

The Role of Customer Segmentation in Engineering Innovation: A Green Approach

¹Shreya Issar, ²Abhishek Purohit, ³Anu Sharma

^{1,2}Department of Computer Science & Engineering, University College of Engineering and Technology, Bikaner Technical university, Bikaner, Rajasthan, India

³Department of HEAS (Chemistry), University College of Engineering and Technology, Bikaner Technical university, Bikaner, Rajasthan, India

ABSTRACT

In this digital era the behaviour of customers has undergone a profound transformation due to easy accessibility of internet and smart engineering technologies. Many industries, businesses, companies and organizations are offering the services to them. The market is very competitive so there is a need to adopt a technology as per customer expectations to cope up with these problems. The customer segmentation is one the best way to rectify these problems. This method is based upon to grouping customers based on similarities to tailor offerings and experiences. With the help of these segments, they can better meet specific needs and can build a best customer relationship. This paper explores how traditional and advanced customer segmentation methods such as K-Means Clustering, Principal Component Analysis, decision Tree Classifier, Elbow Method etc. to engineering innovation and highlights the intersection of customer analytics and engineering design, provide customize solutions.

Keywords: Customer Segmentation, K-Means Clustering, Principal Component Analysis (PCA), Decision Tree Classifier, Marketing Strategy.

INTRODUCTION

Customer segmentation plays a significant role in driving engineering innovation by enabling organizations, companies and other sector to customize their products, services. It is the process of dividing customers into groups on the basis of some parameters such as their characteristics, behaviours, needs, and preferences. With the help of they can tailor their products, services, marketing, and communication strategies to each segment in proper manner and can make a good customer relationship. This innovative technique helps personalize marketing efforts, improve customer engagement, and embellish the customer experience. With the understanding different customer segments, businesses can deliver more relevant content, personalize offers, and optimize their strategies for better results.

Customers have specific characteristics, choices, preferences, and expectations so reorganization of these variable is necessary for effective outcome. Customer segmentation is the method which allows businesses to gain deeper insights into individual customer segments, creating tailored content that fulfil their unique needs and challenges. It increases the consumer engagement by advertisements and marketing messages via suitable platform. If the customer found a good review of a product, then it increases his satisfaction with that product. Customer segmentation eliminates wasted time, money and improve return on investment.

Thus, customer segmentation is a powerful tool for engineering innovation, enabling companies, industries, businesses to create more relevant, effective, and profitable products. Its application is continuously increasing in recent decades. The brief of application is shown in this Table 1.

Table:1 Applications of Customer segmentation in various sectors

Sectors	Applications
Retail & E-commerce	Personalized product recommendations for example Amazon, Flipkart, Spotify etc.
Banking & Insurance	Financial recommendations, Investment strategies
Healthcare	Health programs for example diet plans, weight management etc.
Environmental Campaigns	Designing eco-awareness campaigns
Transportation	Fare plans such as travellers by purpose or location, need logistic etc.
Smart Cities	Public safety, Governance utilities and energy management strategies adjusted for different user segments
Education	Scholarship programs for students with different backgrounds and needs

LITERATURE REVIEW

The area of customer segmentation has evolved significantly over the past decades. There is transition from traditional demographic-based methods to smart engineering innovation methods. Kotler & Keller studied about the customer groups as per age, income, and location. The emergence of data mining techniques has been found in further studies and new methods for customer segmentation based on behavioral and transactional data was introduced.

Clustering algorithms, particularly K-Means, was widely used due to its efficiency and ease of interpretation. MacQueen (1967) was first who introduced the K-Means algorithm. Punj & Stewart, 1983; Dolnicar, 2003 have validated the utility of K-Means for marketing segmentation for large datasets. Elbow method was also used to resolve the challenges arose in K means algorithms. The other engineering innovation was also investigated in this direction for example Principal Component Analysis have been developed by Pearson (1901) and later expanded by Hotelling (1933), reduces the number of features by transforming them into a set of linearly uncorrelated components. It was a good tool in customer segmentation which preserved the variance in the dataset and eliminating redundancy. As clustering is an unsupervised learning task, external validation techniques, including the use of supervised classifiers, have been studied by Rendón et al., 2011.

One of the important method decision trees a transparent method for evaluating clustering performance by providing interpretable rules based on feature splits was invented by Leo Breiman and his coworkers. S. Dutta et.al. did his work in the same field and gave a better input in the research of customer segmentation. A. Kuusik et. al. studied in telecommunication sector. Many others studied in various field such as banking sector, life insurance company and e-business managements and found fruitful results.

