



E-Commerce Product Rating Based on Customer Review Mining

Pankaj Hatwar¹, Sneha Dhepe², Sneha Fale³, Supriya Gedam⁴, Nitish Kumar Minz⁵
Assistant Professor¹
Department of CT
PCE, Nagpur, India

Abstract:

Online reviews have become an important factor when people make purchase and business decisions. Seller selling products on the web often ask or take reviews from customers about the products that they have purchased. As e-commerce is growing and becoming popular day-by-day, the number of reviews received from customer about the product grows rapidly. For a popular product, the reviews can go upto thousands. The increasing popularity of online reviews also stimulates the business of fake review writing, which refers to paid human writers producing deceptive reviews to influence readers' opinions. Our project tackles this problem by building a classifier that takes the review text and the basic information of its reviewer as input and outputs whether the review is reliable. This creates difficulty for the potential customer to read them and to make a decision whether to buy or not the product. Problems also arise for the manufacturer of the product to keep track and to manage customer opinions. And also additional difficulties are faced by the manufacturer because many other merchants sites may sell the same product at good ratings and the manufacturer normally produces many kinds of products. In this research, we aim to summarize all the customer reviews of a product and compare the products based on reviews can be done on one place. This summarization task is different from traditional text summarization, because we only mine the information of that product on which the customers have expressed their opinions and whether the opinions are positive or negative. We do not summarize the reviews by selecting a rewrite some of the original comment, from the reviews to capture the main points as in the classic text summarization. Our experimental results using reviews of a number of products sold online demonstrate the effectiveness of the techniques. Application performs the best with a detection accuracy of 81.92%.

Keywords: Amazon E-Commerce dataset, Dataset acquisition, Support Vector Machine.

I. INTRODUCTION

“What other people thoughts are and their thinking” has always been an important source of information for most of us during the decision-making process. With the rapid expansion of e-commerce, many products are sold on the Web, and many people are also buying products online. In order to enhance customer satisfaction, requirements and online shopping experience, it has become a common practice for online merchants to enable their customers to suggest opinions on the products that they have purchased. With more and more common users becoming comfortable with the Web, a growing number of people are writing reviews and posting them which are becoming beneficial for others.

As a result, the number of reviews that a product receives grows rapidly. Some popular products can get hundreds of reviews at some large merchant sites. In order to provide users with more reliable review information, we aim to build a classification system to detect fake reviews. The input to our algorithm is a review and the related information of the reviewer. Our application will give you the promising reviews by filtering them from other sites. And then you can decide what you want to buy or not.II.

II. LITRATURE REVIEW

A successful attack needs to take into consideration the factors that are most influential to its success. In our work, we identify and examine the following two factors:

1. Authenticity: In order for a fake review to be convincing, and thus less likely to be detected, it needs to be as authentic-looking as possible.

2. Impact: The injected fake review needs to be written in a way that maximizes the (positive or negative) impact on the target's reputation. In the following sections, we formalize and discuss each of these two factors in detail. A number of significant cases have been reported in the news regarding this issue and media investigations revealed that the number of fake reviewers is on the rise as there are many sites which pay for every fake review written on different websites. In copy and close copy reviews were thought to be fake surveys in model building. In order to solve this malignant problem, we propose an interactive semi-supervised model to identify fake reviews which is evaluated later on using real life data and compared with some sophisticated prior research work. A lot of consumers, when searching online for something to buy, will take a look at an online review or rating for a product. First, look for products with a high average user rating, many reviews, and not a lot of variance in the rating scores. It is very important to know if the reviews are genuine because fake reviews lead to false reputation of the product and mislead the user. Generating a system which can detect fake review is very helpful and has a bright scope in the e-commerce industry.

