SEGMENTATION OF CONSUMER USING UNSUPERVISED LEARNING

Arti Ranjan Preeta Rajiv Sivaraman***

ABSTRACT:
Customer segmentation is to collect customer related data to generalize some insight so as to solve business challenges. Data around customers like how they browse, what they choose from where they log in, how they buy things, this is something personal detail of the customer and studying the behaviour of the customer is customer segmentation. After getting all the details most of the companies or business industry provides, customer related market activity which will help not only companies but also customers too for win situations. This is a target market analysis process. In this study market customer segmentation is implemented using k means clustering.

Keywords: ***Cluster analysis, Consumer segmentation, k means.***

INTRODUCTION

Customer segmentation helps in better categorization of customers depending on their behaviour and characteristics. It reduces the risk of pitfall in marketing strategy as this analysis directly hit the right group of people. This is possible due to machine learning and artificial intelligence which monitor each detail of the customer and recommend according to that. Find the statistical pattern in the data, discover recurring patterns from the data and analyse it [1,2]. To collect data for customer segmentation we need to monitor data which can be (RFM) Reoccurrence-when was the last visited, Frequency -how many times customer visited and Monetary-How much money do they spend. For loan companies it is necessary to know about their customer that when a customer is last visited on their site and for how much duration, how frequently they visit and how much they spent their [3]. Which helps companies to find the customer needs and helps in creating new advertisements or schemes beneficial
for customers? Generally company follows STP model that is segmentation, Targeting, and positing. It always works with B2C business models like supermarkets, online shopping sites. Segmentation- Dividing the population into groups that share similar characteristics has comparable purchase behaviour and response similarly to different purchase behaviour. Like sometime people are interested to buy during sale time [4]. For new product development a new data is never available in the market so demographic and Geographic data age, income, education level and others need to be understood. When we have historic data like purchase quantity, time of purchase, purchase quantity and product rating it helps to much clear representation of data. Targeting-evaluating from all segments and deciding which segment to focus on. Here consideration factors are segment size, expected growth and competitor's offering [5].
After finalizing the segment we have to check which product is closely matched with the characteristic to meet customer requirements. And draft advertisement, discount offering according to that.

Types of Segments:

1. Geographic segmentation It is about regions, countries, continents customer segmentation. After separating into regions, we need to check which is more profitable and also analyse customer culture, environment and location [6].

2. Demographic: This analyses customer income, size of family, education, occupation, family which help to find out consumer needs.

3. Psychographic: This will judge about the customer's way of living, lifestyle, beliefs which help to understand the company offer for their customer.
4. Behavioural segmentation: This will observe behaviour of a person based on their income level.

Literature review of Customer segmentation

In this section we will focus on previous study in this area by other researchers. The purpose of customer segmentation is to separate out the groups of customers with similar demand or need that can be based on their characteristics like their buying habits, life standard and many more[7]. Few researchers worked over customer segmentation in different areas such as Marketing, Finance, healthcare. In the current era customer segmentation is a must for growing business strategy, knowing customer needs and choice throughout the year. By awareness of customer requirements a company can launch products at the correct time [8].
According to Sandstorm [9] customers of different significance must be in the market. Companies need to analyse each customer from non-profitable to highly profitable to evaluate the proportionality over the profits. Special discounts and schemes will be launched for the customer like

**Research Scholar, Galgotias University, Greater Noida **Assistant Professor, Asian School of Business, affiliated to CCSU*

