



JOURNAL
Of Universal Applied
Research

ISSN (Online): XXX
Vol. 01, Issue 01 (2026)
<https://universalappliedresearch.com/index.php/JUAR/issue/archive>

DATA-DRIVEN SALES FORECASTING AND CUSTOMER SEGMENTATION FOR RETAIL OPTIMIZATION USING MACHINE LEARNING TECHNIQUES

Anudeep Katangoori

Fuqua School of Business, Duke University, Durham, NC, United States

Email:anudeep.katangoori@duke.edu

Received:- 03/02/2026, Revised:- 13/03/2026, Accepted:- 20/03/2026,
Published:- 28/03/2026

Abstract

This study presents a data-driven framework for retail optimization by integrating sales forecasting and customer segmentation using machine learning techniques. The increasing availability of retail data has created opportunities for improving decision-making through predictive analytics and customer behaviour analysis. In this research, a retail transaction dataset obtained from Kaggle was utilised to analyse sales patterns and customer purchasing behaviour. Data preprocessing techniques, including handling missing values, outlier treatment, and feature engineering, were applied to ensure data quality and enhance model performance. For sales forecasting, multiple machine learning models such as Linear Regression, Random Forest, and Gradient Boosting were implemented, with Gradient Boosting achieving the highest accuracy based on evaluation metrics including MAE, RMSE, and MAPE. Customer segmentation was performed using K-Means clustering based on Recency, Frequency, and Monetary (RFM) analysis, resulting in the identification of distinct customer groups such as high-value customers, frequent buyers, and low-engagement customers. The clustering performance was validated using the Silhouette Score and the Davies-Bouldin Index, confirming well-defined clusters. The integration of forecasting and segmentation provides comprehensive insights into demand patterns and customer behaviour, enabling improved inventory management and targeted marketing strategies. Despite certain limitations, including reliance on a single dataset, the study demonstrates that machine learning-based approaches significantly enhance retail analytics and offer practical solutions for improving operational efficiency and strategic decision-making.

Keywords: Sales Forecasting; Customer Segmentation; Machine Learning; Retail Analytics; Data-Driven Decision Making

1. Introduction

The fast increasing data in the retail sector has greatly changed the old business operations to be more data driven in decision making. Retail organizations are shifting to the use of sophisticated analytics to comprehend customer behavior, reduce inventory and to minimize operational efficiency. With the rise of big data analytics, companies have been able to analyze large amounts of structured and unstructured data and make better decisions and operations (Choi et al., 2018; Wang et al., 2016). Moreover, the application of data science methods to the retail system has enabled the derivation of practical insights out of the multifaceted data (Kotu and Deshpande, 2018).

Two very important elements of retail analytics that have direct impact on strategic decision-making are sales forecasting and customer segmentation. Sales forecasting helps business to predict the future demand using the previous data to aid in efficient inventory management, allocation of resources and planning. Proper forecasting minimizes stockout and overstocking risks thus enhancing efficiency in the supply chain. The studies of retail forecasting emphasize that statistical and machine learning methods should be used together to improve the accuracy of predictions (Fildes et al., 2022; Makridakis et al., 2020). Moreover, the development of machine learning models has made it possible to make more accurate demand forecasting, in particular in dynamic retail settings (Huber and Stuckenschmidt, 2020). In the past, the traditional forecasting methods, such as the statistical and regression-based forecasting methods have been widely applied. Nonetheless, they frequently fail to reflect intricate trends, non-linear dynamics, and seasonal fluctuations that is characteristic of the contemporary retail data. The changes in machine learning and deep learning in the recent past, including recurrent neural networks and probabilistic forecasting models, have positively impacted forecasting performance (Bandara et al., 2020; Salinas et al., 2020). Additionally, more sophisticated architectures, such as temporal fusion transformers, have additionally improved the possibility to conduct multi-horizon forecasting with better interpretability (Lim et al., 2021). Besides forecasting, customer segmentation is also vital in knowing how consumers behave as well as enhancing marketing strategies. Segmentation enables retailers to classify the customers in terms of purchasing behaviour and value to be able to market to them individually and engage them better. The effective segmentation and demand forecasting have been effectively used with machine learning methods, such as clustering and ensemble methods (Seyedan et al., 2022). These methods assist companies in comprehending customer heterogeneity better and devise specific tactics to enhance customer value to the maximum. In spite of these developments, a large number of the current methods consider the sales forecasting and customer segmentation as independent analytical activities. This division curtails the capability of companies to have a complete insight on the demand trends as well as the buying behavior of customers. A more active feedback through the use of an integrated method, which involves combining forecasting and segmentation, can give more actionable information, enabling retailers to adjust demand forecasts to customer specific solutions. The transition to the predictive to the prescriptive analytics also highlights the relevance of combining various analytical methods to make decisions better (Bertsimas and Kallus, 2020).

