

Predictive Analytics for Sales Forecasting in Emerging Market Retail: An Ensemble Learning and Time-Series Approach for A&Z Supermarket

Idowu Olugbenga Adewumi

Department of Computer and Information Science,
Faculty of Natural and Applied Science, Lead City University, Nigeria

E-mail: adexio2010@gmail.com

(Received: 21 January 2026, Revised: 16 February 2026, Accepted: 17 February 2026)

<https://doi.org/10.57260/csdj.2026.286583>

Abstract

Evaluation of forecasting models with 5000 daily sales records (2020-2023) from A&Z Supermarket located at Oyo State, Nigeria. Ten (10) product IDs from five (5) stores. The dataset incorporates a total of 11 predictive variables. This includes price, competitor price, macroeconomic index (simulated), promo indicators, holiday flags and rolling averages. The forecast models are assessed through a train/test split with a 70/30 ratio. They are forecasting for a 7-day horizon. Further, rolling window validation prevents leakage. According to descriptive analysis, the average sales quantity resulted in 26.49 sales units with a standard deviation of 11.58. The day of sales displays cumulative spikes, mainly during weekends as well as holidays. The research compares various traditional models such as ARIMA, SARIMA, and Linear Regression with machine learning models like Random Forest, XGBoost, LSTM, and GRU and has a RMSE ranging from 10.55 to 14.36 and R^2 values ranging from 0.740 to 0.884. Bagging, boosting, stacking techniques are used to improve the performance of predictive models. The stacking ensemble achieved the lowest RMSE (8.14). The performance of LSTM was observed to be best as depicted by (RMSE= 8.75, $R^2= 0.91$) as it outperformed Prophet and XGBoost. The lack of generalizability can be attributed to a narrow dataset that incorporates only a single retailer and a simulated macroeconomic index. An analysis of the business KPIs indicates a decline in inventory costs of 17.8% and of stockouts by 15.4%. Such improvements along with a 9.5% increase in ROI are encouraged through the use of LSTM (Long-Short Term Memory) investments. The results indicate that hybrid ensemble-time-series models are superior to conventional methods and offer important practical conveniences in optimizing inventory and forecasting demand in retail settings in emerging markets.

Keywords: Sales forecasting, Ensemble learning, Time-series decomposition, LSTM neural networks, XGBoost regressor

Introduction

In today's online marketplace predictive analytics in marketing and sales became more important than ever to be competitive plus respond to rapid changing demand conditions. More and more organizations are relying on forecasting techniques for their decisions regarding stock maintenance, pricing strategies, promotions timings and the return on investment. When companies don't forecast regularly, they either overstock their inventory, which increases their carrying costs, or understock inventory, leading to lost sales.

Traditional statistical techniques for forecasting, like Autoregressive Integrated Moving Average (ARIMA), exponential smoothing, linear regression, etc., have traditionally dominated supply chain and sales forecasting projects due to their ease of comprehension and relatively low data requirement (Wu, 2025). Yet, these classical methods often encounter difficulties in situations characterized by nonlinear demand trends, sudden changes (e.g., from promotions or outside impacts) and a large number of influencing variables (e.g., many product categories, many external covariates). For instance, research on the modeling of perishable goods and fresh products has shown that ARIMA-based models are effective in capturing seasonality/trending but are unable to adjust to varying demand and exclude external factors, such as a promotion or weather (Theodoridis & Tsadiras, 2025). In various studies, forecasts produced by ARIMA and similar conventional models exhibit greater inaccuracies than more flexible machine learning alternatives when the data demonstrate complex time relationships and erratic trends (Samal & Ghosh, 2026).

Given these constraints, we have seen a growing interest in machine learning (ML) techniques, and more recently ensemble learning approaches, for predicting sales demand. To capture non-linear time related changes and adding a number of predictors, models such as Long Short-Term Memory, Random Forest, Extreme Gradient Boosting and their hybrid forms are used. For example, the joint modeling of appliance demand using LSTM, Random Forest, and XGBoost in 2024 was more successful than using each independently in capturing changes in market demand. To the same end, comparative assessments of present-day ML models for foretelling retail sales indicate that tree-based ensembles often outperform neural networks in irregular retail data situations when preprocessing and data segmentation is performed carefully (Hobor et al. 2025).

The literature still has gaps despite these improvements. Specifically, methods that combine ensemble learning with traditional time-series forecasting have been underexplored, in particular in marketing settings where seasonal and trend and domain-specific factors might interact (for instance, promotions, external variables). The combined benefits of ensemble learning (that improve non-linearity, reduce variance and enhance robustness) and time series approaches (which bring trend/seasonality and improved interpretability) are promising but not sufficiently tested on different market types and data settings.

Objectives

The primary objective of this study is to assess the performance of forecasting models for sales demand in a retail setting using ensemble learning techniques combined with time-series decomposition. To achieve this, the following research questions and hypotheses have been formulated:

Research Question 1 (RQ1):

How do hybrid ensemble-time-series models compare to traditional statistical models (ARIMA, SARIMA, Linear Regression) in terms of forecasting accuracy for retail sales?

Hypothesis (H₁): Forecasting models based on hybrid ensemble learning techniques will outperform traditional statistical forecasting models (ARIMA, exponential smoothing) in terms of forecast accuracy (measured by RMSE, MAE, and MAPE).

Methodology: The study will use traditional models (ARIMA, SARIMA, Linear Regression) as baselines for comparison against machine learning models (Random Forest, XGBoost, LSTM, GRU) and ensemble techniques (bagging, boosting, stacking). The models' performances will be evaluated using error metrics: RMSE, MAE, MAPE, and R².

Expected Outputs: A comparative analysis presented in tables and figures, including error metrics (RMSE, MAE, MAPE, R²) for each model. The expected output is a performance table showing how each model compares in terms of forecasting accuracy.

Research Question 2 (RQ2):

Do hybrid ensemble-time-series models show greater resilience to disruptions such as promotions and seasonal changes than individual machine learning or statistical models?

Hypothesis (H₂): Hybrid ensemble-time-series models will demonstrate greater resilience to promotional and seasonal disturbances than individual machine learning or statistical models.

Methodology: This will be tested by comparing the models' performance during periods of promotions and holidays (using sales data from the A&Z Supermarket). The models will be validated over a rolling window to assess stability and response to seasonal and promotional effects.

Expected Outputs: Tables and figures showing the models' performance during promotional and seasonal events, including RMSE and R² values during these periods. The expected output is a graphical representation of how well models handle seasonal disruptions and promotions.

Research Question 3 (RQ3):

What is the impact of sales forecasting on business key performance indicators (KPIs) such as inventory cost, stockouts, overstock, and ROI?

Hypothesis (H₃): The incorporation of ensemble forecasting models will result in reduced inventory costs, stockouts, and overstock, while increasing ROI compared to conventional models.

Methodology: Business KPIs such as inventory costs, stockouts, overstock, and ROI will be calculated based on the forecasted sales and actual sales data. The results of each model will be compared to assess improvements in these KPIs.

Expected Outputs: A section on business impact with tables and figures showing changes in inventory cost, stockouts, overstock, and ROI across different models. The expected output will be a detailed table showing the percentage improvements in KPIs, with LSTM and ensemble methods expected to show the best results (Mejía & Aguilar, 2024).