many ecommerce sites do [10] state due to customer segmentation not only production industry as well as service is also getting benefit in work and helps in improvement of services products[11]. Venkata et al. 2017 used segment specific strategies and proposed a model of margin cost to serve matrices which is beneficial to small retailers.Juni Nurma Sari et.al in paper [12] authors have classified the data for customer segmentation in two categories i.e. Internal data and External data. They had categorized the customer profile and transaction history data as internal data and data like cookies, server log and survey data were categorized in external data. Mento, supervised clustering, RF techniques are some categorized methods. Jing Wu et.al in their paper [13] had used credit card transaction data for model building and prepared predictive models at segment-level utilizing pattern based clustering approach. They devised two matrices i.e. Fluctuate-rate matrix and Monetary matrix and performed clustering on both of the matrices to discover various customer characteristics. Further, they used those characteristics to build consumption based consumer segmentation models. Kareena et.al in paper [14] proposed hybrid classifier technique using Decision tree and KNN for customer behavior analysis which outperformed the performance of previously accepted Naïve Bayes model by many researchers. Naive Bayes has 74.11% accuracy whereas Hybrid classifier has an accuracy of 90.75% which is much better. A.Salini et.al [15] compared the performance of 3 individual classification algorithms i.e. Random Forest, Support Vector Machine and Logistic Regression with that of Majority Voting algorithm and proved successful in achieving better Precision, Recall, F-Measure and Accuracy as well. K Means clustering is used to solve the clustering of customers to increase the revenue of the company.

First we need to indicate the number of clusters(K) to get the final output. The algorithm selecting k objects from the data set and finds the centroids of the cluster. Find out eculicleam distance which the difference between cluster mean and object. Assign the new mean to the cluster and then find out the new mean until we get an updated cluster mean. Then using the elbow method, find out the cluster needed to solve this problem. We use python programming language to implement this problem and imported yellow brick package for visualization sklearn package for K-means, and principal components analysis. Used standard scalar for pre-processing of the data. Dataset includes 5 columns customer id, gender, age, Annual income and spending score. The below table1 contains the dataset.

Table 1. Market customer data

Customer ID		Gender	Age	Annual Income (k$)	Spending Score (1-100)	age_cat	income_cat	score_cat
0	1	Male	19	15	15	Student	very rich	low
1	2	Male	21	15	15	Student	very rich	hight
2	3	Female	20	16	16	Student	very rich	very low
3	4	Female	23	16	16	Student	very rich	moderate
4	5	Female	31	17	17	young professional	very rich	

The data set contains numeric columns of age, annual income and spending score. So first I need to convert the numeric column into categorical. We categories the annual income as poor who have less than 25k salary, rich who have income range between 25 to 75k annually. Very rich for those who have greater than 75k annual income similarly age column is also categorized People less than 25 age are considered as students, 25-45 young professionals,46-65 professional and above 65 are considered to be in the retired category. Similarly for spending score the people having less than 20k spending amount considered to be very low spending, spending range 20 to 40 considered as low,40-60 considered as moderate,60 to 80 as high and above 80 is very high.

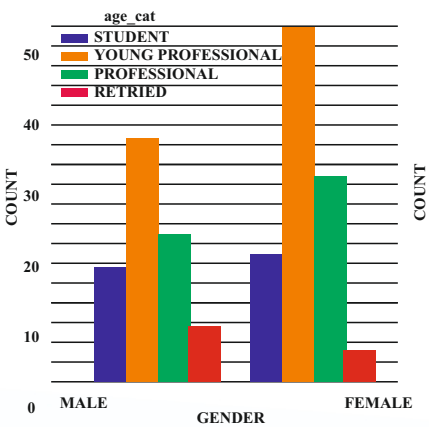


Fig 1: Age category based on gender
In fig1. The graph shows that young professionals in both males and females have the highest purchase.

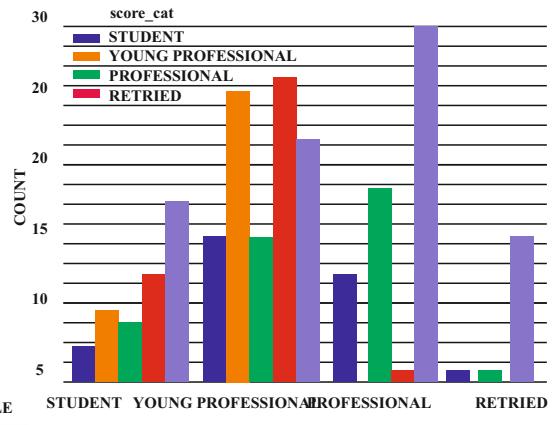


Fig2. Spending score based on category

*Research Scholar, Galgotias University, Greater Noida **Assistant Professor, Asian School of Business, affiliated to CCSU

Applied label encoder to transform the data and the after that applied elbow method to find out the cluster value and optimal value is 6. Elbow method is a validation and interpretation within the cluster to find the appropriate number of clusters in the dataset. Calculated sum of square error for different k value.

