INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

Implementation towards Fake Reviews Detection for Online Product using Machine Learning

Vaidehi Patil, Sakshi Baldwa, Rutuja Shinde, Saurabh Jagdale, Dr. A. V. Markad

Department of Information Technology, AVCOE, Ahmednagar, Maharashtra, India

Department of Information Technology, AVCOE, Ahmednagar, Maharashtra, India

Department of Information Technology, AVCOE, Ahmednagar, Maharashtra, India

Department of Information Technology, AVCOE, Ahmednagar, Maharashtra, India

Department of Information Technology, AVCOE, Ahmednagar, Maharashtra, India

ABSTRACT: The development of e-commerce platforms has given people a new way to generate and consume a great deal of infor- mation on the web. The arrival of online e-commerce platforms, people need not run to out for the basic need of regular stuff because of their fast and efficient features. As now a day's people are counting on online products therefore the importance of a review goes higher. For selecting a product, a customer undergo thousands of reviews to know a product, Thus Product review helps for marketing products online. Online users rely on reviews before making decisions about any product and service. As such, the credibility of online reviews is crucial for businesses and can directly affect companies' reputation and profitability. That is why some businesses are paying spammers to post fake reviews. These fake reviews exploit consumer purchasing decisions. The proposed system helps to detect the fake online reviews and remove it thus helping customer to review the product according to original review and ratings. In this work, reviews published are identified by searching for the particular keyword and then the polarity of review is evaluated as positive and negative. The work also identifies the Ip address of particular device , if more than three reviews are posted from same Ip address, the user is blocked. The reviews are evaluated based on feature selection of each score words. In order to select the best features Naive Bayes Classifier (NBC) is used for training and testing the features of a words and also evaluating the polarity of each review. Performance evaluation parameters such as accuracy, precision and time is taken into consideration and compared with three machine learning classifiers, namely, Random Forest, Naive Bayes and Support Vector Machine(SVM) are calculated to determine fake review.

KEYWORDS: Machine Learning, SVM, Naive Bayes, Random Forest

I. INTRODUCTION

Technologies are changing rapidly. Old technologies are continuously being replaced by new and sophisticated ones. These new technologies are enabling people to have their work done efficiently. Such an evolution of technology is online marketplace. We can shop and make reservation using online websites. Almost, everyone of us checks out reviews before purchasing some products or services. Hence, online reviews have become a great source of reputation for the companies. Also, they have large impact on advertisement and promotion of products and services. With the spread of online marketplace, fake online reviews are becoming great matter of concern. People can make false reviews for promotion of their own products that harms the actual users. Also, competitive companies can try to damage each others reputation by providing fake negative reviews. Researchers have been studying about many approaches for detection of these fake online reviews. Some approaches are review content based and some are based on behavior of the user who is posting reviews. Content based study focuses on what is written on the review that is the text of the review where user behavior based method focuses on country, ip-address, number of posts of the reviewer etc. Most of the proposed approaches are supervised classification models. Few researchers, also have worked with semi-supervised models. Semi-supervised methods are being introduced for lack of reliable labeling of the reviews.

Amazon is the most popular microblogging platform in the world. It is also the fastest growing social network platform and has a dominant position in the area of mircroblogging. More than 500 million registered users post 340 million Amazon messages every day, sharing their opinions and daily activities. Compared with regular microblogging platforms, Amazon messages are much shorter. You are only allowed to post 140 characters or less in one Amazon message. With all of the advantages mentioned above, Amazon thus has become a powerful platform with many kinds

of information from worldwide breaking news to purchasing products at home.

Paper is organized as follows. Section II describes about the related work done earlier for the system to be developed. Section III presents method used and algorithms used for the detection. Section IV presents experimental results showing results of images tested. Finally, Section V presents conclusion.

