Data Science 0 (0) 1
IOS Press


# A Deep Learning Approach for Durian Price Prediction Using Bidirectional LSTM with Seasonal Analysis

Pattharaporn Thongnim [a] and Nopparat Panngam [b,*]
[a] *Department of Mathematics, Burapha University, Chonburi, Thailand*
*E-mail: pattharaporn@buu.ac.th; ORCID: https://orcid.org/0000-0001-8904-3979*
[b] *Department of Mathematics, Burapha University, Chonburi, Thailand*
*E-mail: noparatk@buu.ac.th; ORCID: https://orcid.org/0009-0008-7947-6963*

**Abstract.**
This research presents a deep learning approach for predicting durian prices using Bidirectional Long Short-Term Memory (BiLSTM) networks with seasonal analysis. The study addresses two critical challenges in agricultural price forecasting which include handling missing values in seasonal time series data and capturing complex temporal dependencies. The proposed methodology employs Prophet model for missing value imputation incorporating multiple seasonal components and Fourier terms to maintain temporal integrity. A BiLSTM architecture is implemented with a 25-day sliding window optimized to capture market cycles and seasonal patterns. Therefore, the model is evaluated using 10-fold cross-validation on durian price data from Thailand's primary trading hub during 2023-2024 harvest seasons. The findings demonstrate strong predictive performance with R-squared values of 0.9254 $\pm$ 0.0439 on validation data and Mean Squared Error of 65.3470 $\pm$ 33.1819. The model successfully captures key market phases including initial stability, early decline, mid-season plateau and peak season dynamics. The integration of seasonal features with bidirectional processing enables accurate prediction of price movements providing valuable insights for stakeholders across the durian supply chain.

Keywords: Bidirectional LSTM, Deep learning, Prophet model, Imputation, Durian price

## 1. Introduction

In developing countries, agricultural commodities serve as fundamental drivers of economic growth and development. These countries play an increasingly vital role in global food security contributing significantly to the world's agricultural production and trade. The agricultural sector provides essential food resources and generates substantial employment opportunities and foreign exchange earnings through exports [10]. As global food demand grows rapidly, developing countries with rich farmland are becoming more important in the world's food supply chain [18]. This growing role requires better farming management and accurate price forecasting to support steady economic growth [17].

A tropical fruit known as Durian in Southeast Asia has become a major economic crop that drives Thailand's agricultural exports [2, 20]. The fruit's unique characteristics and growing global demand have led to significant market volatility making price prediction increasingly important for stakeholders

*Corresponding author. E-mail: noparatk@buu.ac.th.

across the supply chain. In recent years, Thailand has established itself as the world's largest durian exporter with both production and export values reaching record breaking levels [8, 13]. However, the seasonal growing patterns and external market factors make it difficult to maintain consistent prices throughout the year.

Missing values in time series data present significant challenges for data analysis and forecasting [4, 15] particularly in agricultural datasets [11, 14, 25]. These gaps can occur due to various reasons such as data collection failures, sensor malfunctions, holidays and market closures. The presence of missing values can disrupt the temporal continuity of the data making it difficult to identify trends, patterns and seasonal variations. The impact of missing values becomes more complex when dealing with seasonal time series data [23, 24]. Gaps in the data can break the seasonal patterns making it challenging to understand and model the underlying seasonal components. Traditional handling methods such as deletion of incomplete records are often unsuitable for time series analysis as they can remove valuable temporal information and disrupt the sequential nature of the data [3, 21]. Moreover, the relationship between missing values and seasonal patterns needs careful consideration.

Therefore, data imputation techniques for time series data with seasonal patterns have evolved significantly ranging from traditional mathematical approaches to advanced machine learning methods. Various strategies have been tested to deal with missing values while maintaining the time based patterns and seasonal features of the data. Linear interpolation assumes a straight line between known points while polynomial interpolation fits a polynomial curve through the data points [5]. Spline interpolation offers more flexibility by fitting piecewise polynomial functions making it particularly suitable for capturing seasonal fluctuations [22].

Moreover, sophisticated approaches leverage statistical and machine learning techniques. Moving average methods and exponential smoothing can effectively handle missing values while accounting for trends and seasonality [9]. The Prophet model has gained popularity for its ability to handle missing values in time series data while automatically detecting seasonal patterns and trend changes. Its decomposition approach, separating trend, seasonality and holiday effects [1, 12] that makes it particularly effective for agricultural price data with regular seasonal variations.

Advanced machine learning methods offer powerful alternatives for missing value imputation. A gradient boosting framework (XGBoost) can handle missing values through its built in mechanisms while capturing complex non-linear relationships in the data [16]. Random Forest imputation leverages multiple decision trees to predict missing values considering various features including seasonal indicators. Hybrid approaches combining multiple techniques have shown promising results [19]. For example, combining traditional interpolation methods with machine learning models can leverage the strengths of both approaches. These hybrid solutions often provide better results than single methods especially for complex seasonal patterns.

