International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

AI-Based Retail Business Analysis using K-Means Clustering

Gurushankar S[1] , Dr.B.Leelavathi[2]

[1] Undergraduate Student [2] Professor Department of Computer Technology, Dr.N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India

E-Mail:guruselvi1206@gmail.com , get2leelavathi@gmail.com

https://doi.org/10.55041/ijsmt.v2i3.192

Cite this Article: S, G. (2026). AI-Based Retail Business Analysis using K-Means Clustering. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.192

Abstract

Artificial Intelligence (AI) has been identified as an emerging technology in the retail industry that has the potential to offer data-driven insights into customer purchase behavior. Today's retail environment is capable of storing large amounts of customer purchase behavior in the form of customer transactional data. However, it is not possible to leverage statistical methods to perform analysis on large customer transactional data. This paper aims to perform an AI- based analysis of real customer purchase behavior data consisting of 1,800 transactions collected during 2023-25. Machine learning methods will be used to perform an analysis of customer purchase behavior. Customers will be segmented into high-value, medium-value, and low-value segments using K-Means classification. This paper aims to show that AI-based customer segmentation is useful in decision-making in the retail industry, as suggested by previous studies on AI-based customer segmentation.

Keywords: Artificial Intelligence, Retail Analytics, Customer Purchase History, Machine Learning, Customer Segmentation.

I. INTRODUCTION

The retail sector has witnessed a rapid digital revolution fueled by the adoption of Artificial Intelligence (AI), big data analytics, and automation technologies. The availability of point-of-sale (POS), loyalty programs, mobile apps, and e-commerce sites has resulted in the generation of large volumes of customer purchase history data. The data contains rich details about customer behavior, frequency of purchases, product preferences, and spending habits. The key challenge lies in extracting useful insights from large volumes of data using traditional statistical and rule-based approaches.The traditional retail analytics practices have faced several limitations in handling larger data sets, unstructured data, and in identifying non-linear relationships between variables. With the ever- increasing competition in the retail industry and the changing nature of customer behavior, there is a need for better analytical practices that can provide better and quicker results, thereby facilitating better sdecision-making. Hence, the need of the hour is the development of intelligent data-driven practices that can effectively analyze transactional data.

Artificial Intelligence and machine learning methods help to overcome these challenges effectively through automated data processing, pattern detection, and predictive modeling. AI-assisted analytics help the retailer to identify unseen patterns in customer shopping habits, increase the accuracy of demand forecasting models, manage

inventory effectively, and increase customer engagement through targeted marketing campaigns [1], [2]. Past research studies proved that the implementation of AI in the retail industry results in improved operational efficiency, inventory cost reductions, and customer satisfaction through targeted decision-making [3], [4].

Among the various applications of AI in retail analytics, customer segmentation is a vital area for understanding customer diversity and developing strategic customer clusters. Customer segmentation enables the differentiation of high-value customers from low-value customers through the identification of customer spending patterns, spending value, and frequency of transactions. Unsupervised machine learning algorithms can be effectively applied for customer segmentation as they do not require any labeled information for discovering the natural groupings in the data.

Considering the Indian retail industry, the adoption of AI is gradually increasing with the rapid growth of digitalization, the development of the e-commerce environment, and the usage of digital payment systems. However, the number of studies using AI techniques for real-world customer transactional data in Indian retail environments is limited [5]. Most of the existing studies are based on virtual data or conceptual models, which are not useful for real- world retail environments.

This research aims to contribute to the existing body of knowledge by using AI and machine learning techniques for real-world customer purchase history data with 1,800 retail transactions collected from 2023 to 2025. The objective of the research is to show the potential of AI techniques in customer segmentation using the most commonly used AI technique, K-Means clustering, for customer segments based on specific features of customer behavior, including the monetary value of the purchase, purchase frequency, and the quantity of products bought by the customers.

II. LITERATURE REVIEW

In recent times, Artificial Intelligence (AI) has emerged as an important facilitator in the transformation of retail business analysis. This is because traditional methods of retail analysis face difficulties in processing large volumes of data, whereas AI-based systems can easily process customer behavior, sales trends, and inventory details in real time [1].

According to recent studies, AI-based business intelligence software has been found to greatly contribute to the overall performance of the retail industry. According to Mirza et al. [1], AI-based business intelligence software has been found to increase the efficiency of inventory management by 20-30% and customer retention by 15-25%. Machine learning algorithms like regression, decision tree, and neural network help in making accurate forecasts regarding customer behavior, sales trends, etc. [2].

In addition, AI has been found to greatly contribute to the overall performance of the retail industry in terms of customer behavior analysis. For instance, AI-based recommendation systems help in providing personalized shopping experiences to customers, resulting in an increase in overall sales. For example, the overall revenue generated by Amazon is greatly attributed to AI-based recommendation systems.Studies confirm that personalized marketing using AI has a positive impact on customer purchase intention.

