AI for Demand Prediction in Retail and E-Commerce

In this report, we explore cutting-edge AI applied in demand forecasting in retail and e-commerce. Being able to forecast demand accurately is very important for efficient supply chain management, inventory optimization and strategic decision making in the retail industry. This has a direct impact on customer satisfaction, company profits and operations.

We discuss seven research papers in this report that employ innovative techniques to the challenge of demand forecasting in retail and ecommerce. The techniques include traditional statistical methods, machine learning ensembles, deep learning and graph neural networks. These studies use data from different sources like historical sales, weather patterns, macroeconomic indicators, promotions, and consumer sentiment. These works generally portray how AI is transforming retail and ecommerce demand forecasting.

Demand Models for Supermarket Demand Forecasting

For retail and E-commerce businesses, demand forecasting is an essential work for balancing product availability, and minimising waste. Inaccurate forecasts can result in either lost loyal customers due to stockouts or increased costs due to overstocking, such as milk and bread in supermarkets.

In the traditional demand forecasting, Kerzel (2023) provides a sample entry point into this discussion. The author examines a model called a negative binomial distribution (NBD), as known as a Gamma-Poisson distribution, which is long-standing assumption that supermarket demand can be represented by Poisson process, in which the number of purchases in a given period follows a Poisson distribution and the interarrival times of purchases follow an exponential distribution. Because it is based on mathematical formular, using this assumption to predict future demand makes results interpretable. The NBD assumes customers have different personal buying rates (modelled by a Gamma distribution), and given their rate, each customer's purchases follow a Poisson process. This mixture creates more realistic demand variability than assuming all customers behave identically (Poisson distribution).

However, while the NBD captures consumer heterogeneity more effectively than Poisson distribution, the author also reveals its limitations. Firstly, the overall assumption that purchases in a supermarket is reasonably well to meet real data, but the NBD has failed to capture scheduled shopping behaviour, such as people would have more purchases at seven-day and fourteen-day intervals, which reflects NBD cannot predict customers' weekly shopping routines. Secondly, the heterogeneity across product categories: perishable and ready-to-eat foods, milk products and durable goods, such as household cleaning products. Statical distribution cannot capture the diversity of shopping behaviours across those product categories.

This is precisely where artificial intelligence, and in particular deep learning, offers significant potential. Deep learning models such as Long Short-Term Memory (LSTM) networks, Temporal Convolutional Networks (TCN), and Transformer-based architectures have already demonstrated strong performance in a wide range of sequence forecasting tasks. Unlike traditional statistical models, these methods do not rely on fixed distributional assumptions. Instead, they can automatically capture non-linear dependencies, long-term temporal dynamics, and interactions among multiple covariates. For example, LSTMs are able to learn from long historical purchase sequences to identify recurrent patterns, while Transformer-based models can scale to very large datasets and capture both short- and long-range dependencies in consumer behaviour. Moreover, probabilistic forecasting frameworks built on deep learning, such as DeepAR, allow practitioners to generate not only point predictions but also full predictive distributions, enabling better decision-making under uncertainty.

<u>Predictive analytics for demand forecasting: A deep learning-based decision support system</u>

Punia, S., & Shankar, S. (2022) introduces a novel ensemble model which combines a deep learning model (LSTM based) with a machine learning model (Random Forest). LSTMs capture linear and nonlinear temporal variations while random forest is used for capturing covariate-based variations in demand data. The model was tested on 4235 demand series from the packaged food products sector. The features of the model included structured and unstructured data like point-of-sales, promotions, weather, economic indicators, and internet media. The approach outperformed benchmark models across multiple error metrics like mean error, mean absolute error and mean squared error.

The proposed framework allows making forecasts for short, median and long-term planning. It uses this input data – point-of-sale, promotion, time-based, store-based and external indicators. The methodology followed includes performing dimensionality reduction using PCA, followed by modelling univariate sales time-series using LSTM and multivariate relationships using Random Forest. Multivariate relationships include sales data and the principal components. To combine predictions from the 2 algorithms, a genetic algorithm is used for deciding weights assigned to the predictions of both algorithm predictions. The objective function of GA is to minimize mean absolute error.

