Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250402046
Volume 4, Issue2 (May-August 2025), PP: 349-352.
www.indjcst.com
ISSN No: 2583-5300

Analysis of Customer Segmentation Using Unsupervised Learning

Adin Mohan[1], Shreyas J[2], Yatheen[3]
[1,2,3] *Department of Computer Science Engineering-Data Science, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India.*

To Cite this Article: *Adin Mohan[1], Shreyas J[2], Yatheen[3], "Analysis of Customer Segmentation Using Unsupervised Learning", Indian Journal of Computer Science and Technology, Volume 04, Issue 02 (May-August 2025), PP: 349-352.*

Abstract: *In today's data-driven world, understanding customers better is key to delivering personalized experiences and staying competitive. This project dives into how unsupervised machine learning, especially clustering techniques, can be used to group customers based on their behavior and predicted future actions. We focus on behavioral and predictive segmentation—two powerful approaches that help uncover what customers do and what they're likely to do next.*

The process starts with collecting rich customer data, including their interactions, purchases, demographics, and preferences. After cleaning and preparing the data through feature engineering, we apply clustering algorithms like K-Means, Hierarchical Clustering, and DBSCAN to uncover natural patterns and customer groups—without needing any predefined labels.

The outcome is a set of meaningful customer segments that can be used dynamically or statically across various business functions. These insights help companies tailor their marketing, boost customer engagement, and make smarter business decisions. By combining behavioral signals with predictive models, this approach offers a scalable way to segment customers more intelligently and effectively.

Key Words: *Customer Segmentation, Unsupervised Learning, Behavioral Segmentation, Predictive Segmentation, Clustering Algorithms, Machine Learning, K-Means, DBSCAN.*

I.INTRODUCTION

In today's data-driven market, businesses have access to more customer information than ever before. But the real challenge lies in making sense of that data to understand customer behavior and deliver personalized experiences. This project focuses on solving that challenge using unsupervised machine learning to identify meaningful customer segments based on behaviour and predictive patterns. Customer segmentation using machine learning plays a crucial role in identifying patterns in customer behavior and grouping them based on shared attributes[1]. Unsupervised algorithms such as K- Means, DBSCAN, and Hierarchical Clustering have proven to be effective tools for this purpose, especially when applied to behavioral and transactional data[5]. Integrating clustering with behavioral models like RFM analysis enables businesses to discover high-value customers and design personalized marketing strategies[6]. Furthermore, combining unsupervised and predictive learning techniques, such as genetic algorithms and time-series clustering, enhances the accuracy and adaptability of segmentation models[4]. Overall, machine learning-based segmentation supports data- driven decision-making, improves customer engagement, and contributes significantly to business growth in competitive markets[3].

In this project, we use techniques like K-Means, Hierarchical Clustering, and DBSCAN to analyse and group customers without the need for predefined labels. These clustering methods allow us to uncover natural patterns in customer behavior—such as purchasing habits, frequency, and spending—which are captured through RFM (Recency, Frequency, Monetary) analysis and other derived features.

The entire analysis is carried out using the Python programming language, leveraging powerful libraries such as Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, and SciPy. Development and experimentation are done in Jupyter Notebook and Google Colab, making the workflow both efficient and collaborative. For visualizing the final segmentation results, tools like Power BI or Tableau may also be used optionally.

By combining behavioral and predictive segmentation with proven clustering techniques, this project aims to help businesses gain deeper insights into customer preferences, support data-driven marketing strategies, and deliver more personalized services.

II.LITERATURE SURVEY

A. "Customer Segmentation Using Machine Learning & Hybrid Model", Rajendra Arakh, Raja Shoaib, Sweta Kriplani, Vanshika Yadav, Swechchha Agrawal, IJSDR - International Journal of Scientific Development and Research, Year: 2024.[3]

This paper presents a hybrid approach that combines supervised and unsupervised machine learning techniques for customer segmentation. It emphasizes the importance of using both labeled and unlabeled data to improve segmentation accuracy and adapt to dynamic customer behavior. The methodology integrates clustering algorithms such as K-Means with

supervised models like Decision Trees and Support Vector Machines (SVM), refining segment boundaries through domain knowledge and iterative analysis. The hybrid model demonstrated improved precision and relevance in customer segmentation, helping businesses craft more targeted strategies and deliver personalized customer experiences, ultimately enhancing loyalty and driving growth.

