Jurnal Teknik Industri, Vol. 28, No. 1, June 2026
ISSN 1411-2485 print / ISSN 2087-7439 online
DOI: https://doi.org/10.9744/jti.28.1.1-12

Optimization of XGBoost Hyperparameters Using Three-Dimensional Learning AVOA for Retail Demand Prediction

Alza Noor Ibrahim, Dhea Qurrotun Nada, Rudi Nurdiansyah*, Andoko
Mechanical and Industrial Engineering Department, Universitas Negeri Malang,
Jl. Semarang No. 5, Malang 65145, Indonesia
Email: alza.noor.2305166@students.um.ac.id, dhea.qurrotun.2305166@students.um.ac.id,
rudi.nurdiansyah.ft@um.ac.id*, andoko.ft@um.ac.id
*Corresponding author

Abstract: Accurate demand forecasting is critical for retail supply chains, particularly in the Fast-Moving Consumer Goods (FMCG) sector, where even small discrepancies between predicted and actual demand can lead to excess inventory or stock shortages. This study proposes a hybrid TDLAVOA–XGBoost model that adaptively optimizes key hyperparameters to improve forecasting accuracy and stability. The analysis is conducted using 990 FMCG inventory records from a publicly available dataset to examine the impact of metaheuristic-based optimization on model performance. The TDLAVOA algorithm identifies an effective hyperparameter configuration (*max_depth* = 3, *learning_rate* = 0.01, *n_estimators* = 100, *gamma* = 1.97, *subsample* = 0.57, and *colsample_bytree* = 0.66), enabling the proposed model to achieve an RMSE of 22.53 ± 0.50 and an MAE of 19.32 ± 0.33. Compared with the default XGBoost baseline, this represents a substantial reduction in prediction error and variability. Comparative results show that TDLAVOA–XGBoost achieves performance comparable to SARIMAX and demonstrates superior accuracy relative to deep learning models, including LSTM and MLP, for limited-sample tabular FMCG demand data. Statistical validation using one-way ANOVA and Tukey's HSD confirms that the performance differences among models are statistically significant ($p < 0.0001$). Overall, the findings indicate that TDLAVOA–XGBoost provides a practical and reliable approach for supporting data-driven inventory planning in retail environments.

Keywords: XGBoost, TDLAVOA, optimization, forecasting, retail.

Introduction

The rapid growth of the modern retail industry has created a strong need for demand forecasting systems that can respond to constantly changing market conditions. When forecasts deviate from actual sales, retailers face familiar problems: excess inventory that increases operational costs and stock shortages that reduce customer satisfaction and lead to lost sales [1, 2]. These issues are particularly pronounced in the Fast-Moving Consumer Goods (FMCG) sector, where demand patterns shift quickly due to promotions, seasonal effects, and evolving consumer preferences [3, 4]. In such a dynamic environment, forecasting methods capable of capturing nonlinear and unstable demand patterns are essential for maintaining efficient inventory operations.

Extreme Gradient Boosting (XGBoost) is widely used for nonlinear prediction tasks because it constructs decision trees sequentially and reduces errors at each stage of the boosting process [4, 5]. The algorithm incorporates several mechanisms, including regularization, column subsampling, and shrinkage, which help control overfitting and improve computational efficiency on large datasets [6]. However, its predictive performance still depends heavily on hyperparameter configuration, as these parameters determine tree-depth, learning behavior, and the model's ability to generalize to new data.

The main challenge in using XGBoost lies in its strong sensitivity to hyperparameter configuration [7]. Parameters such as *learning_rate*, *max_depth*, *n_estimators*, and *gamma* directly influence how well the model generalizes to new data [8]. The *learning_rate* controls the magnitude of weight updates at each boosting iteration, allowing the model to learn at a controlled pace. The *max_depth* parameter defines the depth of each decision tree and therefore affects model complexity and the risk of overfitting. The *n_estimators* parameter determines the number of trees constructed sequentially to reduce residual errors, while *gamma* acts as a regularization term that specifies the minimum loss reduction required for a split. This ensures that new branches are created only when they contribute meaningfully to predictive accuracy. When these parameters are poorly configured, the model may overfit or underfit, resulting in substantial performance degradation [6, 9]. This challenge is further complicated by the wide and nonlinear hyperparameter search space, which makes the optimization process computationally demanding. Traditional approaches such as grid search and random search often become inefficient in this context because they explore the search space in a limited manner and require significant computational effort [10]. Similar challenges have been reported in other optimization problems, where parameter tuning plays a critical role in convergence performance, such as in metaheuristic-based optimization for flow shop scheduling [11].

To overcome these limitations, various metaheuristic optimization algorithms have been applied to tune XGBoost hyperparameters, as metaheuristics consistently demonstrate strong performance in solving complex optimization problems across engineering domains. Methods such as the Genetic Algorithm (GA), Grey Wolf Optimizer (GWO), and Harris Hawks Optimizer (HHO) have reported notable improvements in model accuracy and training efficiency [8, 13]. However, many of these algorithms still struggle to maintain a stable balance between global exploration and local exploitation, which can lead to premature convergence before reaching the global optimum [3, 10]. Reference [9] introduced a hybrid Three-Dimensional Learning African Vulture Optimization Algorithm (TDLAVOA)–XGBoost model for predicting cutting tool wear in turning and milling operations and showed that this approach outperformed GA–XGBoost, offering improvements in both accuracy and stability. TDLAVOA integrates three core learning mechanisms that enhance population adaptability throughout the search process [4, 14]. Nevertheless, the previous study did not fully examine the systematic optimization of XGBoost hyperparameters within this framework.