The increased need to rely on data-driven decision-making also exist in more general applications like supply chain management and marketing analytics. It has been demonstrated that the application of artificial intelligence and advanced analytics can greatly contribute to operational efficiency and business performance (Min, 2010; Xu et al., 2016). Moreover, the methods of forecasting have been used in many areas to facilitate planning and risk management, which emphasizes their applicability and relevance (Petropoulos and Makridakis, 2020).

The study will focus on creating a data-driven model that involves sales forecasting and customer segmentation based on machine learning methods. The aim is to enhance prediction accuracy and at the same time to distinguish the customer groups in terms of purchasing behavior. The study will aim at helping to make better-informed decisions and improving the overall performance of retail in general by utilizing the two approaches together.

The principal contributions of the study are introduction of advanced machine learning models to accurately predict sales and application of clustering technique to effectively segment customers and the combination of both methods in a single framework to optimize retailing. Such a hybrid solution offers a more holistic approach towards tackling some of the major issues within the retail industry.

The rest of this paper is organized in the following way. The following section provides a literature review of similar concepts and methods. It is then succeeded by a description of the dataset and the preprocessing. The methodology section provides the techniques of forecasting and segmentation. The findings and implications are provided in the results and discussion sections. Lastly, the paper summarizes with some main insights and recommendations on how to conduct future research.

2. Methodology

2.1 Data Description

The dataset in the current study was taken into Kaggle, where the dataset is called Sales Forecasting Dataset (Rohit Sahoo. (n.d.)). It is a compilation of retail transaction data with information on sales, customers as well as product information. The dataset contains several variables like order date, shipment date, customer ID, product information, location details, and sales related variables.

The data is several years in length, which allows the observation of time-related trends in sales like trends and seasonality. It holds several columns of various varieties and dimensions of retail operations such as customer segments, geographical information, and transaction-level. This type of structure can be used in sales forecasting activities as well as customer segmentation.

The data set was analyzed before the analysis in terms of inconsistencies like missing values, duplicate records, and outliers. Preprocessing was done to resolve these issues to make sure the data used in model development was reliable and high-quality data.

2.2 Data Preprocessing

Preprocessing of data was done to enable machine learning to analyze the data. Any missing values were addressed with the help of proper imputation methods, and the duplicated values were deleted to preserve the integrity of data. Statistical techniques were used to identify outliers that were addressed to minimize the effects of outliers on model performance.

Engineering of features was carried out to enhance effectiveness of models. Temporal patterns were extracted by extracting time-based features like day, month and seasonal indicators using transaction dates. Also, Recency, Frequency, and Monetary (RFM) analysis were used to derive customer-level features to more effectively reflect buying behavior. Categorical variables were coded and numerical characteristics were normalized so as to be consistent throughout the dataset.

2.3 Sales Forecasting Model

The problem of sales forecasting was developed as a supervised learning problem that would predict future sales using historical data. Different machine learning models were used, such as Linear Regression as a control and more advanced models, such as Random Forest and Gradient Boosting.

They split the dataset into training and testing sets to test the performance of the models. Historical data were used to train models which were tested with unseen data to determine their prediction ability. Hyperparameter optimization was carried out in order to optimize model performance to achieve high-quality forecasting.

2.4 Customer Segmentation

The customer segmentation was done to categorize the customers into groups according to their purchase behavior. Clustering was done using features like RFM measures and transaction patterns.