In addition, the traditional methods of price forecasting such as statistical analysis and time series models often struggle to capture the complex patterns and seasonality inherent in agricultural commodity prices. This limitation has driven the need for more sophisticated prediction approaches that can handle non-linear relationships and temporal dependencies. Deep learning techniques, particularly Long Short-Term Memory (LSTM) networks have shown promising results in time series forecasting across various domains [6]. The bidirectional variant of LSTM offers enhanced capability by processing data in both forward and backward directions [7] potentially improving the model's ability to capture seasonal patterns and long-term dependencies in price movements.

This research presents a new approach to durian price prediction by implementing a Bidirectional LSTM model with seasonal analysis. The study focuses on historical price data from the Thai durian market during 2023-2024 incorporating seasonal factors and market trends. The proposed model utilizes K-fold cross-validation to ensure robust performance evaluation and reliability of predictions. By analyzing both historical patterns and seasonal variations, this approach aims to provide more accurate price forecasts that can benefit farmers and other stakeholders in the durian industry.

## 2. Methodology

### *2.1. Data Preparation*

The data preparation process focused on durian price data collected from the Chanthaburi market, Thailand's primary durian trading hub. The initial dataset spans a two year period from 2024 to 2025 with daily price records during the peak harvest seasons (April-June). The original data has gaps because markets are closed and data collection is incomplete which required an advanced method to fill in the missing information. Table 1 presents the characteristics of the initial dataset.

Table 1
Initial Dataset Characteristics

| Season | Total Days | Missing Values | Percentage (%) |
|---|---|---|---|
| Apr-Jun 2023 | 91 | 61 | 67.03 |
| Apr-Jun 2024 | 91 | 59 | 64.84 |

The initial analysis of price dataset spanning from April to June in both 2023 and 2024 revealed significant data issues. The dataset demonstrates a high proportion of missing values with approximately 67.03% missing data in the 2023 period (61 out of 91 days) and 64.84% in the 2024 period (59 out of 91 days). These substantial gaps in the data are primarily attributed to inconsistent data collection procedures and technical limitations in the data gathering process. To address these data quality concerns, the proposed solution implements an interpolation strategy using Prophet model which is a robust time series forecasting tool. Prophet is particularly well-suited for this task due to its ability to handle missing values in time series data while accounting for multiple seasonality patterns and trends [1].

In addition, the preprocessed data is subjected to feature engineering procedures for model implementation. This included temporal feature creation with day and month indicators followed by normalization using Min-Max scaler. Table 2 shows the structure of the final processed dataset used for model training.

Table 2
Processed Dataset Structure

| Feature | Description |
|---|---|
| Date | Daily timestamps |
| Price | Prophet imputed prices |
| Days | Cumulative day index |
| Month | Seasonal month indicator |

The processed dataset comprises four key features essential for the time series analysis. The date column contains daily timestamps that track temporal progression while price represents the daily price values that have been cleaned and imputed using the Prophet algorithm to handle missing data. The days feature serves as a continuous counter of days from the start of the dataset enabling trend analysis. Additionally, the month feature is included to capture seasonal patterns that occur on a monthly basis allowing the model to identify and learn recurring price fluctuations throughout the year.

## 2.2. Data Preprocessing

### 2.2.1. Missing Value Treatment with Prophet Model

The missing value imputation process employed Prophet model incorporating multiple seasonal components and Fourier terms to capture the complex patterns in durian price data. The implementation consisted of several key steps designed to maintain the temporal integrity and seasonal characteristics of the price series. This additive regression model can be expressed as

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{1}$$

where $g(t)$ is the trend component, $s(t)$ is the seasonal component with Fourier series, $h(t)$ represents holiday effects and $\epsilon_t$ is the error term. For seasonal components with Fourier series as

$$s(t) = \sum_{n=1}^{N} \left( a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right) \tag{2}$$

where $N$ is the number of Fourier terms, $P$ is the period and $a_n, b_n$ are the Fourier coefficients. The imputation process began with data segregation where existing price points are used for model training and missing values are identified for imputation. The Prophet model is trained on non missing values learning the underlying patterns and seasonality in the price data. The model then generated predictions for the missing time points maintaining consistency with the observed seasonal patterns and trends. The effectiveness of this approach is validated through visual inspection of the imputed values against existing patterns and through statistical validation of the predictions' alignment with historical price ranges.

