Journal of Kryvyi Rih National University

Vol. 24, No. 1, 2026

ISSN 2306-5451

e-ISSN 2519-2736

Journal homepage: https://journal-knu.com.ua/en

Article's History: Received: 24.09.2025 Revised: 10.03.2026 Accepted: 28.04.2026 Published: 29.05.2026

UDC 004.8

DOI: 10.31721/2306-5451-2026-1-24-35-53

Comparative analysis of machine learning models for retail demand forecasting considering promotional activity and seasonality

Nataliya Boyko*

PhD in Economic Sciences, Associate Professor
Lviv Polytechnic National University
79000, 12 Stepan Bandera Str., Lviv, Ukraine
Stepan Gzhytskyi National University of Veterinary Medicine and Biotechnologies Lviv
80381, 1 V. Velykyi Str., Dubliany, Ukraine
https://orcid.org/0000-0002-6962-9363

Arsen Dolichnyi

Student
Lviv Polytechnic National University
79000, 12 Stepan Bandera Str., Lviv, Ukraine
https://orcid.org/0009-0008-1605-7382

Abstract. The relevance of the study is driven by the complexity of modelling consumer behaviour in an unstable market environment, where forecast accuracy directly impacts inventory optimisation and the reduction of logistical costs. The aim of this study was to develop and substantiate an AI-based intelligent system for demand forecasting and procurement optimisation in retail, ensuring consistent product availability while simultaneously reducing excess inventory and waste. The research methodology was based on a rigorous time-based data split, which allows for evaluating the models' generalisation ability in a realistic "train on the past, test on the future" scenario. To establish a robust point of reference, a hierarchy of baseline benchmarks was developed, including naive forecasts and models with weekly lags. The study conducted a comparative analysis of Ridge linear regression, the Random Forest ensemble method, and gradient-boosted decision trees (XGBoost, LightGBM). Experimental results confirmed a significant advantage of non-linear models over traditional linear approaches, which fail to adequately reproduce threshold effects and complex interactions between promotions and seasonality. It was established that the LightGBM model demonstrates the best stability and accuracy metrics, ensuring minimal error on the validation set while maintaining resistance to overfitting. Specifically, it was found that XGBoost's high efficiency on training data is often a result of over-adaptation to noise, making LightGBM more suitable for practical business applications. The practical significance of the findings lies in the potential for automating procurement processes, ensuring balanced inventory levels and increasing the economic efficiency of retail networks

Keywords: predictive analytics; feature engineering; gradient boosting; time series; mean absolute error; root mean squared error; business analytics

Introduction

The modern retail market is characterised by an extremely high level of dynamism and uncertainty, which places stringent demands on the accuracy of procurement planning processes. The core problem lies in the

Suggested Citation:

Boyko, N.,& Dolichnyi, A. (2026). Comparative analysis of machine learning models for retail demand forecasting considering promotional activity and seasonality. *Journal of Kryvyi Rih National University*, 24(1), 35-53. doi: 10.31721/2306-5451-2026-1-24-35-53.

*Corresponding author (nataliya.i.boyko@gmail.com)

fact that traditional inventory management approaches often prove incapable of adequately responding to sharp fluctuations in consumer demand driven by seasonal factors, the intensity of marketing activities, and the local specifics of individual retail outlets. In product categories with short shelf lives, any forecasting error inevitably leads to significant financial losses: either through stockouts, which result in lost sales and reduced customer loyalty, or through overstocking, which translates into direct losses from the disposal of spoiled goods. The situation is further complicated by the fragmentation of corporate data, where the lack of an integrated connection between sales history, inventory levels, and promotional plans hinders the formation of a unified decision-making cycle. Thus, there is a critical need to transition from manual reactive management to the creation of intelligent systems capable of automating order generation based on probabilistic analysis of large datasets. The relevance of this study is determined by the necessity to develop such integrated solutions that allow for maintaining an optimal balance between service levels and the minimisation of total costs for retail chains in real-time.

Modern retail operates in an environment of high demand volatility, necessitating a transition from traditional planning methods to sophisticated decision support systems. As noted by S. Moret *et al*. (2020), strategic planning under uncertainty requires the development of robust optimisation frameworks capable of integrating large datasets. This is particularly relevant for the retail industry, where multivariate time series (including prices, promotions, and external factors) form the basis for comparative analysis of forecasting models, according to M. Haque *et al*. (2023). The implementation of artificial intelligence (AI) has already demonstrated its effectiveness in the Ukrainian banking sector, as researchers V. Solodkyi & Yu. Polichuk (2023) highlighted, yet in retail, the primary challenge lies in handling high-dimensional data and identifying anomalies in consumer behaviour.

To ensure forecast accuracy, the use of hybrid architectures, such as CNN-LSTM, which the researchers A. Agga *et al*. (2022) and A. Djaafari *et al*. (2022) successfully applied for short-term forecasting in energy sectors, is critical and can be adapted to model sales dynamics. Furthermore, according to H. Xie *et al*. (2022), time-series analysis based on multi-granularity neighbour residual networks allows for the efficient detection of deviations caused by sharp changes in market conditions. A separate aspect is the digitalisation of the economy and the formation of "smart" management systems. The researcher I. Uninets (2021) noted that the subjective disposition of the smart economy emphasises the role of intelligent systems in transforming business processes. Specifically, researchers Yu. Yan *et al*. (2024) suggested that the use of multi-criteria decision-making (MCDM) methods and fuzzy logic allows for balancing sustainability goals with economic efficiency. In resource-constrained environments, according to K. Bülbül & N. Noyan (2021), multi-stage stochastic programming models help optimise resource provisioning, which is highly relevant for logistics and warehouse inventory management.

Modern forecasting approaches also include ensemble learning methods. Researchers B.A. Tama & S. Lim (2021) and M. Rashid *et al*. (2022) concluded that using tree-based stacking ensembles provides higher stability of results compared to single models. This allows not only for predicting sales volumes but also for assessing risks (Expected Shortfall), which A.-P. Fortin *et al*. (2023) noted is crucial for financial procurement planning. However, despite a significant number of theoretical developments, practical implementation experience in 2024-2025 indicates that the primary source of inefficiency remains data fragmentation. The disconnect between sales records, actual inventory levels, and marketing activity plans (promotions) prevents the establishment of a transparent decision-making cycle. The aim of this research was to develop and substantiate an AI-driven system for demand forecasting and procurement optimisation that ensures stable product availability on shelves while simultaneously reducing excess stock and write-offs.

Materials and Methods

Dataset preparation and mathematical formalisation of the forecasting task

The research employed a multi-level methodology, integrating data mining, mathematical modelling, and robust optimisation to build a decision support system for retail procurement. The study utilises the "Corporación Favorita" dataset (Corporación Favorita grocery..., 2017), which provided granular "store-item-day" records, including sales history, promotions, product attributes, store characteristics, transaction flows, and macroeconomic indicators. To ensure computational efficiency, the 125-million-record dataset was systematically sub-sampled. The scope was restricted to high-turnover "Produce" and "Dairy" categories within the Quito metropolitan area over a 24-month period. This duration allowed for the identification of seasonality and trends while minimising logistical noise. To prevent data leakage and simulate real-world conditions, a chronological partition was applied: 21 months for training, 2 months for hyperparameter tuning, and 1 month for final evaluation. This high-density dataset supports the implementation of advanced machine learning ensembles and optimisation algorithms, enabling the formalisation of the forecasting task as a su pervised learning problem within a multi-dimensional state space.

To formalise the training and optimisation processes, the dataset was to be represented as a tensor of observations. Let $S = \{s_1, s_2, \ldots, s_n\}$ be the set of stores and

$I = \{i_1, i_2, \ldots, i_m\}$ be the set of items. To describe each data point at time $t \in \{1, \ldots, T\}$, the following vectors are to be defined. For each object $j = (s, i)$ at time t (1):

$$V_{j,t} = \langle y_{j,t}, p_{j,t} \rangle, \quad (1)$$

where $y_{j,t} \in \mathbb{R}$ is the sales volume (*unit_sales*); $p_{j,t} \in \{0,1\}$ is the promotion indicator (*onpromotion*). Attributes that do not change over time, but determine categorical affiliation (2):

$$C_j = \langle fam_j, per_j, city_s, type_s, clu_s \rangle, \quad (2)$$

where *fam* is the product family; *per* is the perishability; *clu* is the store cluster. Variables common to many objects at time t (3):

$$E_t = \langle o_t, tr_{s,t}, h_t \rangle, \quad (3)$$

where o_t is the oil price; $tr_{s,t}$ is the transaction count for store s; $h_t \in \{0,1\}$ is the holiday/event flag. The common input vector for the machine learning model $X_{j,t}$ is formed by concatenating the lag values of the target variable, static characteristics, and external factors (4):

$$X_{j,t} = \left[\underbrace{y_{j,t-1}, \cdots, y_{j,t-k}}_{Lags}, \underbrace{MA_{j,t,w}}_{Moving\ Averages}, C_j, p_{j,t}, E_t \right], \quad (4)$$

where $MA_{j,t,w}$ is the moving average over the window w. Thus, the forecasting problem is reduced to finding the approximating function f (5):

$$\hat{y}_{j,t+1} = f(X_{j,t}; \Theta), \quad (5)$$

where Θ is the vector of model parameters that minimises the loss function over the entire "store-item-day" space. By consolidating these heterogeneous variables into a structured feature matrix, a comprehensive digital representation of the retail environment is achieved. This mathematical formalisation ensured that the machine learning models receive not only historical sales data but also the crucial contextual information – ranging from local holidays to global economic indicators – that drives consumer behaviour. The transformation of raw records into a unified tensor of observations $X_{j,t}$ provided the necessary consistency for objective model comparison and subsequent integration with the optimisation module.