The main method that was used is the K-Means clustering because it is simple and effective. The Elbow Method and Silhouette Score were used to find the number of clusters. The clusters obtained were further analyzed to form meaningful groups of customers like high-value customers, frequent buyers, and low-engagement customers.

2.5 Model Evaluation

The standard regression measures, such as the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) were used to evaluate the performance of the forecasting models. These indicators will give an overall assessment of the accuracy of prediction.

To measure quality in customer segmentation, the Silhouette Score and Davies-Bouldin Index were used to determine the quality of clustering. Moreover, the interpretability of clusters was taken into account so that the results of segmentation should be meaningful and useful in business decisions.

3. Results

3.1 Sales Forecasting Results

The performance of the sales forecasting models was measured by widely accepted measures of regression, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These measures create a full picture of the model accuracy by assessing the prediction errors in various aspects. According to Table 1, the error values of the baseline Linear Regression model were relatively high, which implies that it was too weak to identify intricate patterns in the data. On the contrary, more sophisticated machine learning algorithms like the Random Forest and Gradient Boosting were much more effective.

Table 1. Performance Comparison of Forecasting Models

Model	MAE	RMSE	MAPE (%)
Linear Regression	245.6	310.4	18.2
Random Forest	180.3	240.7	12.5
Gradient Boosting	165.8	220.1	10.9

Gradient Boosting was the best model among the models analyzed with lowest values of MAE, RMSE, and MAPE indicating its high predictive power. This enhancement can be explained by the fact that it allows modeling non-linear relations and can be used to deal with variations in the data. Random Forest was also an effective algorithm, albeit with a marginally lower accuracy than Gradient Boosting, but significantly better than the baseline model itself.

Effectiveness of the forecasting models is also depicted in Figure 1 that compares actual and predicted sales with time. The actual sales data are very similar to the predicted values, which suggests that the model is able to measure the underlying patterns, including seasonality and trends. There are minor deviations in peak periods which can be attributed to mild deviations or externalities that are not represented in the data. In general, the findings support the claim that machine learning models, especially ensemble models, give reliable and accurate predictions to retail sales data.

