E-COMMERCE PRODUCT REVIEWS USING SENTIMENTAL ANALYSIS

[1]S.Devi , [2]S.Bharath Kumar, [3]S.Keerthivel ,[4]S.Ranjith
[1]Assistant Professor, [2]Student, [3]Student, [4]Student
[1]Department of Information Technology,
[1] Coimbatore Institute of Technology, Coimbatore, India

Abstract: Sentiment analysis is such a groundwork space that understands and extracts the opinion from the given review and also the analysis method includes Natural Language Processing (NLP), linguistics, text analytics and classifying the polarity of the opinion. With in the field of sentiment analysis there are several algorithms exist to tackle natural language processing issues. On scrutiny with numerous algorithms like GA(Genetic Algorithm) and PCA(Principal component Analysis), it's that combining SVM(Support Vector Machine) with PSO(Particle swarm optimization) offers an improved accuracy for optimization for product reviews characteristic pretend product or not. In this paper we have used NLP for text preprocessing, Naive Bayes classifier for rating classification, and SVM algorithm to detect fake reviews and spam by providing a better accuracy.

Index Terms – Natural Language Processing, Support Vector Machine, Naïve Bayes Classifier

I. INTRODUCTION

Sentiment analysis is very helpful in numerous things. However it's terribly tough method attributable to the quality concerned within the human language. it's got many variants like grammatical, cultural etc. Humans will simply interpret statements like "My order been delayed. Very good". However it should be tough for the machine to know. Equally word "thin" could also be taken as positive with regard to laptop computer however it should be negative once it involves flat wall. Therefore to administer the proper call sentiment analysis should be someday a lot of business specific. Every single day large amount of knowledge, reviews or opinions have gotten hold on at intervals the websites of social media or e- services at intervals the sort of raw data. To work with those information correct ways in which required. Most of the ways in which either focus on verbs, nouns, adverbs or adjectives. A recent study has shown that combination of adverbs and adjectives in sentiment analysis is higher than adjectives alone. But no work has centered on all the potential combos of adverbs, adjectives and verbs.

II. LITERATURE SURVEY

Chonghui Guo [1] gave a ranking based method for online reviews based on the different aspects of the products, which has both objective as well as subjective sentiment values. Weights of these aspects are determined with LDA (Linear Discriminant Analysis) topic model to calculate the objective sentiment value of the product. Consumers personalized preferences are taken into consideration for calculating total scores of alternative products.

Rekha [2] proposed some common features like Camera, Battery, Screen, Sounds, Design and hardware/software performance, of three handsets and then identified and collected related reviews as dataset for their proposed algorithm. They collected positive as well as negative words that are used in reviews frequently. For mining of reviews, they used MATLAB as the platform for simulation. Their proposed algorithm counts the positive and negative features for aquiring the reviews results.

Ms. Payal Yadav [3] focused more on how emoticons play an important role in sentiment analysis. The factors that affect sentiment analysis were analyzed briefly in this paper. Various issues such as sarcasm detection, multilingualism, handling acronyms, slang language, lexical variation and dynamic dictionary handling are been discussed.

Sankar.H [4] gave a comprehensive overview on Sentiment analysis using NLP techniques. Sentiment analysis is one of the application of NLP which aims at identifying the sentiment of the given textual resources. To achieve optimal results we propose a new approach called Hybrid Weighted Word2Vec (HWW2V) text representation. Which is collection of BOW, Weighted Word2Vec and Sentiment lexicon based representation.

S.Rajalakshmi has brought out the process involved in sentiment analysis. The investigation was about the various methods for performing sentiment analysis. It also showed the various tools that were used to explain the process involved in analyzing sentiment.

Kai Yang [6] constructed the domain sentiment dictionary using external textual data. Many classification models can be used to classify documents according to their opinion. These individual models have strengths and weakness. They proposed a highly effective hybrid model combining different single models to overcome their weaknesses. The experimental results showed that their hybrid model outperforms baseline single models.

III. RESEARCH METHODOLOGY

3.1 Sentimental Analysis

Sentiment analysis is one of the important feature of NLP. Sentiment analysis has played a vital role in recent years. It is the mining of text which extracts information in source material. With the ability of algorithms to analyze text. It allows us identify customer sentiments towards products, brands or services in online conversations and feedback. It detects polarity within the text. It can identify issues in real time. It uses stemming, tokenization, parts of speech tagging, lexicons. It also undergoes many comparisons with emojis.

3.2 Naïve Bayes Classifier

Naive Bayes Classifier uses feature vector and analyzes them individually as they are equally independent of each other. The conditional probability is defined as

$$P(X|y_j) = \Pi_{i=1}^{m} P(x_i|y_j) \tag{1}$$

Feature vector 'X' defined as $X=\{x_1, x_2, ... x_m\}$ and y_j is the class label. There are some features like emoticons, keywords, counting positive and negative words, that are utilized for this classification.

3.3 Natural Language Processing

Text classification is a process of categorizing the texts into structured groups. By using this, the text classifiers can analyze text and assign a set of pre-defined tags or categories based on its content automatically.
Unstructured text like emails, websites, and social media like twitter, facebook are hard to extract value from the data unless it is organized in a particular way. Doing so used to be a difficult and expensive process since it required spending time and resources to manually sort the data or creating handcrafted rules that are difficult to maintain. Text classification with NLP have proven to be a great alternative as it is fast, cost-effective, and scalable way.