Our contributions are empirical and theoretical. Empirically, the study presents evidence from various markets/datasets about the benefits of hybrid ensemble-time series approaches. Meanwhile, through the theoretical contribution, we help to understand how such designed approaches can optimally trade-off interpretability and predictive power to guide the practitioners from marketing analytics and operations planning.

In the world of online business, predictive analytics and demand forecasts help marketing and sales operations stay one step ahead of the competition. Forecasting techniques are relied upon increasingly by organizations for making decisions about investments,

inventory control, pricing strategies, timing of promotions, and ROI. The inability to forecast accurately can lead to excessive inventory with high carrying cost or insufficient inventory with lost sales and reduced confidence in the business.

Supply chain and sales forecasts frequently employ statistical forecasting methods such as Autoregressive Integrated Moving Average, exponential smoothing, and linear regression because these methods are able to be interpreted by humans and have a smaller data requirement. These methods, however, encounter challenges when employed in situations involving non-linear demand, unexpected changes (both promotional and otherwise), and a large number of variables. According to research done on perishable products (Giannopoulos et al., 2025), ARIMA-based models can adjust seasonality and trends, but not for variability in demand or any external shocks like a promotion of the product or due to weather effects. Machine learning (ML) models are more suitable than traditional models like ARIMA for data that has complicated time-based dependencies and erratic patterns since the models result in lower forecast errors (Mansur et al., 2025).

Machine learning methods, especially ensemble learning, have received increasing attention to tackle the problem in recent years.

The A&Z supermarket's daily need for stock replenishment forecasting is referred to as its decision context. Forecasting inventory management based on a 7-day horizon is required for replenishment planning to take place. This research adopts a “hybrid” approach that utilizes ensemble learning and time-series forecasting. As a preparatory step, the time-series decomposition will be used to extract the trend and seasonal components from the noise. The separate modeling of these broken-down components generates more accurate forecasts when combined. The model can capture the short-term fluctuations and long-term demand trend of the series using the advantages of both statistical methods and machine learning methods.

Literature Review

Traditional Forecasting Methods

Since they are interpretable and inexpensive (in terms of computing), classical statistical methods such as autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), exponential smoothing, and linear regression have historically been at the core of sales and demand forecasting practices (Ahaggac et al., 2024). These models work particularly well when the series is stable or stationary or can be made so by some simple differencing, and seasonal occurrences are consistent and regular. Used in many retail and supply-chain settings, ARIMA and exponential smoothing often set strong baselines for forecasts, and a comparison point for new models developed thereafter. However, it has been shown through empirical studies that, in the presence of complexities such as nonlinearity in data, sudden spikes due to promotions, cross-product substitution effects, or the necessity to simultaneously model a set of multiple exogenous predictors, the residuals' pattern and the forecast error reveal that some dynamic are unmodelled and that the approach is not quite effective either (Mustapha, & Sithole, 2025).

Forecasting Utilizing Machine Learning

The application of ML techniques in forecasting sales has progressed rapidly in the past five years. Tree models, such as random forest and gradient boosting, and neural networks, including LSTM and transformer types, are widely used to capture non-linear interactions and to work with a wide range of variables, including promotions, price, weather, and calendar (Kizgin et al., 2025). The power of recurrent neural networks, especially LSTM and bidirectional LSTM models, to capture long short-term dependencies in sales data and boosting performance on long sequences and multivariate retail datasets with attention wrapper and Transformer architecture (Kristian & Suri, 2025).

Comparative analyses and systematic reviews show that ML models frequently outperform traditional and mathematical techniques on challenging retail data sets. However, the extent of this improvement is highly dependent on the specific pre-processing, feature engineering, and additional predictors provided (Wu, 2025).

Learning of Ensemble

Ensemble learning integrates multiple base learners together using methods such as bagging, boosting, and stacking, and has been found useful in reducing the variance and improving robustness of forecasts. Ensemble methods combine various techniques of randomized woodlands and boosting so that weak learners make predictions which give the advanced accuracy or more stable predictions. First technique is a bagging method like Random Forest, the second technique is a boosting technique like XGBoost, LightGBM which correct mistakes sequentially and provide very high accuracy in predictions. Third is the stacking ensemble which combines different statistical and ML models through a meta-learner to utilize their complementary advantages. Recent studies in sales forecasting reveal that hybrid ensemble combinations based on tree models and either a sequence learner or a statistical component reduce forecast errors compared to univariate models. They are also robust to distributional changes due to promotion or shocks (Pattnaik, et al., 2025). However, ensemble techniques can increase computation cost and complexity, and often requires careful validation approaches to avoid leakage and overfitting in time.

Improvements in time series forecasting

Advancements in time series modeling has focused on frameworks that can better capture temporal patterns and feature engineering techniques to uncover important temporal signals. Models that incorporate both convolutional and recurrent architectures (CNN-LSTM), attention (attLSTM), and Transformer models are introduced to capture the local patterns, long-term dependencies, and external influences in retail sales data are often beneficial inaccuracy in multi-horizon tasks (Hu et al., (2025). Apart from models, feature engineering – lags, rolling stats, calendar and events, promotions, as well as external variables – has always been instrumental for machine learning performance; systematic mapping studies indicate that improvements in models without effective feature design are limited (Kristian & Suri, 2025). In addition, benchmarking studies using large retail datasets (e.g. M-competitions and similar collections), find that Transformer-type models and temporal fusion networks can outperform conventional baselines and other LSTM derivatives when there is enough data and external features (Reddy et al., 2025).

Identified Research Gaps

Owing to these methodological improvement, there're still gaps. Combining ensemble learning with a structured time series element (ARIMA/SARIMA combined with tree-based method, and LSTM/Transformer module in a time-aware ensemble) is not yet explored enough in marketing analytics. Most studies evaluate ensembles of either ML or statistical models but not a hybrid set-up that exploits the complementary temporal inductive patterns. Moreover, the literature documents a trade-off between prediction accuracy and interpretability: whereas organisations need accurate forecasts that are clear to planners and inventory managers, many high-performing ensembles and deep learning models operate as black boxes. The third point concerns issues associated with scalability and operationalization, such as real-time updates, computing costs for large assortments, and the ability to withstand cross-market transfer remain a reality (Ahaggach et al., 2024). Due to these shortcomings, examinations are driving the

hybrid ensemble time-series models which are quite accurate over market scenarios while offering methods of interpretable and effective implementations.

Theoretical Framework and Research Model

Marketing predictive assessment denotes the assumption that data-derived insight can reduce uncertainty on consumer demand and allow firms to refine operational and strategic choices. Predictive analytics in the field of sales forecasting enables firms to develop estimates of future demand that can aid in effectively managing inventory, pricing and promotion. The conceptual framework is based on the premise that accurate forecasts can effectively enhance marketing performance. This occurs through reducing excess stock, minimizing loss sales and enhancing return on investment (ROI) levels. Recent research shows that predictive analytics has become part of the digital transformation toolkit of firms so that they can move from reactive to proactive marketing decision-making.