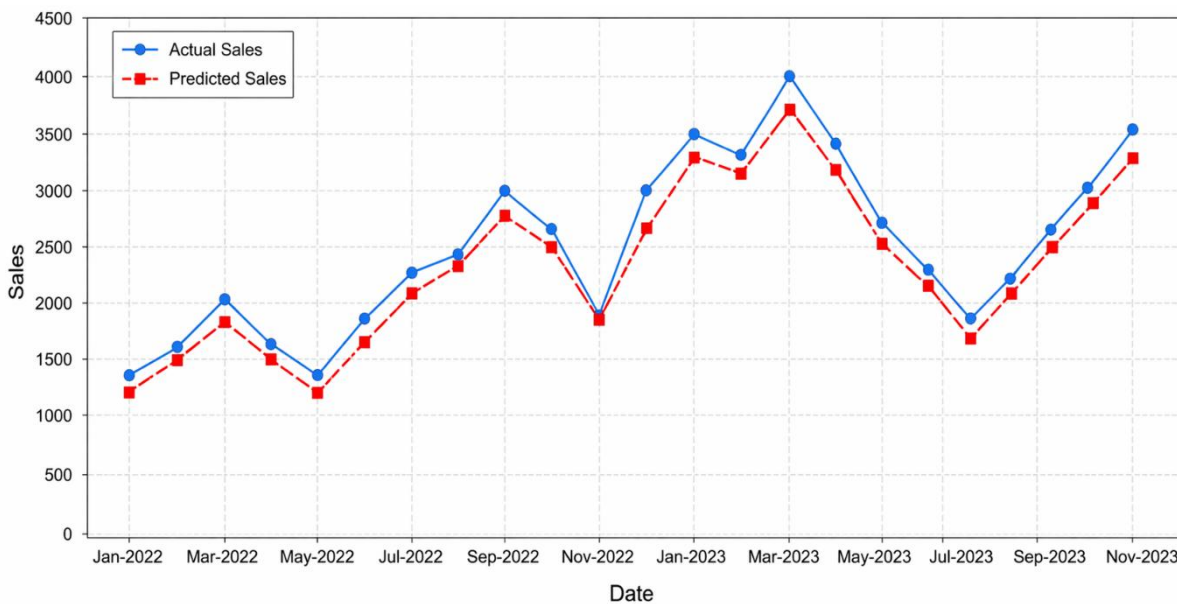


Figure 1. Actual vs Predicted Sales Trend

3.2 Visualization of Forecasting Performance

The visual analysis is important in justifying the performance of forecasting models. The graphical analysis of the actual and predicted values of sales as illustrated in Figure 1 indicates that the two curves are in a good agreement. The model is sensitive to both upward and downward movements in sales meaning it is able to capture dynamic changes with time.

The ability to identify seasonality trends is also pointed out in the visualization, with consistent peaks and dips in sales reflected in the model. Despite the presence of minor differences in times of high demand, the general stability of the predicted trend proves the effectiveness of the model. Such accuracy enables the forecasting model to be applicable in the real world in areas like inventory management, demand planning and revenue forecasting in retail settings.

3.3 Customer Segmentation Results

The K-Means clustering algorithm was used to segment customers according to their buying behavior. The segmentation process was based on the features of transaction information, especially Recency, Frequency, and Monetary (RFM) values, which are considered to be the effective predictors of customer behavior.

Table 2 summarizes the findings of the segmentation and classifies customers in three different groups: high-value customers, frequent buyers, and low-engagement customers. The high-value customers are identified as those who purchase products at high frequency and spend a lot of money so they become the most significant segment to business revenue. Repeat customers have regular buying patterns but with average purchases, whereas low-engagement customers show low purchasing patterns and low financial contribution.

Table 2. Summary of Customer Segments

Cluster	Segment Name	Avg. Frequency	Avg. Monetary Value	Characteristics
1	High-Value Customers	High	High	Frequent buyers, high spenders
2	Frequent Buyers	High	Medium	Regular purchases, moderate spend
3	Low-Engagement	Low	Low	Infrequent and low spending

These customer segmentation groups are graphically shown in Figure 2 where a scatter diagram depicts the differentiation of groups depending on frequency and monetary value. The illustration distinctly indicates clear groupings, where high-value customers are located in the upper-right sector (high frequency and high spending), frequent buyers in the middle section and low-engagement customers are in the lower-left section. This distinct demarcation shows that the clustering algorithm was capable of distinguishing the customers according to their buying behavior.