B. "An Approach Based on Data Mining and Genetic Algorithm to Optimizing Time Series Clustering for Efficient Segmentation of Customer Behavior", Hodjat Hojatollah Hamidi, Bahare Haghi, Computers in Human Behavior Reports, Year: 2024.[4]

This research proposes a dynamic time-series-based segmentation model that evolves with customer behavior over time, improving the recognition of changing patterns through optimized feature weighting. Customer transaction histories are modeled as time series, and a Genetic Algorithm is employed to fine-tune feature importance. Spectral Clustering is used for the segmentation task, supported by other techniques such as Hierarchical, Fuzzy C-Means, and K-Means Clustering. The proposed framework significantly outperformed static segmentation models, offering a more adaptable and accurate method for identifying meaningful customer groups and supporting the development of customized marketing strategies.

C. "Customer Segmentation using K-Means Clustering", Hemashree Kilari, Sailesh Edara, Guna Ratna Sai Yarra, Dileep Varma Gadhiraju, IJERT, GITAM University, Year: 2022.[2]

This study explores how machine learning, particularly K- Means clustering, can uncover hidden patterns in customer datasets to assist business decision-making. Customers are segmented based on demographic and behavioral attributes like age, gender, and location. The Elbow Method and Silhouette Score are used to determine the optimal number of clusters for effective segmentation. The results highlight the effectiveness of K-Means in distinguishing customer groups, enabling businesses to tailor their marketing efforts to specific audiences and improve personalization in services and promotions.

D. "Segmenting Bank Customers via RFM Model and Unsupervised Machine Learning", Musadig Aliyev, Elvin Ahmadov, Habil Gadirli, Arzu Mammadova, Emin Alasgarov, ADA University, Year: 2020.[15]

This paper presents a machine learning-based approach to customer segmentation by combining RFM (Recency, Frequency, Monetary) modeling with unsupervised learning techniques to better understand consumer behavior and improve marketing efficiency. The authors analyze two datasets—an online retail dataset and a mall customer dataset—using RFM feature extraction and clustering algorithms such as K-Means, Mini-Batch K-Means, Hierarchical Clustering, Birch, and Spectral Clustering. Preprocessing steps include missing value treatment, normalization, and feature transformation using cohort analysis and correlation heatmaps. Validation of clustering effectiveness is conducted using the Silhouette Score, Davies- Bouldin Index, and Elbow Method. The findings show that RFM-based clustering, particularly with K-Means and k=5, is effective in identifying distinct customer segments. The study concludes that combining demographic features with RFM metrics offers actionable insights, enabling businesses to develop more targeted retention strategies and personalized marketing campaigns**.**

E. "Customer Segmentation using Machine Learning", Razia Sulthana A, Supraja P, Anukriti Jaiswal, Sairamesh L, University of Greenwich, SRMIST, CMU, Anna University, Year: 2024.[1]

This study investigates customer segmentation using unsupervised learning techniques, primarily K-Means and DBSCAN, applied to an e-commerce dataset. The research begins by cleaning and preprocessing customer data, followed by feature engineering based on RFM values (Recency, Frequency, and Monetary). K-means clustering helps categorize customers into spending segments, while DBSCAN identifies outliers. The study concludes that combining these clustering techniques provides a richer understanding of customer types, aiding in targeted marketing, loyalty campaigns, and business growth strategies.

F. "Data-Driven Market Segmentation in Hospitality Using Unsupervised Machine Learning", Rik van Leeuwen, Ger Koole, ScienceDirect, Vrije Universiteit Amsterdam, Year: 2021.[5]

This paper explores market segmentation in the hospitality sector using hierarchical clustering. The authors use a real - world dataset from a hotel chain and perform feature engineering on variables like booking behavior, stay length, and spending. Clustering results are validated using internal evaluation metrics and business interpretability. The conclusion emphasizes that traditional persona-based segmentation is inferior to data-driven approaches, which allow for automated, flexible, and scalable customer understanding across diverse guest types.

G. "Application Development for Customer Segmentation Using an Unsupervised Learning Algorithm", M. Nirmala, M. Shah Makzoom, Hindusthan College of Engineering and Technology, Year: 2023.[14]

This practical study presents the development of a customer segmentation tool using K-means clustering. The authors focus on creating an end-to-end system that automates customer data preprocessing, model training, and cluster visualization. Implemented in Python, the system allows users to upload data and receive segmented output with visual plots. The paper concludes that such tools democratize access to data science capabilities and support small businesses in deploying intelligent segmentation without deep technical expertise.

