Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

# DETECTION OF FAKE ONLINE REVIEWS

**Dr. R.KAVITHA** Associate professor, Department of Computer Science and Engineering, Siddhartha Institute of Engineering & Technology,VinobhaNagar,Ibrahimpatnam, RR– 501 506, Telangana, India. Email:naaccse4@gmail.com

**G. UDAY KUMAR** Assistant Professor, Department of Computer Science and Engineering, Siddhartha Institute of Engineering & Technology, Vinobha Nagar, Ibrahimpatnam, RR– 501 506, Telangana, India. Email: uday518@siddhartha.ac.in

**N.PRAVALIKA** M.Tech Student, Department of Computer Science and Engineering, Siddhartha Institute of Engineering & Technology, Vinobha Nagar, Ibrahimpatnam, RR– 501 506, Telangana, India. Email: pravalikanaddunuri39@gmail.com

**Abstract:** Nowadays, when somebody wants to make some decisions about a product or a service everyone goes with the reviews as it has become an essential part of decision making. When a customer wants to order a product on an e commerce website firstly everyone checks the review section in detail and further proceeds for decision making about the product. If the reviews posted were satisfactory for the customer he may order the product thus reviews become a reputed parameter for the businesses and companies and also a great source of information for the customers. Every customer thinks that the reviews he/she is seeing is authentic and any manipulation in that from any individuals or any rival companies which may lead to fake data which will be labeled as fake reviews. This type of attempt if not noticed may let us think about the gen-unity of the data. So these reviews are the most important parameter for the businesses and companies. There exist some groups or persons who make use of these reviews to forge customers for their own interest or damage their competitors reputation. In order to solve this problem we uses Machine learning techniques(Supervised and semi-supervised) to detect whether the given review is fake or not with high accuracy. Along with this objective we also focus on developing models which need less data to train. Since we can't always be able to get labeled data we use semi-supervised machine learning to make use of unlabeled data. It is understandable our model should be capable of giving results in reasonably less time. .In this paper we proposed many classification algorithm like Support Vector Machine algorithm (SVM) , Random Forest algorithm (RF) and deep neural network.

## I.INTRODUCTION

### 1.1. ABOUT THE PROJECT

Technologies move quickly. New, more advanced technologies are constantly replacing outdated ones. People are able to do their jobs correctly thanks to this new technology. The web marketplace is one such technological development. We might also use websites to make reservations and purchases. Before purchasing any particular goods or services, we check almost all reviews. As a result, online reviews have emerged as an invaluable source of brand recognition for businesses. Additionally, they have a significant impact on product and service marketing and merchandising. Fake online reviews are becoming a serious problem as the internet market expands. It is possible for people to create fake reviews in order to sell their products, which hurts real users. Additionally, rival organizations may use deceptive negative reviews to harm one another's reputations. In order to locate those fake online reviews, researchers have tried novel approaches. Some strategies are determined solely by the content of the reviews, while others are determined by

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

the user's actions while writing the reviews. The user behavior approach focuses on the u S, whereas the content-based search focuses on what is written in the review, i.e. the assessment text. A ., IP address, the number of guides created by reviewers, and so on The majority of the methods that have been suggested are supervised type models. In addition, only a few researchers have made use of semi-supervised models. Because there aren't enough reliable ways to evaluate critiques, semi-supervised methods are being introduced. In this newsletter, we provide specific category strategies for locating fake online reviews. While some of these strategies are monitored, others can only be partially monitored. For semi-supervised learning, we employ an expectation maximization algorithm. To improve category performance, guide vector machines (SVMs) and the naive statistical Bayesian classifier are used in our research. Examiner-based complete procedures' content has received particular attention. We used word count, sentiment polarity, and review length as characteristics.

