International Journal of Artificial Intelligence & Machine Learning (IJAIML)
Volume 3, Issue 2, July-Dec 2024, pp. 89-94. Article ID: IJAIML_03_02_006
Available online at https://iaeme.com/Home/issue/IJAIML?Volume=3&Issue=2
Impact Factor (2024): 3.56 (Based on Google Scholar Citation)
Journal ID: 9339-1263, DOI: https://doi.org/10.5281/zenodo.13627403

CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING FOR TARGETED MARKETING IN BANKING

Karthika Gopalakrishnan
Data Scientist, USA

ABSTRACT

Customer segmentation plays a pivotal role in the banking industry, allowing institutions to customize marketing strategies and offers for specific customer segments. This paper investigates the use of the K-Means clustering algorithm, a machine learning technique, to group customers based on crucial financial attributes including account balance, balance checking frequency, purchase patterns, cash advances, and purchase frequency. The objective is to form well-defined clusters that reveal distinct customer profiles. By harnessing these insights, banks can design targeted marketing campaigns and personalized offers, enhancing customer engagement, fostering loyalty, and ultimately driving profitability.

Keywords: Customer Segmentation, Machine Learning, K-Means Clustering, Banking, Targeted Marketing, Customer Engagement.

Cite this Article: Karthika Gopalakrishnan, Customer Segmentation Using K-Means Clustering for Targeted Marketing in Banking, International Journal of Artificial Intelligence & Machine Learning (IJAIML), 3(2), 2024, pp. 89-94.
https://iaeme.com/MasterAdmin/Journal_uploads/IJAIML/VOLUME_3_ISSUE_2/IJAIML_03_02_006.pdf

1. INTRODUCTION

In the fiercely competitive banking industry, gaining a deep understanding of customer behavior is crucial for successful marketing and customer retention strategies. Traditional customer segmentation methods, which primarily rely on demographic factors, often overlook the intricate nuances of customer behavior. However, with the rise of big data and machine learning, banks now have the capability to analyze extensive transactional data to uncover patterns and segment customers into more meaningful and actionable groups. This paper explores the application of the K-Means clustering algorithm for customer segmentation within a banking context, illustrating how this approach can significantly enhance targeted marketing efforts and boost customer engagement.

2. DATASET DESCRIPTION

The dataset used in this study contains 8,950 records and 18 attributes related to customer transactions and financial behavior. Each record represents a unique customer, identified by a customer ID. The key attributes used for segmentation include:

- **Balance**: The current balance in the customer's account.
- **Balance Frequency**: How frequently the customer checks or updates their balance.
- **Purchases**: The total amount spent by the customer.
- **One-Off Purchases**: The amount spent on one-time purchases.
- **Installments Purchases**: The amount spent on installment purchases.
- **Cash Advance**: The amount withdrawn as cash advances.
- **Purchases Frequency**: How often the customer makes purchases.
- **One-Off Purchases Frequency**: The frequency of one-time purchases.
- **Purchases Installments Frequency**: The frequency of installment-based purchases.
- **Cash Advance Frequency**: How often the customer takes cash advances.
- **Credit Limit**: The maximum credit amount the customer is allowed.
- **Payments**: The total payments made by the customer.
- **Minimum Payments**: The minimum amount paid by the customer.
- **Percentage of Full Payment**: The percentage of total payments that were full payments.
- **Tenure**: The duration in months the customer has been with the bank.

This comprehensive dataset provides a detailed view of customer behavior, which is crucial for effectively segmenting customers into meaningful groups.

3. METHODOLOGY

Traditional customer engagement strategies encounter several obstacles

3.1. K-Means Clustering Algorithm

K-Means clustering is a widely used unsupervised machine learning algorithm that partitions data into K distinct clusters. Each cluster contains customers with similar attributes. The K-Means algorithm works as follows:

- Initialization: Randomly select K centroids from the data.
- Assignment: Assign each data point to the nearest centroid, forming K clusters.
- Update: Recalculate the centroids as the mean of all data points in each cluster.
- Repeat: Repeat the assignment and update steps until the centroids stabilize and do not change significantly.

The goal of K-Means clustering is to minimize the variance within each cluster while maximizing the variance between clusters.

3.2. Determining the Value of K Using the Silhouette Method

The value of K, or the number of clusters, is a critical parameter in K-Means clustering. In this study, the optimal number of clusters was determined using the silhouette method. The silhouette method measures how similar each point in one cluster is to points in the nearest cluster, providing an index that ranges from -1 to 1. A higher silhouette score indicates better-defined and more distinct clusters. By applying the silhouette method across a range of K values, the optimal K was identified as 3, which provided the highest average silhouette score for the dataset.