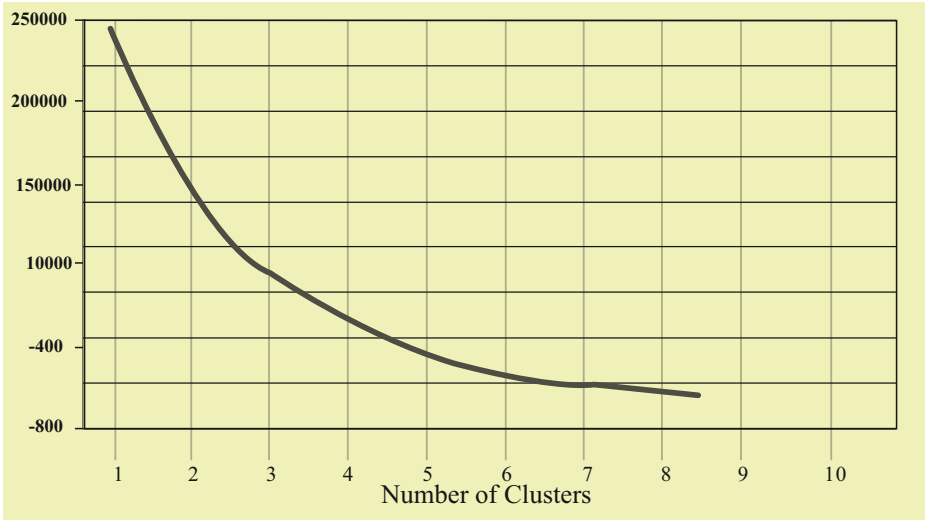


Fig 3.Sum of Square error for different value of k

Using silhouette coefficients we have measured the quality of clustering operation for different k values.

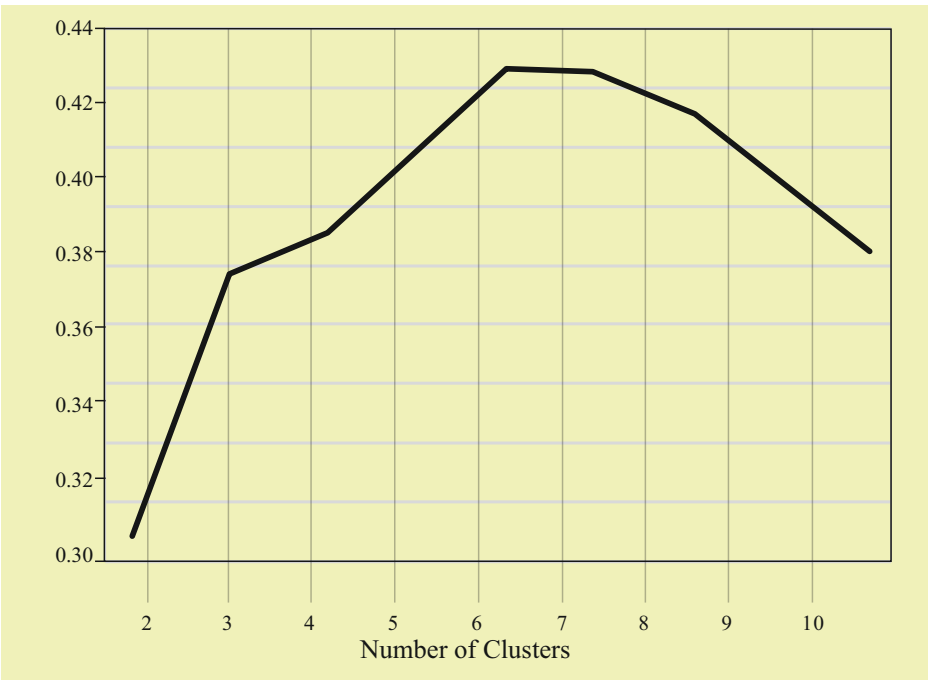


Fig4. Silhouette Coefficient

Using yellow brick cluster silhouette coefficient is visualized which is as shown in fig5.

