RESEARCH ARTICLE                                                                                    OPEN ACCESS

# Spam Review Detection Using Linguistic and Spammer Behavioral Methods

## Mrs R Jhansi Rani MCA [1], Karanam Mohan Kumar [2]

[1] Asst. Professor, Department Computer Application
[2] Student, Department of Computer Application
[1], [2] Chadalawada Ramanamma Engineering College (Autonomous)

**ABSTRACT**
Online reviews regarding different products or services have become the main source to determine public opinions. Consequently, manufacturers and sellers are extremely concerned with customer reviews as these have a direct impact on their businesses. Unfortunately, to gain profits or fame, spam reviews are written to promote or demote targeted products or services. This practice is known as review spamming. In recent years, the spam review detection problem has gained much attention from communities and researchers, but still there is a need to perform experiments on real-world large-scale review datasets. This can help to analyze the impact of widespread opinion spam in online reviews. In this work, two different spam review detection methods have been proposed: (1) Spam Review Detection using Behavioral Method (SRD-BM) utilizes thirteen different spammer's behavioral features to calculate the review spam score which is then used to identify spammers and spam reviews, and (2) Spam Review Detection using Linguistic Method (SRD-LM) works on the content of the reviews and utilizes transformation, feature selection and classification to identify the spam reviews.
**Keywords: -** Spammer, Span detection, Online reviews, Spam reviews.

## I.   INTRODUCTION

Review spam is usually related to email and web spam. The web spam is used to attract people by manipulating the content of the page so that the web page will be ranked highly by the search engines. Email spam is mainly used for advertising purposes. However, spam reviews are different in a sense as these give the wrong opinion about a product/ service and it is very difficult to detect spam reviews manually. Therefore, existing web spam or email spam detection techniques are not suitable for spam review detection. Spam review detection is a challenging task as no one can detect a review as spam by simply reading its text. Review websites are usually open to public reviews. Therefore, any user can act as spammer to write spam reviews about any product and/or service. Spam reviews appear as legitimate until different spammer behavioral features and/or the review text is analyzed to identify the spam reviews. Based on these perspectives, existing approaches of Spam Review Detection (SRD) utilizes spammer behavioral features or linguistic features for the detection of spammers and spam reviews respectively.

The linguistic feature considers review text to identify the reviews as spam or not spam; whereas behavioral features reflect the behavior of reviewer in terms of time stamp of review, review rating, user profile, etc. From the literature review, it has been observed that existing approaches either adapted linguistic methods or utilized behavioral characteristics separately to identify the spammers and spam reviews. Most of the existing works have only utilized the uni-gram linguistic approach to classify reviews. Usually, the uni-gram approach produces good results but fails in some cases. For example, in the following review; ''This hotel is not good'' when analyzed through the uni-gram approach, gives the popularity of the review as neutral with one positive word ''good'' and one negative word ''not''. But when the same review is analyzed using a bi-gram approach, it gives a negative impression due to the use of the words ''not good''.

Considering this limitation, this research intends to utilize N-gram approach to accurately analyze spam reviews. Similarly, many existing approaches ignored several important behavioral features while developing behavioral models for spammer detection. Therefore, there is still a need to employ all existing behavioral and linguistic features to accurately filter out spam and not-spam reviews. The aim of this work is to develop an SRD model adapting a vast set of behavioral and linguistic features on large-scale real-world dataset. In this study, the investigation about the spam review is based on 26.7 million reviews and 15.4 million

reviewers from Amazon.com. However, the main limitation of this domain is that the available datasets are unlabelled, the same is the case with Amazon dataset. To tackle this problem, the proposed approach first formulates a procedure of Spam Review Detection using Behavioral Methods (SRD-BM) to create a labelled dataset. This labelled dataset, then, utilizes Spam Review Detection using Linguistic Method (SRD-LM) to train the classifiers. Specifically, the proposed approaches incorporated linguistics features, such as N-gram techniques, and a number of spammer behavioral features, such as activity window, review count, the ratio of a positive review, the ratio of negative reviews, the ratio of the first review and the review length, for developing the spam review detection model.