To address this gap, the present study applies the TDLAVOA–XGBoost model to FMCG inventory demand forecasting and focuses on optimizing its hyperparameters to improve prediction accuracy and stability. Three categories of benchmark models are considered: metaheuristic-based models, deep learning models, and statistical models. In the metaheuristic group, GA–XGBoost, GWO–XGBoost, and HHO–XGBoost are selected because they have demonstrated reliable hyperparameter optimization performance in various predictive settings. For example, GA–XGBoost is used in [14] for sensor signal quality prediction and achieves an 8% improvement over standard XGBoost. GWO–XGBoost, applied in atmospheric data analysis [15], improves prediction stability, while HHO–XGBoost, used for FMCG supply chain inventory allocation forecasting[16] delivers competitive results compared with other evolutionary approaches. In the deep learning category, Long Short-Term Memory (LSTM) and Multi-Layer Perceptron (MLP) models are included because of their widespread use in time-series prediction and demand forecasting. LSTM, as demonstrated in [17] for electrical power engineering data, effectively captures complex temporal dependencies, whereas MLP, used in [18] for nonlinear data classification, shows strong generalization capability. The Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX) model is used as the statistical baseline because it is widely adopted in demand forecasting applications, including food supply chain modeling [19].

This study aims to improve the accuracy of retail product demand forecasting using FMCG datasets by adaptively optimizing XGBoost hyperparameters through TDLAVOA. The findings are expected to contribute to the development of more efficient, stable, and accurate forecasting models that support data-driven inventory planning in the retail sector.

Methods

Research Workflow

This study follows a series of systematic stages to develop an accurate and stable retail demand prediction model based on inventory data. The process begins with a literature review and data collection to understand the characteristics of Fast-Moving Consumer Goods (FMCG) demand and the factors influencing its fluctuations. The next stage involves data preprocessing, including data cleaning, normalization, and splitting the dataset into training and testing subsets to ensure suitability for model development. After data preparation, the study applies a hybrid Three-Dimensional Learning African Vulture Optimization Algorithm–Extreme Gradient Boosting (TDLAVOA–XGBoost) model to adaptively optimize selected XGBoost hyperparameters. The performance of the optimized model is then compared with several benchmark algorithms GA–XGBoost, GWO–XGBoost, HHO–XGBoost, SARIMAX, LSTM, and MLP to evaluate the accuracy and stability of the forecasting results.

All programming and experiments are conducted in Visual Studio Code (VS Code) using Python 3.10. The implementation relies on several libraries, including Scikit-learn, XGBoost, NumPy, Pandas, and Matplotlib, for model development, analysis, and visualization. Initial data processing is performed in Microsoft Excel, while advanced statistical analyses such as ANOVA and Tukey's HSD tests together with performance visualizations, are conducted using OriginPro 2024. To ensure reproducibility, all experiments use a fixed random seed (seed = 42) for data splitting, model initialization, and metaheuristic population generation.

Model performance is evaluated using two key metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), which provide quantitative measures of prediction accuracy. Overall, the research workflow includes data processing, model implementation, hyperparameter optimization, performance evaluation, and statistical validation, ensuring that the results remain robust and reliable.

XGBoost Architecture

Extreme Gradient Boosting (XGBoost) is a decision tree–based ensemble learning method that constructs multiple trees sequentially, where each new tree aims to correct the errors of the previous one [5]. The algorithm is designed to improve computational efficiency and reduce overfitting through the use of regularization terms and constraints on tree complexity [20]. A general overview of the XGBoost architecture is presented in Figure 1. The process begins by passing input data through a sequence of decision trees built one after another. Each subsequent tree is trained to minimize the residual error from earlier trees, enabling the model to capture complex relationships among variables while reducing prediction bias.

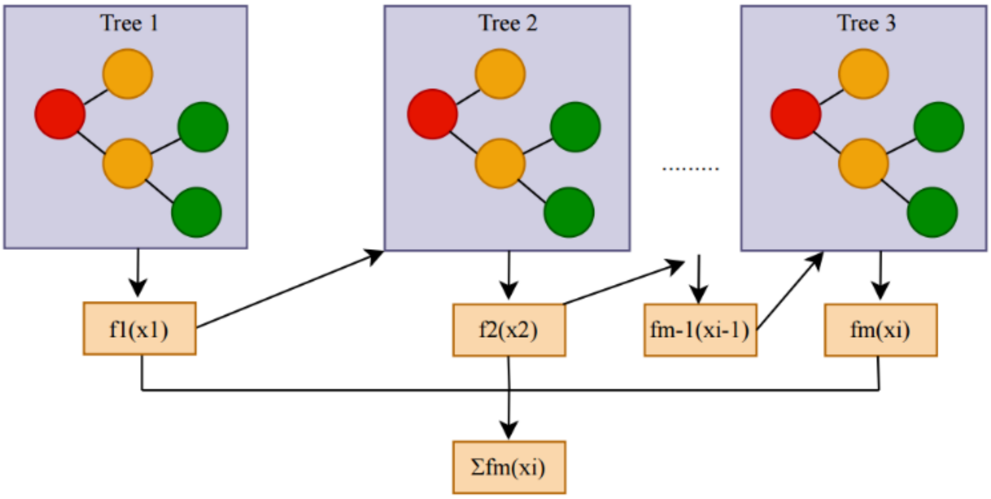

Figure 1. XGBoost model architecture

The XGBoost objective function consists of two components: the training loss and a regularization term, as shown in Equation (1). In this expression, l represents the loss function, while $\Omega(f_t)$ controls the complexity.

$$Obj(\theta) = \sum_{i-1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{1}$$

The regularization term is defined in Equation (2), where T denotes the number of leaves in the decision tree, λ is the regularization coefficient, and w_j represents the weight of the j-th leaf. This term helps balance accuracy and model complexity, promoting more stable learning.

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \textstyle\sum_{j-1}^{T} w_j^2 \tag{2}$$

The optimal weight for each leaf is obtained by minimizing the objective function, as expressed in Equation (3). In Equation (3), g_i and h_i denote the first and second derivatives of the loss function with respect to the previous prediction, and I_j represents the set of data points assigned to leaf j:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{3}$$

$$\hat{y}_i = \textstyle\sum_{t-1}^{T} f_t(X_i) \tag{4}$$

The final model prediction is generated by summing the outputs of all decision trees constructed during the boosting process, as described in Equation (4).