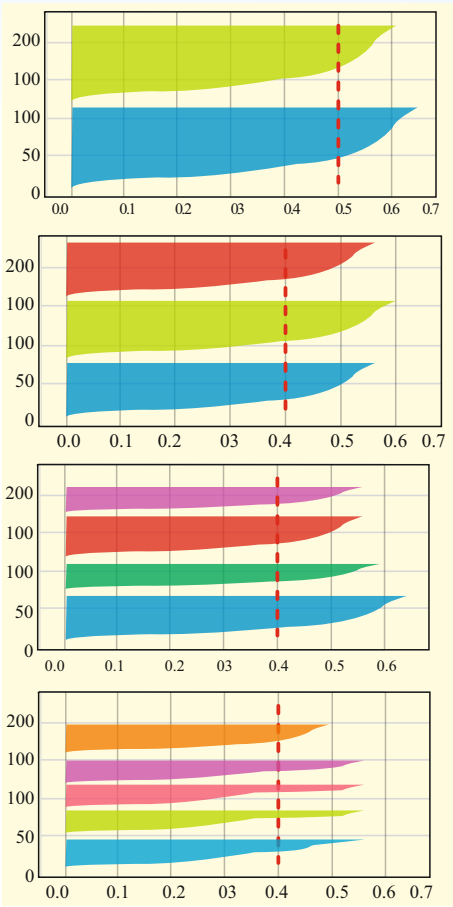


fig5: silhouette coefficient instance with KMeans instance

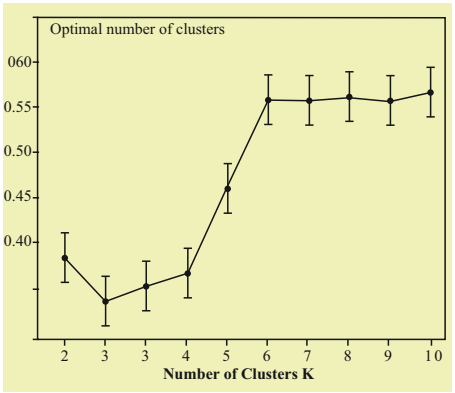
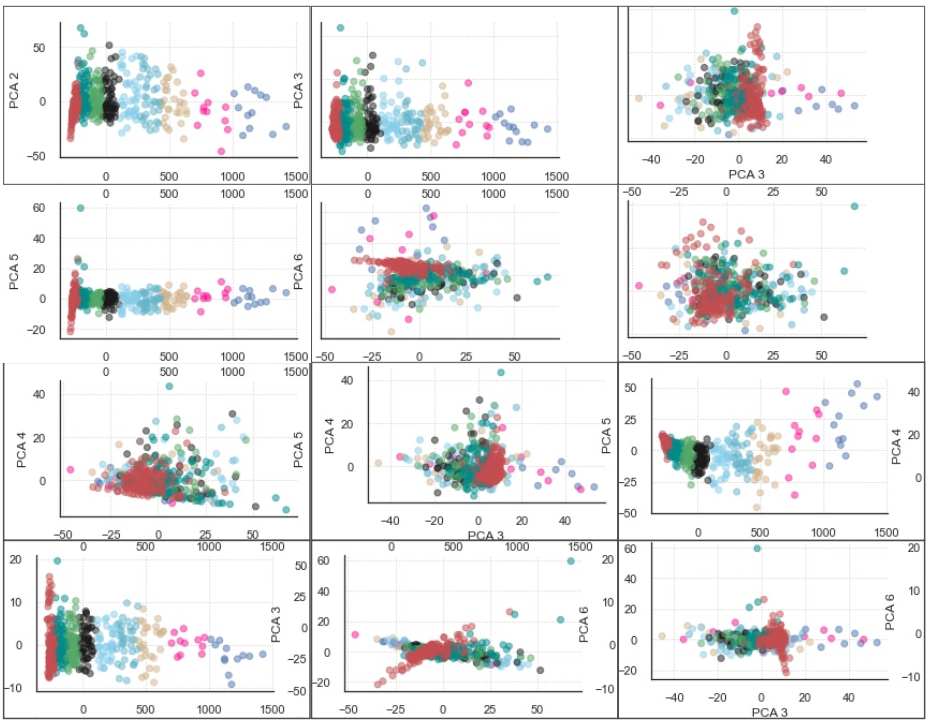


