Journal of Applied Data Sciences
Vol. 7, No. 3, September 2026, pp. 2146-2160

An IoT-Driven Hybrid Stacking Ensemble with Deep Meta-Learning for Vending Machine Sales Forecasting

Yulisman[1,*], Zupri Henra Hartomi[2], Rian Ordila[3], Uci Rahmalisa[4], Arie Linarta[5], Yuda Irawan[6]

[1, 2,3,4,6]*Computer Science, Universitas Hang Tuah Pekanbaru, Pekanbaru, Indonesia*

[5]*Computer Science, Universitas Dumai, Dumai, Indonesia*

(Received: January 9, 2026; Revised: March 15, 2026; Accepted: June 20, 2026; Available online: August 9, 2026)

Abstract

Accurate sales prediction is essential for optimizing inventory management and supporting dynamic pricing strategies in the retail industry, particularly for vending machines (VMs) integrated with IoT technologies. The availability of real-time transactional and environmental data from IoT sensors provides opportunities to improve forecasting accuracy by capturing complex temporal patterns and external influences on consumer behavior. However, traditional time series models and single machine learning approaches often struggle to model nonlinear relationships and long-term dependencies in such data. This study proposes a hybrid stacking ensemble model that integrates machine learning and deep learning techniques to enhance the prediction of daily sales volume per Stock Keeping Unit (SKU) in IoT-enabled vending machines. The proposed framework employs Random Forest Regressor (RF), Support Vector Regression (SVR), and XGBoost Regressor (XGB) as Level-0 base learners. Their predictions, along with corresponding residuals, are utilized as meta-features for a Long Short-Term Memory (LSTM)-based meta-learner, enabling effective modeling of both nonlinear and temporal characteristics. The model incorporates diverse features derived from IoT data, including lagged sales, rolling statistics, temporal attributes (day of week and weekend indicators), and environmental variables such as temperature and humidity collected from IoT sensors. Hyperparameter optimization of the LSTM meta-model is performed using Optuna to improve model stability and generalization. The proposed approach is evaluated using 10-Fold Time Series Cross-Validation to preserve temporal data structure. Experimental results show that the proposed model achieves an R^2 of 0.9967 and an RMSE of 0.0899, outperforming the best individual base model, XGBoost (R^2 = 0.9946, RMSE = 0.1121). Although the improvement is marginal, it consistently demonstrates the advantage of combining machine learning and deep learning through a stacking ensemble strategy. These findings indicate that integrating meta-features, residual learning, and IoT-based feature engineering can improve predictive performance and support adaptive decision-making in real-time vending machine operations.

Keywords: Vending Machine, Sales Prediction, IoT, Stacking Ensemble, Machine Learning, Deep Learning.

1. Introduction

Sales forecasting for vending machines is a crucial component of inventory management and automated service operations because the accuracy of predictions affects restocking schedules, operating costs, and product availability for consumers [1], [2]. IoT enables connected devices to collect, transmit, and analyze real-time data for monitoring, automation, and smarter decision-making [3], [4]. As the adoption of IoT-based vending machines increases, sales transaction data can now be combined with real-time environmental information, such as temperature and humidity, which has been proven to have a significant influence on consumer purchasing behavior [5]. Therefore, the integration of IoT into sales forecasting systems presents new opportunities to improve forecasting accuracy and support more adaptive decision-making.

Traditional time series forecasting methods such as ARIMA and VAR have long been used in sales forecasting. However, these methods are limited in capturing non-linear patterns and complex dynamics in vending machine sales transaction data [6]. Previous studies utilized neural network-based models to capture complex demand patterns, meta-learning strategies have been explored to improve predictive accuracy in time series forecasting [7]. The integration of IoT data with machine learning models has shown promising result for real-time retail prediction [8]. Previous studies have shown that ensemble-based approaches improve forecasting robustness in automated retail systems [9], [10]. Despite these advancements, most existing studies still rely on single-model or limited hybrid approaches, which may

***Corresponding author: Yulisman (yulismanaziera27@gmail.com)**
DOI: https://doi.org/10.47738/jads.v7i3.1395

ISSN 2723-6471

not fully capture both nonlinear relationships and temporal dependencies in IoT-driven environments [11]. However, ML models tend to be less effective at leveraging the long-term temporal dependencies inherent in time series data [12]. This shortcoming in detecting long-term temporal constraints and non-linear relationships drives the need for approaches explicitly designed to increasingly capture the dynamics of time-ordered data. Therefore, deep learning models centered on recurrent neural networks are increasingly being used as a solution to optimize the accuracy of sales transaction forecasts for more complex and temporal data types [13].

To overcome these limitations, Long Short-Term Memory (LSTM)-based deep learning is widely used due to its ability to learn long-term sequential patterns in sales data [14]. LSTM has proven to be effective in the domains of retail sales forecasting, energy, and real-time IoT systems [15]. However, the single use of LSTM often faces challenges in generalization stability, especially when the dataset is heterogeneous and influenced by external factors. Deep learning excels at automatically learning complex nonlinear patterns and long-term dependencies from large, high-dimensional data, enabling more accurate and robust predictions[16]. Ensemble learning strategies offer solutions by combining the strengths of various algorithms to improve model accuracy and robustness. One of the latest approaches, stacking ensemble, utilizes several base models to generate initial predictions, which are then combined through a meta model to produce a more accurate final estimate [17], [18], [19]. In the context of time series-based predictions, the use of hybrid models that combine machine learning as base learners and LSTM as meta learners has shown great potential in improving prediction performance.

In addition, hyperparameter optimization using algorithms such as Optuna further strengthens model performance by ensuring the selection of the best parameters to achieve faster convergence and prevent overfitting [20], [21]. Evaluation using quantitative metrics such as MSE, RMSE, MAE, MAPE, and R^2 also allows for a comprehensive assessment of prediction accuracy [22]. Furthermore, IoT integration that generates real-time data enables the model learning cycle to be adaptive, responsive to market dynamics, and relevant to modern vending machine management. The implementation of a real-time data-driven adaptive prediction architecture improves forecasting and supports more effective decision-making, such as inventory management and vending machine replenishment scheduling. This approach develops an end-to-end intelligent system that integrates forecasting, optimization, and automated control.

Despite the promising performance of machine learning and deep learning approaches in sales forecasting, several limitations remain. Traditional machine learning models are effective in capturing nonlinear relationships but often fail to fully exploit temporal dependencies in sequential data. On the other hand, deep learning models such as LSTM are capable of modeling long-term temporal patterns but may suffer from instability and limited generalization when dealing with heterogeneous data influenced by external factors. Furthermore, existing studies often utilize single-model approaches or simple hybrid frameworks without effectively integrating residual information and multi-source IoT-based features into a unified predictive architecture. Therefore, there is a clear research gap in developing a robust and adaptive forecasting model that can simultaneously capture nonlinear relationships, temporal dependencies, and contextual information derived from IoT environments. To address this gap, this study proposes a hybrid stacking ensemble framework that integrates multiple machine learning models with a deep learning-based meta-learner, enhanced by residual learning and IoT-driven feature engineering.