**1.2. PROJECT DESCRIPTION**

Numerous types of content have proliferated as a result of social media's growing popularity and social network (e. G. User-generated content (UGC) is content that is immediately generated by users and includes audio, video, and text. We are all able to share content on social media almost without any form of external influence thanks to Web 2.0 technologies. This approach implies that it is impractical to a priori verify the credibility of the generated content and the credibility of the sources. Review Spam Detection's goal in this context is to easily identify and identify fake reviews, feedback, blogs, social media posts, scams, and deceptive messages. Techniques for detecting fake reviews were specifically suggested for positive review websites like TripAdvisor1 and Yelp2, where user reviews have a significant impact on people who visit the website online to seek advice.

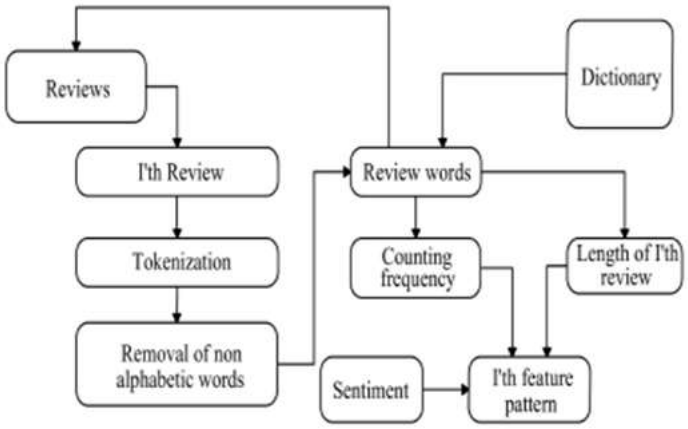


**Fig1.: System Architecture**

**1.3. MODULES**

**1.3.1. Admin**

In this module, the administrator has to log in with a legitimate username and password. After successfully logging in, he can carry out various operations, e.g. B. View All Users and Permissions, Add Products, View All Products, View All Product Ratings, View All Price Ratings, View All Product Recommendations.

**1.3.2. User**

There are n user numbers in this module. The consumer must sign in earlier than performing sure operations. After a hit registration, he should log in with a licensed username and password. If you have successfully logged in, numerous operations together with View My Profile, My Accounts,

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

View All Products, View All Purchased Items and Products, View All Featured Items and Products can be achieved for me

### 1.3.3. Data Preparation

In this undertaking we use how web sites like Olo and Travels Guide have been used. The rating is tied to a 5 big name rating. Each product is related to a class label and a text description. Upload all client critiques for analysis

### 1.3.4. Analyzing Review

To discover fake on-line critiques, we begin with uncooked text information. A dataset that has already been flagged with the aid of preceding researchers eliminates needless textual content from the data, which include B. Articles and prepositions. This textual statistics is then transformed to numeric statistics to make it appropriate for the classifier. The important and important functions are extracted, then the classification procedure starts offevolved. The reviewer records is amassed and analyzed by means of the EM algorithm

### 1.3.5. Detection of Fake

The tokenization process is being used to begin each evaluation. Then candidate function words are generated and pointless phrases are eliminated. Each candidate function word is compared to the dictionary, and its frequency is added to the virtual phrase map-corresponding feature vector column when its function in the dictionary is found. The duration of the examination is measured and added to the function vector along with the count number fee. In the end, a sentiment score of is added to the information set's feature vector. Within the feature vector, we labeled negative moods as 0 and positive moods as excellent.

## II.EXISTING SYSTEM

Content-based techniques consciousness on exam content. Behavioral studies specializes in the reviewer, which incorporates traits of the reviewer. Spotting misleading on line evaluations is generally taken into consideration a class problem, and a popular technique is to use supervised text type strategies.

