A Novel Approach to Predict Product Rating Behaviour Analytics

[1]U. Abinaya, [2]Dr. B. Umadevi

[1]Research Scholar, [2]Assistant Professor & Head,
[1]P.G. & Research Department of Computer Science
[1]Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India.
[2]P.G. & Research Department of Computer Science
[2]Raja Doraisingam Govt. Arts College, Sivaganga, Tamil Nadu, India.

_____

Abstract

In our new modern business trend, every business organizations which makes online business needs some critics to promote their product. This critic is referred as customer rating. This may be rated by star values. The customer review is important to improve the services of companies, which have both close opinion and open opinion. The direct emotions of each customer are refereed as open opinion. However, the company has many contents or group to evaluation themselves by rating and total rating for a type of services which there are many customer who needs to review. The main issue is some customer many contradict in rating from their comments. Amazon is the world's largest online retailers where millions of customers purchase and review products on its website. However, many customers may not show their interest to rate products after the purchase. In this paper, the two products of two different companies smart phone data has been considered for analysis.

Index Terms - E-commerce, DBSCAN, Star rating, Review.

I. Introduction

E-Commerce sites are ahead in acquiring esteem across all over the world. The E-commerce provides sophisticated environment for the Customers by providing its easiness, convenience, reliability, and rapidness. There are a number of Online shopping websites that are available on the internet, such as Amazon, Flip kart, Snap deal, Jabong, Myntra, Paytm, Zovi, etc. These websites gives a multiple choices for their customers to buy products in analytical way to get the product in very less prize rather than the open market. A lot of attractive and day-to-day useful products like books, electronic goods, home appliances, clothing, and footwear are sold from these sites. These websites provide an option to the customers to write their review about their product that they buy from these sites.

These reviews and feedback from customers helps other buyers and manufacturers of the product to enhance their quality and service. The star rating and comments by the reviewers gives and extended support for the customers who are in dilemma to buy a product. Since star rating gives a position and rank for the product in the market space. Potential buyers can make decisions based on the reviews of customers who have purchased and experienced the product. The manufacturers of the product will be able to know the minor or major drawbacks of the product from the reviews which helps the manufacturers to get a chance to release the updated version of the product which satisfies the reviews that are mentioned in the websites. Hence online reviews play a significant role in understanding the customer's voice. Sentiment analysis and opinion mining through [1] machine learning algorithms offer a great possibility in automating the process of gathering, processing, and making sense of the data. By studying the reviews of customer about products helps both shoppers as well as E-commerce companies too.

In this paper, the product review rating by customers and the key data mining algorithms that have been used to solve it are discussed. The paper is structured as follows. The section one describes about introduction. Section 2 deals about background study and its related works. The methodology of the research work is explained in section three. In section four portraits the Experiment results. Finally the paper is concluded in last section.

II. BACKGROUND AND RELATED WORKS

E-commerce sites are growing [2] rapidly; similarly opinion mining is also grabbing the rapidness as well because increase in sales and profit of a company is main motive and for that they must have the knowledge about their customer buying behaviour. There are so many tools and techniques; these are used for mining large database but for mining opinions of customer about products a unique technique is necessary. The data mining techniques are commonly build models that are used to predict future data trends. There are many authors proposed data mining technique, an overview is:

2.1 Literature Review

In this paper, diversified approach using various data mining techniques are collected to analyse the rating behaviour of product at various levels. The study related to data mining for extracting and predicting the product rating analysis used in various models and the comprehensive literature review of various researchers' works are stated below:

Pratik Pede, , Ankit Damania, Gaurav Shah and Satyajeet Ghole proposed [3] to analyse a new product rating approach for mathematically and graphically analyzing sales of same type of
products from different manufactures and with most frequent combination of items.

Anurag Manni and Naman Jaiswal developed intelligent sentiment analysis that can calculate exact sentiment and give an exact rating [4] in order to get a correct review for the product. Their work mainly focused to understand the importance of sentiment analysis, language processing concepts in order to provide better intelligence to e-commerce systems.

Wararat Songpan proposed the analysis [5] and prediction rating from customer reviews who commented as open opinion using probability's classifier model. His study focused mainly of customer review's hotel in open comments for training data to classify comments as positive or negative called opinion mining. In addition to that, the classifier model have been calculated and also made a trend analysis to give the rating using Naive bayes techniques which have the classification accuracy to 94.37% compared with decision tree Techniques.

Lovenika Kushwaha and Sunil Damodar Rathod proposed for opinion mining [6] and feature extraction of product reviews. The objective is to encourage the customers and assist them in choosing the right product. It is based on natural language processing, opinion mining and AdaBoost classifier. The results from their proposed methods were highly effective and efficient in performing their tasks. They also made an attempt to improve the accuracy of our opinion polarity detection and feature extraction among other techniques.

