forecasting

MDPI

Article

Hybrid Clustering for Retail Demand Forecasting: Combining Rule-Based and Machine Learning Methods

Jung-Hyuk Kim [1] and Nam-Wook Cho [2,*]

[1] Graduate School of Public Policy and IT, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea; junghyukkim@seoultech.ac.kr
[2] Department of Industrial and Information Systems Engineering, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea
* Correspondence: nwcho@seoultech.ac.kr; Tel.: +82-2-970-6448

Highlights

What are the main findings?

- This study proposes an adaptive hybrid clustering framework that integrates rule-based and machine learning approaches to address the intermittent and heterogeneous demand patterns characteristic of FMCG retail environments.
- The results demonstrate that hybrid forecasting models incorporating demand-pattern embeddings consistently achieve superior accuracy compared with single-algorithm approaches across all identified demand segments.

What are the implications of the main findings?

- Since no single clustering method demonstrates universal superiority, practitioners are advised to adopt a context-sensitive strategy, selecting rule-based or machine learning approaches based on the characteristics of demand patterns.
- A diagnostic heuristic derived from preliminary clustering statistics can reduce experimental overhead by up to 50%, facilitating more resource-efficient model selection in large-scale retail settings.

Abstract

Retail demand forecasting for fast-moving consumer goods (FMCGs) presents significant challenges due to high product variety, demand intermittency, and uncertainty, which prevent any single model from capturing the diverse demand patterns. To address these challenges, this study proposes a hybrid clustering framework that integrates rule-based (Syntetos–Boylan Classification) and machine learning (ML) approaches, combining time-series embeddings with unsupervised learning to segment products by demand structure. Building on this framework, forecasting is conducted through a two-phase methodology: selecting optimal baseline algorithms per cluster (Phase 1), then enhancing them with embedding-based hybrid models (Phase 2). The effectiveness of this approach is demonstrated using a large-scale real-world dataset comprising over 3.8 million weekly sales records from 12,661 products across 691 stores. Results show that the proposed method improves forecasting accuracy by approximately 5–15% compared to conventional models. Furthermore, model performance varies with demand volatility, as different model–embedding combinations perform best under different conditions. Finally, the proposed diagnostic heuristic reduces experimental effort by 25–50%. Comparative analysis reveals that ML-based clustering outperforms rule-based methods under stable demand, whereas rule-based clustering is superior under high demand uncertainty, confirming that no single

Check for updates

Received: 21 February 2026
Revised: 21 April 2026
Accepted: 23 April 2026
Published: 27 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

clustering paradigm is universally optimal. These findings demonstrate the practical value of adaptive hybrid frameworks for FMCGs demand forecasting.

Keywords: hybrid clustering for retail demand forecasting; time-series embedding; clustering; FMCG retail; Random Forest; XGBoost; PatchTST; GAF-CNN; Syntetos–Boylan Classification (SBC); machine learning

1. Introduction

In the Fast-Moving Consumer Goods (FMCGs) retail sector, where consumption is rapid and product lifecycles are short, accurate demand forecasting is crucial. Inaccurate demand forecasts lead to stockouts and excess inventory, resulting in cost losses and customer dissatisfaction for businesses [1]. FMCGs retail data comprises tens of thousands of items at the Stock Keeping Unit (SKU) level, and accurate forecasting is challenging due to various factors, including seasonality, demand uncertainty, promotional effects, and new product launches [2].

In the multi-product retail industry, sales patterns vary widely across products, making it difficult for a single forecasting model to capture the time series characteristics of all items [3]. To overcome this limitation, a hybrid model is used that clusters items based on their demand characteristics and then applies optimal forecasting models to each cluster. When clustering makes patterns within clusters more homogeneous, it enables model training tailored to those characteristics, thereby improving forecasting performance for individual products [4].

Recent research on retail demand forecasting has employed clustering-based hybrid techniques to address the limitations of single forecasting methods. Among these approaches, two primary clustering strategies have emerged in the literature.

The first approach, rule-based clustering, utilizes the Syntetos–Boylan Classification (SBC), which categorizes demand into four types based on two statistical indicators: Average Demand Interval (ADI) and Squared Coefficient of Variation (CV^2). Appropriate forecasting algorithms are then applied to each category [5,6]. This approach demonstrates strong forecasting performance when demand intervals and uncertainty exhibit clear, distinguishable patterns.

The second approach employs machine learning (ML)-based clustering, which groups products using unsupervised learning algorithms such as K-means clustering, Hierarchical Agglomerative Clustering (HAC), and Gaussian Mixture Models (GMM), followed by applying tailored forecasting algorithms to each cluster [7,8]. Unlike rule-based methods, ML methods capture latent similarities in demand patterns directly from the data, enabling effective segmentation even when demand patterns are ambiguous or exhibit complex, multidimensional characteristics. While rule-based methods excel with well-defined demand structures, ML methods offer greater flexibility for datasets with intricate or overlapping demand patterns. However, the relative performance of these two approaches remains dataset-dependent, motivating the need for hybrid strategies that can adaptively leverage both methodologies.

While both methods have independently demonstrated excellent forecasting performance in previous studies [5,6,9,10], direct comparisons of these two clustering hybrid methods are lacking. The hybrid clustering for retail demand forecasting proposed in this study defines and enables the selection of these methods within a single framework. Through this, it aims to improve cluster-specific demand forecasting performance

by effectively clustering both data with categorized demand patterns and data with complex patterns.

This study addresses the research question: how can rule-based and machine-learning clustering strategies be selectively applied to improve retail demand forecasting under heterogeneous uncertainty conditions? We propose a hybrid framework integrating Syntetos–Boylan Classification with embedding-based unsupervised learning. The framework is evaluated using weekly point-of-sale (POS) data from prominent FMCGs retailers in South Korea, encompassing 691 stores, 12,661 products across two distribution centers, and more than 3.8 million transaction records. This dataset features diverse sales distribution characteristics, providing a robust testbed for examining clustering effectiveness under varying demand patterns. The results demonstrate that clustering performance depends critically on data characteristics rather than a single optimal approach, offering practical guidance for complex retail forecasting environments.

The paper's contributions can be summarized as follows.

- Practical Applicability: The hybrid clustering for retail demand forecasting proposed in this study overcomes the limitations of the existing clustering-based hybrid method. Comparing the unvalidated performance of rule-based and ML methods enables selecting effective methodologies based on data characteristics. Additionally, the practical approach, which utilizes actual sales data, enables the research results to be directly applicable to real-world practice.
- Utilization of Embedding-based Representation Learning: Embedding-based representation learning is a core strength of this study. Specifically, time series embeddings transform product-specific sales patterns into fixed-length vectors, thereby improving clustering when combined with unsupervised learning methods during the ML clustering phase. In both rule-based and ML methods, these embeddings are then used in the forecasting phase with baseline algorithms and exogenous variables to enhance product-specific forecasting.
- Explainable Model: This study used XGBoost-based Feature Importance analysis to identify key factors influencing prediction results. Using retail data without promotions, it clearly revealed variables affecting actual demand.
- Feature Diversity: Furthermore, the diversity of features is a key contributor to improved forecasting in this study. By integrating various data types including time series, sales (domestic/import classification, category, first shipment date, sales start date, price), economic (Consumer Price Index (CPI), Unemployment Rate, West Texas Intermediate (WTI), retail sales index), and weather data (average temperature, average relative humidity, average wind speed) the study enhances demand forecasting performance.

This study is structured as follows. Section 2 reviews the literature on FMCGs retail demand forecasting methodologies. Section 3 describes the entire process of the proposed research methodology. It proceeds in the order of dataset, preprocessing, clustering, forecasting, and evaluation. Section 4 evaluates the forecasting performance of the proposed framework and presents the optimal algorithm. Section 5 presents the paper's conclusions and provides insights for future research.

2. Related Work

In the FMCGs retail sector, accurately forecasting product demand is crucial. This prevents excess inventory and stockouts while enhancing customer satisfaction. The accuracy of demand forecasting enables supply chain management, inventory level maintenance, and production planning [11]. Traditional demand forecasting methods predict based on historical data but have limitations in capturing nonlinear relationships and external

factors [12]. Recently, machine learning and deep learning techniques have been applied to demand forecasting, improving forecasting performance [13]. Hybrid models have gained attention as methods that combine multiple forecasting techniques to leverage the strengths of each approach while compensating for their limitations [14].

2.1. Traditional Forecasting Methods

Demand forecasting methods are classified into three categories: time-series analysis, regression analysis, and machine learning [15]. Traditional methods include time-series and regression analyses. Time series analysis predicts future demand by modeling trends and seasonality using historical data. AutoRegressive Integrated Moving Average (ARIMA) and Holt–Winters Exponential Smoothing (HW) are established statistical methods well-suited for forecasting time series with trend or seasonal characteristics [16,17]. Regression analysis predicts by assuming relationships between sales demand and external variables. Linear regression assumes linear relationships between variables, while multiple regression considers multiple variables simultaneously [18]. However, these methods are limited by multicollinearity, sensitivity to outliers, and their inability to represent nonlinear relationships [19].

2.2. Machine Learning-Based Forecasting Methods

Machine learning-based methods have advanced demand forecasting by learning complex patterns and nonlinear relationships from data. Tree-based models derive predicted values by splitting input variables. Random Forest (RF) employs bagging techniques to generate multiple decision trees and then averages their results. Gradient Boosting Machine (GBM) sequentially corrects errors. Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost) demonstrate excellent performance in demand forecasting [20].

Deep learning methods enable data abstraction using multi-layer neural networks. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models are used for time-series forecasting, which is particularly suitable for data that exhibit temporal dependencies. LSTMs and Gated Recurrent Units (GRUs) learn long-term patterns through gating mechanisms [21]. Recently, transformer-based models have gained attention for their strong long-term forecasting performance, outperforming RNNs through self-attention mechanisms [22].

2.3. Hybrid Models

Hybrid models achieve superior performance and robustness compared to single models by combining multiple forecasting techniques [23]. These models can be classified into seven types: (1) Statistical-machine learning hybrids generate linear predictions using statistical models, then capture nonlinear residuals with machine learning or deep learning models [24,25]. (2) Clustering-based approaches group similar demand patterns and apply optimal models to each cluster [26,27]. (3) Ensemble methods aggregate predictions from multiple models to generate final forecasts [28,29]. (4) Decomposition-based methods separate time series into components and apply appropriate models to each [30,31]. (5) Parallel architectures execute different models simultaneously and combine their outputs [32,33]. (6) Meta-learning approaches use a meta-learner to integrate predictions from multiple base models [34,35]. (7) Embedding-based methods transform time series into vector representations, enabling models to capture complex patterns [36,37].

2.4. Research Trends in Clustering-Based Hybrid Demand Forecasting for Retail

This section examines research trends in clustering-based hybrid methodologies. Table 1 summarizes the research trends in clustering-based hybrid demand forecasting for the retail sector.

