas, hence their application needs some experience.

VI. CONCLUSION AND FUTURE SCOPE

The results discussed in this article are the comparison of two models developed to justify the model performance for this "Amazon's yelp" dataset and their relevance to deploy these models in real-time software. Hence Random forests model performed well compared to the Naïve Bayes algorithm by a large margin. The fake review detection problem is addressed fairly and gives a fair insight into its legality and need, the purpose is to select an algorithm to fulfill the task of fake review detection and its elimination. In future work, hybrid models and new models can be tried for the fake review detection model. By using Google co-lab and NVIDIA graphics GPU, the research can speed up the process of execution.

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 SameetSonawane, 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.

[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.

[9] Xie S, Wang G, Lin S, Yu PS (2012) Review spam detection via temporal pattern discovery. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 823–831.

[10] 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.

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

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

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

[14] Joseph, S. I. T. (2019). SURVEY OF DATA MINING ALGORITHM'S FOR INTELLIGENT COMPUTING SYSTEM. Journal of trends in Computer Science and Smart technology (TCSST),1(01), 14-24.