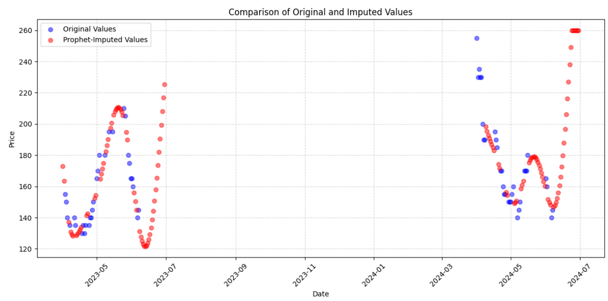



Fig. 1. Comparison of Original and Imputed Values

Fig. 1 shows that the final dataset combining original and imputed values provided a complete time series that preserved the essential characteristics of durian price movements while maintaining temporal continuity. This approach ensured that subsequent analyses and modeling efforts would have a robust foundation of continuous data points, essential for accurate time series prediction.

The imputation results demonstrate strong adherence to market dynamics with the maximum price constraint of 260 THB serving as a critical boundary condition. This threshold is carefully implemented to maintain economic plausibility while preserving the seasonal patterns identified by the Prophet model. The imputed values (shown in red) seamlessly integrate with the existing data points (shown in blue) following the natural price progression observed in the durian market. Notably, the model captures both gradual price transitions and seasonal fluctuations reflecting the typical market behavior during the April-June harvest period. The continuity between original and imputed values suggests that the Prophet model effectively learned the underlying price patterns providing a reliable dataset for subsequent LSTM modeling.

#### 2.2.2. Dataset Structure and Feature Engineering

Following missing value imputation, temporal features are engineered to capture both linear time progression and seasonal patterns as

$$\mathbf{X} = \{Days_t, Month_t, Price_t\} \tag{3}$$

where $Days_t$ represents the cumulative days from the start, $Month_t$ indicates the seasonal month (4-6) and $Price_t$ is the durian price at time $t$.

Table 3
Processed Dataset Structure

| Feature | Description |
|---|---|
| Date | Daily timestamps |
| Price | Prophet-imputed prices |
| Days | Cumulative day index |
| Month | Seasonal month indicator |

Table 3 shows the processed dataset that contains both original and normalized features structured for time series analysis. The temporal elements include daily timestamps and a cumulative day index to track time progression while seasonal patterns are captured through monthly indicators. The price data which has been preprocessed using Prophet for imputation of missing values serves as the main variable of interest.

For temporal feature engineering, the following transformations are applied as

$$\begin{aligned} Date_t &= \text{timestamp of day } t \\ Price_t &= \text{Prophet predicted price at time } t \\ Days_t &= t - t_{min} \\ Month_t &= \begin{cases} 4 & \text{for April} \\ 5 & \text{for May} \\ 6 & \text{for June} \end{cases} \end{aligned}$$

where $t$ represents each time point in the series, $t_{min}$ is the first day in the dataset and $Month_t$ represents the seasonal indicator for the durian harvest period from April to June.

#### 2.2.3. Data Normalization Process

The Min-Max scaler transformation is applied to normalize all features to [0,1] as

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4}$$

where $x$ represents the original value of the feature (Days, Month and Price), $x_{min}$ is the minimum value in the dataset for that feature and $x_{max}$ is the maximum value.

Table 4

Feature Scaling Parameters and Rationale

| Feature | Original Range | Scaled Range | Purpose |
|---|---|---|---|
| Days | [0, 91] | [0, 1] | Linear time progression |
| Month | [4, 6] | [0, 1] | Seasonal pattern recognition |
| Price | $[min_p, max_p]$ | [0, 1] | Value normalization |

Table 4 presents the scaling parameters applied to each feature in the dataset. The Days feature represents temporal progression within the April-June harvest period. This feature is normalized from its original range [0, 91] to [0, 1] to facilitate the model's learning of linear temporal patterns. The month feature spans from April (4) to June (6) and is normalized to [0,1] to facilitate the recognition of seasonal patterns. In this normalized scale, April corresponds to 0 and June to 1, creating a standardized representation of the harvest season progression. The Price feature represents durian prices ranging from minimum ($min_p$) to maximum ($max_p$) values in Thai Baht. The values are normalized to [0,1] to maintain numerical stability during model training and ensure compatibility with other temporal features. Therefore, this comprehensive scaling approach ensures that all features contribute proportionally to the model's learning process while maintaining their relative patterns and relationships.

Given the linear progression of days from April to June as

$$f(t) = \frac{t - t_{min}}{t_{max} - t_{min}} \tag{5}$$

where $f(t)$ is the normalized time value, $t$ is the current day, $t_{min} = 0$ (April 1) and $t_{max} = 91$ (June 30). This produces a linear progression from 0 to 1.

$$f(t) = \frac{t - 0}{91 - 0} = \frac{t}{91} \tag{6}$$

where $t \in [0, 91]$.

### 2.3. LSTM Model Implementation

For the LSTM model, sequences are created using a sliding window approach with a 25-day time step. The model employs a 25-day sliding window which is carefully selected based on multiple domain-specific considerations. This window size holds particular significance in durian price prediction as it aligns with natural market cycles including lunar month trading patterns in Thai agricultural markets.