## II.  RELATEDWORKS

Guoliang He, Ming Zhong & Qingxi Peng Proposed Detecting Professional Spam Reviews. Online reviews about the purchase of products or services provided have become the main source of users' opinions. In order to gain profit or fame, usually spam reviews are written to promote or demote a few target products or services. This practice is known as review spamming. In the past few years, a variety of methods have been suggested in order to solve the issue of spam reviews. In this study, the researchers carry out a comprehensive review of existing studies on spam review detection using the Systematic Literature Review (SLR) approach. Overall, 76 existing studies are reviewed and analyzed. The researchers evaluated the studies based on how features are extracted from review datasets and different methods and techniques that are employed to solve the review spam detection problem. Moreover, this study analyzes different metrics that are used for the evaluation of the review spam detection methods. This literature review identified two major feature extraction techniques and two different approaches to review spam detection. In addition, this study has identified different performance metrics that are commonly used to evaluate the accuracy of the review spam detection models. Lastly, this work presents an overall discussion about different feature extraction approaches from review datasets, the proposed taxonomy of spam review detection approaches, evaluation measures, and publicly available review datasets

SimranBajaj, NiharikaGarg and Sandeep KumarSingh proposed A Novel User-based Spam Review Detection. Taking into account the popularity of sites like Yelp, TripAdvisor or Foursquare- posting online reviews is a very popular way to share opinion on social media websites. 90% of consumer reviews do have an influence on the public. But the trustworthiness of these reviews is still an open issue. The existing researches have focused on the sentiment analysis to detect spam reviews but neglected the personal characteristics of a person posting reviews. This work has focused on spam detection using personal characteristics rather than the reviews. Majority of E-commerce sites describe a customer superficially using his ID (name, email ID). But that is not sufficient to identify the uniqueness of a customer. This work has used two additional attributes of the customer to detect spam reviews like his geographical location and the IP address of the device with which he is accessing different resources on Internet. In addition, we have also proposed a content analysis method to attack non-reviews using spam dictionary.

ToanOng, MichaelMannino and DawnGregg proposed Linguistic characteristics of shill reviews Description: This exploratory study investigates the linguistic characteristics of shill reviews and develops a tool for extracting product features from the text of product reviews. Shill reviews are increasingly used to manipulate the reputation of products sold on websites. To overcome limitations of identifying shill reviews, collected shill reviews as primary data from students posing as shills. Using semi-automated natural language processing techniques, compared shill reviews and normal reviews on informativeness, subjectivity and readability. The results showed evidence of substantial differences between shill reviews and normal reviews in both subjectivity and readability. Informativeness appears to be a mixed separator of shill and normal reviews so additional studies may be necessary. Overall, the study provides improved understanding of shill reviews and demonstrates a method to extract and classify features from product reviews with an eventual goal to increase effectiveness of review filtering methods.

In existing system offered a text mining model by using the unsupervised approach and features, relying upon the time integration among multiple time durations. In addition, this model was integrated with the semantic language model for spotting spam reviews and used a Yelp dataset. In existing system have suggested that the author spamicity unsupervised model has been based on features such as the review posting rate and temporal pattern. The

model produced two clusters: spammers and truthful users. The datasets were gathered from the Chinese website Dianping4 to train the proposed model. In the existing system have observed a network model for spam review detection. In their work, the correlation among users and products was captured and the algorithm was used to recognize the spam reviews.The existing system have applied the semantic language model to identify spam reviews. The authors used the Support Vector Machine classifier to train the proposed method. existing system used a supervised learning approach with a co-training method to highlight spammers based on linguistic features. existing system proposed a classification method that used N-gram characters as a linguistic feature. Moreover,the proposed method used the Naïve Bayes to classify spam and not-spam reviews. existing system have designed a dataset for spam review detection, employing a crowd source through AMT (Amazon Mechanical Turk). The authors found that the classifier performed better by adding elements such as psycholinguistic features. in existing system used statistically based features for the Extreme Gradient Boost Model and Generalized Boosted Regression Model to evaluate multilingual datasets (i.e., the Malay and English languages). It was observed by the experimental results that the