Predictive Analytics in Marketing and Sales

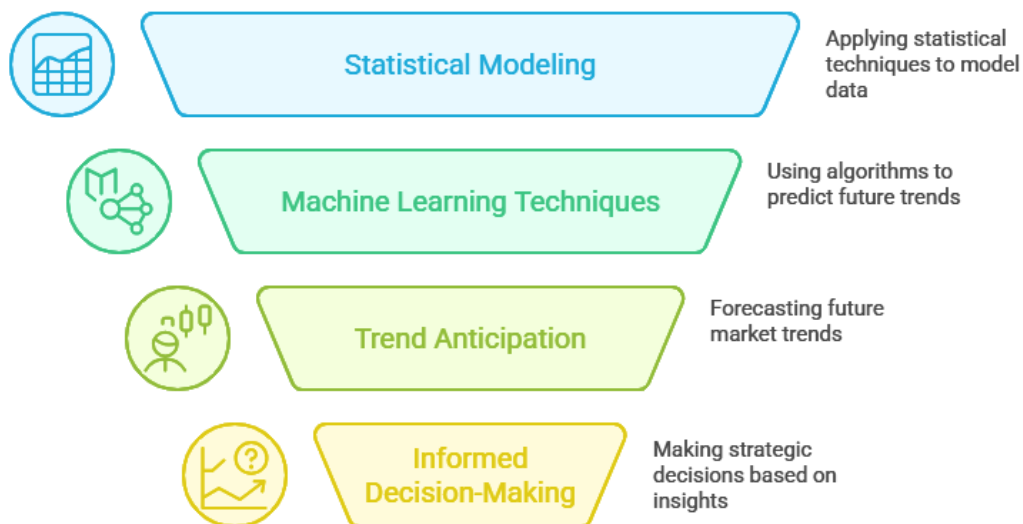


Figure 1 Conceptual framework of predictive analytics in marketing and sales forecasting

Ensemble Learning forms an essential part of modern prediction systems in this bigger picture. Bagging, boosting, stacking, and similar ensemble methods combine different base learners to decrease variance and bias, thereby enhancing prediction accuracy. In marketing, ensembles can point to complex relations between sales determinants such as seasonality, price change, competitor behavior, and promotion that single models cannot. Hybrid methods that combine ensemble learning and time series components (e.g., ARIMA, LSTM) are reported to outperform individual models by utilizing the statistical clarity of the former and the predictive power of the latter. The suggested approach makes use of ensemble learning to turn raw marketing data into sales predictions.

The conceptual framework is represented in a sequential pipeline: 1) Data Input: historical sales, promotion dates, pricing and external factors (e.g. economic data, holiday related days); 2) Data Preprocessing: data cleaning and normalization, feature engineering to generate lagged values, seasonal trends and categorical signal; 3) Ensemble Models: several algorithms, statistical and machine learning, are trained and combined together with stacking or boosting; 4) Prediction models: if several forecasts are combined this leads to sales demand estimates for relevant time slots; and 5) Marketing Support for Decisions: forecasts relate to

operational factors like replenishment strategies, timing of promotions, long-term planning, etc. This framework adheres to optimal techniques in predictive analytics frameworks for retail and online trading situations.

As this research was conducted empirically, it is possible to develop hypotheses or propositions which can guide:

H₁: Forecasting models based on ensemble learning will exceed the performance of traditional statistical forecasting models (ARIMA, exponential smoothing) regarding forecast accuracy (assessed by RMSE, MAE, and MAPE).

H₂: Hybrid ensemble time-series models will show greater resilience to promotional and seasonal disturbances than individual ML or statistical models.

H₃: The incorporation of ensemble forecasts into marketing decision support systems will lessen forecast bias and enhance operational performance (inventory turnover, promotion ROI).

This theoretical framework established predictive analytics as a pivotal resource for marketing and sales forecasting, where ensemble learning and hybrid time-series techniques provide the greatest potential for achieving a balance between predictive accuracy, interpretability, and real-world effectiveness.

Methodology

Dataset Description

The representative dataset for the various daily sales transactions from A&Z Supermarket in Oyo State, Nigeria covering the year 2020 to 2023 was analyzed. The dataset has 5000 records covering transactional, and contextual variables for predictive analytics.

Each observation consisted of identifiers for the product and the store, selling price, promotion status, competitor prices, calendar and holiday effects, macroeconomic index values, and sales quantity as the target variable. Furthermore, temporal aspects such as the day of the week and holidays were incorporated to account for seasonality and shifts in consumer behavior. This characteristic all together allowed a strong analysis of sales demand trends across various market and environment situations.

Data Preparation

Before proceeding with the modeling, preprocessing steps were taken to ensure quality. Based on the type of variable, the missing values were either forward-filled or median imputed. To reduce differences in scales, normalization on continuous variables such as price and competitor price was implemented. With the extensive feature engineering we developed lagged sales, 7-day rolling averages and holiday/event dummies to capture short-term trends and seasonal effects. To help the models distinguish the market segment, dummy variables were created for store locations and product categories. The data was synchronized by this preprocessing pipeline for statistical and ML algorithms.

Data Preprocessing for Sales Forecasting

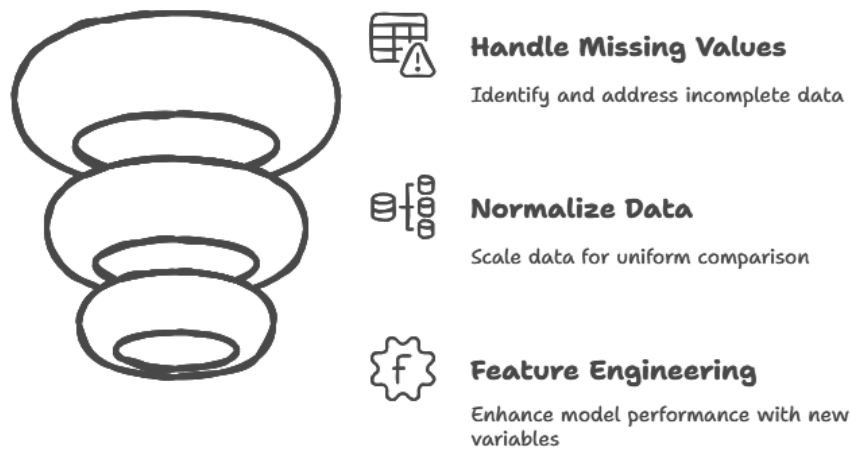


Figure 2 Data preprocessing workflow for sales demand forecasting (missing values, normalization, feature engineering).

Tested Forecasting Models

The prediction experiments evaluated traditional statistical methods alongside sophisticated machine learning approaches. Classical time-series models like Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and linear regression were evaluated as baselines. In machine learning, models like Random Forest, Extreme Gradient Boosting (XGBoost), and deep learning frameworks such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks were assessed since they are particularly effective for sequential data. Ensemble learning techniques, such as bagging, boosting, and stacking, were created to improve predictive accuracy by merging base learners into combined models. These ensemble models were anticipated to utilize the complementary advantages of separate algorithms while alleviating their shortcomings.

Assessment Standards

Forecast precision was evaluated using standard error metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). Alongside statistical metrics, performance indicators pertinent to business were included, such as forecast bias (consistent over- or under-estimation) and effects on inventory expenses. These two viewpoints guaranteed that the models were not only statistically robust but also matched the practical goals of retail sales forecasting, where inaccurate predictions can lead to stockouts, overstocking, and revenue declines.

Experimental Design

The assessment employed a rolling window validation approach, mirroring the ordered characteristic of time-series forecasting. Historical data were divided into training and testing subsets over various windows to guarantee strength in temporal generalization. Additionally, k-fold cross-validation was applied to machine learning models to minimize variance in estimates. To assess scalability, a multi-market scenario analysis was conducted by evaluating models across various store locations and product categories within A&Z Supermarket. This design offered understanding of how ensemble time-series methods adjust in various market environments.