Shyam Nair, Neeraj Pal, Prathamesh Parab, Tejas Watamwar and C.S. Lifna aimed to automate the process of gathering the reviews, analysing the sentiments in them, and presenting them in the form of informative charts and graphs and give recommendations, which can benefit both the customers as well as the businesses.

L. Jack and Y.D. Tsai presented a framework [7] for using text mining to gather customer feedback. Text mining techniques are used to aggregate the top attributes associated with groups of devices, laptops and tablets, as well as individual devices. A case study comparison of three devices compares and contrasts positive and negative aspects mentioned by the users, which is useful to improve future generations of products. Manufacturers can incorporate and review product attributes when a product is launched and over time correct product issues, understand customer requirements, and maintain customer satisfaction.

III. METHODOLOGY

The applications of various data mining [8] techniques which are adopted as a methodology to predict [9] product review rating behaviour of customer. The most predictive data models applied are artificial neural networks, Naïve Bayes, K-Nearest Neighbour (KNN), Support Vector Machine, logistic regression, classification trees, classification [10] and regression trees and discriminant analysis.

The E-commerce sites provide as many details about the products as possible on their web page. The main objectives of the studies are to make it easier for customers to make decisions from the star rating of products data provided by the e-commerce site in the following aspects:

1) To evaluate the effectiveness of different models with respect to the product features in the review mention by customer.

2) To identify the factors that influences the recommendation of the product.

The following section describes the popular models which are used to predict the product rating behaviour by customer reviews. The steps involved in Predicting the rating of product is explained in Figure 1.

Today Data mining becomes the most promising research area. Data mining supports set of techniques that can be used to extract relevant and interesting knowledge from data. Data mining techniques [11] such as association, rule mining, classification, clustering and prediction.

Clustering is a data mining technique used for making group of abstract objects into classes of similar objects are grouped into one cluster and dissimilar objects are grouped in another Cluster. Clustering is a data mining technique have its considerable amount of research in science, information technology field, medical science, image processing, document classification, clustering analysis is also used in banking industries also. Our research draws the power of clustering techniques such as K-means algorithm and DBSCAN (density based) algorithms for partioning the data. The Classification algorithm J48 (C4.5) is applied over the clustering data to predict the rating behaviour of product. The step involved in predicted the rating behaviour which is given in Figure 1. The algorithms are discussed below:

3.1 K-means Clustering

The K-means algorithm is one of the clustering algorithms whose use is widespread. K-means, which belongs to the class of partitioning algorithms, has two main advantages: it is very easy to implement and it takes little time to run, which makes it suitable for large data sets. The K-means is the most common partitioned clustering algorithm which is used to partition n observations into K clusters in which each observation belong to cluster with nearest mean. It is simple, non supervised iterative learning method. The idea behind classifying set of data objects into K number of clusters where K is fixed initially. It first fixes initial group centriods. Then assign each object to the group that has closest centroid. Once all the objects are assigned it recalculate positions of centriods. Again repeat the same process until centroids not change.

3.2 DBSCAN

DBSCAN (Density Based Spatial clustering of application with noise) is to create clusters with minimum size and density. Density is defined as the minimum number of points within a certain distance of each other. It requires two parameters: epsilon (eps) and minimum points (minPts). DBSCAN does not require you to know the number of clusters in the data a priori. DBSCAN does not have a bias towards a particular cluster shape or size. DBSCAN is resistant to noise and provides a means of filtering for noise if desired. DBSCAN does not respond well to high dimensional data. As dimensionality increases, so does the relative distance between points making it harder to perform density analysis. DBSCAN does not respond well to data sets with varying densities.
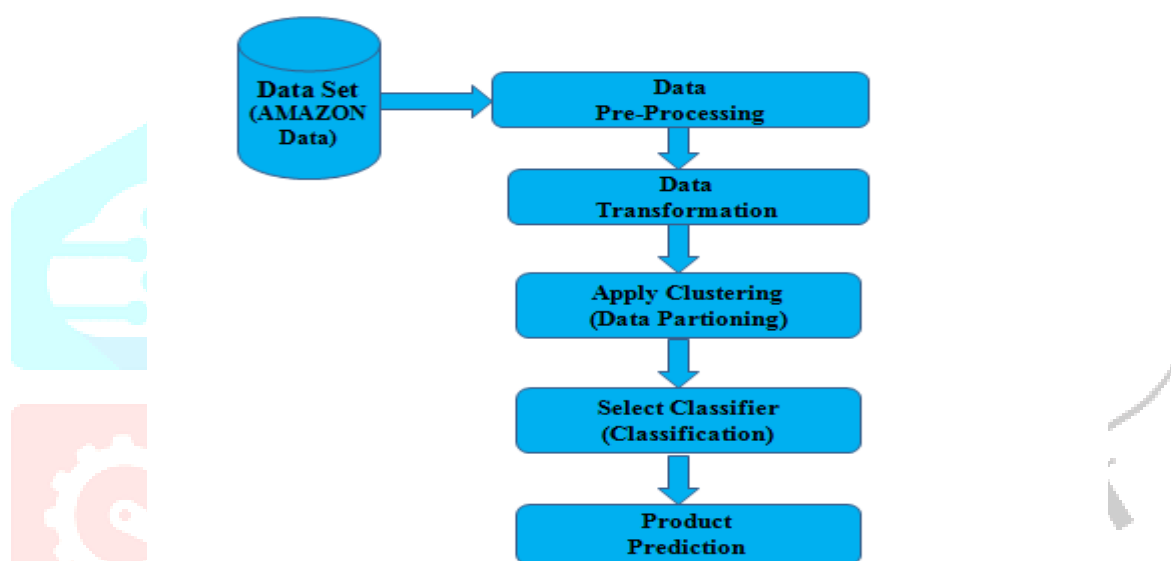