For generating long-term and short-term forecasts, weekly forecasts are aggregated or disaggregated using 2 separate algorithms. For long-term forecasts, a hierarchical reconciliation algorithm is used while for short-term forecasts, the author proposes using a temporal disaggregation method which is also called a demand sensing algorithm. This algorithm uses real-time high frequency data in a GLS model along with weekly sales data for making short-term forecasts.

To describe the nature of the big data the authors used, they describe it using the 5Vs framework of Big Data. The data comprises weekly sales data of 55 food items sold at 77 stores over a 3 years period thereby resulting in a dataset of 4,235 individual demand series. Two years of data were used for training the models and the remaining one year was kept aside as a test set.

The authors used 3 metrics for evaluation – RMAE, RMSE and RME. Based on the results – LSTM+RF model significantly outperformed others with RMAE of 0.3569 and RMSE of 0.4638.

Based on the results achieved, the authors highlight the many benefits offered by the system including cost savings and operational efficiency, data-driven decision making, organizational alignment and broad applicability of the system to other domains.

AI in E-Commerce Demand Forecasting: Algorithms, Metrics, and Insights

The paper by Aye, da Silva, and Pereira Mafia (2024) reviews forty-two studies on demand forecasting in retail and e-commerce. It shows a clear shift from old-fashioned statistical models like ARIMA and regression toward AI-based systems. These newer methods are not only more accurate but also more flexible, which matters a lot in today's unpredictable e-commerce market.

Traditional methods still have some use. For example, ARIMA or regression can work for stable, seasonal trends, but they fail when customer behaviour changes quickly. That is where machine learning and deep learning models come in. Random Forests, Gradient Boosting, and SVMs handle nonlinear patterns much better, while deep learning models like LSTMs and GRUs are great for time-series data. Hybrid approaches, such as combining ARIMA with LSTM, seem to give the best of both worlds. More experimental methods, like GANs, are even being used to simulate rare demand situations.

The studies compared in the paper mostly used error-based metrics. They were Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). This help measure how far predictions are from reality. RMSE is especially tough on big mistakes, while MAPE shows errors in percentage terms, which businesses find easier to understand. A few studies also reported R^2 or accuracy.

Here is a summary of how algorithms and metrics lined up:

Algorithm Type	Examples	Metrics	Performance
Traditional	ARIMA, Regression	MAE, RMSE, MAPE	Works only in stable markets
Machine Learning	Random Forest, GBM, SVM	MAE, RMSE, MAPE	Better than traditional, handles complexity
Deep Learning	LSTM, GRU, CNN	MAE, RMSE, MAPE	Lowest errors, best for time-series
Hybrid	ARIMA+LSTM, CNN-LSTM	MAE, RMSE, MAPE, R^2	Strongest balance
Generative AI	GANs, Transformers	Experimental	Promising, still early

Overall, the paper makes it clear: AI-based methods beat traditional models almost every time. Deep learning and hybrid methods achieve the lowest errors, especially when extra data like weather, holidays, or economic signals are included. In fact, accuracy can improve by 20% compared with models that only use sales history.

In conclusion, AI has changed demand forecasting from a simple statistical task into a core business strategy. While challenges remain—like data quality, cost, and making models more transparent—the direction is obvious. Forecasting in e-commerce will rely more on AI, not less.

Applying Machine Learning in Retail Demand Prediction—A Comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning

Introduction: The study examines the relative effectiveness of deep learning (more especially, LSTM) and tree-based ensemble learning techniques for predicting retail demand, particularly for perishable commodities. Using a sizable real-world dataset from a supermarket, the authors use both external (weather, COVID-19 lockdowns) and internal (pricing, promotions) aspects as characteristics. Fruits, fresh meat, and soft drinks are the three types of perishable goods they test.

This Study: This study evaluates the effectiveness of Artificial Intelligence (AI) techniques, specifically tree-based ensemble methods (Random Forest, XGBoost) and Long Short-Term Memory (LSTM) networks, for demand prediction in retail and e-commerce settings.

Context & Importance: Accurate demand forecasting is crucial in retail and e-commerce. It helps manage inventory, optimize supply chains, reduce overstock and stockouts, and improve customer satisfaction. Traditional statistical methods often struggle with the complex, non-linear, and seasonal nature of retail data. AI, especially machine learning and deep learning, can model these complex patterns more effectively.

