DOI: https://doi.org/10.63712/bpsrj-v2i1p002

CUSTOMER SEGMENTATION USING MACHINE LEARNING TECHNIQUES

R. Amrutha, Department of Information Science and Engineering, AMC Engineering College, Bangalore

ABSTRACT:

Customer segmentation plays a critical role in enabling organizations to deliver personalized financial products and services. This study presents a machine learning–driven framework for segmenting customers based on demographic attributes, transactional behavior, and usage patterns. The proposed system employs clustering algorithms such as K-means, Gaussian Mixture Models, DBSCAN, spectral clustering, and agglomerative clustering to identify meaningful customer groups.

A dataset comprising usage information from approximately 9,000 active credit card users over a six-month period is utilized. After preprocessing and feature selection, clustering models are trained and evaluated to determine the most suitable segmentation strategy. The resulting customer segments are integrated into a Streamlit-based web application that provides customized recommendations related to savings plans, loan options, and wealth management strategies.
The experimental results demonstrate that machine learning–based segmentation enhances the understanding of customer behavior and supports data-driven decision-making for financial institutions. The proposed approach improves customer satisfaction, strengthens engagement, and provides a scalable foundation for personalized financial advisory systems

Keywords: Customer Segmentation, Machine Learning, Clustering Algorithms, Python, Streamlit, K-Means.

INTRODUCTION:

In modern competitive markets, understanding customer diversity is essential for delivering personalized services. Customer segmentation refers to the process of dividing a customer base into distinct groups that share similar characteristics, behaviors, or preferences. These characteristics may include demographic factors, purchasing patterns, financial behavior, and online interactions.

Traditional segmentation techniques often rely on predefined rules and manual analysis, which may fail to capture complex relationships within large datasets. The rapid growth of digital transactions and customer data has increased the need for automated and data-driven segmentation methods. Machine learning techniques, particularly unsupervised learning algorithms, provide an effective solution by identifying hidden patterns and similarities among customers.

In the financial domain, customer segmentation enables institutions to design targeted marketing strategies, improve customer retention, and offer customized financial products. By leveraging machine learning, segmentation becomes more

DOI: https://doi.org/10.63712/bpsrj-v2i1p002

accurate, scalable, and adaptable to changing customer behavior. This study focuses on applying clustering algorithms to credit card usage data and integrating the results into an interactive recommendation system

AIM AND OBJECTIVES

Aim

The primary aim of this study is to develop an effective customer segmentation framework using machine learning techniques to classify credit card users based on their behavioral and financial patterns, and to provide personalized financial recommendations through an interactive application.

Objectives

The specific objectives of the study are:

- To collect and preprocess customer credit card usage data for analytical modeling.
- To identify significant behavioral and financial features influencing customer segmentation.
- To apply and compare multiple machine learning clustering algorithms for customer classification.
- To determine the optimal number of customer segments using appropriate validation techniques.
- To analyze and interpret the characteristics of each customer segment.
- To design and deploy a Streamlit-based application for delivering personalized recommendations related to savings, loans, and wealth management

LITERATURE REVIEW

Customer segmentation has been widely studied across marketing, finance, and data analytics domains. Previous research highlights that treating all customers uniformly is ineffective due to variations in preferences, income levels, spending behavior, and financial goals. As competition intensifies, organizations increasingly rely on customer profiling to enhance retention and profitability.

Several studies emphasize the role of data-driven techniques in improving segmentation accuracy. Machine learning approaches, particularly clustering methods, have been shown to outperform traditional statistical techniques by uncovering non-linear relationships in customer data. Researchers have explored various algorithms, including K-means, hierarchical clustering, Gaussian mixture models, and density-based methods, each offering distinct advantages depending on data structure and application requirements.

Big data analytics has further expanded the scope of customer segmentation by enabling the processing of large-scale, high-velocity datasets. The integration of transactional data, behavioral metrics, and demographic attributes has improved predictive capabilities and personalization outcomes. Recent studies also highlight the importance of combining segmentation with customer lifetime value analysis and personalized targeting to enhance marketing effectiveness.