METHODS OF CUSTOMER SEGMENTATION

The different traditional and advanced methods are used for customer segmentation. The customers can divide into smaller subgroups by few characteristics. These common characteristics help in creating customer profiles

which leads to increasingly personalized marketing messages that guide customers. The traditional methods are summarised in Table: 2. This method are as follows-

Behavioural Segmentation

Behavioural segmentation is on the basis of customer actions and interactions with product or services provided by company. This segmentation is categorized by frequent activities, purchase history, product usage, and routines etc. It is focus on add, emails and reminder on social platform to regular basis.

Geographic Segmentation

Geographic segmentation is on the basis of customers location, country size, states, cities, towns, etc. This segmentation method is useful for companies which are operating in multiple locations within the same city or country.

Table:2 Traditional methods of customer segmentation

Factors	Characteristics
Behavioural	User status, purchase occasion, product, readiness state, rate, loyalty status and benefits
Geographic	Region, population size, country size, climate and city size
Demographic	Age, gender, family size, life cycle, income, occupation, education, religion, race and nationality
Psychographic	Social class, life style and personality

Demographic Segmentation

Demographic segmentation is based on customers income, education level, gender, and age. This segmentation is beneficial is useful for ecommerce, fashion-based companies that sells both mid-range and luxury range products for women and men.

Psychographic Segmentation

Psychographic segmentation is based on customers including personality, habits, beliefs, and interests. It is useful for lifestyle brands. This segmentation deal with the emotional and mental aspects influencing consumer behavior and can create more targeted and effective product development strategies.

The advanced data mining methods have also been emerged as essential tools in customer segmentation. This modern approach involves processing vast datasets from databases using intelligent solutions such as K-Means Clustering, Principal Component Analysis, decision Tree Classifier, Machine Learning, Elbow Method, kernel methods and Taguchi method etc. These methods are highly effective and time-efficient.

The advanced customer segmentation methods are as follows-

K-Means Clustering:In essence, K-means clustering is a powerful tool for grouping data based on similarity, with a wide range of applications in various fields, but it's important to be aware of its limitations, especially regarding the choice of the number of clusters and the shape of the cluster

This method is used to organize data into groups based on their similarity. The algorithm works by picking few central point's randomly called centroids (Figure 1). After this the similar centroid forming a cluster. After all the points are assigned to a cluster the centroids are updated by finding the average position of the points in each cluster. This process repeats until the centroids stop changing forming clusters. The goal of clustering is to divide the data points into clusters so that similar data points belong to same group. This is useful for online store.



(a) Unlabelled Data

(b) Labelled Clusters

Figure: 1 K- Means Clustering

Principal Component Analysis:

When dataset is very big then this method reduces the dimensionality of these dataset and create a set of uncorrelated variables. These uncorrelated variables also called principal component are linear functions of those in the original dataset. It is the oldest method and widely used because information loss is minimum and preserving the variability. This is useful to detect anomalies in credit card transactions, sensor readings, or network traffic etc.

Decision Tree Classifier

It is the processing of data by cleaning and creating variables from existing data. It uses a tree-like graph where each node represents a decision point based on a customer's characteristics such as age, buying behaviors etc. It evaluates the conditions at each step to divide customers into smaller groups (Figure 2). It is a valuable tool for retail business and can increase customers satisfaction, loyalty, and profitability etc.

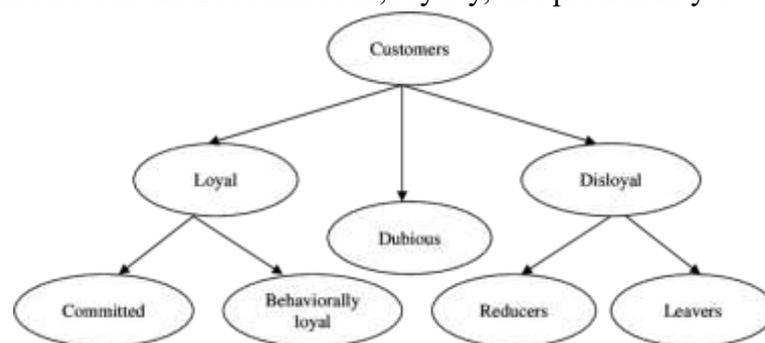


Figure: 2 Decision Tree Classifier

METHODOLOGY:

Customer segmentation is a strategic process that divide customers into distinct groups on the basis of shared characteristics such as location, behaviors, size and needs etc. This methodology enables businesses to tailor their services. It involves following steps (Figure 3:).

Data Preprocessing: It involves data cleaning for handling missing value, reducing noise and duplicate data which is followed by data integration and transformation. scaled the dataset for better performance.