Based on this background, this study proposes a Hybrid Stacking Ensemble of Machine Learning and LSTM with IoT Feature Integration approach for vending machine sales forecasting. By utilizing Random Forest, SVR, and XGBoost as base models and LSTM as a meta model, this system is expected to produce more accurate, adaptive, and applicable predictions in supporting real-time vending machine inventory management.

2. Literature Review

2.1. Overview of the Proposed Methodology

This study developed a vending machine sales prediction model based on stacking ensemble learning by utilizing a combination of machine learning and deep learning. The model development process was carried out in stages, starting from the pre-processing stage, base model training, meta feature formation, to LSTM-based meta model training optimized using Optuna, Adam optimizer, and Early Stopping techniques. The following figure 1 is the model development process:

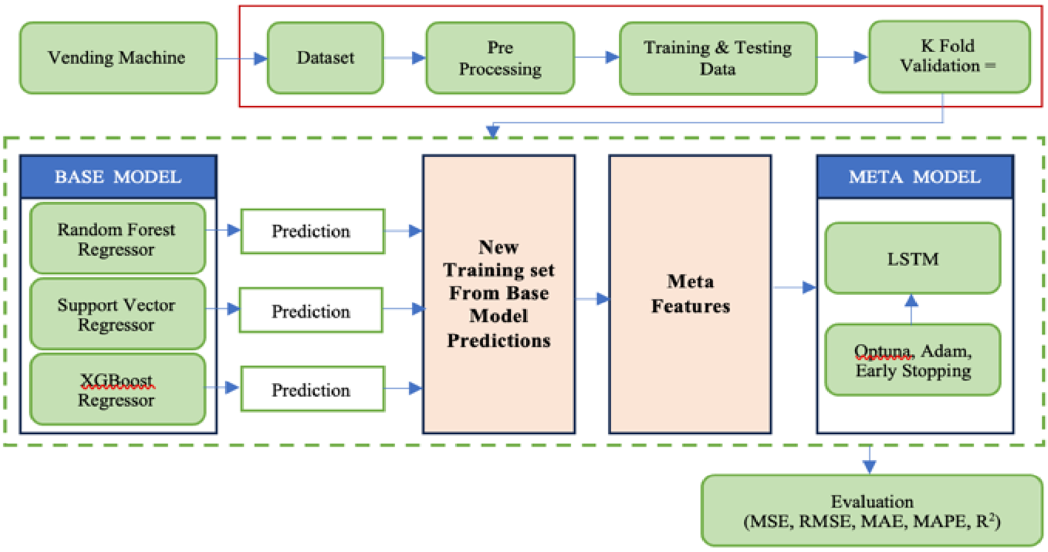

Figure 1. Flow of Model Development

The figure 1 shows the development process of a stacking ensemble learning-based vending machine sales prediction model integrated with an Internet of Things (IoT) system. Sales transaction data and environmental information such as temperature and humidity are collected in real time through sensors and IoT modules on vending machines, then stored and processed as a dataset. The pre-processing stage includes data cleaning, time feature formation, lag variable addition, rolling statistics, and normalization. The processed data is then divided into training data and test data using the K-Fold Cross Validation mechanism so that the evaluation results are more stable and representative. With IoT integration, the field data obtained is always updated and can be directly utilized in the model training and testing cycle.

At the modeling stage, three base models, namely Random Forest Regressor, Support Vector Regressor, and XGBoost Regressor, are used to generate initial predictions which are then combined into meta features. These meta features are utilized by an LSTM-based meta model that is specifically trained to learn time series patterns. The meta model training process is optimized using Optuna to find the best hyperparameters, with the support of the Adam optimizer, which accelerates convergence, and Early Stopping to prevent overfitting. The final model results are evaluated using several key metrics such as MSE, RMSE, MAE, MAPE, and R^2 to provide a comprehensive overview of the accuracy and reliability of predictions. The integration of IoT, stacking ensemble, and deep learning-based optimization makes the prediction system smarter, more adaptive, and more applicable in real-time vending machine management.

2.2. Feature Engineering and Selection

At the Feature Engineering and Selection stage, daily vending machine sales data is further processed to generate variables that can represent temporal patterns and consumption behavior. This process includes the creation of time variables such as day_of_week and is_weekend, the addition of lag variables (lag_1, lag_2, lag_3, lag_7, lag_14) to capture historical dependencies, as well as rolling statistics (rolling_mean and rolling_std in 3, 7, and 14-day windows) to record sales trends and fluctuations. In addition, aggregate variables such as total_demand_day and sku_share were created to measure total daily demand and the relative contribution of each product. The feature selection stage was carried out using Random Forest Regressor by utilizing feature importance to identify the variables that had the most influence on the target (total_sold). This approach ensures that only relevant features are retained, thereby improving model performance and reducing the risk of overfitting. The final features used in model development include: avg_price, day_of_week, is_weekend, sku_encoded, lag_1, lag_2, lag_3, lag_7, lag_14, rolling_mean_3, rolling_mean_7, rolling_mean_14, rolling_std_3, rolling_std_7, rolling_std_14, total_demand_day, dan sku_share.

Feature selection is performed using the feature importance scores generated by the Random Forest Regressor[23]. This method is chosen due to its ability to effectively capture nonlinear relationships and interactions between variables, as well as its robustness to noise and multicollinearity. Compared to alternative approaches such as mutual information, which primarily measures pairwise dependency, or SHAP-based methods, which are computationally more expensive for large datasets, Random Forest importance provides a practical balance between interpretability and computational efficiency. To ensure consistency and avoid arbitrary selection, a threshold-based approach is applied. Specifically, only features with importance values greater than the mean importance score are retained for further modeling. This criterion allows the model to focus on the most relevant variables while reducing redundancy and the risk of overfitting.

2.3. Modelling

This study applies a stacking ensemble learning approach by combining several regression models as meta-feature generators and one advanced model as a meta-learner. Stacking is an ensemble learning technique that combines predictions from multiple models using a meta-learner to improve overall predictive performance[24]. The base models used include Random Forest Regressor, Support Vector Regressor (SVR), and XGBoost Regressor, while Long Short-Term Memory (LSTM) is used as the meta-learner. Random Forest Regressor is a bagging-based ensemble method that builds multiple decision trees using random subsets of data and features to improve accuracy and reduce overfitting. This method is effective in capturing nonlinear relationships and has high stability against data noise. Support Vector Regressor (SVR) is an extension of Support Vector Machine that aims to build a regression function with optimal margin using the concept of ε-insensitive loss and kernel function. SVR is capable of providing good generalization on nonlinear and high-dimensional data. XGBoost Regressor is a gradient-based boosting algorithm designed for computational efficiency and high prediction performance through loss function optimization and regularization. This method is very effective in handling complex and large-scale data. Long Short-Term Memory (LSTM) is a variant of Recurrent Neural Network (RNN) that is capable of learning long-term dependencies in sequential data through a gate mechanism. LSTM is widely used in time series modeling due to its ability to overcome the vanishing gradient problem. The stacking ensemble approach allows the combination of the strengths of various models by using the predictions of the base models as input for the meta-learner, thereby significantly improving the accuracy and stability of predictions.