Fig6. GAP statistic of k values

Using Gap statistic intracluster distance is calculated for values of K.as shown in fig6 there is a large g a p between 4 and 5 k values which is considered to be good.

*Research Scholar, Galgotias University, Greater Noida **Assistant Professor, Asian School of Business, affiliated to CCSU

Applied label encoder to transform the data and the after that applied elbow method to find out the cluster value and optimal value is 6. Elbow method is a validation and interpretation within the cluster to find the appropriate number of clusters in the dataset. Calculated sum of square error for different k value.



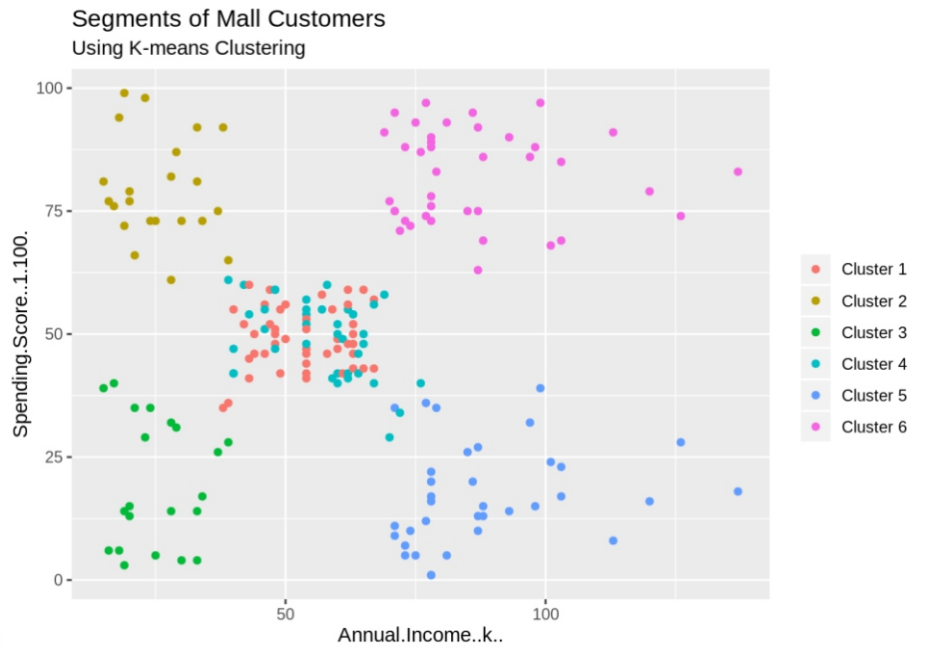


Fig7. PCA filters the noise data and expresses the amount of variance.

Fig8. show the plots between spending score and annual income. Where cluster 1 and 4 represent a medium annual income and spending score. cluster 2 represents low annual income but high spending score. cluster 3 represents low annual income and spending score. cluster 5 represents middle to high annual income with less spending score. Cluster 6 represents high annual income and spending score.