The chosen length enables effective capture of price momentum and trend reversals that typically manifest over 3-4 weeks during the harvest season. Additionally, this duration allows the model to track seasonal progression effectively capturing price variations across early, peak and late season supply dynamics. The sliding window sequence can be formally expressed as

$$X_i = \{[Days_{t-24:t}, Month_{t-24:t}]\}, y_i = Price_{t+1} \tag{7}$$

where $X_i$ represents the input sequence and $y_i$ is the target price for prediction. From a technical perspective, the implementation utilizes a sliding window mechanism that advances one day at a time through the dataset. In the April to June period spanning approximately 91 days, the 25-day window generates $N = T - w + 1$ sequences where $T = 91$ is the total number of days and $w = 25$ is the window size resulting in 66 training sequences per season. These overlapping sequences preserve the temporal continuity of price movements ensuring that the model captures both short-term fluctuations and longer-term trends in the durian market. This approach creates a balance between capturing detailed price patterns and maintaining an adequate sample size for robust model training while satisfying.

$$\{X_i, y_i\}_{i=1}^{N}, \tag{8}$$

where $N = 91 - 25 + 1 = 67$. This approach ensures that the model learns both short-term price fluctuations and medium-term seasonal patterns while maintaining the temporal dependencies crucial for accurate price prediction in the volatile durian market.

### 2.4. Model Architecture

The proposed model architecture implements a deep Bidirectional LSTM network designed specifically for durian price prediction. The architecture consists of multiple layers optimized for capturing both forward and backward temporal dependencies in the price data as

The detailed layer structure is as follows

(1) **First Bidirectional LSTM Layer**

- 150 units with sequence return
- Processes 25 time steps of temporal features
- Bidirectional processing captures future and past dependencies
- Followed by 30% dropout for regularization

(2) **Second Bidirectional LSTM Layer**

- 150 units without sequence return
- Further refines temporal patterns
- Followed by 30% dropout

(3) **Dense Layer**

- 50 units with ReLU activation
- Transforms LSTM outputs into refined features

(4) **Output Layer**

- Single unit for price prediction
- Linear activation for regression task

The model is compiled with the following configurations

- **Optimizer:** Adam with initial learning rate of 0.001
- **Loss Function:** Mean Squared Error (MSE)
- **Callbacks**
  - * Early Stopping with patience=10
  - * ReduceLROnPlateau with factor=0.5, patience=5

Training parameters include

- **Epochs** = 100
- **Batch Size** = 15
- **K-Fold Splits** = 10

The selection of this architecture is driven by several key capabilities essential for accurate durian price prediction. The bidirectional processing enables the model to capture temporal dependencies in both forward and backward directions for understanding price movement patterns. The model's deep structure combined with seasonal feature inputs effectively handles the inherent seasonality in durian price data. To ensure robust performance, dropout layers are strategically placed to prevent overfitting while the adaptive learning rate mechanism to ensures optimal convergence during training. Furthermore, the implementation of k-fold cross-validation maintains model generalization by validating performance across different data subsets making the model reliable for various market conditions throughout the durian season.

### *2.5. Evaluation Metrics*

The performance of the Bidirectional LSTM model is evaluated using multiple metrics to ensure comprehensive assessment of its predictive capabilities. Three primary metrics are employed as

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

where MSE (Mean Squared Error) measures the average squared difference between predicted prices ($\hat{y}_i$) and actual prices ($y_i$) providing a measure of prediction accuracy that penalizes larger errors more heavily.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{10}$$

The Root Mean Square Error (RMSE) provides an intuitive measure of prediction error in the same unit as durian prices (Thai Baht) making it directly interpretable for stakeholders.

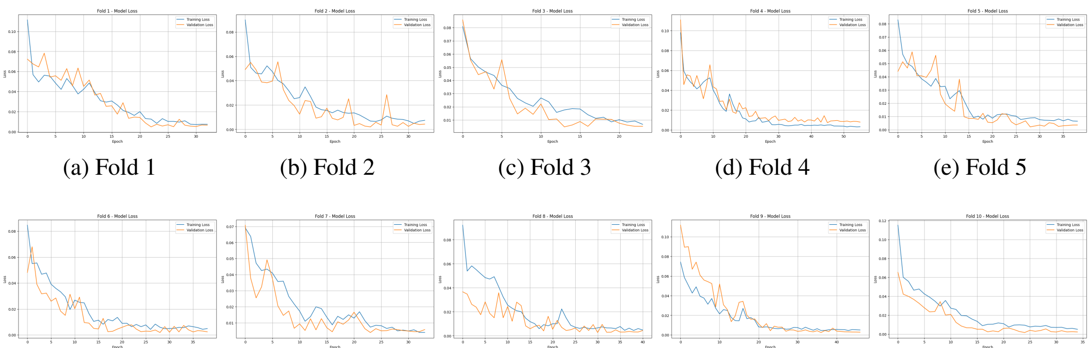
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (11)$$

The R-squared ($R^2$) indicates the proportion of price variance explained by the model where values closer to 1 indicate better fit. To ensure robust evaluation, these metrics are calculated across all 10 folds of the cross-validation. The use of both training and validation metrics helps identify potential overfitting and underfitting while the cross-validation approach ensures the model's performance consistency across different data subsets. The combination of these metrics provides a comprehensive view of the model's ability to capture and predict durian price patterns.