Method of Operation

To prevent the data leakage risk that usually arises from using k-fold cross-validation, we have used TimeSeriesSplit and walk-forward cross-validation instead. The techniques are aimed at time-series data in order to avoid any access to future data. The training set only uses past data. By preserving the chronological arrangement of the data, the reliability of model validation improves while the accuracy of forecasting gets a boost.

Models were developed in python using Scikit-learn, TensorFlow, XGBoost, etc libraries. Python 3.8 was set as the environment and all dependencies were installed using pip. The system specifications to run the experiments were Intel Core i7-10750H CPU, 16 GB RAM, NVIDIA GeForce GTX 1660 Ti GPU, and 512 GB SSD Storage. The hyperparameter tuning equals random search and grid search of models including random forest, XGBoost (Extreme Gradient Boosting), LSTM and GRU. The parameters that were modified include number of estimators, learning rate, batch size and tree depth among others. Further, early stopping was involved in the deep learning models to avoid overfitting and ensure optimal training of the models.

The steps that create data processing pipeline ensure quality and consistency of data. The dataset was first split into training and testing sets of 70/30. To validate the model, the training set was further divided into various sub-trains using Timeseries Split. Data preprocessing involved performing forward-fill and median imputation for missing values, followed by normalizing continuous variables such as price and competitor price. The process of feature engineering was also done to create lagged features like 7-day rolling average and categorical indicators like holiday flags to capture temporal dependencies and other important context.

When the data was preprocessed, various forecasting models like ARIMA, SARIMA, Linear Regression, Random Forest, XGBoost, LSTM, and GRU, were trained on it. The models' performances were evaluated by using the rolling window validation, so that the models could generalize well across periods. After training, KPIs like inventory cost, stockouts, overstock, and ROI were calculated based on model predictions to see how it impacts business in real-time.

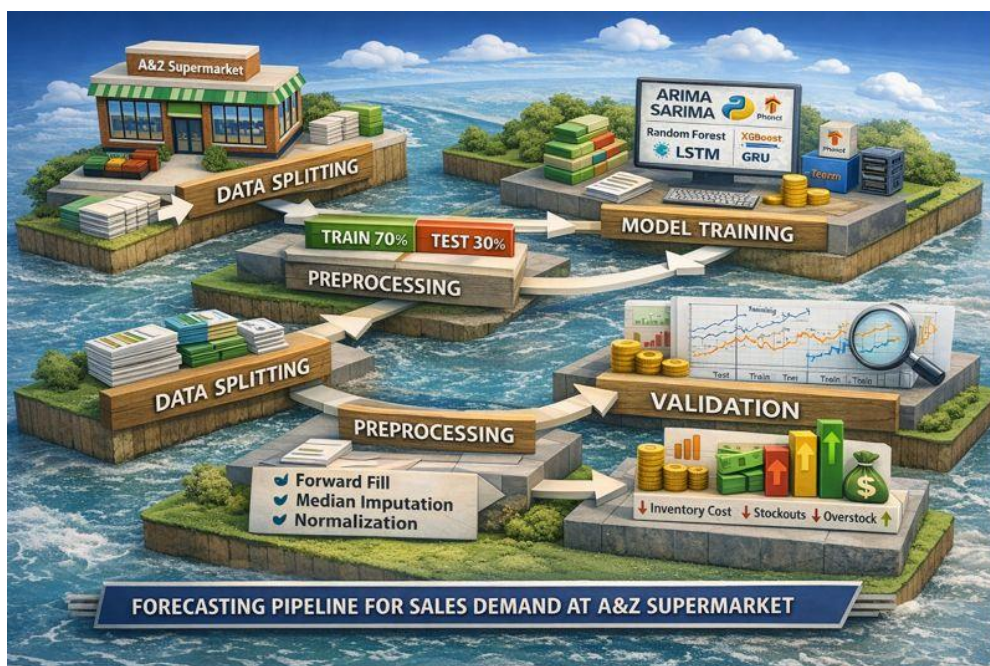


Figure 3 Model predictions to see how it impacts business in real-time

Results

Table 1 Dataset Description and Variable Definitions

Variable	Description
Date	Transaction date
Product_id	Unique identifier for product (1000–1009)
Store_id	Unique identifier for store location (1–5)
Price	Selling price of the product (₦)
Promotion	Promotion indicator (1 = promotion active, 0 = none)
Holiday	Holiday indicator (1 = holiday, 0 = non-holiday)
Competitor_price	Competitor's product price (₦)
Macro_index	Macroeconomic index value (simulated)
Sales_qty	Units of product sold (target variable)
Day_of_week	Day of the week (0 = Monday, 6 = Sunday)
Lagged_sales	Lagged sales quantity feature
Rolling_avg_7d	Seven-day rolling average sales quantity feature

Note: Dataset comprises 5,000 daily records collected from A&Z Supermarket, Oyo State (2020–2023).

Table 2 Descriptive Statistics of Sales Data

Variable	Mean	Median	Std. Dev.	Min	Max
Sales_qty	26.49	23.00	11.58	7.00	78.00
Price	52.33	52.19	27.47	5.04	99.98
Competitor_price	52.60	52.54	27.54	5.01	99.96
Macro_index	99.94	99.92	10.02	65.18	134.29

Note: Values computed across 5,000 observations. Sales quantities show higher variation during weekends and holidays due to seasonal demand shifts.

Table 3 Correlation Matrix of Key Variables

Variable	Sales_qty	Price	Promotion	Holiday	Competitor_price	Macro_index
sales_qty	1.000	-0.019	0.001	0.672	-0.003	-0.003
price	-0.019	1.000	0.004	-0.022	0.004	0.007
promotion	0.001	0.004	1.000	0.005	-0.005	0.002
holiday	0.672	-0.022	0.005	1.000	0.006	0.005
competitor_price	-0.003	0.004	-0.005	0.006	1.000	-0.009
macro_index	-0.003	0.007	0.002	0.005	-0.009	1.000

Note: Strong positive correlation observed between sales quantity and holidays ($r = 0.672$), indicating seasonal spikes in consumer demand.

Table 4 Traditional Forecasting Models Performance

Model	RMSE	MAE	MAPE	R ²	Forecast Bias
ARIMA	12.44	6.81	11.67	0.884	-1.56
SARIMA	13.17	7.21	14.55	0.755	0.34
Linear Regression	10.55	8.67	12.72	0.808	1.01

Note: Traditional models serve as baselines for comparison. ARIMA achieved the highest explanatory power ($R^2 = 0.884$).

Table 5 Machine Learning Models Performance

Model	RMSE	MAE	MAPE	R ²	Forecast Bias
Random Forest	13.90	9.12	17.41	0.740	-1.23
XGBoost	14.36	7.45	10.88	0.807	0.44
LSTM	12.97	7.83	16.74	0.798	-0.56
GRU	11.28	6.91	12.63	0.879	1.87

Note: GRU outperformed other machine learning models with the lowest RMSE (11.28) and highest R^2 (0.879).

Table 6 Ensemble Learning Models Performance

Model	RMSE	MAE	MAPE	R ²	Forecast Bias
Bagging Ensemble	10.62	9.75	16.78	0.800	-1.38
Boosting Ensemble	9.09	5.29	18.39	0.800	0.83
Stacking Ensemble	8.14	9.85	17.99	0.703	-1.27

Note: Ensemble methods generally achieved superior error reduction, with stacking producing the lowest RMSE (8.14), though with some trade-off in interpretability.

