Journal of Information & Systems Management

DLINE JOURNALS

ISSN: 2230 – 8776

JISM 2026: 16 (2)
https://doi.org/10.6025/jism/2026/16/2/75-87

Advanced Retail Analytics Using Market Basket Mining, Product Networks, and Time-Series Forecasting

Dit Suthiwong
Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
s5507011966036@email.kmutnb.ac.th

ABSTRACT

The rapid expansion of digital retail has generated extensive transactional data, creating unprecedented opportunities for advanced, data-driven decision-making. Traditional Market Basket Analysis (MBA) and association rule mining frequently struggle with large, sparse datasets, often producing trivial rules while overlooking temporal and network-level purchasing dynamics. To address these limitations, this study introduces an integrated analytical framework that synergizes association rule mining, product network analytics, temporal modeling, clustering, and time-series forecasting. Leveraging a synthetic dataset of 30,000 grocery transactions spanning four months, the pipeline systematically extracts meaningful product affinities and visualizes complex purchasing pathways through association rule graphs and Sankey diagrams. A product co-occurrence network subsequently identifies structural purchasing communities and central hub items that drive strategic cross-selling opportunities. Temporal trend analysis captures fluctuating customer transaction volumes, while a regression-based forecasting model extrapolates future demand trajectories to optimize inventory planning and workforce scheduling. By synthesizing these complementary analytical layers, the framework transcends conventional basket analysis to deliver a holistic understanding of retail ecosystems. The findings demonstrate how multi-dimensional analytics significantly enhance next-basket recommendation engines, inform store layout optimization, and improve demand prediction accuracy. Ultimately, this unified approach equips retailers with actionable insights to personalize marketing strategies, streamline operational efficiency, and adapt proactively to evolving consumer behaviors in highly competitive market environments.

Keywords: Market Basket Analysis, Association Rule Mining, Product Network Analytics, Time-Series Forecasting, Customer Purchasing Behavior, Retail Analytics, Recommendation Systems, Clustering Techniques

Received: 30 September 2025, Revised 30 December 2026, Accepted 18 January 2026

1. Introduction

The rapid growth of e-commerce and digital retail platforms has generated massive volumes of transactional data, creating new opportunities for data-driven decision-making in retail analytics. Retail organisations increasingly rely on advanced analytical techniques to understand customer purchasing behaviour, optimise product placement, improve recommendation systems, and forecast future demand. Among these techniques, Market Basket Analysis (MBA) has become one of the most widely adopted approaches for identifying hidden relationships among products and understanding customer buying patterns.

Market basket prediction, which involves recommending items for a customer's next purchase based on current or historical shopping behavior, has emerged as an important application of intelligent retail systems. According to [1], Market Basket Analysis commonly employs association rule mining techniques on large-scale sales transaction data to identify meaningful relationships among purchased items.

Market basket analysis has evolved into a benchmark analytical technique for understanding market structures and product interdependencies. Traditional association rule mining methods identify frequent itemsets and generate association rules that reveal co-purchasing behavior among products. Despite its widespread adoption, conventional association rule mining faces several challenges when applied to large, sparse retail datasets.

Tan [2] argued that association rule mining is not always the most suitable method for analysing large market-basket datasets because the corresponding transaction matrices are typically sparse and computationally expensive to process. These limitations often lead to the generation of numerous trivial rules with limited practical insight. To overcome these challenges, Tan summarized real-world sales transaction data into a time-series representation, thereby enabling temporal analysis of purchasing behavior. Nevertheless, purely time-series-based approaches may also struggle with highly complex, large-scale market-basket datasets.

To address these limitations, the present study adopts an integrated analytical framework that combines association analysis, temporal modelling, network analytics, clustering techniques, and forecasting methods. This integrated approach is expected to provide a more comprehensive understanding of customer purchasing ecosystems and support intelligent retail decision-making.

