



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Customer Segmentation Using K-Means Algorithm

S.Venkatasubramanian¹, Sriram.S², Surya Prakash.S³, Veluganes.V.S⁴, Vignesh.S⁵

Associate Professor, Dept. of C.S.E, Saranathan College of Engineering, Trichy, India¹

U.G Students Dept. of C.S.E, Saranathan College of Engineering, Trichy, India^{2,3,4,5}

ABSTRACT: Customer segmentation plays a key role in making business decisions. In the competitive field, it is very important to satisfy the customer needs and to identify the potential customer and these things should be done in the right time in the right manner. This is where machine learning comes into play, various algorithms are applied for unraveling the hidden patterns in the data for better decision-making in the future. This eludes the concept of which segment to target is made unequivocal by applying segmentation. The process of segmenting the customers with similar behaviors into the same segment and with different patterns into different segments is called “Customer Segmentation”. In this paper, the K-Means algorithm has been implemented to segment the customers and finally get the result of clusters obtained from the algorithm. A python program has been developed and the program is been trained by applying a standard scaler onto a data set having around 25,000 data from an E-commerce website.

KEYWORDS: Data analytics; segmentation; k-means; E-commerce; clustering;

I. INTRODUCTION

The concept of a collection is based on an analysis of the market and the relationships within it, or on the knowledge of a selected segment of customers.

Customer Segmentation is an unsupervised method of targeting the customers in order to increase sales and market goods in a better way. It is an important factor in developing a collection, seeing that the competition in the global market is quite high. For this reason, businesses need to be more concerned about meeting customer needs, there being no average customer and no two people who are exactly alike. Having a desirable position in highly competitive markets requires making effort to produce a collection, which will meet the customer’s needs customer satisfy them. It is necessary to know which type of approach is most effective for a particular target group and to identify the needs, wishes, and expectations of a particular customer segment in order to match products and communication with these factors. The result is higher customer satisfaction, as well as the identification of strategic opportunities.

This segmentation can directly or indirectly influence the marketing strategy as it opens many new paths to discover for which segment of product will be good, on what day most sale happened, providing discounts for a specific segment, etc. Customer segmentation allows companies to visualize what actually the customers are buying which will prompt the companies to better serve their customers resulting in customer satisfaction, it also allows the companies to find who their target customers are and improvise their marketing tactics to generate more revenues from them.

II. RELATED WORK

[1] Customer segmentation using K-Means algorithm In this paper, the clustering algorithm used is K-means algorithm which is the partitioning algorithm, to segment the customers according to the similar characteristics. To determine the optimal clusters, elbow method is used. [2] Review on customer technique on ecommerce This paper will review customer segmentation using data, methods and process from customer segmentation research. Data can be processed using one of several methods: Business Rule, Magneto, Customer Profiling, Quantile Membership, RFM Cell Classification Grouping, Supervised Clustering, Customer Likeness Clustering, Purchase Affinity Clustering and Unsupervised Clustering. In this paper, those methods were classified into Unsupervised technique and the process was generalized in determining the business objective, collecting data, data preparation, variable analysis, data processing, and performance evaluation. Customer behaviour in accessing e commerce when viewing a product on e commerce was recorded in server log with time. Duration when seeing the product can be used as customer interest in the product so that it can be used as a variable in customer segmentation. [3] Customer segmentation using clustering and data mining This research paper is a comprehensive report of k-means clustering technique and SPSS Tool to develop a real time

and online system for a particular super market to predict sales in various annual seasonal cycles. [4] A Two phase clustering method for intelligent customers This paper consists of reviews the previous studies related to customer segmentation. This part presents the limitations of existing studies and explains the background reasons of this study. Proposes a clustering model for customer segmentation. Consequently, the results of applying the model in the case of banking industry will be presented. Finally, Concludes the paper with some general discussions and an agenda for further research. [5] Customer segmentation using machine learning In this paper, different clustering approach has been presented in order to segment the customer and apply the different marketing strategies accordingly. The possibility of hybrid combination of clustering algorithm can outperform individual model has also been discussed.[6] Datamining on customer segmentation: A review Data mining is used to extract important information from the bulk of data to save it and summarize it in effective manner. The hidden information can be extracted from the large set of data. The goal of this paper is to investigate the methods that are used for efficient grouping of data. The grouping must be done in its manner that the group can recognize the group members and group can also recognize still, not grouped member so far. Different approaches for customer segmentation in data mining are: clustering and subgroup discovery. Because of some limitations and scope of the clustering techniques, it leads to further refinements in methodology in data mining.

