

# BDA For Commercial Furtherance By Using Elevate Performance Algorithms

M.Arumugam, S.Deepa, Dr.C.Jayanthi

**Abstract :** Generally, any business organization aims to make a profit, in doing so, the organization manages many efforts to make a profit. One of the most important of those efforts is to attract customers and augment the number of customers. Therefore, the business wants to gather customers' perceptions of their organization, and then they improve their organization with that helps of concept. But due to the growth of social networks, they have had customers post their comments online, but, although the comments of the customers are to keep growing, it is very sluggish to have too much time to read them all. This functionality is called big data analysis, but analyzing large data or hidden images is a difficult task, therefore, a high performance mechanism is required to perform these processes. Two steps are performed in this paper to analyze big data or consumers' comments, one of them is the separation of data, because the data can be in any form, and some data may be good, some data may be wrong. Business organizations can improve their business with good data, it is hoped that this data separation method will be more effective to remove some harmful data. The second method is an important method of this effort, because classifying data is the second method, this paper, therefore, uses a very high-performance method for this process. It is therefore believed that this new initiative will be very useful and reliable for business enterprises.

**Keywords:** Big Data Analysis (BDA), Customers' Comments, Separation of Data, Classification of Data.

## 1. INTRODUCTION

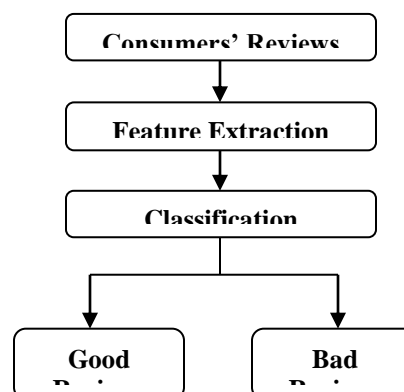
Businesses most rely on the analysis of their data to predict consumers' response and recommend their products to customers. Therefore although analyzing such a large data is a difficult task but it must be calculated very accurately. Although there are many ways to analyze these, it is important to select the most appropriate method and run them correctly. We are known that for a business to grow further, it is essential to get customer feedback. A number of companies are using a variety of methods for that. With the growth of business during this period, they are competing to make customers happy with other business. Customers are very eager to share ideas about their purchased products. They are increasingly using social media to share because of the growth in social media. And because it's so easy to share in that method, customers are increasingly sharing their opinion. This makes it very difficult and time consuming for business companies to read customer data. Without it, the lack of time could not read the opinion of some customers. Businesses are looking for better solutions to solve such problems. So first of all, we need to collect the relevant data of customers from social media and then analyze them. Thus, you can grow the company by reading the customers' feedback and doing the right service. In other papers, this paper will see polls about this. In [1], some discounts and concessions are made, with the intention of appealing to the minds of consumers. After being so attracted to them they fulfill their desire without giving up and make them happy. To make it easier for them to read their comments, Tata has

done with the Meaning system, but there are some mistakes in it. In [2], the DBSCAN algorithm was used to divide the data into clusters. But their area is large and difficult to design.

In [4], predictive models have been used to classify consumers' reviews. However, this may not be effective for large companies. Paper [4] based on business companies; the DMT algorithm was used to classify users' data. But that is not possible. In Paper [5], machine learning was used to analyze the data of electricity consumers. But their classification was wrong at times. In [6], The SVM algorithm was used here to track customers' behavior. But this does not go hand in hand with some applications. In [7], the researcher used a method to analyze the behavior of filters but it consumes more energy. So this proposed method is used here to solve such problems. We will see it in detail in the following section.

## 2. PROPOSED METHODOLOGY

In this proposed mechanism, two methods are utilized, one is feature extraction and the other is classification. So first, for feature extraction, consumers' need to input feedback into the feature extraction process, then once that algorithm has extracted the features of the concept; its output should be input to the classifier. Finally, the publication of the classifier is classified as good reviews and other bad reviews.



- Assistant Professor, Department of Computer Technology-PG, Kongu Engineering College, Perundurai, Erode-638060, Tamilnadu, India. E-Mail:maructpg@gmail.com, Mobile: +9842778152.
- Assistant Professor, Department of Computer Technology-PG, Kongu Engineering College, Perundurai, Erode-638060, Tamilnadu, India. E-Mail:sdeepakec@gmail.com, Mobile: +9600421666.
- Assistant Professor, Department of Computer Science, Government Arts College, Karur-639005, Tamilnadu, India. E-Mail:jayanthibabu117@gmail.com, Mobile: +9659862117.