2. Literature Review

2.1 Market Basket Analysis and Association Rule Mining

Several studies have explored the application of Market Basket Analysis to discover purchasing patterns and improve recommendation systems. [3] proposed a hybrid framework that integrates machine learning algorithms with traditional association rule mining to improve the overall effectiveness of market basket analysis. Using a publicly available dataset, the study investigated extending association rules and integrating deep learning techniques to enhance analytical performance.

Similarly, Agarwal, [4] employed Market Basket Analysis using the Apriori algorithm to identify associations among products from transactional datasets. The study highlighted the simplicity, efficiency, and practical applicability of the Apriori algorithm for introductory transaction data analysis.

Chaturvedi [5] developed a framework that combines traditional Market Basket Analysis techniques with Social Network Influence analysis. The study utilized the Apriori algorithm alongside machine learning approaches to compare the effectiveness of conventional association analysis with socially enhanced recommendation mechanisms.

By analysing customer transaction data and purchasing behavior, Shinde [6] identified relationships among products using Association Rule Mining, Clustering, and Classification techniques. The study aimed to improve recommendation accuracy and enhance customer satisfaction through data-driven personalization.

2.2 Customer Segmentation and Recommendation Systems

Dwarakanath [7] focused on the development of a machine learning-based system for customer segmentation and personalized product recommendation using e-commerce data. The research emphasized the importance of understanding customer behavior and purchasing patterns to enhance personalization strategies and improve user experience.

The study further noted that traditional segmentation techniques, which are often based on demographic characteristics or historical purchases, may fail to capture subtle customer preferences and real-time behavioral dynamics. To overcome these limitations, the proposed framework employed advanced ensemble learning techniques, particularly stacking models, to improve prediction accuracy.

Guidotti [1] introduced the concept of Temporal Annotated Recurring Sequences (TARS) and developed the TARS Based Predictor (TBP) for next-basket recommendation. The TBP model utilizes temporal purchase patterns to estimate customer inventory levels and recommend the most necessary items for subsequent purchases.

2.3 Time-Series Forecasting and Clustering Approaches

Time-series forecasting has become increasingly important in retail demand analysis and sales prediction. [8] highlighted the significance of time-series modelling for improving revenue generation and strategic planning in retail environments. Similarly, Leonard [9] demonstrated the effectiveness of forecasting techniques for optimizing business performance.

Recent studies have emphasized the role of deep learning models in retail forecasting. Abbasimehr [10] reported that Long Short-Term Memory (LSTM) networks can effectively capture temporal dependencies in sales data, while [Pavlyshenko [8] demonstrated the effectiveness of model stacking techniques in improving forecasting accuracy. Other studies further confirmed the growing importance of advanced forecasting techniques in demand prediction.

In addition to forecasting, clustering methods have also been widely explored for pattern discovery and customer segmentation. Martinez Alvarez et al. [11] applied k-means clustering to electricity demand datasets using internal validity indices. However, these indices were unable to effectively capture concurrent directional changes within the data.

Bandara et al. [12] proposed feature based clustering methods for Recurrent Neural Networks (RNNs) that improved interpretability and resilience to noisy data. Nevertheless, the framework aggregated short-term fluctuations that are often essential for dynamic retail pricing analysis.

Javed et al. [13] benchmarked multiple clustering algorithms on static datasets; however, the study lacked the retail sales and pricing context necessary for e-commerce applications.

Image clustering approaches have also attracted research attention. Studies such as [14, 15] combined deep feature learning with k-means clustering [16] for image representation and grouping. Although these methods demonstrated strong performance in image based domains, their application to time-series retail datasets remains limited, with most implementations focusing primarily on satellite imagery [17, 18].

2.4 Research Gap

Although substantial progress has been made in market basket analysis, recommendation systems, clustering, and forecasting, several limitations remain in the existing literature. Traditional association rule mining methods often generate large numbers of trivial associations when applied to sparse transactional datasets. Similarly, many forecasting models focus exclusively on individual product trajectories while overlooking group-level purchasing dynamics and interconnected product ecosystems.