H. "Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation using Clustering Ensemble and Spectral Clustering", Nouri Hicham, Sabri Karim, Hassan II University of Casablanca, The SAI, Year: 2023.[13]

This comprehensive paper compares various unsupervised clustering techniques, including K-means, spectral clustering, and ensemble methods, to analyze their effectiveness in customer segmentation. The authors use synthetic and real datasets to benchmark clustering quality using silhouette score, Davies-Bouldin index, and Calinski-Harabasz score. Their analysis shows that spectral and ensemble clustering outperform basic K-means in cluster coherence and business interpretability. The paper concludes by recommending hybrid and ensemble models for businesses facing complex, high-dimensional customer data.

III.METHODS

To carry out customer segmentation using unsupervised learning, we followed a structured and practical approach that aligns with industry practices, such as those demonstrated in Amazon's machine learning customer segmentation pipeline. This approach includes data collection, data processing, clustering, and application of the resulting customer segments.

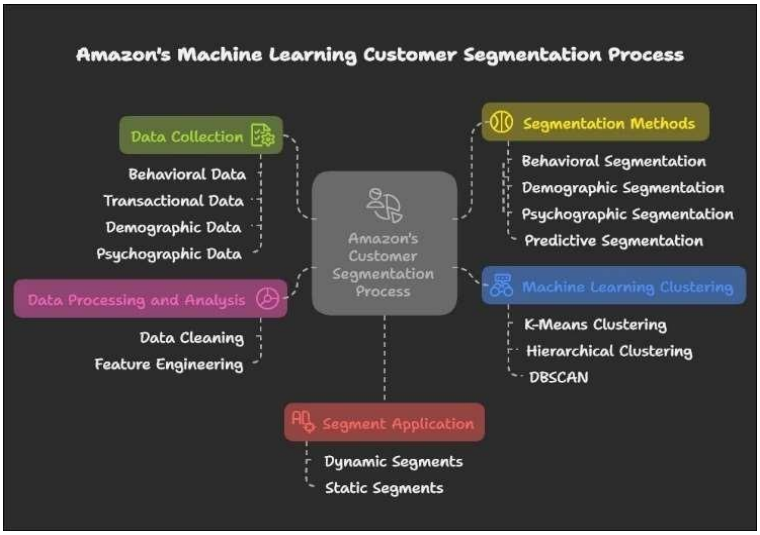

Figure 1: Amazon's Machine Learning Customer Segmentation Process

1. Gathering the Data

The process begins with collecting diverse types of customer data including behavioral data (e.g., browsing habits), transactional data (e.g., purchase history), demographic information (e.g., age, gender), and psychographic attributes (e.g., interests, lifestyle). In some cases, predictive indicators such as churn probability or customer lifetime value are also included to enrich the dataset. These datasets are sourced from open data repositories or generated synthetically to simulate real- world business conditions.

2. Cleaning and Preparing the Data

The raw data undergoes preprocessing to ensure quality and consistency. This involves:

- Removing duplicates and correcting missing or inaccurate entries
- Encoding categorical variables into numerical format
- Normalizing or scaling numerical features
- Engineering new features such as RFM scores or churn likelihood for deeper analysis

3. Applying Clustering Algorithms

With the data prepared, we applied three unsupervised machine learning techniques to identify meaningful customer groups:
K-Means Clustering: Efficiently segments customers into well-defined groups based on similarity
Hierarchical Clustering: Builds a tree-like structure to visualize relationships between clusters
DBSCAN: Identifies clusters of varying shapes and isolates noise or outliers
To optimize the number of clusters, methods such as the Elbow Method, Silhouette Coefficient, and Dendrograms were employed.

4. Enhancing with Predictive Modeling

To improve segmentation accuracy, we integrated basic predictive models like Logistic Regression and Decision Trees. These models predicted future behaviours (e.g., churn), and their outputs were used as features in the clustering step, resulting in forward-looking and actionable segments.

5. Visualizing the Segments

Finally, we used visualization tools like Power BI, Matplotlib, and Seaborn to represent the segmentation results. Visuals included:
5.1 Clear identification of customer groups
5.2 Distinctive behavioral patterns for each segment
5.3 Insights on segment value and potential risk
5.4 Recommendations for targeted business strategies

These visualizations made the complex insights accessible and practical for marketing, customer service, and strategic decision-making.

