Detection Of Fake Reviews Using Machine Learning Algorithm

Article *in* International Journal of Future Generation Communication and Networking · March 2020

CITATIONS
4

READS
4,335

5 authors, including:

Sanjay Laxmanrao Kurkute
Pravara Rural Engineering College Loni
6 PUBLICATIONS **10** CITATIONS

SEE PROFILE

International Journal of Future Generation Communication and Networking
Vol. 13, No. 1, (2020), pp. 415-419

Detection Of Fake Reviews Using Machine Learning Algorithm

Girish .J.Navale Dr. Sanjay Kurkute Amey.R.Kamble Kshitija Shinde

AISSMS IOIT,Pune

Abstract

Large part of the customers can give review after buying from which ever shop they want. Be it online buying or offline retail buying. When a customer buys a product online they check reviews of that product which are very important in today's e-commerce products for decision-making. As writing fake reviews comes with money gain, there has been a huge increase in misleading opinion on particular product reviews on their websites .Misleading review is an dangerous review. Giving positive reviews to target the customers that attract customers and gain increase in sales. There as negative review of a product could cause lesser demand of that product which may reduce its sales. These misleading reviews are dangerous to a product reputation. Here by making use of Machine Learning Algorithm. Such as SVM, Which is a Supervised Learning algorithm. We have predicted the review .i.e. a review is fake or not. Our objective is to decide whether a review is fake or genuine one.

Keywords: *Machine Learning, Support Vector Machines, Supervised Learning*

1. Introduction

Reviews are unit of measurements which are really very much important for success or failure in sales of a product. Reviews are being manipulated or created misleading for positive or negative impacts. In today's digital world dishonest opinion has become a dangerous problem to customers & E-commerce giants. Detection of these fake reviews is really an important and difficult task. These fake reviewers are paid to write these reviews. It is a difficult situation for a common customer to identify the dishonest reviews from genuine ones, by just reading each review .Because they are written in such a way that they look genuine. The customers are really dependent on this reviews to making their decisions to buy a products from this online seller or not.

2. Literature Survey:

1. Hyadri et al.[2] Their work mainly focuses on systematically analyzing and categorizing the models that mainly detect reviews spam. Their work proceeds to gain more in terms of accuracy and results.Different types of detection techniques have different types of their strengths and weaknesses.So with only systematically analyzing we cannot get more accurate results
2. Li et al.[6] The paper mostly proposes on three types of new features which include review density, semantic and emotion which gives the model paramaters to construct algorithm for each feature. But implementing all the new features would be difficut task as emotional review could be a mixed opnion review.
3. Hassan and Islam [9] In Their paper they have introduces some semi-supervised and supervised text mining models to detect fake online reviews as well they have gained efficiency of both techniques on dataset.But compared to text-mining SVM algorithm

ISSN: 2233-7857 IJFGCN

performs better.And when can train it for heavy datasets.It has highest accuracy compared to text-mining. .

Machine Learning technique used for training classifier

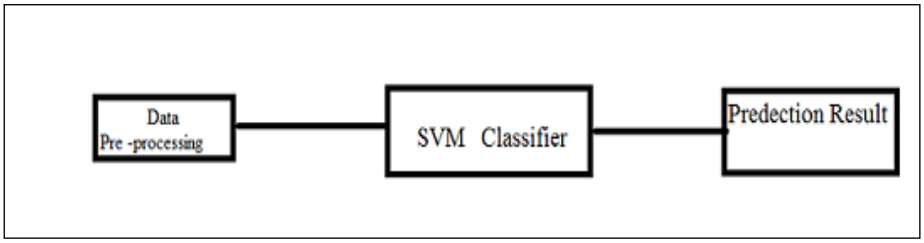


Figure 1: Working of Machine Learning Model

Here the three phases of Machine Learning Model for gaining prediction are described below.

Data Pre-processing: In this Section various actions are carried on data Some of the actions are like smoothing of data, Noise filtering. Removing the null values from the data etc.

SVM Classifier: SVM is defined as Support Vector Machine Classifier. SVM is a supervised Learning algorithm which can be used for categorization of the data. Support Vector Machine is fast and reliable algorithm which enhances the process for detection of fake reviews.

Prediction Of Result: By applying the SVM we predict the results of given data which is been classified for prediction.