Existing clustering methods also exhibit limitations in capturing concurrent changes in sales and pricing behavior, particularly within dynamic retail environments. Furthermore, image based clustering techniques and deep learning frameworks have rarely been adapted for retail oriented time series analysis. [19]

Therefore, there remains a need for a unified analytical framework that integrates association analysis, clustering, network analytics, and time-series forecasting to capture both individual and collective purchasing behaviors in retail systems.

2.5 Proposed Framework

The proposed framework addresses these research gaps by integrating association rule mining, temporal analysis, product network analytics, clustering techniques, and forecasting models into a comprehensive retail analytics architecture.

The framework first applies Market Basket Analysis to identify frequent itemsets and product relationships within transaction datasets. Network analytics are then employed to model product interactions and purchasing ecosystems. Temporal modelling techniques are subsequently used to capture evolving purchasing patterns.

To improve product grouping, the framework adapts image clustering concepts using a custom distance metric capable of capturing concurrent price sales shifts. This enables more accurate and domain specific clustering of retail products.

Finally, advanced forecasting models, including machine learning and deep learning approaches, are incorporated to predict future sales trends and customer purchasing behavior. By combining these analytical layers, the proposed framework provides a more holistic understanding of retail dynamics and supports intelligent business decision making.

3. Dataset

3.1 Dataset Overview

The Retail Transaction Dataset [20] comprises 30,000 unique synthetic retail transactions collected from a simulated grocery store environment. The data was synthetically generated to emulate realistic customer shopping baskets and purchase patterns, making it particularly suitable for market basket analysis, association rule mining, product affinity modeling, recommendation systems, and other retail analytics tasks.

The dataset is provided as a single CSV file (products.csv) containing 538,441 bytes. It is fully synthetic and does not contain any personally identifiable information. The data is licensed under CC0: Public Domain, allowing unrestricted use for research and commercial purposes.

3.2 Data Structure and Variables

Each row represents a single customer transaction (shopping basket) and includes the following fields:

• TransactionID: Unique identifier for each transaction (integer, range 1–30,000).

• CustomerID: Anonymized customer identifier (e.g., C546, C385), allowing for potential multi-transaction customer behavior analysis.

• Products: A comma-separated string listing all items purchased in that transaction. Products span diverse grocery categories including:

• Beverages (Soda, Juice, Cola, Water, Milk)

• Dairy & Frozen (Cheese, Yogurt, Ice Cream, Butter)

• Proteins (Chicken, Sausage, Fish, Minced Meat, Eggs)

• Pantry Staples (Cereal, Rice, Beans, Chickpeas, Lentil, Bread, Flatbread with Meat)

• Produce (Orange, Banana, Strawberry, Apple, Potato, Tomato, Onion, Cucumber)

• Snacks & Sweets (Chips, Chocolate, Cookie, Cracker, Pizza)

• Household & Personal Care (Soap, Shampoo, Detergent, Dish Sponge)

• Timestamp: Date of the transaction in YYYY-MM-DD format (spanning January to April 2025).

3.3 Summary Statistics

The dataset captures 30,000 transactions with an average basket size of approximately 7–8 products per transaction (the exact mean can be derived from the parsed data). Key characteristics include:

• High-frequency items: Cereal appears in over 23% of transactions, followed by Ice Cream, Chicken, Soda, Juice, Cheese, Soap, and Beans (each appearing in ~16–17% of baskets).

• Strongest product association: Cereal and Milk co-occur in 2,852 transactions (lift > 2.5), representing a prominent breakfast bundle.

• Temporal coverage: Transactions are distributed across four months in 2025, enabling basic seasonality or trend analysis if desired.

• Sparsity: Typical of market basket data, with thousands of unique product combinations but recurring strong co-occurrence patterns.

3.4 Data Generation and Suitability for Research

The synthetic nature of the dataset was designed to incorporate realistic conditional purchase probabilities (e.g., dairy items frequently purchased with breakfast staples, proteins with sides). This controlled generation process ensures clean, noise-free data ideal for benchmarking algorithms in:

- Association rule mining (e.g., Apriori, FP-Growth)
- Graph-based product network analysis
- Collaborative filtering and next-basket recommendation
- Customer segmentation based on basket composition

Because the data is synthetic, researchers should note that it does not reflect real-world external factors such as pricing, promotions, store layout, or macroeconomic influences, though internal purchase affinities are modeled realistically.

