IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

METHOD FOR TRACKING NEGATIVE PRODUCT REVIEWS

Dr.T.Bhaskar[1], V.Apoorva Gayathri[2] V.Rithika Reddy[3], G.Kamali[4] ,

[1]Associate Professor, Department of CSE- Artificial Intelligence & Machine Learning, CMRCET, Hyderabad, Telangana

[2,3,4]B. Tech Students, Department of CSE-Artificial Intelligence & Machine Learning, CMRCET, Hyderabad, Telangana

***Abstract*—** Right now, the amount of data on the internet is expanding at an exponential rate. Every day, a lot of data is produced via social media, including reviews, comments, and client opinions. Without any mining operations performed on it, this enormous volume of user-generated data is useless. Since there are a lot of fraudulent reviews, opinion mining techniques ought to include spam detection in order to yield accurate results. A lot of consumers these days use social media comments to shape their purchasing decisions for goods and services. Because so many phony or bogus reviews have been published by individuals or groups for a variety of reasons, opinion identifying spam is a tiresome and challenging task. In order to deceive users or automated detection systems, they fabricate evaluations that either elevate or denigrate specific products in an effort to promote them or damage their reputations. The suggested method uses Ontology, Naïve Bayes-based Spam word dictionary, brand-only review detection, IP address tracking, and geolocation.

Keywords— Machine learning, Opinion Mining, Text Mining

I. INTRODUCTION

E-commerce is one sector that is growing really quickly. E-commerce typically allows users to leave evaluations about the services they have received. It is possible to obtain knowledge from the fact that these reviews exist. Companies can use it, for instance, to decide how best to build their goods or services. Sadly, some people have sought to manipulate the significance of reviews by fabricating fake ones in an effort to either boost the product's popularity or undermine it. They converse online about their ideas. It is common sense for people to research products before making a purchase. Customers can evaluate several brands and choose a product of interest based on reviews. A customer's perception of the goods may be altered by these internet reviews. If these evaluations are accurate, consumers will be better able to choose the right product to meet their needs. However, if reviews are falsified or modified, this could mislead users. Our development of a system that uses the text and rating properties of reviews to identify fraudulent ones for products is aided by this. Data mining techniques will be applied to measure the honesty rating and the degree of a false review. An algorithm has the potential to monitor customer reviews by analyzing the themes and sentiments expressed in online reviews. It can also identify and eliminate fraudulent reviews.

II. LITERATURE REVIEW

"A study on review manipulation classification using decision tree"

A research study on the classification of review manipulation using decision trees has emerged as a significant topic in e-commerce. With an increasing number of customers relying on personal comments from online communities and e-commerce platforms to make purchasing decisions, the reliability of these reviews has become paramount. Unfortunately, some businesses resort to creating fake reviews to influence customer behavior and boost sales. Detecting such manipulated reviews poses a

challenge for customers. Hence, this study employs Decision Tree (DT) algorithms to enhance the accuracy of identifying review manipulation by introducing eight potential attributes associated with manipulated reviews. Additionally, correlation analysis is used to identify important factors in detecting manipulated reviews, and knowledge rules are extracted. Finally, the effectiveness of the proposed method is demonstrated through a case study involving online user comments on smart phones

."**Multiple Aspect ranking using the Good Grief Algorithm"**

We aim to tackle the challenge of analyzing diverse opinions within a text that pertain to different aspects, such as food quality, ambiance, and service, commonly found in restaurant reviews. Our approach frames this task as a multiple aspect ranking problem, seeking to generate numerical scores for each aspect. We introduce an algorithm that simultaneously learns ranking models for each aspect, taking into account the interdependencies among the assigned ranks. By analyzing meta relations between opinions, such as agreement and contrast, this algorithm improves the prediction of individual rankers.We demonstrate that our agreement-based joint model offers greater expressive power compared to individual ranking models. Empirical findings further validate the effectiveness of our model, showing significant improvement over both individual rankers and a leading joint ranking model.

"Analyzing and Detecting review spam"

The extraction, classification, and summarization of opinions from product reviews, forum posts, and blogs have been extensively researched and applied across various domains. However, a significant aspect that remains unexplored is the issue of opinion spam or the reliability of online opinions. This paper addresses this gap, particularly within the realm of product reviews. Despite extensive investigation into web page spam and email spam, there is a notable absence of studies focusing on review spam. We demonstrate that review spam presents unique challenges distinct from those of web page and email spam, necessitating different detection techniques. Leveraging analysis of 5.8 million reviews and 2.14 million reviewers from Amazon.com, we illustrate the pervasive nature of review spam. Our paper introduces a categorization of spam reviews and proposes several techniques for their detection.

"A Joint Model of Text and Aspect Ratings for Sentiment Summarization"

Numerical ratings are commonly found alongside online reviews, representing users' evaluations of various aspects of a service or product. We introduce a statistical model designed to identify corresponding topics within the text and extract textual evidence from reviews that align with each aspect rating. This addresses a key challenge in aspect-based sentiment summarization, as outlined in previous work (Hu and Liu, 2004a). Our model demonstrates strong accuracy without the need for explicitly labeled data, relying solely on user-provided opinion ratings. Furthermore, our proposed approach is versatile and applicable to segmentation tasks in other contexts where sequential data is accompanied by correlated signals.