Table 1. Research Trends in Clustering-Based Hybrid Demand Forecasting.

Ref.	Year	Data Source	Clustering Algorithms	Forecasting Algorithms	Performance Indicators	Feature
[38]	2000	Fashion Retail	K-Median	Linear Regression	revenue	Location, Sales, AVG Temp
[39]	2012	Supermarket Retail	K-Means	ARIMA, SARIMA, ANN	decrease in inventory	Customer/Transaction information
[5]	2014	Royal Air Force	SBC	KH, KH-SES	MAPE	Sales
[40]	2015	Food Retail	Bipartite Graph	Bayesian Network, MLP	MAPE	Warehouse/Product Properties, Sales
[41]	2019	Omni-channel Retail	K-Means	ANN	MSE	Online/Offline Sales
[7]	2020	IT e-commerce Retail	K-Means	OneR, Naive Bayes, KNN, RIPPER, C4.5, Rules-6	MAPE	e-commerce Sales
[42]	2020	Online Retail (Stock&Buy)	ClustAvg	Theta, ARIMA, MLP	Accuracy	Sales
[4]	2021	Fashion Retail	K-Means	ELM, SVR	MAPE, RMSE	Online/Offline Sales, Weather
[9]	2021	Kaggle Supermarket Retail	HAC	RF, XGBoost, LSTM+RF	RMSE, MAE	Transactions, Items, Stores, Holiday events, Oil prices
[6]	2022	M5 Walmart, Spare Parts Retail	SBC	SES, ARIMA, CROSTON	Inventory Decision Insights	Sales
[8]	2022	IT Hardware Retail	K-Means, AHC, GMM	ARIMA, RNN-LSTM	Cost	Sales/Stock/Customer Information
[25]	2022	Sports Retail	K-Means	LSTM, Prophet, Bayesian	Accuracy	Sales, Stores, Customers, Products, Delivery
[43]	2022	Fashion Retail	K-Means	SVM, RF, NN	MAE, RMSE	Sales
[44]	2022	Online Retail	K-Means	GLM	Accuracy	Sales, Product
[45]	2023	Bicycle accessories Retail	HAC	Regression	Accuracy	Sales, Product, Promotion
[46]	2023	Omni-channel Retail	DTW	ANN	RMSE	Sales
[47]	2023	Kaggle Retail	K-Means	Linear Regression, RF, XGBoost, LSTM	Accuracy	Sales, Customer, Product
[48]	2024	Food & Beverage Retail	GMM, HAC	RF	Accuracy	Sales, Customer, Region, Distribution, Product, Promotion
[49]	2024	Kaggle Retail	LSTM	RF	Accuracy	Sales, Product, Location
[50]	2025	Walmart Retail	K-Means	ARIMA	Improve inventory management	Sales, Product, Promotion

Studies from the early 2000s relied on rule-based classification and clustering techniques. These include research on store-level demand optimization using K-Median clustering [38], customer segmentation studies using ARIMA and Multi-Layer Perceptron (MLP) [39], Bipartite Graph clustering research [40], and the ClustAvg approach [42].

In the 2010s, data-driven, unsupervised learning techniques gained mainstream adoption. K-Means clustering was applied to customer and product segmentation using integrated online and offline sales data [41,43,44,47,50], while Hierarchical Agglomerative clustering (HAC) and Gaussian Mixture Model (GMM) were utilized for datasets with high demand uncertainty [8,9,48]. Dynamic Time Warping (DTW)-based pattern clustering represents a methodological extension that leverages time-series similarity [46].

Prediction algorithms have also evolved. Initially, traditional statistical techniques such as ARIMA, Simple Exponential Smoothing (SES), and regression analysis were used [6], and subsequently, machine learning-based approaches, including RF, XGBoost, Support Vector Machine (SVM), and Support Vector Regression (SVR) were introduced [9,43,47,48]. Recently, deep learning-based time series forecasting methods, such as LSTM and RNN-LSTM, have emerged [8,25,46]. There has also been research on ensemble architectures that combine LSTM and RF [9].

From a data perspective, most studies included sales and transaction data (e.g., sales volume, price, net sales) [44,46,50]. Subsequently, customer and store attributes (customer ID, store ID, location, demographic information, etc.) were integrated [38,39,47,48]. Recently, exogenous variables such as online/offline channel characteristics [4,41,46], promotions and events, weather, and oil prices have been incorporated [9,45,48].

Clustering-based hybrid demand forecasting in retail has evolved from simple rule-based approaches to attribute-based methods (K-Means clustering (K-Means), HAC, GMM) and pattern-based methods (DTW), while prediction algorithms have progressed from statistical methods to machine learning, deep learning, and ensemble approaches. Data has expanded beyond a sales- and customer-centric focus to a multidimensional framework encompassing channels, promotions, and external environmental factors. This suggests that hybrid approaches are becoming increasingly important in retail inventory management and decision support.

3. Methodology

This study proposes a hybrid clustering approach for retail demand forecasting that effectively captures diverse demand patterns in FMCGs retail data. The proposed methodology selectively utilizes rule-based and ML methods to construct prediction models for each cluster.

The rule-based method clusters data using the Syntetos–Boylan Classification (SBC). The ML method targets items that are difficult to categorize into clear demand types. First, time-series data are embedded to represent pattern similarity among products. Subsequently, product groups are clustered through unsupervised learning.

For both methods, the forecasting stage consists of two phases. Phase 1 selects an appropriate prediction algorithm for each cluster from ARIMA, RF, XGBoost, LSTM, and Autoformer.

In Phase 2, a hybrid model combining embeddings with the selected optimal algorithm is applied, and the resulting product-level predictions are compared across clusters. The overall framework is illustrated in Figure 1.

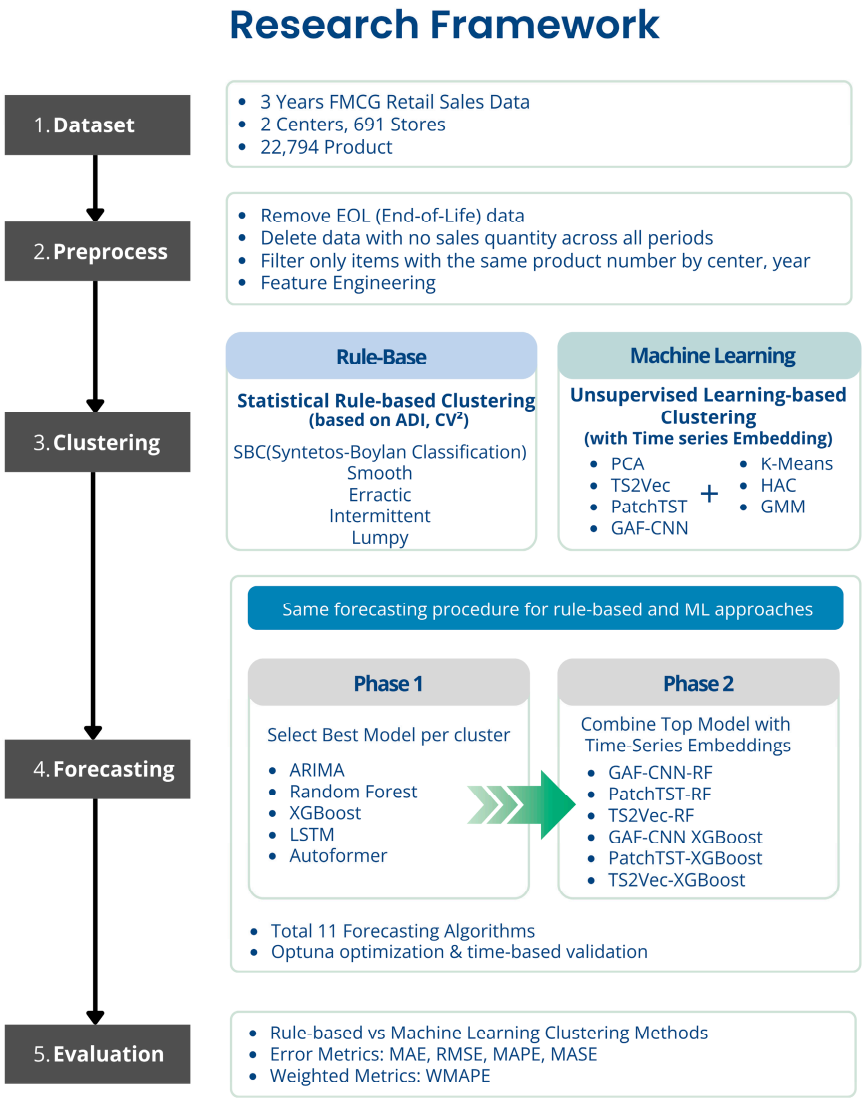


Figure 1. Research Framework.

3.1. Dataset

The dataset comprises three years (2020–2022) of weekly POS data from two distribution centers in the Korean FMCGs industry, covering 22,794 retail products across nine categories: tools, hobbies, stationery, food, cleaning, beauty, props, cooking, and storage. Each record includes domestic import classification, category, product number, product name, sales unit price, initial release date, start date, and sales quantity.

3.2. Data Preprocessing and Feature Engineering

The preprocessing stage involves constructing the modeling dataset through several steps. First, discontinued products and those with zero sales were excluded. Products were filtered to retain only those with consistent product numbers across all centers and years. Lag features were constructed, including one-week prior sales and a four-week moving average of sales. A holiday count variable was added to capture fluctuations in weekend sales. External variables—average temperature, humidity, wind speed, CPI, unemployment rate, WTI, and retail sales index—were incorporated. The complete feature set is illustrated in Table 2.

Table 2. Summary of Feature Engineering Variables.

Sortation	NO	Column	Description	Data Type
Time Series Feature (8)	1	DATE	DATE	DATETIME64
	2	YEAR	YEAR	INT64
	3	MONTH	MONTH	INT64
	4	WEEK	WEEK	INT64
	5	TOTAL_HOLIDAY_CNT	Number of Holidays	INT64
	6	LAG1	Sales Quantity 1 week ago	FLOAT64
	7	LAG1_4W_ROLLING_AVG	Sales Quantity over the Past 4 weeks	FLOAT64
	8	W_QTY	Sales Quantity	INT64
Product Feature (7)	9	CATEGORY	Product Category	OBJECT
	10	PRODUCT_CODE	Product Code	INT64
	11	PRODUCT_NAME	Product Name	OBJECT
	12	PRICE	Unit Price	INT64
	13	FIRST_SHIPMENT_DATE	Initial Release Date	INT64
	14	START_SALES_DATE	Sales Start Date	INT64
	15	ORIGIN_TYPE	Domestic/Import Classification	OBJECT
Weather Feature (3)	16	AVG_TEMPERATURE	Average Temperature	FLOAT64
	17	AVG_HUMIDITY	Average Humidity	FLOAT64
	18	AVG_WIND_SPEED	Average Wind Speed	FLOAT64
Economy Feature (4)	19	CPI	Consumer Price Index	FLOAT64
	20	UNEMPLOYMENT_RATE	Unemployment Rate	FLOAT64
	21	OIL_PRICE	West Texas Intermediate	FLOAT64
	22	RETAIL_SALES_INDEX	Retail Sales Index	FLOAT64

Categorical variables require encoding for machine learning algorithms [51], while numerical variables must be normalized to prevent disproportionate influence during model training. Normalization is particularly critical for distance-based and optimization-based models [52]. In this study, one-hot encoding was applied to categorical variables. Standard scaling was used during clustering to identify sales patterns, while Min-Max scaling was employed in the product-level forecasting stage.