Data and Scope: The paper uses real-world retail sales data, including time-series information on product demand over various periods. The authors prepare the data with feature engineering to capture temporal effects such as the day of the week, seasonality, promotions, and product metadata.

Findings: XGBoost consistently outperformed both LSTM and Random Forest in prediction accuracy across most metrics, including RMSE and MAE. LSTM performed reasonably well, particularly for products with strong timedependent sales patterns. Tree-based models were more robust and easier to train, especially with limited data or fewer temporal dependencies. In e-commerce settings where external features, like promotions and product features, are important, tree-based models used these features more effectively.

Implications for AI in Retail: AI techniques like XGBoost offer high accuracy, efficiency, and interpretability, making them suitable for realworld retail applications. Deep learning (LSTM) can be useful in scenarios with rich time-series data and longer sequences but may not always justify the complexity in retail situations. The study suggests that hybrid models or ensemble strategies that combine both approaches could potentially provide even better performance. Retailers and e-commerce platforms should choose their models based on data characteristics, business constraints, and forecasting horizons.

Sales Prediction Scheme Using RFM based Clustering and Regressor Model for Ecommerce Company

A machine learning-based sales prediction system created for e-commerce companies improves demand forecasting and hence supports data-driven marketing decisions. Different from typical time-series forecasting methods, the proposed approach incorporates heterogeneous data sources, such as product, regional, shipping, transaction, and customer behavior features. An important contribution of this study is the use of recently created customer-centric features, where RFM analysis is offered for the Recency, Frequency, and Monetary measures of customer actions, which are then employed in a K-means clustering procedure allowing the segmentation of customers into groups that exhibit distinct purchase behaviors.

The RFM clustering procedure yields three clusters: new customers with low engagement, occasional buyers with medium engagement, and loyal customers with high engagement. The clusters are included in supervised regressions through two strategies. First, separate models are built for every cluster, allowing predictions to be adapted specially to each customer group dynamic. Second, cluster membership is one-hot encoded as additional explanatory variables in the full data set. Both strategies aim to increase the predictive power by injecting customer behavior insights directly into the forecasting models.

Multiple regression algorithms have been analyzed, including linear regression, Random Forest, and XGBoost methods, upon the enriched datasets. Standard error measures, such as RMSE and R^2, are used for performance assessment. It is found that the XGBoost is always better performing than the other models, with high predictive accuracy, attaining R^2 values ranging from 98 to 99% and low RMSE scores. This confers that the use of unsupervised customer segmentation in conjunction with supervised regression techniques is highly successful in capturing both behavioral heterogeneity and the structural patterns of e-commerce sales data.

Now, a host of pragmatic benefits are made available for e-commerce companies with the implementation of our framework. Incorporating customer segmentation criteria within the predictive modeling permits the extraction of sufficiently rich knowledge to support targeted marketing strategies. For instance, marketing campaigns might be designed for high-value loyal customers, whilst engagement strategies could be implemented for new or occasional buyers. Moreover, the framework may be used to plan promotions at times when sales are expected to be low, so the establishments could contend with demand fluctuations in advance. Apart from marketing, the predictive insights from the forecast model could be applied in inventory management, logistics, and pricing schemes, thus making it an extensive tool in business decision-making.

These limitations of the proposed framework are also discussed in the paper: limited scope in marketing-oriented applications; a single data set used for testing purposes; and no exploration of other business functions, such as supply chain optimization or procurement planning. Further, though, relatively static clustering means customer group assignments are not dynamically updated as behavior changes, which may limit its applicability in the long run in a fast-changing e-commerce environment.

Based on the suggestions for future research directions, areas of interest concern the automation of cluster updates to reflect changes in customer behavior; testing across several datasets for generalizability purposes; and implementation of the forecasts in BI dashboards at an operational level in real time. Another avenue worthy of exploration is establishing a direct connection between prediction results and operational processes, such as inventory replenishment and logistics scheduling. Both of these would enhance robustness and practical utility of the suggested framework.

In conclusion, the paper proposes a novel fusion of unsupervised clustering and supervised regression models for sales prediction in e-commerce. Customer-oriented features together with machine learning techniques optimize predictive power, while also suggesting marketing and business avenues for realization.