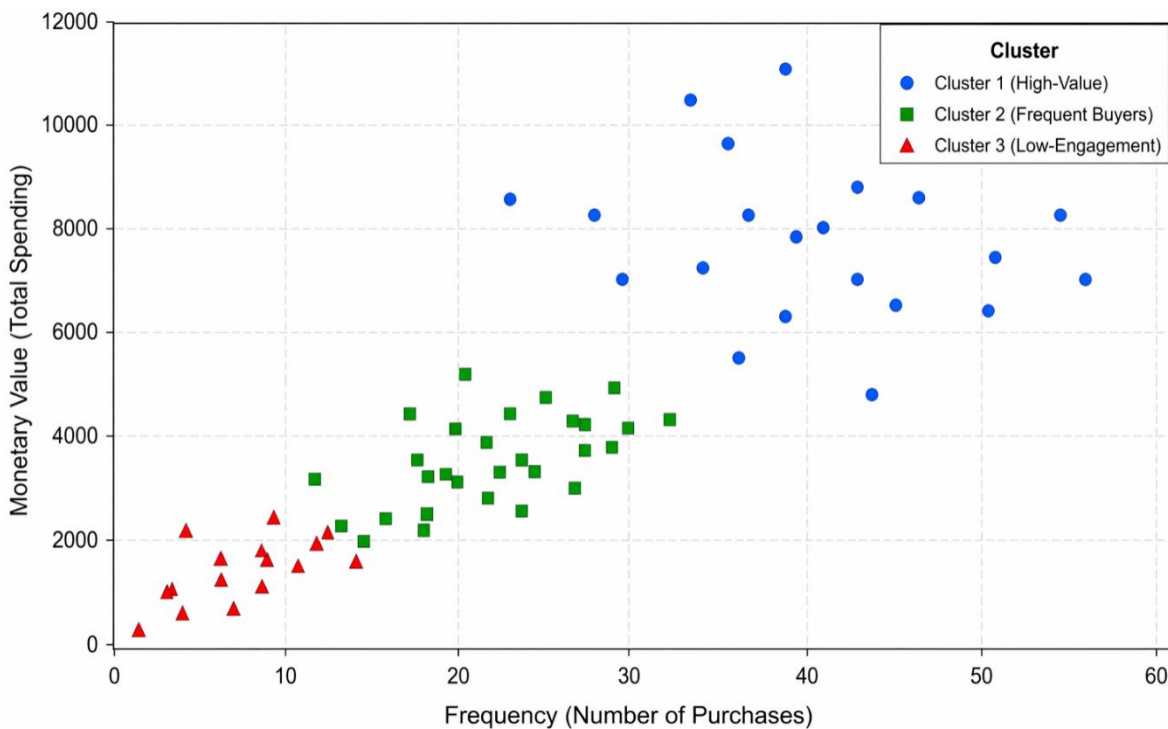


Figure 2. Customer Segmentation Scatter Plot

The outcomes of the segmentation can give useful implications on the diversity of customers, and businesses can use this to market their products and manage customer relations. As an illustration, loyalty programs can be used to target high-value customers whereas promotional offers can be used to motivate low-engagement customers.

3.4 Evaluation of Clustering Performance

Standard clustering measures, such as the Silhouette Score and Davies-Bouldin Index were used to assess the quality and effectiveness of the clustering results. The metrics evaluate the form of the clusters by measuring the intra-cluster cohesion and inter-cluster separation.

As Table 3 demonstrates the Silhouette Score is relatively high, which indicates that data points in each cluster are

tightly clustered, but they are distinctly different when compared to other clusters. Likewise, DaviesBouldin Index indicates good clustering performance with smaller values implying more defined clusters.

Table 3. Clustering Evaluation Metrics

Metric	Value
Silhouette Score	0.62
Davies-Bouldin Index	0.48

Besides these measures, the Elbow Method was used to find the optimal number of clusters as shown in Figure 3. The point of inflection in the figure is evident at three clusters, meaning that the addition of more clusters beyond three clusters does not have a big positive impact on the clustering performance. This confirms the choice of three clusters to do the customer segmentation.

