

# Customer Segmentation Using Machine Learning

Prof. Nikhil Patankar <sup>a,1</sup>, Soham Dixit <sup>a</sup>, Akshay Bhamare <sup>a</sup>, Ashutosh Darpel <sup>a</sup> and Ritik Raina <sup>a</sup>

<sup>a</sup>Dept. Of Information Technology Sanjivani College of Engineering, Kopergaon-423601 (MH), India

**Abstract.** Nowadays Customer segmentation became very popular method for dividing company's customers for retaining customers and making profit out of them, in the following study customers of different of organizations are classified on the basis of their behavioral characteristics such as spending and income, by taking behavioral aspects into consideration makes these methods an efficient one as compares to others. For this classification a machine algorithm named as k-means clustering algorithm is used and based on the behavioral characteristic's customers are classified. Formed clusters help the company to target individual customer and advertise the content to them through marketing campaign and social media sites which they are really interested in.

**Keywords.** Machine learning, Customer segmentation, K-means algorithm

## 1. Introduction

Today many of the businesses are going online and, in this case, online marketing is becoming essential to hold customers, but during this, considering all customers as same and targeting all of them with similar marketing strategy is not very efficient way rather it's also annoys the customers by neglecting his or her individuality, so customer segmentation is becoming very popular and also became the efficient solution for this existing problem. Customer segmentation is defined as dividing company's customers on the basis of demographic (age, gender, marital status) and behavioral (types of products ordered, annual income) aspects. Since demographic characteristics does not emphasize on individuality of customer because same age groups may have different interests so behavioral aspects is a better approach for customer segmentation as its focus on individuality and we can do proper segmentation with the help of it.

## 2. Literature Survey

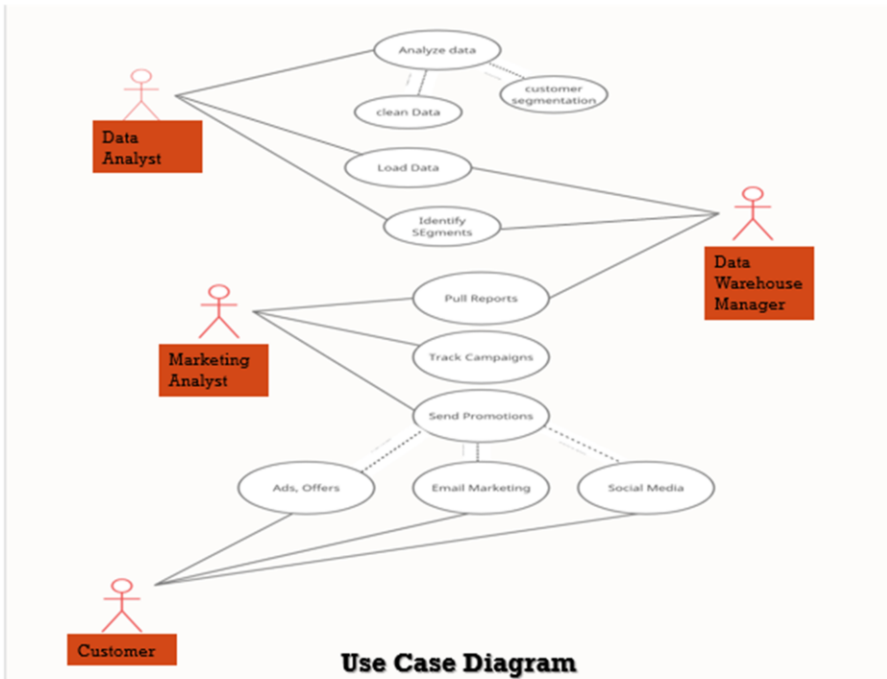
[1] A solution is proposed as distinguish the customers group into two groups named as premium and standard with the help of machine learning methods named as NEM, LiRM and LoRM [2].

---

<sup>1</sup> Prof. Nikhil Patankar, Sanjivani College of Engineering, Kopergaon, India.  
Email: patankarnikhil@sanjivani.org.in

Tushar Kansal, Suraj Bahuguna, Vishal Singh, Tanupriya Choudhury. “Customer Segmentation using K-means Clustering”, International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS).2018, In this paper customer segmentation on Telecom customers is achieved by using information such as age, interest, etc. with the help of cluster analysis method.