This vending machine sales prediction model was built using a stacking ensemble learning approach with three base models, namely Random Forest Regressor (RF), Support Vector Regressor (SVR), and Extreme Gradient Boosting (XGBoost). Each base model was trained on preprocessed data to generate initial $\hat{y}RF, \hat{y}SVR, \hat{y}XGB$. In addition, residual error is calculated for each base model using the formula:

$$\boldsymbol{e_i^{(m)}} = y_i - \hat{y}_i^{(m)} \tag{1}$$

with y_i is the actual value, $\hat{y}_i^{(m)}$ prediction from the base model m, and $e_i^{(m)}$ residual. The combination of predictions and residuals from all base models forms meta-features.:

$$Z = [\hat{y}RF, \hat{y}SVR, \hat{y}XGB, e^{(RF)}, e^{(SVR)}, e^{(XGB)}] \tag{2}$$

which is then used as input for the meta model. The meta model used is LSTM because of its ability to capture temporal patterns from sequential data. The meta-features Z input is converted into a three-dimensional tensor (samples, timesteps, features) to match the LSTM architecture. LSTM then processes the sequence with the following internal equation:

$$f_t = \sigma\left(W_f . [h_{t-1}, x_t] + b_f\right), i_t = \sigma\left(W_i . [h_{t-1}, x_t] + b_{fi}\right) \tag{3}$$

$$\tilde{C}_t = \tanh(W_c . [h_{t-1}, x_t] + b_c), C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{4}$$

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o), h_t = o_t \odot \tanh(C_t) \tag{5}$$

with f_t, i_t, o_t, respectively as forget, input, and output gates, C_t cell state, as well as h_t hidden state. Output LSTM continued to the Dense layer to produce the final prediction.

Model optimization was performed using the Adam Optimizer with the Mean Squared Error (MSE) loss function:

$$\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i)^2 \tag{6}$$

In addition, the meta model training process is equipped with Dropout to prevent overfitting and Early Stopping to stop earlier if the loss does not improve. To obtain the best hyperparameter configuration (number of LSTM units, dropout, learning rate, batch size, and epochs), Optuna is used with the Bayesian Optimization principle. In each trial, Optuna selects a combination of parameters.

$\theta = \{units, dropout, lr, batch, epoch\}$, then calculate the score with cross-validation using the coefficient of determination R^2:

ISSN 2723-6471

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{7}$$

Optuna then minimizes the negative value of R2 so that the model obtains the best configuration, which is ultimately used in the final training stage.

2.4. Evaluation

Model evaluation was performed using a 10-Fold Time Series Cross-Validation mechanism to ensure that the resulting performance was more stable and not biased towards the data sequence. Each fold produced evaluation metrics on the base model (Random Forest, SVR, and XGBoost) and meta model (Stacking LSTM). The evaluation results were then averaged to obtain an overview of the overall model performance.

In this study, the 10-Fold Time Series Cross-Validation is implemented using an expanding window approach to preserve the temporal ordering of the data and prevent data leakage[25]. Specifically, in each fold, the training set consists of all observations up to a certain time point, while the validation set includes the immediately following observations. As the folds progress, the training window expands by incorporating new data, while the validation window moves forward in time. This approach ensures that the model is always trained on past data and evaluated on future data, which reflects real-world forecasting conditions. By strictly maintaining chronological order, the proposed validation strategy prevents information leakage from future observations into the training process, thereby providing a more reliable and realistic assessment of model performance.

The main metrics used consist of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R2). The formula for each metric is written as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{10}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{12}$$

From the evaluation results, it can be seen that the Stacking LSTM meta model consistently produces lower MSE, RMSE, MAE, and MAPE values compared to the base model, as well as values $\boldsymbol{R^2}$ higher. This shows that integrating the predictions and residuals from the base model provides significant additional information for the meta model, thereby improving the accuracy of vending machine sales predictions. The evaluation results are visualized through time series graphs (actual vs. predicted), loss per epoch, and regression plots between actual and predicted values. These graphs reinforce the quantitative results by showing that the model's predictions consistently follow the actual sales pattern, the training error steadily decreases over epochs, and the regression scatter points approach the diagonal line y = x.

The selection of evaluation metrics in this study is intended to provide a comprehensive assessment of model performance from multiple perspectives. RMSE is used to penalize larger prediction errors more heavily, making it suitable for capturing significant deviations in sales forecasting. MAE provides a more robust measure of average error that is less sensitive to outliers, while R^2 reflects the proportion of variance in the target variable explained by the model. In addition, MAPE is included to express prediction errors in percentage terms, which is useful for interpretability in real-world business contexts. However, it is important to note that MAPE can be sensitive to small actual values, potentially leading to inflated error percentages when the target values are close to zero. To mitigate this limitation, MAPE is interpreted alongside other metrics such as RMSE and MAE to ensure a balanced and reliable evaluation of model performance.

2.5. IoT Sensor Data Acquisition

The design of the IoT system used in this study can be seen in the figure 2.

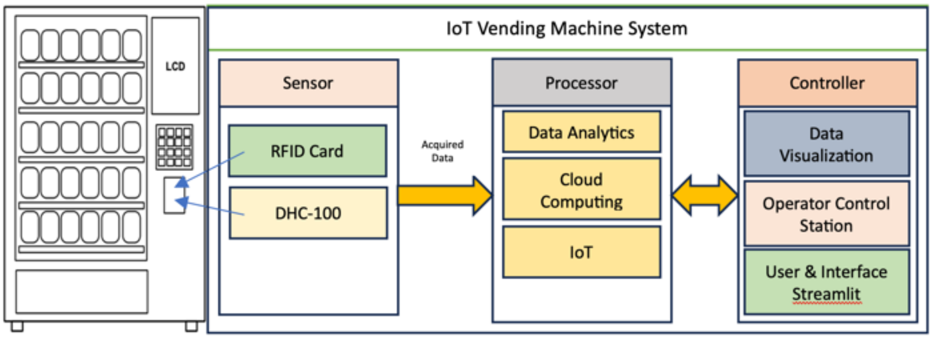

Figure 2. IoT Device Design

The figure 2 above illustrates the architecture of an IoT sensor data acquisition system for real-time sales prediction. This system consists of environmental sensors used to measure temperature and humidity within the vending machine environment, along with an RFID-based identification system to record transaction activities. The environmental sensors continuously capture real-time conditions that may influence consumer purchasing behavior, so that sales data can be recorded in real-time into the system. en processed and transmitted through the IoT module or microcontroller to the cloud infrastructure for storage and further analysis. The collected data is processed in an integrated manner using analytics modules and cloud computing, then the results are visualized through a Streamlit-based user interface. This system also allows operator interaction through station control to monitor sales predictions in real-time.

2.6. Finalization and Evaluation of the Model

The final stage of this research involved the finalization and evaluation of the proposed hybrid stacking ensemble model. Three base models Random Forest Regressor, Support Vector Regressor (SVR), and XGBoost Regressor were trained to generate meta-features, which subsequently served as input for the LSTM-based meta-learner. This design enabled the system to combine the strengths of tree-based, kernel-based, and boosting algorithms with the temporal learning capacity of deep learning models.