III. PROPOSED ALGORITHM

A. Design Considerations:

- Developing K-Means model for clustering
- Plotting center point smartly instead of choosing some random point
- K-Number of clusters is chosen not randomly instead we use elbow method.
- All data points should be processed and cleansed from redundant value

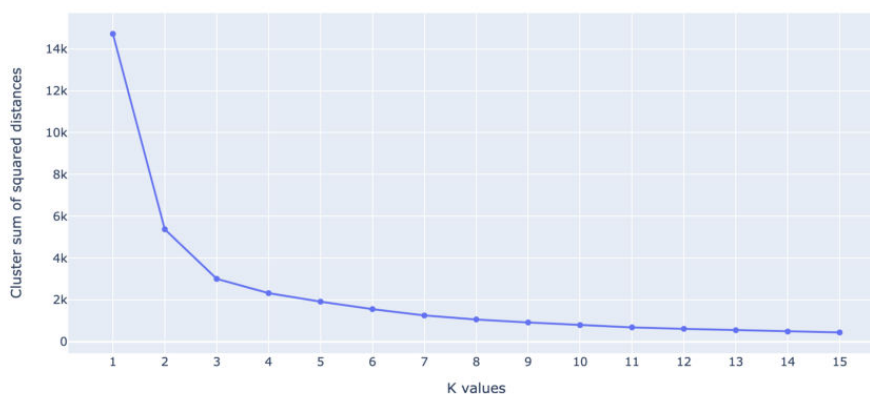
B. Description of the Proposed Algorithm:

Aim of the proposed algorithm is to cluster the customers based on their behaviour in more efficient way. The proposed algorithm is consists of two main steps.

Step 1: Calculating K value:

We are going to choose k which is the most important hyperparameter of K-means. First, we select k value against the optimization criteria of the K-means, inertia, using elbow method. We are going to build different K-means models with k values 1 to 15, and save the corresponding inertia values. When we plot inertia against k value,

Within Cluster Sum of Squared Distances VS K Values



With the elbow method, we are going to select the k value where the decrease in the inertia stabilizes. 36 When k=1 inertia is at the highest, meaning data is not grouped yet. Inertia decreases steeply until k=2. Between k=2 and 4, the curve continues to decrease fast. At k=3, the descent stabilizes and continues linearly afterwards, forming an elbow at k=3. This points out the optimal number of customer group is 3.

Step 2: Updating the K Value in our K-Means model

Let's plug in the k=3 to K-means and visualize how customer groups are created. The K-Means model will cluster the customers based on the features we calculated since the value of k is 3 we get three clustered groups of customers. Each customer group is indicated in different colour. After that, based on the segmentation we store those customer information in three separate CSV files and provide that as an output to the user. So initially we had a raw dataset consist of redundant data of customers, then after pre-processing and analysis we have segmented the information about the customers as, favourable (with high spending and low return rate), Moderate (with moderate spending and average return rate) and least favourable (with low spending and high return rate). By using this information user can provide offers and coupons to certain customers for better benefit.

eq. (3)

IV. PSEUDO CODE

Step 1: randomly choose k examples as initial centroids.

Step 2: while true:

```

Create k clusters by assigning each
    Example to closest centroid
Compute k new centroid by averaging
    Examples In each cluster
If centroid don't change:
    break
    
```

Step 3: Plot those inertia values in elbow method and find the elbow point

Step 4: Update that point k value to k-means model and perform clustering

Step 5: Visualize the output clustering

Step 6: Store those grouped customers in separate file

Step 7: End.

V. SIMULATION RESULTS

This simulation involves three different colour of nodes which are red, blue and green where red is for the customers who have been most favourable with high spending and low return rate. Blue is for the customers who have been made not that much of a frequent purchase by giving offers and coupons company can able to change these customers as favourable customers. And at last, there is a group of customers mentioned as green node which is for the customers who have been made least purchase.