3.5 Preprocessing Notes

For analysis, the Products column should be split into individual items and either one-hot encoded or used to construct a transaction-item matrix. No missing values are present. The dataset is ready for immediate use in Python (pandas, mlxtend, networkx) or R.

4. Analysis

4.1 Association Rule Graph

The association rule graph visualizes the relationships among frequently co-purchased products extracted from the transactional dataset. Each node in the graph represents a product, while the edges denote strong associations between products based on transaction co-occurrence frequency. Products connected by thicker or denser edges indicate stronger purchasing relationships among customers.

The graph was generated using product-pair co-occurrence analysis, where highly frequent product combinations were selected and visualized as a network structure.

The figure reveals significant product affinity patterns within customer purchasing behavior. Several products appear as central hubs, indicating that they are frequently purchased alongside many other items. Such products may represent essential or complementary retail goods.

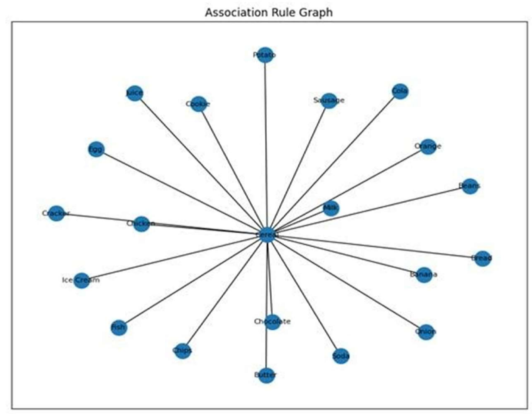

Figure 1. Association Rule Graph

Dense clusters in the graph indicate groups of products commonly purchased together, suggesting natural consumer shopping bundles. These associations can support:

- Cross-selling strategies
- Product placement optimization
- Bundle recommendation systems
- Intelligent retail marketing

The existence of highly interconnected nodes also demonstrates that customer purchasing behavior is not random but follows identifiable consumption patterns.

4.2 Time-Series Trend Analysis

The time-series trend figure illustrates the variation in transaction volume across time using timestamp information from the dataset. The x-axis represents chronological time, while the y-axis represents the number of transactions recorded during each period.

The visualization captures temporal purchasing activity and overall retail transaction dynamics.

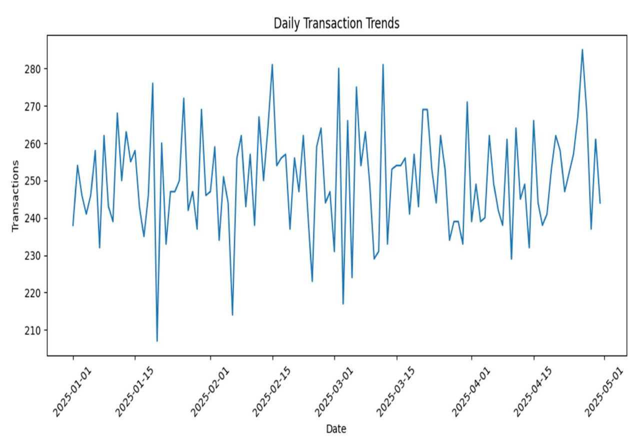

Figure 2. Time-Series Trend Analysis

The figure shows fluctuations in transaction activity over time, indicating changes in customer purchasing intensity. Peaks in the curve correspond to periods of higher shopping activity, while valleys indicate reduced transaction volume.

These temporal patterns may result from:

- Seasonal demand changes
- Weekend or holiday shopping behavior
- Promotional campaigns
- Consumer purchasing cycles

The overall trend provides insights into retail demand evolution and customer engagement behavior. If the graph shows gradual growth, it suggests increasing platform usage or customer activity. Conversely, declining trends may indicate reduced purchasing engagement during certain periods.

