Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN 1989 –9572

DOI: 10.47750/jett.2023.14.02.058

FAKE REVIEW DETECTION USING SUPERVISED MACHINE LEARNING

[1]Dr.P.Veeresh
[2]Nagaladinne Namira Jahan
[3]Boggula Mamatha
[4]Bommasetty Mahalakshmi
[5]Mallampalli Kongathi Poornachandrika

Journal for Educators,Teachers and Trainers, Vol.14 (2)

https://jett.labosfor.com/

Date of reception: 18 Jan 2023

Date of revision: 09 Feb 2023

Date of acceptance: 21 Mar 2023

Dr.P.Veeresh, Nagaladinne Namira Jahan,Boggula Mamatha, Bommasetty Mahalakshmi, Mallampalli Kongathi Poornachandrika (2023). FAKE REVIEW DETECTION USING SUPERVISED MACHINE LEARNING.Journal for Educators, Teachers and Trainers,Vol. 14(2). 630-633

Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers, Vol. 14(2)
ISSN 1989 –9572
https://jett.labosfor.com/

FAKE REVIEW DETECTION USING SUPERVISED MACHINE LEARNING

[1]Dr.P.Veeresh ,[2]Nagaladinne Namira Jahan ,[3]Boggula Mamatha,[4]Bommasetty Mahalakshmi,[5]Mallampalli Kongathi Poornachandrika
[1]Professor &HOD,[2]Students
Department Of CSE
St. Johns College of Engineering & Technology, Errakota, Yemmiganur

ABSTRACT
Online evaluations have a significant influence on contemporary business and trade. The buying of internet items is mostly influenced by user reviews. Therefore, people or organizations with opportunistic tendencies attempt to alter product evaluations for their own gain. This research presents a range of semi-supervised and supervised text mining models for identifying fraudulent online reviews. It also evaluates the effectiveness of both approaches using a dataset of hotel reviews.

I. INTRODUCTION

The pace of technological advancements is accelerating. Outdated technology are constantly being replaced by novel and advanced ones. These emerging technologies are facilitating individuals to do their tasks with high efficiency. An online marketplace represents a significant advancement in technology. Online websites provide the convenience of shopping and making reservations. Practically all of us examine evaluations prior to buying items or services. Therefore, internet evaluations have emerged as a valuable means of assessing the reputation of firms. Additionally, they have a significant influence on the advertising and marketing of goods and services. The proliferation of online marketplaces has led to a significant increase in the prevalence of fraudulent online reviews, which is a major cause for worry. Individuals have the ability to provide deceptive evaluations in order to promote their own items, which might have detrimental effects on genuine consumers. Furthermore, rival firms may attempt to undermine one another's image by disseminating fabricated unfavorable evaluations.

Researchers have been investigating several methods for detecting fraudulent internet reviews. There are two types of techniques: content-based and user behavior-based. Content-based approaches focus on analyzing the substance of the reviews, while user behavior-based approaches analyze the behavior of the person who is submitting the reviews. Content-based analysis focuses on the textual content of a review, whereas user behavior-based methods include factors such as the reviewer's nationality, IP address, and number of postings. The majority of the suggested methodologies consist of supervised classification models. Only a few number of researchers have also used semi-supervised models. Semi-supervised techniques are being used due to the absence of dependable labeling for the reviews.

This study presents several classification methods for identifying fraudulent online reviews, including both semi-supervised and supervised techniques. Semi-supervised learning employs the Expectation-maximization technique. We use the Statistical Naive Bayes classifier and Support Vector Machines (SVM) as classifiers in our study to enhance the classification performance. Our primary emphasis has been on the substance of the review-based techniques. The features used include word frequency count, emotion polarity, and review length.

II. EXISTINGSYSTEM

- Content based methods focus on what is the content of the review. That is the text of the review or what is told in it. Heydari et al. [2] have attempted to detect spam review by analyzing the linguistic features of the review. Ott et al. [3] used three techniques to perform classification. These three techniques are- genre identification, detection of psycholinguistic deception and text categorization.
- Behavior feature based study focuses on the reviewer that includes characteristics of the person who is giving the review. Lim et al. [7] addressed the problem of review spammer detection, or finding users who are the source of spam reviews. People who post intentional fake reviews have significantly different behavior than the normal user. They have identified the following deceptive rating and review behaviors.
- Deceptive online review detection is generally considered as a classification problem and one popular approach is to use supervised text classification techniques [5]. These techniques are robust if the training is performed using large datasets of labeled instances from both classes, deceptive opinions (positive instances) and truthful opinions (negative examples) [8]. Some researchers also used semi-supervised classification techniques.

Disadvantages

- In the existing work, the system uses only to semi-supervised learning.
- Only Text Classification as sentiment text and it never finds fake review.