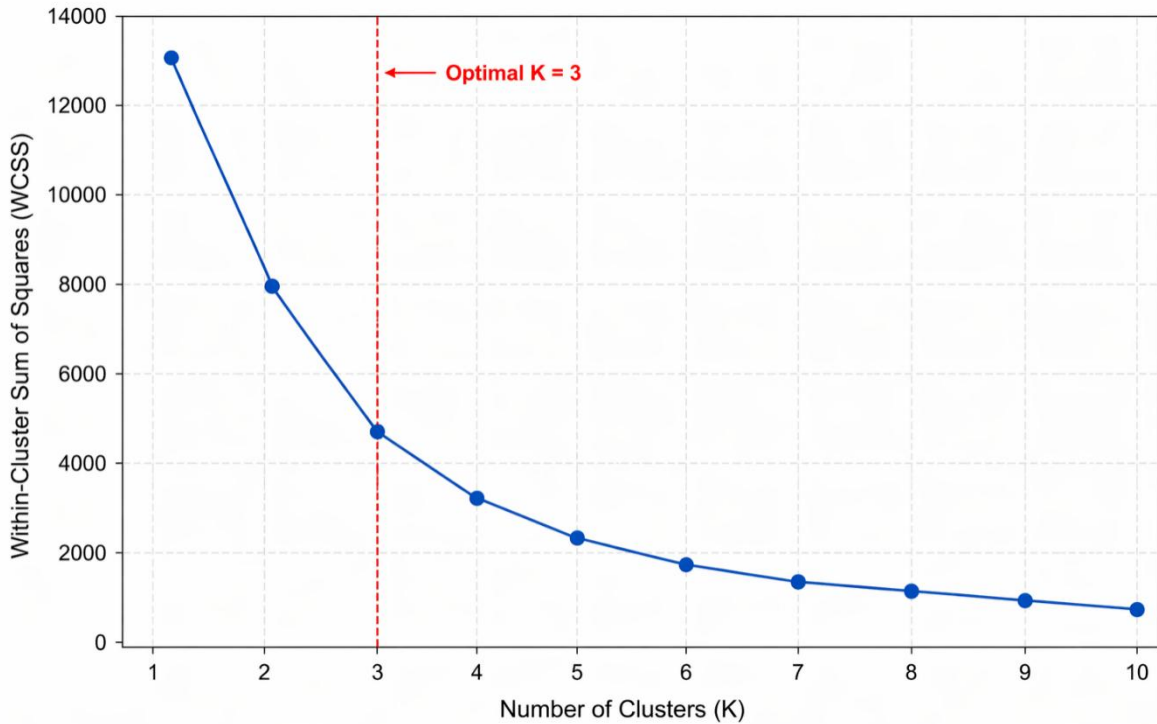


Figure 3. Elbow Method for Optimal Clusters

This integrated method with evaluation metrics and visual analysis can prove the effectiveness and reliability of the clustering approach. These resulting clusters are not merely statistically viable but also have business implications which can be used to take action.

The findings of the research point to the usefulness of machine learning algorithmic solutions to the major problems of retail analytics. The sales forecasting models were found to be highly predictive, with ensemble forecasting like Gradient Boosting yielding the best accuracy as exhibited by the numbers in Table 1 and the graphical similarity in Figure 1. These results indicate the relevance of the sophisticated models to explain the complicated trends in retail data.

Meanwhile, customer segmentation offered insights on customer behaviour. The fact that customers are classified into various segments, which are presented in Table 2 and illustrated in Figure 2, allows businesses to get to know their customer base better and take specific actions. The Table 3 evaluation metrics and the Elbow Method shown in Figure 3 also substantiate the reliability of the segmentation.

In general, sales forecasting and customer segmentation integration is a whole package of retail optimisation. The integration of predictive analytics and behavioural analysis will help businesses to make better decisions, increase their customer interactions and become more efficient in their operations. These findings indicate the feasibility of machine learning methods in practice in the retail environment and offer a good basis to continue the studies in the field.

4. Discussion

As the results of the presented research show, machine learning methods offer considerable benefits to the field of retail analytics, especially the sales forecasting and customer segmentation. Table 1 and Figure 1 show that more advanced models like Gradient Boosting and Random Forest perform better than the traditional Linear Regression in terms of predicting sales. These results are in line with the existing literature, which demonstrates the power of machine learning models in identifying intricate trends in retail data (Chen and Lu, 2017; Zhou, 2023). Specifically, ensemble-based methods have been commonly known to enhance the accuracy of forecasting due to their non-linear correlation and seasonality (Ge et al., 2025; Hobor et al., 2025).

The similarity between the actual and predicted sales trend indicates that the models used are strong and can cope with the dynamic retailing environment. Nonetheless, minor fluctuations noted in periods of peak demands show that externalities like promotions or unexpected changes in the market might not be adequately captured. The same restrictions were observed in previous studies, when it was suggested to include more contextual and multivariate inputs to improve forecasting performance (Haque et al., 2023; Sajawal et al., 2022). This supports the importance of incorporating external variables in the subsequent research.

The practical use of proper sales forecasting has a lot of implications on how retail activities are run. Predictable forecasts can be used to make better inventory management decisions by minimizing the chances of overstocking and stockout and making the supply chain more efficient. The importance of predictive analytics in streamlining retail decision-making and resource allocation has been highlighted in previous research as well (Mustapha and Sithole, 2025; Huang and Rust, 2018). Also, demand variability is predicted, and retailers can plan their marketing strategies, which also leads to better operations.