IV.CONCLUSION

Based on the literature survey conducted on **"Analysis of Customer Segmentation Using Unsupervised Learning"**, it is evident that unsupervised learning techniques are highly effective in identifying meaningful customer segments by uncovering hidden patterns in behavioral and transactional data. The survey highlighted the strengths of various clustering approaches in analyzing attributes such as purchase frequency, recency, monetary value, and time-based trends. As a result, three key methods—K-Means Clustering, DBSCAN, and Hierarchical Clustering—have been selected for implementation in the upcoming project. These algorithms were chosen for their ability to handle different data distributions, discover natural groupings, and support a range of business objectives. In the next phase, these techniques will be applied to real-world customer datasets to generate actionable insights, enabling more targeted marketing strategies, improved personalization, and data-driven decision-making.

References

[1]. Razia Sulthana A, Supraja P, Anukriti Jaiswal, Sairamesh L, Customer Segmentation using Machine Learning, University of Greenwich, SRM Institute of Science and Technology, Carnegie Mellon University, Anna University, Year: 2024.

[2]. Hemashree Kilari, Sailesh Edara, Guna Ratna Sai Yarra, Dileep Varma Gadhiraju, Customer Segmentation using K- Means Clustering, Department of Computer Science Engineering, GITAM University, Andhra Pradesh, India, Year: 2022.

[3]. Prof. Rajendra Arakh, Raja Shoaib, Prof. Sweta Kriplani, Vanshika Yadav, Swechchha Agrawal, Customer Segmentation Using Machine Learning & Hybrid Model, Department of Computer Science & Engineering, Shri Ram Institute of Technology Jabalpur, IJSDR, Year: 2024.

[4]. Hodjat Hojatollah Hamidi, Bahare Haghi, An Approach Based on Data Mining and Genetic Algorithm to Optimizing Time Series Clustering for Efficient Segmentation of Customer Behavior, Computers in Human Behavior Reports, Year: 2024.

[5]. Rik van Leeuwen, Ger Koole, Data-Driven Market Segmentation in Hospitality Using Unsupervised Machine Learning, Ireckonu & Vrije Universiteit Amsterdam, ScienceDirect, Year: 2021.

[6]. Israa Lewaaelhamd, Customer Segmentation Using Machine Learning Model: An Application of RFM Analysis, Department of Business Administration, The British University in Egypt, Year: 2023.

[7]. G. Durga Dhikshith, T. Akash, B. Phani Kumar, Dr. S. S. Subashka Ramesh, Segmentation Study on Customers Data Based on Machine Learning Algorithms, IRJMETS, Year: 2022.

[8]. Yash Parab, Jugal Dave, Customer Segmentation Using Machine Learning: A Comprehensive Research Study, IJNRD, Year: 2023.

[9]. R. Amutha, Aazmaan Ahmed Khan, Customer Segmentation Using Machine Learning Techniques, AMC Engineering College, Bangalore, India, Year: 2023.

[10]. Dr. Y. Suresh, Dr. J. Senthilkumar, Dr. V. Mohanraj, Dr. Selvaraj Kesavan, Customer Segmentation Using Machine Learning in Python, Sona College of Technology, India, Year: 2023.

[11]. V. Vijilesh, A. Harini, M. Hari Dharshini, R. Priyadharshini, Customer Segmentation Using Machine Learning, Kumaraguru College of Technology, IJSRCSEIT, Year: 2021.

[12]. Dr. Huma Lone, Dr. Prajakta Warale, Cluster Analysis: Application of K-Means and Agglomerative Clustering for Customer Segmentation, Rajgad Institute of Management Research, Journal of Positive School Psychology, Year: 2022.

[13]. Nouri Hicham, Sabri Karim, Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation using Clustering Ensemble and Spectral Clustering, Faculty of Legal Economic and Social Sciences AIN SEBAA, Hassan II University of Casablanca, Morocco, The Society of AI and Research, Year: 2023.

[14]. "Application Development for Customer Segmentation Using an Unsupervised Learning Algorithm", M. Nirmala, M. Shah Makzoom, Hindusthan College of Engineering and Technology, Year: 2023.

[15]. Segmenting Bank Customers via RFM Model and Unsupervised Machine Learning", Musadig Aliyev, Elvin Ahmadov, Habil Gadirli, Arzu Mammadova, Emin Alasgarov, ADA University, Year: 2020