### 3. Use Case Diagram



**Figure 1.** Use case Diagram

Use case diagram of proposed system consist of 4 users 1. Data Analyst 2. Marketing Analyst 3. Data Warehouse Manager 4. Customer in figure 1 And 6 use cases,

1. Analyze Data: analyst has the access to loaded data and analyst clean the data and perform analysis to form clusters.
2. Load Data: analyst log into database & view data & load into memory to work on it.
3. Identify Segments: analyst form report for segmented customer data and send to data warehouse and marketing analyst can access that data to form marketing strategies.
4. Pull Reports: marketing team can view & make edits on the reports, data for report is pulled from DW system.

5. Track Campaigns: The customer's interaction tracked by marketing team for success report.
6. Send Promotions: Marketing team send promotions through mail, social media ads, paid ads, coupons.

#### 4. K-means Clustering Algorithm

K-means Clustering is a clustering Algorithm in which we are given with data points with its data set and features and the mechanism is to categories those data points into clusters as per their similarities.

The algorithm forms K clusters based on its similarity. To calculate the similarity K-means uses Euclidean distance measurement method.

Steps

- i. In first step, we randomly initialize k points.
- ii. K-means classifier categorizes each data point to its nearest mean and rewrite the mean's coordinates.
- iii. Iteration is continuing up till all data points are classified.

#### 5. Proposed System

In our system we including annual income and total spending as a feature for classification in figure2

1. **Data Gathering:** first, Data analyst fetch data required for analysis from database, format data i.e., remove all NA values from data & make data ready for processing.
2. **Feature Extraction:** Selects features which makes model more accurate, in our case features are annual income and spending score for efficient analysis.
3. **K-means Classifier:** After that, K means classifier performs clustering with respect to features provided to it,
4. **Hyper Parameter Tuning:** during forming groups to select optimal no of clusters we applied hyper parameter tuning which is achieved by Elbow method to choose optimal no of clusters.

below graph is for elbow method which shows curve is getting flatter after 5 which indicates that 5 is optimal no of clusters we can form for better classification.

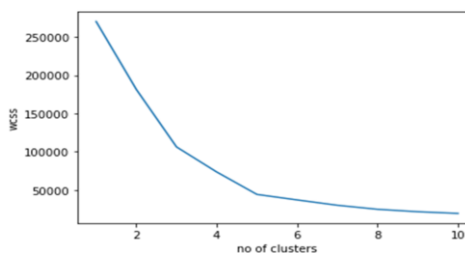


Figure 2. Elbow Method

- 5. **Data Visualization:** With the formed clusters marketing team can make different strategies for better targeting customers in figure 3.

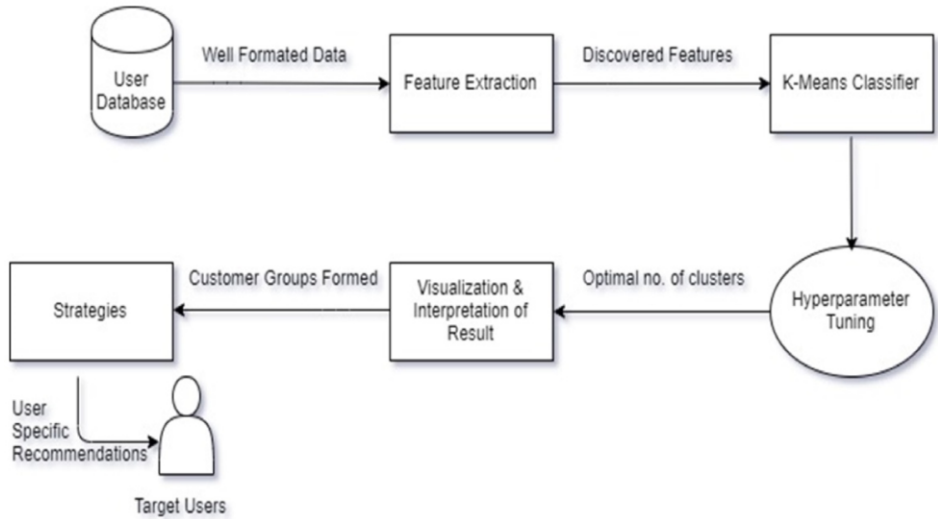


Figure 3. Flow of operation

### 6. Results

After analysis of data and classifying customers with features annual income and spending score, we got clusters of customers & with formed clusters marketing team form strategies for customers specific recommendation to make value out of them in figure 4.