**Fig.1: Flow of Proposed Method**

**1. Feature Extraction**

Here the PCA algorithm is used to extract the opinions of consumers. This is because the PCA algorithm minimizes the dimensions of the data and extracts the features more precisely. That is, consumers will give good feedback or give bad feedback or give mix both feedbacks. Reviews are give in more so, that will very difficult to read and understand. So this PCA is used to separate them. This method divides only two types of good and bad of reviews. That is, if customers post both as a mix, it divides that mix reviews into two categories - good and bad. Example reviews:

- 1) It is very good.
- 2) It is very good but the cloth is very poor quality.
- 3) It is very nice.
- 4) It is very bad.
- 5) This was very useful for me.
- 6) The room is quite dirty, but the room design is very amazing.

**Fig.2: Consumers' Comments**

In the example above, a few are good reviews and a few with bad reviews and a few with mixed both. This PCA algorithm takes the concepts that are both mixed, and then divide separately, which are given below. That means, So with this method, categorizing them to classifier is very simple. The problem is solving in mathematically by algorithm is given as follows.

Steps for PCA mechanism:

STEP 1: Normalizing Consumer's Data

In this step, the reason for normalizing the feedback is that the PCA algorithm will works perfectly. If seeing how this is done, the respective means are subtracted from the numbers in their respective columns. In this case, if there are good and bad comments, all the good comments become 'G' and all the bad comments become 'B'.

Here,

G = Good reviews

B = Bad reviews

So this creates the Data Set and its Mean is '0'.

STEP 2: Covariance Calculation

Since this data set is two-dimensional, it creates a 2x2 covariance matrix.

$$\text{Matrix (Covariance)} = \begin{bmatrix} \text{Var [G1]} & \text{Cov [G1, G2]} \\ \text{Cov [G2, G1]} & \text{Var [G2]} \end{bmatrix}$$

That means,

$$\text{Var [G1]} = \text{Cov [G1, G1]}$$

$$\text{Var [G2]} = \text{Cov [G2, G2]}$$

STEP 3: Calculate the Eigen values And Eigenvectors This

S.No	Original Reviews	Good Reviews	Bad reviews
1	It is very good.	It is very good.	
2	It is very good	It is very good	
3	It is very nice.	It is very nice.	
4	It is very bad.		It is very bad.
5	This was very useful for me.	This was very useful for me.	
6	The room is quite dirty		The room is quite dirty
7	the cloth is very poor quality.		the cloth is very poor quality
8	the room design is very amazing.	the room design is very amazing.	

**Fig.3: Extracted Comments**

Eigen value and the eigenvectors are calculated for the covariance matrix and since this is a square matrix this can be calculated.  $\lambda$  is the matrix A's Eigen value here. The equation is:

$$\text{Det} [\lambda I - A] = 0$$

'I' is the identity matrix in dimensions. This is a necessary condition for the matrix to subtract. 'Det' is used to determine the team. These Eigen values can be found by solving the corresponding eigenvectors 'v' here.

$$[\lambda I - A] * v = 0$$

STEP 4: Creating feature vectors by selecting components

First, we need to sort these Eigen values from big numbers to small numbers. In doing so, the elements are placed in importance. This is where dimensional reduction takes place. There are 'n' Eigen values and 'n' eigenvectors associated with a data set of 'n' variables. The key component of the dataset is the Eigen vectors associated with the highest values of the Eigen values. And the analysis of this dataset depends on how many of these Eigen values are chosen. To reduce the dimensions, only the first 'p' eigenvalues are selected and the others rejected. As a result, a lot of information is lost. Very little information is lost if the eigenvalues are small. Since in this is a two-dimensional, the two values that are most similar to this eigenvalue can be taken.

$$\text{Feature Vector} = (\lambda_1, \lambda_2)$$

STEP 5: Generating Principal Components

In this final step, the main components are created using all the math of hitherto done. To create it, take the swap of the feature vector, and it is multiplied to the left with transfer for the measured version of the original data set.

$$\text{NewData} = \text{FeatureVector}^T * \text{ScaledData}^T$$

That means,

NewData = a matrix with key elements.

FeatureVector = a matrix created using the selected eigenvectors.

ScaledData = the measured edition of the original dataset.