II. RELATED WORK

A. Related To Fake Reviews

1) Title : "Detection of fake online reviews using semi- supervised and supervised learning"
Author: R. Hassan and M. R. Islam, Year: 2019 IEEE Authors have shown several semi-supervised and supervised text mining techniques for detecting fake online reviews in this research. They have combined features from several research works to create a better feature set.

2) Title :"Towards understanding and detecting fake reviews in app stores"
Author: Daniel Martens · Walid Maalej Year: 2019 Springer
Authors analyzed the market of fake review providers and their fake reviewing strategies and found that developers buy reviews to relatively expensive prices of a few dollars or deal with reviews in exchange portals

3) :"Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Learning"
Author: Naveen Kumar, Deepak Venugopal, Liangfei Qiu Subodha Kumar
Year: 2018 Taylor Francis
Authors propose a novel hierarchical supervised- learning approach to increase the likelihood of detecting anomalies by analyzing several user features and then characterizing their collective behavior in a unified manner.

4) "Semi-Supervised Learning Based Fake Review Detec- tion,"
Author: H. Deng et al., Year: 2017 IEEE
Authors consider both the metadata features and content related features to construct a semi-supervised learn- ing based fake review classifier. Firstly, they use the similarity characteristics of the text to determine a set of true negative cases or fake reviews and extract the characteristic vector from multiple aspects.

B. Related to Classifier

1) A pattern based approach for multiclass sentiment analysis in twitter
Authors have developed SENTA architecture, which concentrates on classifying of text into more classes. SENTA aims to achieve multi-class classification by providing an easy to-use graphical interface and also achieves good accuracy on classification with multi-class classification

2) Understand short texts by harvesting and analyzing semantic knowledge
This paper discussed about understanding short text, which is quite challenging and crucial. Short texts are in large quantity and are ambiguous and noisy. Knowledge intensive approaches concentrate on semantics in the tasks of text segmentation, part-of-speech tagging and concept labeling Semantic knowledge is indispensable for short text understanding and knowledge intensive approaches are efficient and effective in discovering semantics of short texts

3) Universal, unsupervis ed (rule based), uncovered sentiment analysis
This paper addresses a unsupervised approach for mul- tilingual sentiment analysis including a compositional syntax based rules

4) Sentiment analysis of the conflict of ages series Authors have discussed why it is important to use text mining tool to provide the distinct perspective on the conflict of ages series.They have also concluded about the characteristics of the writer

5) Semeval 2017 task 5: Fine grained sentiment analysis on financial microblogs and news,"
Paper describes the system submitted to the FineGrained Sentiment Analysis on financial Microblogs and News tasks. They also cite the need to explore domain specific sentiment lexicons and use the deep learning method to improve the

performance.

6)   Sentiment analysis algorithms and application s: A survey
This paper explains uses of Sentiment analysis or opin- ion mining i.e. to study people's emotions, attitudes, sentiments, evaluations, moods, and opinions

7)   Affective computing and sentiment analysis
This paper stated that the Sentiment analysis is the stem of natural language processing and machine learning methods, which is the current trending research area in the text mining. It is the important source of decision making and it can be extracted, identified, evaluated from the online sentimental reviews

8)   **Semeval 2015 task 10: Sentiment analysis in twitter** Paper provides the information about the Sentiment analysis using twitter as The content of user generated opinions in the social media such as facebook, twitter, review sites, etc are . These opinions can be tapped and used as business intelligence for various uses such as marketing, prediction, etc.

9)   An ensemble sentiment classificati on system of twitter data for airline services analysis
The system described in this paper uses a twitter data and performs the sentiment classification for airline services analysis

10)   Compariso n research on text methods on preprocessi ng twitter sentiment analysis
This paper explains the various methods of preprocess- ing of twitter data for sentiment analysis and there comparison

III. PROPOSED METHODOLOGY

A.   Methodology

1)   Review Acquisition : The review is captured. 2)Preprocessing : The review is tokenized and cleaned.
3)Processing:

- Feature Extraction : Feature for fake reviews is extracted.
- Classification : The Model classifies reviews into different category .

4)Polarity : reviews are recognized.