Beyond regularization, XGBoost employs several additional strategies to stabilize training. These include shrinkage through the learning rate parameter η, as well as row and column subsampling (*subsample* and *colsample_bytree*), both of which help reduce overfitting and limit the model's dependence on specific subsets of data [6]. In this study, six hyperparameters are optimized: *max_depth*, *learning_rate*, *n_estimators*, *gamma*, *subsample*, and *colsample_bytree*. Together, these parameters determine tree depth, learning rate, the number of boosting rounds, the minimum loss reduction required for splitting, and the proportion of rows and columns sampled in each iteration. Identifying an effective combination of these parameters is essential for balancing bias, variance, and computational cost, making metaheuristic optimization a suitable approach for this task [21].

Metaheuristic Algorithm

Metaheuristic algorithms are population-based global optimization methods designed to solve complex nonlinear problems without requiring derivatives of the objective function [22]. They are widely used as alternatives to conventional approaches such as gradient descent or grid search, which may become trapped in local optima and require substantial computational effort. Two key mechanisms underpin metaheuristic optimization: exploration and exploitation. Exploration allows the algorithm to search the solution space broadly and avoid local optima, whereas exploitation focuses on refining promising regions to improve accuracy and accelerate convergence [23]. Maintaining an effective balance between these two mechanisms is essential for achieving robust optimization performance.

In this study, four metaheuristic algorithms are employed: the Genetic Algorithm (GA), Grey Wolf Optimizer (GWO), Harris Hawks Optimizer (HHO), and the Three-Dimensional Learning African Vulture Optimization Algorithm (TDLAVOA). These algorithms are selected because they offer a reasonable balance between convergence speed and adaptability to variations in the objective function. GA is an evolutionary method that mimics natural selection through operations such as selection, crossover, and mutation. These processes generate diverse candidate solutions, enabling GA to explore wide and non-convex search spaces effectively [14]. GWO, on the other hand, models the social hierarchy and cooperative hunting strategy of gray wolves. Its leadership structure, which consists of alpha, beta, and delta wolves, guides the rest of the population toward the search target and enables the algorithm to maintain a stable balance between exploration and exploitation throughout the optimization process [24]. Harris Hawks Optimizer (HHO) is inspired by the cooperative hunting strategy of Harris's hawks. The algorithm employs adaptive mechanisms, including soft besiege and hard besiege phases, which adjust the intensity of exploitation based on the hawks' proximity to the prey. This adaptability helps HHO avoid local optima and achieve faster convergence [16]. In contrast, TDLAVOA is an enhanced version of the African Vulture Optimization Algorithm (AVOA) that incorporates two additional learning mechanisms: Three-Dimensional Learning (TDL) and Reverse Elite Learning (REL). TDL expands exploration by enabling dynamic three-dimensional interactions among individuals, while REL preserves population diversity by maintaining elite solutions and reducing the likelihood of premature convergence. Together, these mechanisms provide TDLAVOA with strong adaptive capabilities in balancing exploration and exploitation [9, 26].

Three-Dimensional Learning African Vulture Optimization Algorithm (TDLAVOA)

The Three-Dimensional Learning African Vulture Optimization Algorithm (TDLAVOA) is developed based on the adaptive behavior of African vultures as they search for prey and adjust their flight strategies according to their energy levels and distance from the target [3]. The algorithm operates through five main stages: population initialization, position updating, three-dimensional learning, reverse elite learning, and termination criteria [9].

$$\boldsymbol{X_i = X_{min} + rand(0,1)} \times (\boldsymbol{X_{max} - X_{min}}) \tag{5}$$

In the first stage, population initialization, several individuals—referred to as vultures—are generated randomly within the parameter search space. Each individual $\boldsymbol{X_i}$represents one candidate combination of XGBoost hyperparameters, and its initial value is computed using Equation (5). The fitness value for each individual is then calculated based on its performance when evaluated within the XGBoost model.

The second stage, position updating, adjusts each individual's position according to its energy value $\boldsymbol{E}$, which is determined using Equation (6). This energy value dictates the search phase: when $|\boldsymbol{E}| \geq \mathbf{1}$, the algorithm enters the global exploration phase Equation (7); when $|\boldsymbol{E}| < \mathbf{1}$, it switches to the local exploitation phase Equation (8). During exploration, individuals update their positions using the average population position $\boldsymbol{X}_{\mathrm{mean}}$ and the current best solution $\boldsymbol{X}_{\mathrm{best}}$. In contrast, during exploitation, new candidate positions are generated using two randomly selected individuals $\boldsymbol{X_j}$ and $\boldsymbol{X_k}$, allowing the search to refine promising areas more precisely.

$$\boldsymbol{E} = \mathbf{2}\boldsymbol{C} \times \boldsymbol{rand} - \boldsymbol{C} \tag{6}$$

$$\boldsymbol{X_i^{t+1}} = \boldsymbol{X_{best}^t} \times (\mathbf{1} - \boldsymbol{E}) + \boldsymbol{rand}(\mathbf{0},\mathbf{1}) \times (\boldsymbol{X_{mean}} - \boldsymbol{X_i^t}) \tag{7}$$

$$\boldsymbol{X_i^{t+1}} = \boldsymbol{X_{best}^t} + \boldsymbol{E} \times (\boldsymbol{X_j} - \boldsymbol{X_k}) \tag{8}$$

As shown in Figure 2, the interaction between exploration, exploitation, and the three-dimensional learning zone illustrates how TDLAVOA adapts its search behavior over time. Yellow points represent vulture individuals, orange lines indicate exploration movements based on Equation (7), and green lines represent exploitation movements from Equation (8). At the center, the Three-Dimensional Learning Zone integrates the strategies defined in Equations (9)–(11), helping maintain population diversity and stabilize convergence. The green arrow on the right illustrates how the optimized parameters from TDLAVOA are then fed into XGBoost to enhance predictive performance.