III. EXISTING WORK

When performing any type of internet shopping, many of the users will spend their quality time into reading other user reviews if they are available. Clearly consumers value the feedback given by

other users as do the companies that sell such products. Blogs, websites, discussion boards etc. are a repository of customer suggestions which are valuable and important sources of textual data. Therefore, today's individuals and older ones extensively rely on reviews available on line. It means that people make their decisions of whether to purchase the products or not by analyzing and reflecting the existing opinions on those products. The fact that is if the potential customer or users gets a genuine overall impression of a product by considering the present affect for that product, it is highly probable that he will actually purchase the product. Normally if the percentage of positive and effective opinions is considerable, it is likely that the overall impression will be highly positive. Likewise, if the overall impression is not proper, it is doubtful that they don't buy the product. Now the customers can write any opinion text, this can motivates the individuals, and organizations to give undeserving fake opinions to promote or not to credit some target products, services, organizations, individuals, and even ideas without disclosing their true intentions. These faked opinion information is called opinion fake. In a dataset of movie review is used to detect fake reviews by using Classification Algorithms through Sentiment Analysis. Popular supervised classifiers: Perceptron algorithms are used. In first dataset is collected from the source then is preprocessed, then after which our classification algorithms are trained which is then followed by the detection process, and then a confusion matrix is generated which classifies the reviews as number of correct and incorrect predictions of value. The dataset was classified by using the confusion matrix, and also a very important part of the study because the reviews were classified from datasets whether they are fake or real reviews.

IV. PROPOSED METHODOLOGY

As most of the people require review about a product before spending their money on the product. So people come across various reviews in the website but these reviews are genuine or fake is not identified by the user. They give good reviews for many different products manufactured by their own firm. User will not be able to find out whether the review is genuine or fake. The system will find out whether the review is genuine or fake. The user will view various products and will give review about the product. And the user will get genuine reviews about product.

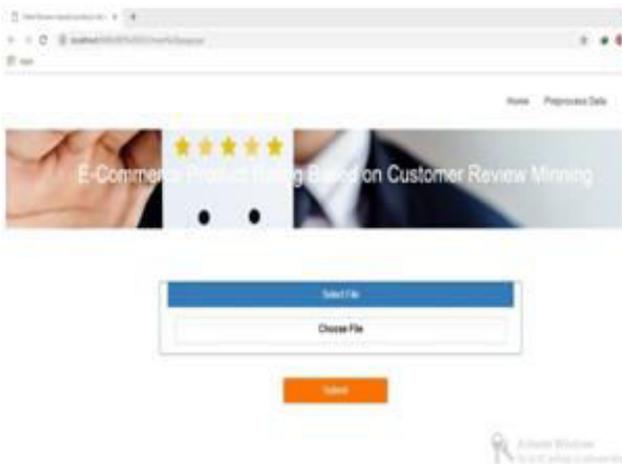
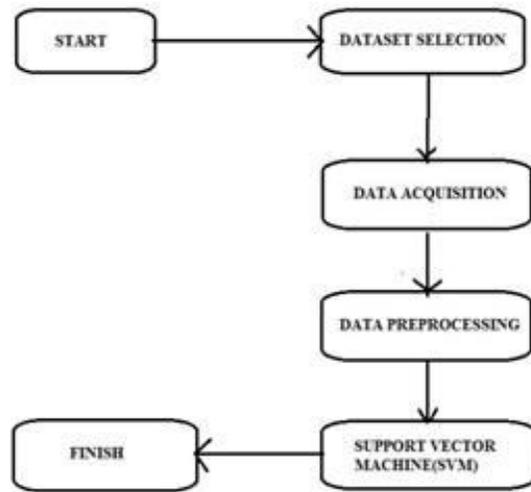


Figure.1. Homepage of the application.

• FLOW CHART OF PROPOSED WORK



• DATASETS

In our proposed system, the review datasets of the products are directly taken in online from the amazon website www.amazon.com by using the import.io tool After feeding the online data set we have to download a .csv file of our desired product, We can collect the reviews of numerous users for the product and now these reviews can be viewed by product wise. With the help of online review data sets, the user's opinions are getting collected and we can get the users feedback for the products. After applying the data reprocessing technique and the stop words are removed .Now the wordlist is going to be viewed comparing this word list with a bag of words containing Good opinions and Bad opinions after this opinion mining we can get the high pinioned products .Those products can be viewed in the browser.