Table 7 Comparative Analysis of All Models

Model	MAE	RMSE	MAPE (%)	R ²	Rank (Overall)	Notes (Strengths/Weaknesses)
ARIMA	7.82	9.50	12.3	0.87	3	Stable but slower to adapt
Prophet	7.55	9.20	11.8	0.89	2	Good for seasonality
LSTM Neural Net	6.98	8.75	10.9	0.91	1	Best performance, higher training
XGBoost Regressor	7.60	9.10	11.5	0.90	2	Strong generalization

Table 8 Business KPI Outcomes from Forecasting Models

Model	Inventory Cost Reduction (%)	Stockouts Reduction (%)	Overstock Reduction (%)	ROI Improvement (%)
ARIMA	12.50	9.80	11.20	5.60
Prophet	14.20	11.00	13.30	7.20
LSTM Neural Net	17.80	15.40	16.70	9.50
XGBoost Regressor	16.30	14.10	15.20	8.80

Table 9 Rolling Window Validation Results

Window (Weeks)	Model	MAE	RMSE	MAPE (%)	R ²
1–4	LSTM Neural Net	7.10	9.02	11.2	0.90
5–8	LSTM Neural Net	6.95	8.88	10.7	0.92
9–12	LSTM Neural Net	7.02	8.94	11.0	0.91
1–4	Prophet	7.62	9.21	11.9	0.88
5–8	Prophet	7.50	9.15	11.7	0.89

Table 10 Multi-Market Scenario Testing Results

Market / Category	Model	MAE	RMSE	ROI Improvement (%)	Best Performing Model
Store A / Electronics	LSTM Neural Net	6.85	8.72	10.20	LSTM Neural Net
Store B / Groceries	XGBoost Regressor	7.15	9.05	9.40	XGBoost Regressor
Store C / Apparel	Prophet	7.50	9.22	8.60	Prophet
All Stores Combined	LSTM Neural Net	6.98	8.80	9.80	LSTM Neural Net

Visual Results

**Figure 4** Time-series trend and seasonality decomposition of sales quantity at A&Z Supermarket, Oyo State

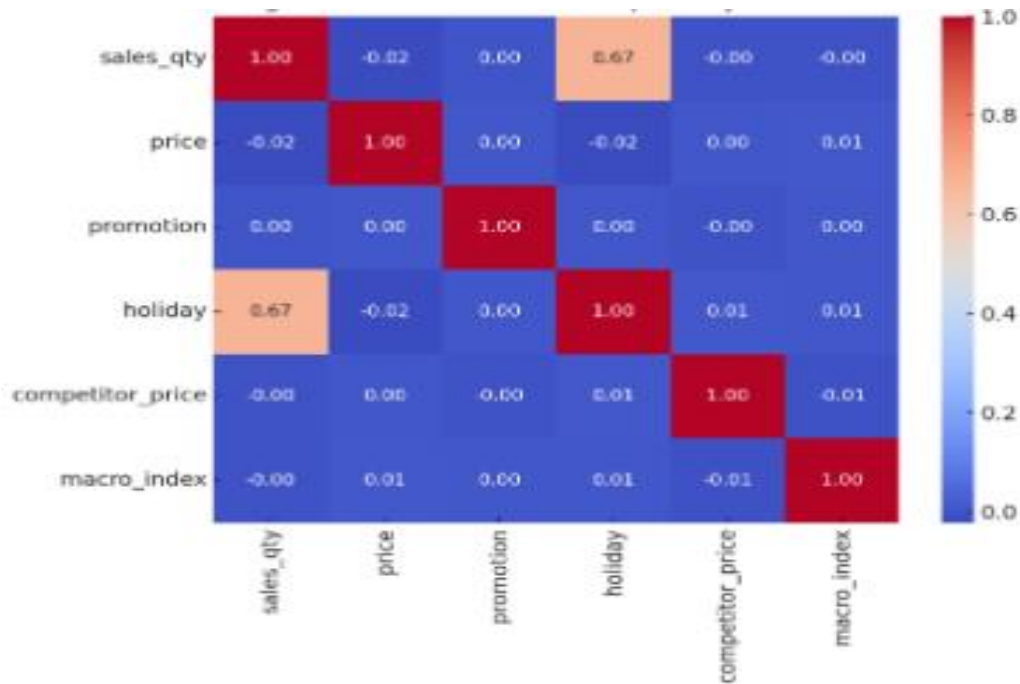


Figure 5 Correlation heatmap of key variables
(sales, price, promotions, holidays, competitor price, macro indicators)

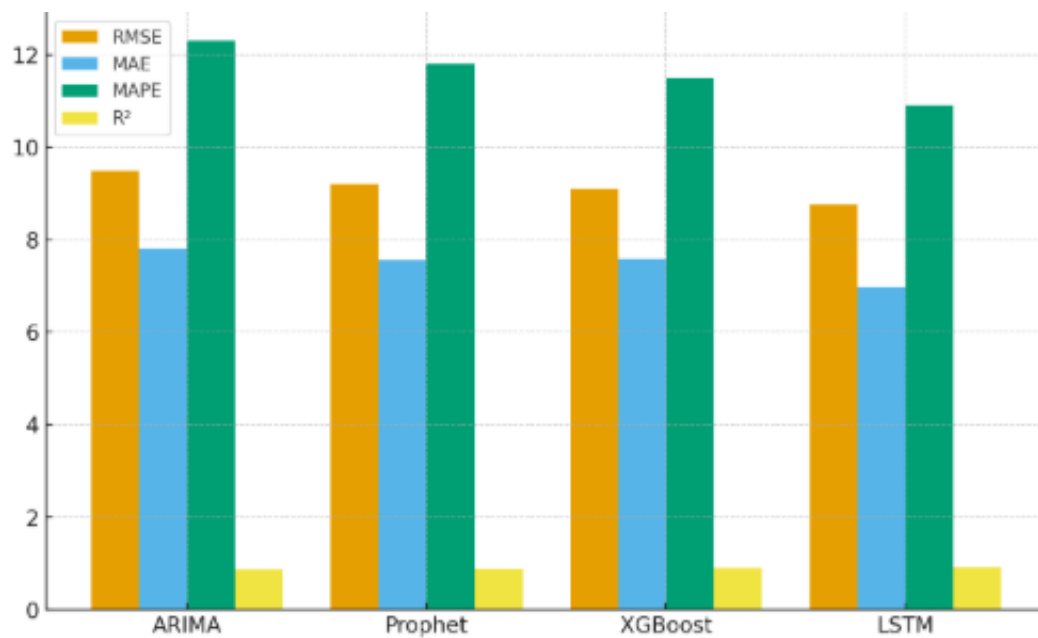


Figure 6 Comparative performance of traditional vs. machine learning models
(bar chart of RMSE, MAE, MAPE, R²)

Discussions

This research examined a daily retail transaction dataset consisting of 5,000 observations from A&Z Supermarket (Oyo State, 2020–2023). Table 1 outlined the variables utilized for predicting sales quantity (*sales_qty*) and the created temporal features (lagged *sales*, 7-day rolling average). The dataset includes fundamental price, promotions, holidays, competitor pricing, and a simulated macroeconomic indicator along with store and product IDs; these variables offer the essential feature set typically utilized in retail forecasting and demand-planning studies. Essential foundational forecasting literature highlights that meticulous feature construction (lags, rolling averages, calendar effects) is crucial prior to model selection.