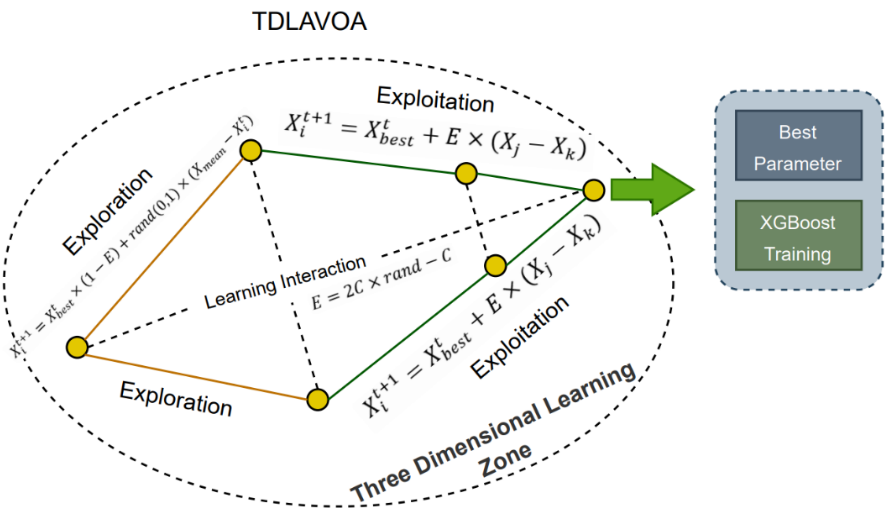

Figure 2. Illustration of the TDLAVOA optimization mechanism

The third stage, Three-Dimensional Learning (TDL), combines three complementary strategies: Experience-Based Learning (EBL), Neighbor-Based Learning (NBL), and Reverse-Based Learning (RBL). These strategies are expressed in Equations (9)–(11). EBL enables individuals to learn from their own best experiences, helping accelerate convergence. NBL incorporates information from neighboring individuals to maintain diversity. RBL uses the reverse position relative to the search space boundaries to help individuals escape from local optima. Together, these strategies create a dynamic three-dimensional learning zone that strengthens exploration while maintaining the overall search direction.

$$\boldsymbol{X_{EBL}^{t+1}} = \boldsymbol{X_i^t} + \boldsymbol{r_1}(\boldsymbol{X_{best}} - \boldsymbol{X_i^t}) \tag{9}$$

$$\boldsymbol{X_{NBL}^{t+1}} = \boldsymbol{X_i^t} + \boldsymbol{r_2}(\boldsymbol{X_{neighbor}} - \boldsymbol{X_i^t}) \tag{10}$$

$$\boldsymbol{X_{RBL}^{t+1}} = \boldsymbol{X_{max}} + \boldsymbol{X_{min}} - \boldsymbol{X_i^t} \tag{11}$$

The fourth stage, Reverse Elite Learning (REL), updates elite individuals, referring to those with the highest fitness values, by generating shadow solutions around the optimal point as defined in Equation (12). The parameter $\boldsymbol{\alpha}$acts as an exploration control factor that adjusts the search intensity near elite solutions.

$$\boldsymbol{X}_{elite}^{t+1} = \boldsymbol{X}_{best}^{t} + \boldsymbol{\alpha} \times (\boldsymbol{X}_{RBL}^{t} - \boldsymbol{X}_{best}^{t}) \tag{12}$$

The final stage, termination criteria, determines when the optimization process ends. The algorithm stops when the maximum number of iterations is reached or when no meaningful improvement occurs in the fitness value. The best solution obtained at the end of the process is then used as the optimal configuration for the XGBoost hyperparameters.

Benchmark Model

A set of comparative models is used to evaluate the performance of the proposed TDLAVOA–XGBoost framework, as summarized in Table 1. The selected models represent diverse forecasting paradigms, including metaheuristic-optimized machine learning models, deep learning architectures, and statistical time-series methods, enabling a comprehensive and balanced comparison. All benchmark models are implemented with clearly defined parameter settings to ensure fairness and transparency in the evaluation process. Through this multi-paradigm benchmarking strategy, the study systematically assesses the accuracy, stability, and robustness of the proposed TDLAVOA–XGBoost model relative to established forecasting approaches.

Table 1. Comparative model

Category	Model	Description	Parameters
ML Ensemble	TDLAVOA-XGBoost	Hybrid model for retail demand forecasting with adaptive hyperparameter optimization.	pop = 20, iter = 30, α = 0.5; mechanisms: AVOA + TDL (EBL, NBL, RBL) + REL
	GA-XGBoost	Used in energy consumption prediction studies, improving hyperparameter search efficiency and reducing prediction error [14].	pop = 20, iter = 30, CR = 0.8, MR = 0.1
	GWO-XGBoost	Applied in energy demand forecasting and industrial scheduling, maintaining stable performance under noisy data [24].	pop = 20, iter = 30, control parameter a ϵ [2 → 0]
	HHO-XGBoost	Utilized for short-term electricity load forecasting, achieving fast convergence on non-convex problems [16].	pop = 20, iter = 30, energy parameter E ϵ [-2, 2]
	XGBoost	Baseline model commonly used to assess the impact of hyperparameter optimization in retail forecasting [26].	Default hyperparameters
Deep Learning	LSTM	Used in consumer behavior and retail demand prediction to capture long-term temporal dependencies [17].	LSTM(50–1), lookback = 5, Adam (lr = 0.001), epochs = 100, batch = 32
	MLP	Applied in electronic product sales and financial forecasting, modeling simple nonlinear relationships [18].	Dense(64–32–1), ReLU, Adam (lr = 0.001), epochs = 100, batch = 32
Time Series	SARIMAX	Used in grocery sales and energy demand forecasting, combining seasonal patterns with external variables [19].	(p,d,q) = (1,1,1); (P,D,Q,s) = (1,1,1,7); enforce_stationarity = False

Evaluation Metrics

To evaluate the predictive performance of each model, this study employs standard evaluation metrics that quantify both accuracy and prediction error. These metrics enable a comprehensive comparison across models. Their definitions and corresponding mathematical formulations are presented below.

Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is defined as the square root of the Mean Square Error (MSE). It expresses prediction error in the same scale as the target variable, RMSE is easier to interpret in practical applications. A lower RMSE value indicates higher predictive accuracy. RMSE is calculated using Equation (13).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{13}$$

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) measures the average magnitude of the absolute differences between predicted and actual values. Similar to RMSE, lower MAE values indicate better model performance. MAE is defined in Equation (14).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (14)$$

Validation Statistics

Statistical validation is conducted to determine whether performance differences among the models are statistically significant. The Shapiro–Wilk test is first used to assess the normality of the data distribution, while Bartlett's test evaluates the homogeneity of variances to verify whether parametric assumptions are satisfied. When both assumptions are held, a one-way Analysis of Variance (ANOVA) is applied to examine whether there are significant differences in average performance across models. If the ANOVA indicates statistical significance, the Tukey Honest Significant Difference (Tukey HSD) post-hoc test is performed to identify which specific model pairs differ. If the assumptions of normality or homogeneity are violated, non-parametric alternatives—such as the Friedman and Nemenyi tests—are employed to ensure a robust comparison. These validation procedures follow standard practices in optimization research, where ANOVA and Tukey HSD are commonly used to evaluate performance differences among metaheuristic algorithms [27].

Results and Discussions

Data Collection and Preprocessing

The data used in this study are obtained from Kaggle.com (https://www.kaggle.com/datasets/a-ndrexibiza/grocery-sales-dataset), which contains sales transactions from the Fast-Moving Consumer Goods (FMCG) sector. This dataset is selected because it reflects the complex dynamics of modern retail operations, including variations in product demand, seasonal fluctuations, and sensitivity to pricing and shelf life [1, 3]. A total of 990 inventory samples were used for the machine learning–based demand prediction experiments.

The preprocessing stage begins with data cleaning to address missing values and outliers. Missing numerical values are imputed using the median, while missing categorical values are filled with the mode to preserve the overall structure of the dataset. Outliers are identified using the Interquartile Range (IQR) method and removed if they fall outside the interval $Q1 - 1.5 \times IQR$ to $Q3 + 1.5 \times IQR$ [28]. Feature engineering is then performed to incorporate temporal variables such as day of the week, month of the year, and lag features, allowing the model to better capture time-dependent demand patterns [14].

Data normalization is applied using the Z-score method to ensure consistent feature scaling and reduce bias during model training. Model evaluation is conducted using a time-series cross-validation approach based on a rolling window scheme, which preserves the chronological order of the data and prevents temporal leakage [29]. Through these preprocessing and validation steps, the dataset remains clean, temporally consistent, and ready for training with XGBoost and its metaheuristic-optimized variants.

Optimization of Hyperparameters

Optimization is performed to identify the optimal hyperparameter configuration for XGBoost using four metaheuristic algorithms: TDLAVOA, GA, GWO, and HHO. The search range for each parameter follows the limits presented in Table 2, which are selected to balance exploration capability and computational efficiency, as suggested in [9].

Table 2. Hyperparameter search limits

Hyperparameters	Definition	Lower bound	Upper bound
max_depth	Maximum depth of the decision trees, controlling model complexity	3	12
learning_rate	Step size that determines how much each new tree influences the model	0.01	0.3
n_estimators	Number of trees constructed sequentially during boosting	50	500
gamma	Minimum loss reduction required to make a further split	0	5
subsample	Proportion of training samples used for each tree	0.5	1
Colsample_bytree	Proportion of features sampled when constructing each tree	0.5	1

The optimization results, summarized in Table 3, indicate that the TDLAVOA–XGBoost model produces the most stable hyperparameter configuration: *max_depth* = 3, *learning_rate* = 0.01, *n_estimators* = 100, *gamma* = 1.97, *subsample* = 0.75, and *colsample_bytree* = 0.77. In comparison, GA–XGBoost selects a higher number of estimators (178) with *gamma* = 2.33, while GWO–XGBoost and HHO–XGBoost generate very large *n_estimators* values (243 and 253, respectively), accompanied by *subsample* and *colsample_bytree* values close to 1.0.

Table 3. Optimized Hyperparameter Values for Metaheuristic–XGBoost Models

Hyperparameters	TDLAVOA_XGB	GA_XGB	GWO_XGB	HHO_XGB
Max_depth	3	4	3	3
Learning_rate	0.01	0.02	0.01	0.01
n_estimator	100	178	243	253
Gamma	1.97	2.33	0.71	2.45
Subsample	0.75	0.92	0.99	0.99
Colsample_bytree	0.77	0.60	0.99	0.75

These results suggest that TDLAVOA identifies a simpler yet highly effective hyperparameter configuration, whereas GA and HHO tend to favor larger parameter values, indicating stronger exploitation of the search space. This pattern highlights the role of the three-dimensional learning mechanism in TDLAVOA, which promotes a balanced interaction between exploration and exploitation during optimization [4, 10]. By integrating experience-based, neighbor-based, and reverse-based learning strategies, TDLAVOA regulates parameter growth while maintaining stable convergence behavior. This capability allows the algorithm to consistently identify parameter settings that generalize well across different data splits. The stability of the selected configurations is consistent with findings reported in [9], where TDLAVOA demonstrates a strong ability to avoid premature convergence while maintaining adaptability compared with other population-based optimization methods.

Comparative Model Performance

Model performance is evaluated using two primary metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The results of 100 independent runs are summarized in Table 4. Among all models, TDLAVOA–XGBoost achieves the strongest overall performance, with an RMSE of 22.53 ± 0.50 and an MAE of 19.32 ± 0.33. Compared with the default XGBoost baseline (RMSE = 26.35 ± 0.81; MAE = 21.94 ± 0.59), this corresponds to an improvement of approximately 14.5% in RMSE and 11.9% in MAE. GA–XGBoost, GWO–XGBoost, and HHO–XGBoost achieve RMSE values of 23.17 ± 0.87, 22.86 ± 0.81, and 23.07 ± 0.86, respectively, with MAE values ranging from 19.79 to 19.99. In terms of computational time, TDLAVOA–XGBoost requires 1419 s, whereas GA–XGBoost, GWO–XGBoost, and HHO–XGBoost complete the optimization in 925 s, 780 s, and 776 s, respectively. As external benchmarks, the LSTM model produces an RMSE of 25.91 ± 1.29 and an MAE of 22.75 ± 1.31, while the SARIMAX model achieves an RMSE of 22.76 ± 0.77 and an MAE of 19.69 ± 0.84.