Extreme Gradient Boost Model performed better for the English review dataset and the Generalized Boosted Regression Model performed better for the Malay dataset. Kumar. In existing system have proposed a hierarchical supervised- learning method. This method analyzed reviewer's behavioral features and their interactions using multivariate distribution. In the existing system recommended a supervised model based on reviewer features to identify spam reviews. In the existing system used various rule-based machine learning algorithms. Moreover, the authors compared the effectiveness of the proposed method through a Ten-Fold cross-validation training model for sentiment classification. existing system performed different experiments using the threshold-based method to identify spam reviews. The authors proposed different time-sensitive features to find spam reviews as early as possible and trained the model by using the SVM classifier.the feature-based sparse additive generative model and the SVM classifier to discover the general rule for spam review detection. In the existing system, the system is using Low performance supervised learning approach with a co-training method to highlight spammers based on linguistic features. This system is less performance due to lack of Spam Review Detection which is not using Behavioral Method (SRD-BM).

## III. PROPOSED SYSTEM ARCHITECTURE

A framework of the proposed SRD-BM in which the system process starts with the identification and calculation of spammer behavioral features in unlabelled Amazon review dataset. This calculation is performed on all reviews of the dataset based on the equations listed in the proposed system. The average score of respective reviews in dataset is calculated using normalized values of each behavioral feature. This average score is then used to calculate accuracy of spam review identification using mean value method. Next, to identify the importance of each behavioral feature, the process continues by dropping each feature one-by-one and recalculates the updated average score, named as drop score.

The accuracy achieved using average score is compared with that of drop score. If the achieved accuracy is dropped by 5% than a weight of ``2" is assigned to that specific dropped behavioral feature otherwise a weight of ``1" is assigned. Similarly, all behavioral features are assigned weights based on their importance in the dataset. Next, spam score of each review is calculated with respect to the assigned weights to each behavioral feature. This spam score is then compared with a variable threshold to highlight the review as spam or not spam.
The proposed SRD-BM executes in four phases: (1) First it calculates the normalized value (0-1) of each spammer behavioral feature. (2) Based on these values, it calculates the mean score for each review and the overall accuracy of the complete dataset. (3) Next, it assesses the impact of each behavioral feature by following dropping feature method and assigns a weight according to the importance of each behavioral feature. (4) Finally, it calculates spam score using weighted behavioral features and identifies spam and not spam reviews using different threshold values.
In this proposed system there are two modules they are:

1. Review Analyst
2. User

**Review Analyst**: The following are the functionalities provided by the Review Analyst:

1. Login

2. Data splitting
3. Train data with SVM and LR
4. View product data set details
5. Search product data set details
6. View spam reviews
7. View positive reviews
8. View negative reviews
9. Generate product score results
10. Logout