To ensure robustness, the training and evaluation were conducted using 10-Fold Cross-Validation. This strategy provides a more representative performance estimate by preserving the temporal order of data, thus avoiding data leakage issues common in time series prediction tasks. Each fold produced evaluation metrics for both base learners and the final stacking-LSTM, enabling a comparative analysis of performance consistency. The meta-model training process utilized the Adam optimizer with a learning rate of 0.001 and incorporated Early Stopping to prevent overfitting by halting the training once the model's loss ceased to improve. Hyperparameter tuning was further refined through Optuna optimization, ensuring the LSTM configuration (units, batch size, and epochs) was optimal for the specific dataset.

3. Results and Discussion

3.1. Dataset and Preprocessing

The dataset used in this study consists of 10,000 vending machine sales transaction data collected during the period from 2023 to 2025. Raw data includes transaction time information, Stock Keeping Unit (SKU), number of items sold, and daily average price. During the pre-processing stage, a series of important steps are carried out, starting from data cleaning to address missing or duplicate values, the creation of time features such as day of week and is weekend, as well as the addition of lag variables (lag-1, lag-2, lag-3, lag-7, lag-14) and rolling statistics (mean and standard deviation) to represent historical sales patterns. In addition, daily total demand and SKU share are also calculated to assess the contribution of each product to total sales. Product category data is converted into numerical values through Label Encoding, then all features are normalized using StandardScaler to make them more suitable for model training. To evaluate the relevance and influence of each variable, Random Forest Regressor was used in the initial stage as a feature importance analysis, so that the features that contribute most to sales predictions can be identified and optimally utilized in further model development. To provide a more comprehensive understanding of the dataset and ensure model robustness, a detailed analysis of data characteristics was conducted. The dataset consists of 10,000 transaction records collected from 2023 to 2025, with the target variable (total_sold) showing a moderately varied distribution across different SKUs and time periods. Descriptive analysis indicates that the data does not exhibit extreme skewness, making it suitable for regression-based modeling. In terms of data completeness, the proportion of missing values is minimal (less than 2% of the total observations). These missing values were handled during preprocessing using appropriate imputation techniques to maintain data integrity and avoid bias in model training. Although class imbalance is typically associated with classification problems, the distribution of the target variable and feature values was further examined to ensure that no specific SKU, time period, or price range dominates the dataset. The analysis confirms that the dataset is relatively well-balanced, supporting stable model training and generalization. This comprehensive data assessment enhances the reliability of the proposed stacking ensemble model. The following figure 3 shows the daily sales trends for each SKU in vending machines during the period from 2023 to 2025:

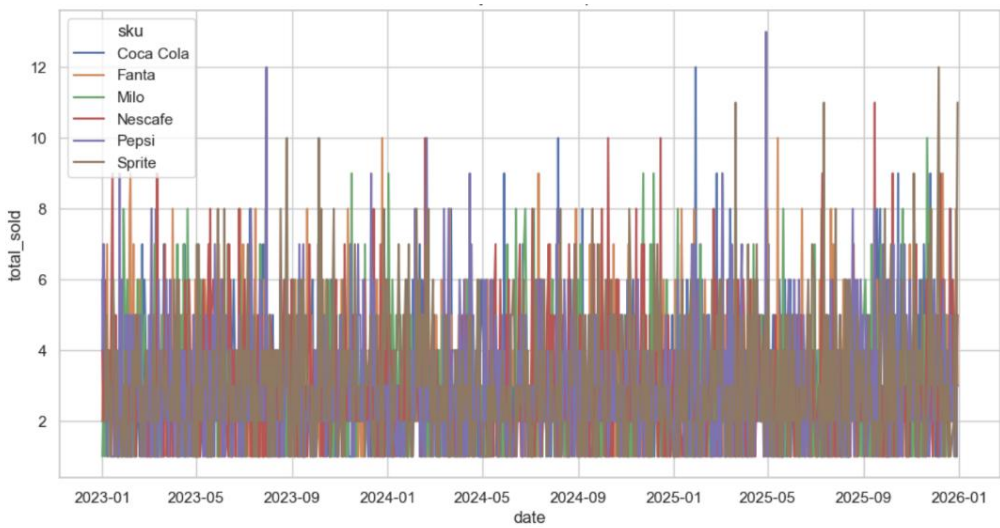

Figure 3. Daily Sales Trends per SKU

The figure 3 above shows that daily sales trends for each SKU tend to fluctuate throughout the 2023–2025 period, indicating variations in consumption influenced by seasonal factors and consumer behavior. Products such as Coca Cola and Pepsi appear to experience more frequent sales spikes than other SKUs, indicating higher popularity and market appeal. This variation pattern is important to analyze further because it can be the basis for strengthening lag features, rolling statistics, and temporal variables in building machine learning-based sales prediction models. The following figure 4 shows the correlation between variables in the preprocessed dataset:

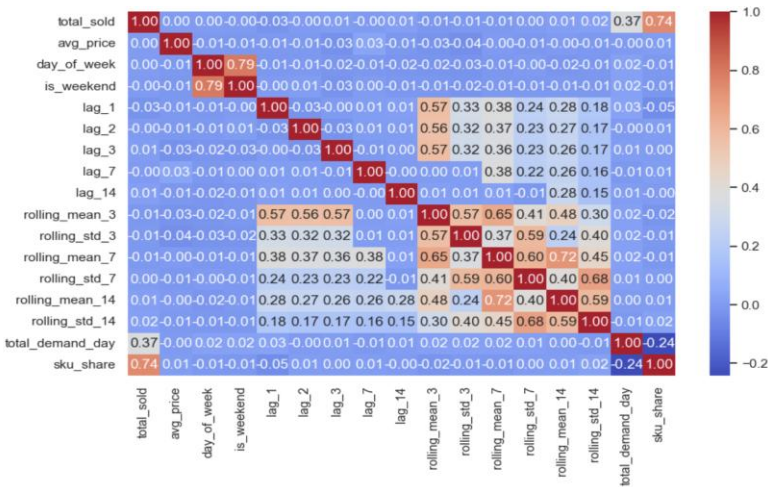

Figure 4. Correlation Heatmap between Variables

The figure 4 above shows that the total_sold variable has a fairly strong positive correlation with sku_share (0.74) and avg_price (0.37), so these two variables can be considered significant in predicting sales. Lag features such as lag_1, lag_2, and lag_3 also show moderate positive correlations with total_sold, confirming the importance of historical sales

ISSN 2723-6471

patterns in model formation. Meanwhile, low correlations in variables such as rolling_std indicate a smaller contribution, but are still relevant for adding variety to the information in the prediction model.

3.2. Performance of Base Models and Stacking Ensemble with LSTM Meta-Learner

To ensure transparency and reproducibility of the experimental results, it is essential to report the training parameters used by the best-performing model of each method. Therefore, the following table 1 presents the key training parameters of the optimal models employed in this study.