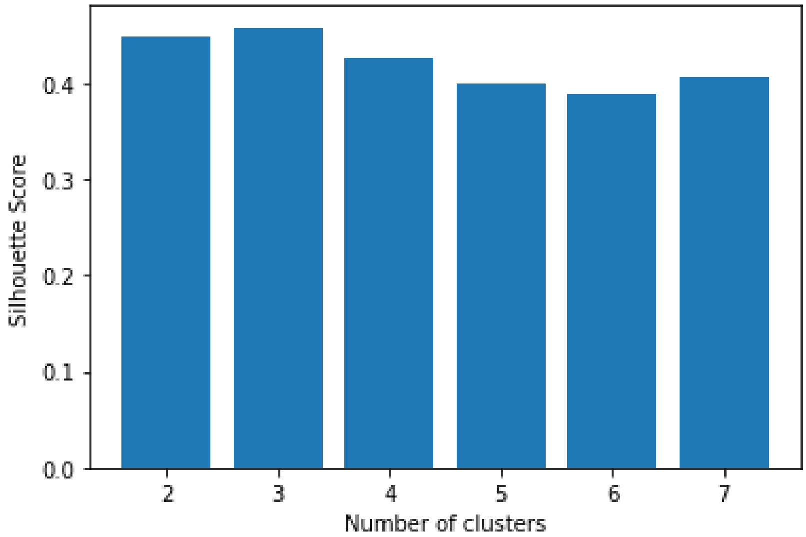

Figure 1: Determination of Clusters using Silhouette Score

1. RESULTS AND DISCUSSION

4.2. Cluster Analysis

Using the K-Means algorithm with K=3, the customers were classified into the following three clusters:

- Cluster 0: Low Balance, Small Spenders, Low Credit Limit

  This cluster represents a large group of customers who maintain low account balances and have low spending levels. These customers generally have the lowest credit limits, indicating a conservative approach to spending. This segment might include individuals who use their accounts primarily for basic transactions and may not rely heavily on credit.

- Cluster 1: Medium Balance, High Spenders, High Credit Limit

  This cluster is smaller in size and comprises customers with medium account balances who engage in high spending activities. These customers have the highest credit limits, which aligns with their spending behavior. They likely represent a more affluent customer base that uses credit for significant purchases and may benefit from premium banking services.

- Cluster 2: High Balance, High Cash Advance, Low Purchase Frequency, High Credit Limit

  This small cluster includes customers with high account balances and frequent cash advances, but with low purchase frequency. These customers also have high credit limits, suggesting that they may use their credit cards as a form of loan rather than for regular spending. This segment might consist of customers who require liquidity for large expenses but do not regularly make purchases.