Integration procedure and feature engineering

The table join was implemented by sequential merge operations on the keys *store_nbr*, *item_nbr*, and *date*. To transform the unidimensional time series into a multidimensional structure capable of reflecting complex nonlinear dependencies, a feature engineering stage was performed: Autoregressive block (Lags); Rolling Windows; Logarithmic transformation. To transform these disparate data sources into a format suitable for machine learning, the structure of the dataset was formalised within a mathematical framework. The use of lagged variables was based on the assumption of demand autocorrelation. A set of lags $L = \{7, 14, 21, 28\}$ was defined to correspond with the weekly cyclicality of the retail sector. Mathematically, a lagged feature is defined using the lag operator (shift operator) (6):

$$L^k y_{j,t} = y_{j,t-k}. \quad (6)$$

Selecting k as a multiple of 7 allows the model to account for intra-week seasonality (e.g., consistently higher sales on Saturdays), thereby minimising the residual variance that static features cannot explain. To neutralise random fluctuations (white noise) and isolate underlying trends, moving averages are applied. The feature $MA_{j,t,w}$ is calculated as follows (7):

$$MA_{j,t,w} = \frac{1}{w} \sum_{i=1}^{w} y_{j,t-i}, \quad (7)$$

where $w \in \{7, 14, 30\}$ represents the window size. A window of $w = 7$ captures short-term fluctuations, such as immediate consumer response to the start of a promotional campaign. A window of $w = 30$ reflects long-term dynamics in product popularity or monthly seasonal demand shifts. Retail sales distributions typically exhibit a "heavy tail" (a high frequency of low-demand items and a few instances of extremely high demand). To stabilise the variance of the results, the following transformation is applied (8):

$$z_{j,t} = ln(1 + y_{j,t}). \quad (8)$$

This enables the models to better approximate relative (percentage) changes rather than absolute ones, which is critical for minimising the Root Mean Square Error (RMSE) across the entire range of values. The impact of a promotional activity $p_{j,t}$ is modelled not merely as a linear offset but as a multiplicative effect. In the logarithmic feature space, this is represented as an additive interaction (9):

$$\hat{z}_{j,t} = \cdots + \beta \cdot p_{j,t} + \gamma \cdot (p_{j,t} \cdot MA_{j,t,w}), \quad (9)$$

where γ accounts for the fact that the promotional effect on a high-demand item (high *MA*) is significantly larger than on an item with low baseline demand. After the preparation procedures, the final feature matrix $X_{j,t}$ included the following: an Autoregressive Block (4 lagged variables), a Trend Block (3 rolling average features), an Exogenous Block (oil prices, transaction counts, and holiday flags), and a Categorical Block (5 features: store, item, category, city, and promotion status). This data structure reduces the entropy of the input signal, enabling gradient boosting algorithms such as XGBoost and LightGBM to converge more efficiently.

Mathematical representation of the input vector and modelling approach

The final input vector $X_{j,t}$ for the machine learning models was constructed by concatenating all extracted feature blocks (10):

$$X_{j,t} = [Lags_{j,t}, MA_{j,t,w}, C_j, p_{i,t}, E_t]. \quad (10)$$

Consequently, the forecasting problem is reduced to finding an approximating function f that minimises the error across the entire "store-item-day" space (11):

$$\hat{y}_{j,t+1} = f(X_{j,t}; \theta). \quad (11)$$

The established data architecture (Fig. 1) ensures a reduction in the entropy of the input signal, enabling gradient boosting algorithms to achieve high forecasting accuracy.

Subsequently, Figures 2 was analysed systematic and integrated into a single working dataset to build the demand forecasting models.

```
train: (125497040, 6) – daily sales history
items: (4100, 4) – product catalogue
stores: (54, 5) – shop directory
transactions: (83488, 3) – number of transactions by shop
holidays: (350, 6) – calendar of events and holidays
oil: (1218, 2) – daily oil prices
```

a

```
--- train.csv (main sales table)---
      id date store_nbr item_nbr unit_sales onpromotion
  0    0    2013-01-01      25      103665     7.0      NaN
  1    1    2013-01-01      25      105574     1.0      NaN
  2    2    2013-01-01      25      105575     2.0      NaN
  3    3    2013-01-01      25      108079     1.0      NaN
  4    4    2013-01-01      25      108701     1.0      NaN
Columns:
- date – date of sale
- store_nbr – shop number
- item_nbr – product number
- unit_sales – number of units sold
- onpromotion – was the product on offer that day
```

b

Figure 1. Architectural diagram of the intelligent demand forecasting and inventory replenishment system
Notes: a – data ingestion and preprocessing pipeline featuring anomaly detection and feature engineering; b – predictive engine and optimisation module for generating automated procurement orders
Source: created by the authors

```
--- items.csv (product catalogue)---
     item_nbr family class perishable
0       96995    GROCERY I      1093         0
1       99197    GROCERY I      1067         0
2      103501     CLEANING      3008         0
3      103520    GROCERY I      1028         0
4      103665  BREAD/BAKERY     2712         1
Columns:
- item_nbr - product number
- family - product family (category, e.g. Produce, Dairy, etc.)
- class 0150 - internal product class
- perishable - if the goods are perishable; 0 if not
```

a

```
--- stores.csv (shop directory)---
 store_nbr      city                          state type cluster
  0     1       Quito                         Pichincha   D    13
  1     2       Quito                         Pichincha   D    13
  2     3       Quito                         Pichincha   D    8
  3     4       Quito                         Pichincha   D    9
  4     5   Santo Domingo  Santo Domingo de los  Tsachilas   D    4
Columns:
- store_nbr - shop number
- state - region (state/province)
- type - type of shop (format)
- cluster - shop cluster (a grouping of similar shops)
```

b

```
--- transactions.csv (transactions)---
         date         store_nbr    transactions
0     2013-01-01         25           770
1     2013-01-02         1            2111
2     2013-01-02         2            2358
3     2013-01-02         3            3487
4     2013-01-02         4            1922
Columns:
- store_nbr - shop number
- transactions - the number of receipts (transactions) in the
shop on that day
```

c

```
--- holidays_events.csv (events and holidays)---
      date      type    locale  locale_name       description       \
0  2012-03-02 Holiday   Local      Manta          Fundacion de Manta
1  2012-04-01 Holiday Regional   Cotopaxi       Provincializacion de
                                                      Cotopaxi
2  2012-04-12 Holiday   Local      Cuenca         Fundacion de Cuenca
3  2012-04-14 Holiday   Local     Libertad   Cantonizacion de Libertad
4  2012-04-21 Holiday   Local     Riobamba   Cantonizacion de Riobamba
   transferred
0    False
1    False
2    False
3    False
4    False
Contains information on:
- holidays
- special events
- working/non-working days
which may affect demand
```

d

Figure 2. Comparison of actual sales data and forecasting results of various machine learning models
Notes: a – items; b – stores; c – transactions; d – holidays_events
Source: created by the authors

The primary objective is to forecast daily unit sales at the granular "store-item-date" level to support inventory optimisation and supply planning. This panel time series task utilises a global modelling approach, training on network-wide data to capture localised demand patterns. To ensure high predictive accuracy, the methodology integrates calendar and promotional features, supplemented by engineered historical variables such as moving averages and lagged demand. The modelling strategy employs a hierarchical approach to balance complexity and interpretability. Linear models serve as an essential baseline, providing computational efficiency and clear variable coefficients, particularly when applied to log-transformed data. However, due to their limited capacity to handle complex non-linearities, the core of the research relies on tree-based ensemble methods – specifically Random Forest, XGBoost, and LightGBM. These architectures are preferred for their robust ability to process heterogeneous feature sets and model intricate interactions, providing the accuracy required for effective retail demand forecasting.

Procurement optimisation module

The proposed intelligent retail management system was based on a two-tier decision-making architecture. This approach allows for the separation of market uncertainty processing into two sequential stages: statistical modelling of future demand and normative optimisation of logistical processes. The following provided a detailed mathematical formalisation of the system's functional components: the intelligent demand forecasting module and the resource optimisation module. The first level of the proposed system focused on solving the predictive analytics problem – constructing high-precision consumer demand forecasts. The functional purpose of this module was to approximate complex, non-linear dependencies between a vector of multi-factor input features and the resulting sales volume.

Below is a formalised description of the data structure and model training objective functions. Let $Y = \{y_1, y_2, \ldots, y_t\}$ be a time series of sales volumes for a specific product. The forecasting problem is reduced to finding a function f that minimises the expected error over the forecasting horizon h (12).

$$\hat{y}_{t+h} = f(X_t, P_t, S_t, E^t) + \epsilon, \quad (12)$$

where $\hat{y}_{t+h}$ is the predicted demand for the period $t + h$; $X_t = \{y_{t-1}, \ldots, y_{t-n}\}$ is the vector of autoregressive features (lags of 7, 14, and 30 days); P_t is the marketing vector (presence of promotions, discount levels); S_t represents calendar features (seasonality, day of the week, holidays); E_t denotes external factors (oil price, inflation index), ϵ is the random noise.

To train the gradient boosting models (XGBoost, LightGBM), the RMSE or LogLoss loss function (in case of log-transformation of the target variable) is used (13):

$$L(\theta) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - f(x_i;\theta)\right)^2} \to \min_{\theta}, \quad (13)$$

where N is the total number of observations in the training set; y_i represents the actual sales value for the i-th observation; $f(x_i; \theta)$ denotes the predicted value generated by the model for the input vector x_i with parameters θ; $(y_i - f(x_i; \theta))^2$ is the squared error, which penalises larger deviations more heavily, ensuring the model's robustness against significant forecasting errors. The next critically important stage of the system's operation is the transformation of the obtained predicted values $\hat{y}$ into specific management decisions regarding inventory replenishment volumes. While forecast accuracy is a necessary condition, it does not inherently account for the complex structure of logistical costs, warehouse capacity constraints, or the financial risks associated with stockouts.

To address this challenge, a prescriptive optimisation module was developed to integrate the results of intelligent demand analysis into a normative inventory management model. This facilitates a transition from passive future prediction to the active formation of an optimal procurement strategy based on minimising total operational costs. The function of this module was to transform the obtained point forecast $\hat{y}$ into a specific decision regarding the procurement volume Q, while accounting for economic risks and resource constraints. The objective function is to minimise the total logistical costs (Z) for each SKU over the planning horizon T (14):

$$Z = \sum_{t=1}^{T}(C_h \cdot I_t + C_s \cdot S_t + C_o \cdot O_t) \to \min, \quad (14)$$

where t is the index of the calculation period (e.g., day or week); T is the total planning horizon (the number of periods for which the procurement schedule is generated); i is the index of a specific Stock Keeping Unit (SKU), applied when the model considers a multi-product assortment. C_h (Holding Cost) is the unit cost of storing one item of inventory during one period t. This includes warehouse rent, utilities, insurance, and capital opportunity costs. C_s (Shortage Cost) is the penalty per unit of shortage (lost sales). This represents the sum of unrealised profit and potential losses due to decreased customer loyalty. Typically, $C_s \gg C_h \cdot C_o$ (Ordering Cost) is the fixed cost associated with placing and delivering a single order.

These costs are independent of the batch size and are incurred per transaction (administrative and logistical setup costs). I_t (Inventory Level) is the volume of stock available in the warehouse at the end of period . This is a state variable that must be non-negative ($I_t \geqslant 0$). S_t (Shortage Quantity) is the volume of unsatisfied demand in period t. This occurs when the sum of current inventory and new arrivals is less than the predicted demand $\hat{y}_t$. O_t (Ordering Indicator) is a binary decision variable ($O_t \in \{0,1\}$), where $O_t = 1$ if an order is placed in period t and $O_t = 0$ otherwise. The system of constraints includes several key components. Inventory Balance Constraint defines the stock level at the end of period t is defined by the following recursive equation (15):

$$I_t = I_{t-1} + Q_t - \hat{y} + S_t, \quad (15)$$

where $I_t \geqslant 0$ is the physical inventory level at the end of the period; $S_t \geqslant 0$ represents the shortage volume (backorder), which occurs if the actual demand exceeds the available stock. This constraint ensures the continuity of material flows by linking the previous state of the system with current replenishment and demand.

