Customer Persona Segmentation using Machine Learning

A. Venkateshwara Rao[1], A. Akhila[2], G. Sravanthi[3], A. Jairam[4], M. Ashrit[5]

Associate Professor[1] and UG Scholar[2,3,4,5]

Dadi Institute of Engineering & Technology, Anakapalle, Visakhapatnam, India

Abstract: *A stylish woman influencer, who always does online shopping 12 hours a day on, online shopping websites like Amazon, Flipkart, Myntra, etc. as she, filters the category of her styling to look more attractive. For the overnights, she faced the problem of filtering every product but which was the trend the product would be out of stock in the market. So, We made to conclude thatthe markets in various industries shouldn't Make the product out of stock or overflow. Analyze the industry and provide enhanced accuracy for improving the strategies of the market which increase profits.*

Keywords: KMeans clustering, Exploratory Data Analysis (EDA)-RFM- Analysis, Stream lit library.

I. INTRODUCTION

Customer segmentation is a crucial aspect of marketing strategies, aiming to divide a heterogeneous customer base into distinct groups with similar characteristics andneeds. This abstract will explore various methodologies used in customer segmentation, including demographic, psychographic, behavioral, and geographic approaches. Additionally, it will discuss the importance of segmentation in tailoring marketing strategies, enhancing customer satisfaction, and ultimately improving business performance. Pivotal factors include:

Business pretensions and objects

- What are you hoping to achieve by segmenting your guests? (e.g., increase deals, ameliorate client retention, epitomize marketing juggernauts)
- What specific business questions are you trying to answer through segmentation? (e.g.," Which client parts are most profitable?"," Who are our most at threat guests?") client Data Collection and Management
- What data do you have available about your guests?(e.g., demographics, purchase history, website gest, check responses)
- How will you collect and organize this data for analysis? (e.g., CRM system, internal databases, selling robotization platforms)
- Data quality is pivotal for accurate segmentation. ensure your data is clean, complete, and up- to- date. Segmentation Bases (Variables)
- What client characteristics will you use to member your guests? These can be Demographic Age, gender, income, position, family size, etc. Behavioural Purchase history, frequency of purchase, product preferences, channel operation, website gest, etc. Psychographic Values, interests, cultures, stations, etc. (frequently requires fresh data sources like check requirements- grounded client requirements and problems that your product or service addresses.
- Value-grounded the value a client contributesto your business in terms of profit, profitability, or continuance value. Segmentation ways There are colorful approaches to segmenting guests, and the stylish choice depends on your data and objects. Some common ways include RFM analysis Recency, frequency, Monetary Value. K-means clustering Groups guests grounded on analogous characteristics. Hierarchical clustering Creates a scale of client parts grounded on similarity. Decisiontrees parts guests grounded on a series of decision rules.

Client Member description and Profiling

- Once you've linked parts, define each member with clear characteristics and a descriptive marker.
- Develop client biographies that paint a picture of each member, including typical demographics, actions, requirements, and values.Practicable perceptivity andoperations
- How will you use the client parts you've linked?
- Develop targeted marketing juggernauts, product recommendations, pricing strategies, and client service approaches acclimatized to each member.
- Measure the effectiveness of your segmentation strategy and upgrade it over time as your business and client base evolve.

II. LITERATURE SURVEY

Customer segmentation is critical in today's competitive landscape. It allows businesses to understand their customer base more effectively and tailor strategies accordingly. By segmenting customers, businesses can: Develop targeted marketing campaigns that resonate with specific segments. Offer products and services that cater to specific customer needs. Optimize pricing strategies for different segments. Enhance customer service experiences based on segment characteristics.

Key Research Areas:

Segmentation Techniques and Algorithms:

- Advanced Clustering Techniques: Explore clustering algorithms beyond K-means and Hierarchical clustering, like Density-Based Spatial Clustering of Applications with Noise (DBSCAN) or Self-Organizing Maps (SOM). These might be suitable for handling complex data structures ornoisy data.
- Supervised Learning for Segmentation: Investigate how supervised machine learning algorithms like Support Vector Machines (SVMs) or Random Forests can be used for customer segmentation by training them on labelled data withdesired customer segments.
- Hybrid Approaches: Evaluate research on combining traditional segmentation techniques with machine learning components for improved segmentation accuracy and insights. 1.7 Segmentation Bases and Customer Insights:
- Beyond Traditional Bases: Research the use ofnon-traditional segmentation bases like social mediaengagement, sentiment analysis of customer reviews, or website clickstream data. These can provide unique insights into customer behavior and preferences.
- Incorporating Customer Lifetime Value (CLV): Explore methods for integrating CLV calculations into segmentation models to identify high-value customer segments for targeted retentionstrategies.
- Dynamic Segmentation: Investigate approaches for building dynamicsegmentation models that can adapt to changingcustomer behaviour over time. This could involveincorporating real-time data streams.

