ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

E-Commerce Fake Review Detection Using Data-Analyst

P.Premadevi[1] , R.Jansirani[2] , R.Preethi[3] , S.Sowntharya[4], J.Sumithra[5]

AP, Department of CSE, Angel College of Engineering and Technology, Tamil Nadu, India [1] ,

IV year, Department of CSE, Angel College of Engineering and Technology, Tamil Nadu, India [2,3,4,5]

ABSTRACT-Online consumer reviews have become a baseline for new consumers to tryout a business or a new product .online reviews are the most valuable sources of information about customer opinions and are consider the pillars of which the reputation of an organization is built .From a customer's perspective, review information is key to making a proper decision regarding an online purchase . Nowadays, review sites are more and more confronted with a spread of misinformation .This paper proposes, Naive bayes algorithm is used to evaluate the proposed feature set and compare it against the neural network algorithm in detecting fraud .The feature set considers the review on the platform to determine the user is committing fraud .The naive Bayesian algorithm helps in classifying the feature set with other feature set to detect fraud.

KEYWORDS: Fake review ,Naive bayes classifier, Machine learning technique

I. INTRODUCTION

As the technology of e-commerce continues to grow rapidly, the impact of reviews on online increase . For consumers, the reviews become an essential way for them to get more information of the product quality Helping them to make purchase decision. For business owners, the reviews given effort to improve and enhance their business. The feedback from the customer reviews is helpful for product improvements.However the reviews may not always betruthfully provided and fake reviews generally exist. The business owners may pay someone to write good reviews about their own product or bad reviews about their competitor's product. These Fake reviews lead the consumer making wrong judgments about the quality. Therefore, detecting fake reviews is important and is also a challenging issue in both industrial and academic. The document used in this work are obtained from a dataset of product reviews that have been collected from ecommerce platform.Then the document is classified as real positive and real negative reviews or fake negative and fake positive reviews. Fake negative and fake positive reviews by fraudsters who try to play their competitors existing system can lead to financial gains for them. This, unfortunately reviews gives strong incentives to write fake review that attempts to intentionally mislead readers by providing unfair reviews to several products for the purpose of damaging their reputation. Detecting such fake reviews is a significant challenge. For example, fake consumer reviews in a e-commercesector are not only affecting individual consumers but also corrupt purchaser's confidence in online shopping.It is hard to distinguish these two reviews whether they are fake or not for regular reviews. However, it is possible to detect and classify these reviews with machine learning techniques with the help of information other than the textual content itself.The only information exposed to internet uses is the review text and these nearly impossible for them to identify whether particular review is fake or not by just analyzing this raw text.Actually,textual content is not only the information that can be obtained from online reviews. In machine learning technique we are using supervised learning method. This technique performs data mining mission of concluding a function from labeled training data. The training data consist of set of training examples. In supervised learning, each example is a pair of consisting an input object(Typically a vector) and a desired output value(Also called the supervisory signal).

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

This paper describes a simple fake review detection method based on one of the machine learning algorithms-Naive Bayes classifier. The goal of the research is to examine how this particular method works for this particular problem given a manually labeled dataset and to support the idea of using machine learning algorithm for fake review detection. In this paper Naïve Bayes classifier specifically used for fake review detection. Also, the developed system was tested on a relatively on a new dataset, which gave an opportunity to evaluate its accuracy on a recent data by calculating its probability value.we can find the fake review effectively by using parameters like (1)Purchased customer(Either online or offline), (2)Review count(An user is allowed to post limited number of reviews), (3)Time duration(Time between product purchase and review posting).We select a supervised algorithm to evaluate and feature set as we are more focused on determining the characteristics which lead to fraud. We evaluate the features we selected by comparing the performance with other well- known feature sets, we include each of over features one after another and observe the changes in the performance metrics such as accuracy, precision, recall and F1 score with supervised learning, we focused on understanding the characteristics of the user which lead to review fraud.

