Characterizing Products' Outcome by Sentiment Analysis and Predicting Early Reviewers, using Early Reviews

K Mohan Krishna[1], Md Pardaz Banu[2], P Priyanka[3], M Pravallika[4], P Pushpa Latha[5]

[1]Asscociate Professor Mohan Krishna, Dept. Of CSE, VVIT, Andhra Pradesh, India
[2]Student Mohammad Pardaz Banu, Dept. Of CSE, VVIT, Andhra Pradesh, India
[3]Student Padinam Priyanka, Dept. of CSE, VVIT, Andhra Pradesh, India
[4]Student Meda Pravallika, Dept. of CSE, VVIT, Andhra Pradesh, India
[5]Padavala Pushpa, Dept. of CSE, VVIT, Andhra Pradesh, India

---***---

Abstract - *Now-a-days, people are depending more on reviews while purchasing a product online. A product's lifetime is divided into three consecutive stages, namely early, majority and laggards. The reviews that are posted during the early lifetime of a product are called early reviews and the users who post them are called early reviewers. It is known that the early reviews majorly influence product popularity. Based on the reviews of these early reviewers, a product's sale is dependent. Hence, the success or failure of a product can be determined by the early reviews. In our research paper, we are using the SVM model, which is a supervised learning algorithm to find out what products are getting successful and what products are getting failed by using the data of different products' early reviews from e-commerce websites. To find early reviewers, we are using a statistics-based method, PER.*

*Keywords***: Early reviewers, Early reviews, E-Commerce, SVM model(Support Vector Machine), supervised learning.**

1.INTRODUCTION

E-commerce websites have provided users a feature to share their experiences of the usage of the products they purchased, that is posting online reviews. These online reviews contain important information about the products and thus, they have become an important source of information or help for customers to decide whether to buy a product or not.

According to [2], it has been reported that about 71% of global online shoppers read online reviews before purchasing a product. A product's reviews are divided into three stages namely early, majority and laggards. The reviews that are posted during the early lifetime of a product i.e., soon after the product release are called early reviews.

In [3], it is described that these early reviews have a high impact on the sales of a product. In [4], it is described that even though early reviewers contribute a small proportion of reviews, their opinions can determine the success or failure of new products and services.

Based on the information, we can infer that early reviews are important for product marketing and sales and they determine whether a product will be successful in the market or not. Hence, in this paper, we are analysing the success/failure of a product using their corresponding early reviews. This helps the enterprises decide what type of products to launch or develop that is adjusting marketing strategies and improving their profits.

The early reviews are those reviews that are posted during the early lifetime of a product, that is soon after the launch of a product into the market. In our research, we are considering only early reviews for determining the success/failure of the products.

We are using the Support Vector Machine model to do sentiment analysis of the early reviews of products.

1.1 Related Work

[1]In Characterizing and Predicting Early Reviewers for Effective Product Marketing on E-Commerce Websites by Ting Bai, Wanye Xin Zhao, Yulan He, Jian-Yun Nie, Ji-Rong Wen, those users are predicted who might post an early review of a given product.

[2]N. V. Nielsen, briefed said that it has been reported that about 71% of global online shoppers read online reviews before purchasing a product.A product's reviews are divided into three stages namely early, majority and laggards. The reviews that are posted during the early lifetime of a product i.e., soon after the product release are called early reviews.

[3]W. D. J. Salganik M J, Dodds P S in ASONAM, 2016 described that these early reviews have a high impact on subsequent product sales.

[5] E. Gilbert and K. Karahalios, it was briefed that the number of spam reviews has increasingly grown on e-commerce websites, and it was found that about 10% to 15% of reviews echoed earlier reviews and might be posted by review spammers. It is possible that spam reviews are posted to give biased or false opinions on some products so as to influence the consumers' perception of the products by directly or indirectly

inflating or damaging the product's reputation. The existence of spam reviews could lead to erroneous conclusions in our study.

Margin based embedded model can characterize both user comparison relations and the information from the product side.

2. PROPOSED METHOD

2.1 PER Method

It is a Statistics-based method in which the approach is to calculate the number of times (or the ratio) that a user has acted as an early reviewer in history data. It is based on the intuition that if a user has posted many early reviews in the past, he/she is likely to post early reviews on a new product too.

It follows the method of ranking the users based on the ratio that a user has acted as an early reviewer. PER is defined as:

$$PER(u) = NER(u) / NR(u)$$

where,

NR: Number of reviews that they have previously posted.

NER: Number of times that a user has previously acted as an early reviewer.

2.2 Early Spam Deviation Spamming model

In the current system there is limited work of spam detection models to find spam reviews in user review. In the current system by using *Early deviation spamming* model to detect the spam reviews. Here in our proposed framework we are taking consideration of this advantages and adding two more methods of spam detection model.