Logistical Ordering Rule: to account for supply chain requirements and transport capacities, the order volume Q_t is constrained as follows (16):

$$Q_{min} \cdot O_t \leqslant Q_t \leqslant Q_{max} \cdot O_t, \quad (16)$$

where $O_t \in \{0,1\}$ is a binary decision variable: 1 if a procurement order is placed, and 0 otherwise; Q_{min} and Q_{max} define the minimum and maximum allowable delivery batch sizes, respectively. This constraint prevents economically inefficient small-scale deliveries and aligns the model with realistic logistical standards. Warehouse Capacity Constraint considers physical space

limitations within the storage facility, the following condition must be met for all n products (17):

$$\sum_{i=1}^{n} v_i \cdot I_{it} \leqslant V_{total}, \quad (17)$$

where v_i is the specific unit volume of the i-th product (SKU); V_{total} is the total usable capacity of the warehouse. This ensures that the cumulative volume of all stored inventory items does not exceed the physical capacity of the warehouse at any point in time t.

Demand modelling and forecast evaluation

As a potential extension, sequence-oriented neural networks may be explored to capture complex patterns, though they remain secondary to the core methodology due to high computational demands and limited interpretability. Consequently, the project focuses on a hierarchy of regression models for short-term daily demand forecasting at the "store-item-day" level, ranging from linear benchmarks to sophisticated tree ensembles. By integrating diverse features such as sales history, promotions, and macroeconomic indicators, this tiered approach enables a rigorous evaluation of the incremental benefits of model complexity.

Model performance is assessed using standard regression metrics, including Mean Absolute Percentage Error (MAPE), RMSE, and the coefficient of determination (R^2). While MAPE and RMSE quantify predictive deviation, R^2 serves as a critical indicator of the proportion of variance explained by the model, where values approaching 1 denote high explanatory power. By establishing these statistical benchmarks, the study effectively compares various machine learning architectures to ensure accuracy before integrating the final forecasts into the procurement optimisation module.

▼ Mean actual value (18):

$$\bar{A} = \frac{1}{n}(1/n) \cdot \sum_{i=1}^{n} A_t, \quad (18)$$

where $\bar{A}$ – mean actual value, used in this study to normalise scale-dependent errors and calculate the relative reliability of forecasts across different product categories; A_t – actual demand value at time t; n – total number of observation periods within the test dataset.

▼ Mean Absolute Percentage Error (19):

$$MAPE = (1/n) \cdot \sum_{i=1}^{n} \left|\frac{A_t - F_t}{A_t}\right|, \quad (19)$$

where A_t – actual value; F_t – forecast value; n – number of observation periods.

▼ Root Mean Square Error (20):

$$RMSE = \sqrt{(1/n) \cdot \sum_{i=1}^{n} (A_t - F_t)^2}. \quad (20)$$

▼ Total Sum of Squares (21):

$$TSS = \sum_{i=1}^{n} (A_t - \bar{A})^2. \quad (21)$$

where A_t – actual demand value at time t; $\bar{A}$ – mean actual value calculated across the observation period; n – number of observations.

▼ Residual sum of squares (22):

$$RSS = \sum_{i=1}^{n} (A_t - F_t)^2, \quad (22)$$

where F_t – forecast value generated by the model for the same period. The coefficient of determination is then computed as follows (23):

$$R^2 = 1 - \frac{RSS}{TSS}, \quad (23)$$

where R^2 is used as an additional summarising metric to evaluate what proportion of demand fluctuations the model can explain through the features used. However, the primary emphasis in interpreting the results is placed on MAE, RMSE, and MAPE, as these metrics most directly reflect the expected forecast error in quantitative and relative terms.

In addition to standard deviation metrics, the Weighted Average Percentage Error (WAPE) is employed to evaluate forecasting accuracy across the entire product assortment. Unlike MAPE, this metric avoids the problem of division by zero (in cases of zero sales) and mitigates the impact of outliers in low-volume products by assigning greater weight to high-velocity SKUs (24):

$$WAPE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} y_i}, \quad (24)$$

where y_i is the actual sales volume of the i-th product; $\hat{y}_i$ is the predicted sales volume of the i-th product; n is the number of observations (products or time periods). In conclusion, the selected set of metrics forms a multi-layered framework for evaluating the performance of the forecasting models. While the coefficient of determination (R^2) and the sums of squares (TSS, RSS) provide insights into the overall explanatory power of the models, metrics such as MAPE, RMSE, and WAPE offer a practical understanding of error in both quantitative and percentage terms. The application of WAPE was of particular importance, as it enables a weighted assessment of forecasting quality across the entire product assortment, mitigating distortions caused by low-velocity items.

Results

Demand forecasting and impact analysis on key economic indicators

The experimental stage of this study was aimed at verifying the developed mathematical framework and conducting a comparative performance evaluation of the selected machine learning models. The utilisation of the engineered feature matrix $X_{j,t}$, which incorporates autoregressive lags, rolling averages, and external economic indicators, enabled comprehensive model testing within a realistic "one-step-ahead" forecasting scenario. The evaluation of the results was carried out according to a two-tier scheme.

At the first level, the predictive accuracy of regression algorithms (Ridge, XGBoost, LightGBM) was analysed using statistical error metrics. At the second

level, the outputs from the best-performing predictive model were integrated into the Mixed-Integer Linear Programming (MILP) procurement optimisation module to assess their impact on key economic indicators: stockout levels, waste volumes (write-offs), and total logistical costs. To quantitatively validate the effectiveness of the proposed feature engineering approach, a rigorous benchmarking of the selected models was performed. The comparative analysis of their predictive performance, measured across several key error metrics, is visualised in Figure 3. This graphical representation allows for a clear identification of the top-performing algorithms in terms of stability and deviation from the actual demand.

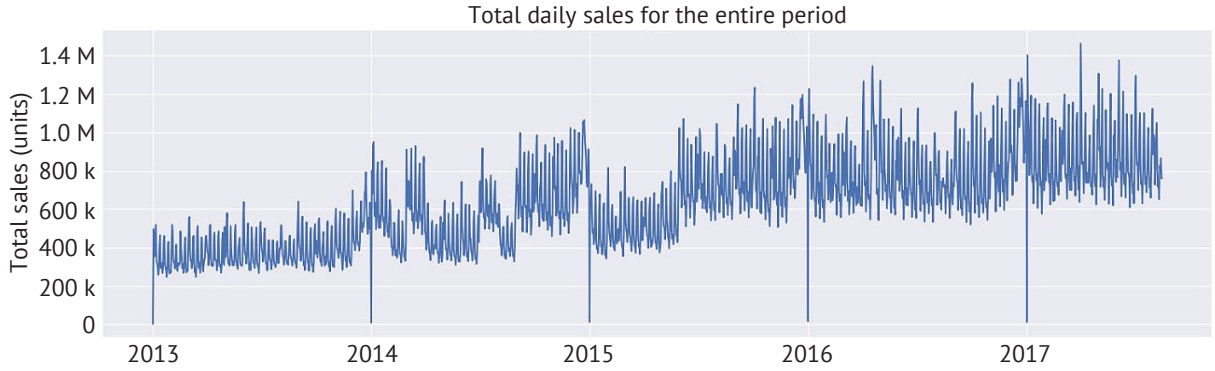

Figure 3. Daily sales history graph

Source: created by the authors

According to the presented chart, a distinct seasonality in daily sales dynamics can be observed. Regular fluctuations in demand are evident across the timeline: periods of growth and decline repeat consistently from year to year. Furthermore, a general upward trend in sales volume is observed over time, indicating a gradual expansion of the retail chain's operations or shifts in consumer behaviour. The chart also reveals sharp peaks and troughs, which may be attributed to holidays, promotional activities, or other external factors. Thus, the data demonstrate both short-term fluctuations and long-term trends, which are essential for building an accurate demand forecasting model (Fig. 4a). As shown in the chart, total sales demonstrate steady growth between 2013 and 2016. This may indicate an expansion of the product range, an increase in the number of customers, or the growth of the retail chain. Following the peak value in 2016, a certain decline is observed in 2017, however, this is due to the fact that the data for 2017 is only available through August (Fig. 4a). The chart (Fig. 4b) displays a clear seasonality: sales increase almost every year towards the end of the year, particularly in November and December. Additionally, dips can be observed in the summer months during certain years, indicating temporary fluctuations in demand. The chart (Fig. 5) indicates that the clusters differ significantly in terms of sales levels, with Cluster 14 making the largest contribution and substantially outperforming the others. The lowest sales volumes are observed in Clusters 1 and 15, which may suggest lower commercial activity or a smaller scale of stores within these groups.

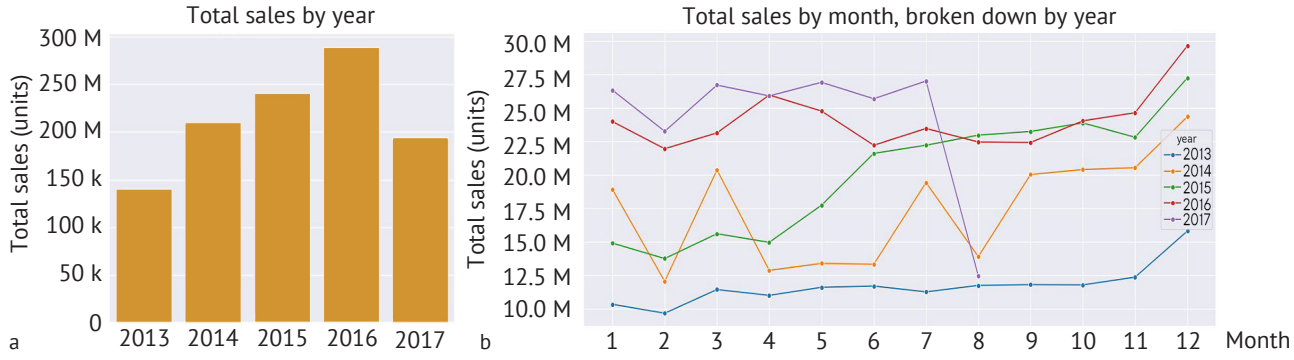

Figure 4. Evaluation of forecasting accuracy across different machine learning models using standard error metrics

Notes: a – total sales by year; b – total monthly sales by year

Source: created by the authors

Figure 5. Top 10 clusters by sales volume

Source: created by the authors

In this dataset, a cluster refers to a group of stores categorised by similar characteristics. Favorita classifies stores into clusters based on location, size, neighbourhood type, customer activity, and assortment to analyse their behaviour as distinct segments. This approach allows for monitoring the sales dynamics of different store groups and identifying similar demand patterns. The chart (Fig. 6) shows that the number of transactions exhibits regular seasonal fluctuations, with sharp increases during peak periods (holidays and sales). There are also days with near-zero activity, which likely indicate store closures or technical data inconsistencies. The scatter plot (Fig. 7a) reveals a noticeable negative correlation: during periods of lower oil prices, total sales were predominantly higher. This may indicate a dependence of consumer spending on energy price levels. The correlation matrix (Fig. 7b) shows that sales are strongly correlated with the number of transactions and the average number of items per purchase, which is logical and expected. Furthermore, there is a notable negative correlation with oil prices, confirming the potential influence of macroeconomic factors on purchasing activity.