II.RELATED WORK

A. Impact of reviewer social interaction on online consumer review fraud detection

Kunalgoswami,et,al[1]:This paper illustrated that some businesses or reviews uses these reviews to spread fake information about the information or product. Neural network algorithm is used to evaluate the proposed feature set and compare it against these state of the art features set in detecting fraud. The feature set considers uses social interaction and the yelp platform to determine if the user is committing fraud. The parameters used in this paper are content, time of posting,user. The F1 score obtained using neural network for detecting fraud value is 0.95.

B.Opinion spam detection in online reviews

Ajay rastogi,et.al[2]:This paper illustrated that spammers exploit these review platforms illegally because of incentive involved in writing fake reviews. This paper analysis and categorized as opinion spamming to three detection targets: opinion spam, opinion spammers, and collusive opinion spammer groups. Our survey introduces the challenges of automatic fake review detection using natural language processing. The parameters used in the papers are the product id, posting time and date, verified purchase tag.

C.Opinion spam detection using review and review centric features

Lijo V.P[3]: This paper illustrated that review centric features and reviewer centric features will be consider for feature selection. Review centric features include length of review, Time of posting and the reviewer centric features include percentage of positive reviews. This is implemented using hadoop architecture. HDFS is used to store the datasets. Analysis and model building will be doing with the help of open source software H2O.

D.Feature analysis for fake review detection through supervised classification

Macro viviani,et.al[4]: This paper illustrated that an article aims at providing an analysis of the main review and reviewer centric features that have been proposed of to now in the literature to detect fake reviews, in particular from those approaches that employee supervised machine learning techniques. This work intends and estimates some additional new features that can be suitable to categorize genuine and fake reviews. A supervised classifier based on random forest have been implemented, by considering both well known& new features, and a large scale labeled datasets.

E.A framework for review spam detection research

Mohammabalitavakoll[5]:This paper illustrated that spammers generate and post fake reviews in order to promote or demote brands and mislead potential customers. We have attempted to create a high quality framework to make a clear vision on review spam detection method**.**In addition this report contains a comprehensive collection of detection metrics used in proposed spam detection approaches. These metrics is extremely applicable for developing novel detection methods.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

F.Fake news detection using naive bayes classifier

Mykhailogranik,et.al[6]: This paper illustrated that fake news detection of Facebook comments.This paper describes simple fake news detection based on the one of the artificial intelligence algorithm-Naïve bayes classifier.This approach implemented as a software system and tested against a dataset of Facebook comments. Fake news detection developed system was tested on a relatively new dataset, which gave an opportunity to evaluate it is performance on a recent data.We achieved classification accuracy of approximately 74% of a testset which is a decent result considering the relative simplicity of a model

G.Semisupervised learning based fake review detection

Ningluo,et.al[7]: This paper illustrated that the existence of fake reviews makes the consumer cannot make their right judgment of sellers, which also causes the credibility of the platform downgraded. It is nearly impossible to make the correct annotation by reading only a small portion of comments based on the classifier trained under the traditional method. In this paper, the proposed a new algorithm to identify fake reviews based on semi supervised learning method. Real data based experiment have demonstrated that the proposed method can achieve desired performance.

H. Opinion mining andopinion spam detection

Pooja.H,et.al[8]: This paper illustrated that opinions are very important for consumer and for manufacturer too.Because based on these opinions any one can decide whether products are useful or not.This paper aims to represent a literature survey on opinion mining and opinion spam.Opinion mining has three important components: opinion holder,opinion object,opinion orientation. The parameters used in this paper are review length,rating,date of posting,review id,reviewer id.

I.Opinion spam detection:A review

Salma farooq,et.al [9]: This paper illustrated that opinion spam detection in the review context is to identify every fake review,fake reviewer,and fake reviewer group. In which supervised spam detection and unsupervised spam detection is used to find fake reviews.In supervised spam detection review centric feature,reviewer centric feature,and products centric feature are used to identify the fake reviews.In unsupervised spam detection the features observed in A) Author features are Content similarity, maximum number of reviews reviewing activity B) Review features are Extreme rating, Rating deviation ,Early time frame.