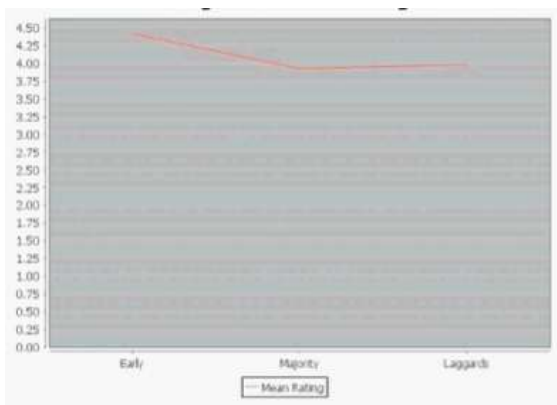- Average Content Similarity

- Rate Deviation

Fig -1: Comparisons of rating scores by early, majority, laggards

2.2.1 Average Content Similarity

First we need to check for a review if any other reviews are there for the same product by the same user then we need to check otherwise it is not spam.

   Similarity content score

> 0.4 > 0.5 => not spam.
>
> 0.7 > 0.5 => spam.

For checking similarity of text we are using Levenshtein Distance Algorithm.

2.2.2 Rate Deviation

In this method, we will calculate using the below formula.

Score = 1 - (User Rate of item - Avg Rate of item) / 4

If score > 0.5 then it is considered as not spam.

If score < 0.5 then it is considered spam.

For Example, consider user rate of item is 5 and average rate of an item is 1.5 then,

score = 1 - (5 - 1.5)/4;

score = 1 - (3.5 / 4)

score = 0.125

As the score is less than 0.5 hence, we can say false (spam).

3.2 SVM for Sentimental Analysis

Support Vector Machine model is a supervised machine learning algorithm used for both classification and regression.

It is a technique that performs on the text to determine whether the author's intentions towards a product is positive, negative or neutral.
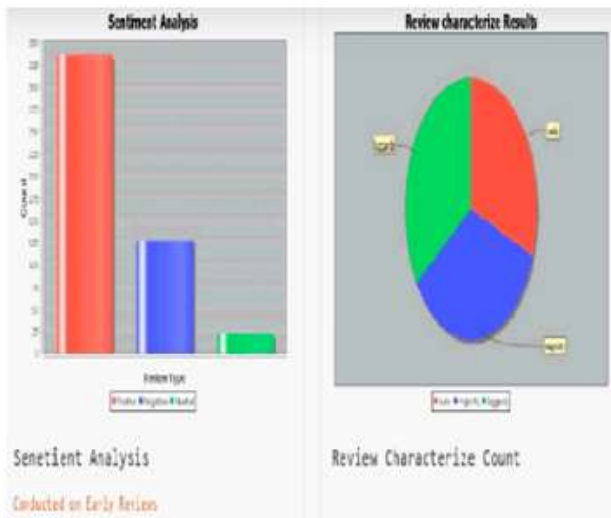
Fig -2: Sentimental Analysis conducted on early reviews

Steps needed to perform this technique is as follows:

- Gathering perfect data for training and testing

- Vectorizing the data

- Creating a linear SVM model to train and then predict.

3. CONCLUSIONS

We have built a model using machine learning algorithm, SVM to characterize the users' experience of products using the reviews data i.e., to do sentiment analysis of the reviews of the products.

While doing the analysis, we have only considered early reviews since early reviews tend to provide more accurate and useful information about the product. We have removed spam reviews from the data to avoid errors in the results. We have also found early reviewers of products using the PER method.

In future, we would like to find early reviewers for each type of product separately using doc2vec model. Also, we have just found but not used the early reviewers' data which can be further used to establish communication channels with the early reviewers.

REFERENCES

[1] Ting Bai, Wanye Xin Zhao, Yulan He, Jian-Yun Nie, Ji-Rong Wen, "Characterizing and Predicting Early Reviewers for Effective Product Marketing on E-Commerce Websites", IEEE, 2018.

[2] N. V. Nielsen, "E-commerce: Evolution or revolution in the fast- moving consumer goods World," nngroup. com, 2014.

[3] W. D. J. Salganik M J, Dodds P S, "Experimental study of in- equality and unpredictability in an artificial cultural market," in ASONAM, 2016, pp. 529–532.

[4] R. Peres, E. Muller, and V. Mahajan, "Innovation diffusion and new product growth models: A critical review and research directions," International Journal of Research in Marketing, vol. 27, no. 2, pp. 91 – 106, 2010.

[5] E. Gilbert and K. Karahalios, "Understanding deja reviewers" in CSCW, 2010, pp. 225–228.

[6] J. McAuley and A. Yang, "Addressing complex and subjective product-related queries with customer reviews," in WWW, 2016, pp. 625–635.

[7] L. A. Fourt and J. W. Woodlock, "Early prediction of market success for new grocery products." Journal of Marketing, vol. 25, no. 2, pp. 31 – 38, 1960.

[8] B. W. O, "Reference group influence on product and brand purchase decisions," Journal of Consumer Research, vol. 9, pp. 183–194, 1982.