Dimensionality Reduction:

Dimensional reduction is carried out by using principal component analysis to reduce the number of features while retaining important information

Hyperparameter Tuning:

Apply the Elbow method to find the optimal number of clusters for K-Means clustering.

Model Training:

Trained the K-Means clustering model using the identified optimal number of clusters. And found the centroid or cluster centers for each segment.

Model and Data Persistence:

Saved these trained K-Means model and the scaled dataset using joblib for future use.

Model Evaluation:

After this use decision tree classifier to evaluate the model's accuracy and split the data into training and testing sets. Now trained the decision tree classifier on the training data and tested the accuracy of the model on the testing data. Saved the trained decision tree model for future use.

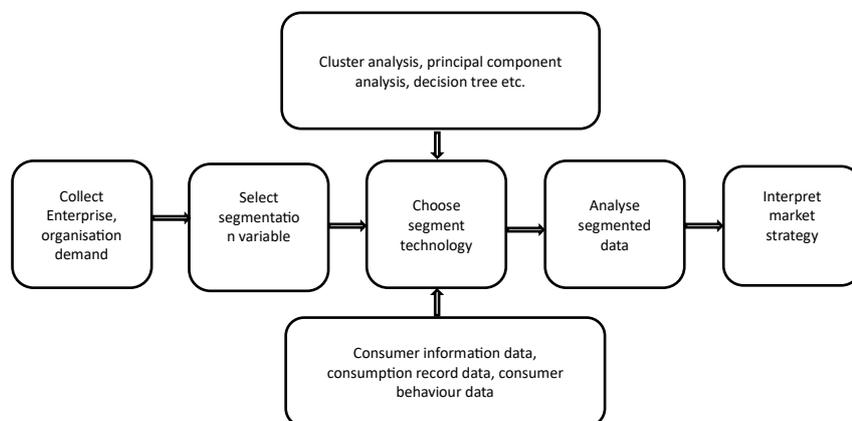


Figure 3: Process of Customer Segmentation

DATA ANALYSIS AND RESULTS

Datasets obtained after the data preprocessing phase was ready for dimensionality reduction. Principal Component Analysis method applied for dimensional reduction. The smaller set of principal components retained better variance. This step was critical for improving the computational efficiency of the model and enhancing the interpretability of the clustering results. Elbow method was used for determining the optimal

number of clusters (K) for the K-Means algorithm. To further validation a decision tree classifier was used to predict cluster memberships. The dataset was split into training and testing sets with good ratio.

It has been observed that data segmented into four clusters (K=4), which are represented by different colors that is red (cluster 0), green (cluster 1), blue (cluster 2), black (cluster 3). Red clusters are densely packed and denote low-spending or low-activity customers. They may be new customers, low-value buyers and those with minimal transactions. Green clusters are slightly more spread. These customers show moderate activity and potential growth segment if targeted with some offers. Blue clusters spread out along the PCA1 axis and are most dispersed cluster indicates high variance in behavior means high-spending customers, frequent shoppers. They can be considered the premium segment. Black clusters are most densely packed and positioned slightly upward along PCA2 could represent frequent but average spenders or a core loyal segment. These are important for consistent business (Figure 4).

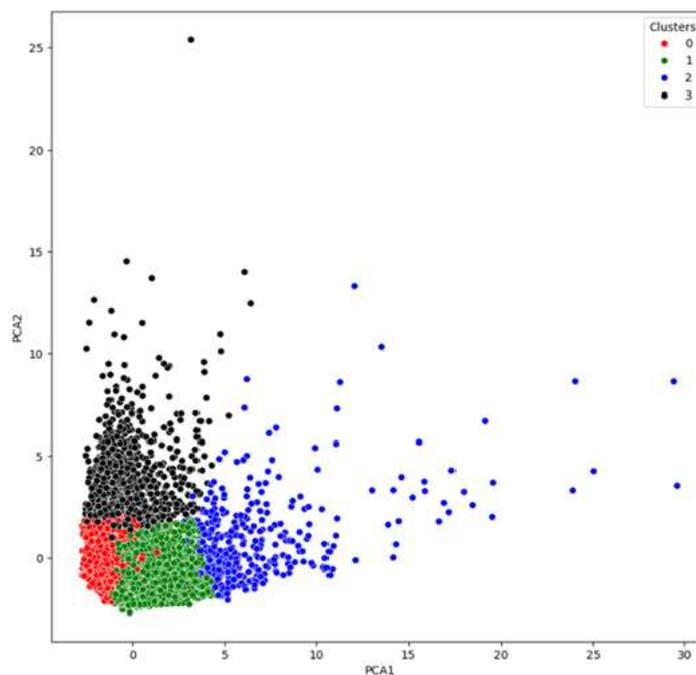


Figure 4: K - means Clustering

It has been found that customer segmented in cluster 0 or red are low balance most customers and majority have balances under 5,000. Their balance frequency concentrated near 1.0, indicates regular balance usage that is very active. The purchases are also very low or most customers have near-zero purchases. From this we can make the interpretation that they have low balance, low purchases but frequently using balance or small but frequent transaction customers. They are budget-conscious or daily-use segment (Figure 5).