J.Sentiment analysis of amazon reviews using naive bayes on laptop products with mongoDBand R

Sanaprasanth,et.al [10]: In this paper used naive bayes classifier algorithm and semantic decision tree to classify the polarization of explanations given on e-commerce websites.They used a web crawlerto fetch a comment on a particular web page.The spelling correction is done to make the most sensible comment for knowing the polarity of words using world net dictionary. And then stemming is performed to remove the stop words. After classifying the positive and negative words using naive bayes classifier, the overall polarity is calculated using decision tree.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

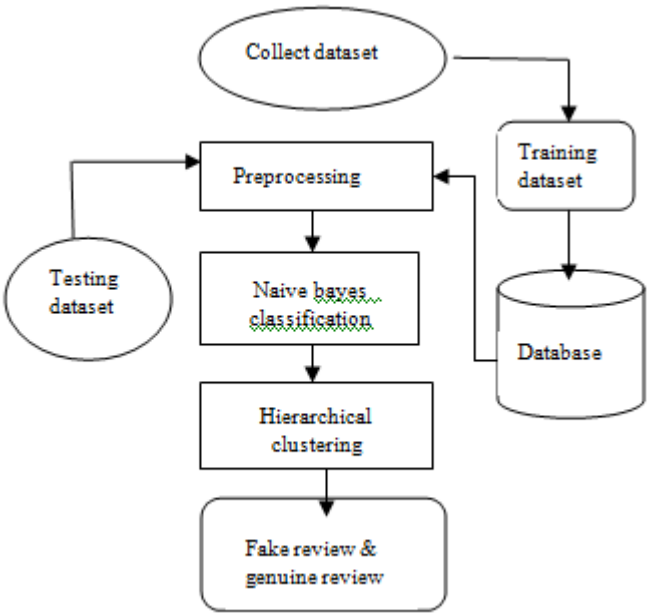
(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

III.PROPOSED MODEL

A.SYSTEM ARCHITECTURE



B.Description

The dataset is collected from a particular platform.We have to separate data into a training set and testing set most of the data is used for training, and a smaller portion of data is used for testing. The training and test data are stored the database. In preprocessing the extracted data's are preprocessed by removing stop words, short forms and emotions. All short forms will be replaced with full words so that it is understandable for all the users. Naïve bayes is algorithm that analyzes data and recognizes patterns, used for classification and regression analysis. Naïve bayes algorithm is based on the concept of review that defines decision boundaries. A review is one that separates between a set of objects having different class memberships. A schematic example: fake and original. After the preprocessing the data's are classified into good review and bad review. The words are identified based on the keywords to classify the reviews.After classification process over than going to do the clustering process. In this process we are going to group the reviews. The original and fake reviews are grouped and separated each other. After that clustering process over then we will get the result for the given input data text file.

C.Algorithm

Naive bayes algorithm

Naïve bayes(NB) is used as an classifier in various real world problems. It is based on the bayes theorem of probability. It is mainly used to predict the class of unknown dataset. The assumption is that the occurrence of a particular feature in a class is isolated to the existence of any other feature. All the properties individually subsidize of to the probability. Bayes theorem:

$P(A/B)=[P(B/A)]/P(B)$,

Where P(A) : Prior probability of hypothesis A

P(B) : Prior probability of training data B

P(A/B) : Probability of A given B(also known as the Posterior Probability of the class A(target) when B is the given predictor(attribute))

P(B/A) : Probability of B given A.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

D. Parameters used

a.Review count

The purchaser can give maximum three reviews for a product. The reviews can be either positive or negative. If the customer gives more than three reviews then the review will be ignored.

b.Time duration

The time between the product purchasedand the time of posting the review.

c.Purchased customer

Only purchased customer is allowed to provide the review .The customer will be either online or offline customer.