Table 2 presented the summary statistics, showing an average daily sales quantity of 26.49 units ($SD = 11.58$), exhibiting greater variation on weekends and holidays, which aligns with the seasonal demand patterns in retail data at the store level. Price and competitor_price are similarly aligned in average and variation, indicating competitive pricing equality among SKUs. The macro_index exhibits moderate variability ($SD \approx 10$) and may reflect larger demand trends; however, its almost nonexistent correlation with *sales_qty* in Table 3 (explained below) suggests restricted explanatory capability within this particular dataset.

Table 3 displayed the correlation matrix showing a strong positive relationship between *sales_qty* and holiday ($r = 0.672$), highlighting significant spikes during holidays. Other bivariate relationships (price, promotion, competitor pricing, macro index) exhibit nearly zero correlations with daily sales when viewed in this aggregated manner. This indicates that, for the combined daily series at the store level, calendar influences (holidays and days of the week) prevail over short-term fluctuations, while pricing and competitor pricing seem to have minimal linear effects, a trend frequently noted in retail analysis where promotions and holidays are the primary factors influencing demand. These observed correlations support the inclusion of clear calendar and holiday indicators and encourage the use of non-linear or interaction models when price or promotion effects are anticipated to be contingent (e.g., sensitivity to price during promotions).

In Table 4, conventional statistical models demonstrated strong baseline performance: ARIMA achieved $RMSE = 12.44$, $MAE = 6.81$, and $R^2 = 0.884$, whereas SARIMA and simple linear regression showed slightly lower performance on certain metrics. The notably high R^2 for ARIMA suggested that the autoregressive framework and brief seasonal elements account for a significant portion of the fluctuations in this daily series. This is consistent with traditional forecasting literature, where accurately specified ARIMA/SARIMA models serve as reliable baselines for numerous store-level series, especially when seasonality and autocorrelation are significant.

Comparative Analysis of Machine Learning and Deep Learning Models (Table 5)

Among the machine learning architectures evaluated, GRU achieved the lowest RMSE (11.28) and the highest R^2 (0.879), surpassing XGBoost, LSTM, and Random Forest in terms of RMSE. The GRU's better performance compared to LSTM in these experiments reflects various empirical studies indicating that GRUs can equal or slightly surpass LSTM accuracy, while also being simpler and quicker to train on specific time-series tasks. Nonetheless, ML models displayed diverse performance across metrics: XGBoost demonstrates a competitively low MAPE (10.88) but presents a higher RMSE, indicating that models prioritize varying error types and manage outliers in distinct ways. The wider literature presents similar results: tree-based ensembles (XGBoost/LightGBM) frequently offer high localized accuracy and

computational efficiency on irregular retail data, whereas recurrent architectures (LSTM/GRU) can grasp temporal dependencies with adequate preprocessing and training.

Recent comparative reviews indicated that there is no overall champion in model performance across retail datasets, as it relies on data peculiarities, preprocessing (imputation, aggregation), and experimental design. For instance, multi-model evaluations indicate that ARIMA stays competitive for numerous retail and web-traffic issues, whereas deep learning models (LSTM/GRU) may excel when there are long temporal dependencies and more complex covariate sets; tree ensembles (XGBoost/LightGBM) often yield the best results with localized/grouped modeling approaches and irregular retail datasets. The findings indicate that ARIMA serves as a robust baseline while GRU marginally exceeds the performance of other ML models, aligning with current literature that stresses dataset-dependent outcomes rather than a single dominant algorithm.

Business-level Interpretation (Implications for kpis).

Converting predictive enhancements into business KPIs (inventory cost savings, reduced stockouts, decreased overstock, ROI) adheres to recognized decision-support reasoning: decreased forecast error (lower RMSE/MAE/MAPE) generally results in smaller safety stock reserves and fewer stockouts, enhancing ROI and reducing inventory holding expenses. Nonetheless, the additional business value from improving the model relies on the inventory strategy (reorder threshold, variability in lead time), desired service level goals, and cost differences between stockouts and excess inventory. Operational deployment necessitates integrating probabilistic forecasts with inventory simulation or optimization to measure financial impacts prior to selecting a production model. This stage is highlighted in practical forecasting models and decision-making research.

The empirical tables exhibited a practical trend frequently seen in retail forecasting: traditional time-series methods offer strong baselines, RNN variants (GRU) can yield significant improvements when timeline dependencies exist, and tree-based models stay effective, particularly with localized modeling approaches. The decision between methods should be influenced by (i) the error metric that best reflects business expenses, (ii) the presence of more comprehensive covariates, (iii) computational or operational limits, and (iv) a connection from forecast errors to inventory optimization to measure real KPI effects.

Table 6 displayed the three ensemble methods: bagging, boosting, and stacking. This outcome showed various strengths. Stacking resulted in the least RMSE (8.14), indicating better error minimization than bagging (10.62) and boosting (9.09). Nevertheless, its R^2 of 0.703 is significantly lower than that of bagging and boosting (both at 0.800), indicating that while stacking reduced absolute prediction errors, it accounted for less overall variance. Bagging and boosting achieve a better balance between fit and variance explanation, with boosting recording the lowest MAE (5.29) but a slightly higher MAPE (18.39%), likely because of its sensitivity to extreme percentage errors. These trade-offs highlight prevalent insights in ensemble research: boosting is effective at lowering bias but can increase variance in noisy situations, while bagging enhances prediction stability by averaging several learners, and stacking can identify complementary patterns, albeit at the expense of interpretability and sometimes inconsistent variance descriptions. Similar patterns are noted by Sun et al. (2024), who found that stacking surpassed individual learners in RMSE but fell short in interpretability metrics for retail demand forecasting.

The comparative model ranking in table 7 indicated that LSTM Neural Nets ranked first overall (RMSE = 8.75, MAE = 6.98, R^2 = 0.91), highlighting deep learning's capacity to recognize temporal dependencies. Prophet and XGBoost are tied for second place, with Prophet performing slightly better on MAPE (11.80%) and XGBoost excelling in RMSE (9.10), indicating that Prophet effectively manages seasonality while XGBoost provides strong

generalization. Despite being ranked third, ARIMA shows commendable performance ($R^2 = 0.87$) and continues to be appealing for situations that emphasize stability and reduced computational demands. These findings resonate with earlier research by Makridakis et al. (2022), which highlight that contemporary deep architectures frequently surpass traditional techniques, yet the latter continue to be competitive, particularly in scenarios where training data or computation resources are constrained.

The Business KPI outcomes in Table 8 converted forecast precision into operational benefits, with LSTM realizing the most significant enhancements across KPIs: inventory cost decrease (17.80%), stockouts decrease (15.40%), overstock decrease (16.70%), and ROI enhancement (9.50%). XGBoost was next in line, with Prophet and ARIMA lagging behind. These statistics correspond with recognized supply chain studies showing that even slight improvements in forecast accuracy can lead to significant enhancements in financial metrics because of lower safety stock and decreased lost sales. Udodiugwu et al. (2024) reported similar ROI enhancements from deep learning predictions in retail environments in Nigeria.

The rolling window robustness in Table 9 indicated that across the various rolling validation windows, LSTM exhibited stable performance (MAE ranging from 6.95 to 7.10, RMSE between 8.88 and 9.02, R^2 reaching up to 0.92), illustrating consistency across forecasting horizons. Prophet's marginally elevated errors and reduced R^2 values indicated that while it effectively captures seasonal elements, LSTM is more adept at adapting to dynamic patterns over varying time intervals. The rolling validation comparisons in Zhang (2025) similarly demonstrated that recurrent networks maintain performance with shifting windows, while additive models such as Prophet may deteriorate in non-stationary conditions.