## 3. Results

### 3.1. Training Process

The training process is conducted using 10-fold cross-validation over 100 epochs with early stopping criteria. Fig. 2 shows the learning progression for a representative fold showing the convergence of training and validation losses.



(a) Fold 1 (b) Fold 2 (c) Fold 3 (d) Fold 4 (e) Fold 5

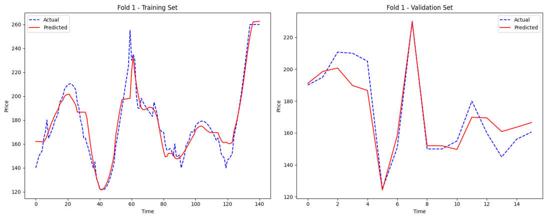
(f) Fold 6 (g) Fold 7 (h) Fold 8 (i) Fold 9 (j) Fold 10

Fig. 2. Training and Validation Loss Curves for Each Fold in 10-Fold Cross-Validation

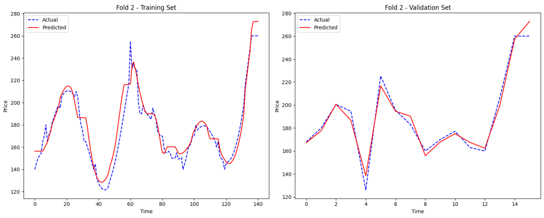
The model typically converged within 40-60 epochs, as demonstrated by the early stopping mechanism. The learning rate adaptation is triggered approximately 2-3 times during training, effectively managing the optimization process. Fig. 3 shows the model's prediction performance on both training and validation sets.

Table 5
Summary of Model Performance Metrics Across 10-Fold Cross-Validation

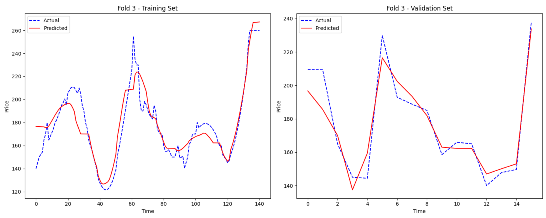
| Metric | Training | Validation |
|---|---|---|
| R-squared ($R^2$) | 0.9153 ± 0.0243 | 0.9254 ± 0.0439 |
| MSE | 89.8651 ± 26.9683 | 65.3470 ± 33.1819 |