IV.IMPLEMENTATION

a.Review retrieval and preprocessing

Data are collected from a particular platform(yelp,amazon,etc..)The extract the data are preprocessed by removal of stop words,emotion stickers,short forms. In this process we are going to remove the verbal related words like is, was, etc…Those words don't give any useful information about the reviews. Separating dataset into training and testing datasets. Most of the data is used for training and a smaller portion of data is used for testing.

b.Probability prediction using naive bayes

Naive bayes algorithm analyzes data and recognizes patterns, user for classificationand recursive algorithms. Naive bayes algorithm is based on the concept of review that defines decision boundaries. For probability calculation, convert dataset table into frequency table. Create likelihood table by finding probability. Use naive bayes equation to calculate the posterior probability. Based on probability values it predicts the result. A confusion matrix is used to calculate accuracy.

c.Hierarchical clustering

After the classification process over then going to do the clustering process.In this process we are going to group the reviews.The original and fake reviews are grouped and separated each other. After that clustering process over then will be get the result for the given input data text file.

d. Fake review detection

Gather the genuine review from clustering process. Check the parameters: Purchased customer(check whether the customer is online or offline customers),Review count(Allow to user to provide limited number of reviews),Time duration(The time between product purchased and time of posting the reviews).based on these parameters that identifies the fake reviews.

V.CONCLUSION

Instead of some thousand goods in a super supply,customers may choose among millions of products in an online store to satisfy the personalization difficulties.In this paper, we use naive bayes algorithm to classify the polarity of comments given on e-commerce websites. First, we use a web crawler to fetch reviews on a particular webpage. Then preprocessing is performed to remove stop words. After categorizing the positive and negative words using naive bayes algorithm, the overall division is calculated. The research showed,that event quite simple machine learning algorithm such as naive bayes classifier may show a good result on such an important problem as fake review detection. Therefore, the results of this research suggest even more, that machine learning algorithm, may be successfully used to tackle this important problem.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

VI. FUTURE WORK

A possible direction for future work is to use text classification algorithms, which has to provide a higher accuracy than naïve bayes algorithm. It possible to extend this study to apply the same architecture for different bench marks dataset using different prediction models

REFERENCES

[1]Kunalgowsami,Youngheepark,Chungsik song"Impact of reviewer social interaction on online consumer review fraud detection",journal of bigdata 2017.
[2]Ajay rastogi,Monicamehrotra,"Opinion spam detection in online reviews",journal of information and knowledge management 2017.
[3]LijoV.P,Dilshadominic"Opinion spam detection using review and review centric features",IEEE international conference on power,control,signaland instrumentation engineering 2017.
[4]Macro viviani,Julienfontanaravi, Gabriella pasi"Feature analysis for fake review detection through supervised classification",2017 international conference on data science and advanced analysis.
[5]Mohammabalitavakoll, Atefehheytary,Zuriatiismail,Naomiesalim,"A framework for review spam detection research",International journal of computer and information technology 2016.
[6]MykhailoGranik, VolodymyrMesyura"Fake news detection using naive bayes classifier",2017 IEEE first Ukraine conference on electrical and communication engineering.
[7]Ningluo,uaxundeng,lingfengzhao,yuanliu,Zingweiwang"Semi supervised learning based fake review detection",2017 IEEE international conference on ubiquitous computing and communication
[8]Pooja.H,Tosil M.bhalobia,"Opinion mining and opinion spam detection" International research journal of engineering and technology 2017.
[9]Salma farooq,Hillalahmadahandai,"Opinion spam detection: A review",International journal of enginnering research and development.
[10]Sanaprasanthshakthi,Mohan Kamal Hassan and R Sasikala,"Sentiment analysis of amazon reviews using naive bayes on laptop products with mongoDB and R",IOP conference series metrials science and engineering.