Such temporal analyses are valuable for:

- Demand forecasting
- Inventory management
- Workforce planning
- Retail campaign scheduling

4.3 Product Relationships

The Sankey diagram visualizes the flow and strength of relationships between frequently co-purchased products. Each node represents a product category or item, while the connecting flows represent transaction relationships between products.

The width of each flow corresponds to the magnitude of co-occurrence frequency between product pairs.

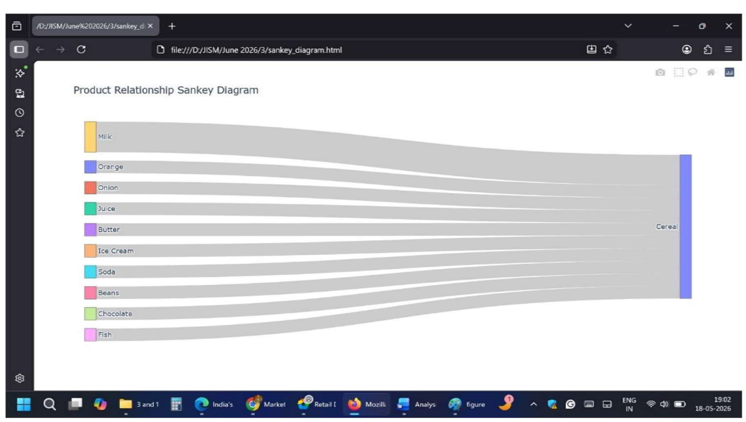

Figure 3. Sankey Diagram of Product Relationships

The Sankey diagram highlights dominant purchasing pathways within customer shopping baskets. Wider flows indicate stronger transactional relationships, suggesting that customers consistently purchase those products together.

The visualization demonstrates:

- Product transition behavior
- Consumer purchasing pathways
- Frequently connected retail items
- Dominant shopping combinations

Products with multiple outgoing or incoming flows can be interpreted as influential products within the retail ecosystem. These products may act as:

- Anchor products
- Gateway purchases
- Complementary goods

This analysis is highly useful for:

- Recommendation engines
- Store layout design
- Basket optimization
- Promotional targeting

The figure also reveals hidden transactional dependencies that may not be immediately observable using traditional statistical summaries.

4.4 Product Co-occurrence Network Graph

The product co-occurrence network graph models products as interconnected nodes based on simultaneous occurrence within customer transactions. Edges between nodes represent co-purchase relationships, while network density reflects the complexity of customer purchasing interactions.

The network was constructed using graph-based retail analytics techniques.

The network graph demonstrates the structural organization of product interactions within the retail dataset. Highly connected nodes indicate products with strong purchasing centrality, meaning they are commonly purchased alongside numerous other items.

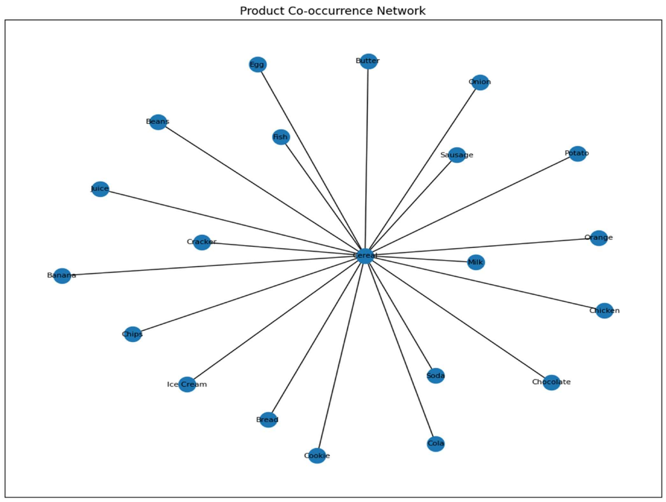

Figure 4. Product Co-occurrence Network Graph

Clusters within the network suggest:

- Product communities
- Consumer preference groups
- Thematic shopping behavior

The presence of tightly connected subnetworks indicates that customers tend to purchase products in structured groups rather than isolated selections.

