International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Research Paper on Customer Segmentation Using K-Means Clustering And RFM Analysis

Mrs. S. N. Firame (Guide)
Department of Information Technology
Sinhgad Institute of Technology and Science
Pune, India

Mr. Harshvardhan Yadav
Department of Information Technology
Sinhgad Institute of Technology and Science
Pune, India

Mr. Jayesh Kadam
Department of Information Technology
Sinhgad Institute of Technology and Science
Pune, India

Mr. Aditya More
Department of Information Technology
Sinhgad Institute of Technology and Science
Pune, India

Mr. Vinay Wakchaure
Department of Information Technology
Sinhgad Institute of Technology and Science
Pune, India

***Abstract*—** Customer segmentation, a crucial step in data driven marketing and personalization, entails breaking up a customer base into discrete groups that share characteristics or behaviors. Simple demographic divisions, like age or income, have given way to more intricate behavioral and psychographic groupings, influenced by preferences, lifestyle choices, and purchasing patterns, as businesses work to better understand their clientele. Businesses can now create highly refined customer segments using advanced analytical techniques and vast amounts of data thanks to technological advancements. These segments inform personalized customer journeys, product recommendations, and targeted marketing campaigns. This section talks about the increasing importance of customer segmentation and how machine learning has increased the range of possible uses for it, allowing businesses to make well-informed decisions supported by data.

INTRODUCTION

The strategic practice of customer segmentation entails breaking up a client base into discrete groups based on shared traits, habits, or requirements. Businesses that want to target particular customer segments with customized marketing, product offerings, and customer engagement strategies must use this segmentation. In the past, segmentation was based on demographic information such as location, income, and age, which gave a broad picture but frequently lacked specificity. More complex segmentations based on behavioral patterns, past purchases, and psychographics are now possible thanks to the revolution in data science and machine learning. Businesses can now analyze large and complex datasets using sophisticated machine learning models like k-means clustering, decision trees, and neural networks, which 1 uncover patterns and relationships that lead to more individualized and successful marketing campaigns. Customer segmentation is becoming a more dynamic and real-time practice as data availability and processing power increase, giving businesses crucial insights into the wants and needs of their customers. In the era of data driven marketing, this paper examines the development, methods, and uses of customer segmentation, emphasizing the contribution of machine learning to improving segmentation impact and accuracy.

LITERATURE REVIEW

Customer Segmentation and Targeting by Data Science Method (2020) - With an emphasis on increasing profitability through sophisticated algorithms, Inwook Moon investigates the application of quantitative data science techniques for customer segmentation in this study. The approach incorporates k-means clustering, a well-liked segmentation technique that reveals unique customer profiles by grouping customers according to shared characteristics. Neural networks and other supervised learning algorithms are also used to forecast consumer behavior, which is useful for focused advertising campaigns. The main conclusions of the study highlight how using data science in segmentation can greatly increase a business's profitability by offering useful insights into the most lucrative client segments. Businesses can identify and rank high-value clients and make well-informed strategic decisions by enhancing data management procedures and utilizing advanced analytical tools. Creating customized marketing strategies and increasing overall business efficiency are two advantages of this segmentation technique. Moon's research demonstrates how data driven approaches are essential for comprehending consumer preferences and matching goods and services to market

needs, which eventually results in a competitive edge.

Customer Segmentation Using Machine Learning (2022) - Razia Sultana A, Anukriti Jaiswal, Supraja P., and Sairamesh L. examine a number of machine learning clustering techniques for customer segmentation in this study. Their research offers a comparative analysis, analyzing the advantages and disadvantages of various approaches to ascertain whether they are appropriate for customer segmentation. Case studies from the banking and retail industries are included in the study to demonstrate how clustering has been effectively used to improve customer engagement. The efficiency of k-means clustering, a method notable for its capacity to group clients with comparable buying habits and preferences, is among the main conclusions. Businesses can use data-driven marketing strategies to target more individualized interactions with particular customer groups by segmenting their customer base in this manner. Businesses hoping to improve customer satisfaction through more focused marketing initiatives will find this study especially helpful. The authors demonstrate how applying machine learning to segmentation not only aids in comprehending a range of client needs but also provides useful, tangible advantages for companies looking to gain a competitive edge.