3.3. Clustering Methods for Demand Forecasting

This study employs rule-based and ML-based clustering methods for demand forecasting. Clustering groups products with similar sales patterns, enabling cluster-specific prediction models. The rule-based method classifies products using predefined statistical criteria, while the ML method applies unsupervised embedding-based learning to identify data patterns. Clustering performance for the ML approach was evaluated using the Silhouette Coefficient (SC) and the Davies–Bouldin Index (DBI) [53]. The Silhouette Coefficient is an internal evaluation metric that simultaneously measures how closely each data point is

bound to other points within the same cluster and how well it is separated from points in other clusters. Its values range from −1 to 1, with higher values indicating a more clearly defined cluster structure [53].

In contrast, the Davies–Bouldin Index is an evaluation metric that assesses inter-cluster similarity by jointly considering inter-cluster separation and intra-cluster cohesion, with lower values indicating better separation between clusters and superior clustering quality [53]. Rule-based methods are effective for clear demand patterns, whereas ML-based methods excel with irregular or complex patterns [5–8]. The rule-based and ML methods are detailed in Sections 3.3.1 and 3.3.2, respectively.

3.3.1. Rule-Based Clustering

The rule-based method categorizes products into four types based on Average Demand Interval (ADI) and Squared Coefficient of Variation (CV^2) [7]: Smooth (low ADI, low CV^2), Intermittent (high ADI, low CV^2), Erratic (low ADI, high CV^2), and Lumpy (high ADI, high CV^2). This approach enables pattern differentiation, including the identification of intermittent demand, and facilitates cluster-specific model selection.

3.3.2. Machine Learning-Based Clustering

Machine learning-based clustering employs embedding-based unsupervised learning to transform time series into a latent vector space for pattern-based clustering. Direct comparison of raw time series is challenging due to the data's multidimensional and irregular nature, necessitating the use of embedding techniques [36]. While Principal Component Analysis (PCA) reduces dimensions and improves computational efficiency, it inadequately captures temporal flow and patterns [54]. Time series embedding preserves temporal structure in vector space, enabling effective pattern separation during clustering and facilitating model learning [55,56]. Although widely used in healthcare, finance, and sensor analysis, time series embedding remains underexplored in demand forecasting. This study compares PCA with advanced time-series embedding techniques to assess their effectiveness.

Four embedding techniques are employed:

Principal Component Analysis (PCA): Extracts principal components to reduce dimensionality while preserving core characteristics [54].

Gramian Angular Field–Convolutional Neural Network (GAF-CNN): Converts time series into GAF images representing temporal correlations as two-dimensional matrices, then extracts patterns via CNN while preserving temporal dependencies [36].

Patch Time Series Transformer (PatchTST): Divides time series into fixed-length patches processed by a Transformer encoder to capture both local and long-term patterns through self-attention mechanisms [37].

Time Series to Vector (TS2Vec): Generates latent vectors that capture global and local contexts via contrastive learning, providing generalized representations for prediction and analysis tasks [57].

Embedding vectors serve as input for three clustering algorithms:

K-Means clustering: A center-based algorithm that divides data into K clusters and iteratively updates each centroid to maximize within-cluster similarity [58].

Hierarchical Agglomerative clustering: A bottom-up algorithm that starts from individual data points and progressively merges similar clusters [59].

Gaussian Mixture Model: A probabilistic clustering technique that assumes data is composed of a mixture of multiple normal distributions [60].

3.4. Hybrid Forecasting Framework

The forecasting stage consists of three phases: training, validation, and testing, with both rule-based and ML methods following the same procedure. Phase 1 identifies the optimal model among five algorithms (ARIMA, RF, XGBoost, LSTM, and Autoformer), while Phase 2 evaluates hybrid models combining the selected model with embedding techniques (GAF-CNN, PatchTST, and TS2Vec).

Training. Sales quantity is used as the target variable, with 15 input features including time-series, product, seasonal, and economic variables. The dataset is split chronologically into training (January 2020–June 2022, 83.4%), validation (July–September 2022, 8.3%), and testing (October–December 2022, 8.3%) sets. One-hot encoding is applied to categorical variables, and min–max normalization is applied to numerical variables. To ensure fair comparison, the same data split is used across all stages. Clustering is performed on training data only to prevent data leakage, after which forecasting is conducted for each cluster. In Phase 1, five models are trained to identify the optimal model per cluster. In Phase 2, three embedding methods are integrated with the selected model, resulting in 11 hybrid models. Time-series embeddings transform input sequences into latent representations, thereby improving generalization and stability [56].

Validation. The validation set is used to tune hyperparameters and prevent overfitting. Hyperparameters are optimized using Optuna's Tree-structured Parzen Estimator (TPE) to minimize RMSE [61].

Testing. The test set is used to evaluate the final model performance using the optimized parameters. The results are presented in the following Evaluation section.

3.5. Forecasting Performance Evaluation

In the evaluation phase, the performance of the rule-based and ML methods was compared. Product-level forecasting accuracy was assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE) [34,45]. Each metric is defined as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{1}$$

MAE measures the average magnitude of forecasting errors.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2} \tag{2}$$

RMSE assigns greater weight to larger errors due to the squaring operation.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100 \tag{3}$$

MAPE evaluates relative forecasting accuracy as a percentage of actual values.

$$\mathrm{MASE} = \frac{\frac{1}{n}\sum|y_t - \hat{y}_t|}{\frac{1}{n-1}\sum|y_t - y_{t-1}|} \tag{4}$$

MASE scales forecast errors relative to a naive benchmark, enabling comparison across different time series.

Because the two methods employ different clustering approaches, simple average metrics may not adequately reflect differences in sales volume across products. Therefore,

this study also adopts the Weighted Mean Absolute Percentage Error (WMAPE) [62], defined as:

$$\text{WMAPE} = \frac{\sum_{i=1}^{K} n_i \cdot \text{MAPE}_i}{\sum_{i=1}^{K} n_i} \quad (5)$$

WMAPE evaluates overall forecasting performance by weighting each cluster according to its size.

4. Results and Discussion

4.1. Evaluation of ML-Based Clustering

In Section 4.1, clustering performance evaluation experiments for the ML method were conducted. The ML method transformed the input time series into vectors using time-series embeddings, including PCA, GAF-CNN, PatchTST, and TS2Vec. Subsequently, the embedding vectors were used as input for unsupervised learning-based clustering, applying K-Means, HAC, and GMM to cluster products with similar sales patterns. Clustering performance was evaluated using the Silhouette Coefficient (SC) and Davies-Bouldin Index (DBI), and the evaluation results are presented in Table 3.

Table 3. Performance Evaluation of Machine Learning-Based Clustering Methods.

NO	Embedding	Model	Feature Variables	K	SC	DBI
A Center						
1	PCA	K-Means	Time Series	3	0.4846	0.9419
2	PCA	K-Means	Time series + Sales	3	0.4785	0.9378
3	TS2Vec	K-Means	Time Series	3	0.5933	0.6866
4	PatchTST	K-Means	Time Series	5	0.6038	0.6218
5	GAF-CNN	K-Means	Time Series	2	0.5055	0.8910
6	PCA	HAC	Time Series	3	0.4192	0.9428
7	TS2Vec	HAC	Time Series	3	0.5902	0.6586
8	PatchTST	HAC	Time Series	5	0.5287	0.6359
9	PCA	HAC	Time series + Sales	3	0.3920	0.9592
10	GAF-CNN	HAC	Time Series	2	0.6434	0.7152
11	PCA	GMM	Time Series	3	0.1324	1.7658
12	TS2Vec	GMM	Time Series	3	0.1869	1.5709
13	PatchTST	GMM	Time Series	3	0.6488	0.8826
14	PCA	GMM	Time series + Sales	3	0.1446	1.6398
15	GAF-CNN	GMM	Time Series	2	0.5336	0.7954
B Center						
1	PCA	K-Means	Time Series	3	0.4699	0.9352
2	PCA	K-Means	Time series + Sales	3	0.4655	0.9288
3	TS2Vec	K-Means	Time Series	3	0.5782	0.7021
4	PatchTST	K-Means	Time Series	3	0.5950	0.6570
5	GAF-CNN	K-Means	Time Series	2	0.5358	0.8310
6	PCA	HAC	Time Series	3	0.5801	0.9680
7	TS2Vec	HAC	Time Series	3	0.5453	0.6849

Table 3. *Cont.*

NO	Embedding	Model	Feature Variables	K	SC	DBI
8	PatchTST	HAC	Time Series	4	0.5565	0.7410
9	PCA	HAC	Time series + Sales	3	0.4533	0.9360
10	GAF-CNN	HAC	Time Series	2	0.6023	0.7152
11	PCA	GMM	Time Series	3	0.2489	1.7731
12	TS2Vec	GMM	Time Series	3	0.1227	1.8953
13	PatchTST	GMM	Time Series	4	0.6339	0.6165
14	PCA	GMM	Time series + Sales	3	0.2179	1.6418
15	GAF-CNN	GMM	Time Series	2	0.4791	0.9863

Table 3 shows that PatchTST–GMM achieves the best clustering performance at both centers: Center A (K = 3, SC = 0.6488, DBI = 0.8826) and Center B (K = 4, SC = 0.6339, DBI = 0.6165), indicating strong cluster separation and consistency.

Clustering simplifies the data structure by grouping observations with similar patterns, thereby reducing heterogeneity and enabling more consistent pattern learning within each cluster. While clustering quality is evaluated using metrics such as SC and DBI [53], these reflect structural validity rather than forecasting performance. Accordingly, clustering is treated as a preprocessing step, and its effectiveness is ultimately assessed based on forecasting accuracy. This perspective aligns with Bala (2012), who showed that clustering-based demand partitioning improves forecasting accuracy compared to single-model approaches [39].

In contrast, PCA–GMM shows the poorest performance (Center A: SC = 0.1324, DBI = 1.7658), indicating its limitation in capturing temporal patterns. Although GAF-CNN–HAC achieves relatively high SC, it produces only two clusters from 12,661 products, limiting its ability to capture pattern diversity.

From an embedding perspective, time-series–based methods (TS2Vec, PatchTST, and GAF-CNN) outperform PCA-based approaches, demonstrating their effectiveness in capturing temporal dynamics. In addition, the single time-series experiment outperforms the multi-variable setting, suggesting that ML-based clustering can achieve stable performance even with minimal input features.