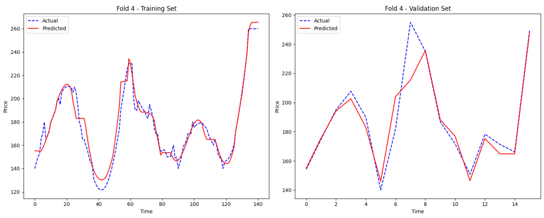

(a) Fold 1



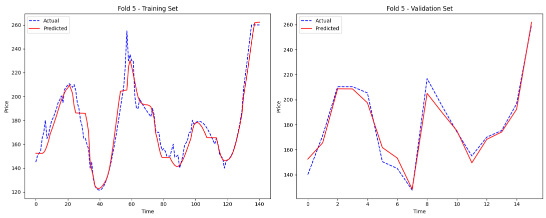

(b) Fold 2



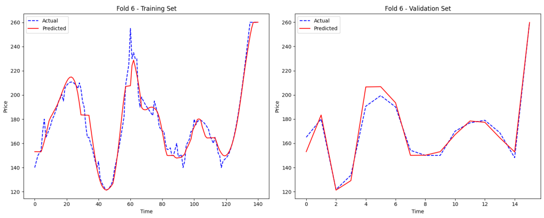

(c) Fold 3



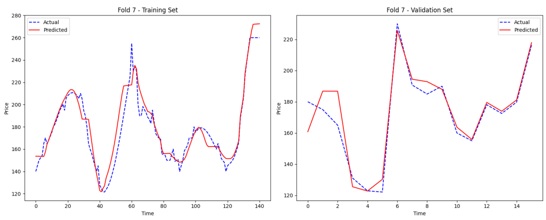

(d) Fold 4



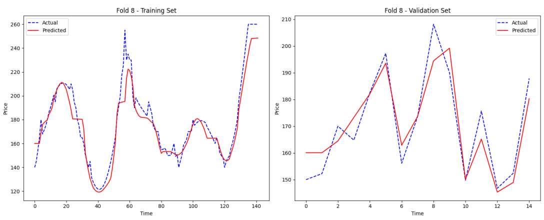

(e) Fold 5



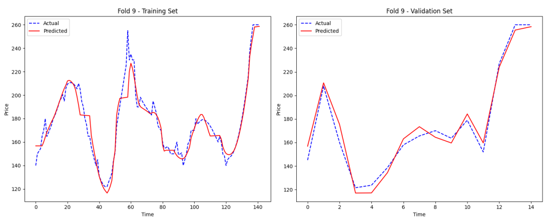

(f) Fold 6



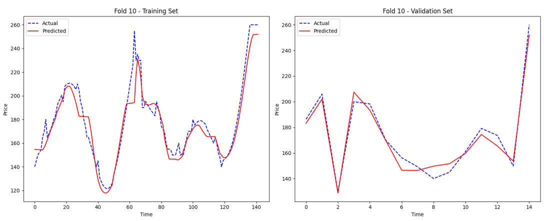

(g) Fold 7




(h) Fold 8




(i) Fold 9




(j) Fold 10

Fig. 3. Comparison of Actual vs Predicted Prices for All Folds. Each subfigure shows Training (left) and Validation (right) results.

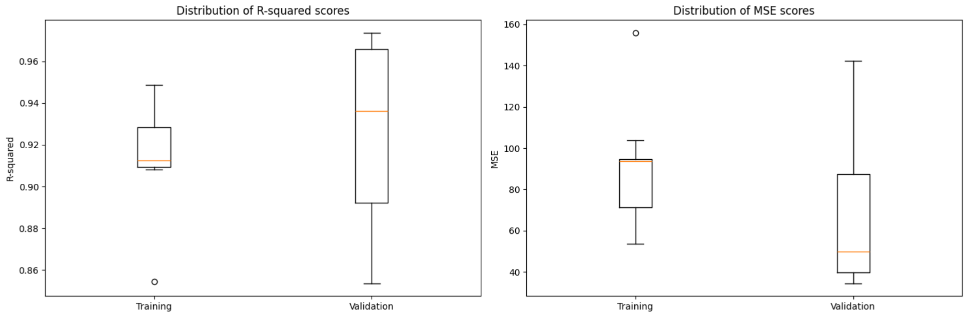



Fig. 4. Distribution of Performance Metrics Across K-Fold Validation

The cross-validation results demonstrate consistent performance across folds as shown in Fig. 4. The model demonstrated robust performance across both training and validation sets as shown in Table 5. The high R-squared values (training: $0.9153 \pm 0.0243$, validation: $0.9254 \pm 0.0439$) indicate that the model explains approximately 92% of the variance in durian prices suggesting excellent predictive capability. The close alignment between training and validation R-squared values with only a 1.01% difference in their means. This indicates that the model generalizes well to unseen data without overfitting. The Mean Squared Error (MSE) values show better performance on the validation set ($65.3470 \pm 33.1819$) compared to the training set ($89.8651 \pm 26.9683$) though with higher variability as indicated by the larger standard deviation. This pattern suggests that while the model maintains consistent predictive accuracy across different data subsets, its performance can vary depending on the specific characteristics of the validation data. The relatively small standard deviations in R-squared values across folds further confirm the model's stability and reliability for durian price prediction.

The training process revealed several important insights about the model's performance and characteristics. The consistent convergence across all folds demonstrated the model's stable learning behavior while the small gap between training and validation metrics confirmed that the dropout layers effectively prevented overfitting. Furthermore, the bidirectional architecture proved successful in capturing price patterns in both temporal directions enhancing the model's ability to understand complex price movements. The significant contribution of seasonal features to prediction accuracy highlighted the importance of incorporating temporal patterns in durian price forecasting, particularly given the highly seasonal nature of durian production and market dynamics.

Table 6

Training Process Summary

| Metric | Mean | Standard Deviation |
|---|---|---|
| Convergence Epoch | 35.0 | ±0.0 |
| Final Learning Rate | 1.3e-04 | 0.0e+00 |
| Training Time (min) | 0.0 | ±0.0 |

The training process demonstrated consistent performance across all folds. The model consistently converged at epoch 35 indicating stable learning behavior. The final learning rate settled at $1.3 \times 10^{-4}$ suggesting appropriate adaptation of the learning process. However, the training time measurement requires recalibration as it shows zero values due to timing resolution issues in the measurement implementation.