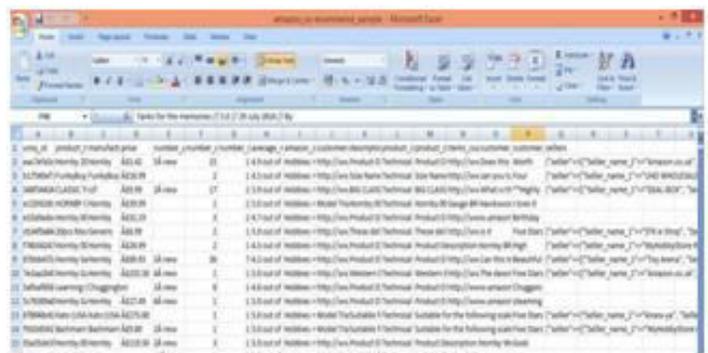


Figure.2. AMAZON Dataset.

• DATA SHEET PREPROCESSING

Ability of a computer to recognize handwritten character is a fascinating area of research due to the peculiarities involved in handwritten characters. Algorithm for Offline handwritten Character recognition differs as a result of diversities involved in writing with various language script. In a task of handwritten character recognition preprocessing and segmentation are two main phases and preliminary steps to be performed on acquired handwritten images. Achieving higher performance in handwritten character recognition depends on feature extraction process, which is highly influenced by preprocessing phase.

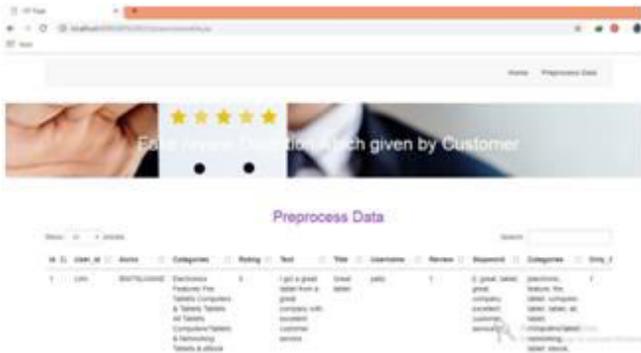


Figure.3. preprocessed data of dataset.

• **SUPPORT VECTOR MACHINE(SVM)**

SVM in machine learning is a supervised learning model with the related learning algorithm, which examines data and identifies patterns, which is used for regression and classification analysis. Recently, many classification algorithms have been proposed, but SVM is still one of the most widely and most popular used classifiers. Applying the kernel equations arranges the data instances in such a way within the multi-dimensional space, that there is a hyper-plane that separates data instances of one kind from those of another. The kernel equations may be any function that transforms the linearly non-separable data in one domain into another domain where the instances become linearly separable. Kernel equations may be linear, quadratic, Gaussian, or anything else that achieves this particular purpose. Once we manage to divide the data into two distinct categories, our aim is to get the best hyper-plane to separate the two types of instances.

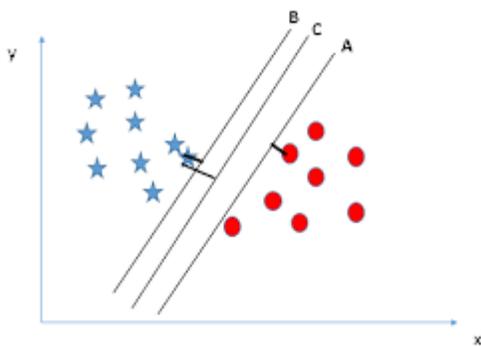


Figure4. photo of n-dimensional points in space.

The data instances that were not linearly separable in the original domain have become linearly separable in the new domain, due to the application of a function (kernel) that transforms the position of the data points from one domain to another. This is the basic idea behind Support Vector Machines and their kernel techniques. Whenever a new instance is encountered in the original domain, the same kernel function is applied to this instance too, and its position in the new domain is found out.