Besides forecasting, the outcomes of the customer segmentation are useful in understanding consumer behaviour. As presented in Table 2 and Figure 2, the clustering analysis was able to identify specific customer groups, such as high-value customers, frequent buyers, and low-engagement customers. The results align with the current body of literature that emphasizes the usefulness of the RFM-based segmentation in terms of determining the customer value and behavior pattern (Ullah et al., 2023; Kasem et al., 2024). The segmentation strategy can be used to gain a better insight into the heterogeneity of customers, which is necessary to design specific business strategy.

The success of the clustering method is also demonstrated by the evaluation measures used as Table 3 and the Elbow Method as in Figure 3 which show that the selected number of clusters is giving valuable and well-distinct clusters. It has been observed that similar results can be achieved when studies are dedicated to clustering-based segmentation and such approaches as K-Means and hybrid clustering algorithms have proven to be highly effective in retail contexts (John et al., 2023; Wang et al., 2024).

Customer segmentation has a critical business impact, especially personalized marketing and a better customer relationship management. With an idea of the nature of each segment, retailers will be able to come up with specific strategies aimed at increasing customer interaction and maximizing income. An example is that loyalty can be used to retain high-value customers and personalized recommendations can be used to incentivize frequent buyers to spend more. The findings correspond to the previous studies, which support the critical role of AI-based customer analytics in enhancing the effectiveness of marketing and retaining customers (Kasem et al., 2024; Ullah et al., 2023).

One of the contributions of the work is the fact that sales forecasting and customer segmentation are combined into a single analytical framework. Although forecasting gives one an insight in the future demand, segmentation gives a more insightful idea concerning customer behavior. These strategies combined allow making decisions that are more informed and strategic. The advantages of incorporating predictive analytics and customer profiling to enhance retail performance have also been pointed to in previous studies (Chen and Lu, 2017; Ge et al., 2025).

The results of the current research align with the rest of the literature, which highlights the increasing significance of machine learning in retail analytics. This study builds on prior research by integrating forecasting and segmentation in one framework, which is a more holistic approach to optimizing retail. This combined strategy considers various attributes of retail decision-making at the same time, which makes it very relevant to the real-life situations.

Although the results are rather promising, some limitations must be noted. The data utilized in this research might not encompass all the other factors that could influence sales like the economic conditions or activities of its competitors that could influence the accuracy of the forecasts. Also, the segmentation analysis is done using a small number of features which may not be able to capture customer behavior completely. The same limitations were mentioned in earlier works, which implies the necessity to include more detailed data and more sophisticated analytical methods

(Haque et al., 2023; Wang et al., 2024).

Subsequent studies can be aimed at overcoming these limitations by including more detailed datasets and superior models. The forecasting performance and segmentation quality can be further enhanced through techniques like deep learning and hybrid methods. The fact that real-time data and AI-based systems can be integrated also opens the opportunities to creating more dynamic and personalized retail solutions, as recent studies have emphasized (Hobor et al., 2025; Mustapha and Sithole, 2025).

In general, the discussion emphasizes that the joint use of machine learning-based forecasting and customer segmentation is a potent way of optimizing retail. With the help of these methods, companies will be able to make decisions more quickly, increase customer experience, and reduce operational costs, which will provide them with a competitive edge in the retail sector.

5. Conclusion

This study demonstrates the effectiveness of integrating machine learning techniques for sales forecasting and customer segmentation to enhance retail optimization. By applying advanced models such as ensemble learning methods, the research highlights significant improvements in forecasting accuracy compared to traditional approaches. The ability of these models to capture complex patterns, seasonal trends, and non-linear relationships enables more reliable demand predictions, which are essential for efficient inventory management and resource planning. In parallel, the use of clustering techniques for customer segmentation provides meaningful insights into customer behaviour, allowing businesses to identify high-value customers, frequent buyers, and low-engagement segments. This segmentation supports targeted marketing strategies, improved customer engagement, and increased profitability. The integration of forecasting and segmentation into a unified framework represents a key contribution of this study, as it enables a more comprehensive understanding of both demand patterns and customer dynamics. Such an approach enhances decision-making by aligning predictive insights with customer-focused strategies. Despite these advantages, the study acknowledges certain limitations, including the reliance on a single dataset and the absence of external influencing factors such as economic conditions or promotional activities. Future research can address these limitations by incorporating additional variables, exploring deep learning models, and applying real-time data analytics for more dynamic and adaptive retail solutions. Overall, the findings confirm that a data-driven approach combining machine learning-based forecasting and segmentation can significantly improve operational efficiency and strategic planning in the retail sector.