III. EXISTING SYSTEMS

A. Opinion Mining Using Ontological Spam Detection

"Fake Product Review Monitoring System" aims to combat fraudulent reviews on the internet by employing various techniques such as opinion mining, ontological spam detection, and classification methods like Naïve Bayes. This system utilizes user IP address tracking and account verification to identify reviews originating from spammers, focusing specifically on detecting reviews that solely emphasize brand rather than the product itself. It incorporates a negative word dictionary to classify reviews with an excess of negative terms as spam, as well as employs ontology to categorize reviews discussing unrelated topics as spam. By utilizing sentiment analysis algorithms and data mining techniques, the system aims to accurately identify and block fraudulent reviews, thereby assisting users in finding reliable product evaluations.

B. Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining

In order to detect fraudulent reviews and locate the user, Kohli, Mishra, and Gupta submitted a paper titled "Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining." The majority of customers look for product reviews before making a purchase. Thus, users encounter a variety of evaluations on the website, but they are unable to determine whether they are real or fraudulent. Some product companies post positive ratings on review websites in an effort to increase the product's visibility; these individuals are part of the Social Media Optimization team. For a wide range of goods produced by their own company, they provide positive reviews. The user

won't be able to determine if the review is authentic or fraudulent. Using Opinion Mining, a system called "Fake Product Review Monitoring and Removal for Genuine Online Product Reviews" is introduced to identify bogus reviews on the website. This technology will detect the IP address in order to determine whether phone reviews were posted by the social media optimization team. Using his user ID and password, the user will check in to the system, see the available products, and provide a product evaluation. The user will also receive accurate product reviews. In order for it to be validated, he must input the email address from which he is reviewing while doing so. He won't be able to voice his ideas again if he posts a fraudulent review, as his ID will be blocked. How the system operates is as follows: The system's administrator will add products. In order to access the system, users must input their email address and OTP number. After logging in, the user can view the product and leave a review for it. The user's identity will be checked before reviews are posted. Additionally, if a reviewer's email address is spammed, the administrator will ban them. The fraudulent review will be removed by the admin. Login as the administrator: Admins use their admin ID and password to log into the system. Product Addition: The system's administrator will add the product. Delete Review: The review that the system has identified as fraudulent will be removed by the admin. User Login: Using his user ID and password, the user will log in to the system. See the item: - The product will be viewed. Post Review: The product may be reviewed by users.

IV. LIMITATIONS OF EXISTING SYSTEMS

Limited Context Understanding: Fake review monitoring systems often struggle with understanding the context in which reviews are posted. While they may be effective at identifying certain linguistic patterns associated with fake reviews, they may fail to grasp the subtleties of language and context that can indicate genuine or fake intent. For example, a sarcastic or humorous review may be mistaken for fake content if the system cannot accurately interpret the tone and context of the review.

Vulnerability to Noise and Ambiguity: Fake review monitoring systems are vulnerable to noise and ambiguity in the data, which can lead to inaccuracies in detection. Noise refers to irrelevant or misleading information that can obscure genuine patterns, while ambiguity refers to situations where the meaning of a review is unclear or open to interpretation. Both noise and ambiguity can make it challenging for the system to accurately differentiate between genuine and fake reviews, increasing the risk of false positives or false negatives.

High False Positive Rates: One of the significant limitations of existing fake review monitoring systems is the prevalence of false positives, where legitimate reviews are incorrectly flagged as fake. High false positive rates can undermine trust in the system and lead to frustration among users whose reviews are mistakenly removed or flagged. Addressing false positive rates requires improving the accuracy and precision of detection algorithms, as well as incorporating additional contextual information beyond linguistic features

V. SYSTEM REQUIREMENTS

C. Hardware Requirements:

1. RAM: 256MB (min)
2. Hard Disk: 20GB
3. Processor: Pentium IV (min)
4. Speed: 1.1GHz
5. Monitor: SVGA

D. Software Requirements:

1. Operating system: Windows 10(min)
2. Coding Language: HTML, CSS, Python

E. Libraries:

1. Tkinter
2. sklearn
3. Numpy
4. Pandas
5. Nltk
6. Django
7. Matplotlib
8. RE

VI. METHODOLOGY

To implement this project author has used following modules

1) Data Collection: With this module, we will import the AMAZON reviews dataset into the application.

2) Data Preprocessing: using this module we will read all reviews and then remove stop words, special symbols, punctuation and numeric data from all reviews and after applying Preprocessing we will extract features from all reviews.

3) Features Extraction: here we will apply TF-IDF (term frequency Inverse Document Frequency) algorithm to convert string reviews into numeric vector Each word's frequency will be represented in the vector instead of the words

4) SVM Algorithm: We'll utilize the SVM algorithm on the TF-IDF vector for training, followed by testing the trained SVM model with test data to determine prediction accuracy.