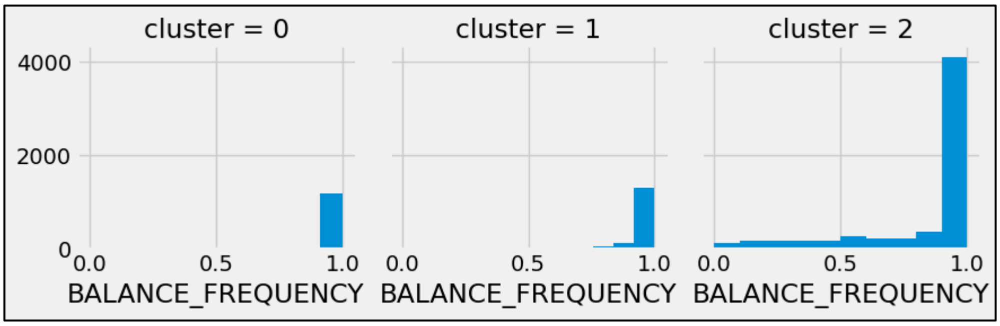

Figure 2: Cluster Representation - Balance Frequency

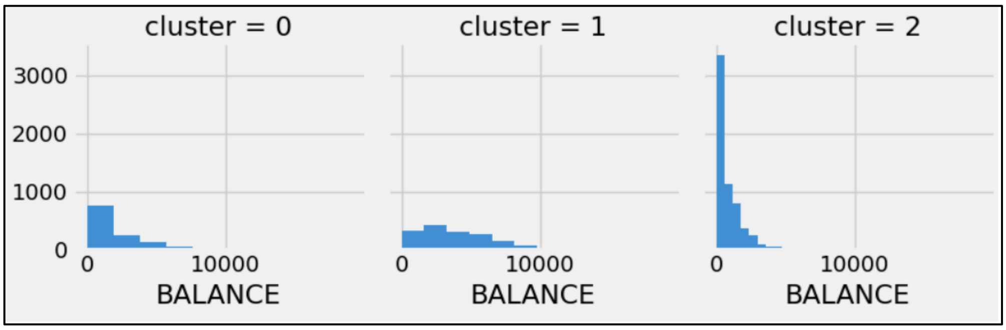

Figure 3: Cluster Representation – Balance

4.2. Implications for Targeted Marketing

The identification of these distinct customer clusters through K-Means clustering allows banks to develop and implement highly targeted marketing strategies. By understanding the unique characteristics and needs of each cluster, banks can tailor their products, services, and marketing messages to resonate more effectively with each group, leading to enhanced customer engagement and satisfaction. Below is an elaboration on how each cluster can be targeted:

- Cluster 0: Low Balance, Small Spenders, Low Credit Limit

  Marketing Focus: Customers in this cluster are typically conservative spenders with low credit limits. Marketing strategies for this group should emphasize basic financial products that align with their conservative spending habits. For instance, promoting low-interest credit cards, basic savings accounts, and budgeting tools can appeal to this segment. Banks can also encourage these customers to engage more with the bank's services by offering rewards for increasing their balance or frequency of transactions, thus fostering greater loyalty.

- Cluster 1: Medium Balance, High Spenders, High Credit Limit

  Marketing Focus: This cluster includes customers with higher spending power and higher credit limits. These customers are more likely to be interested in premium financial products and services. Banks can target this segment with personalized offers for high-reward credit cards, exclusive rewards programs, and investment products. Additionally, offering tailored financial advisory services can help these customers manage their finances more effectively, further increasing their loyalty and lifetime value to the bank. Special offers like cashback on high-value purchases or invitations to exclusive events could also resonate well with this group.

- Cluster 2: High Balance, High Cash Advance, Low Purchase Frequency, High Credit Limit

  Marketing Focus: Customers in this cluster are characterized by high balances, frequent cash advances, and low purchase frequency. This behavior suggests that these customers may be using their credit cards primarily as a source of liquidity rather than for everyday purchases. Marketing efforts for this segment could focus on promoting products that support their need for liquidity, such as personal loans with flexible repayment terms, lower interest rates on cash advances, or balance transfer offers. Banks can also consider offering financial planning services that help these customers manage their cash flow more effectively, potentially encouraging them to increase their engagement with other banking products.

5. CONCLUSION

The application of K-Means clustering to customer segmentation in the banking sector demonstrates the value of machine learning in identifying distinct customer profiles. By leveraging these insights, banks can implement targeted marketing strategies that cater to the specific needs and behaviors of different customer groups. This targeted approach not only enhances customer engagement but also fosters loyalty and drives profitability by ensuring that customers receive relevant offers and services that resonate with their financial habits and needs.

REFERENCES

[1] Kotu, V., & Deshpande, B. (2014). Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner. In 2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr) (pp. 57-64). IEEE. doi:10.1109/CIFEr.2014.6924068

[2] Alam, S., Shakil, K. A., & Sethi, N. (2019). Analysis and Comparison of Various Machine Learning Techniques for Customer Segmentation. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 208-214). IEEE. doi:10.1109/ICCIKE47802.2019.9004261

[3] Ghelichi, Z., Jalali, A., & Sadeghi, M. (2017). Customer segmentation in banking industry using clustering and data mining techniques. In 2017 7th International Conference on Computer and Knowledge Engineering (ICCKE) (pp. 287-292). IEEE. doi:10.1109/ICCKE.2017.8167902

[4] Rathod, R., & Khanna, A. (2019). A Review on Customer Segmentation Techniques Using Machine Learning and Their Applications. In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 424-428). IEEE. doi:10.1109/CONFLUENCE.2019.8776901

[5] Farzaneh, R., & Mohammadi, S. (2017). Customer segmentation using K-means and K-medoids clustering algorithms: A case study of a retail chain. In 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 1108-1112). IEEE. doi:10.1109/IEEM.2017.8290044

Citation: Karthika Gopalakrishnan, Customer Segmentation Using K-Means Clustering for Targeted Marketing in Banking, International Journal of Artificial Intelligence & Machine Learning (IJAIML), 3(2), 2024, pp. 89-94

Article Link:
https://iaeme.com/MasterAdmin/Journal_uploads/IJAIML/VOLUME_3_ISSUE_2/IJAIML_03_02_006.pdf

Abstract:
https://iaeme.com/Home/article_id/IJAIML_03_02_006

editor@iaeme.com