Figure 1: Steps involved in predicting the rating behaviour of Product

3.2 J48 (C 4.5)

The J48 (C4.5) technique is one of the decision tree families that can produce both decision tree and rule-sets and construct a tree for the purpose of improving prediction accuracy. The J48 classifier is among the most popular and powerful decision tree classifiers. The J48 creates an initial tree using the divide-and-conquer algorithm.

IV. EXPERIMENT RESULTS

Consumer's feedback can be collected in different manner for different products. In this research, the Amazon on line selling goods reviews is particularly a reliable source of data for capturing consumer perceptions. The research mainly focused on analysing the product review rating of consumers for Samsung and Redmi Smart phones.

The factors to be considered in the analysis is understanding what is really important to the consumers, what are the positive reviews or Critical reviews (Negative) which is affected by product reviews, and what specifically consumers ratings when reviewing the products. For that the sample data is collected from Amazon [12] which is given in Figure 2.

At the outset a sample of 20 models are only considered for analysis with different attributes. The attributes are Product-Name, Price, Customer-ID, 1-Star, 2-Star, 3-Star, 4-Star, 5-Star, Positive-Reviews, Negative-Reviews and Number-Of-Customers. The data set is pre-processed; The K-means and DBSCAN algorithms are applied over the data to partitioning the data into different (k=5) clusters. For each cluster the rating of the two products such as Samsung and Redmi is anlaysed. From the cluster data, the product review by different customer is anlaysed through the classification technique J48 (C4.5) based on the rating.

From the classification, the tree is generated which gives the rules to predict the product accurately. The outcome of the classification primarily focused only on the highest rating i.e. 5-Star rating for the clustered data. Since the 5-Star is the highest rating to give better intuition to the consumers.

The results from the classification algorithm is compared with the actual 5-Star rating of AMAZON are used to identify or predict the product which has the 5-Star rating by different customer's reviews. The actual positive and critical (negative) review rating of customer for AMAZON data is compared with the rating output of the clustering algorithm is analysed. Finally, the error accuracy by the classification algorithm J48(C 4.5) for the cluster data is also analysed. The interpretations are given in Figures 3 to 11.