### 3.2. Evaluation Metrics

Table 7
Detailed Performance Metrics for Each Fold

| Fold | Training | | | Validation | | |
|---|---|---|---|---|---|---|
| | $R^2$ | MSE | RMSE | $R^2$ | MSE | RMSE |
| 1 | 0.9130 | 93.9716 | 9.6939 | 0.8891 | 93.2995 | 9.6592 |
| 2 | 0.9086 | 93.5255 | 9.6709 | 0.9678 | 38.8361 | 6.2319 |
| 3 | 0.8545 | 155.9636 | 12.4885 | 0.9016 | 91.5806 | 9.5698 |
| 4 | 0.9485 | 53.6709 | 7.3260 | 0.8664 | 142.1813 | 11.9240 |
| 5 | 0.9259 | 77.6471 | 8.8118 | 0.9623 | 42.5006 | 6.5192 |
| 6 | 0.9415 | 62.3041 | 7.8933 | 0.9622 | 35.9177 | 5.9931 |
| 7 | 0.9120 | 94.9813 | 9.7458 | 0.9102 | 75.1719 | 8.6702 |
| 8 | 0.9080 | 103.6660 | 10.1817 | 0.8536 | 51.7899 | 7.1965 |
| 9 | 0.9291 | 69.2784 | 8.3234 | 0.9737 | 47.8027 | 6.9139 |
| 10 | 0.9115 | 93.6422 | 9.6769 | 0.9668 | 34.3903 | 5.8643 |
| Mean | 0.9153 | 89.8651 | 9.3812 | 0.9254 | 65.3471 | 7.8542 |
| Std | ±0.0251 | ±27.1634 | ±1.4047 | ±0.0442 | ±34.2825 | ±1.9748 |

The model's performance metrics across all folds demonstrate strong predictive capability and consistency. The R-squared values show high accuracy in both training (0.9153 $\pm$ 0.0251) and validation (0.9254 $\pm$ 0.0442) sets indicating that the model explains approximately 92% of the variance in durian prices. The Mean Squared Error (MSE) and its square root (RMSE) values reveal better performance in the validation set (65.3471 $\pm$ 34.2825 of MSE, 7.8542 $\pm$ 1.9748 of RMSE) compared to the training set (89.8651 $\pm$ 27.1634 of MSE, 9.3812 $\pm$ 1.4047 of RMSE) suggesting good generalization without overfitting.

### 3.3. Future Price Prediction

The price predictions for the durian harvest season (April-June) of 2025 are generated using the trained BiLSTM model. The model predicts significant price fluctuations following seasonal patterns observed in the historical data.

Table 8
Selected Price Predictions for Key Periods in 2025

| Period | Date Range | Predicted Price Range (THB) |
|---|---|---|
| Early Season | April 26-30 | 156.19 - 156.29 |
| Price Decline | May 5-10 | 152.34 - 149.85 |
| Mid-Season Stability | May 27-June 1 | 170.37 - 170.45 |
| Peak Season | June 25-30 | 263.14 - 266.40 |

The predicted price trajectory shows distinctive patterns throughout the harvest season in Table 8. The period begins with a phase of initial stability in late April where prices maintain a steady level around 156 THB. Moving into early May, the market experiences a gradual decline with prices decreasing to their lowest point at 149.85 THB. As the season progresses to late May, prices enter a mid-season plateau phase, stabilizing around 170 THB. The most significant price movement occurs during the peak season

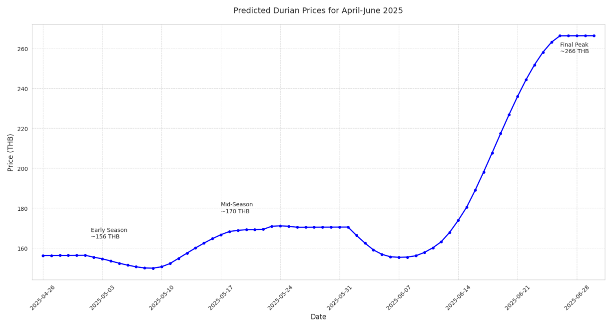



Fig. 5. Predicted Durian Price Trends for April-June 2025

in late June where prices surge dramatically from 155 THB in early June to reach their highest point at approximately 266 THB.

Fig.5 shows that these predictions suggest a significant upward trend in durian prices throughout the 2025 season with the most substantial increases occurring in late June. However, these forecasts should be interpreted with consideration of external factors such as weather conditions, market dynamics and potential economic changes that could influence actual price movements.

## 4. Discussions

The results from this study offer several significant insights into both the technical implementation and practical implications of using Bidirectional LSTM for durian price prediction. The model's performance, demonstrated by high R-squared values ($0.9153 \pm 0.0243$ for training, $0.9254 \pm 0.0439$ for validation) which indicates robust predictive capabilities. The close alignment between training and validation metrics with only a 1.01% difference, suggests effective generalization without overfitting. This performance stability is particularly noteworthy given the complex nature of agricultural price data. The marginally better performance on the validation set indicates that the model successfully captures underlying price patterns rather than merely memorizing training data.