Overall, the literature confirms that machine learning–based segmentation provides a robust foundation for personalized financial services and customer relationship management

MATERIAL AND METHODS:

Data Source

The dataset used in this study consists of transactional and behavioral data from approximately 9,000 active credit card customers over a six-month period. The data includes multiple numerical attributes representing spending behavior, payment patterns, credit utilization, and account tenure.

Data Preprocessing

Data preprocessing was performed to ensure accuracy and consistency. Missing values were handled using appropriate imputation techniques, and irrelevant or redundant attributes were removed. Feature scaling was applied to normalize numerical variables, enabling fair distance computation during clustering.

Feature Selection

Exploratory data analysis was conducted to identify meaningful features contributing to customer behavior. Key attributes such as balance, purchase frequency, installment purchases, cash advances, credit limit, payments, and tenure were selected for modeling.

Machine Learning Techniques

Unsupervised machine learning algorithms were employed for customer segmentation, including:

- K-Means Clustering
- Gaussian Mixture Model (GMM)
- DBSCAN
- Agglomerative Clustering
- Spectral Clustering

Among these, K-Means clustering was primarily used due to its simplicity, scalability, and interpretability.

Model Optimization

The Elbow Method was applied to determine the optimal number of clusters by analyzing within-cluster sum of squares (WCSS). Model performance was evaluated through cluster visualization and interpretability of segment characteristics.

System Implementation

The final segmentation model was integrated into a Streamlit-based web application. The application enables users to input customer details and receive personalized financial recommendations based on their assigned segment

PROBLEM STATEMENT

Financial institutions manage a diverse population of credit card users with varying spending habits, repayment behaviors, and financial needs. Understanding these variations is challenging due to the volume and complexity of customer data.

DOI: https://doi.org/10.63712/bpsrj-v2i1p002

This study aims to analyze the usage patterns of approximately 9,000 active credit card users over a six-month period and classify them into distinct customer segments using machine learning techniques. The objective is to identify meaningful behavioral patterns that can support personalized recommendations related to loans, savings, and wealth management.

An additional challenge lies in translating analytical results into actionable insights. To address this, the proposed system integrates the segmentation model into a Streamlit-based application that interacts with users and provides recommendations tailored to their assigned segment.

PROPOSED METHODOLOGY

The proposed methodology follows a systematic machine learning pipeline for customer segmentation and recommendation generation.

Data Collection and Preprocessing

Customer data is collected from multiple sources, including transaction records, demographic information, and behavioral metrics. The dataset is cleaned to handle missing values, inconsistencies, and outliers. Feature engineering techniques are applied to derive meaningful variables that represent customer behavior effectively.

Model Selection

Several clustering algorithms are evaluated for segmentation, including K-means, DBSCAN, agglomerative clustering, spectral clustering, and Gaussian mixture models. The selection of algorithms is based on data characteristics, scalability, and interpretability.

Training and Validation

The preprocessed dataset is divided into training and validation subsets. Models are trained by tuning hyperparameters and optimizing clustering performance. The elbow method is used to determine the optimal number of clusters for K-means.

Customer Segmentation

The trained models are applied to segment customers into distinct groups. Each cluster represents customers with similar financial behavior and usage patterns.

Interpretation and Visualization

Cluster characteristics are analyzed to understand spending habits, credit utilization, and repayment trends. Visualization techniques are used to present segmentation results clearly and intuitively.

Deployment

The final segmentation model is deployed through a Streamlit web application. Users can input their details and receive personalized financial recommendations aligned with their segment profile.

Continuous Improvement

User feedback and new data are continuously incorporated to refine the segmentation model and improve recommendation accuracy over time.