Evaluation Metrics and Validation Strategies:

- Advanced Segmentation Evaluation Metrics: Move beyond basic measures like distinctiveness and delve into metrics likesilhouette coefficient for cluster quality orentropy for information gain in segmentation models.
- Statistical Validation Techniques: Researchstatistical methods like hypothesis testing or model
- selection techniques (e.g., cross- validation)for robust validation of your segmentation models.
- Data Quality and Bias Considerations: Explore the impact of data quality and potential biases on segmentation results. Identify techniquesfor data cleaning and bias mitigation strategies.

Customer Segmentation Applications and Case Studies:

- Industry-Specific Applications: Focus on case studies that demonstrate successful customer segmentation within your specific industry. This helps you understand best practices and potential challenges in your domain.

- Omnichannel Segmentation Strategies: Explore how companies are segmenting customers by considering interactions across multiple channels (online, offline, mobile). This is crucial in today's omnichannel marketing environment.
- Segmentation for Customer Experience (CX) Optimization: Investigate how customer segmentation can be leveraged to personalize and improve customer experiences based on specific segment needs and preferences.

Emerging Trends and FutureDirections:

- Explainable AI (XAI) for Segmentation: Research the use of XAI techniques to understand the reasoning behind complex AI powered segmentation models. This transparency can be valuable for interpreting results and building trust.
- Real-time Customer Segmentation and Personalization: Explore how realtime data streams (e.g., website activity, and social media interactions) can be used for dynamic, real-time segmentation and personalized customer experiences
- Privacy-Preserving Segmentation Methods: Investigate techniquesfor customer segmentation that comply with evolving data privacy regulations like GDPR and CCPA. This ensures responsible handling of customer data.

III. EXISTING SYSTEM

Understand the client rested on their purchasingdata. This analysis will help to deal with terms to target guests with some strategies. A.S.M. Shahadat Hossainetal. Once the categorization of the points is attained, the coming step is to use them to construct the clusters. DBSCAN takes up a core point and also looks atthe points that are inside its Epsilon compass circle and assigns a Cluster marker to those points. As clustering is unsupervised knowledge, you need to dissect each cluster and have a description concerning business data because clustering is always guided by some business rules.

We can also change the value forepsandmin_sample to tune themodel and make clusters more in shapes.

IV. METHODOLOGY

DATASETS:

"Customer Persona Segmentation Dataset"(Kaggle): This dataset provides a large collection of labelled customers for training and testing customer segmentation models.

Data Preparation:

Data Collection: Gather customer data relevant to your segmentation goals. This could include demographics (age, income), purchase history (frequency, amount, product category), website behaviour (clicks, pages viewed), or other relevant datapoints.

Data Cleaning: Ensure your data is clean and free of errors like missing values or inconsistencies. Techniques like imputation or data deletion might be necessary.

Data Transformation: If necessary, transform your data to ensure features are on a similar scale. This is crucial for K-means as it relies on distances between data points. Techniques like normalization or standardization can be used.

Feature Selection:

Identify the most relevant customer features that will contribute meaningfully to segmentation. This helps focus the algorithm and avoid irrelevant data points influencing cluster formation. 1.2.3 Feature selection can be based on domain knowledge, correlationanalysis, or feature importance scores from other machine learning models (if applicable).

Defining the Number of Clusters (k):

K-means requires pre-defining the number of customer segments (k). There's no single "best" way to determine k, so some exploration is needed. Here are some approaches: 1.2.4 Elbow Method: Plotthe sum of squared distances (within-cluster variance) for different k values. The "elbow" in the curve often indicates the optimal k where adding more clusters doesn't significantly reduce variance.

Silhouette Analysis: This method calculates a silhouette score for each data point, which represents how well it's placed within its assigned cluster. Higher average silhouettescores indicate a better clustering solution.

Domain Knowledge: Business insights aboutyour customer base can also guide the selection ofa reasonable number of segments.

K-Means Algorithm Implementation:

The K-means algorithm follows these steps:

Randomly initialize k centroids: These are the centre points of each cluster, initially chosen atrandom positions within the data space.

Assign data points to nearest centroids: Each data point is assigned to the cluster with the nearest centroid based on distance (usually Euclidean distance).