Table 4. Model performance test results (mean)

Model	RMSE	Std (±)	MAE	Std (±)	Runtime (s)
TDLAVOA-XGBoost	22.53	0.50	19.32	0.33	1419
GA-XGBoost	23.17	0.87	19.99	0.74	925
GWO-XGBoost	22.86	0.81	19.79	0.80	780
HHO-XGBoost	23.07	0.86	19.93	0.82	776
XGBoost	26.35	0.81	21.94	0.59	-
LSTM	25.91	1.29	22.75	1.31	-
MLP	24.59	1.38	20.60	1.13	-
SARIMAX	22.76	0.77	19.69	0.84	-

These results indicate that TDLAVOA–XGBoost delivers highly competitive predictive performance compared with both traditional time-series and machine-learning approaches. This finding is consistent with previous studies showing that metaheuristic-based optimization can enhance the robustness of ensemble learning models in nonlinear forecasting tasks [9]. Relative to the default XGBoost baseline, the proposed model demonstrates a clear reduction in prediction error while maintaining lower variability across repeated runs, indicating improved stability in the hyperparameter search process [15]. Although its predictive accuracy is comparable to that of SARIMAX, TDLAVOA–XGBoost provides greater flexibility in capturing nonlinear relationships inherent in FMCG demand data, which are often difficult to model using linear statistical assumptions. Moreover, compared with deep learning models, particularly LSTM and MLP, the proposed approach yields substantially lower prediction errors. This observation aligns with prior findings that tree-based ensemble models often generalize more effectively than deep neural networks when trained on limited datasets [25].

The strong and consistent performance of TDLAVOA–XGBoost can be attributed to the integration of Three-Dimensional Learning (TDL) and Reverse Elite Learning (REL), which improve the balance between exploration and exploitation during hyperparameter optimization and help mitigate premature convergence [9]. This improved search capability comes with increased computational cost. The runtime of TDLAVOA–XGBoost is approximately 53% higher than that of GA–XGBoost and about 82% higher than that of GWO–XGBoost, reflecting the additional learning stages incorporated into the algorithm. Despite this increase, TDLAVOA demonstrates stable convergence behavior and consistently reaches near-optimal solutions within the first 60–70% of the maximum iteration budget. This suggests that the additional learning mechanisms primarily enhance convergence stability rather than introducing excessive exploratory overhead. From a computational complexity perspective, the overall training cost is governed mainly by population size, iteration budget, and repeated evaluation of the XGBoost model. This represents a practical trade-off between computational effort and solution quality, which is commonly observed in population-based metaheuristic optimization methods [7].

Statistical Validation and Model Ranking

Preliminary assumption tests are conducted to verify the suitability of parametric statistical analysis. The Shapiro–Wilk test indicates that the residuals of all models follow a normal distribution ($p > 0.004167$), and Bartlett's test confirms the homogeneity of variances across groups ($p = 0.530$). With these assumptions satisfied, a one-way ANOVA is performed, yielding an F-statistic of 620.9 with $p < 0.0001$. This result indicates statistically significant differences in RMSE performance among the evaluated models. Figure 3 presents the RMSE distribution for each model along with Tukey HSD group labels. The metaheuristic-based XGBoost models—TDLAVOA–XGBoost, GA–XGBoost, GWO–XGBoost, and HHO–XGBoost—together with SARIMAX, are classified in the top-performing group ("a"). The deep learning models, LSTM and MLP, form the intermediate group ("b"), while the default XGBoost model is assigned to the lowest-performing group ("c"). These results indicate that models incorporating metaheuristic-based optimization or statistical time-series approaches achieve significantly lower RMSE values than the unoptimized XGBoost baseline.

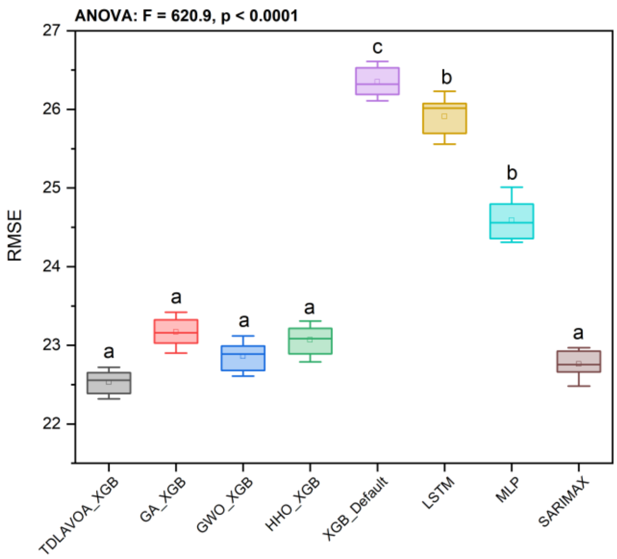

Figure 3. Results of the ANOVA–Tukey test on the RMSE values across models.

The TDLAVOA–XGBoost model achieves the lowest RMSE among all evaluated models, indicating superior predictive accuracy compared with the default XGBoost baseline. In contrast to the higher error and variability observed in the unoptimized XGBoost model, the reduced RMSE and standard deviation achieved by TDLAVOA–XGBoost highlight the effectiveness of incorporating advanced learning strategies into the optimization process. This improvement suggests that the integration of three-dimensional learning and reverse elite learning mechanisms enhances the algorithm's ability to explore the hyperparameter search space while maintaining convergence stability. Similar findings are reported in [27], where metaheuristic-based models evaluated using ANOVA and Tukey HSD demonstrate stronger robustness and consistency than conventional tuning approaches. The consistency of these results with previous studies [15] and [7] further supports the conclusion that combining population-based optimization algorithms with XGBoost improves convergence behavior and strengthens generalization capability in nonlinear forecasting tasks. Moreover, the relatively low dispersion of RMSE values achieved by TDLAVOA–XGBoost reflects a well-balanced exploration and exploitation process, which is a fundamental principle of modern evolutionary optimization methods.