Overall, these results demonstrate that embedding-based clustering, particularly PatchTST–GMM, is most effective for capturing heterogeneous demand patterns.

4.2. Clustering Results: Rule-Based vs. ML Methods

Table 4 shows the cluster distribution from the rule-based method (SBC).

Table 4. Rule-based Method (SBC) Cluster Results.

Center	K	Product Quantity			
		Cluster 1 (Smooth)	Cluster 2 (Intermittent)	Cluster 3 (Erratic)	Cluster 4 (Lumpy)
A	4	9451	117	2139	954
B	4	9608	121	2017	915

Both Center A and Center B formed 4 clusters, with the Smooth type dominating at both centers. In contrast, the Intermittent and Lumpy types contained fewer products, indicating that irregular demand is concentrated in specific product groups.

Figure 2 presents the weekly average sales (W_QTY) trends for each cluster classified by the rule-based method (SBC), with the time axis expressed in YearWeek format. The *x*-axis denotes the year and week (e.g., 202001 corresponds to the first week of 2020), spanning January 2020 to December 2022.

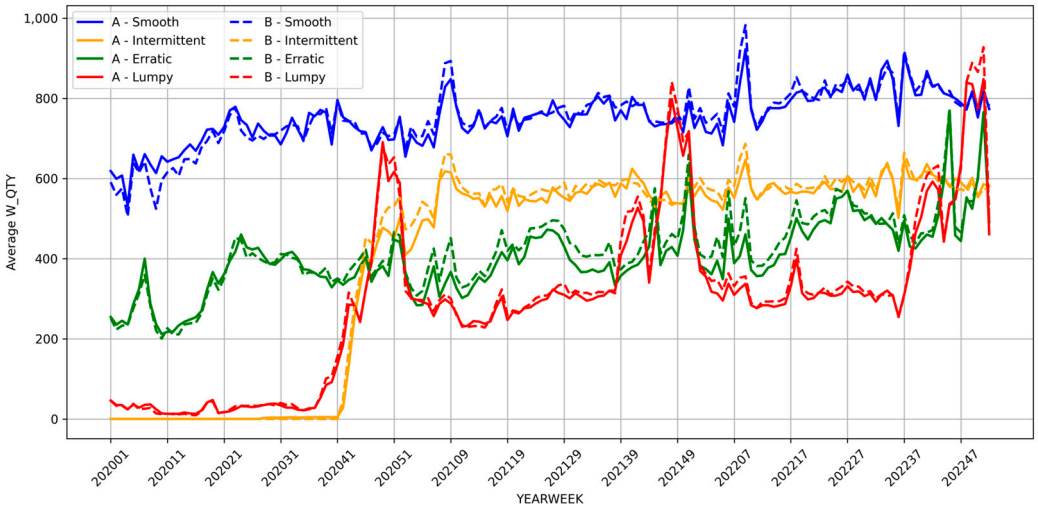


Figure 2. Average Sales Volume over Time Clusters from the Rule-Based Method.

The four clusters are clearly differentiated in terms of average level and variability, with similar pattern structures observed across both Center A and Center B. The Smooth cluster exhibits the highest and most stable demand, whereas the Intermittent, Erratic, and Lumpy clusters show lower averages with greater variability.

A noticeable increase in demand occurs after approximately week 40 in some clusters, attributed to the time lag between product launch and actual shipment. In addition, recurring demand spikes around weeks 40, 100, and 150 reflect seasonal effects, particularly pre-Christmas purchasing patterns typical in FMCGs retail data.

By incorporating both year and week information, structural shifts and seasonal patterns can be more clearly identified. Table 5 presents the product clustering results obtained using the ML method (PatchTST–GMM).

Table 5. ML Method (PatchTST-GMM) Cluster Results.

Center	K	Product Quantity			
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
A	3	2513	10,008	140	-
B	4	10,211	1801	214	435

Center A data were partitioned into 3 clusters, while Center B data were partitioned into 4 clusters. For Center A, the majority of products (79%) were assigned to Cluster 2, with substantially fewer products in Clusters 1 and 3. In contrast, Center B exhibited a more balanced distribution across four clusters, indicating greater heterogeneity in demand patterns. These results demonstrate that embedding-based clustering can autonomously identify and segment latent demand structures without predefined categories, adapting the cluster composition to the intrinsic characteristics of each dataset.

Figure 3 presents the weekly average sales (W_QTY) trends for each cluster classified by the ML-based method (PatchTST–GMM), with the time axis expressed in YearWeek format.

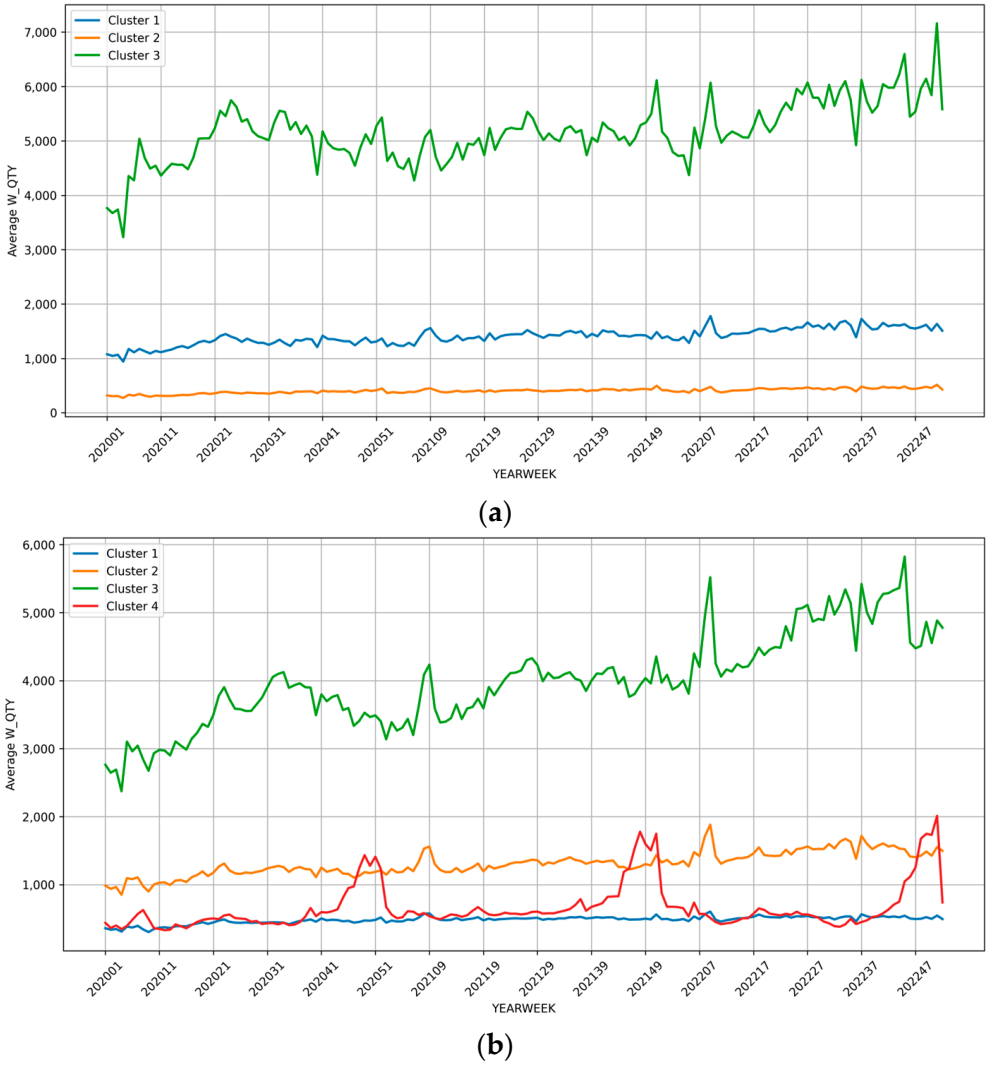


Figure 3. Average Sales Volume over Time by Clusters from ML Method (PatchTST-GMM). (**a**) Center A; (**b**) Center B.

Center A forms three clusters and Center B forms four clusters, each exhibiting distinct time-series characteristics and variability. Several clusters exhibit recurring short-term spikes and fluctuations, suggesting that ML-based clustering can capture latent demand patterns that rule-based methods miss.

Structural changes are also observed in certain clusters over time, reflecting non-stationarity in demand. Such changes may lead to distribution shifts between training and test periods, potentially affecting forecasting performance. To address this, the proposed clustering-based forecasting approach accounts for demand heterogeneity and structural changes, thereby enabling more stable, robust forecasting performance.

4.3. Performance Evaluation of Cluster-Level Forecasting Models

In Section 4.3, the performance of the forecasting algorithms, both rule-based and ML methods, was evaluated. The performance of 11 forecasting algorithms was compared, including single models in Phase 1 and hybrid embedding-combined models in Phase 2. The prediction results for each cluster are summarized in Tables 6–9.

Tables 6 and 7 compare forecasting performance across clusters using the rule-based method. Across 11 models, Phase 2 models incorporating time-series embeddings consistently outperform Phase 1 single models.

Table 6. Forecasting Results for Rule-Based Clusters (Center A).

Cluster (Product Quantity)	Evaluation	Phase 1 (Baseline Models)					Phase 2 (Proposed Models)					
							GAF-CNN		PatchTST		TS2Vec	
		ARIMA	RF	XGBoost	LSTM	Autoformer	XGBoost	RF	XGBoost	RF	XGBoost	RF
Cluster_1 (9451)	MAE	249.72	**164.83**	167.39	194.14	187.71	144.30	176.33	150.67	**143.11**	195.26	188.49
	RMSE	267.94	**182.65**	186.81	211.27	207.07	162.78	195.00	170.57	**163.87**	217.11	207.69
	MAPE	33.51	**23.78**	23.88	27.81	27.88	22.29	26.11	22.87	**22.08**	27.58	27.08
	MASE	2.85	**1.93**	1.94	2.25	2.12	1.62	1.98	1.67	**1.59**	2.13	2.08
Cluster_2 (117)	MAE	219.87	**84.81**	85.36	87.77	86.31	79.62	91.42	80.14	**74.40**	107.30	101.20
	RMSE	228.59	**93.52**	95.68	97.49	96.64	90.83	101.02	91.72	**84.45**	118.09	111.04
	MAPE	37.01	**14.09**	14.59	16.95	14.97	13.78	15.08	13.54	**12.52**	17.24	16.35
	MASE	4.54	**2.01**	2.08	2.04	1.84	1.75	1.94	1.74	**1.57**	2.16	2.03
Cluster_3 (2139)	MAE	538.26	268.12	**260.91**	429.59	351.77	245.42	342.29	**240.97**	256.65	348.58	321.90
	RMSE	601.13	329.70	**323.01**	489.66	411.27	311.45	409.58	**299.27**	318.44	416.21	384.66
	MAPE	64.60	51.52	**49.93**	62.49	55.10	46.63	57.46	**46.75**	49.69	54.50	54.17
	MASE	4.14	2.51	**2.43**	3.69	2.90	2.11	3.08	**2.12**	2.26	3.12	2.94
Cluster_4 (954)	MAE	494.67	344.53	**284.63**	507.02	435.01	378.21	434.45	**316.94**	343.04	477.27	388.87
	RMSE	572.37	404.03	**345.16**	590.65	504.97	456.39	509.96	**379.87**	412.59	567.51	461.58
	MAPE	64.62	55.59	**51.97**	56.12	57.14	51.71	59.23	**48.57**	54.59	57.01	55.15
	MASE	4.37	3.41	**2.96**	4.24	3.97	3.66	4.36	**3.34**	3.38	4.67	3.99

Table 7. Forecasting Results for Rule-Based Clusters (Center B).