DOI: https://doi.org/10.63712/bpsrj-v2i1p002

PERFORMANCE ANALYSIS

Input Description

The dataset contains behavioral information for nearly 9,000 active credit card users, represented by 18 numerical features such as balance, purchase frequency, cash advance usage, credit limit, payment history, and tenure.

Output and Results

The elbow method is applied to identify the optimal number of clusters. The clustering results reveal distinct customer segments with varying financial behaviors. Visualization of clustered data confirms clear separation among groups.

The segmented output is used to generate targeted recommendations for savings plans, loan eligibility, and wealth management strategies, demonstrating the practical applicability of the model.

RESULTS:

The experimental evaluation shows that machine learning–based clustering effectively identifies meaningful customer segments. K-means clustering, supported by the elbow method, provides stable and interpretable results. Other clustering methods complement the analysis by handling overlapping clusters and noise.

The deployed Streamlit application successfully delivers real-time, personalized financial recommendations. The results indicate improved customer understanding and enhanced potential for targeted service delivery

DISCUSSION

The results of this study demonstrate that machine learning–based customer segmentation is an effective approach for understanding diverse financial behaviors among credit card users. The clustering algorithms successfully grouped customers into distinct segments based on spending habits, repayment behavior, and credit usage patterns.

K-Means clustering, supported by the elbow method, produced well-defined and interpretable clusters, making it suitable for financial decision-making applications. Other algorithms such as GMM and DBSCAN provided complementary insights, particularly in handling overlapping customer behaviors and outliers.

The integration of segmentation results into an interactive Streamlit application highlights the practical applicability of the proposed framework. Personalized recommendations related to savings plans, loan options, and wealth management strategies enable financial institutions to align their offerings with customer needs and risk profiles.

Overall, the findings indicate that automated customer segmentation enhances customer engagement, improves service personalization, and supports data-driven marketing strategies. The framework can be further extended by incorporating additional behavioral variables and

DOI: https://doi.org/10.63712/bpsrj-v2i1p002

real-time data to improve accuracy and adaptability

CONCLUSION

This study demonstrates the effectiveness of machine learning techniques for customer segmentation in the financial domain. By analyzing credit card usage data, the proposed system successfully classifies customers into distinct segments and provides personalized financial recommendations.

The integration of clustering algorithms with an interactive web application enables financial institutions to align their services with individual customer needs and risk preferences. This personalized approach enhances customer satisfaction, loyalty, and engagement while supporting data-driven business strategies.

Future work may include incorporating additional data sources, advanced deep learning techniques, and real-time analytics to further improve segmentation accuracy and recommendation quality

REFERENCE:

1. Blanchard, T., Bhavnagar, P., & Behera, T. (2019). Marketing Analytics with Python. Packt Publishing.
2. Griva, A., Bardaki, C., Pramatari, K., & Papakiriakopoulos, D. (2018). Customer segmentation using market basket data. Expert Systems with Applications, 100, 1–16.
3. Hong, T., & Kim, E. (2011). Customer classification based on online purchasing behavior. Expert Systems with Applications, 39(2), 2127–2131.
4. Hwang, Y. H. (2019). Hands-On Data Science for Marketing. Packt Publishing.
5. Premkanth, P. (2012). Market segmentation and its impact on customer satisfaction. Global Journal of Management and Business Research, 12(1).
6. McKinsey Global Institute. (2011). Big Data: The Next Frontier for Innovation and Productivity.
7. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. ACM Computing Surveys, 31(3), 264–323.
8. Jiang, T., & Tuzhilin, A. (2008). Improving personalization through optimal customer segmentation. IEEE Transactions on Knowledge and Data Engineering, 21(3), 305–320.
9. Kashwan, K. R., & Velu, C. (2013). Customer segmentation using clustering techniques. International Journal of Computer Theory and Engineering, 5(6), 856–860.
10. Ballestar, M. T., Grau-Carles, P., & Sainz, J. (2018). Customer segmentation in e-commerce. Journal of Business Research, 88, 407–414.

Source of Support: Nil. **Conflicts of Interest:** None