## 2.1 PROPOSED SYSTEM

The aim is to propose different semi-supervised and supervised text mining fashions to hit upon faux on-line critiques and to examine the performance of the two techniques on a dataset of product reviews. In this article, we gift one of a kind category strategies to locate fake on line critiques, some of which might be partially monitored and others are monitored. Focuses in particular on the content material of evaluation-based strategies. We used word matter, sentiment polarity, and evaluation length as a feature. Project Provides a classification method to come across faux on line critiques. Project Uses a supervised approach to pick out a faux scoring algorithm. They use the Naive Bayes classifier and are used as counterfeit detection classifiers, which improves class overall performance. It makes a speciality of detecting faux content primarily based on consumer metrics and verification procedures.

## III.SYSTEMREQUIREMENTS

### SOFTWARE REQUIREMENTS

| | | |
|---|---|---|
| Operating system | : | Windows 7 Ultimate. |
| Coding Language | : | Python. |
| Front-End | : | Python. |
| Back-End | : | Django-ORM |

### HARDWARE REQUIREMENTS

| | | |
|---|---|---|
| Processor | : | Intel Core I5 |
| Speed | : | 2.1 GHZ |

| | | |
|---|---|---|
| RAM | : | 4 GB (min) |
| Hard Disk | : | 500 GB (min) |

.

## IV.IMPLEMENTATION

The software program implementation segment focuses on translating layout necessities into supply code. The principal goal of the implementation is to create internal documentation of the source code in order that debugging, testing and adjustments can be rolled returned and the conformance of the code to the specification may be without problems confirmed. It enables to keep the supply code as easy and easy as feasible. Good programs are characterised by way of simplicity, readability and elegance. Complexity, brilliance, and vagueness are signs and symptoms of bad design and poorly focused questioning. A persistent approach, respectable programming style, right manuals, internal feedback and the opportunities of modern pc languages assist to enhance the transparency of the supply code. Because single-in, single-out additives allow you to recognize application behavior by analyzing the code from begin to complete, they're on the coronary heart of structured coding. While strictly adhering to this concept can reason troubles, it increases questions about the temporal and spatial performance of the code. Single-entry, unmarried-exit programs may require repetitive pieces of code or repeated calls to subroutines. In such circumstances, the use of this assemble could prevent loops from exiting too early and branching into exception-managing code. To deal with implementation realities, we on occasion violate this perception, even though our intention is not to inspire awful coding practices. The fashion of coding in laptop programming is clear within the templates that programmers use to talk a desired action or end result. While awful programming fashion can undermine the dreams of a brilliant language, right programming style can triumph over the shortcomings of simple computer languages. Good programming fashion ought to offer code this is clean, attractive, and easy to understand.

## V. CONCLUSION

For the purpose of identifying fake online reviews, we have demonstrated a variety of semi-supervised and supervised text mining techniques in this study. By combining elements from a number of research studies, we have enhanced the feature set. In addition, we experimented with a brand-new classifier that was not utilized in the prior work. We were able to make Jiten et al. better as a result. compared the accuracy of previous semi supervised methods and found that the supervised Naive Bayes classifier delivers the highest accuracy. This ensures that our dataset is correctly labeled because we are aware that the semi-supervised model performs well when reliable labeling is not available.

## FUTURE SCOPE

Only user reviews have been the focus of our research. In the future, user behaviors and texts can be combined to create a better classification model. To increase the dataset's precision, sophisticated tokenization preprocessing tools can be utilized. A larger data set can be used to assess the proposed methodology's efficacy. Only English reviews will benefit from this research. It is possible to do so for Bangla and a number of other languages.

## 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] J. W. Pennebaker, M. E. Francis, and R. J. Booth, ”Linguistic Inquiry and Word Count: Liwc,” vol. 71, 2001.
[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] J. Li, M. Ott, C. Cardie, and E. Hovy, “Towards a general rule for identifying deceptive opinion spam,” in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.
[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] J. Karimpour, A. A. Noroozi, and S. Alizadeh, “Web spam detection by learning from small labeled samples,” International Journal of Computer Applications, vol. 50, no. 21, pp. 1–5, July 2012.