AI-Driven Demand Prediction with Spatial GNNs

Accurate demand prediction is crucial for success in retail and e-commerce. It affects inventory management, logistics, and pricing. Traditional statistical models and basic machine learning methods often struggle with the complex mix of time and space factors that influence demand across different regions and product categories.

J. Li (2023) proposes an advanced AI framework using Spatial Graph Neural Networks (SGNNs). In this method, retail locations or online regions are modeled as graph nodes, while relational factors, such as proximity, demographics, and purchasing patterns, form the network edges. This setup enables SGNNs to understand both the spatial connections among stores and the time factors like seasonality, promotions, or sudden spikes in demand by combining graph and sequence learning techniques.

Performance Metrics - The SGNN model showed a significant improvement in predictive accuracy compared to traditional models and deep learning standards. Studies on SGNNs in retail demand prediction indicate:

Root Mean Squared Error (RMSE): Recent work shows that SGNNs have reduced RMSE by up to 15% compared to LSTM and by 5 to 10% against ARIMA. For instance, if ARIMA's RMSE is 0.61 and LSTM's is 0.54, SGNNs typically achieve RMSE values around 0.46 to 0.49, showing fewer and smaller forecasting errors.

Mean Absolute Error (MAE): SGNNs also consistently outperform benchmarks in this area, with absolute errors dropping from 0.37 (ARIMA) and 0.34 (LSTM) to about 0.29 in published SGNN studies.

Explained Variance / R^2: SGNNs showed higher R^2 scores, often above 0.90, indicating that their forecasts closely match the recorded demand patterns, particularly in dense, interconnected retail settings where spatial relationships are strong influencers.

Business Impact - SGNN-powered forecasts lead to better inventory planning, fewer stockouts, and lower excess costs. They also improve supply chain responsiveness, allowing for more effective targeted promotions, which ultimately increases profitability and customer satisfaction.

The paper highlights that how relationships between stores are defined (graph construction) is crucial for SGNN's performance and calls for future research to develop methods for even larger networks.

Retail Demand Forecasting: A Comparative Study for Multivariate Time Series

Accurate demand forecasting is very important for retail businesses. It directly affects supply chain efficiency, inventory management, and overall profitability. Previously, most forecasting approaches have relied only on historical sales data. However, this method overlooks the influence of broader economic conditions on consumer purchasing behavior. This paper reflects this gap by incorporating macroeconomic variables Consumer Price Index (CPI), Index of Consumer Sentiment (ICS), and unemployment rates into retail demand forecasting models.

Early forecasting methods, such as SARIMA, were widely applied because of their ability to capture seasonality and trends. More recent approaches support machine learning (ML) techniques such as Random Forests, Support Vector Regression, and Neural Networks, which outperform classical models. Studies also suggest that integrating consumer confidence and macroeconomic indicators enhances predictive accuracy. However, prior research often tested only one model (e.g., LSTM) and lacked broad comparative evidence. This study fills that gap by applying multiple machine learning models with and without macroeconomic variables.

The dataset used comes from Walmart (USA), covering five years, 3,049 products, and 10 stores across three states. It includes product details, prices, store data, promotions, events, and calendar effects. To enrich predictions, macroeconomic data were added from reliable sources such as the World Bank and the University of Michigan.

The authors applied preprocessing techniques such as filtering outdated records, imputing missing prices, and generating lagged/rolling features to capture sales trends and seasonality. Independent variables included both product-level attributes (e.g., sales, prices) and external indicators (CPI, ICS, unemployment).

Five machine learning models were compared: Lasso Regression, Ridge Regression, LightGBM, XGBoost, and Decision Tree Regression. Each was trained twice, first using only historical sales data, and then with added macroeconomic features. Performance was observed by using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

- **Overall performance:** Incorporating macroeconomic features yielded modest but consistent improvements in predictive accuracy for most models. For example, Lasso improved from RMSE **1.80239** to **1.79865** and MAE **0.88665** to **0.88479**, while Ridge improved slightly from RMSE **1.73869** to **1.73848** with MAE values almost unchanged around **0.845**.
- **Best performing model:** The best model is LightGBM with macroeconomic variables, as it achieved the lowest error values, improving RMSE from **1.71740** to **1.71504** and MAE from **0.84859** to **0.84742** compared to without macro variables.
- **Decision Tree regression:** Performed worse when macroeconomic features were added, as RMSE increased from **2.35698** to **2.36479** and MAE from **1.00046** to **1.00170**, suggesting some models may not capture these variables effectively.
- **Feature importance:** Across models, variables such as rolling averages, lagged sales, and product prices remained strong predictors. However, macroeconomic indicators (especially **CPI** and **ICS**) also appeared in the top features, demonstrating their explanatory power.