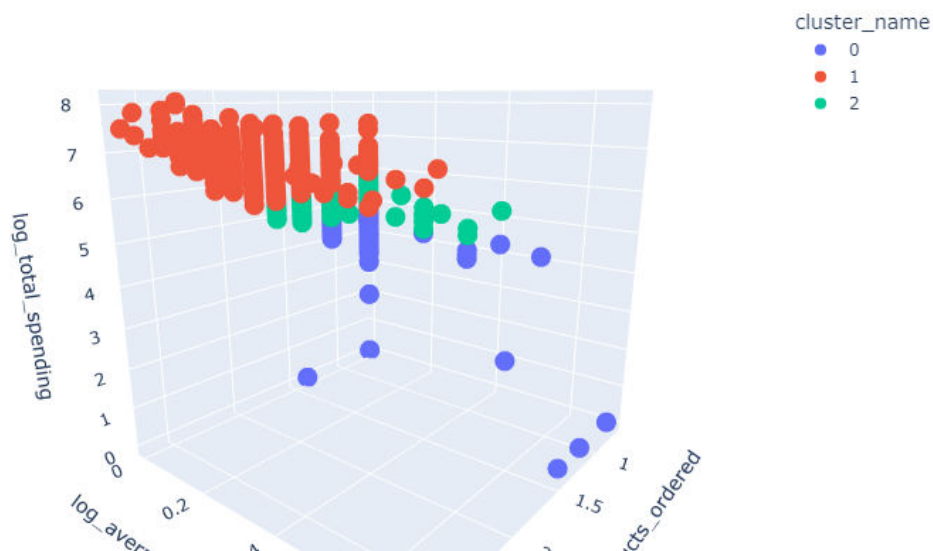


Fig.1 Visualization of customer segmentation



VI. CONCLUSION AND FUTURE WORK

In this project, we illustrated analyzing and segmenting the task of e-commerce customer data. Customer segmentation is a way to improve communication with customers, to know customer desires, customer activities so that appropriate communication can be built. Customer segmentation is necessary so that it leads to potential customers, which increases profit. After the clustering and segmentation phases, the model can segment the customer based on their behaviour. Then it will be cached separately based on their segmented characteristics. Many organizations' marketing team are now highly focused on enticing people to their product. In this respect, the deployed project will benefit the marketing team. Customer segmentation problem is approached from a behavioral aspect with the number of products ordered, average return rate and total spending for each customer. Using 3 features made it easy for us to understand, clustering process and visualize the model. In terms of correctness of the model, k means clustering algorithm in the deployed model will not have any outliers that reduces the accuracy of the model, which is a major drawback of existing model DBSCAN

Although the proposed system outperforms the present method in terms of clustering, there is still opportunity for additional validation and enhancement. In future, customers can be segmented based on other types of customer segmentation, such as psychographic segmentation, which is based on a customer's inner or qualitative traits, demographic segmentation, which is based on customer personas, and geographic segmentation, which is based on a customer's physical location. Instead of feeding period record of customer data, in future immediate segmentation can be done with the aid of servers. In future, for simplicity of usage and delivery, the complete deployed model could be constructed into a web application.

REFERENCES

1. Yash Kushwaha, Deepak Prajapati, "Customer Segmentation using KMeans Algorithm".
2. Nurma Sari, Juni & Nugroho, Lukito & Ferdiana, Ridi & Santosa, Paulus. Review on Customer Segmentation Technique on Ecommerce. Advanced Science Letters. 2016.
3. Kishana R. Kashwan and C. M. Velu, "Customer Segmentation Using Clustering and Data Mining Techniques," International Journal of Computer Theory and Engineering vol. 5, no. 6, pp. 856-861, 2013.
4. Namvar, Morteza et al. "A Two Phase Clustering Method for Intelligent Customer Segmentation." 2010 International Conference on Intelligent Systems, Modelling and Simulation (2010): 215-219.
5. E. Saraf, S. Pradhan, S. Joshi and S. Sountharajan, "Behavioral Segmentation with Product Estimation using K-Means Clustering and Seasonal ARIMA," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1641-1648, doi: 10.1109/ICOEI53556.2022.9776834



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details