Customer Segmentation Using K-means Clustering (2021) - This paper, which was written by Rahul Shirole, Laxmiputra Saloke, and Saraswati Jadhav, looks at how k-means clustering is used in customer segmentation, especially in sectors like telecommunications and internet banking. Data collection, preprocessing (to address problems like null values and outliers), and the Elbow Method—which establishes the ideal number of clusters—are all part of the methodology. The authors conclude that k-means is frequently more effective for large datasets after comparing it with other clustering techniques like K Medoids, CLARA, and PAM. By helping businesses to comprehend particular behaviors and preferences within their clientele, this study demonstrates the value of k-means clustering in creating focused marketing strategies. Accurate customer segmentation helps businesses increase customer satisfaction and retention, which makes it a very useful strategy for companies looking to improve their market presence by better understanding the needs of their customers.

Customer Segmentation Using Machine Learning (2021) - Using a variety of machine learning clustering techniques, including k-means, V. Vijilesh, A. Harini, M. Hari Dharshini, and R. Priyadarshini offer a thorough method for customer segmentation in this study. Their study highlights the benefits of employing a variety of clustering techniques to offer a comprehensive understanding of consumer preferences and behavior. The study attempts to capture more intricate patterns in customer data by combining various approaches, which can be challenging to find with a single technique. Businesses that need 2 in-depth knowledge of various client segments in order to create successful strategies will find this method especially helpful. The results of the study show that multi-method clustering promotes improved strategic development, enabling businesses to tailor their products and maximize marketing. A better understanding of customer profiles is facilitated by this combined method approach, which also makes decision-making and customer engagement more efficient. In the end, the study emphasizes how crucial multifaceted and adaptive segmentation models are for companies looking to improve customer satisfaction and marketing efficacy..

METHODOLOGY

1. Data Collection & Preprocessing

The dataset includes customer purchase history with attributes such as customer ID, invoice date, invoice number, and purchase amount. Steps include:

Handling null values

Removing duplicates

Formatting date fields

Converting monetary values into usable numerical form

2. RFM Analysis

Three key features were calculated:

Recency: Days since last purchase

Frequency: Number of purchases

Monetary: Total purchase value

Each customer was scored from 1 to 5 in each category (R, F, M), leading to a composite RFM Score.

3. K-Means Clustering

Using the Elbow Method, the optimal number of clusters was selected.

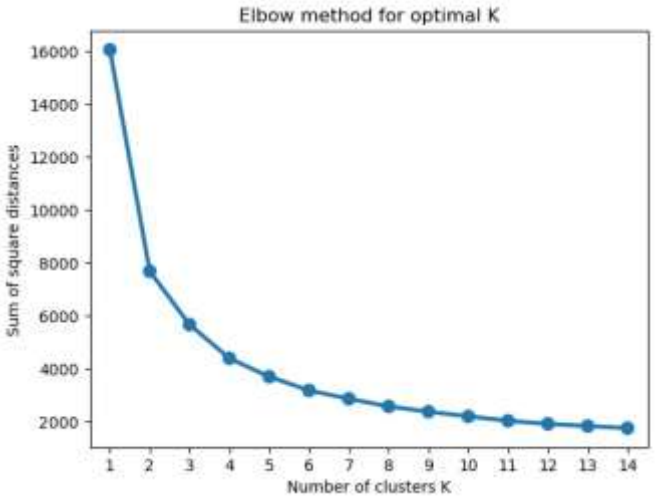

Fig. 1: Elbow Method Graph

K-Means clustering was applied on normalized RFM values.