Recalculate centroids: The centroids arerecomputed as the mean of the points assigned to each cluster.

Repeat steps 2 and 3: These steps are repeated iteratively until the centroids no longer significantly change their positions (convergence).

Evaluation and Interpretation:

Once the algorithm converges, analyse the resulting clusters:

Cluster Characteristics: Describe the key characteristics of each cluster based on the features used for segmentation.
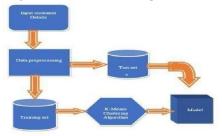
Visualization: Techniques like scatter plots or dimensionality reduction methods (e.g., PCA) can help visualize the separation between clusters.

Business Meaning: Interpret the clusters in the context of your business objectives. How do these segments differ in terms ofcustomer behavior, needs, or value?

Refinement and Iteration:

K-means is sensitive to the initial choice of k. You might need to experiment with different k values and feature selections to find the optimalsegmentation solution.

As you gain more customer data and refine your understanding of your customer base, you might revisit your segmentation strategy and resegment usingK-means or other techniques.

Overview of the Project

The goal of the Data Science Methodology is to provide fundamental answers to the five primary areas of data science projects in a predetermined order. These elementsare:

- Issue to Handle
- Conditions to Gather
- Knowledge to Get Ready
- Modelling to Assess
- Implementation of Feedback
- Train model on the training dataset

But we could we accept these predictions? We can divide our train dataset into two parts, train, and approval, as one method of doing this. We can set up the model for this stage of preparation and include that to create expectationsfor the
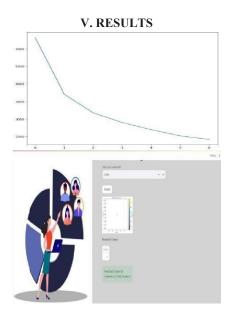
approval phase. Because we have the actual forecasts for the approval portion (which we do not have for the test dataset), we can thus approve our expectations.

V. RESULTS

VI. FUTURE SCOPE

We can evolve our project from a basic K-means segmentation tool to a sophisticated real-time customer segmentation platform with a userfriendly Streamlit interface. This can be a valuable asset for businesses seeking to gain deeper customer insights and deliver a more personalized customer experience.

VII. CONCLUSION

As our dataset was unlabelled, in this project we have opted for internal clustering validation rather than external clustering validation, which depends on someexternal data like labels. Internal cluster validation canbe used for choosing a clustering algorithm that best suits the dataset and can correctly cluster data into its opposite cluster.

Segmentation is an essential tool used by a majority ofbusinesses to understand and categorize customersbased on their spending habits, Customer ID, age, country, gender, etc.

REFERENCES

[1]. Customer Segmentation In Shopping Mall Using Clustering In Machine Learning M.Thirunavakarasu1, Kuncham Pavan Kumar Reddy2, G.Srinivasa Teja3 Https://Www.Irjet.Net/Archives/V9/I3/Irjet-V9i3212.Pdf

[2]. Farrukh, Muhammed Umar, Wainwright, Richard, Crockett, Keeley, McLean, David and Dagnall, Neil (2023) Building Actionable Personas Using Machine Learning Techniques. In: 2022 IEEE Symposium Series on Computational Intelligence (IEEE SSCI), 04 December 2022 - 08

[3]. December 2022, Singapore. https://doi.org/10.1109/SSCI51031. 2022.10022180

[4]. Customer Segmentation Using Machine Learning

[5]. Prof. Nikhil Patankar,1, Soham Dixit, Akshay Bhamare, Ashutosh Darpel, and Ritik Raina Dept. Of Information Technology Sanjivani College of Engineering, Kopargaon423601(MH), India https://www.researchgate.net/public ation/356756320_Customer_Segme ntation_Using_Machi ne_Learning

[6]. Customer Segmentation Using Machine Learning Prof. Nikhil Patankar,1, Soham Dixit, Akshay Bhamare, Ashutosh Darpel, and Ritik Raina Dept. Of Information Technology Sanjivani College of Engineering,

Kopargaon- 423601(MH), India. earning-inpython/https://www.researchgate.net/public ation/356756320_Customer_Segme ntation_Using_Machi ne_Learning

[7]. Farrukh, Muhammed Umar, Wainwright, Richard, Crockett, Keeley, McLean, David and Dagnall, Neil (2023) Building Actionable Personas Using Machine Learning Techniques. In: 2022 IEEE Symposium Series on Computational Intelligence (IEEE SSCI), 04 December 2022 - 08 December 2022, Singapore

[8]. https://doi.org/10.1109/SSCI51031. 2022.10 022180