Louis Savinien Dupuis Journal of Multidisciplinary Research, 2025: V(4) 16–17
DOI: 10.21839/lsdjmr.2025.v4.229

Research Article

Customer Segmentation Using Machine LearningClusteringTechniques

A. Abinaya, K. Bairavi, V. Kanimozhi,G.Akshaya
Department of Computer Science, Immaculate College for Women,Cuddalore
(Received: 09-04-2025; Accepted 08-06-2025; Published Online 03-11-2025)

Abstract

Customer segmentation is a critical component of modern business strategy, enabling organizations to better understand customer behaviour and design targeted marketing initiatives. Traditional segmentation methods often rely on manual analysis and limited demographic factors, which restrict scalability and accuracy. This paper presents a machine learning–based approach for customer segmentation using unsupervised clustering techniques. Algorithms such as K-Means, DBSCAN, and Hierarchical Clustering are applied to customer data containing features such as age, annual income, and spending behaviour. The proposed system automates segmentation, improves scalability, and uncovers hidden patterns within customer data. Experimental results demonstrate that machine learning–based segmentation provides meaningful customer groups that support data-driven marketing decisions and enhance business performance.

Keywords: Customer Segmentation, Machine Learning, Unsupervised Learning, K-Means Clustering, DBSCAN, Business Analytics

Introduction

In today's competitive business environment, understanding customer behaviour is essential for sustainable growth and customer retention. Customer segmentation divides a heterogeneous customer base into homogeneous groups based on shared characteristics such as demographics and purchasing behaviour. This enables businesses to deliver personalized marketing strategies and improve customer satisfaction.

Traditional segmentation methods depend largely on manual analysis and basic statistical techniques, limiting their ability to handle large and complex datasets. With the advancement of machine learning, unsupervised learning techniques have emerged as powerful tools for identifying meaningful customer segments automatically. This paper explores the application of clustering algorithms to segment customers efficiently and accurately.

Background and Related Work

Customer segmentation has long been used in marketing and customer relationship management. Traditional approaches include demographic, geographic, psychographic, and behavioural segmentation. Statistical techniques such as Recency, Frequency, and Monetary (RFM) analysis have also been widely adopted.

Recent studies highlight the effectiveness of machine learning techniques for segmentation due to their ability to process high-dimensional data and discover hidden patterns. Clustering algorithms such as K-Means, DBSCAN, and Hierarchical Clustering have been successfully applied across industries including retail, banking, and e-commerce. However, selecting appropriate algorithms and evaluation metrics remains a challenge.

Machine Learning for Customer Segmentation

Machine learning enables automated and scalable customer segmentation by identifying intrinsic patterns in data without predefined labels. Unsupervised learning techniques are particularly suitable for segmentation tasks. A. Clustering Algorithms

K-Means Clustering: Efficient and simple, partitions data into a predefined number of clusters by minimizing intra-cluster variance.

DBSCAN: Density-based algorithm capable of detecting clusters of arbitrary shape and handling noise.

Hierarchical Clustering: Builds nested clusters and provides dendrogram-based visualization without requiring predefined cluster numbers.

B. Evaluation Metrics

Cluster quality is evaluated using methods such as the Elbow Method, Silhouette Score, and Davies–Bouldin Index to determine optimal cluster selection.

Proposed System

The proposed system applies machine learning– based clustering for customer segmentation using numerical features such as age, annual income, and spending score.

A. System Architecture

The workflow consists of data collection, preprocessing, clustering, evaluation, visualization, and insight generation.

B. Data Preprocessing

Data cleaning, handling missing values, normalization, and feature scaling are performed to ensure accurate clustering results.

C. Visualization

2D and 3D scatter plots are used to visualize customer segments and cluster centroids for business interpretation.

Implementation and Methodology

The Mall Customers Dataset is used for experimentation. Exploratory Data Analysis (EDA) is conducted to understand data distribution and feature relationships. K-Means clustering is implemented as the primary algorithm, with alternative methods explored for comparison. Cluster validity is assessed using standard evaluation metrics.

The methodology involves data collection from a structured dataset, preprocessing through normalization and cleaning, applying K-Means clustering, evaluating optimal cluster numbers, and visualizing results using 2D and 3D plots. The final clusters are analysed to derive business insights.

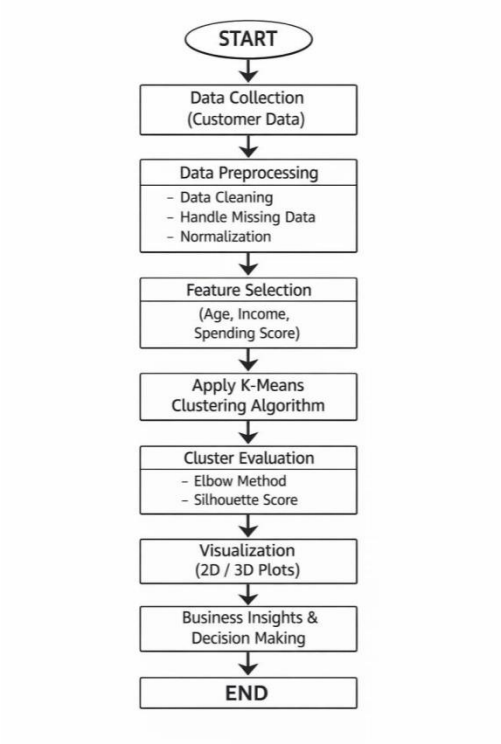


Fig1.1

Results and Discussion

The results demonstrate that machine learning– based clustering effectively segments customers into meaningful groups based on purchasing behaviour and income levels. Visualization confirms clear separation among clusters, enabling actionable business insights such as targeted promotions and personalized offers.

Conclusion

This study demonstrates the effectiveness of machine learning techniques for automated customer segmentation. Compared to traditional methods, the proposed approach offers scalability, adaptability, and improved accuracy. The system supports data-driven marketing strategies and can be extended to various industries and datasets. Future work may include real-time segmentation and integration with live business systems.

References

T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, Springer, 2009.

P. Kotler and K. Keller, Marketing Management, Pearson Education, 2016.

J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2012

Jain, “Data clustering: 50 years beyond K-means,” Pattern Recognition Letters, vol. 31, no. 8, pp. 651–666, 2010.

I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. San Mateo, CA, USA: Morgan Kaufmann, 2016.

J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.

L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, USA: Wiley, 2009.

M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 1996, pp. 226–231.

R. Xu and D. Wunsch, “Survey of clustering algorithms,” *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, May 2005.

P. Kotler and K. L. Keller, *Marketing Management*, 15th ed. Upper Saddle River, NJ, USA: Pearson, 2016.

C. Aggarwal, *Data Mining: The Textbook*. Cham, Switzerland: Springer, 2015.

M. Wedel and W. A. Kamakura, *Market Segmentation: Conceptual and Methodological Foundations*, 2nd ed. Boston, MA, USA: Kluwer Academic Publishers, 2000.

P. Harrigan, E. Evers, M. Miles, and T. Daly, “Customer engagement and the relationship between involvement, engagement, self-brand connection and brand usage intent,” *Journal of Business Research*, vol. 88, pp. 388–396, Jul. 2018.

R. Kohavi and F. Provost, “Glossary of terms,” *Machine Learning*, vol. 30, no. 2–3, pp. 271–274, 1998.

F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.