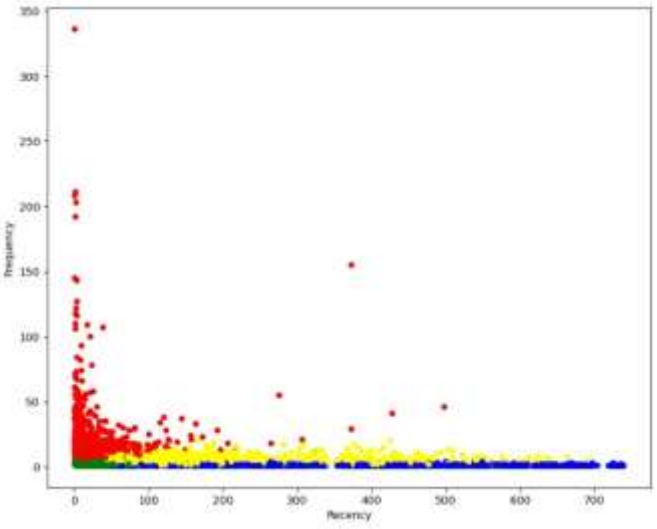

Fig. 2: Customer Cluster Visualization

Figure 3: Signup Page

Figure 4: Login Page

Figure 5: Forgot Password Page

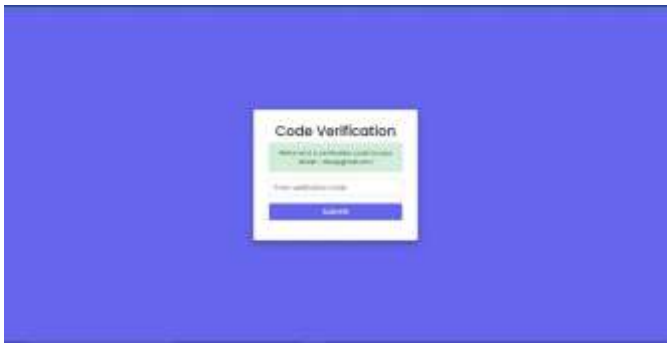

Figure 6: Code Verification Page

Figure 7: PHP XAMPP Server Screenshot

Figure 8: Dashboard View 1

Figure 9: Dashboard View 2

Figure 10: K-Means Info Page

4. Cluster Labeling

Clusters were labeled based on RFM patterns such as:

Loyal Customers

At Risk

Potential Loyalists

New Customers

PROBLEM ANALYSIS

While K-Means Clustering is effective, there are limitations:

1. Static RFM Values: Dynamic customer behavior may not be captured in real time.

2. Elbow Method Subjectivity: Optimal cluster number determination may vary.

3. Feature Scaling Dependency: Clustering quality heavily relies on normalization.

4. Lack of Demographic Insight: Segmentation is behavior-based only.

RESULT AND DISCUSSION

Post-analysis, the clusters revealed insightful patterns:

Cluster 0: High monetary and frequency scores—"Loyal Customers"

Cluster 1: Low recency, low frequency—"Churned Customers"

Cluster 2: Medium scores—"Potential Loyalists"

The model allows marketing teams to allocate budgets more efficiently and improve retention strategies.

FUTURE SCOPE

Integration with real-time dashboards for dynamic segmentation

Use of additional algorithms like DBSCAN or Hierarchical

Clustering

Combine demographic and behavioral data

Recommendation systems based on segments

CONCLUSION

Businesses can still deliver individualized experiences that boost customer satisfaction and loyalty by using customer segmentation, which is still a powerful tool. This study examined both conventional and cutting edge machine learning methods, emphasizing their uses and advantages in practical contexts. Business growth is being driven by segmentation, which is becoming more dynamic and flexible as organizations embrace more advanced data analytics and artificial intelligence capabilities. In the end, successful customer segmentation enables companies to make well-informed, fact-based decisions that suit the requirements and preferences of their clients.

ACKNOWLEDGEMENT

We are deeply grateful to **Dr. A. A. Yadav**, Head of Department, for his steady support and invaluable guidance throughout this work. We also thank **Mrs. Gayathri Ganesan**, our project mentor, whose expert advice and encouragement were essential to completing the Customer Segmentation study. We appreciate our institute for providing the resources that made this research possible, and we acknowledge the collective effort of our team-mates — Harshvardhan Yadav, Jayesh Kadam, Aditya More**,** Vinay Wakchaure — whose dedication and collaboration turned this project into a success.

REFERENCES

1. Inwook Moon, "Customer Segmentation and Targeting by Data Science Method," 2020.
2. Razia Sultana A. et al., "Customer Segmentation Using Machine Learning," 2022.
3. Rahul Shirole et al., "Customer Segmentation Using K-Means Clustering," 2021.
4. Vijilesh et al., "Customer Segmentation Using Machine Learning," 2021.
5. Tushar Kansal et al., "Customer Segmentation Based on RFM Model," IEEE, 2018.
6. Chaohua Liu, "Customer Segmentation Based OnRFM," IEEE, 2011.
7. A.S.M. Shahadat Hossain, "Customer Segmentation using Centroid Based and Density Based Clustering," IEEE, 2017.
8. Nurma Sari et al., "Review on Customer Segmentation Technique on Ecommerce," 2016