**User**: The following are the functionalities provided by the User:

1. Register and login
2. Add product data set details
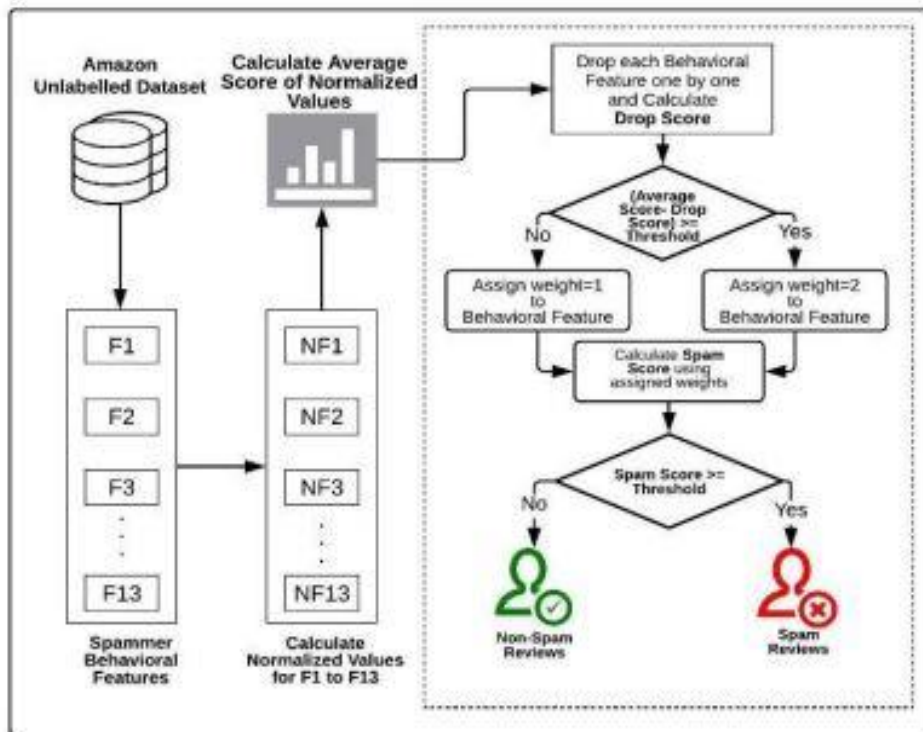3. Search an product data set
4. View your profile
5. Logout



Fig.1 Proposed system architecture

## IV. RESULTS AND DISCUSSION

The output screens obtained after running and executing the system are shown from Fig.2 to Fig.8
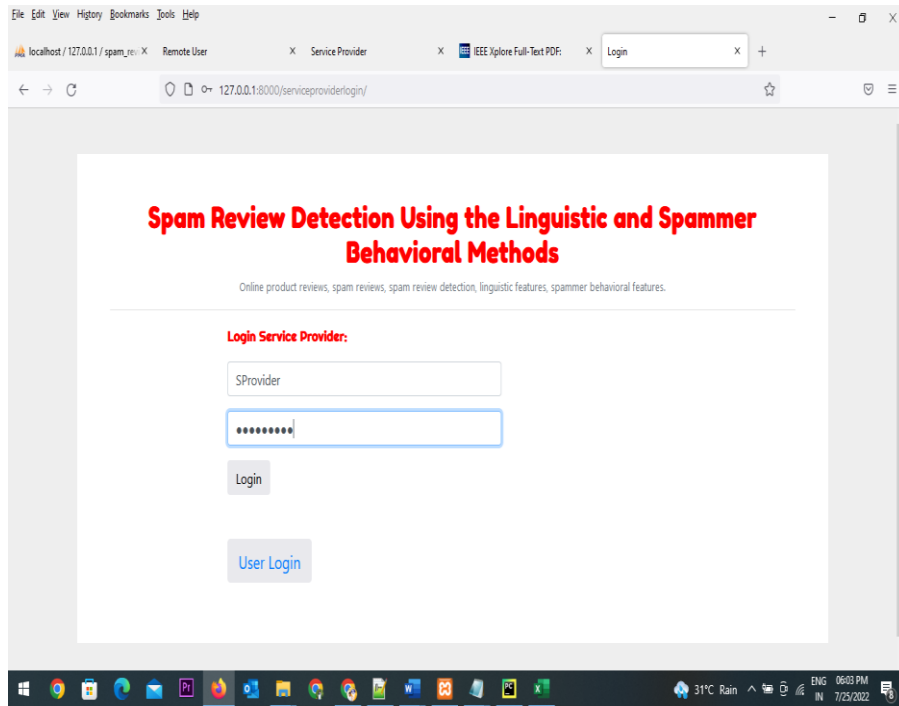
Fig.2 Service Provider Login
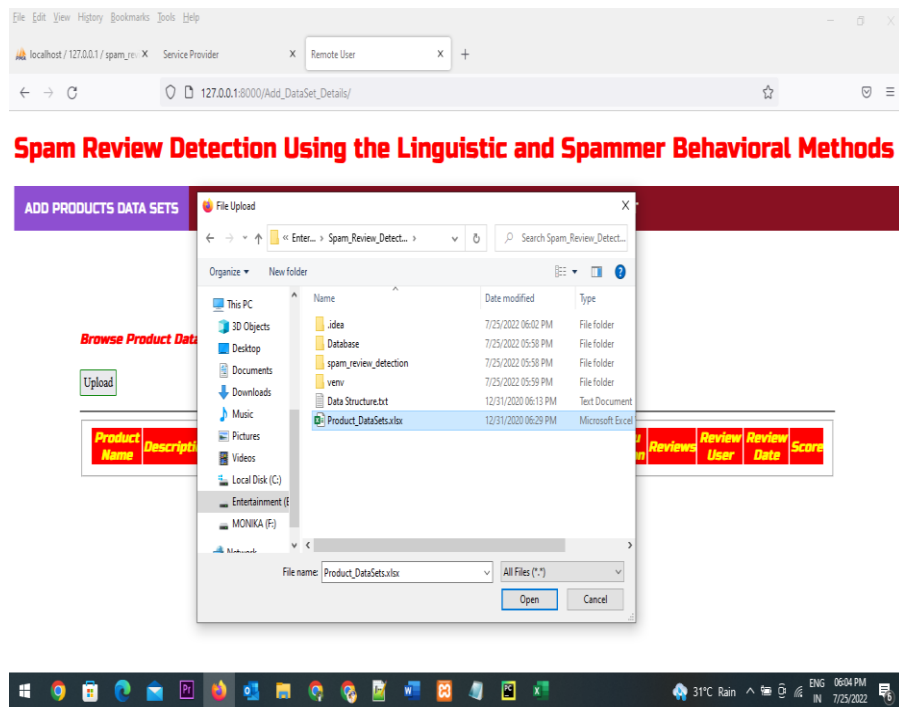