Figure 2 : Sample Data Collected from AMAZON.
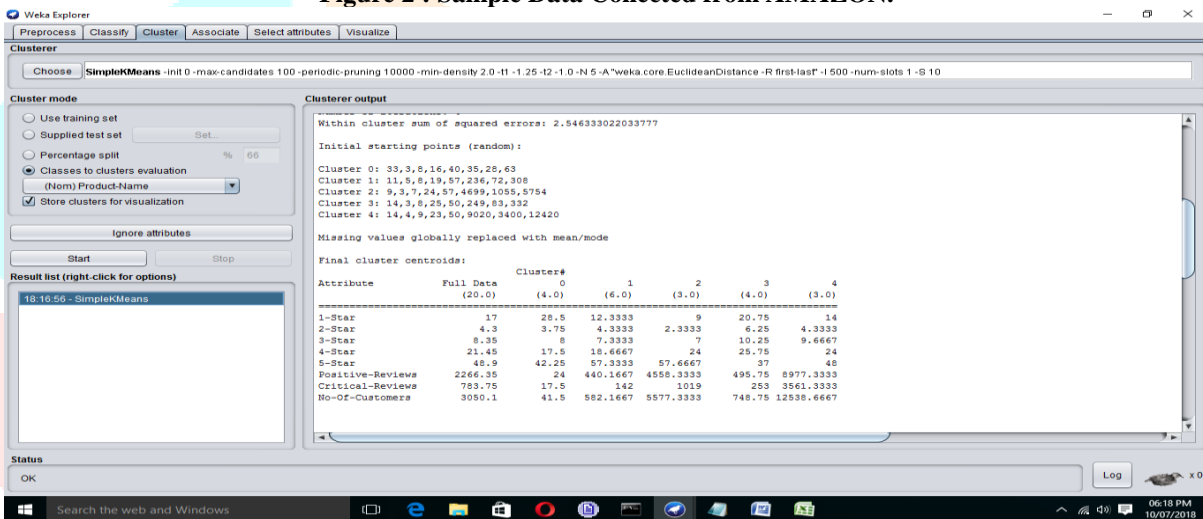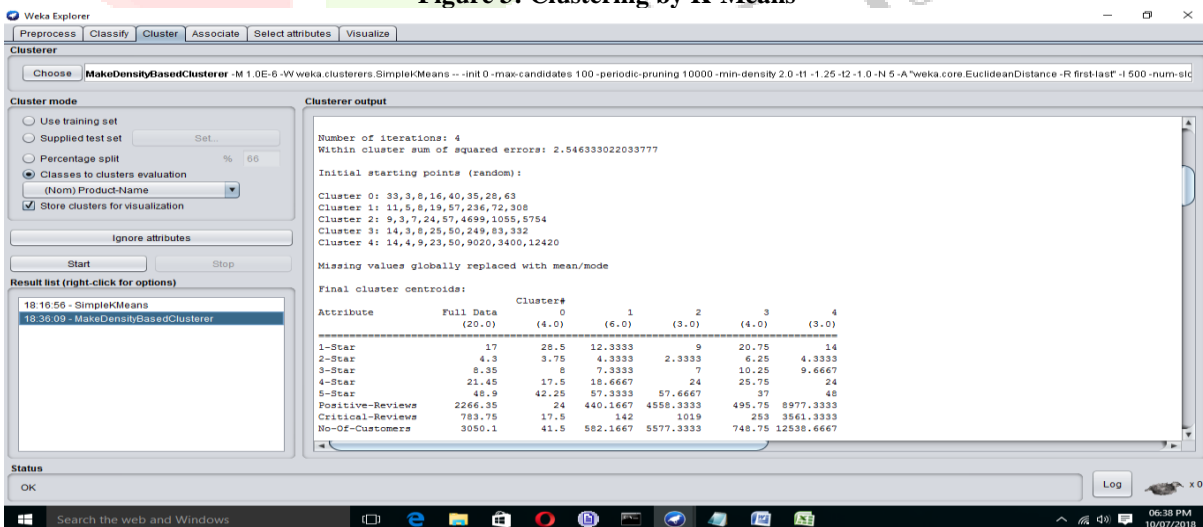

Figure 3: Clustering by K-Means

Figure 4: Clustering by DBSCAN

Table 1: Clustering by K-Means

Cluster-Name	1-Star	2-Star	3-Star	4-Star	5-Star	Positive-Reviews	Critical-Reviews	No-Of-Customers	Product-Name
Cluster-0	28.5	3.75	8	17.5	42.25	24	17.5	41.5	Samsung-Galaxy -J2-2017- (Black&8GB)

Cluster-1	12.3333	4.3333	7.3333	18.667	57.3333	440.1667	142	582.1667	Samsung-Galaxy-J7Nxt-(Gold&16GB)
Cluster-2	9	2.3333	7	24	57.6667	4558.3333	1019	5577.3333	Redmi-5-(Black&32GB)
Cluster-3	20.75	6.25	10.25	25.75	37	495.75	253	748.75	Samsung-Galaxy-J2-Pro-(Silver&16GB)
Cluster-4	14	4.3333	9.6667	24	48	8977.3333	3561.333	12538.6667	Samsung-On7-Pro-(Black)