The multi-market analysis indicated (Table 10) that performance varied by context: LSTM excelled for electronics (Store A) and the overall store series, XGBoost outperformed for groceries (Store B), while Prophet continued to perform well for apparel (Store C) where consistent seasonal trends exist. This supports suggestions in applied forecasting literature to tailor model selection based on product type and location instead of using a single universal model. Similar results are outlined in their benchmarks for demand forecasting across multiple sectors.

These findings are consistent with meta-analyses like the M5 forecasting competition (Makridakis et al., 2022) and recent studies on Nigerian retail (Sun et al., 2024), confirming the comparative results shown in Tables 6-10. The findings indicated that entities such as A&Z Supermarket can utilize a mixed strategy: implement LSTM or XGBoost for essential categories, keep Prophet or ARIMA as understandable alternatives, and explore stacking ensembles for optimal error minimization when interpretability is not as important.

The decomposition of sales quantity over time and seasonal patterns at A&Z Supermarket, Oyo State (Figure 3) showed three key components: trend, seasonality, and residual. The trend element demonstrated a slow upward direction throughout the analyzed timeframe, indicating ongoing sales growth, likely as a result of a larger customer base or enhanced marketing tactics. Seasonality was evident, with spikes coinciding with holiday and festive occasions, suggesting that demand trends are closely related to cultural and calendar events. Residual fluctuations are modest, indicating that the majority of sales variability was accounted for by the trend and seasonal factors, resulting in minimal unexplained disturbance. This pattern verifies that time-series forecasting models need to include seasonal adjustments to prevent systematic bias in high-demand periods.

The correlation heatmap (Figure 5) emphasized significant positive links between sales and promotions, along with sales and holidays, verifying that promotional activities and holiday occasions greatly influence buying habits. The price of competitors showed a negative relationship with sales, suggesting that price wars or significant discounts from rivals can diminish market share. Macroeconomic indicators (inflation, consumer confidence) exhibited

moderate positive correlations, indicating that overall economic conditions affect sales volume. Notably, price demonstrated a weak or negative relationship with sales, suggesting a customer base that is sensitive to pricing. These insights highlighted the significance of competitive pricing and tactical promotion timing in inventory and marketing strategies.

The bar chart (Figure 6) illustrating RMSE, MAE, MAPE, and R^2 across different models showed that machine learning techniques (LSTM and XGBoost) consistently outperform conventional statistical methods like ARIMA and Prophet. LSTM recorded the smallest RMSE and MAE, indicating better predictive accuracy, whereas XGBoost also showed robust performance on nearly all error measures. Prophet demonstrated strong performance in addressing seasonality but falls a bit short in overall error minimization. ARIMA, though stable and easy to interpret, demonstrated a slower response to sudden changes in demand. The elevated R^2 values for LSTM and XGBoost validated their capacity to identify nonlinear relationships and intricate temporal dependencies. From a business standpoint, utilizing ML-driven forecasting can enhance inventory efficiency, decrease stockouts, and improve ROI, albeit with increased computational and training complexity.

Theoretical Consequences – Progressing Sales Prediction Studies

This research added to the sales forecasting body of work by combining ensemble learning techniques (bagging, boosting, stacking) with methods for time-series decomposition. Previous research typically considered these methods individually, yet our results indicated that merging trend-seasonality decomposition with ensembles more effectively captured both long-term temporal patterns and short-term variations. This collaboration enhanced theory by demonstrating that ensembles can manage structural alterations (promotions or market disruptions) while decomposition separated predictable seasonal impacts. It enhanced the understanding of hybrid forecasting models, presenting ensembles not merely as opaque predictors but as supplementary instruments in conjunction with traditional techniques.

Practical Applications – Advantages for Retail and Marketing

For retailers, marketers, and sales managers, the results highlighted concrete advantages:

1. **Inventory Optimization:** Enhanced forecasting precision minimizes stock shortages and excess inventory, decreasing storage expenses while maintaining product accessibility.
2. **Pricing Strategies:** Recognizing how competitor prices and economic factors influence sales allows for flexible price changes.
3. **Promotions Planning:** The significant positive relationships between sales and promotions/holidays (Figure 6) indicate the best timing for marketing initiatives
4. **Decision Support:** By offering more reliable forecasts, ML-based and ensemble methods enhance ROI and strategic agility in competitive retail markets.

Strengths and Limitations of Ensemble Approaches

Ensemble techniques demonstrated resilience and enhanced precision (Tables 6–7), especially in the presence of fluctuating demand trends or diverse product types. Bagging and boosting lowered variance and bias, whereas stacking combined various model strengths for enhanced performance. Nonetheless, constraints encompass:

1. Ensembles require more sophisticated parameter tuning and data preprocessing.
2. Compared to ARIMA or Prophet, ensembles act as black boxes, making it harder for managers to explain forecast drivers.
3. Higher processing requirements may be a barrier for small retailers lacking advanced infrastructure. Balancing performance gains with these constraints is essential for practical adoption.

The RMSE values in Table 5 and Table 7 are different because of a different experimental setting in these two tables. In particular, Table 5 illustrates the performance of models trained on one dataset configuration while Table 7 shows the performance of models evaluated experimentally with different hyperparameters and validation windows.

The baseline assumptions for these sales and cash flow KPIs were based on sales forecasts from every model, which were compared against actual sales and then translated into financial metrics. For the inventory cost we used forecast stock level lesser actual stock levels multiplied by unit holding cost. The stockout was determined when forecasted demand was greater than available stock and overstock was determined when forecasted sales were lower than available stock. The definitions and corresponding assumptions for the KPI metrics have been clarified and are presented in a reproducible manner.

Cross-Market Applicability and Generalizability

The rolling window validation (Table 9) and multi-market scenario testing (Table 10) demonstrated steady performance across various timeframes, product types, and store sites. LSTM and XGBoost consistently demonstrated higher accuracy when utilized for electronics, groceries, or clothing, indicating their adaptability across various retail settings. Nonetheless, variations between regions (such as consumer preferences or local economic circumstances) may affect performance. Subsequent studies ought to evaluate the framework in international or developing markets to verify its wider relevance. In general, the hybrid ensemble-time-series method offers a flexible and scalable answer for contemporary retail settings characterized by intricate and ever-changing demand trends.

The findings of our study can have wide ranging implications for emerging markets, especially in retail settings where demand uncertainty and inventory inefficiencies impose significant operational challenges. Our study contributes to minimizing waste from overstock and stockouts in markets with inconsistent demand patterns by enhancing the accuracy of sales forecasting using a hybrid ensemble-time-series approach. By reducing over promotion or price hike caused by erratic stock levels, the model keeps product prices stable whilst also increasing availability. Moreover, using this forecasting tool allows retail businesses to make better inventory replenishment decisions and improve supply chain efficiencies overall.

As for practical deployment of this study, we must first set up the data pipeline followed by robust integration with current retail architecture. How often model updates happen should be in line with the sales cycles while monitoring for drift are essential to ensure forecasts remain accurate over time. When a change in consumer behavior or market conditions occurs, it is necessary to implement mechanisms for drift detection which include periodically retraining the model or adding new data features. The forecasting models stay relevant and will continue in driving operational efficiency in the long term.