Table 1. Training Parameters of the Best Model for Each Method

Method	**Model**	**Key Training Parameters**
Machine Learning	Random Forest Regressor	n_estimators = 300; max_depth = 20; min_samples_split = 2; min_samples_leaf = 1; max_features = auto; random_state = 42
Machine Learning	Support Vector Regressor (SVR)	kernel = RBF; C = 100; epsilon = 0.1; gamma = scale
Machine Learning	XGBoost Regressor	n_estimators = 500; max_depth = 6; learning_rate = 0.05; subsample = 0.8; colsample_bytree = 0.8; objective = reg:squarederror
Hybrid Stacking Ensemble	LSTM Meta-Learner (Optuna Optimized)	LSTM units = 64; dropout = 0.2; optimizer = Adam; learning_rate = 0.001; batch_size = 32; epochs = 100; early_stopping = patience 10

The model evaluation process was conducted using a 10-Fold Cross Validation scheme to ensure the stability of model performance across variations in training and testing data. Each model was tested using five key metrics: R^2, which measures the model's ability to explain data variance, and four error metrics: MSE, RMSE, MAE, and MAPE, which assess the level of prediction error of each model. The following table 2 shows the average model evaluation results after the 10-fold validation process.

Table 2. Average Evaluation Results of Models Based on 10-Fold Cross Validation

Model	**R^2**	**MSE**	**RMSE**	**MAE**	**MAPE**
Random Forest	0.9899	0.0315	0.1592	0.0379	0.8256
SVM	0.9640	0.1123	0.3219	0.1826	8.7128
XGBoost	0.9946	0.0169	0.1121	0.0221	0.5457
Stacking	0.9882	0.0372	0.1724	0.0771	3.1866

Based on the results in table 2, the XGBoost Regressor model performed best compared to the other models, with an R^2 value of 0.9946, indicating the model's ability to explain variation in sales data exceeding 99%. The resulting error rate was also the lowest, with an MSE of 0.0169, RMSE of 0.1121, MAE of 0.0221, and MAPE of only 0.5457%, indicating a very low prediction error rate. This demonstrates that XGBoost is highly effective in capturing non-linear and complex patterns in vending machine sales data.

The Random Forest and Stacking LSTM models also demonstrated competitive performance, with R^2 values of 0.9899 and 0.9882, respectively, indicating that both models are still capable of representing the relationships between variables very well. Although the SVM model had a fairly good level of accuracy (R^2 = 0.9640), its high MAPE value (8.71%) indicates its sensitivity to extreme fluctuations in sales data. Overall, these results indicate that ensemble learning approaches, particularly XGBoost and Stacking LSTM, provide the most consistent and accurate results in the context of IoT data-based vending machine sales prediction.

To visually evaluate model performance, we compare the actual (target) values with the predicted (output) values for both the training and test data. This visualization aims to assess the model's ability to follow actual sales fluctuation

ISSN 2723-6471

patterns, both during the training and testing stages. The following figure 5 shows the model's prediction results for the training and test data.

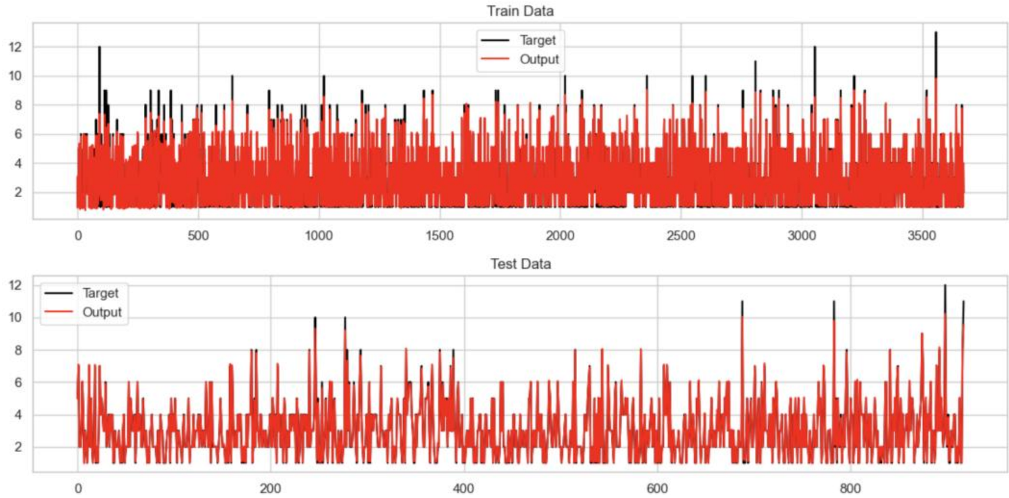

Figure 5. Comparison of Actual and Predicted Sales on Training and Testing Data

The figure 5 above shows that the model follows the actual value change pattern very well on the training data. The prediction line (red) almost completely overlaps the target line (black), indicating a very small error rate. This indicates that the model is able to capture the relationship between features and sales targets effectively. In addition, the consistency between training and testing performance, as reflected by similar error patterns and high R^2 values, suggests that the model generalizes well without significant signs of underfitting or overfitting. Meanwhile, on the test data, the model's prediction pattern remains consistent with the actual trend, although there are slight deviations at some extreme points with sharp sales spikes. These deviations may be influenced by factors that are not explicitly captured in the input features, such as sudden demand fluctuations, promotional activities, or stock variations. In particular, larger prediction errors tend to occur during sharp sales spikes, suggesting that extreme or irregular patterns are more challenging for the model to capture. This observation indicates potential limitations in the available feature set rather than definitive causes, and highlights opportunities for incorporating additional contextual variables in future work. Overall, these results indicate that the Stacking LSTM model is capable of good generalization and has high accuracy in predicting vending machine sales in real time based on historical and IoT data.

To assess the extent to which the model's predictions approximate the actual values, a regression analysis was performed between the target and output variables on the training and testing data. This graph depicts the linear relationship between the actual values and the model's predictions, which is an important indicator for measuring the suitability of the modeling results. The following figure 6 shows the results of the regression test between the target and the model's output.

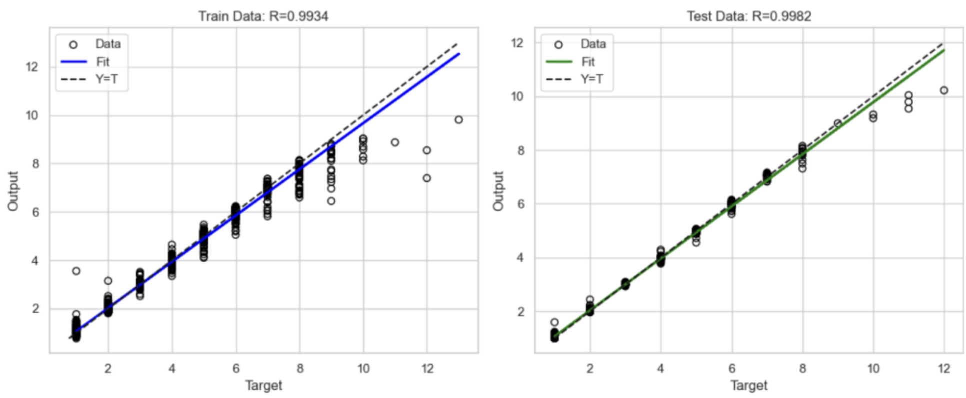

Figure 6. Regression Plot of Actual vs Predicted Values for Training and Testing Data

The figure 6 shows that the model produces a very strong relationship between actual and predicted values, indicated by a correlation coefficient (R) value of 0.9934 for the training data and 0.9982 for the testing data. Most of the data points are spread very close to the diagonal line (Y = T), indicating that the predicted results have minimal deviation from the actual values. This proves that the developed Stacking LSTM model is capable of generalizing very well and has a high level of accuracy in both the training and testing processes.