This figure provides evidence for:

- Consumer behavioral segmentation
- Product ecosystem analysis
- Intelligent recommendation systems
- Graph-based retail analytics

From a business perspective, highly central products may serve as strategic products for:

- Promotion campaigns
- Inventory prioritization
- Demand stabilization

4.5 Forecasting Curve

The forecasting curve presents both observed historical transaction patterns and predicted future transaction behavior. The observed data are plotted alongside the forecasted trend generated using a regression-based predictive model.

The x-axis represents the temporal sequence, while the y-axis represents transaction volume.

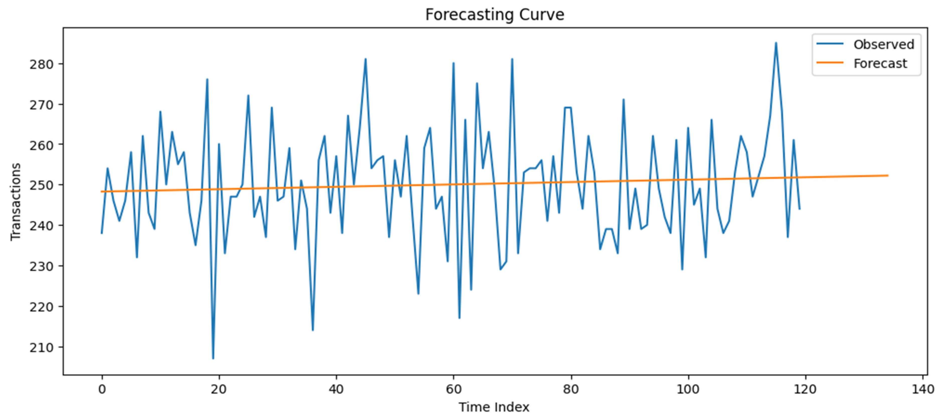

Figure 5. Forecasting Curve

The forecasting model identifies the underlying trend in customer transaction behavior and extrapolates future purchasing activity based on historical patterns.

The forecast curve demonstrates:

• Expected future transaction growth or decline

• Retail demand direction

• Predictive customer activity trends

If the forecast exhibits an upward trajectory, it suggests increasing customer engagement and transaction growth. A downward trend would indicate declining retail activity or reduced demand.

The smooth nature of the prediction curve reflects generalized long-term trends rather than short-term fluctuations. Such forecasting analyses are essential for:

• Inventory planning

• Demand management

• Revenue estimation

• Strategic retail decision-making

The forecasting results also demonstrate the applicability of machine learning and predictive analytics in retail transaction modeling.

5. Conclusion

Advanced retail analytics has become essential for understanding customer purchasing behavior and improving data-driven decision-making in modern retail systems. Although traditional Market Basket Analysis and association rule mining techniques provide valuable insights into product relationships, they face significant limitations when applied to large and sparse transactional datasets.

The reviewed studies demonstrate the growing importance of integrating machine learning, deep learning, clustering, and forecasting techniques into retail analytics frameworks. Existing research has explored recommendation systems, customer segmentation, temporal modelling, and clustering approaches; however, many frameworks remain limited in their ability to simultaneously capture product interdependencies, temporal purchasing dynamics, and group-level behavioral patterns.

The proposed integrated framework addresses these limitations by combining association analysis with other techniques.

References

[1] Guidotti, R., Rossetti, G., Pappalardo, L., Giannotti, F., Pedreschi, D. (2019). Personalized market basket prediction with temporal annotated recurring sequences. *IEEE Transactions on Knowledge and Data Engineering, 31*(11), 2151–2163.

[2] Tan, S. C., Lau, J. P. S. (2014). Time series clustering: A superior alternative for market basket analysis. In T. Herawan, M. Deris, J. Abawajy (Eds.), *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)* (Lecture Notes in Electrical Engineering, Vol. 285). *Springer*. https://doi.org/10.1007/978-981-4585-18-7

[3] Poojari, Rishika Ramesh. (2023). *Enhancing retail strategies: An integrated framework for market basket analysis using Apriori and MLP in consumer behavior modeling* (Master's thesis). *National College of Ireland, Dublin*.