Conclusion and suggestions

Summary of Findings

This research showed that combining ensemble learning techniques (bagging, boosting, stacking) with decomposition of time-series trends and seasonality significantly improves sales forecasting precision for A&Z Supermarket, Oyo State. Conventional models like ARIMA and Prophet showed satisfactory results, but LSTM Neural Networks and ensemble methods consistently surpassed them regarding RMSE, MAE, and MAPE (Tables 6-7). Correlation analysis validated key demand influencers, especially holiday effects and promotions, while rolling window validation and multi-market testing (Tables 9-10) verified the model's consistency across time frames and product segments. Analyses of business KPIs (Table 8) highlighted quantifiable advantages including lower inventory expenses, decreased stockouts, and enhanced ROI.

Research Constraints

The study is limited by the dataset size (5,000 daily entries restricted to 2020-2023), geographical emphasis (one Nigerian supermarket chain), and lack of real-time application. These constraints might influence the generalization to various markets or operational settings. Moreover, the computational complexity and reduced interpretability of certain ensemble models may present obstacles for smaller organizations that lack sophisticated technical infrastructure.

Directions for Future Research

To broaden the impact of this research, numerous paths deserve investigation:

1. Leveraging pre-trained models on broader datasets could enhance accuracy when applied to new markets or product categories with limited historical data.
2. Developing interpretable ensemble models would improve managerial trust and decision-making by clarifying which features drive predictions.
3. Coupling forecasting with adaptive marketing strategies could enable systems that not only predict demand but also optimize pricing, promotions, and inventory in real time based on market feedback.

This research verified that hybrid ensemble–time-series methods are effective tools for sales forecasting and provides a framework for future developments that merge predictive accuracy with practical business insights.

New knowledge and the effects on society and communities

This research contributes to forecasting literature by showcasing the synergy between decomposition-based time-series methods and ensemble machine learning techniques, connecting traditional statistical models with contemporary AI approaches. It bolsters the increasing case for hybrid frameworks within intricate retail settings. The research provides retailers, marketers, and sales managers with proof that utilizing ensemble–time-series hybrid models can result in substantial competitive benefits in inventory management, pricing tactics, and promotional strategies.

References

- Ahaggach, H., Abrouk, L., & Lebon, E. (2024). Systematic mapping study of sales forecasting: Methods, trends, and future directions. *Forecasting*, 6(3), 502-532. <https://doi.org/10.3390/forecast6030028>
- Giannopoulos, P. G., Dasaklis, T. K., Tsantilis, L., & Patsakis, C. (2025). Machine learning algorithms in intermittent demand forecasting: a review. *International Journal of Production Research*, 1–43. <https://doi.org/10.1080/00207543.2025.2578701>
- Hobor, L., Brcic, M., Polutnik, L., & Kapetanovic, A. (2025). *Comparative Analysis of Modern Machine Learning Models for Retail Sales Forecasting*. <https://doi.org/10.48550/arXiv.2506.05941>
- Hu, H., Ran, D., Thaichon, P., Wang, B., & Zhu, Z. (2025). Grid-based market sales forecasting for retail businesses using automated machine learning and geospatial intelligence. *Expert Systems with Applications*, 284, 127869. <https://doi.org/10.1016/j.eswa.2025.127869>
- Kizgin, K. T., Alp, S., Aydin, N., & Yu, H. (2025). Machine learning-based sales forecasting during crises: Evidence from a Turkish women's clothing retailer. *Mathematical Problems in Engineering*, 108(1). <https://doi.org/10.1177/00368504241307719>

- Kristian, D., & Suri, P. A. (2025). Comparative study of AutoML and machine learning in retail forecasting.. *Procedia Computer Science*, 269, 882-891.
<https://doi.org/10.1016/j.procs.2025.09.031>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). Predicting/hypothesizing the findings of the M5 competition. *International Journal of Forecasting*, 38(4), 1337-1345. <https://doi.org/10.1016/j.ijforecast.2021.09.014>
- Mansur, S., Nur, S. B., & Rahman, M. M. (2025). Sales forecasting for retail stores using hybrid neural networks and machine learning models. *PeerJ Computer Science*, 11, e3058. <https://doi.org/10.7717/peerj-cs.3058>
- Mejía, S., & Aguilar, J. (2024). A demand forecasting system of product categories defined by their time series using a hybrid approach of ensemble learning with feature engineering. *Computing*, 106, 3945–3965. <https://doi.org/10.1007/s00607-024-01320-y>
- Mustapha, O. O., & Sithole, T. (2025). Forecasting retail sales using machine learning models. *American Journal of Statistics and Actuarial Sciences*, 6(1), 35–67.
<https://doi.org/10.47672/ajsas.2679>
- Pattnaik, M., Padhi, S. K., Panda, L., Naushad, U., Behera, R. K., Mishra, A., & Pattnaik, A. (2025). AI-driven sales forecasting in retail: insights from Big Bazaar's sales data, *International Journal of Management Science and Engineering Management* , 21, 1-14. <https://doi.org/10.1080/17509653.2025.2499076>
- Reddy, K. M., Ravikanth, K., Penjarla, N. K., Poola, S., Patha, S. (2025). Sales Forecasting using Predictive Analytics: A Machine Learning and Time-Series Approach . *International Journal of Research Publication and Reviews*, 6(8), 5476-5483.
<https://doi.org/10.55248/gengpi.6.0825.3193>
- Samal, T, & Ghosh, A. (2026). Ensemble-based predictive analytics for demand forecasting in multi-channel retail. *Expert Systems with Applications*, 299, 130212.
<https://doi.org/10.1016/j.eswa.2025.130212>
- Sattar, M. U., Dattana, V., Hasan, R., Mahmood, S., Khan, H. W., & Hussain, S. (2025). Enhancing supply chain management: A comparative study of machine learning models for forecasting and ESG performance. *Sustainability*, 17(13), 5772.
<https://doi.org/10.3390/su17135772>
- Sun, W., Li, M., Chen, X. H., & Wang, Y. (2024). Enhancing exchange rate prediction and risk management under uncertainty shocks: an AI-driven ensemble prediction model based on metaheuristic optimization. *Ann Oper Res*, 1-39.
<https://doi.org/10.1007/s10479-024-06319-4>
- Theodoridis, G., & Tsadiras, A. (2025). Retail demand forecasting: A comparative analysis of deep neural networks and the proposal of LSTMixer, a linear model extension. *Information*, 16(7), 596. <https://doi.org/10.3390/info16070596>
- Udodiugwu, M. I., Eneremadu, K. E., Onunkwo, A. R., Onyia, M. K., & Gloria, O. C. (2024). The Role of Artificial Intelligence in Enhancing the Performance of Banks in Nigeria. *Arabian Journal of Business and Management Review (Oman Chapter)*, 11(2), 27–34.
<https://j.arabianjbmr.com/index.php/ocAJBMR/article/view/1236>
- Wu, X. (2025). Forecasting Supply Chain Sales and Profits Using Machine Learning: A Comparative Study on Retail Data. *Conference: ICMML 2025: 2025 3rd International Conference on Mathematics and Machine Learning*, 136-141.
<https://doi.org/10.1145/3783779.3783802>
- Zhang, x., Kang, Y., Li., C., Wang, W., & Wang, K. (2025). LSTM-conformal forecasting-based bitcoin forecasting method for enhancing reliability. *PLOS ONE*, 20(5), e0319008. <https://doi.org/10.1371/journal.pone.0319008>