Conclusions

This study demonstrates that integrating the Three-Dimensional Learning African Vulture Optimization Algorithm (TDLAVOA) with XGBoost improves the accuracy and stability of retail demand forecasting through adaptive hyperparameter optimization. The experimental results show that TDLAVOA–XGBoost achieves highly competitive predictive performance across all evaluated models, with RMSE and MAE values of 22.53 and 19.32, respectively, and the lowest performance variability across repeated runs. Statistical validation using one-way ANOVA and Tukey's HSD confirms that these improvements are statistically significant ($p < 0.0001$), placing the proposed model among the top-performing approaches. The observed performance gains highlight the importance of TDLAVOA's balanced exploration and exploitation mechanism, particularly the three-dimensional learning and reverse elite learning components, in producing robust and stable hyperparameter configurations. Although the enhanced optimization strategy incurs higher computational cost compared with simpler metaheuristic approaches, this trade-off is justified in offline decision-support settings where forecasting accuracy and stability are prioritized. Given limitations related to the relatively small dataset size (990 samples), the focus on a single FMCG domain, and potential constraints in capturing complex temporal dependencies, the findings suggest that TDLAVOA–XGBoost is especially effective for small-scale, tabular retail demand forecasting tasks. Future work may extend the proposed framework to larger and more diverse datasets, integrate it into real-time decision-support systems, and explore hybridization with deep learning architectures to further improve scalability and predictive performance in modern retail and supply chain analytics.

Acknowledgment

This research was funded by the Institute for Research and Community Service (LPPM) of University State of Malang in 2025 through a Student Innovation Grant with Contract Number 24.2.877/UN32.14.1/LT/2025. We are grateful for the support that enabled this research to run smoothly.

AI Declaration

The authors declare that AI-assisted tools (ChatGPT) were used only for minor language editing to improve clarity and readability. These tools were not used for research design, data analysis, model development, mathematical formulation, or interpretation of results. All scientific content, methodology, experiments, and conclusions in this manuscript were developed entirely by the authors.

References

[1] M. S. Hossain and F. Parvin, "A comparative study of various statistical and machine learning models for predicting restaurant demand in Bangladesh," *PLoS One*, vol. 20, no. 6, p. e0325449, Jun. 2025, doi: https://doi.org/10.1371/journal.pone.0325449.

[2] J. Lin, "Application of machine learning in predicting consumer behavior and precision marketing," *PLoS One*, vol. 20, no. 5, p. e0321854, May 2025, doi: https://doi.org/10.1371/journal.pone.0321854.

[3] M. R. Hasan, M. R. Islam, and M. A. Rahman, "Developing and implementing AI-driven models for demand forecasting in US supply chains: A comprehensive approach to enhancing predictive accuracy," *Edelweiss Applied Science and Technology,* vol. 9, no. 1, pp. 1045–1068, Jan. 2025, doi: https://doi.org/10.55214/ 25768484.v9i1.4308.

[4] B. Abdollahzadeh, F. S. Gharehchopogh, and S. Mirjalili, "African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems," *Computers & Industrial Engineering*, vol. 158, p. 107408, Aug. 2021, doi: https://doi.org/10.1016/j.cie.2021.107408.

[5] M. Arantes, W. G. Manteiga, J. Torres, and A. Pinto, "Striking a balance: Navigating the trade-offs between predictive accuracy and interpretability in machine learning models," *Electronic Research Archive*, vol. 33, no. 4, pp. 2092–2117, 2025, doi: https://doi.org/10.3934/era.2025092.

[6] A. Thakur, A. Kumar, S. K. Mishra, S. K. Behera, J. Sethi, S. S. Sahu, and S. K. Swain, "Product length predictions with machine learning: An integrated approach using extreme gradient boosting," *SN Computer Science*, vol. 5, no. 6, p. 659, Jun. 2024, doi: https://doi.org/10.1007/s42979-024-02999-8.

[7] A. Mehdary, A. Chehri, A. Jakimi, and R. Saadane, "Hyperparameter optimization with genetic algorithms and XGBoost: A step forward in smart grid fraud detection," *Sensors*, vol. 24, no. 4, p. 1230, Feb. 2024, doi: https://doi.org/10.3390/s24041230.

[8] A. Faramarzi, M. Heidarinejad, B. Stephens, and S. Mirjalili, "Equilibrium optimizer: A novel optimization algorithm," *Knowledge-Based Systems*, vol. 191, p. 105190, Mar. 2020, doi: https://doi.org/10.1016/j.knosys.2019.105190.

[9] Z. Che, C. Peng, C. Wang, and J. Wang, "A novel integrated TDLAVOA-XGBoost model for tool wear prediction in lathe and milling operations," *Results in Engineering*, vol. 27, p. 105984, Sep. 2025, doi: https://doi.org/10.1016/j.rineng.2025.105984.

[10] A. Binbusayyis and M. Sha, "Stability prediction in smart grid using PSO optimized XGBoost algorithm with dynamic inertia weight updation," *Computer Modeling in Engineering and Sciences*, vol. 142, no. 1, pp. 909–931, 2025, doi: https://doi.org/10.32604/cmes.2024.058202.

[11] R. Nurdiansyah, O. A. W. Rijanto, B. Santosa, and S. E. Wiratno, "An improved differential evolution algorithm for permutation flow shop scheduling problem," *International Journal of Operations Research*, vol. 16, no. 2, Jun. 2019, doi: https://doi.org/10.6886/IJOR.201906_16(2).0002.

[12] M. Dobrojevic, L. Jovanovic, L. Babic, M. Cajic, T. Zivkovic, M. Zivkovic, S. Muthusamy, M. Antonijevic, and N. Bacanin, "Cyberbullying sexism harassment identification by metaheurustics-tuned eXtreme gradient boosting," *Computers, Materials and Continua*, vol. 80, no. 3, pp. 4997–5027, 2024, doi: https://doi.org/10.32604/cmc.2024.054459.