Fig8. clustering of market data

Conclusion and Future Scope

The aim of customer segmentation is to improve the relations with customers and improve company buying and selling margin. From the above section of literature review it has been observed that after knowing your customer it's easy for the organization to provide offers, customer choice products and many more. Sometimes due to limited data and restricted customer details it's difficult to segment customers and differentiate among other customers. The study shows the impact of customer spending on the basis of their earnings. Male to female ratio of spending which predict the behaviour of customer purchase using machine learning algorithms and helps to develop and introduce new marketing schemes or offers. In future segmentation is useful in supply chain, CRM, Optimize network planning to design traffic schemes and product review.

References:

1. Ş. Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2018, pp. 1-6.

2. Dr. V. VenkataNarayana and Dr. A. Kanaka Durga, Customer Segmentation -A Case of Increase in Customer Satisfaction Levels with Segment Specific Customer Service Strategies. International Journal of Computer Engineering & Technology, 8(6), 2017, pp. 89–94.

3. Marleen Onwezen, Chapter 15 - Including Context in Consumer Segmentation: A Literature Overview Shows the What, Why, and How, Editor(s): Gastón Ares, Paula Varela, In Woodhead Publishing Series in Food Science, Technology and Nutrition,Methods in Consumer Research, Volume 1, Woodhead Publishing,2018,Pages 383-400

*Research Scholar, Galgotias University, Greater Noida **Assistant Professor, Asian School of Business, affiliated to CCSU

4. Bottcher, M., Spott, M., Nauck, D. & Kruse, R (2009). Mining changing customer segments in dynamic markets, Expert Systems with Applications 36 (2009) 155–164.

5. Buttle, F. (2009). Customer Relationship Management: Concepts and Technologies.

6. Anderson, J.C., Narus, J. A. and Narayandas, D. (2009). Business Market Management Understanding, Creating and Delivering Value, Pearson Education Prentice Hall: Upper Saddle River
New Jersey, Third edition.

7. Fang E., Palmatier R.W., &Steenkamp, J.B.E.M. (2008). Effect of service transition strategies on firm value. J Mark, 72(4), 1–14.

8. Lambert, M. (1990). Segmentation of Markets Based on Customer Service, International Journal of Physical Distribution & Logistics Management, Vol. 20 Iss 7 pp. 19 – 27

9. Ilung Pranata and Geo' Skinner. "Segmenting and targeting customers through clusters selection & analysis". In: Advanced Computer Science and Information Systems (ICACSIS), 2015 International Conference on. IEEE. 2015, pp. 303–308

10..Minghua Han. "Customer segmentation model based on retail consumer behavior analysis". In: Intelligent Information Technology Application Workshops, 2008. IITAW'08. International Symposium on. IEEE. 2008, pp. 914–917.

11. Adam Elvander. Developing a Recommender System for a Mobile Ecommerce
Application. http://uu.diva-portal.org/smash/record. jsf?pid=diva2%3A826802. Retrieved: 2017-01-19.

12. J. du Toit, R. Davimes, A. Mohamed, K. Patel, and J. M. Nye, Customer Segmentation Using Unsupervised Learning on Daily Energy Load Profiles, Journal of Advances in Information Technology Vol. 7, No. 2, May 2016.

13. JuniNurma Sari, LukitoNugroho, RidiFerdiana, PaulusInsapSantosa: Review on Customer Segmentation Technique on Ecommerce, Journal of Computational and Theoretical Nanoscience, 22(10):3018-3022, Oct-2016.

14. Jing Wu, Zheng Lin: Research on customer segmentation model by clustering, ICEC'05, Proceedings of the 7th International conference on Electronic commerce, Page 316-318, Aug-2005, DOI:10.1145/1089551.1089610.

15. Kareena, Raj Kumar: A Consumer Behavior Prediction Method for E-Commerce Application, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume - 8, Issue – 2S6, Jul-2019

16. A.Salini, U.Jeyapriya: A Majority Vote Based Ensemble Classifier for Predicting Students Academic Performance, International Journal of Pure and Applied Mathematics, Volume 118 No. 24, ISSN: 1314-3395, Mar-2018

*Research Scholar, Galgotias University, Greater Noida **Assistant Professor, Asian School of Business, affiliated to CCSU