3.4 Support Vector Machine

It is a supervised machine learning model which uses classification algorithms. It is defined by a separating hyperplane. With the given set of sample of labeled training data, the algorithm gives an output of an optimal hyperplane which categorizes new examples. In 2D space this hyperplane divides a plane into two 'parts where in each class lies in either side. It uses the discriminative function defined as

$$g(X) = w^T\varphi(X) + b \tag{2}$$

In the above equation, 'X' is the feature vector, 'w' is the weights vector, 'b' is the bias vector, $\varphi()$ is a non linear mapping from input space to high dimensional feature space. 'w' and 'b' are learned by itself in the training set.
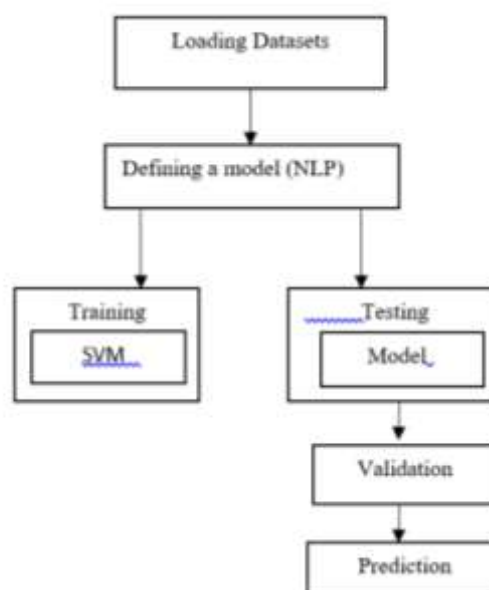

Fig:3.1: System Diagram

IV. FRAMEWORK

The sequence of the proposed framework is as follows:

4.1 Preprocessing

We standardized the data by handling null values,imputation,stemming,lemmatization,POS tagging, n-grams and filtering the dataset. We have done sentiment analysis based on the product review rating by using Naive Bayes classifier.

Table 4.1: Statistics of Dataset used

Dataset	21000
Training	16800
Testing	4200

4.2 Defining an SVM Model

SVM model is a supervised Machine Learning algorithm used for classification-predicting a label or group and regression-predicting a continuous value. It performs classification by finding the hyperplane that differentiate the classes we plot in n-dimensional space. We detected fake reviews in the dataset that were given by the customers. We then detected the spam reviews which has generated automatically using opinion mining.

### 4.3 Training and testing of the datasets

Datasets are separated into training and testing datasets.80% percent of data are used to train the model and it is been repeatedly trained so that the dataset shows an increased accuracy and 20% data are used to test the model.

4.4 Validation & Prediction

The predicted values are found using the above SVM algorithm. We got an accuracy of 80% in fake review detection. Then based on the evaluation of the result and summary are predicted.

V.EXPERIMENTAL RESULTS AND DISCUSSION

The data required for this study was collected from Kaggle website.The constrains that are used for this study includes product name, ratings, and reviews are selected so as to get accurate output. Then the sentimental analysis has been made from the collected datasets using certain specified functions.
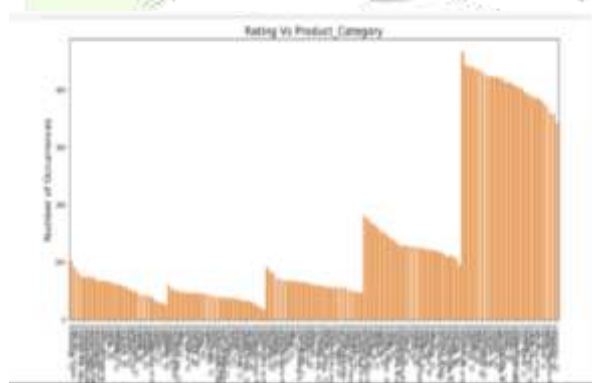
Fig 5.1 Product Categories and their Rating

Fig 5.2 Accuracy of Fake Reviews

VI.CONCLUSION

In this paper a brief overview of categorizing product reviews into positive and negative by text classification and rating. Fake review analysis is done for the motive of identifying most genuine reviews with the maximum percentage of accuracy. In addition to it we have done spam detection detecting the spam reviews which are generated automatically. Different datasets yields extremely different results. This results have been **evaluated** with accuracy of 80%.

REFERENCES

[1] Products ranking through aspect based sentiment analysis of online heterogeneous reviews by Chonghui Guo, Zhonglian Du and XinyuKou**.**
[2] Sentiment analysis of online mobile reviews by Rekha and Dr. Williamjeet Singh
[3] Sentireview: sentiment analysis based on text and emoticons by Ms. Payal Yadav and Prof. Dhatri Pandya**.**
[4] Investigating sentiment analysis using machine learning approach by Sankar.H and Subramaniyaswamy.
[5] A comprehensive survey on sentiment analysis by S.Rajalakshmi, S.Asha, and N.Pazhaniraja
[6] An effective hybrid model for opinion mining and sentiment analysis by Kai Yang, Yi Cai, Dongping Huang, Jingnan Li, Zikai Zhou