



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 6, June 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.512**



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

# Spam Review Analysis and Detection System

**Mansi Pakhale<sup>1</sup>, Swati Pandey<sup>2</sup>, Gauri Patil<sup>3</sup>, Mohan Yelpale<sup>4</sup>**

Student, Dept. of Computer Engineering, NBN Sinhgad College, Maharashtra, India<sup>1</sup>

Student, Dept. of Computer Engineering, NBN Sinhgad College, Maharashtra, India<sup>2</sup>

Student, Dept. of Computer Engineering, NBN Sinhgad College, Maharashtra, India<sup>3</sup>

Professor, Dept. of Computer Engineering, NBN Sinhgad College, Maharashtra, India<sup>4</sup>

**ABSTRACT:** As Customers are mostly dependent on the reviews and comments before purchasing any product. This dependence has prevailed the entries of fake and spam reviews which get camouflaged among authentic ones. These reviews are written to mislead readers or automated detection systems when promoting a target product or downgrading it to boost their reputations. We propose a framework to categorize the comments from the Play Store App into different categories and then analyze the comments. The Opinion mining is also known as Sentiment Analysis is used. Here we use Spam dictionary to identify the spam words in the reviews. In Text Mining we apply several algorithms from Machine learning and detect the categories of the comments or reviews.

**KEYWORDS:** Opinion Mining, Sentiment analysis, Spam Reviews, Machine Learning, Aspect Categorization.

## I. INTRODUCTION

One area of explosive growth is e-commerce. Online stores generally allow their customers to write reviews of the service they have purchased, and these reviews can be used as a source of information. For examples, However, today's technological advancement has made it possible to use the reviews to design new products or services, but unfortunately, the importance of the review is misused by certain parties, who generate fake reviews to promote their products or discredit them. The original product review and rating can be found once you have access to the system. Then, it is the source of satisfaction and reliability for you. A technique for identifying fake reviews is proposed in which the system extracts reviews related to a given product and then analyzes them to identify the most credible reviews for that product. Previous research detects fake reviews using different approaches including identification address, opinion mining and sentiment analysis, machine-learning approach. Here, as an add on work we have used Play Store App and its comments classification tab provided that the link for the app is given, where comments below a certain app on Play Store will be classified as different categories.

## II. RELATED WORK

“Opinion Mining Using Spam Detection” paper which will help us to find out fake reviews by using machine learning algorithms. To find out fake review in the website this “Fake Product Review Monitoring System” system is introduced. This system will find out fake reviews made by the customers and it will block the users. To find out the review is fake or genuine, we will use some classification such as

1. If multiple reviews are coming from the same IP address then the Reviews are considered Spam.
2. Using Account Used to check whether the reviews are done using the same account.
3. If there are more than five negative words in the review, then it was identified as Spam. For example, a user has posted a review: “This product is a piece of junk, the design is bad, the quality is inferior, and the price is as low as at minimum. Here, this sentence consists of 4-5 negative words. So, the system will check the count of negative words, if the count exceeds, then it will be considered as spam review. Therefore, Negative Word Dictionary will be used with customized Sentiment strength algorithm. According to this approach, probability of given review to be Spam is more so it will be considered a Spam.

Second Paper titled “Spam Review Detection Using Deep Learning” where different Deep learning algorithms are used for the classification of reviews and to find its accuracy. Algorithms like LSTM, KNN, TF-IDF and N-gram techniques have been used and proven the accuracy of classification. The deep learning concepts as Multilayered Perceptron (MLP), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) is used in the paper which we have also included heavily in our paper.

### III. METHODOLOGY

1. The proposed system takes the URL of the play store App, and then store all the reviews related to each product in a CSV file. Apply sentiment analysis and add sentiment polarity column. Sentiment is given based on polarity values given below. Polarity > 0 (positive) Polarity < 0 (negative) Polarity = 0 (neutral).
2. The data was pre-processed when it was extracted using Naive Bayes and Decision Tree algorithms, resulting in a csv file of positive and negative reviews and a confusion matrix.[5] Categorization is the main Process we are applying for classification of the comments.
3. We have already created the user review dataset and train that model using the polarity basis method. This dataset trained model is being used for classification of app comments into categories like Positive, Negative or Neutral.
4. The most important feature in text classification is a n-gram that is word based or character centered. System uses this model in order to generate features to classify the reviews. In the proposed approach, bigrams (N = 2) is using to find whether the reviews are deceptive or truthful. After converting each review in bigram form, system is done with the preprocessing step and now system is ready move to the feature extraction technique.
5. A proposed system is meant to classify reviews from into different categories. It collects the data, classifies it using lemmatization and inserts punctuation and stop words in the words of the reviews and then processes it using the Aspect Categorization technique.
6. After Categorization, the comments can be analyzed on the basis of how many positive and negative comments are being made and thus can give the overview of the app to the user before downloading.

### IV. EXPERIMENTAL RESULTS

Pseudocode

Tokenization - This technique removes  
Special character and images.

Initialize feature vector bg feature =[0,0..0]

for token in text.tokenize() do

if token in dict then

token idx =

getIndex(dict,token) bg

feature[token idx]++

else

conti

nue

end if

end

for

Algorithm

Aspect Identification Input:

Collection of sentences =fS1, S2,

S3.. Sng Output:

Aspects assigned to sentences

1. initialize aspects

2. for all sentences do

3. stanford tagger = SPOS (sentences)



4. if NN in stanford tagger then
5. aspects NN
6. end if
7. end for
8. initialize aspects groups
  
9. for all aspects do
10. WordNet sets = WNSS (aspects)
11. if TRUE in WordNet sets then
12. aspects groups aspects
13. end if
14. end for
15. frequent aspects = freq measure (aspects, aspects groups, 10)
16. tree = DT (sentences, frequent aspects)
17. initialize aspect assigned sentences
18. for all sentences do
19. aspect identification = tree (sentences)
20. if TRUE in aspect identification then
21. aspect assigned sentences aspect identification
22. end if
23. end for
24. return aspect assigned sentences

## V. CONCLUSION

We proposed a System which helps the users to Analyze the comments of the app by the Categorization review technique. Use of algorithms like Naïve Bayes, Aspect categorization technique from Sentiment Analysis. Our software will help the user to pay for the right product. We added a feature of classifying Play Store App comments, because of increasing irrelevant comments on unrelated app. It monitors the spam review made on any product/app. And user can be sure about the products availability on that application and reviews too.

## REFERENCES

1. A. Sinha, N. Arora, S. Singh, M. Cheema, and A. Nazir, "Fake Product Review Monitoring Using Opinion Mining," vol. 119, no. 12, pp. 13203– 13209, 2018.
2. Torbet, Georgina. "U.S. Customers Spent over \$6 Billion on Black Friday Purchases." Digital Trends, Digital Trends, 25 Nov. 2018.
3. Sterling, Greg. "Study Finds 61 Percent of Electronics Reviews on Amazon Are 'Fake'." Marketing Land, 19 Dec. 2018.
4. K. Khan, W. Khan, A. Rehman, A. Khan, Asfandiyar. Khan, A. Ullah Khan, B. Saqia, "Urdu Sentiment Analysis," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 9, 2018.





**Impact Factor:  
7.512**



# **INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH**

**IN SCIENCE | ENGINEERING | TECHNOLOGY**

**📞 9940 572 462    📞 6381 907 438    ✉️ [ijirset@gmail.com](mailto:ijirset@gmail.com)**



**[www.ijirset.com](http://www.ijirset.com)**

Scan to save the contact details