B.   System Architecture

validation data include labels to monitor performance metrics of the model, the testing data should be unla- beled. Test data provides a final, real-world check of an unseen dataset to confirm that the ML algorithm was trained effectively.
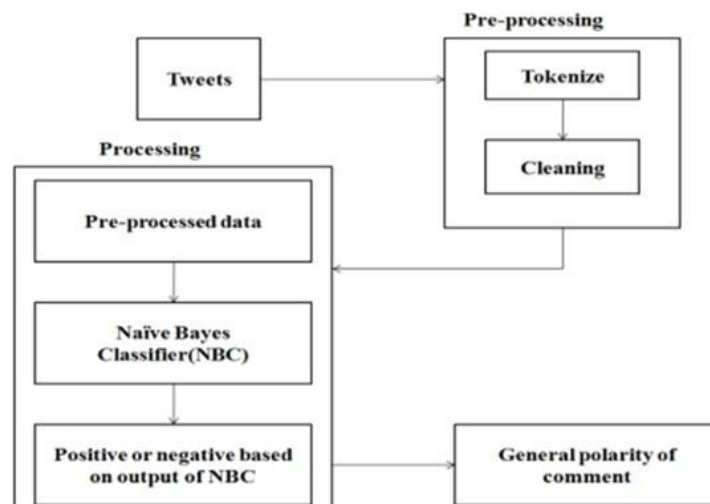
D. Algorithm

1)   SVM

Fig. 1. System Architecture

C. Modules

Input: D dataset, on-demand features, aggregation-based features,

Output: Classification of Application

for each application App-id in D do

Get on-demand features and stored on vector x for App-id

x.add (Get-Features(app-id));

end for

for each application in x vector do

Fetch first feature and stored in b, and other features in w.

hw,b (x) = g (z) here z= (wT x + b)

if (z 0)

assign g(z)=1;

else g(z)=-1;

end if

end for

Fig. 2. SVM

1) Data Preprocessing

A technique which is used to transform the raw data in a useful and efficient format. In System ,reviews are the Input Data . For prepreprocessing , Cleaning and tokenization id done.

2) Data Classification

Classification is Data mining task of the pre- dicting value of Categorical variable. This is done by building model depending on one or more attributes or features.

3) Training and Testing

Different datasets serve different purposes in preparing an algorithm to make predictions and decisions based on real-world data.

Training data.

This type of data builds up the machine learning algorithm. The data scientist feeds the algorithm input data, which corresponds to an expected output. The model evaluates the data repeatedly to learn more about the data's behavior and then adjusts itself to serve its intended purpose.

Test data

After the model is built, testing data once again validates that it can make accurate predictions. If training and Naive Bayes
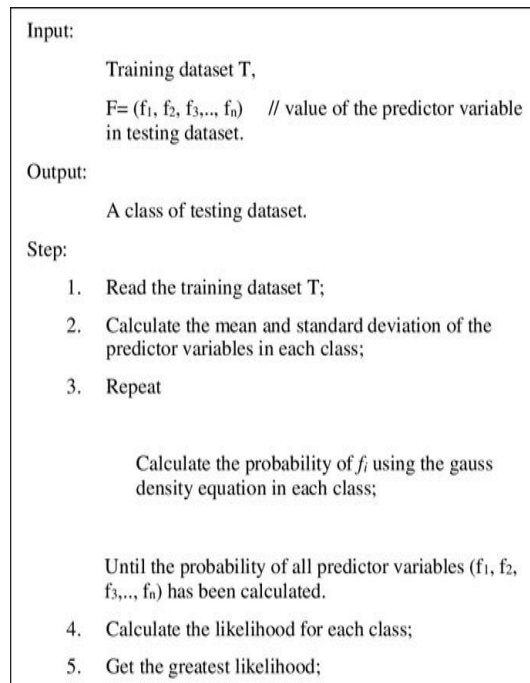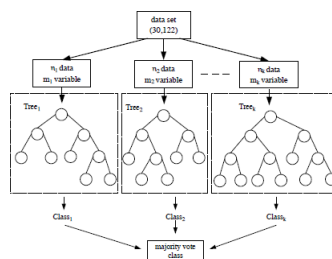
Input:

> Training dataset T,
>
> F= (f₁, f₂, f₃,.., fₙ)   // value of the predictor variable in testing dataset.