Cluster (Product Quantity)	Evaluation	Phase 1 (Baseline Models)					Phase 2 (Proposed Models)					
							GAF-CNN		PatchTST		TS2Vec	
		ARIMA	RF	XGBoost	LSTM	Autoformer	XGBoost	RF	XGBoost	RF	XGBoost	RF
Cluster_1 (9608)	MAE	247.27	**160.37**	160.52	204.52	173.86	140.76	175.64	143.06	**139.50**	193.99	187.78
	RMSE	265.21	**178.00**	180.17	223.55	192.92	158.80	193.84	163.92	**160.64**	215.89	206.51
	MAPE	32.89	**23.11**	23.20	29.14	25.89	21.50	25.54	21.45	**21.19**	27.09	26.82
	MASE	2.83	**1.89**	1.88	2.33	1.98	1.59	1.99	1.61	**1.57**	2.14	2.09
Cluster_2 (121)	MAE	113.98	**81.46**	81.09	84.83	85.38	75.78	89.94	77.98	**73.04**	101.00	100.05
	RMSE	123.74	**91.59**	91.61	93.13	95.83	87.57	100.86	90.00	**83.61**	113.61	110.14
	MAPE	21.21	**14.96**	15.01	14.44	16.50	14.33	16.44	14.68	**13.81**	18.17	18.22
	MASE	2.68	**1.96**	1.97	2.00	1.87	1.67	1.97	1.72	**1.61**	2.09	2.11
Cluster_3 (2017)	MAE	540.17	270.38	**255.92**	414.88	362.22	239.26	314.49	**225.18**	248.65	349.55	314.20
	RMSE	600.93	330.30	**312.57**	480.92	418.20	303.88	377.92	**284.68**	306.31	416.28	375.86
	MAPE	63.97	50.81	**49.21**	62.55	54.68	45.45	54.11	**42.69**	48.66	52.96	51.76
	MASE	4.11	2.49	**2.36**	3.60	3.05	2.09	2.85	**2.00**	2.23	3.08	2.85
Cluster_4 (915)	MAE	538.76	373.24	**333.30**	493.95	466.53	401.90	486.88	335.64	360.39	**319.78**	393.09
	RMSE	618.66	432.83	**395.05**	576.42	537.86	480.18	567.47	394.63	420.36	**378.65**	462.77
	MAPE	64.79	56.52	**53.64**	60.82	57.01	50.66	61.01	47.55	54.82	**47.30**	53.45
	MASE	4.39	3.46	**3.21**	4.44	3.99	3.64	4.58	3.36	3.36	**3.27**	3.80

In Center A, the optimal model varies by cluster. PatchTST–RF performs best for the Smooth (Cluster 1) and Intermittent (Cluster 2) types, whereas PatchTST–XGBoost achieves superior performance for the Erratic (Cluster 3) and Lumpy (Cluster 4) types.

Similar patterns are observed in Center B. PatchTST–RF performs best for the Smooth and Intermittent types, while PatchTST–XGBoost is optimal for the Erratic type, consistent with Center A. For the Lumpy type, TS2Vec–XGBoost shows the best

performance, indicating that both PatchTST and TS2Vec embeddings are effective for high-uncertainty demand.

Table 8. Forecasting Results for PatchTST–GMM Clusters (Center A).

Cluster (Product Quantity)	Evaluation	Phase 1 (Baseline Models)					Phase 2 (Proposed Models)					
							GAF-CNN		PatchTST		TS2Vec	
		ARIMA	RF	XGBoost	LSTM	Autoformer	XGBoost	RF	XGBoost	RF	XGBoost	RF
Cluster_1 (2513)	MAE	594.25	369.31	**358.56**	475.31	389.54	**318.94**	405.52	323.48	323.89	442.94	420.97
	RMSE	643.86	416.08	**408.03**	528.69	438.54	**375.62**	456.93	376.98	372.98	500.89	471.79
	MAPE	35.18	25.42	**24.38**	30.84	27.10	**22.59**	27.56	22.97	23.25	28.94	28.18
	MASE	3.46	2.36	**2.29**	2.82	2.32	**1.91**	2.43	1.96	1.96	2.63	2.51
Cluster_2 (10,008)	MAE	218.89	133.37	**127.45**	184.76	160.09	122.60	152.02	**115.77**	121.24	167.50	155.05
	RMSE	240.09	152.85	**147.35**	206.68	180.89	145.54	173.58	**135.88**	140.60	191.39	175.87
	MAPE	42.37	32.21	**31.48**	38.78	35.23	29.46	34.90	**28.96**	30.43	35.71	34.89
	MASE	3.09	2.07	**2.00**	2.67	2.28	1.79	2.27	**1.74**	1.80	2.46	2.31
Cluster_3 (140)	MAE	2212.97	1497.40	**1434.19**	1917.62	1706.07	1385.69	1704.68	**1314.49**	1361.60	1739.86	1643.32
	RMSE	2542.16	1800.25	**1748.01**	2262.51	2029.11	1731.76	2036.26	**1624.23**	1677.37	2105.41	1973.25
	MAPE	36.04	28.05	**26.82**	33.99	31.03	26.74	31.17	**24.86**	25.93	30.19	29.57
	MASE	3.75	2.73	**2.62**	3.44	2.93	2.40	2.92	**2.32**	2.39	3.03	2.90

Table 9. Forecasting Results for PatchTST–GMM Clusters (Center B).

Cluster (Product Quantity)	Evaluation	Phase 1 (Baseline Models)					Phase 2 (Proposed Models)					
							GAF-CNN		PatchTST		TS2Vec	
		ARIMA	RF	XGBoost	LSTM	Autoformer	XGBoost	RF	XGBoost	RF	XGBoost	RF
Cluster_1 (10,211)	MAE	214.04	129.92	**129.58**	172.54	151.78	161.17	158.25	115.63	**113.75**	161.40	154.28
	RMSE	232.09	**147.10**	148.21	189.95	169.71	176.40	176.64	134.69	**133.70**	182.47	172.67
	MAPE	39.90	**29.88**	29.99	34.75	32.71	34.80	34.28	27.41	**26.92**	32.74	32.66
	MASE	2.97	**1.95**	1.95	2.44	2.16	2.59	2.33	1.69	**1.65**	2.29	2.22
Cluster_2 (1801)	MAE	545.60	336.74	**332.48**	440.27	375.90	**293.81**	366.07	299.57	306.37	417.85	385.64
	RMSE	584.83	376.68	**375.32**	479.81	416.80	**343.00**	407.69	344.84	346.71	468.42	427.69
	MAPE	35.19	24.94	**24.34**	31.20	27.98	**22.10**	26.54	22.44	23.74	29.34	27.87
	MASE	3.66	2.48	**2.45**	3.02	2.60	**2.01**	2.49	2.10	2.12	2.81	2.64
Cluster_3 (214)	MAE	1964.62	1143.86	**1129.15**	1410.41	1399.62	**996.41**	1334.76	1044.68	1050.29	1347.56	1320.13
	RMSE	2200.80	1378.82	**1360.12**	1652.65	1635.47	**1252.10**	1568.43	1302.25	1288.23	1615.57	1555.77
	MAPE	37.17	25.08	**24.17**	30.39	29.79	**21.26**	28.28	22.57	22.85	28.81	27.26
	MASE	4.50	3.07	**2.99**	3.55	3.35	**2.45**	3.33	2.63	2.54	3.39	3.18
Cluster_4 (435)	MAE	967.49	581.70	**502.84**	884.72	771.52	639.71	767.74	**459.61**	534.54	725.10	643.54
	RMSE	1112.19	691.19	**605.04**	1033.78	903.06	773.33	911.58	**559.07**	639.33	877.01	767.45
	MAPE	60.90	52.03	**47.74**	64.52	58.25	49.78	55.80	**42.93**	49.09	52.70	51.41
	MASE	4.65	3.48	**3.01**	4.60	4.12	3.99	4.56	**2.95**	3.35	4.29	3.91

Overall, RF-based models perform better for low-uncertainty demand, whereas XGBoost-based models perform better for high-uncertainty demand. These findings suggest that cluster-specific model selection is more effective than applying a single model across all demand types.

Tables 8 and 9 present forecasting performance across clusters using the PatchTST–GMM method. Consistent with the rule-based method, Phase 2 models combined with time-series embeddings outperformed the single models in Phase 1.

In Center A, XGBoost models combined with time-series embeddings consistently show superior performance across clusters. GAF-CNN–XGBoost performs best in Cluster 1 (low uncertainty), while PatchTST–XGBoost is effective in both low-uncertainty (Cluster 2) and high-uncertainty (Cluster 3) settings.

In Center B, the optimal model varies by demand uncertainty. PatchTST–RF performs best in Cluster 1 (lowest uncertainty), whereas GAF-CNN–XGBoost and PatchTST–XGBoost achieve better performance in clusters with higher uncertainty. Overall, RF-based models are more suitable for low-uncertainty demand, while XGBoost-based models are more effective for high-uncertainty demand.

Due to differences in cluster definitions and scales, direct comparison between rule-based and ML methods is limited. Therefore, a unified comparison using WMAPE across all products is provided in Section 4.6.

Table 10 presents a statistical comparison between the Phase 1 baseline and the proposed Phase 2 models. The results show that the proposed model consistently achieves lower MAE across all clusters. Both the paired *t*-test and the Wilcoxon test confirm that these improvements are statistically significant and robust. In SBC-based clusters, time-series embedding–enhanced RF and XGBoost models generally outperform their baseline counterparts, except in one case (SBC-A Cluster 4), where the baseline model performs better, suggesting that Phase 1 models may occasionally remain competitive depending on data characteristics. In ML-based clusters, Phase 2 models consistently achieve lower MAE with larger improvement margins, suggesting that embedding-based clustering effectively captures complex demand patterns.