[4] Agarwal, U., Afrin, F., Mishra, N., Yang, T., Zhao, Y., Rathore, B. (2026). Market basket analysis for retail businesses: The impact of promotions and discounts on customer buying behaviour. In A. Swaroop, V. Kansal, A. E. Hassanien (Eds.), *Proceedings of Sixth Doctoral Symposium on Computational Intelligence* (Lecture Notes in Networks and Systems, Vol. 1497). *Springer*.

[5] Chaturvedi, M., Loonkar, S., Desai, K. (2025). Enhancing retail strategy through market basket analysis and social network influence. *In 2025 International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA)* (p. 1–6). *IEEE*.

[6] Shinde, A. (2026). *Data mining for retail product recommendation using data analysis*. SSRN. https://doi.org/10.2139/ssrn.6733123.

[7] Dwarakanath, B., Sai Ganesh, S., Mathesh, M., Goutham Ram, N., Suresh Kumar, K. (2025). Investigating customer segmentation dynamics via market basket analysis: An extensive exploration of market insights. In V. S. Rathore, S. Murala, E. Tuba, M. M. Singh (Eds.), *Modern Practices and Trends in Expert Applications and Security* (Lecture Notes in Networks and Systems, Vol. 1379). *Springer*.

[8] Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *Data, 4*(1), 15. https://doi.org/10.3390/data4010015.

[9] Leonard, M. (2000). *Promotional analysis and forecasting for demand planning: A practical time series approach*.

[10] Abbasimehr, H., Shabani, M., Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers Industrial Engineering, 143*, 106435. https://doi.org/10.1016/j.cie.2020.106435.

[11] Martinez Alvarez, F., Troncoso, A., Riquelme, J. C., Aguilar Ruiz, J. S. (2011). Energy time series forecasting based on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering, 23*(8), 1230–1243. https://doi.org/10.1109/TKDE.2010.227.

[12] 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.

[13] Javed, A., Lee, B. S., Rizzo, D. M. (2020). A benchmark study on time series clustering. *Machine Learning with Applications, 1*, 100001.

[14] Ren, Y., Wang, N., Li, M., Xu, Z. (2020). Deep density-based image clustering. *Knowledge-Based Systems, 197*, 105841.

[15] Wu, J., Long, K., Wang, F., Qian, C., Li, C., Lin, Z., Zha, H. (2019). Deep comprehensive correlation mining for image clustering. *In Proceedings of the IEEE/CVF International Conference on Computer Vision* (p. 8150–8159).

[16] Caron, M., Bojanowski, P., Joulin, A., Douze, M. (2018). Deep clustering for unsupervised learning of visual features. *In Proceedings of the European Conference on Computer Vision* (p. 132–149).

[17] Khiali, L., Ndiath, M., Alleaume, S., Ienco, D., Ose, K., Teisseire, M. (2019). Detection of spatio-temporal evolutions on multi-annual satellite image time series: A clustering based approach. *International Journal of Applied Earth Observation and Geoinformation, 74*, 103–119.

[18] Lampert, T., Lafabregue, B., Gançarski, P. (2019). Constrained distance based k-means clustering for satellite image time-series. *In IGARSS 2019 – 2019 IEEE International Geoscience and Remote Sensing Symposium* (p. 2419–2422). *IEEE.*

[19] Alex, S., Satish, K., Anitha, V., Kalaiselvi, S., Santhi, S. (2025). Demand forecasting for market basket analysis. In D. I. Stroe, D. Nasimuddin, S. H. Laskar, S. K. Pandey (Eds.), *Emerging Electronics and Automation* (Lecture Notes in Electrical Engineering, Vol. 1237). Springer.

[20] Coban, B. (2025). *Retail transaction dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/bkcoban/retail-transaction-dataset