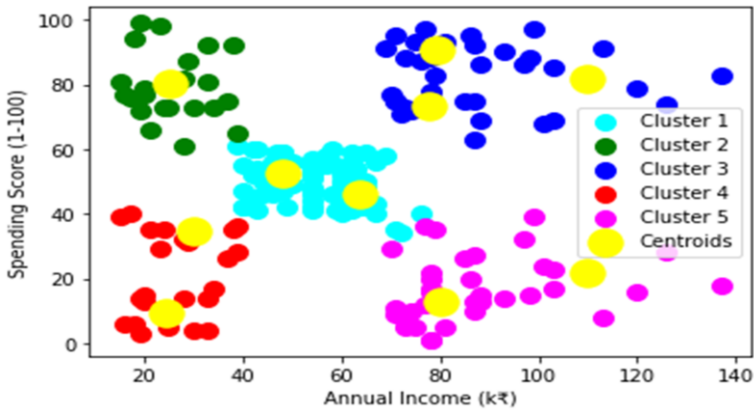


Figure 4. Final Cluster Formed

## 7. Drawback of System

1. Marketing will become expensive.
2. Because of having less no. of customers in a segment problem of limited production occurs.

## 8. Conclusions

Customer segmentation is performed on the company's customers data and with the help of K-means clustering machine learning algorithm customers are divided using features like total spending and annual income, this study also proves that the dividing customers on the basis of behavioral characteristics is a better solution for existing customer segmentation problem and K-means clustering algorithm is identified as a good choice for this approach.

## References

- [1] Sukru Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods", IEEE, Year: 2018.
- [2] Jayant Tikmani, Sudhanshu Tiwari, Sujata Khedkar "Telecom customer segmentation based on cluster analysis An Approach to Customer Classification using k-means", IJIRCCCE, Year: 2015.
- [3] Chinedu Pascal Ezenkwu, Simeon Ozuomba, Constance kalu Electrical/Electronics and Computer Engineering Department, University of Uyo, Akwa Ibom State, Nigeria "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services", IJARAI, Year: 2015.
- [4] Potharaju, S. P., Sreedevi, M., Ande, V. K., & Tirandasu, R. K. (2019). Data mining approach for accelerating the classification accuracy of cardiocography. *Clinical Epidemiology and Global Health*, 7(2), 160-164.
- [5] Yogita Rani and Dr. Harish Rohil "A Study of Hierarchical Clustering Algorithm", IJICT, Year: 2013.
- [6] Omar Kettani, Faycal Ramdani, Benaissa Tadili "An Agglomerative Clustering Method for Large Data Sets", IJCA, Year: 2014.
- [7] Sneha, Chetna Sachdeva, Rajesh Birok "Real Time Object Tracking Using Different Mean Shift Techniques—a Review", IJSCE, Year: 2013. SulekhaGoyat "The basis of market segmentation: a critical review of literature", EJBM, Year: 2011.
- [8] Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In *Cognitive Informatics and Soft Computing* (pp. 247-256). Springer, Singapore.
- [9] Tanupriya Choudhury, Vivek Kumar, Darshika Nigam, Intelligent Classification and Clustering Of Lung and Oral Cancer through Decision Tree and Genetic Algorithm, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2015
- [10] Tanupriya Choudhury, Vivek Kumar, Darshika Nigam, An Innovative and Automatic Lung and Oral Cancer Classification Using Soft Computing Techniques, *International Journal of Computer Science and Mobile Computing*, 2015
- [11] H. Mehta, V.S. Dixit and P. Bedi, "Refinement of recommendations based on user preferences".
- [12] H. Mehta, S.K. Bhatia, V.S. Dixit and P. Bedi, " Collaborative personalized web recommender system using entropy-based similarity measure".
- [13] Rivedi, A., Rai, P., DuVall, S. L., and Daume III, H. (2010, October). 'Exploiting tag and word correlations for improved webpage clustering in Proceedings of the 2nd international workshop on Search and mining user-generated contents (pp. 3-12). ACM.
- [14] Potharaju, S. P., & Sreedevi, M. (2017). A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance. *Journal of Engineering Science & Technology Review*, 10(6).
- [15] Domavicius, G., and Tuzhilin, A. (2015). Context-aware recommender systems. In *Recommender systems handbook* (pp. 191-226). Springer US.

- [16] K. Windler, U. Juttner, S. Michel, S. Maklan, and E. K. Macdonald, "Identifying the right solution customers: A managerial methodology," *Industrial Marketing Management*, vol. 60, pp. 173–186, 2017.
- [17] R. Thakur and L. Workman, "Customer portfolio management (cpm) for improved customer relationship management (crm): Are your customers platinum, gold, silver, or bronze?" *Journal of Business Research*, vol. 69, no. 10, pp. 4095–4102, 2016.