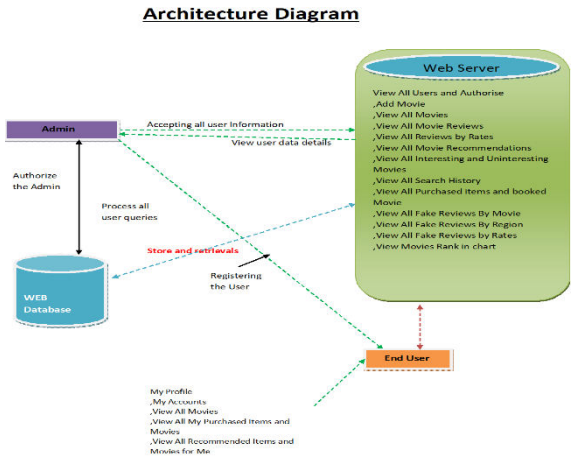
III. PROPOSED SYSTEM

- In the proposed system, each review goes through tokenization process first. Then, unnecessary words are removed and candidate feature words are generated.
- Each candidate feature words are checked against the dictionary and if its entry is available in the dictionary then its frequency is counted and added to the column in the feature vector that corresponds the numeric map of the word.
  Alongside with counting frequency, the length of the review is measured and added to the feature vector.
- Finally, sentiment score which is available in the data set is added in the feature vector. We have assigned negative sentiment as zero valued and positive sentiment as some positive valued in the feature vector.

Advantages

- The system is very fast and effective due to semi-supervised and supervised learning.
- Focused on the content of the review based approaches. As feature we have used word frequency count, sentiment polarity and length of review.

IV. SYSTEM ARCHIECTURE:



V. IMPLEMENTATION

- **Admin**

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as View All Users and Authorise,Add Movie, View All Movies, View All Movie Reviews, View All Reviews by Rates, View All Movie Recommendations, View All Interesting and Uninteresting Movies, View All Search History, View All Purchased items and booked Movie ,View All Fake Reviews By Movie, View All Fake Reviews By Region, View All Fake Reviews by Rates ,View Movies Rank in chart

User

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations like View My Profile, My Accounts, View All Movies, View All My Purchased Items and Movies, View All Recommended Items and Movies for Me.

VI. CONCLUSION

In this study, we have shown several semi-supervised and supervised text mining methods for identifying fraudulent online reviews. We have amalgamated characteristics from many research endeavors to construct an enhanced collection of traits. Additionally, we have experimented with other classifiers that were not used in the prior study. Consequently, we have enhanced the precision of prior semi-supervised methods conducted by Jiten et al. [8]. Additionally, our research has shown that the supervised Naive Bayes classifier yields the maximum level of accuracy. This guarantees that our dataset is accurately labeled, since we are aware that semi-supervised models perform well in situations when trustworthy labeling is unavailable.

Our study has focused only on user reviews. In the future, the integration of user actions and words may be used to create a more advanced categorization model. Utilizing advanced preprocessing technologies for tokenization enhances the precision of the dataset. The suggested approach may be evaluated for a bigger data set to determine its efficacy. This study is only focused on English reviews. It is feasible to do this for Bangla as well as several other languages.

REFERENCES

[1] Chengai Sun, Qiaolin Du and Gang Tian, “Exploiting Product RelatedReview Features for Fake Review Detection,” Mathematical Problems inEngineering, 2016.
[2] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, ”Detection ofreview 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 deceptiveopinion spam by any stretch of the imagination,” in Proceedings ofthe 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] J. W. Pennebaker, M. E. Francis, and R. J. Booth, ”Linguistic Inquiryand Word Count: Liwc,” vol. 71, 2001.
[5] S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deceptiondetection,” in Proceedings of the 50th Annual Meeting of the Associationfor Computational Linguistics: Short Papers, Vol. 2, 2012.
[6] J. Li, M. Ott, C. Cardie, and E. Hovy, “Towards a general rule foridentifying deceptive opinion spam,” in Proceedings of the 52nd AnnualMeeting of the Association for Computational Linguistics (ACL), 2014.
[7] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, “Detectingproduct review spammers using rating behaviors,” in Proceedings ofthe 19th ACM International Conference on Information and KnowledgeManagement (CIKM), 2010.
[8] J. K. Rout, A. Dalmia, and K.-K. R. Choo, “Revisiting semi-supervisedlearning for online deceptive review detection,” IEEE Access, Vol. 5,pp. 1319–1327, 2017.
[9] J. Karimpour, A. A. Noroozi, and S. Alizadeh, “Web spam detection bylearning from small labeled samples,” International Journal of ComputerApplications, vol. 50, no. 21, pp. 1–5, July 2012.