Table 2 : Clustering by DBSCAN

| Cluster-Name	1-Star	2-Star	3-Star	4-Star	5-Star	Positive-Reviews	Critical-Reviews	No-Of-Customers	Product-Name
Cluster-0	28.5	3.75	8	17.5	42.25	24	17.5	41.5	Samsung-Galaxy-J2-2017-(Black&8GB)
Cluster-1	12.3333	4.3333	7.3333	18.667	57.3333	440.1667	142	582.1667	Samsung-Galaxy-J7Nxt-(Gold&16GB)
Cluster-2	9	2.3333	7	24	57.6667	4558.3333	1019	5577.3333	Redmi-5-(Black&32GB)
Cluster-3	20.75	6.25	10.25	25.75	37	495.75	253	748.75	Samsung-Galaxy-J2-Pro-(Silver&16GB)
Cluster-4	14	4.3333	9.6667	24	48	8977.3333	3561.3333	12538.6667	Samsung-On7-Pro-(Black)
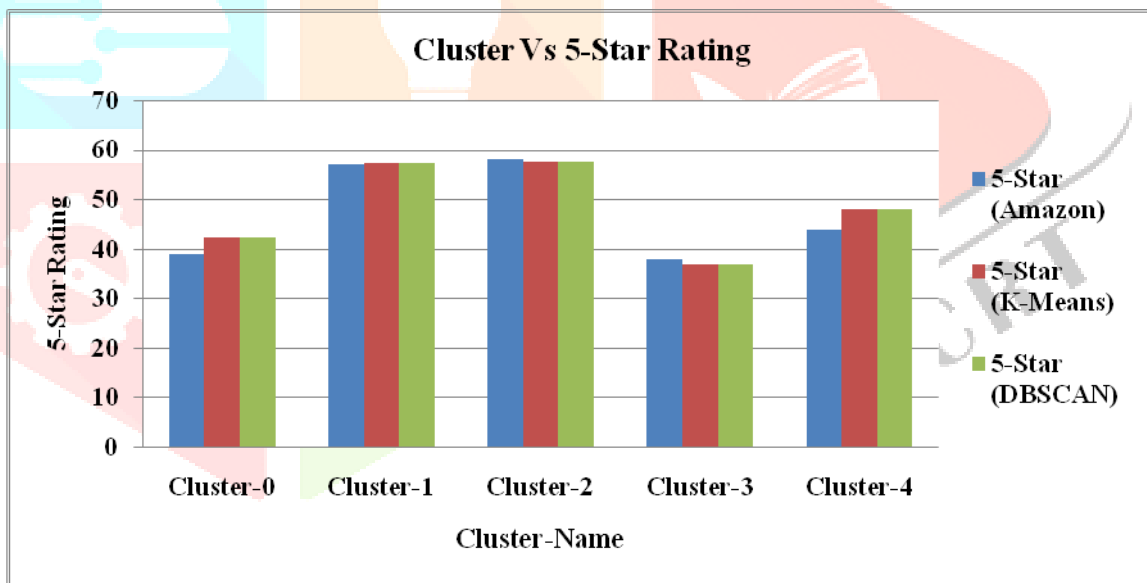
Figure 5 : Cluster Vs 5-Star Rating by Amazon, K-Means and DBSCAN.
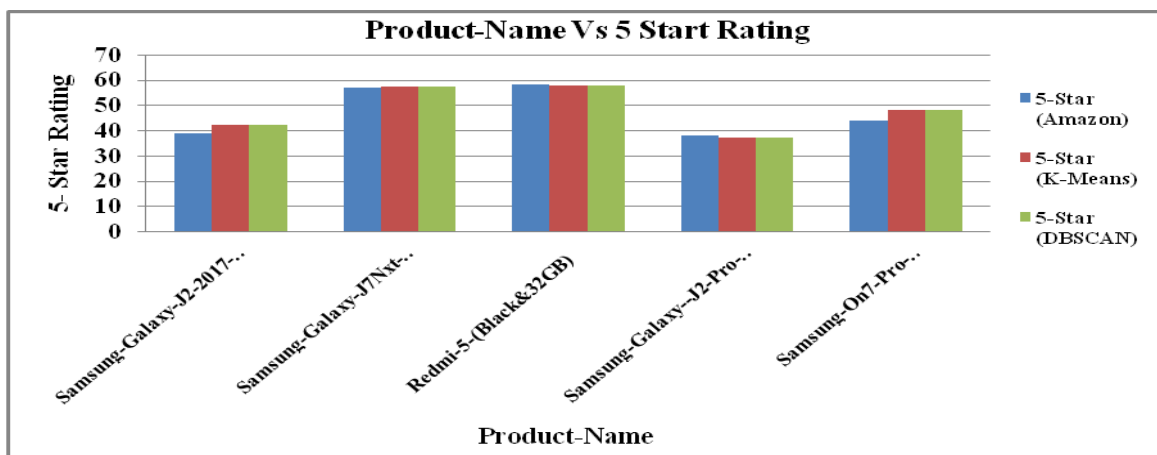

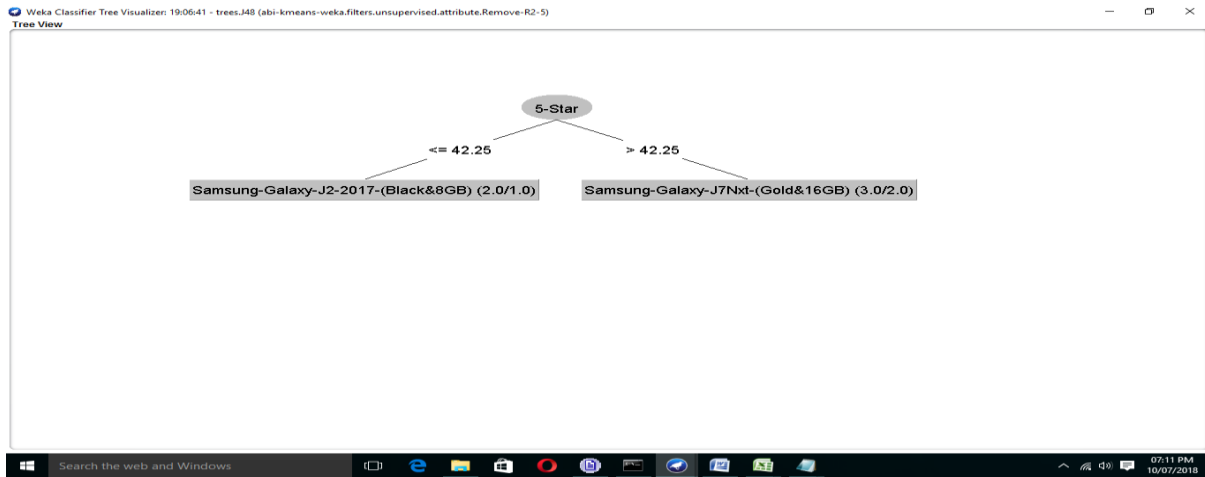**Figure 6 : Product-Name Vs 5-Star Rating by Amazon, K-Means and DBSCAN**