The customer segmented in cluster 1 or green have slightly more balance than cluster 0 but still

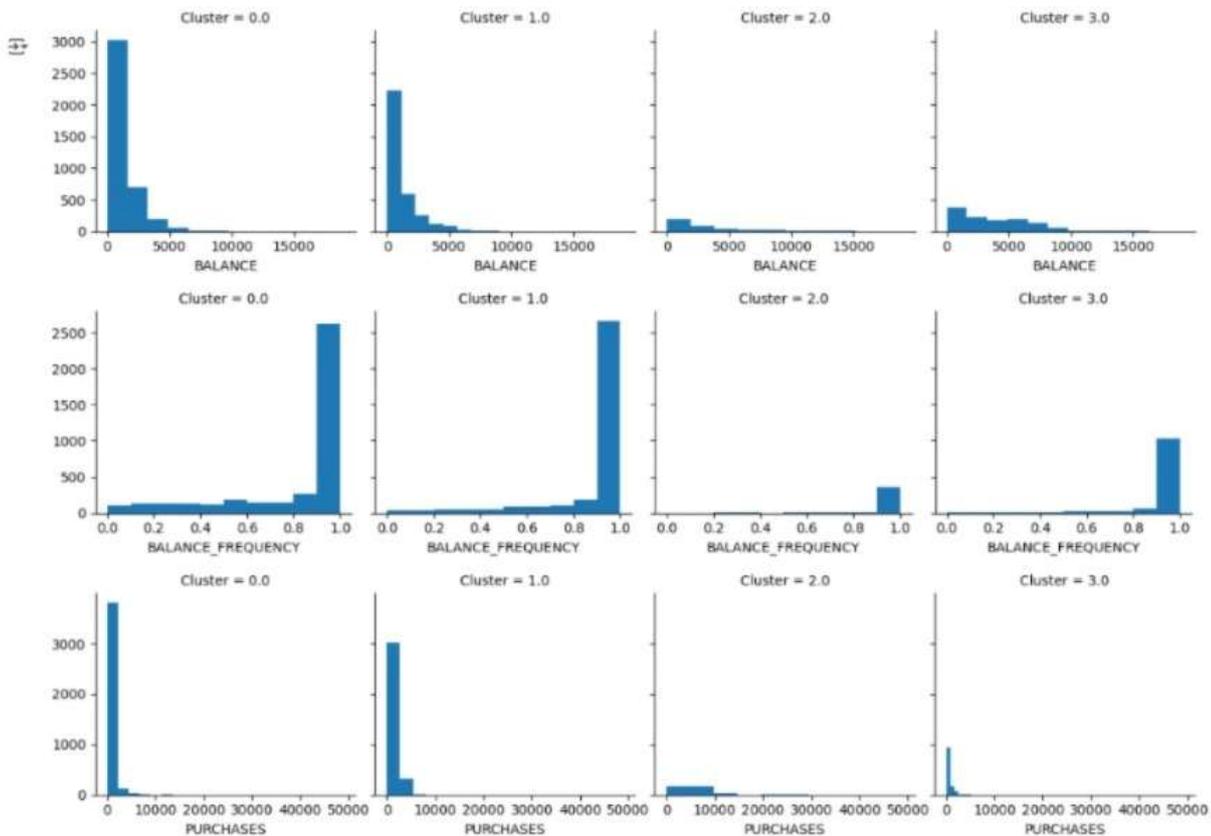


Figure 5: Cluster Representation - Balance, Balance Frequency and Purchase

low overall. Their balance Frequency is very high that is near 1.0 and frequent usage. The purchases are mostly near zero. From this we can make the interpretation that they are similar to Cluster 0 but slightly higher balance and frequent users.

The customer segmented in cluster 2 or blue are spread across the full range and have higher balances. Their balance Frequency varies means some are frequent users and some are less so. Thus, their purchases skewed low but includes customers with higher purchase activity. From this we can make the interpretation that they are high potential customers, possibly diverse spending behavior.