It should be noted that formal statistical comparisons among models within the same phase (e.g., RF versus XGBoost in Phase 1, or GAF-CNN versus PatchTST versus TS2Vec in Phase 2) were not the primary focus of this study. The performance differences within each phase are generally small and consistent across clusters. Moreover, given the large sample sizes involved, even marginal differences may become statistically significant without necessarily implying meaningful practical improvements. Therefore, within-phase model selection in this study is based on MAE rankings rather than formal significance testing, and practitioners are encouraged to consider both statistical and practical significance when selecting among models within the same phase.

Table 10. Statistical Comparison between Phase 1 and Phase 2 Forecasting Models.

Method	Center	Cluster	N	Baseline	Proposed	Baseline MAE	Proposed MAE	Diff MAE	CI	*t*-Test *p*	Wilcoxon *p*
SBC	A	1	122,863	RF	PatchTST-RF	164.43	142.64	21.79	20.84, 22.74	0.000	0.000
	A	2	1521	RF	PatchTST-RF	84.34	73.59	10.75	8.22, 13.28	0.000	0.000
	A	3	27,807	XGBoost	PatchTST-XGBoost	257.16	236.43	20.73	17.45, 24.02	0.000	0.000
	A	4	12,402	XGBoost	PatchTST-XGBoost	280.38	310.19	−29.81	−37.66, −21.95	0.000	0.000
	B	1	124,904	RF	PatchTST-RF	160.00	138.66	21.34	20.48, 22.21	0.000	0.000
	B	2	1573	RF	PatchTST-RF	80.82	72.20	8.62	6.50, 10.75	0.000	0.000
	B	3	26,221	XGBoost	PatchTST-XGBoost	252.35	220.83	31.52	28.47, 34.58	0.000	0.000
	B	4	11,895	XGBoost	TS2Vec-XGBoost	329.76	313.30	16.45	9.78, 23.13	0.000	0.000

Table 10. *Cont.*

Method	Center	Cluster	N	Baseline	Proposed	Baseline MAE	Proposed MAE	Diff MAE	CI	*t*-Test *p*	Wilcoxon *p*
ML	A	1	32,669	XGBoost	GAF-CNN-XGBoost	356.49	315.86	40.63	37.03, 44.23	0.000	0.000
	A	2	130,104	XGBoost	PatchTST-XGBoost	126.43	114.06	12.37	11.66, 13.07	0.000	0.000
	A	3	1820	XGBoost	PatchTST-XGBoost	1426.71	1300.43	126.27	69.00, 183.55	0.000	0.000
	B	1	132,743	RF	PatchTST-RF	129.52	112.96	16.56	15.85, 17.28	0.000	0.000
	B	2	23,413	XGBoost	GAF-CNN-XGBoost	332.08	292.33	39.75	36.07, 43.42	0.000	0.000
	B	3	2782	XGBoost	GAF-CNN-XGBoost	1134.65	998.32	136.32	108.18, 164.47	0.000	0.000
	B	4	5655	XGBoost	PatchTST-XGBoost	481.88	439.23	42.64	33.59, 51.70	0.000	0.000

Overall, these results demonstrate that the proposed model delivers significant and consistent performance gains across diverse clusters and data environments, supporting the effectiveness of the hybrid clustering-based forecasting framework.

4.4. Feature Importance Analysis

Feature importance analysis was conducted to assess each feature's relative contribution to the model's predictive performance. We employed the Gain-based importance metric from XGBoost, which calculates the mean decrease in loss function attributable to splits on a given feature across all trees. Higher Gain values indicate that a feature provides more informative splits, thereby contributing more substantially to prediction accuracy [63]. The feature importance rankings for Centers A and B are summarized in Table 11.

Table 11. Feature Importance Analysis Results.

Feature	A Center Importance (%)	B Center Importance (%)
CATEGORY	30.99	45.76
LAG1	22.66	14.95
START_SALES_DATE	15.15	16.97
LAG1_4W_ROLLING_AVG	13	8.72
PRICE	6.34	4.71
FIRST_SHIPMENT_DATE	4.48	4.3
Subtotal (Top 6 Features)	92.62	95.41
Other Features (n = 9)	7.38	4.59

Other Features include ORIGIN_TYPE, RETAIL_SALES_INDEX, UNEMPLOYMENT_RATE, AVG_HUMIDITY, AVG_TEMPERATURE, CPI, OIL_PRICE, TOTAL_HOLIDAY_CNT, and AVG_WIND_SPEED.

Based on the results in Table 11, both centers showed CATEGORY, LAG1 (Sales quantity 1 week ago), START_SALES_DATE (Date of start of sale), LAG1_4W_ROLLING_AVG (Sales quantity for 4 weeks based on the previous week), PRICE (Unit price of sale), and FIRST_SHIPMENT_DATE (Initial release date) as the most influential variables. When these top six variables were combined, they accounted for 92.6% and 95.4% at Centers A and B, respectively.

At the center level, the importance of variable type differed. Center A showed higher importance for time series variables (LAG1, LAG1_4W_ROLLING_AVG) compared to

Center B. Center B showed higher importance for product-related variables (CATEGORY, START_SALES_DATE) compared to Center A. This indicates that time-series variables strongly influence Center A's demand structure, whereas product-related variables strongly influence Center B's demand structure.

The other nine macroeconomic and weather variables were generally of low importance at both centers, serving merely as auxiliary explanatory variables.

In summary, although the composition of the most important variables was similar across the two centers, the relative contribution weights differed. This demonstrates that, even when handling the same product categories, each center's demand structure shows distinct sales patterns. Utilizing these results, demand managers can make informed decisions based on the most influential variables.

4.5. Comparison of Rule-Based and ML Methods

Table 12 compares the forecasting performance of the rule-based (SBC) and ML (PatchTST–GMM) methods across all products. Cluster-level performance is evaluated using MAPE, while overall performance is assessed using WMAPE and WAMPE to reflect product-level and volume-weighted errors.

Table 12. Comparison of Rule-Based and ML Methods.

Center	Method	Weighted SUM (WMAPE)	WMAPE (%)	Weighted SUM (WAMPE)	WAMPE (%)
A	Rule-Based (SBC)	359,720.55	28.41	304,274.08	24.03
A	ML (PatchTST-GMM)	350,080.75	27.65	299,910.97	23.69
B	Rule-based (SBC)	334,649.76	26.43	289,848.53	22.89
B	ML (PatchTST-GMM)	368,130.97	29.08	311,734.52	24.62

The results show that the optimal method varies by demand characteristics. In Center A, where demand patterns are relatively stable, the ML (PatchTST–GMM) method achieves lower errors. In contrast, in Center B, which exhibits higher demand uncertainty, the rule-based (SBC) method performs better. These findings indicate that no single method is universally optimal.

Overall, the results support the effectiveness of a hybrid approach that adaptively selects clustering and forecasting strategies based on demand patterns. However, a limitation of this study is that the method selection is made post hoc, without a predefined criterion. This limitation highlights the need for a principled selection framework.

Table 13 summarizes the cluster-level statistical characteristics obtained from the ML-based clustering. The results indicate clear differences in demand patterns between the two centers. Center B contains clusters characterized by high variability and extremity, particularly Cluster 4, which exhibits high CV, zero ratio, ADI, kurtosis, and trend volatility, reflecting irregular and highly fluctuating demand patterns. In contrast, products in Center A are largely concentrated in relatively stable clusters with smoother demand behavior.

These differences in demand characteristics explain the performance variations observed in Table 12 and provide the basis for the proposed heuristic framework. Specifically, the presence of high-volatility and extreme clusters suggests the suitability of rule-based methods, whereas more stable demand structures favor ML-based approaches.

Table 14 presents the experimental savings achieved using the proposed diagnostic heuristic. The results show that the heuristic reduces the experimental search space by approximately 25–50% across different volatility scenarios.

Table 13. ML clustering statistics summary table.

Center	**A**			**B**			
Cluster	**1**	**2**	**3**	**1**	**2**	**3**	**4**
number_of_products	2513	10,008	140	10,211	1801	214	435
mean_W_QTY	1411.2	405.3	5197.6	484.4	1317.0	4027.3	669.4
mean_CV_W_QTY	0.4	0.6	0.7	0.6	0.5	0.6	1.6
mean_ZERO_RATIO	0.0	0.1	0.1	0.1	0.0	0.0	0.3
mean_ADI	1.0	1.1	1.1	1.0	1.0	1.1	1.9
mean_MAD	264.9	96.1	824.5	107.6	238.1	697.0	175.2
mean_KURTOSIS	3.1	3.1	7.4	2.7	3.6	6.1	12.2
mean_TREND_VOLATILITY	297.4	118.4	1288.2	117.7	276.6	834.8	473.6

Notes: mean_W_QTY denotes mean weekly sales quantity; mean_CV_W_QTY, mean coefficient of variation in weekly sales quantity; mean_ZERO_RATIO, mean proportion of zero-demand observations; mean_MAD, mean median absolute deviation.

Table 14. Experimental Savings under the Proposed Diagnostic Heuristic.

Case	**Volatility (A, B)**	**Required Experiments**	**Eliminated Experiments**	**Experimental Savings (Count, %)**
Case 1	Low, Low	ML Clustering (A, B) ML Forecasting (A, B)	Rule-based Clustering & Forecasting (A, B)	4/50%
Case 2	Low, High	ML Clustering (A, B) ML Forecasting (A) Rule-based Clustering & Forecasting (B)	ML Forecasting (B) Rule-based Clustering & Forecasting (A)	3/37.5%
Case 3	High, Low	ML Clustering (A, B) Rule-based Clustering & Forecasting (A) ML Forecasting (B)	ML Forecasting (A) Rule-based Clustering & Forecasting (B)	3/37.5%
Case 4	High, High	ML Clustering (A, B) Rule-based Clustering & Forecasting (A, B)	ML Forecasting (A, B)	2/25%

These findings indicate that the volatility structure derived from ML-based clustering can serve as an effective prior diagnostic indicator for selecting between rule-based and ML-based forecasting strategies. This diagnostic approach enables substantial reductions in experimental cost while maintaining performance.

Consistent with existing literature, the results confirm that no single forecasting approach is uniformly superior and that performance depends on data characteristics. This study extends this insight by showing that clustering strategies should also be selected according to demand characteristics.

Previous studies have typically adopted a single clustering paradigm. For example, Petropoulos and Kourentzes [5] and Li et al. [6] focused on rule-based approaches, while Chen and Lu [4] employed ML-based clustering without comparison to rule-based alternatives. As a result, the relative effectiveness of different clustering strategies under varying demand conditions has not been systematically examined.

In contrast, this study jointly considers clustering strategy and forecasting model selection. The results show that ML-based clustering performs better under low demand volatility, whereas rule-based clustering is more effective under high uncertainty. These findings highlight that no single clustering paradigm is universally optimal and emphasize the importance of context-sensitive hybrid approaches.

4.6. Robustness Validation Under Alternative Data Splitting

To assess the impact of temporal data splitting on the findings, an additional robustness analysis was conducted using an alternative time-series split (training: 2020–2021, validation: January–June 2022, test: July–December 2022).