Fig.3 Browse data set
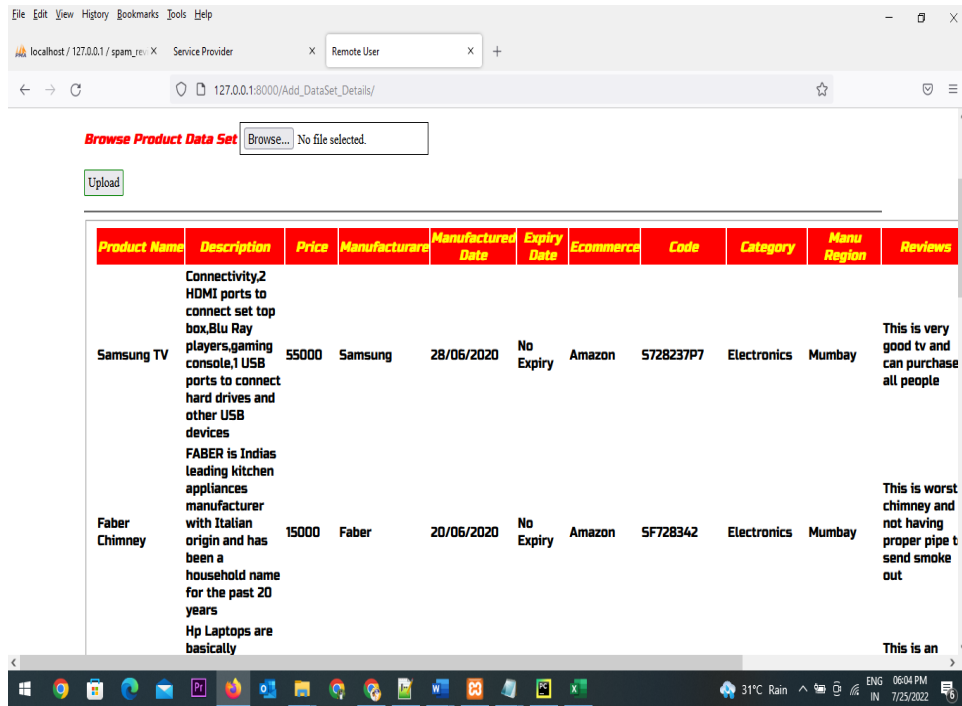
Fig.4 Data set details