Inventory management and demand forecasting techniques are considered to be two of the most effective applications of AI in retail businesses. The use of predictive analytics techniques in AI reduces the probability of inventory shortages, saving businesses extra expenditure on inventory management. AI-based applications in demand forecasting have already been implemented by retailers like Walmart, yielding better efficiency in supply chain management, saving businesses from operational losses [6].

However, there are certain challenges in implementing AI in retail businesses, including data privacy, implementation costs, and a lack of skilled resources. Ethical issues also pose a limitation in implementing AI in retail businesses, especially for small and medium-sized retailers.

From the literature, it is evident that AI is effective in analyzing a retail business; however, there is a lack of research in integrating different AI-based applications in a retail business into a unified decision-making model. The purpose of this study is to bridge this knowledge gap.

III. METHODOLOGY

A. Data Collection

The dataset used in this research was sourced from the Kaggle platform. It is an online platform that is used by many data scientists in carrying out data science and machine learning research. It is an open- access platform, meaning that it is freely accessible by everyone. This dataset is known as the Customer Purchase History dataset. It is a public dataset that is used by many data scientists in carrying out research. It contains 1,800 rows of data that were generated between 2023 and 2025. It is used in the simulation of real-world scenarios in the retail sector. It contains various details regarding the transactions that were carried out by customers. It contains details regarding customer identification, product details, dates of purchase, quantities, prices, modes of payment, and ratings. This dataset is used in the simulation of real scenarios in various stores that deal in various products such as Electronics and Office Supplies. No personally identifiable information is contained in the dataset. It is downloaded from the Kaggle platform in carrying out machine learning-based customer segmentation

Table I: Customer Purchase History Dataset Summary

Parameter	**Description**
Dataset Name	Customer-Purchase-History
Total Transactions	1,800
Unique Customers	1,642
Time Period	2023–2025
Product Categories	Electronics, Office Supplies
Payment Methods	Cash, Card, Gift Card
Data Type	Real transactional retail data

Table II: Description of Dataset Attributes

Column Name	**Description**
Customer ID	Unique identifier for each customer
Transaction ID	Unique identifier for each purchase
Product ID	Unique identifier for each product
Product Category	Category to which the product belongs

Transaction Date	Date of purchase
Quantity	Number of items purchased
Unit Price	Price per unit of product
Total Sales	Total transaction value
Payment Method	Mode of payment
Store/Channel	Purchase channel(online/offline)

B. Feature Extraction for K-Means Clustering

While the data set possesses a number of attributes, only a limited number of numerical and behavior- oriented attributes are necessary for the clustering operation using the K-Means clustering algorithm. In this regard, feature extraction was carried out to extract only the necessary attributes from the overall data set that could represent the customer's purchasing behavior.

From the entire data set, customer-level numerical attributes such as the frequency of purchases, total amount spent, average amount spent in a transaction,and recency were extracted. Categorical attributes were not considered for the clustering operation due to the fact that the K-Means clustering algorithm is a distance computation-based clustering algorithm.

Table III: Selected Features Used for K-Means Clustering

Feature Name	**Derived from column(s)**	**Reason for selection**
Purchase Frequency	Transaction ID	Measures customer activity level
Total Purchase value	Total sales	Represents customer monetary value
Average Transaction value	Total sales	Differentiate high and low spenders
Recency	Transaction Date	Indicates recent customer engagement
Total Quantity Purchased	Quantity	Captures buying volume behaviour

C. K-Means Clustering

K-Means is a widely used unsupervised learning technique that groups data into K clusters based on similarity. It works by assigning data points to the closest cluster centroid and iteratively updating the centroids to minimize intra-cluster distances. In the retail business analysis domain, K-Means is commonly used for customer

segmentation, which allows retailers to segment customers based on purchase behavior, spend, and frequency. However, K-Means is sensitive to the choice of K and the initialization of the centroids.

The methodologies involve data preprocessing, feature extraction, machine learning modeling, and business insight generation. Data preprocessing involves cleaning, normalizing, and aggregating data at the customer level. Feature selection is achieved using a well-known RFM-based model [1], [11].

K-Means clustering is used as an unsupervised learning algorithm to segment customers into three distinct behavioral segments. The algorithm is commonly used in customer segmentation in retail analytics [2], [8], [13].