5) Naïve Bayes Algorithm: We'll employ the Naïve Bayes algorithm on the TF-IDF vector for training, and subsequently, we'll test the trained Naïve Bayes model with the test data to compute prediction accuracy.

6) Decision Tree Algorithm: We will apply Decision Tree algorithm on TF-IDF vector to train Decision Tree algorithm and then we apply test data on Decision Tree trained model to calculate Decision Tree prediction accuracy

7) Detect Sentiment from Test Reviews: With this module, we'll import test reviews, and the machine learning algorithm will predict sentiment for each review. In the provided test reviews dataset, sentiment values are absent, and the machine learning model will predict sentiment for each test instance.

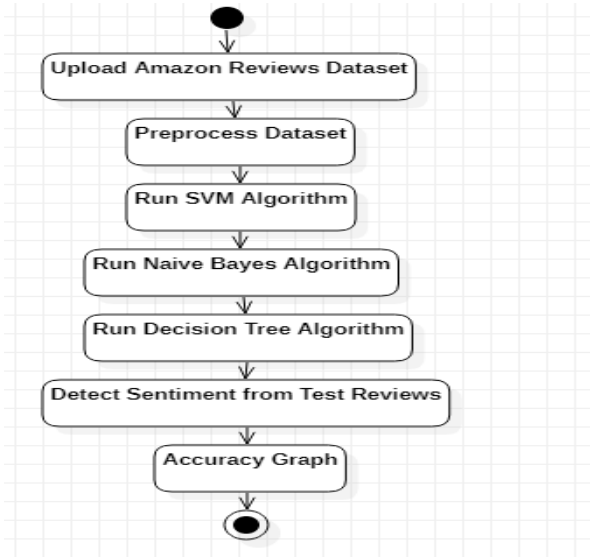
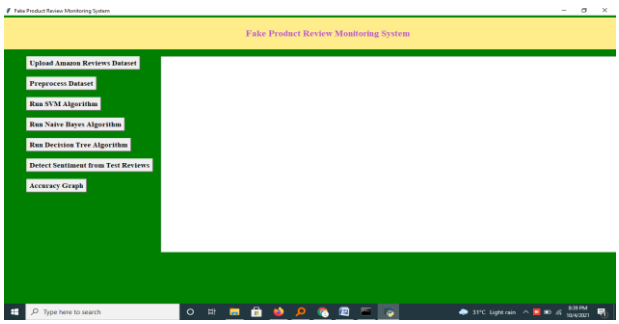
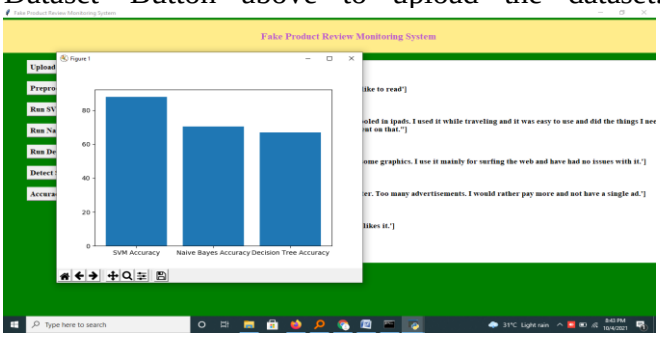


Fig. 1. Flow Chart

VII. RESULTS



Please click the 'Upload Amazon Reviews Dataset' Button above to upload the dataset.



In above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithms and in all 3 algorithms SVM got higher accuracy.

VIII. CONCLUSION

Many techniques for identifying spam reviews have been researched in an effort to improve the accuracy and use of opinion mining. A thorough explanation of the methods currently in use is provided in order to determine whether or not the review is spam. For opinion mining to yield more accurate results, other techniques are integrated, such as IP Address Monitoring and Taxonomy to detect Spam Reviews. A fresh dataset devoid of spam reviews is constructed when the spam reviews in the original dataset are identified, and opinion mining is then done on the newly created Spam Filtered Dataset. Finally, a new approach that employs spam filtered data for opinion mining is proposed, improving the accuracy of detecting spam reviews.

IX. REFERENCES

[1] Raja sree S. Jadhav, Prof. Deepali V. Gore, "A New Approach for Identifying Manipulated Online Reviews using Decision Tree ". (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), pp 1447-1450, 2014

[2] Long- Sheng Chen, Jui-Yu Lin, "A study on Review Manipulation Classification using Decision Tree", Kuala Lumpur, Malaysia, pp 3-5, IEEE conference publication, 2013.

[3] Benjamin Snyder and Regina Brazil, "Multiple Aspect ranking using the Good Grief Algorithm "Computer Science and Artificial Intelligence Laboratory Massachusetts Institute of Technology2007.

[4] Ivan Tetovo, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization "Ivan Department of Computer Science University of Illinois at Urbana, 2011

[5] N. Jindal and B. Liu, "Analysing and detecting review spam," International Conference on Web Search and Data Mining, 2007, pp. 547-552.

[6] N. Jindal and B. Liu, "Opinion spam and analysis," International Conference on Web Search and Data Mining, 2008, pp. 219-230.