Figure 7 : Predicting Product from J48 for K-means clustering data
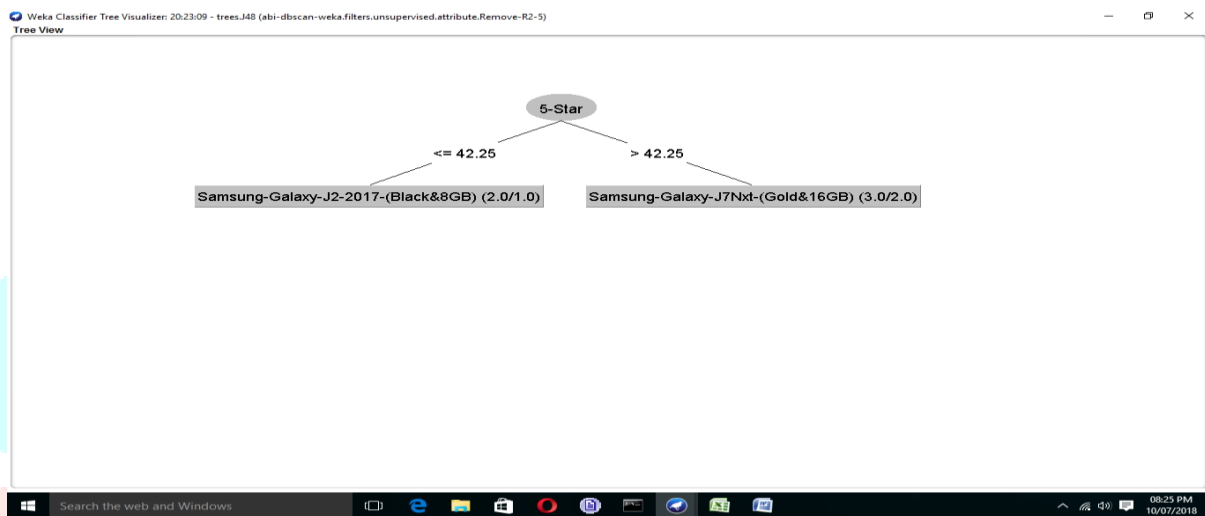

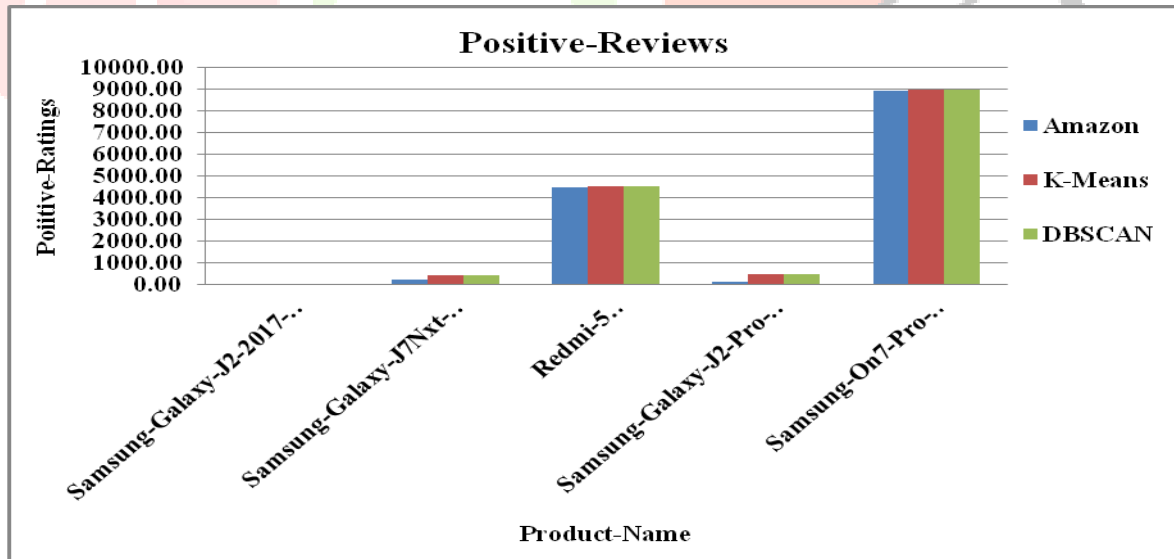**Figure 8 : Predicting Product from J48 for DBSCAN clustering data**