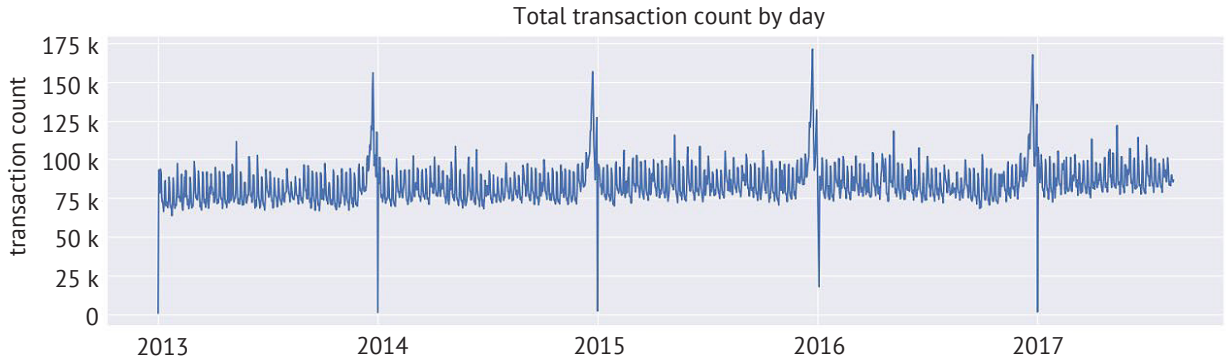

Figure 6. Total transaction count by day

Source: created by the authors

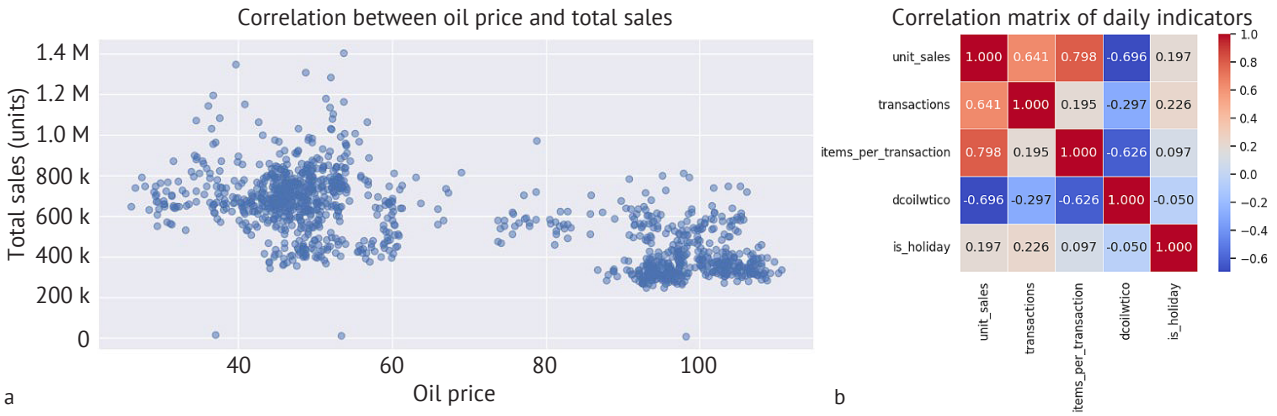

Figure 7. Feature importance analysis for the demand forecasting model

Notes: a – correlation chart between oil price and total sales; b – correlation matrix of daily indicators

Source: created by the authors

To integrate oil prices as a critical macroeconomic indicator, the forecasting model is structured as a supervised learning task within a multi-dimensional state space. By mapping historical sales data alongside internal drivers and external economic factors into a unified "store-item-day" dataset, the model effectively captures the complex dynamics of retail demand. This approach balances computational feasibility with statistical rigor, allowing the system to treat global macroeconomic shifts – such as oil price fluctuations – as dynamic regulators of consumer purchasing power. Ultimately, this demand forecasting framework effectively accounts for seasonality, long-term trends, specific store activity, and broader macroeconomic conditions. By synthesising these diverse variables, the model significantly enhances the accuracy of sales predictions, providing a robust foundation for optimising daily supply planning, inventory management, and overall logistics performance.

Sub-sample construction and data preparation

The data preparation process began with a strategic decision to limit the task to a ma nageable scale, as the full Corporación Favorita dataset contains an excessive number of "store-item-date" combinations, requiring prohibitive computational resources for rapid experimentation. Consequently, a representative sub-sample was formed: a single store and several key product categories were selected to preserve demand diversity while ensuring stable computations on the Central Processing Unit (CPU). For each selected category, a limited number of the most representative items were chosen based on sufficient sales volume and steady dynamics. This filtering approach allows for the significant reduction of noise originating from rare or nearly inactive items, thereby improving the quality of the training signal and the overall robustness of the forecasting models (Fig. 8).

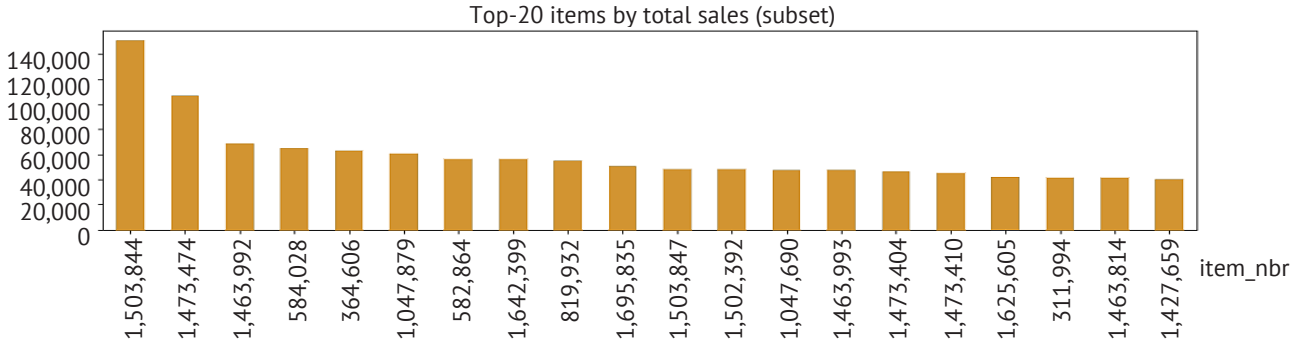

Figure 8. Top 20 items by total sales volume in the sub-sample

Source: created by the authors

To avoid data leakage during sub-sampling, the selection of the most representative items was conducted using only the portion of the data preceding the validation period, rather than the entire history. The reasoning is straightforward: if the most recent days were used to select items, the sub-sample could "overfit" to the future period, artificially inflating model performance metrics. Therefore, the selection was based on historical sales excluding the future validation window. Next, the sales data was organised into a format suitable for time-series analysis: a complete daily panel was constructed, ensuring that all selected items are present for every day. This is crucial because, in real-world retail data, the absence of a record for a specific day does not mean the day did not exist -more often, it indicates zero sales or a gap in registration. A continuous daily structure guarantees that the time series for each item is consistent, allowing for the correct construction of features based on demand history (lags, moving averages, etc.) (Fig. 9).

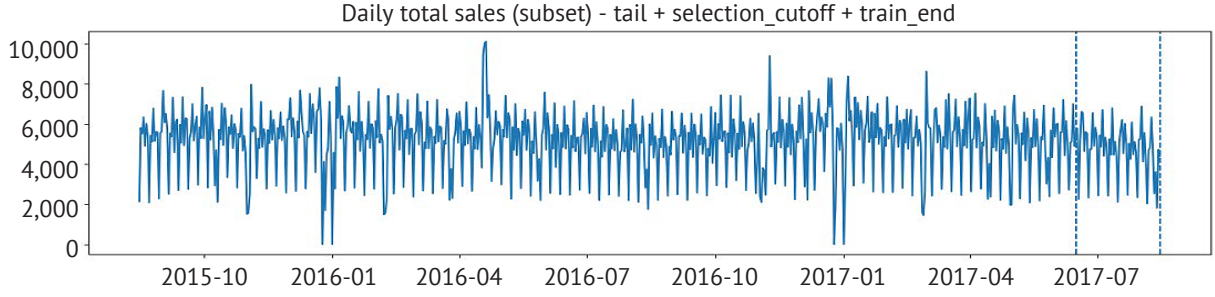

Figure 9. Total sales by product families in the sub-sample

Source: created by the authors

In parallel, the sales values were cleaned and adjusted to a logically consistent form. Since retail data sometimes contains incorrect values (specifically negative numbers due to returns or accounting errors), these were eliminated. Additionally, to stabilise the target variable, a transformation was applied to reduce the impact of rare peak values. Demand often exhibits a "long tail": most days result in small or medium sales, but occasional sharp spikes occur. Stabilising the distribution helps models learn better from typical behaviour rather than overfitting to outliers (Fig. 10).

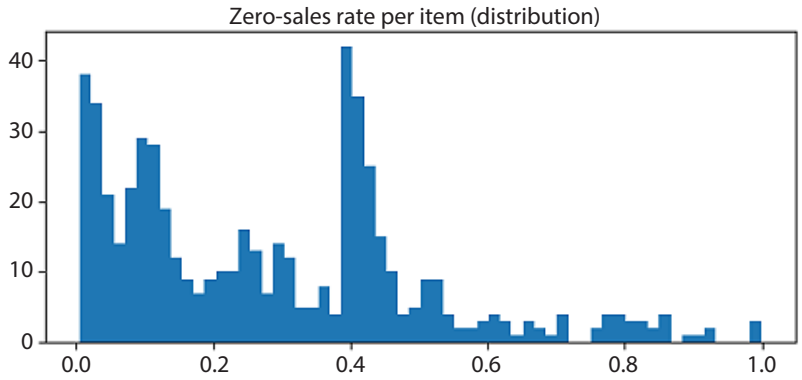

Figure 10. Distribution of the proportion of zero-sales days for each item
Source: created by the authors

Promotional data was integrated separately, as promotions significantly affect demand. Without accounting for this factor, the model typically makes systematic errors: underestimating forecasts on promo days and overestimating them on regular days. Promotional information was included to reflect a real-world forecasting scenario: for future dates, only the promo data known within the planning horizon is used, while missing values are filled conservatively. Next, external and calendar factors were added, including store transactions, the oil price indicator, and holiday/event features. When filling missing values, a cautious strategy was employed to avoid "look-ahead bias": values were propagated only forward in time, and baseline imputations were calculated using only known historical data. For holidays and events, a store-specific mapping was applied, where national-level events were included along with relevant regional and local events (Fig. 11).