Fig. 5 Spam reviewers

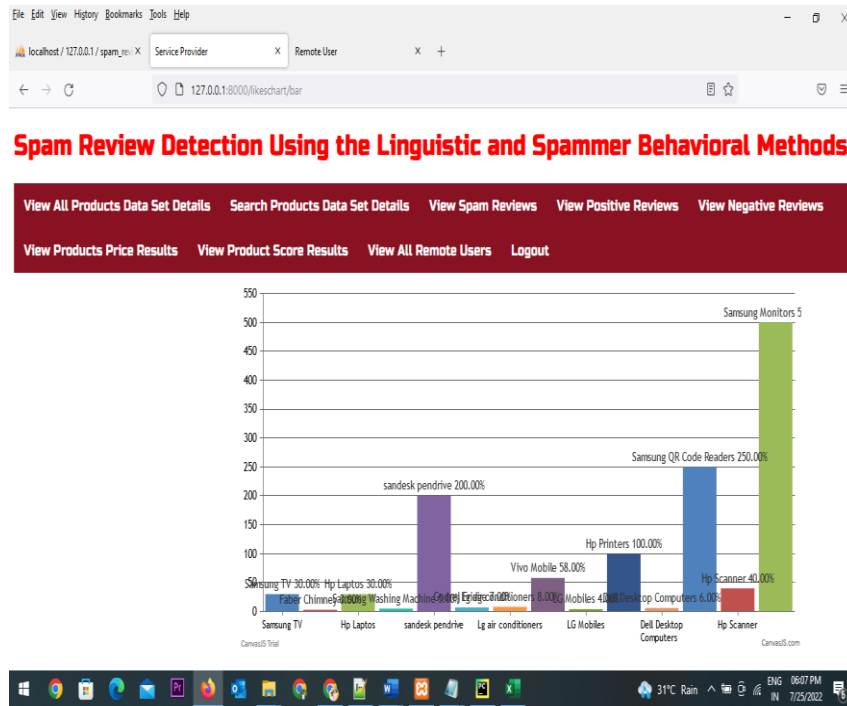Fig.6 Positive reviews



Fig.7 Negative reviews

Fig.8 Product score results

## V. FUTURE SCOPE AND CONCLUSION

In recent years, the spam review detection problem has gained much attention from communities and researchers, but still there is a need to perform experiments on real-world large-scale review datasets. This can help to analyze the impact of widespread opinion spam in online reviews. In this work, two different spam review detection methods have been proposed: Spam Review Detection using Behavioral Method (SRD-BM) utilizes thirteen different spammer's behavioral features to calculate the review spam score which is then used to identify spammers and spam reviews. In the future work,Spam Review Detection using Linguistic Method (SRD-LM) was implemented on the content of the reviews and utilizes transformation, feature selection and classification to identify the spam reviews

## REFERENCES

[1] Gunter Ollmann, "The Phishing Guide Understanding & Preventing Phishing Attacks", IBMInternet Security Systems, 2007.
[2] https://resources.infosecinstitute.com/category/enterprise /phishing/the-phishing-landscape/phishing-data-attackstatistics/#gref
[3] Mahmoud Khonji, Youssef Iraqi, "Phishing Detection: A Literature Survey IEEE, and Andrew Jones, 2013
[4] Mohammad R., Thabtah F. McCluskey L., (2015) Phishing websites dataset. Available: https://archive.ics.uci.edu/ml/datasets/Phishing+Websi tes Accessed January 2016
[5] http://dataaspirant.com/2017/01/30/how-decision-treealgorithm-works/
[6] http://dataaspirant.com/2017/05/22/random-forestalgorithm-machine-learing/
[7] https://www.kdnuggets.com/2016/07/support-vectormachines-simple-explanation.html
[8] www.alexa.com
[9] www.phishtank.com