



AI-DRIVEN CUSTOMER SEGMENTATION USING CLUSTERING TECHNIQUES

¹Manila Gupta, ²Sidra Amjad Ali Shaikh, ³Sagarika Shekhar Srivastava,

⁴Sidra Sameer Solkar, ⁵Ansari Bushra Zahiruddin

¹Assistant Professor, ²⁻⁵Students

¹Department of Computer Engineering,

¹Rizvi College of Engineering, Mumbai, India

Abstract:

Customer segmentation is a fundamental strategy in contemporary marketing that enables organizations to classify customers into distinct groups for targeted engagement and improved decision-making. This research presents an AI-driven customer segmentation framework that integrates machine learning with intelligent automation to analyze behavioral patterns and generate actionable business insights. The proposed system employs the K-Means clustering algorithm to identify meaningful customer segments based on behavioral and transactional attributes, while the Elbow Method and Silhouette Coefficient are used to determine the optimal number of clusters. An AI module powered by gpt-4o-mini is integrated to automatically interpret cluster results, generate natural-language insights, and produce downloadable summary reports for businesses. The combination of clustering, visualization, and AI interpretation produces a comprehensive analytical system that supports personalized marketing, customer retention, and overall business growth.

Keywords: *Customer Segmentation, Artificial Intelligence, K-Means, Clustering, Machine Learning, Marketing Analytics, GPT-based Analysis.*

I. INTRODUCTION

In today's data-driven economy, understanding customer behavior has become crucial for developing effective marketing strategies. Businesses continuously collect vast amounts of data through digital platforms, and the challenge lies in transforming this data into actionable knowledge. Customer segmentation, which divides a heterogeneous customer base into smaller, homogeneous groups, is an essential component of data-driven marketing [1]. It enables personalized communication, efficient product recommendations, and improved customer satisfaction. Traditional segmentation approaches relied heavily on demographic factors such as age, gender, and income, often limiting analytical depth.

Modern techniques leverage data mining, machine learning, and artificial intelligence to uncover hidden behavioral patterns and deeper relationships between customers [2]. Clustering, an unsupervised learning technique, identifies natural groupings within datasets without predefined labels. Algorithms such as K-Means, Hierarchical Clustering, and DBSCAN have proven effective for segmenting customers based on metrics like spending habits, purchase frequency, and product preferences [3]. This study combines these clustering algorithms with artificial intelligence for automated insight generation. The integration of the gpt-4o-mini model provides AI-driven interpretation and automatic report generation, enabling non-technical users to access downloadable insights through natural-language explanations that bridge the gap between data analytics and business usability.

II. LITERATURE REVIEW

Customer segmentation using clustering techniques has been extensively studied in marketing and data science research. K-Means clustering is one of the most widely used algorithms for dividing customers based on recency, frequency, and monetary value (RFM). Studies show that this method helps businesses identify high-value, potential, and at-risk customers, improving campaign effectiveness and return on investment [3]. Hierarchical clustering has also been applied in marketing analysis, revealing nested structures within customer data and uncovering sub-segments [4].

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is another major clustering method effective for discovering clusters of arbitrary shapes and handling outliers [5]. Techniques such as Principal Component Analysis (PCA) have been used to improve clustering performance and interpretability [6]. Hybrid clustering approaches that combine K-Means and Hierarchical Clustering have demonstrated increased stability and segmentation accuracy [7].

Despite these advancements, most prior research focuses on static datasets and lacks AI-based interpretation or visualization. The present study overcomes these limitations by integrating artificial intelligence for automatic business insight generation and including a feature

that allows users to download AI-generated detailed reports and summaries. This innovation makes customer segmentation more interactive, automated, and business-ready.

III. METHODOLOGY

The proposed methodology integrates unsupervised machine learning with AI-driven analytics to achieve intelligent and interpretable customer segmentation. The dataset used for this study consists of 5,000 e-commerce customer records containing attributes such as age, gender, income, spending score, and purchase frequency. Data preprocessing involved cleaning the dataset, handling missing values, encoding categorical variables, and normalizing numerical features using Min-Max scaling to ensure equal influence across parameters. Feature selection identified the most impactful variables, including total amount spent, purchase frequency, average order value, and days since last purchase.

The K-Means clustering algorithm was applied with $k = 5$, producing five customer segments: Premium, High-Value, Core, Standard, and Basic customers. The optimal number of clusters was determined using the Elbow Method and validated with the Silhouette Coefficient to ensure compact and well-separated clusters. Visualization was achieved using the Recharts library to generate interactive scatter plots, line charts, and pie graphs, providing a clear understanding of customer distribution and behavior across clusters. The clustering results also help identify key trends that guide better marketing and retention decisions. The analysis further enables businesses to predict purchasing patterns and improve personalized recommendation systems.