For businesses, this means forecasting systems should not only rely on internal sales history but also factor in external economic conditions. Among the tested methods, tree-based algorithms like LightGBM show the greatest potential. This integration empowers retailers to optimize inventory, reduce costs, and gain a competitive edge in a dynamic market.

Conclusion

Across all the studies reviewed, one thing is clear: demand forecasting in retail and e-commerce has shifted from being a simple statistical task to a data-driven strategic function powered by AI. Traditional methods like Poisson-based models or ARIMA still provide interpretable baselines, but they struggle with the complexity of modern consumer behavior, diverse product categories, and outside factors like promotions, weather, or economic indicators.

The research shows how machine learning, deep learning, hybrid models, and more recently, graph-based neural networks, are filling these gaps. Tree-based ensembles like XGBoost offer efficiency and accuracy when external features matter. Meanwhile, LSTMs and Transformer-based models are good at capturing long-term patterns. Hybrid models and ensemble strategies consistently do better than single techniques, showing the benefits of combining different strengths. Additionally, frameworks like RFM-based clustering bring customer-focused insights into forecasting. This helps predict not just “what” will sell but also “who” will buy. Spatial GNNs broaden the scope by including geographic and relational factors, showing that demand is affected not just by time but also by location.

What ties these methods together is their impact on business results. Better forecasts lead to less waste, improved inventory management, lower costs, targeted marketing, and ultimately, greater customer satisfaction. At the same time, the studies highlight ongoing challenges: ensuring data quality, updating models in changing environments, and balancing accuracy with clarity for practical use.

Overall, the evidence suggests that the future of demand forecasting will not rely on a single model but on flexible, AI-driven systems that combine various data sources and methods. As retail and e-commerce evolve in unpredictable markets, organizations that adopt these AI-driven strategies will gain a significant advantage, turning demand prediction from a reactive necessity into a proactive tool for growth and innovation.

Bibliography

Kerzel, U. (2023). Demand Models for Supermarket Demand Forecasting. International Journal of Supply and Operations Management, 10(1), 89–104. https://doi.org/10.22034/ijsom.2022.109044.2145

Punia, S., & Shankar, S. (2022). Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems, 258*, 109956. https://doi.org/10.1016/j.knosys.2022.109956

Aye, T. U., da Silva, E. R., & Pereira Mafia, M. (2024). Emerging technologies in e-commerce demand forecasting: Trends, practices, and future perspectives. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4917431

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. PLOS ONE, 13(3), e0194889. https://doi.org/10.1371/journal.pone.0194889

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts. https://otexts.com/fpp2/

Nasseri, M., Falatouri, T., Brandtner, P., & Darbanian, F. (2023). Applying Machine Learning in Retail Demand Prediction—A Comparison of Tree-Based Ensembles and Long Short-Term Memory-Based Deep Learning. Applied Sciences, 13(19), 11112–11112. https://doi.org/10.3390/app131911112

Li, J. (2023). Product Demand Prediction with Spatial Graph Neural Networks. International Journal of Artificial Intelligence in Retail, 14(2), 123–138.

Chan, T., & Chen, Y. (2024). Spatial Graph Neural Networks for Accurate Product Demand Prediction. CPL, 12(3), 1–8.

Additional figures and performance metrics adapted from: Probabilistic Demand Forecasting with Graph Neural Networks. arXiv preprint, 2022.

Abellana, D. P. M., Rivero, D. M. C., Aparente, M., et al. (2020). Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: a Philippine scenario. Journal of Tourism Future.

Bakas, D., & Triantafyllou, A. (2019). Volatility forecasting in commodity markets using macro uncertainty. Energy Economics, 8, 79-94.