Sample Collection: Collection of data for training the algorithm is known as sample collection.

Supervised Learning Algorithm:

Supervised Learning is carried out on labeled dataset.
Labeled Data:- Labeled data contains both the input as well as the output parameters which are required by the classifier to train itself.
After gaining the Labeled data we can train our classifier and then when can use our classifier on unlabelled data to obtain the proper analysis report.

Validating the data: The validation of data in machine learning is a confusing term .But it can be put in terms of obtaining an output in specific range. Or we can validate our result with certain amount of parameters. These parameters can validate our predicted result is right or wrong.

In Supervised Learning we require both training of data and validating the data.

Training Dataset: With the help of labeled data we provided training to our SVM Classifier. By giving the classifier an idea that this type of review is fake one or genuine one. With the help of this

ISSN: 2233-7857 IJFGCN

training the classifier can differentiate easily and between large datasets of reviews .After Learning more from labeled data the accuracy of classification is highly increased.

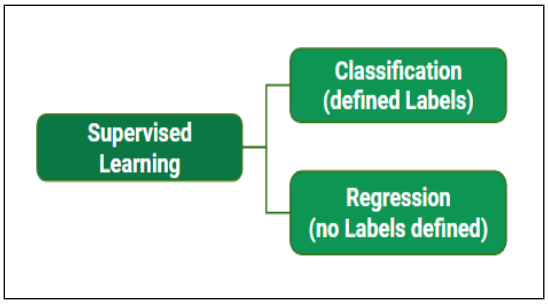


Figure 3 : Supervised Learning Model

SVM ALGORITHM TECHNIQUE

SVM Support Vector Machine is said to be supervised machine learning formula which is used for each classification as well as regression. It's basically used for classification related issues. While in this formula, we frequently try to plot every knowledge of object as a degree in n-dimensional area. . Here n is the range of options we have regarding every feature which is a selected coordinate. After this ,we perform classification by finding the hyper plane that differentiates between the two categories
:

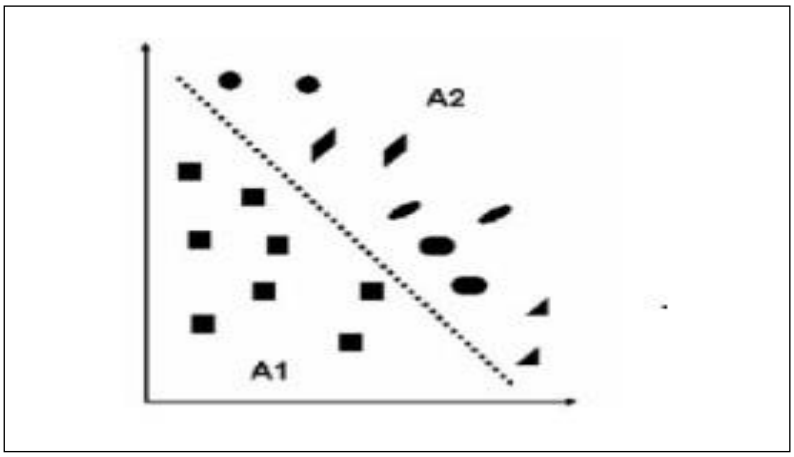


Figure 3 :Hyper plane used for categorization

The main motto for using the Support vector machines is that it is easy producer of hyper plane which separates the dataset into various categories. Which allow us to start with sample drawback. Consider for a given dataset, we have to classify red triangles from blue circles. We need goal form a line that classifies the information into two categories, making a difference between red triangles and blue circles. whereas one will get us a transparent line that separates the two categories, there are several lines which may try this job.
There's not a single one line that will agree on which might perform this task. The objective of SVM totally depends on linear separation in a very high dimension.