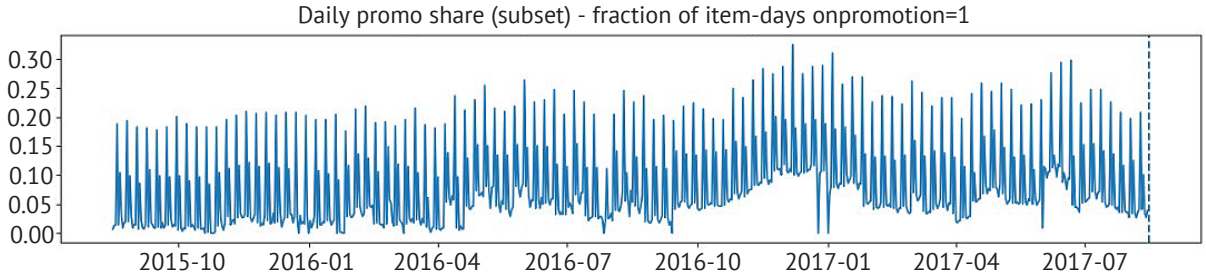

Figure 11. Proportion of item-days with promotions in the sub-sample over tim
Source: created by the authors

After merging all sources into a unified table, a time-based split was performed for training and validation. Validation was conducted on a later time interval to simulate a real-world forecasting task: the model was trained on past data and its performance was evaluated on future dates. This approach provides an unbiased assessment of the model's ability to generalise demand patterns (Fig. 12a). During the feature engineering stage, characteristics were generated to reflect demand history and context, using only historical values relative to each date. Specifically, features representing previous sales values, their averages and volatility within moving windows, exponential smoothing, and promotional intensity indicators from prior days were added. Additionally, aggregated category context was utilised, allowing the model to perceive broader demand trends within a specific product group. Since these features are potentially the most sensitive to data leakage, their construction logic was strictly "from past to future", ensuring that no current or future values were used in the calculations (Fig. 12b).

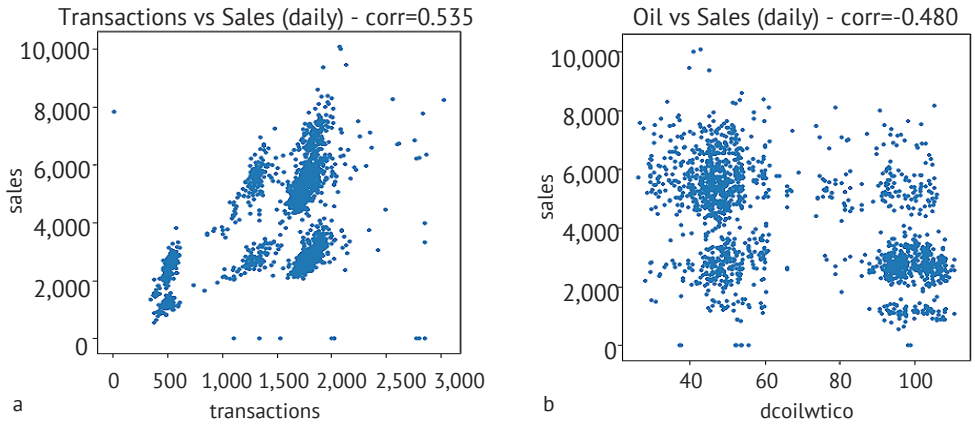

Figure 12. Structural diagram of the deep learning model architecture and training process

Notes: a – relationship between transaction count and total sales (daily values) in the sub-sample; b – relationship between oil price and total sales (daily values) in the sub-sample

Source: created by the authors

As a result of this preparation, a time-consistent training set and a forecasting set for future dates were obtained. The finalised sub-sample ensures a balance between the diversity of demand patterns and the ability to conduct fast and stable experiments. Within the scope of this work, machine learning models are utilised in a unified demand forecasting framework: a forecast for future dates is generated based on sales history and contextual factors (Guo *et al.*, 2022; Kim & Jang, 2023). A consistent feature engineering logic and time-based data splitting are applied across all approaches to ensure a fair comparison, where differences in results are attributable to the properties of the algorithms themselves rather than varying training conditions. Special care is taken to prevent "look-ahead bias": training is performed on earlier dates, while validation is conducted on a later segment that simulates a real-world forecasting scenario where future sales values are unknown.

Before deploying complex algorithms, baseline benchmarks – simple base models – are established. These are necessary to provide a clear point of reference: if a complex model cannot outperform simple rules, it indicates either that the features are insufficiently informative or that the chosen algorithm does not suit the data structure. Naive forecasts reflecting typical retail logic are used as benchmarks: a zero-sales forecast, a global mean forecast, an item-specific mean forecast, and a persistence forecast based on the recent past (e.g., repeating the sales from the same day last week). While these approaches do not require complex training, they often provide respectable quality in tasks involving seasonality, making them an essential minimum standard for evaluation.

Baseline models and performance benchmarking

Alongside the baseline benchmarks, Linear Regression is applied as a simple and interpretable starting point. It is particularly useful when the relationship between features and demand can be approximated as a sum of the influences of individual factors. The linear approach also helps verify feature quality: if even a simple model consistently captures part of the underlying patterns, it confirms high-quality data preparation. To improve training stability, features are scaled to a consistent range to ensure that factors of different magnitudes influence the forecast correctly.

Next, Random Forest is utilised as an ensemble method based on multiple decision trees (Xu *et al.*, 2021; Yin *et al.*, 2023). Its advantage lies in its natural ability to capture non-linear dependencies and feature interactions without requiring an explicit definition of these relationships. In retail demand, this is crucial because responses to promotions, seasonality, and calendar events are often non-linear. Random Forest is also more robust to noise, as the forecast is averaged across many trees, and the construction of each tree involves random selection of data and features, which increases ensemble diversity and reduces the risk of overfitting.

A separate class of models is represented by Gradient Boosted Decision Trees (GBDT), implemented in this work using two popular libraries (Rashid *et al.*, 2022). Gradient boosting builds trees sequentially, with each subsequent tree correcting the errors of the previous ones. This enables the model to capture complex, subtle dependencies between historical sales, promotional activity, seasonal patterns, and context. This approach is typically among the strongest for tabular data with heterogeneous features. To ensure stability and generalisation, the complexity of individual trees is restricted, and the number of trees and the learning rate are controlled to prevent the model from memorising history instead of learning patterns.

In summary, the model suite is designed to cover various levels of complexity and different data assumptions: from naive baseline rules as a minimum

standard, through a simple linear model as an interpretable benchmark, to tree ensembles and gradient boosting as flexible methods for non-linear dependencies. This approach allows for a justified comparison of algorithms under identical conditions, highlighting the role of baseline benchmarks and providing insights into which classes of models are best suited for demand forecasting tasks accounting for promotions, seasonality, and contextual factors.

The simplest benchmark was a "zero-sales" baseline. It is used not as a realistic forecasting method, but as a control point: if a model cannot outperform a constant zero, it indicates that the features, data splitting, or training logic are fundamentally flawed. For this baseline, the following results were obtained: MAE = 8.46, RMSE = 16.97, a coefficient of determination R^2 = -0.33, and a WAPE = 1.00 (meaning the total volume error is 100%). These values are expected: a zero-sales forecast ignores the data structure and fails to explain any demand fluctuations.

BL_Zero: {'MAE' : 8.464005470275879, 'RMSE' : 16.97098173917485, 'R2' : -0.3310892581939697, 'WAPE' : 1.0, 'N' : 33480}

The second baseline benchmark was a forecast based on the global mean sales level from the training set. This is a simple assumption that demand fluctuates around a single average value and does not require item-level detail. For this baseline, the results were: MAE = 7.64, RMSE = 14.75, R^2 = -0.005, and WAPE = 0.90. While this benchmark is noticeably better than the zero forecast, it still explains almost none of the variation is close to zero) because it fails to account for the differences between items and their individual dynamics. This result demonstrates that a simple global average is insufficient for a task characterised by high assortment heterogeneity.

BL_GlobalMean: {'MAE' : 7.644719123840332, 'RMSE' : 14.747846543178188, 'R2' : -0.0051953792572021484, 'WAPE' : 0.9032034873962402, 'N' : 33480}

The next step was a baseline that accounts for individual product differences by predicting each item's specific average sales level (calculated from the training history). This approach partially models structural heterogeneity: different products have different baseline demand scales. The results for this baseline are significantly better: MAE = 5.51, RMSE = 11.19, R^2 = 0.42, and WAPE = 0.65. The emergence of a positive R^2 indicates that accounting for item identity indeed explains a significant portion of the data variance, even without utilising seasonality, promotions, or lags.

BL_ItemMean: {'MAE' : 5.506714344024658, 'RMSE' : 11.189294766999952, 'R2' : 0.42137300968170166, 'WAPE' : 0.6506038308143616, 'N' : 33480}

The strongest among the baseline approaches was the weekly lag baseline, which forecasts demand based on the item's sales values from exactly one week prior. In retail data, this is a very robust concept because demand often exhibits weekly seasonality, where similar days of the week repeat comparable patterns. This baseline achieved the following results: MAE = 4.63, RMSE = 10.40, R^2 = 0.50, and WAPE = 0.55.

Compared to the item-specific mean, it improves performance across all key metrics, confirming the importance of short-term temporal context and demand cyclist. Consequently, the weekly lag baseline should be considered the primary competitor for machine learning models: if an ML model fails to outperform this level, it is either overfitting or not receiving sufficiently informative features. Thus, the baseline models establish a clear scale of complexity and quality: from a completely uninformative forecast to a strong temporal benchmark that leverages weekly recurrence.

Model performance and overfitting analysis

The resulting metrics demonstrate that the most significant contributions to quality improvement at the baseline level come from the transition from a global mean to an item-specific mean, followed by the shift to a weekly lag, which adds temporal context (Fig. 13).