Output:

> A class of testing dataset.

Step:

1. Read the training dataset T;

2. Calculate the mean and standard deviation of the predictor variables in each class;

3. Repeat

> Calculate the probability of f_i using the gauss density equation in each class;
>
> Until the probability of all predictor variables (f₁, f₂, f₃,.., fₙ) has been calculated.

4. Calculate the likelihood for each class;

5. Get the greatest likelihood;

Fig. 3. Naive Bayes

2)   Random Forest

Fig. 4. Random Forest
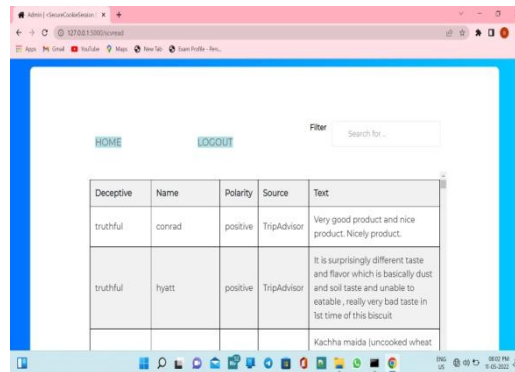
IV. RESULT & DISCUSSIONS
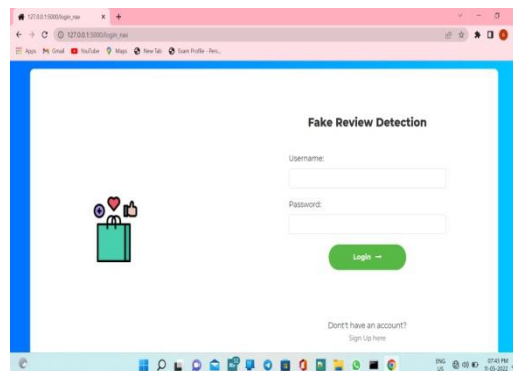
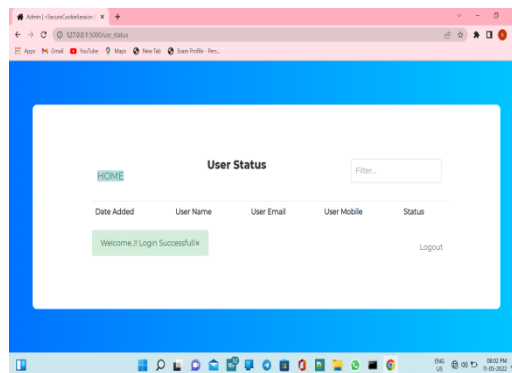
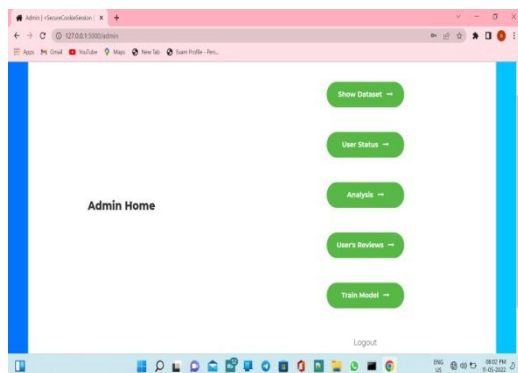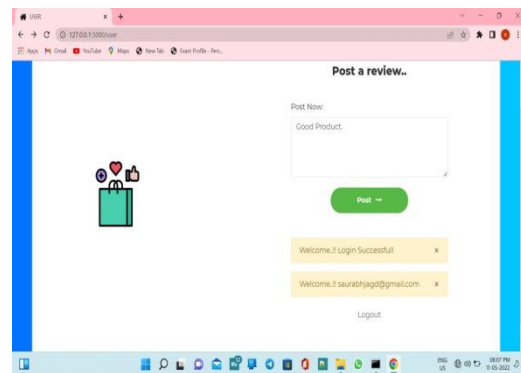
Fig. 5. Result1