The customer segmented in cluster 3 or black have similar pattern to Cluster 2, with some customers holding very high balances. Their balance frequency is around 0.8 for many and are moderately active. Their purchases are mostly low but a few high-spending users exist. From this we can make the interpretation that they are less frequent but high-balance users that is wealthy but occasional spenders.

CONCLUSION

Customer segmentation by leveraging machine learning techniques such as K -means clustering combined with Principal Component Analysis for dimensionality reduction and decision tree classification for validation is a systematic and data-driven approach. The findings of this study have practical implications for businesses aiming to personalize marketing strategies, optimize resource allocation, and enhance customer engagement. Despite the encouraging results, some limitations should be acknowledged. Other smart clustering algorithms, such as DBSCAN or hierarchical clustering, could provide alternative perspectives on customer groupings. In conclusion, it is a framework for customer segmentation provide better platform for businesses, companies in today's dynamic market landscape.

REFERENCES

1. Robert L. Thorndike, "Who belongs in the family?", *Psychometrika*, Volume 18, pages 267–276, (1953).
2. MacQueen, J. "Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, 281-297, (1967).
3. Punj, Girish N & Stewart, David W, "An Interaction Framework of Consumer Decision Making," *Journal of Consumer Research*, Journal of Consumer Research Inc., vol. 10 (2), pages 181-196, September (1983).
4. Pearson, K., "Principal Components Analysis", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6, 559, (1901).
5. Hotelling, H., "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, 24, 417-441, (1933).
<http://dx.doi.org/10.1037/h0071325>
6. Breiman, L., Friedman, J., Olshen, R. and Stone, C., "Classification and Regression Trees", Chapman and Hall, Wadsworth, New York (1984).
7. Harrison T. S., "Mapping customer segments for personal financial services," *International Journal of Bank Marketing*, vol. 12, no. 8, pp. 17–25, (1994).
8. Kuusik A., "How to avoid customers leaving: The case of the Estonian telecommunication industry", *Baltic Journal of Management* 4 (1), 66-79 (2009).
DOI:10.1108/17465260910930458
9. Rendon E. et.al. "Internal versus external cluster validation indexes", *International Journal of Computers, Communications & Control (IJCCC)* 5(1), 27-34, (2011)
10. Kotler, P. and Keller, K.L., "Marketing Management", 14th Edition, Pearson Education, (2012).
11. Hiziroglu A., "Soft computing applications in customer segmentation: state-of-art review and critique". *Expert Syst. Appl.* 40(16), 6491-6507, (2013).
12. Brito P.Q. et al., Customer segmentation in a large database of an online customized fashion business. *Robot. Comput. -Integr. Manuf.* 36, 93–100 (2015).
13. Dutta S, Bhattacharya S, Guin K. K., "Data mining in market segmentation: a literature review and suggestions", In: *Proceedings of Fourth International Conference on Soft Computing for Problem Solving: SocProS 2014*, Volume 1, Springer, pp 87–98, (2015).
14. Jha A. et.al., "Customer Segmentation and Personalization in Banking", *Journal of Emerging Technologies and Innovative Research*, Volume 6 (1), P123-129, (2019).
15. Rahman S. A. et.al., "Customer Segmentation and Profiling for Life Insurance using K-Modes Clustering and Decision tree classifier", *International Journal of Advanced Computer Science and Applications*, Vol. 12, No. 9, (2021).
16. Deng Y. et. al., "A study on e-commerce customer segmentation management based on improved K-means algorithm", *Information Systems and e-Business Management*, 18(4), pp. 497–510, (2020).
<https://doi.org/10.1007/s10257-018-0381-3>
17. Alves Gomes M. et.al., "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases", *Information system and e- business management*, Volume 21 (3), pages 1-44, (2023) DOI:10.1007/s10257-023-00640-4
18. Upadhyay R.U. et.al., "A Review of Customer Segmentation Methods", *International Journal of Advanced Research in Science, Communication and Technology*, Volume 3, Issue 2, March (2023).
19. Shiradwade P. et.al, "Customer Segmentation With K-Means Clustering", *International Research Journal of Modernization in Engineering Technology and Science*, Volume 5 (9)P 661-664, (2023).
20. Seethalakshmi K., "Crafting Customer Segmentation Strategies for Advanced recommendation System" *IEEE Explore*, 24-26 July, (2024), Third International Conference on Electrical, Electronics, information and Communication Technologies (ICEEICT), DOI: 10.1109/ICEEICT61591.2024
21. Gopalakrishnan K., "Customer Segmentation using K- means Clustering for targeted marketing in banking", *International Journal of Artificial Intelligence & Machine Learning*, Volume 3, Issue 2, pp. 89-94, July-Dec (2024).