3.3. Model Optimization and Performance Evaluation

The following figure 7 compares actual data with predictions from a stacking LSTM model equipped with meta-features and optimization using Optuna. This graph demonstrates the model's ability to represent the complex relationship between temporal variables and actual sales results in vending machine data over the observation period.

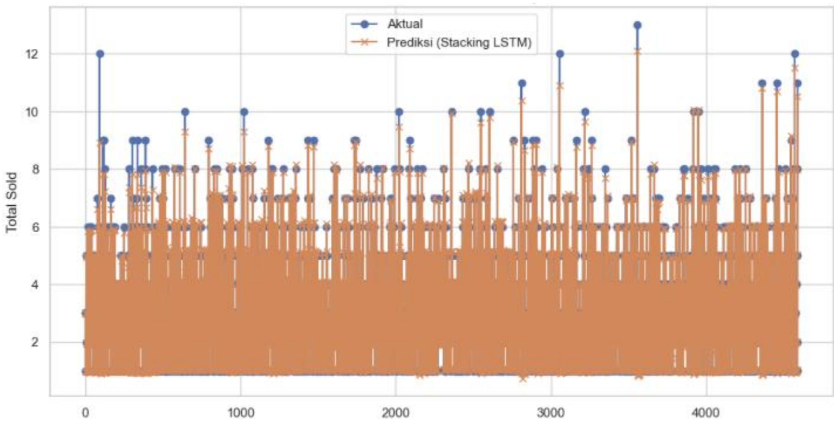

Figure 7. Actual vs Predicted Sales Using Meta Optimized

The visualization on figure 7 results show that the application of meta-features significantly contributes to improving prediction accuracy, as the model can utilize residual information from the base model as additional features to enrich data representation. Furthermore, the hyperparameter optimization process using Optuna helps obtain the best configurations for units, dropout, learning rate, and batch size, resulting in a more stable and efficient LSTM model during training. Overall, the prediction results appear very close to the actual values with minimal deviation, indicating the successful integration of stacking, meta-feature, and Optuna optimization strategies in improving the performance of the IoT-based sales prediction model.

The following graph shows the results of testing the relationship between actual and predicted values from a Stacking LSTM model equipped with meta-features and hyperparameter optimization using Optuna. This figure 8 displays a visualization of the regression between predicted results and actual data in an IoT-based vending machine sales prediction system.

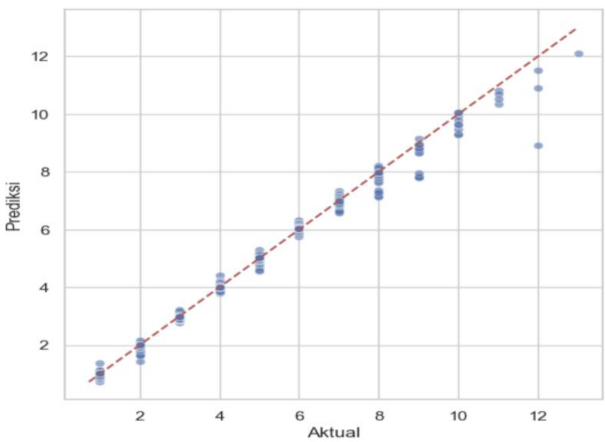

Figure 8. Regression Plot of Actual vs Predicted Values Using Meta-Optimized Stacking LSTM

Based on figure 8 the regression results show that the data points are distributed very close to the diagonal line (Y=T), indicating a high agreement between the actual values and the predicted results. The application of meta-features allows the model to capture the non-linear and complex relationships between base model predictions, while Optuna optimization helps obtain optimal parameter configurations to improve model stability and precision. The dense distribution pattern along the diagonal line reflects strong predictive performance, with relatively small errors and good model generalization ability to new data.

The following are the results of a model evaluation using the 10-Fold Cross Validation scheme, which aims to ensure the stability and consistency of model performance across various data subsets. The evaluation was conducted after applying meta-features and hyperparameter optimization using Optuna on the Stacking LSTM architecture. These test results show a performance comparison between individual models (Random Forest, SVM, and XGBoost) and the combined Stacking model after optimization. The following table 3 shows the average results of the model evaluation.

Table 3. Actual vs Predicted Sales Using Meta Optimized

Model	R^2	MSE	RMSE	MAE	MAPE
Random Forest	0.9899	0.0315	0.1592	0.0379	0.8256
SVM	0.9640	0.1123	0.3219	0.1826	8.7128
XGBoost	0.9946	0.0169	0.1121	0.0221	0.5457
Stacking (Optimization)	0.9967	0.0102	0.0899	0.0540	2.0666

Based on the evaluation results in table 3, the Stacking (Optimization) model demonstrated superior performance compared to other individual models. The R^2 value of 0.9967 indicates the model has excellent predictive ability, while the MSE (0.0102) and RMSE (0.0899) indicate a very small prediction error rate. Furthermore, the MAPE value of 2.0666% indicates high relative accuracy to the actual value. These results demonstrate that the application of meta-features and parameter optimization through Optuna can improve accuracy, reduce errors, and produce a more reliable and efficient vending machine sales prediction model.

The following is a performance comparison between the base models and the Stacking meta-model after optimization using meta-feature augmentation and Optuna hyperparameter tuning. The evaluation was conducted to assess the extent of performance improvement obtained from integrating LSTM as a meta-model in the stacking ensemble approach. The following figure 9 shows the results of the model performance comparison after optimization.

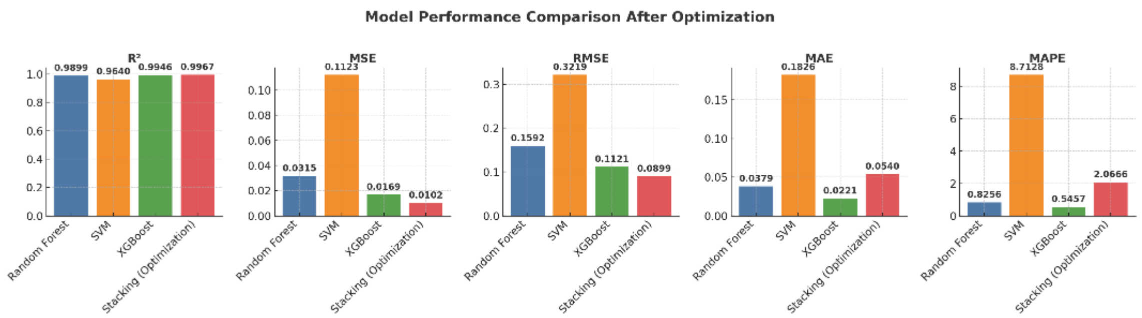

Figure 9. Comparison Between Base and Meta Models After Optimization

The visualization on figure 9 demonstrates that the stacking model significantly outperforms all base models after optimization. It achieves the highest R^2 value (0.9967) and the lowest error metrics (MSE = 0.0102, RMSE = 0.0899, MAE = 0.0540, MAPE = 2.0666), indicating a superior ability to generalize and minimize prediction errors. Compared to XGBoost, which previously had the best performance among the base models, the optimized stacking approach shows a notable reduction in error values. This improvement confirms that the integration of meta features and Optuna-based hyperparameter tuning effectively enhances model precision and robustness in predicting vending machine sales.