In this experiment, the Ridge linear model effectively fails as a candidate for a high-quality forecast, serving instead as a clear illustration of the limitations of the linear approach for demand forecasting. Its validation metrics are inferior not only to tree-based models but even to the stronger weekly seasonality baseline. The reason is that Ridge attempts to describe demand as an approximately linear combination of factors. However, in retail sales, key effects are almost always non-linear: promotions work differently for different products, interact with days of the week and seasons, and context-based reactions often have thresholds (e.g., demand shifts abruptly on specific days or during events). Furthermore, the Lag-7 baseline directly utilises the strongest structure in this data – weekly recurrence – whereas Ridge reproduces it only indirectly through calendar features and averaged statistics. Consequently, the model is unable to capture the actual shape of temporal patterns, leading to weak explanatory power (low R^2) and a performance that lags behind even a simple seasonal heuristic. Interestingly, the gap between training and validation performance for Ridge is not catastrophic; this indicates that the issue is not overfitting, but rather underfitting due to model bias – the model form is simply too basic for this task (Fig. 14).

Random Forest demonstrates significantly more adequate behaviour and stands out as a strong, stable, and reliable option. Its performance on the training set is solid, and the degradation on the validation set is moderate – indicating that the model does not merely fit the training data but effectively transfers

patterns to the future period. The advantage of RF in this task stems from the fact that trees naturally capture non-linear interactions (e.g., different demand regimes depending on promotions, day of the week, and seasonality). It can handle conditional rules such as: "if a promotion is active and it is a weekend, expect different behaviour than on weekdays without a promotion". Furthermore, by its nature, Random Forest is often less prone to severe overfitting because the forecast is averaged across many trees, and the inherent randomness (sub-sampling of both data and features) acts as a form of regularisation. This explains why your RF model shows a narrow gap between training and validation performance – a clear sign of healthy generalisation (Fig. 15).

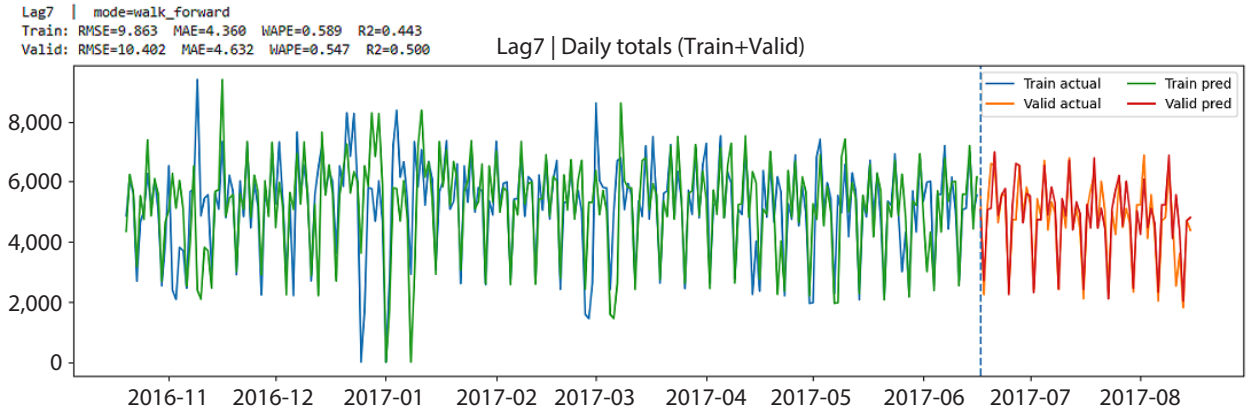

Figure 13. Actual vs. predicted total daily sales for the weekly lag baseline

Source: created by the authors

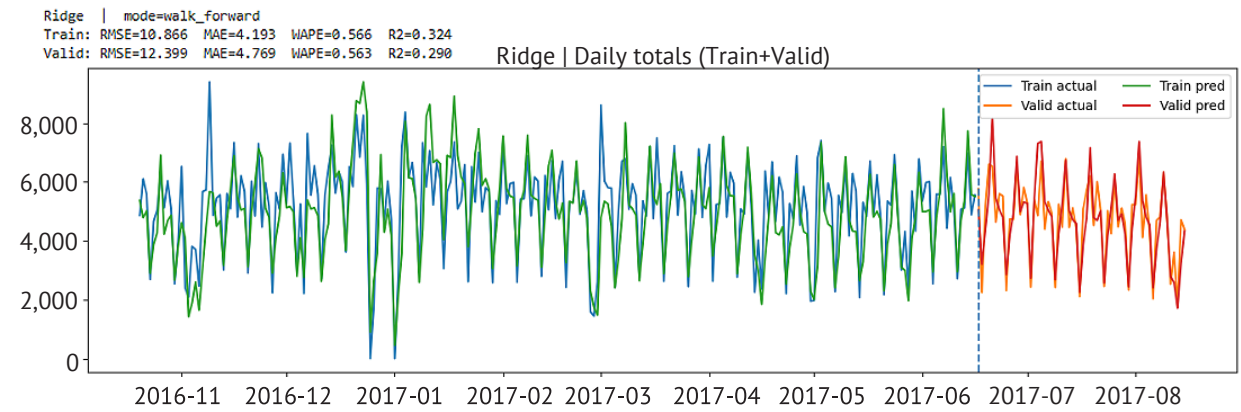

Figure 14. Actual vs. predicted total daily sales for Linear Regression

Source: created by the authors

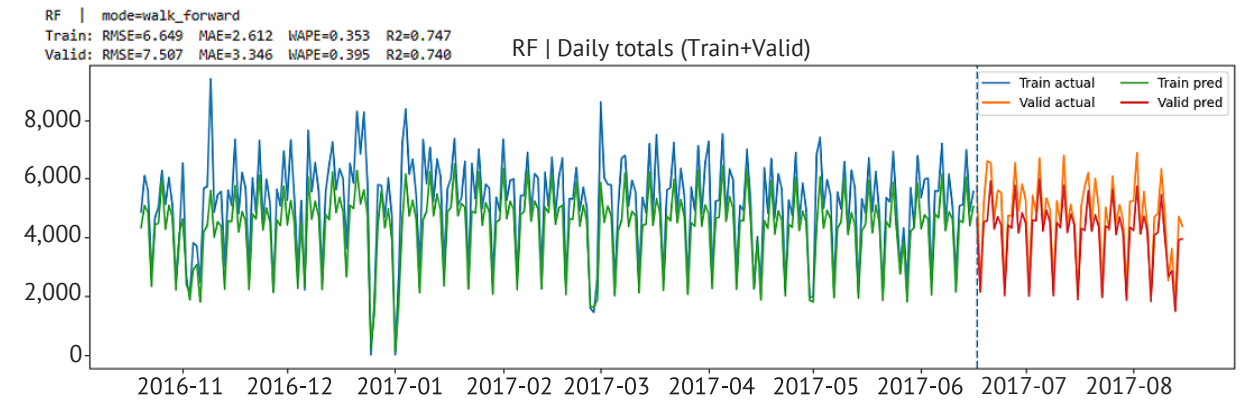

Figure 15. Actual vs. predicted total daily sales for Random Forest

Source: created by the authors

XGBoost exhibits a typical issue associated with the excessive power of boosting: it is nearly perfect on the training set (displaying very low errors and a very high R^2), but on the validation set, its performance returns to the level of other strong models, failing to translate its training success into a corresponding gain in the future. This is a classic profile of overfitting or over-adaptation to the specific details of the training period. Essentially, the model learns not only the general patterns (weekly seasonality, inertia, promotional effects) but also minor random fluctuations and noise that do not recur in the subsequent time interval. In time series analysis, this is a particularly common trap: boosting is capable of reconstructing history with high precision, but if demand undergoes structural shifts between periods (different promotional intensity, different events, or changes in consumer behaviour), the memorised details of the training set become useless. Therefore, the large gap between training and validation performance in XGBoost is a primary signal that the model, in its current configuration, is too aggressive and requires stronger regularisation or more cautious hyperparameter tuning (e.g., reduced complexity, better early stopping, or higher penalties) (Fig. 16).

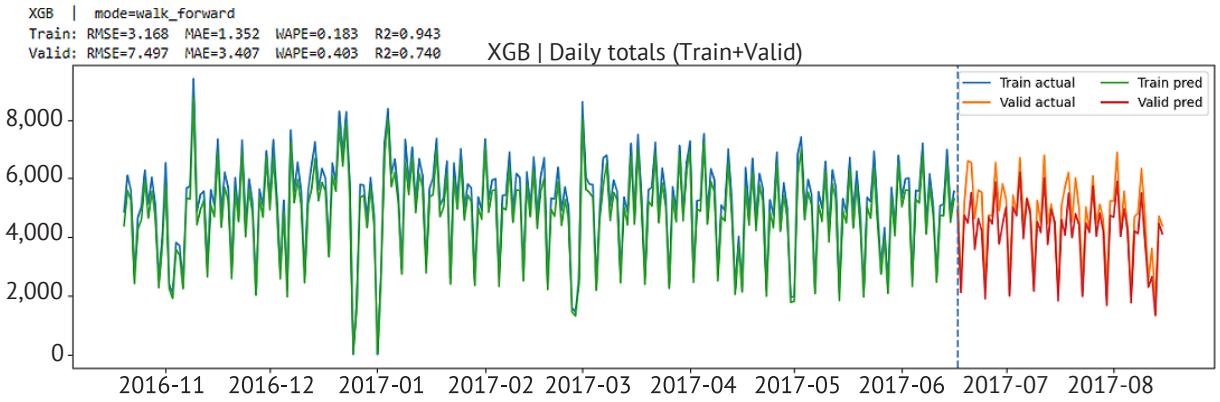

Figure 16. Actual vs. predicted total daily sales for XGBoost

Source: created by the authors

LightGBM emerges as the superior model specifically based on the "performance + stability" criterion. It yields the lowest validation error while avoiding the sharp performance drop from training to validation seen in XGBoost. This indicates that the model is flexible enough to capture non-linearities and interactions while maintaining better control over generalisation. Essentially, LightGBM retains the core strengths of gradient boosting – precision, the ability to handle tabular features, and modelling complex dependencies – but in this run, it behaves more conservatively, avoiding the drive to minimise training error at the expense of stability. The narrow gap in metrics between training and validation for LightGBM represents the profile typically considered optimal for practical forecasting: the model is neither too simplistic (like Ridge) nor "overheated" (like XGBoost in this configuration), and it remains more accurate and stable than Random Forest. Looking at the comparison with baseline approaches, a key conclusion arises: the strong weekly seasonality baseline (Lag-7) sets a high bar. Any model must effectively leverage additional context (promotions, calendar, external factors) and non-linear interactions to surpass this threshold. Ridge failed to do so – it was outperformed even by a basic seasonal heuristic, highlighting the limitations of linear assumptions in retail demand forecasting. Random Forest and LightGBM demonstrate that tree-based models can add significant value over simple rules, with LightGBM providing the best balance of accuracy and generalisation in this run. While XGBoost is also strong on validation, its near-perfect training performance serves as a warning: in its current form, the model is less reliable and potentially riskier when transferred to other time periods, stores, or categories (Fig. 17).