ISSN: 2233-7857 IJFGCN

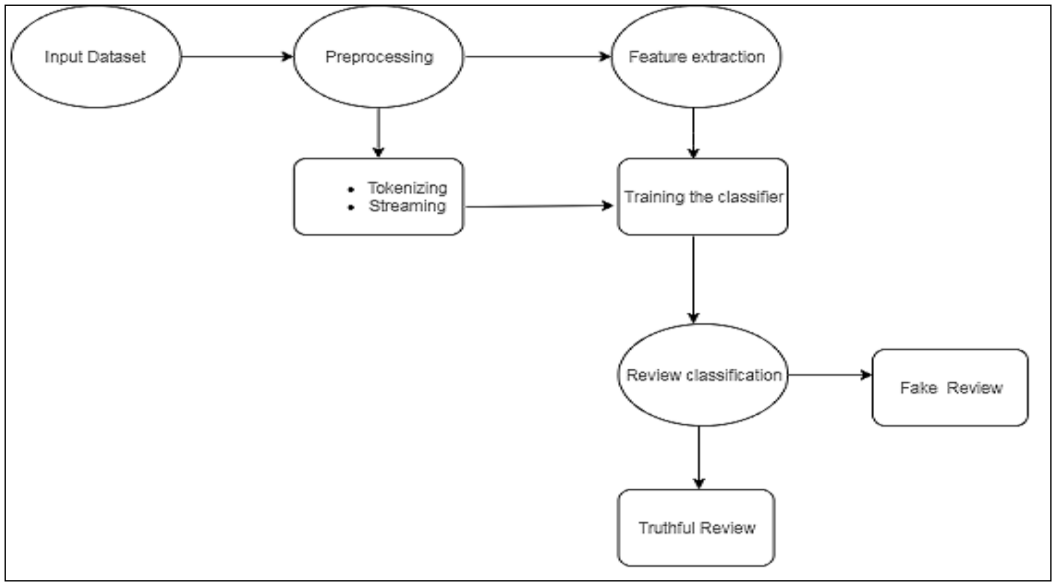
Figure 4 : Matching Score Vector plane formed after using SVM.

Feature area where knowledge square measure are mapped to think about the ultimate non-linearity of the matter. to get a small level of generalization capability.

The margin between hyper plane and the knowledge is maximized. The training of Support Vector Machine is carried with the matching scores of the vector. Hyper-plane is a plane which usually or may be a plane that linearly divides the n-dimensional knowledge points in two part.

If required for the second phase, hyper plane is line, just in case of 3D it's plane. It is combined known as n-dimensional line

Figure 5 : Design of our proposed system



Future work in Machine Learning Application with respect to Fake Review Detection

a. Currently we are considering only the static data and the open-databases which are available .Detecting ,the reviews in real time is future aspect in this project.
b. We have currently used only Support Vector Machine classifier ,in future aspect multiple algorithms can be implemented to check reviews.
c. While using multiple machine learning algorithms we can enhance the accuracy of detection of fake reviews.
d. Training a single classifier at a time. Is totally time consuming , in future aspect we can train multiple classifiers simultaneously. So that multiple classifiers can be trained and used of proper review detection.
e. It would be considered future aspect when we can load multiple datasets and perform review analysis on them.

ISSN: 2233-7857 IJFGCN

2. CONCLUSION

In this paper, we have created a model for fake review detection using Machine Learning algorithms such as SVM (Support Vector Machines). The model which we have created achieves its most elevated correctness. Fake Review detection is a developing research field .With limited amount of open datasets. With this project we are trying to get high accuracy and also reduce the time required to detect the Fake Reviews. Also we can use this project to detect the multiple fake Reviews.

.

References

[1] Chengai Sun, Qiaolin Du and Gang Tian, "Exploiting Product Related Review Features for Fake Review Detection," Mathematical Problems in Engineering, 2016.
[2] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, "Detection of review spam: a survey", Expert Systems with Applications, vol. 42, no.7, pp. 3634–3642, 2015.
[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in Proceedings of the 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.
[4] Prof Vaishali Babanne , Mr Girish Navale " Load Balancer Plug-in Filter"
[5] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2, 2012.
[6] Yuejun Li, Xiao Feng, Shuwu Zhang, Yuejun Li." Detecting Fake Reviews Utilizing Semantic and Emotion Model"
[7] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.
[8] J. K. Rout, A. Dalmia, and K.-K. R. Choo, "Revisiting semi-supervised learning for online deceptive review detection," IEEE Access, Vol. 5,pp. 1319–1327, 2017
[9] Rakibul Hassan, Md. Rabiul Islam. " Detection of fake online reviews using semi-supervised and supervised learning."

ISSN: 2233-7857 IJFGCN