Fig. 7. Result3

Fig. 8. Result4

Fig. 6. Result2

Fig. 9. Result5

V. CONCLUSION

In this work, reviews from product will be extracted using a particular string search. These reviews will be subjected to analysis using RF, SVM and NB classifiers In order to classify them into fake review or Genuine Review. Identification of Ip address of particular device , if more than three reviews are posted from same Ip address, the user is blocked.

REFERENCES

[1]R. Hassan and M. R. Islam, "Detection of fake online reviews using semi-supervised and supervised learning," 2019 International Confer- ence on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1-5, doi: 10.1109/ECACE.2019.8679186.

[2] Martens, D., Maalej, W. Towards understanding and detecting fake reviews in app stores. Empir Software Eng 24, 3316–3355 (2019).

[3] Naveen Kumar, Deepak Venugopal, Liangfei Qiu & Subodha Kumar (2018) Detecting Review Manipulation on Online Platforms with Hi- erarchical Supervised Learning, Journal of Management Information Systems, 35:1, 350-380, DOI: 10.1080/07421222.2018.1440758

[4]H. Deng et al., "Semi-Supervised Learning Based Fake Review Detec- tion," 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), 2017, pp. 1278-1280, doi: 10.1109/ISPA/IUCC.2017.00195.

[5]W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093–1113, 2014.

[6]E. Cambria, "Affective computing and sentiment analysis," IEEE Intel- ligent Systems, vol. 31, no. 2, pp. 102–107, 2016.

[7]S. Rosenthal, P. Nakov, S. Kiritchenko, S. Mohammad, A. Ritter, and V. Stoyanov, "Semeval-2015 task 10: Sentiment analysis in twitter," in Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015), pp. 451–463, 2015.

[8]Y. Wan and Q. Gao, "An ensemble sentiment classification system of twitter data for airline services analysis," in Data Mining Workshop (ICDMW), 2015 IEEE International Conference on, pp. 1318–1325, IEEE, 2015.

[9]Z. Jianqiang and G. Xiaolin, "Comparison research on text preprocessing methods on twitter sentiment analysis," IEEE Access, vol. 5, pp. 2870–2879, 2017.

[10]S. Tokle, S. R. Bellipady, R. Ranjan, and S. Varma, "Energy-efficient wireless sensor networks using learning techniques," Case Studies in Intelligent Computing: Achievements and Trends, pp. 407–426, 2014.

[11]M. Bouazizi and T. Ohtsuki, "A pattern-based approach for multiclass sentiment analysis in twitter," IEEE Access, vol. 5, pp. 20617–20639, 2017.

[12]W. Hua, Z. Wang, H. Wang, K. Zheng, and X. Zhou, "Understand short texts by harvesting and analyzing semantic knowledge," IEEE transactions on Knowledge and data Engineering, vol. 29, no. 3, pp. 499–512, 2017.

[13]D. Vilares, C. Go ´mez-Rodr´ıguez, and M. A. Alonso, "Universal, unsupervised (rule-based), uncovered sentiment analysis," Knowledge- Based Systems, vol. 118, pp. 45–55, 2017.

[14]D. Thomas, "Sentiment analysis of the conflict of ages series," in International Forum, vol. 20, pp. 31–43, 2017.

[15]K. Cortis, A. Freitas, T. Daudert, M. Huerlimann, M. Zarrouk, S. Hand- schuh, and B. Davis, "Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news," in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pp. 519–535, 2017.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 ⬥ 6381 907 438 ⬥ ijircce@gmail.com

Scan to save the contact details