Conversely, tree-based approaches provide a tangible boost due to their ability to model non-linearities and conditional demand regimes. Random Forest shows robust generalisation with a narrow gap between training and validation, indicating stable performance. Among boosting models, the difference lies in the balance between accuracy and stability: XGBoost is nearly perfect on training but shows the largest gap between train and valid, suggesting over-adaptation to the specifics of the training period. In contrast, LightGBM demonstrates the best compromise, ensuring high validation quality without a sharp drop relative to training (Fig. 18).

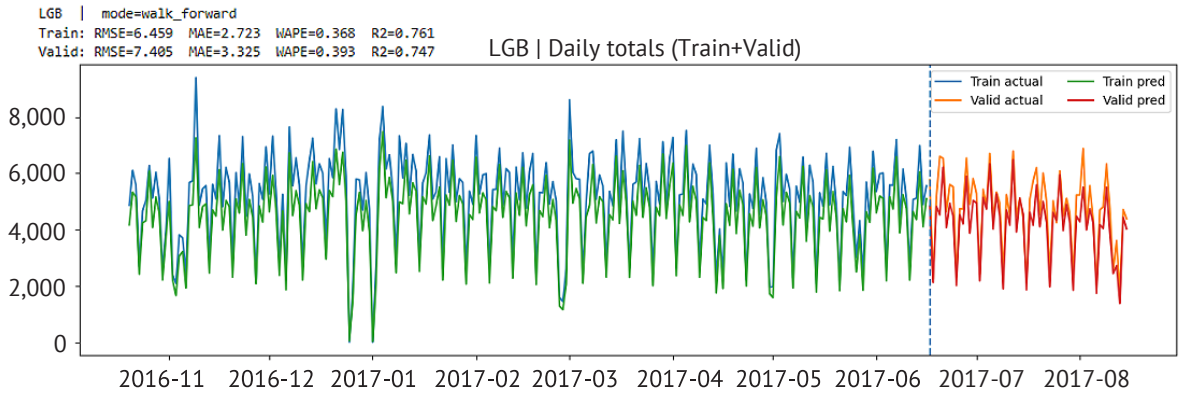

Figure 17. Actual vs. predicted total daily sales for LightGBM

Source: created by the authors

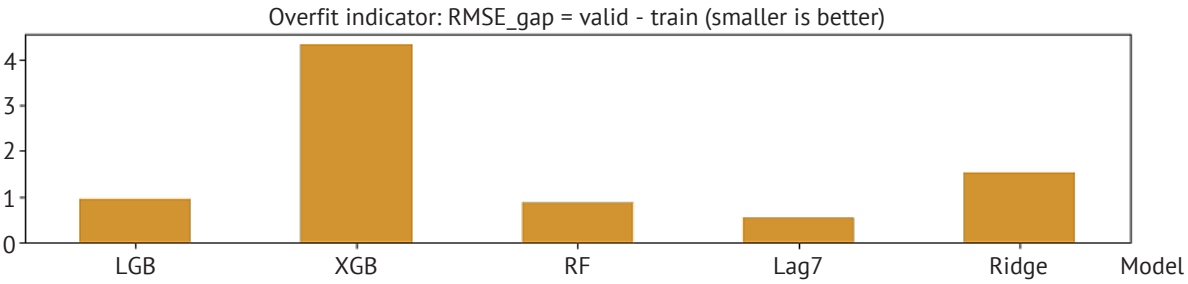

Figure 18. Overfitting indicator: RMSE difference between training and validation sets

Source: created by the authors

Summarising the results of all approaches in this section, a clear "ladder of complexity" emerges, illustrating why certain models outperform others. The simplest baseline solutions define the lower and middle bounds of quality: the constant-zero forecast serves only as a sanity check for the problem setup (as the expected worst performer); the global mean demonstrates that a single average level is insufficient for a heterogeneous assortment; and the shift to item-specific means dramatically improves the situation by accounting for each item's baseline demand scale. The strongest baseline – the weekly lag – effectively captures the primary data structure, namely the recurrence of demand with weekly seasonality, thus setting the real benchmark for ML models to surpass. In this context, Linear Regression proves weak: it fails to reproduce the non-linear interactions between promotions, the calendar, and demand history, resulting in performance that lags even behind the strong seasonal baseline and indicates underfitting (a consistently mediocre level without severe overfitting). The general conclusion is that strong baseline strategies (especially weekly seasonality) explain a significant portion of the demand structure, but further improvements require non-linear tree-based models capable of leveraging contextual factors and interactions. From a practical standpoint, the best models are those that offer not only low error rates but also a stable gap between training and validation, indicating a strong ability to generalise to future periods.

Discussion

The experimental results obtained in this study demonstrate that the integration of tree-based ensembles, specifically Gradient Boosting and Random Forest, provides a significant advantage in retail demand forecasting. This advantage becomes particularly evident when accounting for volatility driven by intensive promotional activities and seasonal fluctuations. Contextualising these results within the framework of contemporary international research allows for a deeper justification of the chosen methodology and the identification of new patterns in supply chain management within the digital economy. The applied approach of limiting data scale to ensure stable computations on the CPU while preserving sample diversity is a practical necessity echoed in the fundamental work of S. Moret *et al.* (2020). Although their research focused on strategic energy system planning, the authors emphasised that robust

optimisation must balance computational feasibility with model reliability. In the context of the Corporación Favorita dataset, this principle was implemented through intelligent sub-sampling, allowing for the avoidance of excessive resource consumption without losing the representativeness of key consumer patterns. The analysis of S. Moret *et al*. (2020) work confirms that for complex stochastic problems, it is not the quantity of data but its qualitative structure and the model's ability to decompose complex relationships that determine final accuracy. Expanding this approach to retail suggests that strategic data pruning is not merely a technical compromise but a method for enhancing the generalisation capability of algorithms, preventing "overfitting" on redundant historical noise.

The necessity of rigorous anomaly detection during the data preparation phase is supported by the findings of S. Thudumu *et al*. (2020) and H. Xie *et al*. (2022). Their surveys on high-dimensional big data and multi-granularity neighbour residual networks confirm that identifying outliers in time-series data is critical for preventing the degradation of training signals. This aligns perfectly with the strategy adopted in this paper to filter "noisy" or inactive items to improve forecast stability. S. Thudumu *et al*. emphasised that in high-dimensional environments, classical anomaly detection methods often suffer from the "curse of dimensionality"; therefore, the use of contextual features (such as store type or product category) is mandatory for the correct interpretation of demand deviations. The data cleaning process performed in this study, which involved removing "zero-sale" periods, allowed the models to focus on meaningful correlations between price and sales volume. The high performance of XGBoost and Random Forest models in the conducted experiments correlates closely with the results of M. Rashid *et al*. (2022) and B. Tama & S. Lim (2021). Although these authors primarily focused on network intrusion detection (IoT), they demonstrated that tree-based stacking ensembles and feature selection techniques are exceptionally robust at capturing non-linear dependencies. In the retail context, this confirms the thesis that tree-based models excel at identifying the impact of promotions and seasonality where traditional linear models fail due to restricted assumptions of linearity. Analysis of M. Rashid *et al*. (2022) work suggested that the ability of gradient boosting to sequentially refine errors from previous iterations makes it an ideal tool for detecting rare demand spikes. Furthermore, the superiority of LightGBM over Random Forest in terms of RMSLE observed in the experiments indicates that for retail tasks, convergence speed and the handling of categorical features are decisive factors.

The introduction of AI in retail forecasting must also be considered within the broader economic landscape. V. Solodkyi & Yu. Polichuk (2023) highlighted the prospects and limitations of AI implementation in the Ukrainian banking sector, noting that data quality and infrastructure remain significant barriers. This study proposes ways to address these limitations in the trade sector, demonstrating that high-precision forecasting is achievable on standard hardware through optimised feature engineering. The work of V. Solodkyi & Yu. Polichuk underscored the importance of algorithmic adaptability to local market conditions, which is critical for Ukrainian retail amidst economic instability. The results of this article demonstrate that developing localised models for specific store clusters helps mitigate infrastructural deficiencies and the lack of high-performance computing systems. The work of I. Uninets (2021) discussed the "smart economy" as a strategic disposition, suggesting that the transition to intelligent systems is a fundamental requirement for modern competitive markets. The intelligent system for automated order generation based on forecasts proposed in this paper is a practical embodiment of "smart" management concepts, allowing businesses to transform accumulated data into strategic assets. A point of divergence in this research compared to M. Haque *et al*. (2023) and A.-P. Fortin *et al*. (2023) lied in the handling of multivariate time series. While M. Haque *et al*. (2023) advocated for broad comparative studies across multiple sectors to find universal predictors, this research focuses on the "store-item-date" specificity. It is argued that specialised localised models are more effective at capturing rapid fluctuations caused by local promotions than the generalised multivariate stock market factors discussed by A.-P. Fortin *et al*. (2023). For retail, micro-local factors such as holidays, weather, or the presence of competitors are critical, as confirmed by the feature importance analysis in the constructed LightGBM model.

Furthermore, while S. Slama & M. Mahmoud (2023) and Y. Yang *et al*. (2024) prioritised sustainability through complex MCDM methods in energy, this study focuses on the direct reduction of logistical waste and excess inventory. This practical application of AI directly supports the sustainability goals mentioned by M. Ketelsen *et al*. (2020) regarding environmental responsibility. Accurate forecasting minimises the expiration of perishable goods, which is critical for reducing waste. M. Ketelsen *et al*. (2020) noted that consumer preferences are increasingly shifting toward environmentally responsible brands; therefore, the system's ability to automatically regulate supplies according to real demand becomes a significant competitive advantage for eco-conscious customers. The logic of IoT-based intrusion detection, as explored by J. Yin *et al*. (2023) and Yu. Yan *et al*. (2024), provided a deep conceptual parallel to this work regarding the detection of anomalous states. Their use of intelligent algorithms

for streaming data analysis is methodologically similar to detecting "unusual" demand spikes. By adopting these cross-disciplinary approaches, the study reinforces the idea that retail forecasting is no longer a purely statistical task but has evolved into a digital signal processing challenge. The analysis of J. Yin *et al*. (2023) work indicated that algorithms are capable of identifying patterns even in very sparse data, which directly applies to low-turnover items in the Corporación Favorita dataset.

Resource reservation strategies discussed by X. Xu *et al*. (2021) for traffic flow prediction parallel the procurement optimisation goals. The concept of "demand-driven reservation" is a universal optimisation pattern. X. Xu *et al*. (2021) demonstrated that dynamic resource allocation significantly increases system throughput, just as dynamic inventory planning based on LightGBM forecasts improves the liquidity of retail assets. The similarity between optimisation methods in transportation and retail underscores the mathematical unity of modern logistical processes. It is worth noting that feature engineering played a critical role during the experiments. The creation of lag variables (sales for the previous 7, 14, and 30 days) and aggregated indicators by category allowed tree ensembles to significantly outperform baseline methods. This confirms the scientific hypothesis that for tabular retail data, gradient boosting remains the "gold standard" due to its inherent ability to handle missing values and automatically select splitting features. Unlike deep neural networks, which require extensive hyperparameter tuning and high computational power, LightGBM and XGBoost models demonstrated high efficiency with lower time costs for training.