The system was implemented as a web-based platform with React.js for the front end and Supabase for backend storage. Users can upload a CSV dataset, view clustering outcomes in real time, and explore results through dynamic dashboards. A key feature of the system is the AI module powered by gpt-4o-mini, which automatically generates descriptive insights for each cluster. These insights are compiled into AI-generated detailed reports and summaries that users can download directly from the dashboard. The platform also integrates automated anomaly detection to highlight unusual customer behavior patterns and ensures secure, scalable data handling through Supabase’s authentication and real-time synchronization. Furthermore, the integration of AI-driven insights enhances decision-making by simplifying complex data interpretations for business users.

IV. ALGORITHM USED

The K-Means clustering algorithm was chosen for its simplicity, scalability, and high performance in handling large datasets. K-Means divides the data into k non-overlapping clusters by minimizing the intra-cluster variance. It initializes k centroids randomly and assigns each data point to the nearest centroid based on Euclidean distance. The centroids are then recalculated as the mean of all points in each cluster, and this process iterates until the centroids stabilize. The objective function of K-Means minimizes the total squared distance between data points and their corresponding centroids [1].

To ensure cluster quality, the Elbow Method was used to identify the optimal value of k by analyzing the point where adding more clusters produced marginal improvement in reducing variance. The Silhouette Score was also employed to evaluate cluster separation, with a score close to 1 indicating well-defined clusters. These validation techniques confirmed the reliability of the chosen $k=5$ model.

V. RESULTS AND DISCUSSION



Fig 5.1: Customer Segmentation Dashboard



Fig 5.2: AI generated detailed report

The K-Means clustering produced five distinct customer groups with clear behavioral characteristics. Premium customers displayed the highest spending and purchase frequency, while High-Value customers maintained consistent engagement with moderate spending. Core customers exhibited balanced purchasing habits, Standard customers demonstrated occasional transactions, and Basic customers showed low engagement requiring reactivation strategies. The overall Silhouette Score was 0.65, indicating strong intra-cluster cohesion and inter-cluster distinction.

The integration of the AI-driven gpt-4o-mini module enhanced the analytical capability of the system by automatically generating human-readable summaries and marketing recommendations for each segment. The AI module created downloadable detailed reports and executive summaries, which summarized cluster behavior, customer value, and potential business actions. These reports are automatically formatted and available for download in PDF or text format, enabling marketing teams to use them for campaign planning, presentations, and documentation. This automation significantly reduces manual analysis time while improving interpretability and accessibility for non-technical stakeholders.

Interactive visualization further strengthened understanding by allowing users to explore patterns across dimensions such as total spending versus frequency or category preference versus lifetime value. The integration of machine learning, visualization, and AI-generated reporting provided a complete data-driven decision-support framework for customer management and marketing strategy.

VI. CONCLUSION AND FUTURE SCOPE

This research successfully implemented an AI-driven customer segmentation framework that integrates clustering algorithms, visualization tools, and artificial intelligence. The system effectively identifies meaningful customer segments and transforms complex analytics into actionable insights through AI-generated detailed reports and downloadable summaries. The inclusion of automatic reporting ensures that both technical and non-technical users can easily interpret results and apply them for strategic decision-making.

Future enhancements may include incorporating advanced clustering techniques such as Gaussian Mixture Models (GMM), DBSCAN, or Hierarchical Clustering to capture complex relationships within data. Real-time customer tracking and predictive modeling could be added to forecast churn and customer lifetime value. Additionally, deploying the framework as a SaaS platform would allow enterprises to perform continuous segmentation and automatically download periodic AI-driven reports. Expanding the model to include psychographic and geographic attributes could further improve segmentation depth and personalization capabilities.

REFERENCES

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2011.
- [2] A. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Pearson, 2005.
- [3] R. Ng, K. Smith, and S. Brown, "K-Means Clustering for Retail Customer Segmentation," *J. Retail Analytics*, vol. 8, no. 2, pp. 45–52, 2019.
- [4] L. Zhao and H. Chen, "Hierarchical Clustering in Marketing Analysis," *Int. J. Data Sci.*, vol. 9, no. 3, pp. 100–110, 2020.
- [5] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Proc. KDD*, pp. 226–231, 1996.
- [6] F. Rodriguez and A. Laio, "Clustering by Fast Search and Find of Density Peaks," *Science*, vol. 344, pp. 1492–1496, 2014.
- [7] P. Kumar and S. Gupta, "Hybrid Clustering Techniques for Customer Segmentation," *Adv. Comput. Intell.*, vol. 15, no. 4, pp. 78–88, 2021.