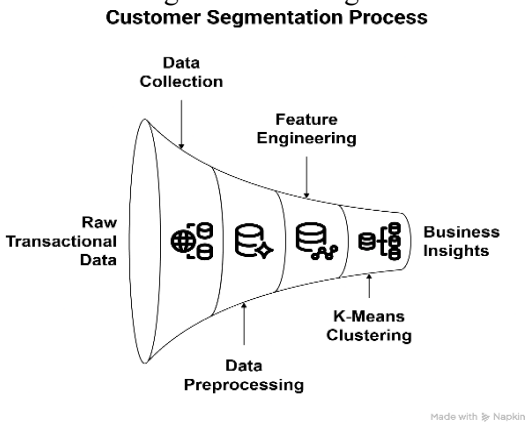


Fig. 1. AI-Based Retail Analytics Framework

IV. RESULTS AND DISCUSSION

Table IV: Customer Segmentation Results Using K-Means Clustering

Cluster	No. of Customers	Interpretation
Cluster 0	690	High-Value Customers
Cluster 1	805	Medium-Value Customers
Cluster 2	147	Low-Value Customers

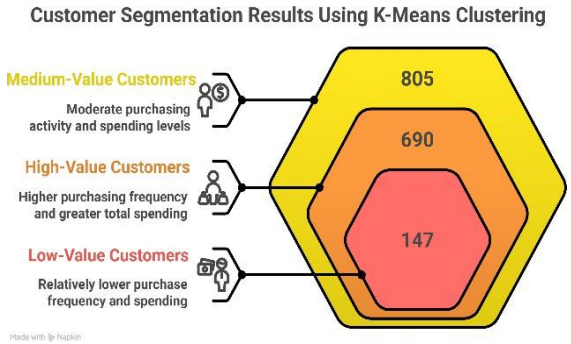
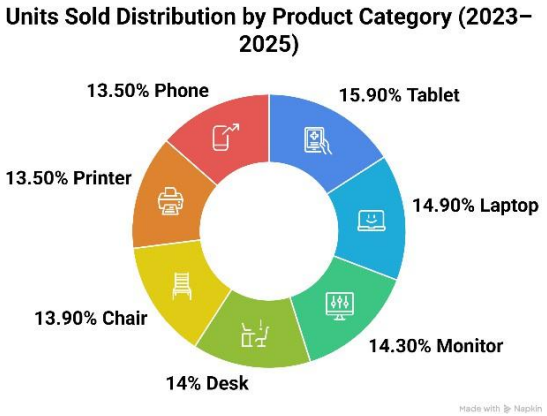


Fig. 2. Customer Segmentation Based on Purchasing Behavior

Fig. 3. Revenue Contribution by Product



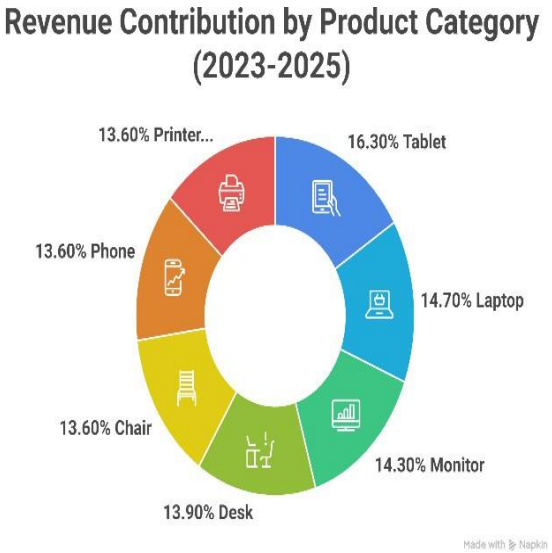


Fig. 4. Units Sold Distribution by Product

A. Product-Level Sales Performance Analysis (2023–2025)

The data collected from the sales between 2023 and 2025 has been analyzed using the K-Means clustering algorithm based on the aggregated features of the products, such as the quantity sold, revenue, and purchase frequency. From the results obtained from the clustering algorithm, the differences in the sales performance of the products have been identified.

The results obtained from the analysis show that Tablet has the highest sales performance compared to the other products in the data. Tablet has the largest share of the total quantity sold and the revenue earned. It has the highest frequency of purchases and the largest share of the revenue earned from the sales of the products. It has the highest performance in the high-performance cluster and is the most strategic product in the data, which has the largest share of the revenue earned in the retail industry.

On the other hand, the results obtained from the analysis show that the product Printer has the lowest sales performance in the data. It has the smallest share of the total quantity sold and the revenue earned. It has the smallest frequency of purchases compared to the other products. It has the lowest performance in the high-performance cluster and has the smallest share of the revenue earned in the retail industry.

The results obtained from the analysis of the data using the K-Means clustering algorithm have been compared using the unit sales and the revenue earned. It has been observed that the results obtained from the analysis show that Tablet has the largest share of the unit sales and the revenue earned. On the other hand, the results obtained from the analysis show that the product Printer has the smallest share of the unit sales and the revenue earned.