V. CONCLUSION

In this paper we presented the first study of fake reviews. We found that spotting the individual fake reviews was quite a difficult task but spotting the groups' was comparatively easier one. We propose the scoring algorithm which are used to

analyze the dataset and form the candidate groups using the process of SVM Classifier and we also proposed some behavioral features which are considered for finding fake reviews. Our analysis and experimental findings on real datasets provide valuable insight on the domain of fake reviews.

VI. FUTURE WORK

Finding the opinion fake from huge amount of unstructured data has become an important research problem. Now business organizations, specialists and academics are putting forward their efforts and ideas to find the best system for opinion fake review. Although, some of the algorithms have been used in fake review analysis gives good results, but still no algorithm can resolve all the challenges and difficulties faced by today's generation. More future work and knowledge is needed on further improving the performance of the fake review. There is a huge need in the industry, in day-to-day life for such applications because every company wants to know how consumers really feel about their products and services and those of their competitors by analyzing true reviews not fake reviews. This research proposes an opinion fake analyzer which automatically classifies input text data into either fake or non-fake category. The chosen algorithm based on simulation work is Support Vector Machine (SVM). A direction for future research is to implement the system and check performance by applying proposed approach to various benchmark data sets. Comparing performance of different classification methods to find the best one for our proposed opinion fake classification method could be another future research direction. However, there exist other kinds of review or reviewer related features that are likely to make a contribution to the prediction task. In the future we will do further investigate different kinds of features to make more accurate predictions.

VII. REFERENCES

[1]. ElshrifElmurngi, AbdelouahedGherbi, "Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques", École de Technologie Supérieure Montreal, Canada, DATA ANALYTICS 2017: The Sixth International Conference on Data Analytics.

[2].RAYMOND Y. K. LAU, S. Y. LIAO, RON CHI WAI KWOK, KAIQUAN XU, YUNQING XIA, YUEFENG LI TextMining and Probabilistic Language Modeling for Online Review Spam Detection ACM Trans. Manag. Inform. Syst. 2, 4, Article 25 (December 2011)

[3]. B. Jiang, R. Cao, B. Chen, "Detecting Product Review Spammer using Activity Model," Proceeding of International Conference on Advanced Computer Science and Electronics Information(ICACSEI 2013), pp. 650-653, 2013.

[4].J. Malbon, "Taking fake online consumer reviews seriously," Journal of Consumer Policy, vol. 36, no. 2, 2013, pp. 139–157.

[5]. B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in Proceedings of the 42nd annual meeting on Association for Computational Linguistics .Association for Computational Linguistics, 2004, p. 271. [Online]. Available from: <http://>

www.cs.cornell.edu/People/pabo/movie%2Dreview%2Ddata /
X. Xu, T. Han, Z. Xu, Y.

[6]. Wang, Y. Liu, "Design and Implementation of Spam Review Detection System", Proceedings of CLSW 2013, LNAI 8229, Berli, pp. 508–518, 2013.

[7]. S. Hassan, M. Rafi, and M. S. Shaikh, "Comparing svm and naive bayes classifiers for text categorization with wikitology as knowledge enrichment," in Multitopic Conference (INMIC), 2011 IEEE 14th International. IEEE, 2011, pp. 31–34. N. Jindal, B. Liu, E. Lim, "Finding Unusual Review Patterns Using Unexpected Rules," Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, Ontario, Canada, pp. 1549-1552, 2010.

[8]. C.-H. Chu, C.-A. Wang, Y.-C. Chang, Y.-W. Wu, Y.-L. Hsieh, and W.-L. Hsu, "Sentiment analysis on chinese movie review with distributed keyword vector representation," in Technologies and Applications of Artificial Intelligence (TAAI), 2016 Conference on. IEEE, 2016, pp. 84–89

[9]. Jeneen Interlandi (February 8, 2010). "The fake-food detectives". Newsweek. Archived from the original on October 21, 2010