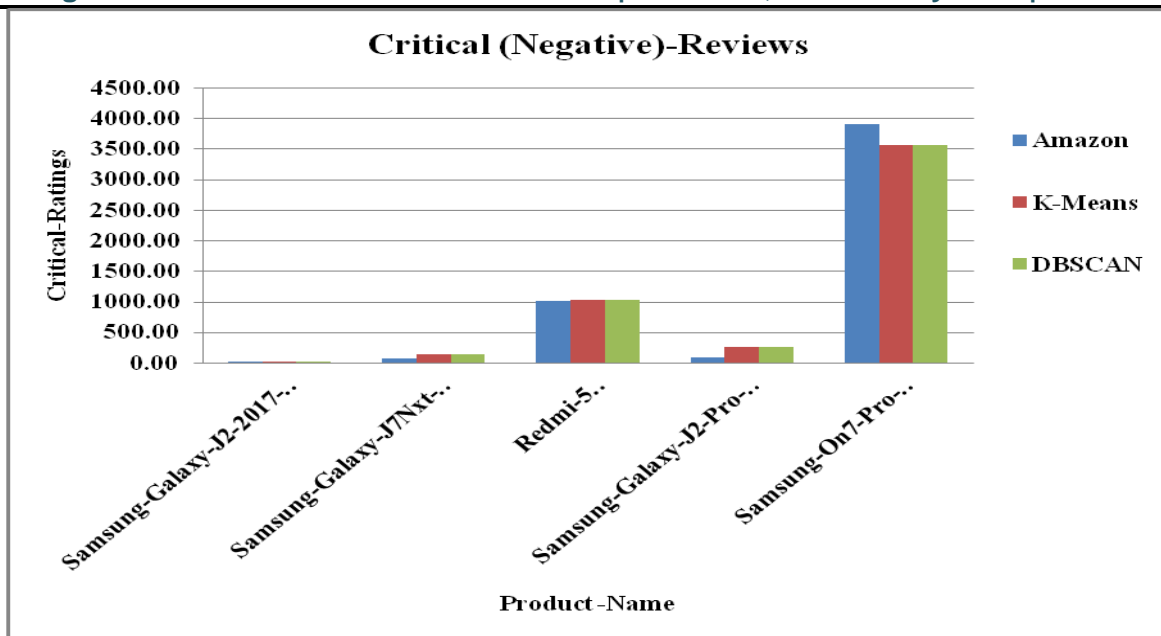**Figure 9 : Comparison of Positive Reviews of AMAZON with Clustering**

Figure 10 : Comparison of Critical (Negative) Reviews of AMAZON with Clustering

Table 3 : Error Accuracy by J48 for cluster-data

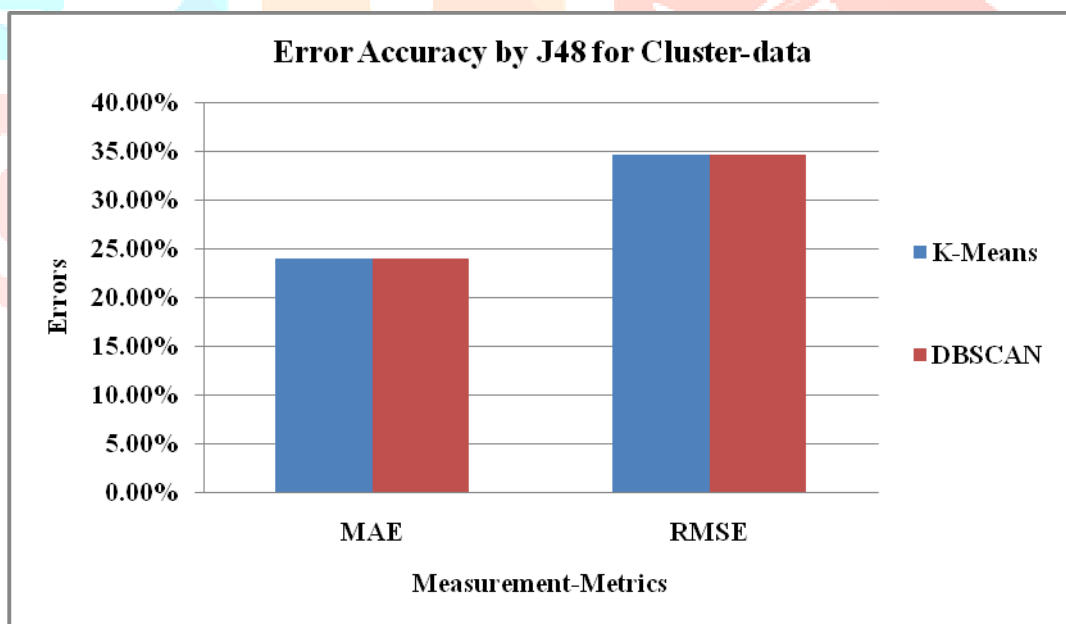| Cluster -Algorithms	MAE	RMSE
K-Means	0.24	0.3464
DBSCAN	0.24	0.3464