**1. Classification**

After extracting consumers 'opinions as good and bad, good comments are categorized as one class and all bad comments are classified as one class. To this classification, the most efficient SVM classifier is used. The following table shows how to characterize these comments by SVM method.

Original Reviews	Good	Bad
It is very good but the cloth is very poor quality.	It is very good	the cloth is very poor quality
The room is quite dirty, but the room design is very amazing.	the room design is very amazing.	The room is quite dirty

Fig.4: Classified Comments

That is to say, in the above table, if you give the extracted comments and other comments to the SVM classifier, it analyzes the feedback and categorizes the good comments into one category and the bad comments into another category. In this classification, the mathematical functions can be found by the following SVM classifier. Generally this classifier will find the hyperplane first it to classify a data more accurately. Steps to SVM mechanism STEP 1: Here are three hyperplanes such as A, B and C. Now we need to find the most appropriate hyperplane to categorize the good comments and the bad comments.

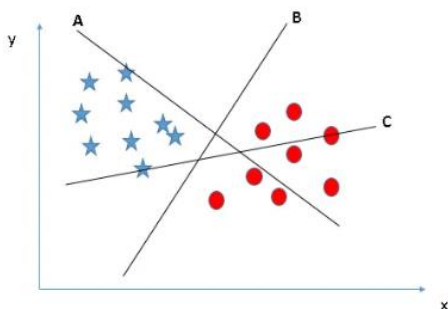


Fig.5: Accurate Classification 1

In the fifth of the figure, the stars are taken as good comments and the circles are all bad comments. Here in figure 5, the hyperplane of B is most correctly classified. STEP 2: The three hyperplanes in Fig. 6 have very well classified the two types. So in this situation, find the most suitable hyperplane.

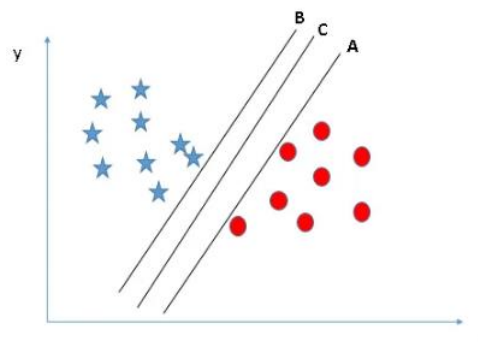


Fig.6: Accurate Classification 2

As shown in the figure 7, the distance between adjacent comments and hyperplanes should be increased, because then you can find the most perfect hyperplane, and that distance is said to be the margin.

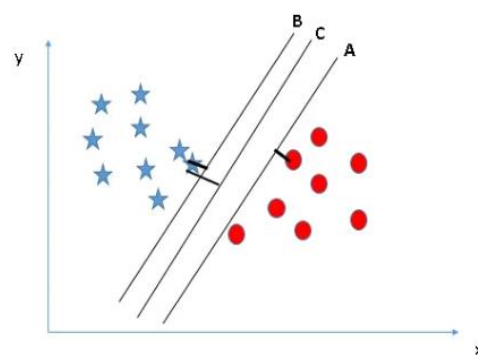


Fig.7: Accurate Classification 3

In Figure 7, the hyperplane of C has greater distance compared to the three hyperplanes. Hence the hyperplane C is called the most correct hyperplane. Highly margin hyperplanes are the ones that divide the data more precisely, if it selects a low-margin hyperplane, which will misclassify it. So this method is more accurately categorized the consumers' opinions because the SVM is chosen a very large margin of hyperplane.

**3. RESULTS AND DISCUSSION**

In this section, this paper compares the methods used here, with other methods. The details of this can be clearly seen by the following figures.