The following is a comparison of model performance between the base models and the Stacking meta-model before and after Optuna-based optimization and meta-feature augmentation. The evaluation was conducted using 10-Fold

Cross Validation to ensure model generalization to varied data. The following figure 10 shows the comparison of model performance before and after optimization.

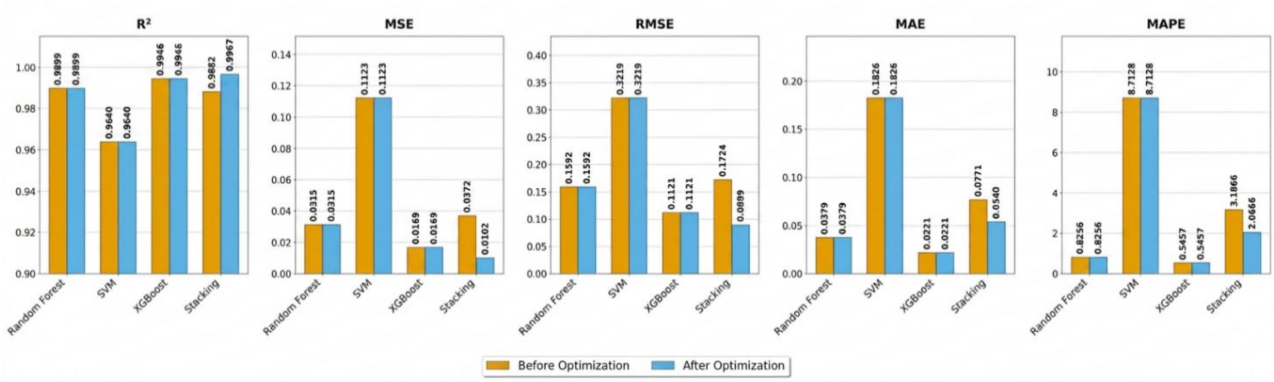

Figure 10. Comparison of Base and Meta Models Before and After Optimization

Based on figure 10 illustrates a significant improvement in model performance after optimization, particularly in the stacking model. The Stacking (Optimization) model achieved the highest R^2 value of 0.9967 and the lowest error rates across all metrics MSE (0.0102), RMSE (0.0899), MAE (0.0540), and MAPE (2.0666). These results demonstrate that the combination of meta-feature augmentation and Optuna-based hyperparameter tuning effectively enhanced prediction accuracy and model robustness compared to the baseline models.

The following image shows the results of testing an IoT system developed to measure direct sales parameters in the field. Data from sensors such as RFID and temperature were successfully transmitted and displayed in real time to Google Sheets. The results of data acquisition can be seen in the following figure 11:

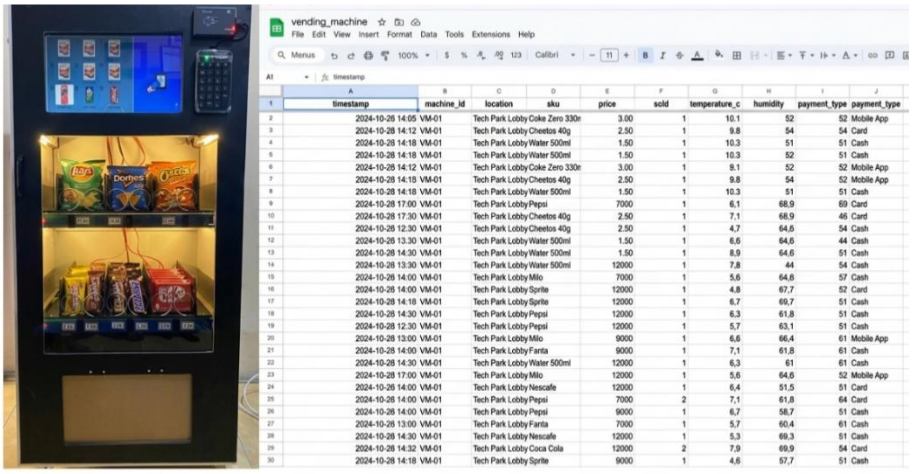

Figure 11. Field Testing and Real-Time Sales Data Acquisition Through IoT

Based on figure 11, this trial proves the successful integration between IoT hardware and Google Sheets-based cloud systems for real-time sales data acquisition. The data transmitted includes five key variables that are highly relevant for sales forecasting: time_spent, machine_id, location, sku, price, sold, temperature_c, humidity, payment_type. The variability of the data recorded at all times shows that the sensors are capable of accurately and continuously capturing environmental dynamics.

The advantage of this system lies in its ability to combine monitoring and forecasting functions in a single integrated platform. Not only does it provide up-to-date inventory status information, but it is also capable of accurately estimating restocking needs. This is particularly relevant in the context of automated inventory management and operational efficiency, where delays in restocking can lead to lost sales, while overstocking increases storage costs and the risk of product expiration. Therefore, this system not only functions as a prediction tool, but also as a decision support system

ISSN 2723-6471

(DSS) that supports data-driven business strategies, strengthens supply chain efficiency, and improves customer satisfaction.

4. Conclusion

Based on the overall experimental results, it can be concluded that the proposed stacking ensemble model with an LSTM meta-learner, optimized using Optuna and enhanced with meta-feature augmentation, demonstrates strong predictive performance in vending machine sales forecasting. The model achieves an R^2 value of 0.9967 with low error metrics (MSE = 0.0102, RMSE = 0.0899, MAE = 0.0540, and MAPE = 2.0666), outperforming individual base models such as Random Forest, SVR, and XGBoost. These results indicate that the integration of meta-features and hyperparameter optimization contributes to improved model accuracy and the ability to capture complex and dynamic sales patterns. However, it is important to note that the model has been evaluated using data from a specific environment, and therefore its generalizability to other vending machine settings, locations, or product distributions has not yet been fully validated. In addition, the model shows potential limitations in capturing extreme fluctuations or irregular sales patterns, which may affect prediction accuracy under certain conditions. Therefore, the applicability of the model to broader real-world scenarios should be interpreted with caution. For future research, it is recommended to further investigate these limitations by evaluating the proposed approach on larger and more diverse datasets across multiple locations to enhance external validity. Moreover, advanced architectures such as Transformer-based models or Temporal Fusion Transformers (TFT) could be explored to better capture long-range dependencies and complex temporal dynamics, particularly in handling irregular or highly fluctuating sales patterns. These directions are expected to improve both predictive performance and model interpretability in real-world time series forecasting applications.

5. Declarations

5.1. Author Contributions

Conceptualization: Y. and Z.H.H.; Methodology: Y.; Software: Y.; Validation: Y., Z.H.H., R.O., and U.R.; Formal Analysis: Y. and Z.H.H.; Investigation: Y.; Data Curation: Y.; Writing Original Draft Preparation: Y.; Writing Review and Editing: Y., Z.H.H., R.O., and U.R.; Visualization: Y.; Supervision: Z.H.H., R.O., U.R., A.L., and Y.I. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Li, F. Xiong, X. Huang, and X. Wen, “Construction and optimization of vending machine decision support system based on improved C4.5 decision tree,” *Heliyon*, vol. 10, no. 3, pp. 1–16, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25024.