Figure 11 : Error Accuracy by J48 for Cluster data

V. CONCLUSION

The online business promoters are nowadays increased many fold on par with the technological revolution. Since people have an open choice to see and select different products by sitting in their home itself. In this online marketing the Amazon is becoming the most prominent organization in the world. The opinion poll is more important for Amazon from customer's perspective. Data for Samsung galaxy, Redmi are collected and analysed. The star rating values with positive and negative feedbacks are collected and applied through algorithms K-means and DBSCAN to organize into various groups. The outcome of the clustering is applied through the J48 (C4.5) classification algorithm. From the classification algorithm, the most liking Smartphone model is identified or classified based on the 5 Star rating behaviour.

REFERENCES

[1]Ali Hasan , Sana Moin , Ahmad Karim and Shahaboddin Shamshirband, 2018, "Machine Learning-Based Sentiment Analysis for Twitter Accounts, Mathematical and Computational Applications volume: 23 Issue: 01.

[2] Lovenika Kushwaha and Sunil Damodar Rathod, June-2016. "Opinion Mining of Customer Reviews based on their Score using Machine Learning Techniques", International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 06.

[3] Zainab Pirani, Zainab Bhavnagarwala, Anuja Marewar and Madhuri Kamble, March 2017, "Analysis and optimization of online sales of products".

[4] Anurag Manni and Naman Jaiswal, "Product Rating Based On Review Using Data Mining",
International Journal of Advance Research, Ideas and Innovations in Technology, ISSN: 2454-132X Impact factor: 4.295 (Volume3, Issue3).

[5] Wararat Songpan, 2017, "The Analysis and Prediction ofCustomer Review Rating Using Opinion Mining", IEEE SERA 2017, June 7-9, London, UK.

[6] Miss Lovenika Kushwaha and Prof. Sunil Damodar Rathod, "New opinion mining technique for online product reviews and features", Volume 2, Issue 4.

[7] Yi-Fang Tsai and Lleyana Jack, July 2015, "Using Text Mining of Amazon Reviews to Explore User-Defined Product Highlights and Issues".

[8] Umadevi B. and Sundar D, Alli Dr P., January 2013, "An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50", *International Journal of Data Mining & Knowledge Management Process (IJDKP),*Vol.3.

[9] Umadevi B, Sundar D., Alli Dr.P., 10th June 2014, "Novel Framework For The Portfolio Determination Using PSO Adopted Clustering Technique", *Journal of Theoretical and Applied Information Technology* , Vol. 64- No.1.

[10] Umadevi B.and Sundar D., Alli Dr.P., 2013, "An Optimized Approach to Predict the Stock Market Behavior and Investment Decision Making using Benchmark Algorithms for Naive Investors",*Computational Intelligence and Computing Research (ICCIC), IEEE International Conference*, 26-28 Dec. 2013,Page(s):1 – 5978-1-4799-1594-1INSPEC Accession Number: 14061140.

[11] Abinaya U and Umadevi Dr. B., "A Review on Product Rating Behaviour Analytics",   International Journal of Computer Science and Mobile Applications, Vol.6 Issue. 6, June- 2018, pg. 10-15, ISSN: 2321-8363.

[12] http://jmcauley.ucsd.edu/data/amazon.

Authors' Biography

U. ABINAYA is an M.Phil Research Scholar in PG & Research Department Of Computer Science, Raja Doraisingam Government Arts College, Sivaganga, Tamilnadu, India. Her research concentration include in Data mining, Machine Learning and its applications.

Dr. B. UMADEVI has received her Doctoral degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, India. Currently working as Assistant Professor & Head- P.G and Research Department of Computer Science, Raja Doraisingam Government Arts College, Sivangai-Tamilnadu, India. She has over 23 years of Teaching Experience and published her research papers in various International, National Journals and Conferences. Her research interests include Data Mining, Soft Computing and Evolutionary Computing. She got the Best Paper Award for her publication in the IEEE International Conference on Computational Intelligence and Computing Research held on 27th Dec 2013 at   VICKRAM College of Engineering and Technology.