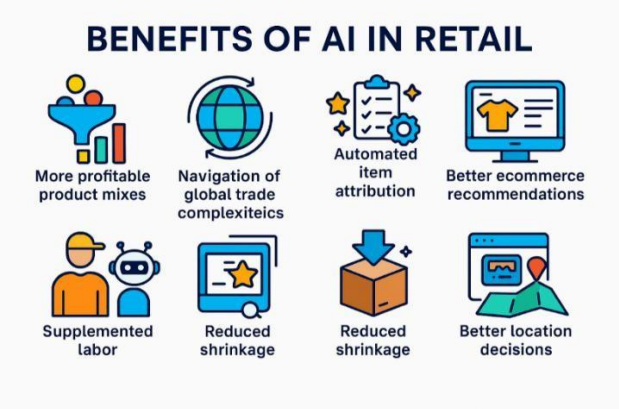


Fig. 4. Benefits of AI in Retail Analytics

AI-based analytics enables retailers to optimize operations, enhance customer satisfaction, and improve profitability, as supported by recent empirical studies.

V. CONCLUSION

This study illustrates the role of Artificial Intelligence in enhancing the analysis of retail businesses using real customer purchase information. The study was conducted using the K-Means algorithm on 1,800 transactions in the retail industry, resulting in the successful segmentation of customers into high-value customers, medium-value customers, and low-value customers.The study proves that the application of Artificial Intelligence in the retail industry for customer segmentation can help businesses understand the spending patterns, purchase frequency, and engagement of customers. The study also proves that the application of Artificial Intelligence in the retail industry results in the following benefits:

- Improved customer segmentation
- Improved decision-making
- Improved marketing strategies
- Improved inventory management
- Improved operational efficiency

Therefore, the role of Artificial Intelligence in the retail industry for the optimization of retail businesses is significant.

VI. FUTURE WORK

Though this study was focused on customer segmentation using K-Means clustering, there are many ways in which future research can be extended:

Sales Forecasting : Using machine learning algorithms to predict sales trends in the future.

Market Basket Analysis : Using association rule mining to find products that customers buy together frequently.

Customer Churn Prediction : Developing models to predict customers who might stop purchasing products in the future.
Sentiment Analysis : Using Natural Language Processing (NLP) techniques to analyze customer feedback and understand customer satisfaction levels.
Real-Time Analytics : Using AI algorithms in real- time retail applications to generate instant analytics.
Larger Data : Using larger datasets, as well as multi-store data, to improve model accuracy.

Further research can be done in using other advanced AI techniques, like deep learning, in customer segmentation.

VII. REFERENCES

[1] D. Grewal, A. L. Roggeveen, and J. Nordfält, "The future of retailing," *Journal of Retailing*, vol. 93, no. 1, pp. 1–6, 2017.
[2] M.-H. Huang and R. T. Rust, "Artificial intelligence in service," *Service Science*, vol. 10, no. 2, pp. 155–172, 2018.
[3] D. Bitra, "The revolutionary impact of AI in modern retail: A technical analysis," *Int. J. Res. Comput. Appl. Inf. Technol.*, vol. 7, no. 2, pp. 1981– 1992, 2024.
[4] S. Swaraj, "The role of artificial intelligence in transforming retail and supply chain management," *IJCRT*, vol. 13, no. 6, pp. 175–182, 2025.
[5] R. Gahlawat, "Role of artificial intelligence in modern retail management," *IJGRIT*, vol. 3, no. 2, pp. 179–190, 2025.
[6] M. Choudhary and R. Milan, "AI-powered innovations in retail marketing," *Int. J. Commerce Manag. Res.*, vol. 11, no. 3, pp. 49–57, 2025.
[7] J. B. Mirza *et al.*, "AI-driven business intelligence in retail," *AIJMR*, vol. 3, no. 1, pp. 1–22, 2025.
[8] V. N. Gaikwad, "AI adoption in retail: Challenges, impact, and strategic responses," *The Chitransh Academic & Research Journal*, vol. 1, no. 5, pp. 42–55, 2025.
[9] M. A. Raji *et al.*, "Real-time data analytics in retail," *GSC Advanced Research and Reviews*, vol. 18, no. 3, pp. 59–65, 2024.
[10] A. Alfian *et al.*, "Customer behavior analysis using real-time analytics in retail," *Computers & Industrial Engineering*, vol. 134, pp. 593–606, 2019.
[11] K. Kumar *et al.*, "Machine learning applications in retail analytics," *Procedia Computer Science*, vol. 167, pp. 1822–1831, 2020.
[12] P. Chatterjee *et al.*, "AI-enabled customer analytics in retail," *Journal of Retail Analytics*, vol. 15, no. 2, pp. 45–58, 2021.
[13] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2019.
[14] V. Pareto, *Manual of Political Economy*, Macmillan, 1971.
[15] H. A. D. M. Arachchi and G. D. Samarasinghe, "AI attributes and consumer purchase intention," *European Journal of Management Studies*, vol. 30, no. 3/4, pp. 243–267, 2025