[2] B. Rahi, F. B. Kawtharani, A. M. Hassan, and H. F. Hassan, “Assessment of the nutritional value of vending machine products and associated purchasing behavior in Lebanese universities,” *Br. Food J.*, vol. 124, no. 11, pp. 3396–3410, Jan. 2022, doi: 10.1108/BFJ-08-2021-0887.

[3] Y. Cao, Y. Ikenoya, T. Kawaguchi, S. Hashimoto, and T. Morino, “A real-time application for the analysis of multi-purpose vending machines with machine learning,” *Sensors*, vol. 23, no. 4, pp. 1935–1948, 2023, doi: 10.3390/s23041935.

[4] H. Zia, M. U. Jawaid, H. S. Fatima, I. U. Hassan, A. Hussain, S. Shahzad, and M. Khurram, "Plastic waste management through the development of a low cost and light weight deep learning based reverse vending machine," *Recycling*, vol. 7, no. 5, pp. 70–82, 2022, doi: 10.3390/recycling7050070.

[5] K. Kvak and M. Straka, "The use of the Internet of Things in the distribution logistics of consumables," *Applied Sciences*, vol. 14, no. 8, pp. 3263–3278, 2024, doi: 10.3390/app14083263.

[6] U. M. Sirisha, M. C. Belavagi, and G. Attigeri, "Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison," *IEEE Access*, vol. 10, no. 2, pp. 124715–124727, 2022, doi: 10.1109/ACCESS.2022.3224938.

[7] F. Xiao, L. Liu, and T. Peng, "Meta-learning for few-shot time series forecasting," *J. Intell. & Fuzzy Syst.*, vol. 43, no. 1, pp. 1–12, Apr. 2022, doi: 10.3233/JIFS-212228.

[8] V. D. Namburi, D. Rajendran, and A. A. Singh, "Machine Learning Algorithms for Enhancing Predictive Analytics in ERP-Enabled Online Retail Platform," *Int. J. Adv. Ind. Eng.*, vol. 10, no. 04 SE-Articles, pp. 65–73, Dec. 2022, doi: 10.14741/.

[9] M. Seyedan, F. Mafakheri, and C. Wang, "Cluster-based demand forecasting using Bayesian model averaging: An ensemble learning approach," *Decis. Anal. J.*, vol. 3, no. 3, pp. 1-13, 2022, doi: https://doi.org/10.1016/j.dajour.2022.100033.

[10] T. Samal and A. Ghosh, "Ensemble-based predictive analytics for demand forecasting in multi-channel retailing," *Expert Syst. Appl.*, vol. 299, no. 3, pp. 1-12, 2026, doi: 10.1016/j.eswa.2025.130212.

[11] B.-J. Kim and I.-W. Nam, "A Review of Hybrid LSTM Models in Smart Cities," *Processes*, vol. 13, no. 7, pp. 1-46, 2025. doi: 10.3390/pr13072298.

[12] J. Kim, H. Kim, H. Kim, D. Lee, and S. Yoon, "A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges," *Artif. Intell. Rev.*, vol. 58, no. 7, pp. 216–222, 2025, doi: 10.1007/s10462-025-11223-9.

[13] D. Li and J. Xin, "Deep learning-driven intelligent pricing model in retail: from sales forecasting to dynamic price optimization," *Soft Comput.*, vol. 28, no. 20, pp. 12281–12297, 2024, doi: 10.1007/s00500-024-09937-z.

[14] G. Hooshmand Pakdel, Y. He, and X. Chen, "Predicting customer demand with deep learning: an LSTM-based approach incorporating customer information," *Int. J. Prod. Res.*, vol. 47, no. 5, pp. 1–13, Feb. 2025, doi: 10.1080/00207543.2025.2468885.

[15] M. W. Hasan, "Design of an IoT model for forecasting energy consumption of residential buildings based on improved long short-term memory (LSTM)," *Meas. Energy*, vol. 5, no. 3, pp. 100033 - 100055, 2025, doi: https://doi.org/10.1016/j.meaene.2024.100033.

[16] I. D. Mienye and T. G. Swart, "A comprehensive review of deep learning: Architectures, recent advances, and applications," *Information,* vol. 15, no. 12, pp. 755–772, 2024, doi: 10.3390/info15120755.

[17] T. Stephan, P. P. D. Paramana, C.-C. Lin, S. Agarwal, and R. Verma, "Federated learning-driven IoT system for automated freshness monitoring in resource-constrained vending carts," *J. Big Data*, vol. 12, no. 1, pp. 33–46, 2025, doi: 10.1186/s40537-025-01063-3.

[18] W. Liang, P. Lu, and M. Wang, "Hybrid Stacking Model with Optuna Optimization for Open Stope Stability Prediction," Rock Mech. Rock Eng., vol. 59, no. 4, pp. 4845-4869, Apr. 2026, doi: https://doi.org/10.1007/s00603-025-05043-0.

[19] A. Jain, S. Khan, K. Koli, H. Taneja, A. Panwar, T. Alshammari, E. M. Obaid, and S. M. Abdu, "PhishNet 1.0: optuna-optimized stacking ensemble with Boruta-based feature selection for phishing URL detection," Sci. Rep., vol. 16, no. 1, pp. 1480-1502, Jan. 2026, doi: https://doi.org/10.1038/s41598-025-31447-7

[20] V. A. Boateng and B. Yang, "A Global Modeling Pruning Ensemble Stacking With Deep Learning and Neural Network Meta-Learner for Passenger Train Delay Prediction," *IEEE Access,* vol. 11, no. May, pp. 62605–62615, 2023, doi: 10.1109/ACCESS.2023.3287975.

[21] A. M. Joseph Raj, S. L. Kruthika, and A. S., "Enhanced meta ensemble stacking approach with XGBoost and optuna based detection of Parkinson's disease," Front. Digit. Health, vol. 8, no. 1, pp. 1799144-1799165, Mar. 2026, doi: https://doi.org/10.3389/fdgth.2026.1799144

[22] A. A. K Abdul Hamid, “Improvement of Time Forecasting Models Using Machine Learning for Future Pandemic Applications Based on COVID-19 Data 2020–2022,” *Diagnostics*, vol. 13, no. 6, pp. 1-32, 2023. doi: 10.3390/diagnostics13061121.

[23] G. W. Kim, C. Y. Ju, H. Seok, and D. H. Lee, “Adaptive Stacking Ensemble Techniques for Early Severity Classification of COVID-19 Patients,” *Appl. Sci.*, vol. 14, no. 7, pp. 1-15, 2024, doi: 10.3390/app14072715.

[24] K. Kent, L. Li, A. O’Fee, A. Iyer, and C. Fleming, “A multimodal evaluation of an Australian university food environment and sustainability practices through student perspectives and a cross-sectional audit,” *Int. J. Sustain. High. Educ*., vol. 27, no. 4, pp. 864–881, Mar. 2025, doi: 10.1108/IJSHE-07-2024-0471.

[25] A. Zaidi, “Predicting wildfires in Algerian forests using machine learning models,” *Heliyon*, vol. 9, no. 7, pp. 1-17, 2023, doi: 10.1016/j.heliyon.2023.e18064.