Cluster-level forecasting performance was re-evaluated, focusing on the representative Phase 1 models (RF, XGBoost) and Phase 2 models (PatchTST–RF, PatchTST–XGBoost). Table 15 summarizes the overall performance across clustering methods and distribution centers for WMAPE and WAMPE.

Table 15. Robustness Validation Results Under Alternative Data Splitting.

Center	Method	Weighted SUM (WMAPE)	WMAPE (%)	Weighted SUM (WAMPE)	WAMPE (%)
A	Rule-based (SBC)	356,626.44	28.17	313,516.71	24.76
A	ML (PatchTST-GMM)	343,664.48	27.14	270,475.29	21.36
B	Rule-based (SBC)	334,585.68	26.43	294,933.17	23.29
B	ML (PatchTST-GMM)	340,044.67	26.86	316,253.30	24.97

The results show that the relative performance patterns remain consistent with the original findings: Center A favors the ML-based approach, while Center B favors the rule-based approach. At the cluster level, PatchTST–RF performs better in low-variability clusters, whereas PatchTST–XGBoost performs better in high-variability clusters. In addition, embedding-based models consistently achieve lower errors than baseline models, indicating their effectiveness in capturing latent demand structures.

Overall, these findings demonstrate that the proposed framework is robust to alternative data splitting and maintains stable performance under potential structural changes, such as those during the COVID-19 period. They further support the importance of selecting forecasting models based on demand characteristics rather than relying on a single universal approach.

5. Conclusions

The purpose of this study is to improve the accuracy of demand forecasting in the FMCGs multi-product retail industry. To this end, a hybrid clustering model was proposed that selectively employs rule-based and ML clustering methods based on data characteristics. Experiments using POS data from two distribution centers, spanning 12,661 products over three years, showed that the ML method (PatchTST-GMM) achieved low error rates in Center A, while the rule-based (SBC) method demonstrated superior performance in Center B. This suggests that, even within the same industry, the optimal framework can vary with data characteristics and sales distribution.

Performance evaluation of 11 algorithms in the forecasting phase revealed that time-series embedding-combined XGBoost models (GAF-CNN-XGBoost, PatchTST-XGBoost, TS2Vec-XGBoost) performed well in clusters with high demand uncertainty, whereas time-

series embedding-combined RF models (PatchTST-RF) showed superior performance in clusters with stable sales patterns. This demonstrates that the two methods can be used in complementary ways, depending on the sales distribution of product groups. The study reconfirmed that the SBC system, traditionally used in intermittent demand forecasting, is effective for patterns with high demand uncertainty and validated that embedding-based unsupervised learning clustering methods can be an effective alternative for learning stable patterns.

This study has the following academic and practical significance: First, we propose a hybrid methodology that selectively integrates rule-based and ML methods to classify products based on demand pattern similarity and enhance forecasting performance in retail environments with heterogeneous demand patterns, validated using actual retail data. Although GAF-CNN, PatchTST, and TS2Vec are increasingly adopted for forecasting applications, their use in FMCGs retail demand forecasting remains limited. This study contributes a validated case demonstrating the effectiveness of state-of-the-art time-series embedding techniques in real retail forecasting problems.

Second, the effectiveness of the traditional intermittent demand classification system (SBC) and the time-series embedding-based unsupervised clustering (PatchTST-GMM) was evaluated using the weighted mean absolute percentage error (WMAPE). Experiments demonstrated that ML-based forecasting is more effective under low demand uncertainty, whereas the rule-based approach outperforms under high demand uncertainty, providing practical grounds for data-driven method selection.

Third, an explainable model was developed through XGBoost Feature Importance analysis. When applied to retailer data without promotional effects, the analysis clearly identified the key independent variables driving predictions, with the top six variables accounting for over 92% of the explained variance.

Fourth, hyperparameter optimization was automated using the Optuna library in Python 3.10.6, minimizing average sales distribution error within each cluster. This approach substantially reduced the burden of repetitive tuning and enabled rapid derivation of optimal hyperparameters tailored to each cluster's demand characteristics.

From the perspective of practitioners and managers, this methodology can serve as a criterion for selecting appropriate approaches for retail demand forecasting. Rule-based and ML methods can be selectively applied based on the data's demand pattern, helping achieve optimal clustering and forecasting. Additionally, variable importance analysis provides insight into which independent variables most contribute to predictions, enabling managers to make informed decisions. Hyperparameter automation through the Optuna library enables rapid response to changing market conditions. The experimental results show that forecasting accuracy varies across product classifications, and managers can use these insights to adjust demand forecasting strategies for different product lines.

This study has several limitations. First, the analysis is based on data from only two distribution centers, which limits generalizability to other countries, distributors, or industry sectors. Second, exogenous variables showed relatively low importance, and important demand drivers—such as promotions, inventory levels, competitors' activities, and local events—were not fully incorporated. As a result, the findings may depend on the specific variable set and model configuration used in this study. Third, the SBC relies on ADI and CV^2 computed from historical demand, making it sensitive to the selected sample period. Consequently, variations in the observation period may lead to different cluster assignments for the same products.

Future research should expand the proposed framework across broader geographic regions and industry sectors to enhance generalizability. As this study focused on weekly forecasting, comparative analyses across short-, medium-, and long-term horizons would

be valuable. Incorporating store-level clustering alongside product clusters could better capture region-specific demand characteristics, and integrating richer contextual variables—such as promotions, price elasticity, and inventory dynamics—would further improve forecasting accuracy and model comprehensiveness.

Author Contributions: Conceptualization, J.-H.K. and N.-W.C.; methodology, J.-H.K.; validation, J.-H.K. and N.-W.C.; formal analysis, J.-H.K.; investigation, N.-W.C.; resources, J.-H.K.; data curation, J.-H.K.; writing—original draft preparation, J.-H.K.; writing—review and editing, N.-W.C.; visualization, J.-H.K.; supervision, J.-H.K.; project administration, N.-W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset analyzed in this study is proprietary and was provided by a retail company under a confidentiality agreement. Due to contractual and commercial sensitivity, the raw data cannot be publicly shared.

Acknowledgments: This research was supported by Seoul National University of Science and Technology.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Al Orbani, M. SKU Time Series Forecasting Methods for FMCGs. Master's Thesis, Rochester Institute of Technology, Dubai, United Arab Emirates, 2022. Available online: https://repository.rit.edu/theses/11172 (accessed on 10 January 2026).
2. Abolghasemi, M.; Gerlach, R.; Tarr, G.; Beh, E. Demand Forecasting in Supply Chain: The Impact of Demand Volatility in the Presence of Promotion. *arXiv* **2019**, arXiv:1909.13084. [CrossRef]
3. Mejia, S.; Aguilar, J. A Demand Forecasting System of Product Categories Defined by Their Time Series Using a Hybrid Approach of Ensemble Learning with Feature Engineering. *Computing* **2024**, *106*, 1765–1784. [CrossRef]
4. Chen, I.-F.; Lu, C.-J. Demand Forecasting for Multichannel Fashion Retailers by Integrating Clustering and Machine Learning Algorithms. *Processes* **2021**, *9*, 1578. [CrossRef]
5. Petropoulos, F.; Kourentzes, N. Forecast Combinations for Intermittent Demand. *J. Oper. Res. Soc.* **2015**, *66*, 914–924. [CrossRef]
6. Li, L.; Kang, Y.; Petropoulos, F.; Li, F. Feature-Based Intermittent Demand Forecast Combinations: Bias, Accuracy and Inventory Implications. *arXiv* **2022**, arXiv:2204.08283. [CrossRef]
7. Afifi, A.A. Demand Forecasting of Short Life Cycle Products Using Data Mining Techniques. In *Artificial Intelligence Applications and Innovations*; Maglogiannis, I., Iliadis, L., Pimenidis, E., Eds.; IFIP Advances in Information and Communication Technology; Springer: Cham, Switzerland, 2020; Volume 583, pp. 151–162. [CrossRef]
8. Puspita, R.; Wulandhari, L. Hardware Sales Forecasting Using Clustering and Machine Learning Approach. *IAES Int. J. Artif. Intell.* **2022**, *11*, 1074–1084. [CrossRef]
9. Paruthipattu, S.P. Demand Forecasting Based on External Factors Using Clustering and Machine Learning. Master's Thesis, National College of Ireland, Dublin, Ireland, 2021. Available online: https://norma.ncirl.ie/id/eprint/6251 (accessed on 1 February 2026).
10. David, E.; Bellot, J.; Le Corff, S. HERMES: Hybrid Error-Corrector Model with Inclusion of External Signals for Nonstationary Fashion Time Series. *arXiv* **2022**, arXiv:2202.03224. [CrossRef]
11. Dincer, K.F.; Turgay, S. Balancing Demand and Supply: Inventory Allocation in FMCG. *Ind. Eng. Innov. Manag.* **2023**, *6*, 41–49. [CrossRef]
12. Olatunji, A.O. Leveraging Data Science for Demand Forecasting and Inventory Management: A Comprehensive Review. *J. Basic Appl. Res. Int.* **2025**, *31*, 29–38. [CrossRef]
13. Khakpour, A. Data Science for Decision Support: Using Machine Learning and Big Data in Sales Forecasting for Production and Retail. Master's Thesis, Østfold University College, Halden, Norway, 2020. Available online: https://hdl.handle.net/11250/2660428 (accessed on 25 March 2026).
14. Suddala, S. Dynamic Demand Forecasting in Supply Chains Using Hybrid ARIMA–LSTM Architectures. *Int. J. Adv. Res.* **2024**, *12*, 1167–1171. [CrossRef]
15. Punia, S.; Singh, S.P.; Madaan, J.K. Predictive Analytics for Demand Forecasting: A Deep Learning-Based Decision Support System. *Knowl.-Based Syst.* **2022**, *258*, 109956. [CrossRef]
16. Fattah, J.; Ezzine, L.; Aman, Z.; El Moussami, H.; Lachhab, A. Forecasting of Demand Using ARIMA Model. *Int. J. Eng. Bus. Manag.* **2018**, *10*, 1847979018808673. [CrossRef]