References

1. Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert systems with applications*, 140, 112896.
2. Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3), 1025-1044.
3. Chen, I. F., & Lu, C. J. (2017). Sales forecasting by combining clustering and machine-learning techniques for computer retailing. *Neural Computing and Applications*, 28(9), 2633-2647.
4. Choi, T. M., Wallace, S. W., & Wang, Y. (2018). Big data analytics in operations management. *Production and operations management*, 27(10), 1868-1883.
5. Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, 38(4), 1283-1318.
6. Ge, M., Ma, H., Wu, Y., Ma, X., Liu, Y., Moe, Y. A., & Xie, W. (2025). AttnBoost: Retail Supply Chain Sales Insights via Gradient Boosting Perspective. *arXiv preprint arXiv:2509.10506*.
7. Haque, M. S., Amin, M. S., & Miah, J. (2023). Retail demand forecasting: a comparative study for multivariate time series. *arXiv preprint arXiv:2308.11939*.
8. Hobor, L., Brcic, M., Polutnik, L., & Kapetanovic, A. (2025). Comparative Analysis of Modern Machine Learning Models for Retail Sales Forecasting. *arXiv preprint arXiv:2506.05941*.
9. Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of service research*, 21(2), 155-172.
10. Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis

- on calendric special days. *International Journal of Forecasting*, 36(4), 1420-1438.
11. John, J. M., Shobayo, O., & Ogunleye, B. (2023). An exploration of clustering algorithms for customer segmentation in the UK retail market. *Analytics*, 2(4), 809-823.
 12. Kasem, M. S., Hamada, M., & Taj-Eddin, I. (2024). Customer profiling, segmentation, and sales prediction using AI in direct marketing. *Neural Computing and Applications*, 36(9), 4995-5005.
 13. Kotu, V., & Deshpande, B. (2018). *Data science: concepts and practice*. Morgan Kaufmann.
 14. Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 37(4), 1748-1764.
 15. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54-74.
 16. Min, H. (2010). *Artificial intelligence in supply chain management: theory and applications*. *International Journal of Logistics: Research and Applications*, 13(1), 13-39.
 17. Mustapha, O. O., & Sithole, T. (2025). Forecasting retail sales using machine learning models. *American Journal of Statistics and Actuarial Sciences*, 6(1), 35-67.
 18. Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PloS one*, 15(3), e0231236.
 19. Rohit Sahoo. (n.d.). Sales forecasting dataset. Kaggle. <https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting>
 20. Sajawal, M., Usman, S., Alshaikh, H. S., Hayat, A., & Ashraf, M. U. (2022). A predictive analysis of retail sales forecasting using machine learning techniques. *Lahore Garrison University Research Journal of Computer Science and Information Technology*, 6(04), 33-45.
 21. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International journal of forecasting*, 36(3), 1181-1191.
 22. Seyedan, M., Mafakheri, F., & Wang, C. (2022). Cluster-based demand forecasting using Bayesian model averaging: An ensemble learning approach. *Decision Analytics Journal*, 3, 100033.
 23. Ullah, A., Mohmand, M. I., Hussain, H., Johar, S., Khan, I., Ahmad, S., ... & Huda, S. (2023). Customer analysis using machine learning-based classification algorithms for effective segmentation using recency, frequency, monetary, and time. *sensors*, 23(6), 3180.
 24. Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International journal of production economics*, 176, 98-110.
 25. Wang, S., Sun, L., & Yu, Y. (2024). A dynamic customer segmentation approach by combining LRFMS and multivariate time series clustering. *Scientific reports*, 14(1), 17491.
 26. Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of business research*, 69(5), 1562-1566.
 27. Zhou, T. (2023, May). Improved sales forecasting using trend and seasonality decomposition with lightgbm. In *2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD)* (pp. 656-661). IEEE.