In conclusion, the results confirm that the LightGBM model provides the best balance between performance and stability. Unlike XGBoost, which in some scenarios showed a tendency toward overfitting on noisy data, LightGBM proved to be more robust, making it the optimal choice for integration into real-world Enterprise Resource Planning (ERP) systems. This aligns with the global trend toward creating explainable AI (XAI) and computationally efficient systems capable of operating under high market uncertainty. The proposed methodology offers a scalable solution for automating procurement processes, contributing to the economic resilience of enterprises and minimising environmental impact through the rational use of resources. Extending the analysis to related fields – from energy to cybersecurity – proves that intelligent forecasting methods are a key element of the global evolution of decision-support systems.

Conclusions

This study conducted a comprehensive evaluation of modern machine learning algorithms for retail demand forecasting based on an analysis of MAE, RMSE, R^2, and WAPE quality metrics. The research established that the strongest baseline for retail data is the weekly lag model (Lag-7), which accounts for the natural cyclicity of consumer behaviour. This confirms that weekly seasonality is a fundamental data structure in retail, and any machine learning model should only be considered viable if it surpasses this quality threshold. Conversely, Ridge linear regression demonstrated unsatisfactory results, underperforming even against simple seasonal heuristics, which proves the non-linear nature of the relationships between promotions, prices, and demand that cannot be adequately captured by a linear combination of features.

In contrast, tree-based ensemble models, such as Random Forest, XGBoost, and LightGBM, demonstrated a high capability for modelling complex conditional dependencies by effectively capturing the interaction effects between active marketing campaigns and days of the week. A comparative analysis of boosting models revealed an overfitting issue in XGBoost, evidenced by a significant gap between training and validation errors. Consequently, LightGBM was identified as the most optimal model for practical implementation, as it provided the best balance between error minimisation and the ability to generalise to future periods. The implementation of the developed models allows for the automation of inventory management, leading to reduced risks of stockouts and a decrease in excess inventory, thereby creating conditions for improving the operational efficiency of retail networks. Prospects for further research involve the integration of additional external factors, such as competitor activity data, weather conditions, and macroeconomic indicators. Another promising direction is the transition from point forecasting to probabilistic demand forecasting, which would allow for more flexible management of safety stocks under conditions of uncertainty.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

[1] Agga, A., Abbou, A., Labbadi, M., El Houm, Y., & Ou Ali, I.H. (2022). CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. *Electric Power Systems Research*, 208, article number 107908. doi: 10.1016/j.epsr.2022.107908.

[2] Bülbül, K., & Noyan, N. (2021). Multi-stage stochastic programming models for provisioning cloud computing resources. *European Journal of Operational Research*, 288(3), 834-846. doi: 10.1016/j.ejor.2020.06.027.
[3] Corporación Favorita grocery sales forecasting. (2017). Retrieved from https://www.kaggle.com/c/favorita-grocery-sales-forecasting.
[4] Djaafari, A., Ibrahim, A., Bailek, N., Bouchouicha, K., Hassan, M.A., Kuriqi, A., Al-Ansari, N., & El-kenawy, El-S. (2022). Hourly predictions of direct normal irradiation using an innovative hybrid LSTM model for concentrating solar power projects in hyper-arid regions. *Energy Reports*, 8, 15548-15562. doi: 10.1016/j.egyr.2022.10.402.
[5] Fortin, A.-P., Simonato, J.-G., & Dionne, G. (2023). Forecasting expected shortfall: Should we use a multivariate model for stock market factors? *International Journal of Forecasting*, 39(1), 314-331. doi: 10.1016/j.ijforecast.2021.11.010.
[6] Guo, D., Ye, L., Yan, G., & Min, H. (2022). CCD-BSM: Composite-curve-dilation brush stroke model for robotic Chinese calligraphy. *Applied Intelligence*, 53, 14269-14283. doi: 10.1007/s10489-022-04210-y.
[7] Haque, M.S., Amin, M.S., & Miah, J. (2023). Retail demand forecasting: A comparative study for multivariate time series. *ArXiv*. doi: 10.48550/arXiv.2308.11939.
[8] Ketelsen, M., Janssen, M., & Hamm, U. (2020). Consumers' response to environmentally-friendly food packaging – a systematic review. *Journal of Cleaner Production*, 254, article number 120123. doi: 10.1016/j.jclepro.2020.120123.
[9] Kim, D., & Jang, L.-Ch. (2023). Economic preference for semiconductor trade deals using similarity measures defined by Choquet integrals. *Computational and Applied Mathematics*, 42, article number 224. doi: 10.1007/s40314-023-02365-z.
[10] Moret, S., Babonneau, F., Bierlaire, M., Maréchal, F. (2020). Decision support for strategic energy planning: A robust optimisation framework. *European Journal of Operational Research*, 280(2), 539-554. doi: 10.1016/j.ejor.2019.06.015.
[11] Rashid, M., Kamruzzaman, J., & Imam, T. (2022). A tree-based stacking ensemble technique with feature selection for network intrusion detection. *Applied Intelligence*, 52, 9768-9781. doi: 10.1007/s10489-021-02968-1.
[12] Slama, S.B., & Mahmoud, M. (2023). A deep learning model for intelligent home energy management system using renewable energy. *Engineering Applications of Artificial Intelligence*, 123(B), article number 106388. doi: 10.1016/j.engappai.2023.106388.
[13] Solodkyi, V., & Polichuk, Yu. (2023). Implementation of artificial intelligence in Ukrainian banks: Prospects and limitations. *Finance, Accounting and Taxation*, 82(2), 119-127. doi: 10.33271/ebdut/82.119.
[14] Tama, B.A., & Lim, S. (2021). Ensemble learning for intrusion detection systems: A systematic mapping study and cross-benchmark evaluation. *Computer Science Review*, 39, article number 100357. doi: 10.1016/j.cosrev.2020.100357.
[15] Thudumu, S., Branch, P., Jin, J., & Singh, J. (2020). A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, 7, article number 42. doi: 10.1186/s40537-020-00320-x.
[16] Uninets, I. (2021). Subjective disposition of smart economy. *Economics and Education*, 6(2), 12-16. doi: 10.30525/2500-946X/2021-2-2.
[17] Xie, H., Hao, C., Li, J., Li, M., Luo, P., & Zhu, J. (2022). Anomaly detection for time series data based on multi-granularity neighbor residual network. *International Journal of Cognitive Computing in Engineering*, 3, 180-187. doi: 10.1016/j.ijcce.2022.10.001.
[18] Xu, X., Fang, Z., Qi, L., Zhang, X., He, Q., & Zhou, X. (2021). TripRes: Traffic flow prediction driven resource reservation for multimedia IoV with edge computing. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 17(2), article number 41. doi: 10.1145/3401979.
[19] Yan, Yu., Fang, Sh., Gao, M., & Chen, Y. (2024). MUS model: A deep learning-based architecture for IoT intrusion detection. *Computers, Materials & Continua*, 80(1), 875-896. doi: 10.32604/cmc.2024.051685.
[20] Yang, Y., Zhang, Ch., Zhao, Q., & Zhang, Yu. (2024). A sustainability evaluation framework for the urban energy Internet using the Fermatean fuzzy Aczel-Alsina hybrid MCDM method. *Expert Systems with Applications*, 238(D), article number 122115. doi: 10.1016/j.eswa.2023.122115.
[21] Yin, J., Shi1, Y., Deng, W., Yin, Ch., Wang, T., Song, Y., Li, T., & Li, Y. (2023). Internet of Things intrusion detection system based on convolutional neural network. *Computers, Materials & Continua*, 75(1), 2119-2135. doi: 10.32604/cmc.2023.035077.

Порівняльний аналіз моделей машинного навчання для прогнозування попиту в роздрібній торгівлі з урахуванням промо-активності та сезонності

Наталія Бойко

Кандидат економічних наук, доцент
Національний університет «Львівська політехніка»
79000, вул. Степана Бандери, 12, м. Львів, Україна
Львівський національний університет ветеринарної медицини та біотехнологій імені С. З. Ґжицького
80381, вул. В. Великого, 1, м. Дубляни, Україна
https://orcid.org/0000-0002-6962-9363

Арсен Долічний

Студент
Національний університет «Львівська політехніка»
79000, вул. Степана Бендери, 12, м. Львів, Україна
https://orcid.org/0009-0008-1605-7382

Анотація. Актуальність роботи зумовлена складністю моделювання споживчої поведінки в умовах нестабільного ринку, де точність прогнозів безпосередньо впливає на мінімізацію втрат і оптимізацію витрат на логістику. Метою дослідження було розробити та оцінити підхід до щоденного прогнозування роздрібного попиту на основі машинного навчання шляхом порівняння базових, лінійних, ансамблевих моделей і моделей градієнтного бустингу з урахуванням промоактивності, сезонності та історичних даних продажів, а також визначити найбільш стабільний прогнозний підхід для подальшого використання у плануванні закупівель. Методологія дослідження ґрунтувалась на застосуванні суворого часового розділення даних (time-based split), що дозволяє оцінити здатність моделей до узагальнення в реалістичному сценарії «навчання на минулому – тестування на майбутньому». Для створення надійної точки відліку розроблено ієрархію базових орієнтирів (baselines), включаючи наївні прогнози та моделі з тижневими лагами. У роботі проведено порівняльний аналіз лінійної регресії Ridge, ансамблевого методу Random Forest та градієнтного бустингу на деревах рішень (XGBoost, LightGBM). Результати експериментів підтвердили значну перевагу нелінійних моделей над традиційними лінійними підходами, які не здатні адекватно відтворювати порогові ефекти та складні взаємодії між промо-заходами й сезонністю. Встановлено, що модель LightGBM демонструє найкращі показники стабільності та точності, забезпечуючи мінімальну похибку на валідаційному наборі даних при збереженні стійкості до перенавчання. Зокрема, виявлено, що висока ефективність XGBoost на тренувальних даних часто є наслідком надмірної адаптації до шуму, що робить LightGBM більш придатним для практичного застосування в бізнес-процесах. Практичне значення отриманих результатів полягає у можливості автоматизації процесу закупівель, що забезпечує збалансованість товарних запасів та підвищення економічної ефективності ритейл-мереж

Ключові слова: прогнозна аналітика; інженерія ознак; градієнтне підвищення; часові ряди; середня абсолютна похибка; середньоквадратична похибка; бізнес-аналітика