17. Tirkes, G.; Guray, C.; Celebi, N. Demand Forecasting: Comparison Between Holt-Winters Model, Trend Analysis and Decomposition Models. *Teh. Vjesn.* **2017**, *24*, 503–509. [CrossRef]
18. Wang, G. Sales Forecasting for Firms Based on Multiple Regression Model. In *Proceedings of the International Conference on Big Data Economy and Digital Management (BDEDM)*; SciTePress: Setúbal, Portugal, 2022; pp. 628–633. [CrossRef]
19. Lukman, A.F.; Farghali, R.A.; Kibria, B.G.; Oluyemi, O.A. Robust-Stein Estimator for Overcoming Outliers and Multicollinearity. *Sci. Rep.* **2023**, *13*, 9066. [CrossRef]
20. Ahmed, A.M. Accelerate Demand Forecasting by Hybridizing CatBoost with the Dingo Optimization Algorithm to Support Supply Chain Conceptual Framework Precisely. *Front. Sustain.* **2024**, *5*, 1388771. [CrossRef]
21. Roy, K.; Ishmam, A.; Abu Taher, K. Demand Forecasting in Smart Grid Using Long Short-Term Memory. *arXiv* **2021**, arXiv:2107.13653. [CrossRef]
22. Oliveira, J.M.; Ramos, P. Evaluating the Effectiveness of Time Series Transformers for Demand Forecasting in Retail. *Mathematics* **2024**, *12*, 2728. [CrossRef]
23. Zhang, G.P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing* **2003**, *50*, 159–175. [CrossRef]
24. Falatouri, T.; Darbanian, F.; Brandtner, P.; Udokwu, C. Predictive Analytics for Demand Forecasting—A Comparison of SARIMA and LSTM in Retail SCM. *Procedia Comput. Sci.* **2022**, *200*, 993–1003. [CrossRef]
25. Liu, Z.; Zhang, Z.; Zhang, W. A Hybrid Framework Integrating Traditional Models and Deep Learning for Multi-Scale Time Series Forecasting. *Entropy* **2025**, *27*, 695. [CrossRef]
26. Seyedan, M.; Mafakheri, F.; Wang, C. Cluster-Based Demand Forecasting Using Bayesian Model Averaging: An Ensemble Learning Approach. *Decis. Anal. J.* **2022**, *3*, 100033. [CrossRef]
27. Ozturk, Z.K.; Cetin, Y.; Isik, Y.; Cicek, Z.I.E. Demand Forecasting with Clustering and Artificial Neural Networks Methods: An Application for Stock Keeping Units. In *Modeling, Dynamics, Optimization and Bioeconomics IV*; Pinto, A., Zilberman, D., Eds.; Springer Proceedings in Mathematics & Statistics; Springer: Cham, Switzerland, 2021; Volume 365, pp. 275–292. [CrossRef]
28. Duan, G.; Dong, J. Construction of Ensemble Learning Model for Home Appliance Demand Forecasting. *Appl. Sci.* **2024**, *14*, 7658. [CrossRef]
29. Zhang, Y.; Ren, G.; Liu, X.; Gao, G.; Zhu, M. Ensemble Learning-Based Modeling and Short-Term Forecasting Algorithm for Time Series with Small Sample. *Eng. Rep.* **2021**, *4*, e12486. [CrossRef]
30. Smyl, S. A Hybrid Method of Exponential Smoothing and Recurrent Neural Networks for Time Series Forecasting. *Int. J. Forecast.* **2020**, *36*, 75–85. [CrossRef]
31. Büyükşahin, Ü.Ç.; Ertekin, Ş. Improving Forecasting Accuracy of Time Series Data Using a New ARIMA–ANN Hybrid Method and Empirical Mode Decomposition. *arXiv* **2018**, arXiv:1812.11526. [CrossRef]
32. Rhif, M.; Ben Abbes, A.; Martínez, B.; Farah, I.R. Veg-W2TCN: A Parallel Hybrid Forecasting Framework for Non-Stationary Time Series Using Wavelet and Temporal Convolution Network Model. *Appl. Soft Comput.* **2023**, *137*, 110172. [CrossRef]
33. Hassanpouri Baesmat, K.; Farrokhi, Z.; Chmaj, G.; Regentova, E.E. Parallel Multi-Model Energy Demand Forecasting with Cloud Redundancy: Leveraging Trend Correction, Feature Selection, and Machine Learning. *Forecasting* **2025**, *7*, 25. [CrossRef]
34. Cawood, P.; van Zyl, T.L. Feature-Weighted Stacking for Nonseasonal Time Series Forecasts: A Case Study of the COVID-19 Epidemic Curves. *arXiv* **2021**, arXiv:2108.08723. [CrossRef]
35. Godahewa, R.; Bergmeir, C.; Webb, G.I.; Montero-Manso, P. An Accurate and Fully-Automated Ensemble Model for Weekly Time Series Forecasting. *arXiv* **2020**, arXiv:2010.08158. [CrossRef]
36. Molina-Tenorio, Y.; Prieto-Guerrero, A.; Rodriguez-Colina, E.; Vásquez-Toledo, L.A.; Olvera-Guerrero, O.A. Gramian Angular Field and Convolutional Neural Networks for Real-Time Multiband Spectrum Sensing in Cognitive Radio Networks. *Sensors* **2025**, *25*, 3580. [CrossRef]
37. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. *arXiv* **2022**, arXiv:2211.14730. [CrossRef]
38. Fisher, M.; Rajaram, K. Accurate Retail Testing of Fashion Merchandise: Methodology and Application. *Mark. Sci.* **2000**, *19*, 266–278. [CrossRef]
39. Bala, P.K. Improving Inventory Performance with Clustering-Based Demand Forecasts. *J. Model. Manag.* **2012**, *7*, 23–37. [CrossRef]
40. İşlek, İ.; Öğüdücü, Ş.G. A Retail Demand Forecasting Model Based on Data Mining Techniques. In Proceedings of the 2015 IEEE 24th International Symposium on Industrial Electronics (ISIE), Buzios, Brazil, 3–5 June 2015; pp. 55–60. [CrossRef]
41. Pereira, M.M.; Frazzon, E.M. Towards a Predictive Approach for Omni-Channel Retailing Supply Chains. *IFAC-PapersOnLine* **2019**, *52*, 844–850. [CrossRef]
42. Benhamida, F.Z.; Kaddouri, O.; Ouhrouche, T.; Benaichouche, M.; Casado-Mansilla, D.; López-de-Ipiña, D. Stock&Buy: A New Demand Forecasting Tool for Inventory Control. In Proceedings of the 2020 5th International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 23–26 September 2020; pp. 1–6. [CrossRef]

43. Giri, C.; Chen, Y. Deep Learning for Demand Forecasting in the Fashion and Apparel Retail Industry. *Forecasting* **2022**, *4*, 565–581. [CrossRef]
44. Cohen, M.C.; Zhang, R.; Jiao, K. Data Aggregation and Demand Prediction. *Oper. Res.* **2022**, *70*, 2597–2618. [CrossRef]
45. van Ruitenbeek, R.E.; Koole, G.M.; Bhulai, S. A Hierarchical Agglomerative Clustering for Product Sales Forecasting. *Decis. Anal. J.* **2023**, *8*, 100318. [CrossRef]
46. Soltani, M.; Khatami Firouzabadi, S.M.A.; Amiri, M.; Hajian Heidary, M. Proposing an Integrated Approach for Omnichannel Demand Forecasting Using Machine Learning–Time Series Clustering with Dynamic Time Warping Algorithm and Artificial Neural Networks. *Res. Prod. Oper. Manag.* **2023**, *14*, 121–140. [CrossRef]
47. Malik, A.; Dargar, G.; Sharma, A.; Pandey, P. Predictive Analysis for Retail Shops Using Machine Learning for Maximizing Revenue. In Proceedings of the 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 17–19 May 2023; pp. 126–133. [CrossRef]
48. Mitra, R.; Saha, P.; Tiwari, M.K. Sales Forecasting of a Food and Beverage Company Using Deep Clustering Frameworks. *Int. J. Prod. Res.* **2023**, *62*, 3320–3332. [CrossRef]
49. Mikkilineni, B.S.; Madala, U.; Bonthagorla, R.S.; Parikala, Y.P.; Kumar, V.P.; Kishore, V.K. An Experimental Study on Prediction of Revenue and Customer Segmentation. In *Proceedings of the 2024 8th International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 4–6 January 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 500–507. [CrossRef]
50. Stylianou, T.; Pantelidou, A. A Machine Learning Approach to Consumer Behavior in Supermarket Analytics. *Decis. Anal. J.* **2025**, *16*, 100600. [CrossRef]
51. Poslavskaya, E.; Korolev, A. Encoding Categorical Data: Is There Yet Anything 'Hotter' Than One-Hot Encoding? *arXiv* **2023**, arXiv:2312.16930. [CrossRef]
52. Pinheiro, J.M.H.; Oliveira, S.V.B.; Silva, T.H.S.; Saraiva, P.A.R.; de Souza, E.F.; Godoy, R.V.; Ambrosio, L.A.; Becker, M. The Impact of Feature Scaling in Machine Learning: Effects on Regression and Classification Tasks. *arXiv* **2025**, arXiv:2506.08274. [CrossRef]
53. Yin, H.; Aryani, A.; Petrie, S.; Nambissan, A.; Astudillo, A.; Cao, S. A Rapid Review of Clustering Algorithms. *arXiv* **2024**, arXiv:2401.07389. [CrossRef]
54. Gao, J.; Hu, W.; Chen, Y. Revisiting PCA for Time Series Reduction in Temporal Dimension. *arXiv* **2024**, arXiv:2412.19423. [CrossRef]
55. Liang, Z.; Zhang, J.; Liang, C.; Wang, H.; Liang, Z.; Pan, L. A Shapelet-Based Framework for Unsupervised Multivariate Time Series Representation Learning. *Proc. VLDB Endow.* **2023**, *17*, 386–399. [CrossRef]
56. Irani, H.; Ghahremani, Y.; Kermani, A.; Metsis, V. Time Series Embedding Methods for Classification Tasks: A Review. *arXiv* **2025**, arXiv:2501.13392. [CrossRef]
57. Yue, Z.; Wang, Y.; Duan, J.; Yang, T.; Huang, C.; Tong, Y.; Xu, B. TS2Vec: Towards Universal Representation of Time Series. *arXiv* **2021**, arXiv:2106.10466. [CrossRef]
58. Kanungo, T.; Mount, D.M.; Netanyahu, N.S.; Piatko, C.D.; Silverman, R.; Wu, A.Y. An Efficient K-Means Clustering Algorithm: Analysis and Implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881–892. [CrossRef]
59. Murtagh, F.; Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classif.* **2014**, *31*, 274–295. [CrossRef]
60. Syntetos, A.A.; Boylan, J.E.; Croston, J.D. On the Categorization of Demand Patterns. *J. Oper. Res. Soc.* **2005**, *56*, 495–503. [CrossRef]
61. Garine, R. Enhanced E-Commerce Demand Prediction through Ensemble Models and Optuna-Based Hyperparameter Optimization. In Proceedings of the 2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI), Wardha, India, 29–30 November 2024; pp. 1–7. [CrossRef]
62. Chukwuemeka, U.M.; Nnalue, A.D.; Obiekwe, S.J.; Maruf, F.A.; Anakor, A.C.; Moses, M.O.; Amaechi, C.; Okonkwo, U.P.; Amaechi, I.A. Comparative Validity Assessment of Three Android Step Counter Applications: A Semi-Structured Laboratory-Based Study. *BMC Digit. Health* **2025**, *3*, 20. [CrossRef] [PubMed]
63. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.