[13] M. Todorovic, N. Stanisic, M. Zivkovic, N. Bacanin, V. Simic, and E. B. Tirkolaee, "Improving audit opinion prediction accuracy using metaheuristics-tuned XGBoost algorithm with interpretable results through SHAP value analysis," *Applied Soft Computing*, vol. 149, p. 110955, Dec. 2023, doi: https://doi.org/10.1016/j.asoc.2023.110955.

[14] K. K. Yun, S. W. Yoon, and D. Won, "Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process," *Expert Systems with Applications*, vol. 186, p. 115716, Dec. 2021, doi: https://doi.org/10.1016/j.eswa.2021.115716.

[15] L. Jovanovic, G. Jovanovic, M. Perisic, F. Alimpic, S. Stanisic, N. Bacanin, M. Zivkovic, and A. Stojic, "The explainable potential of coupling metaheuristics-optimized-XGBoost and SHAP in revealing VOCs' environmental fate," *Atmosphere*, vol. 14, no. 1, p. 109, Jan. 2023, doi: https://doi.org/10.3390/atmos14010109.

[16] B. Yang, H. Zhang, Z. Zhao, X. Chen, Y. Chen, and Y. Yang, "Optimization of temperature uniformity for multi microwave source alternating heating based on superpermutation planning enhanced by Harris Hawks optimization," *International Journal of Thermal Sciences*, vol. 206, p. 109349, Dec. 2024, doi: https://doi.org/10.1016/j.ijthermalsci.2024.109349.

[17] C. Jia, X. Kong, M. Wang, Z. Yu, H. Li, Y. Jiang, J. Hu, and C. Ai, "Application of attention mechanism-based LSTM neural network in stratigraphy identification," *Results in Engineering*, vol. 27, p. 105267, Sep. 2025, doi: https://doi.org/10.1016/j.rineng.2025.105267.

[18] F. A. Chiappini, M. R. Alcaraz, and L. Forzani, "A bootstrap-assisted methodology for the estimation of prediction uncertainty in multilayer perceptron-based calibration," *Analytica Chimica Acta*, vol. 1353, p. 343954, Jun. 2025, doi: https://doi.org/10.1016/j.aca.2025.343954.

[19] P. Manigandan, M. S. Alam, M. Alharthi, U. Khan, K. Alagirisamy, D. Pachiyappan, and A. Rehman, "Forecasting natural gas production and consumption in United States evidence from SARIMA and SARIMAX models," *Energies*, vol. 14, no. 19, p. 6021, Sep. 2021, doi: https://doi.org/10.3390/en14196021.

[20] T.-K. Yu, I.-C. Chang, S.-D. Chen, H.-L. Chen, and T.-Y. Yu, "Predicting potential soil and groundwater contamination risks from gas stations using three machine learning models (XGBoost, LightGBM, and Random Forest)," *Process Safety and Environmental Protection*, vol. 199, p. 107249, Jul. 2025, doi: https://doi.org/10.1016/j.psep.2025.107249.

[21] D. Kumar, G. Soni, B. Ramtiyal, and L. Vijayvargy, "Data-driven approach for rational allocation of inventory in a FMCG supply chain," *International Journal of System Assurance Engineering and Management*, 2024, doi: https://doi.org/10.1007/s13198-024-02519-0.

[22] S. K. Panda and S. N. Mohanty, "Time series forecasting and modeling of food demand supply chain based on regressors analysis," *IEEE Access*, vol. 11, pp. 42679–42700, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3266275.

[23] S. Sawangarreerak, P. Thanathamathee, P. Lakkanawanit, and N. S. A. Wahab, "Anchor-based explainable and causal artificial intelligence for enhancing financial predictions of future earnings," *IEEE Access*, vol. 13, pp. 61026–61047, 2025, doi: https://doi.org/10.1109/ACCESS.2025.3557264.

[24] Y. Duan, Y. Mao, Y. Guo, X. Wang, and S. Gao, "COVID-19 propagation prediction model using improved grey wolf optimization algorithms in combination with XGBoost and Bagging-integrated learning," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–13, Jun. 2022, doi: https://doi.org/10.1155/2022/1314459.

[25] Z. Nabavi, M. Mirzehi, and H. Dehghani, "Reliable novel hybrid extreme gradient boosting for forecasting copper prices using meta-heuristic algorithms: A thirty-year analysis," *Resources Policy*, vol. 90, p. 104784, Mar. 2024, doi: https://doi.org/10.1016/j.resourpol.2024.104784.

[26] M. S. Vhatkar, P. S. Mahajan, R. D. Raut, N. Cheikhrouhou, and S. Ghoshal, "Optimized hyperparameters for retail sales forecasting using grid search," *Engineering Applications of Artificial Intelligence*, vol. 158, p. 111472, Oct. 2025, doi: https://doi.org/10.1016/j.engappai.2025.111472.

[27] R. Fauzi, A. Priansyah, P. K. Puspadewa, S. M. Z. Awal, H. T. Nguyen, and A. P. Rifai, "Optimizing vehicle routing for perishable products with time window constraints: A case study on bread distribution," *Jurnal Teknik Industri: Jurnal Keilmuan dan Aplikasi Teknik Industri*, vol. 27, no. 1, pp. 1–20, Jan. 2025, doi: https://doi.org/10.9744/jti.27.1.1-20.

[28] L. Zhou, Q. Zhu, Q. Chen, P. Wang, and H. Huang, "Predicting hospital outpatient volume using XGBoost: a machine learning approach," *Scientific Reports*, vol. 15, no. 1, p. 17028, May 2025, doi: https://doi.org/10.1038/s41598-025-01265-y.

[29] J. Qiu, "An analysis of model evaluation with cross-validation: Techniques, applications, and recent advances," *Advances in Economics Management and Political Science*, vol. 99, no. 1, pp. 69–72, Sep. 2024, doi: https://doi.org/10.54254/2754-1169/99/2024OX0213.