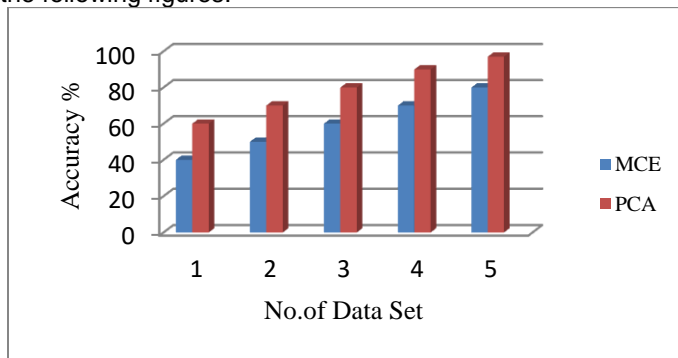
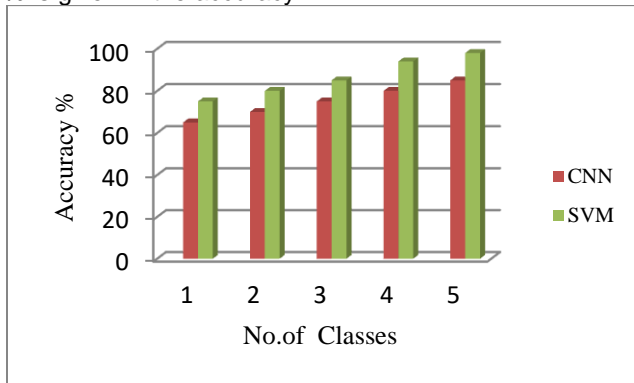


Fig.8: PCA Comparison

In Figure 8, the PCA algorithm is compared to the MCE algorithm. This is because both are methods of classifying

features. In doing so, the PCA algorithm is able to more efficiently and accurately extracted consumers' reviews than MCE algorithm. In this, the MCE 80% and the PCA algorithm 90% is given in the accuracy.



**Fig.9: SVM Comparison**

Similarly the methods used for classification are also compared. In comparison, the SVM system is more efficient than the CNN. This is elaborated on in figure 9. So for such problems, SVM has proved to be very good. Here, the SVM is given as 95% accuracy and the 85% accuracy.

#### 4. CONCLUSION

This paper wants to conclude is that; this paper is made with a focus on the concept of business growth. Business growth depends on customer feedback, so this paper categorized the customer's opinion extract and then categorized them. This paper has successfully completed both of those actions. This is because the two methods used for these actions have done their job precisely. So this new system will be very useful for business companies to grow further, because there are millions of people in this world. It is unlikely that the owners of those companies will read the comments that all send to the establishments by consumers. So this proposed method solves that problem.

#### REFERENCES

- [1] M. R. Zakiullah and M. S. Islam, "Design of a low cost and convenient Hadoop application for extracting consumer behavior", 20th International Conference of Computer and Information Technology (ICCIT) - 08 February 2018.
- [2] Liping Zhang, Song Deng, Shiyue Li, "Analysis of power consumer behavior based on the complementation of K-means and DBSCAN", IEEE Conference on Energy Internet and Energy System Integration (EI2) - 04 January 2018.
- [3] Li Haiyan ; Zhang Hui ; Zhang FangHong ; Guo Meijing, "An analysis of consumers' cognition and reading time prediction in digital reading", IEEE/ACIS 16th International Conference on Computer and Information Science - 29 June 2017.
- [4] Edwin Camilleri ; Shah Miah, "A Consumer Analytics Framework for Enabling Data-Driven Marketing Intervention Methods", 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE) - 11 October 2018.
- [5] Yingchen Zhang, Rui Yang, Kaiqing Zhang, Huaiguang Jiang, Jun Jason Zhang, "Consumption Behavior Analytics-Aided Energy Forecasting and

Dispatch", IEEE Intelligent Systems ( Volume: 32 , Issue: 4 , 2017 ) - 17 August 2017.

- [6] K. Maheswari ; P. Packia Amutha Priya, "Predicting customer behavior in online shopping using SVM classifier", IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) - 01 March 2018.
- [7] Hafize Fidan ; Atanaska Teneva ; Stanko Stankov ; Eva Dimitrova, "Consumers' Behavior of Restaurant Selection", International Conference on High Technology for Sustainable Development (HiTech) - 10 December 2018.
- [8] Vijayabanu ; R. Gayathri ; S. SaiPrashanthi ; S. Harshitha Paarvathi, "Consumer Perception towards Consumer Ethnocentrism", International Conference on Current Trends towards Converging Technologies (ICCTCT) - 29 November 2018.
- [9] Artem Priemyshev ; Vyacheslav Zhukov ; Mikhail Komarov, "The Behavior of Consumers in the Shop Based on the Data from "Locastor" System", IEEE 20th Conference on Business Informatics (CBI) - 03 September 2018.
- [10] Muqaddas Gull ; Arshi Pervaiz, "Customer Behavior Analysis Towards Online Shopping using Data Mining", 5th International Multi-Topic ICT Conference (IMTIC) - 20 September 2018.