The model demonstrates particular strength in capturing seasonal patterns and market dynamics. The predicted price trajectory shows four distinct phases such as initial stability, early decline, mid-season plateau and peak season surge that align with known market behaviors in the Thai durian industry. This capability is enhanced by the bidirectional architecture which enables the model to learn both forward and backward temporal dependencies. The accurate prediction of these market phases provides valuable insights for stakeholders across the durian supply chain. The model's ability to identify price stability periods and transition points is particularly relevant for market participants making strategic decisions.

The practical implications of these findings extend across various stakeholder groups in the durian industry. Farmers can utilize the predictions to optimize harvest timing and marketing decisions, potentially maximizing their returns by aligning harvest periods with predicted price peaks. Traders benefit from improved understanding of expected price stability periods and transition points enabling more effective inventory management and contract negotiations. Market regulators can leverage these predictions for more informed market intervention planning and formulation.

However, several limitations of the current study should be acknowledged. The model's reliance on historical data from 2023 to 2024 may limit its ability to capture long-term market evolution patterns. Additionally, while the model shows strong performance in predicting price movements, it does not explicitly account for external factors such as weather conditions and global market dynamics that could significantly impact prices. The 25-day prediction window though effective for current patterns might require adjustment for capturing longer-term market trends.

Future research directions could address these limitations by incorporating additional features such as weather data and international market indicators extending the temporal scope of the training data and developing ensemble approaches that combine multiple prediction methods. Such enhancements could further improve the model's robustness and practical utility in real world applications. The integration of external factors and broader market indicators could provide a more comprehensive framework for understanding and predicting durian price movements.

## 5. Conclusions

This research presents a novel approach to durian price prediction by implementing a Bidirectional LSTM model enhanced with seasonal analysis. The study makes several significant contributions to both the methodological and practical aspects of agricultural price forecasting. The primary achievements can be summarized in three key areas.

First, the proposed Bidirectional LSTM architecture demonstrates exceptional predictive accuracy achieving R-squared values of $0.9254 \pm 0.0439$ on validation data. This high performance validates the effectiveness of combining bidirectional processing with seasonal feature engineering for capturing complex price patterns. The model's ability to maintain consistent performance across multiple cross-validation folds indicates its robustness and reliability.

Second, the study successfully addresses the critical challenge of missing data in agricultural price series through an innovative application of the Prophet model. The implemented imputation strategy, with a maximum price constraint of 260 THB, preserves essential market dynamics while maintaining economic plausibility. This approach provides a practical solution for handling data gaps common in agricultural market data.

Third, the developed model offers practical value for various stakeholders in the durian industry. The accurate prediction of seasonal price patterns including the identification of key market phases from April to June. It provides actionable insights for farmers, traders and market regulators. The model's ability to forecast price movements with high precision can significantly improve decision-making in harvest timing, inventory management and market planning.

The successful implementation of this model demonstrates the potential of deep learning approaches in agricultural price forecasting. Future research could extend this work by incorporating additional external factors and exploring longer prediction horizons. This study contributes to the growing body of research on agricultural price prediction and provides a robust framework for similar applications in other agricultural commodities.

## References

[1] S. Arslan, A hybrid forecasting model using LSTM and Prophet for energy consumption with decomposition of time series data, *PeerJ Computer Science* **8** (2022), e1001.

[2] P. Bhuthong, P. Rattanasimanon and N. Udomkiattikul, AGRICULTURAL EXPORTS FROM THAILAND TO CHINA, *Agricultural Trade between China and the Greater Mekong Subregion Countries: A Value Chain Analysis* (2022), 59.

[3] T.L. Dean and D.V. McDermott, Temporal data base management, *Artificial intelligence* **32**(1) (1987), 1–55.

[4] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago and O. Tabona, A survey on missing data in machine learning, *Journal of Big data* **8** (2021), 1–37.

[5] J. Honaker and G. King, What to do about missing values in time-series cross-section data, *American journal of political science* **54**(2) (2010), 561–581.

[6] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu and H. Zhang, Deep learning with long short-term memory for time series prediction, *IEEE Communications Magazine* **57**(6) (2019), 114–119.

[7] Y. Imrana, Y. Xiang, L. Ali and Z. Abdul-Rauf, A bidirectional LSTM deep learning approach for intrusion detection, *Expert Systems with Applications* **185** (2021), 115524.

[8] N. Jaemjarus, K. Rattanamanee, K. Bhanityanakorn and M. de Correa, The Guidelines for Developing Competitiveness of Thai Agricultural Products: A Case Study of Durian Production in 14 Southern Provinces, Thailand, *Revista de Gestão Social e Ambiental* **18**(3) (2024), e04702–e04702.

[9] L. Kumar, S. Khedlekar and U. Khedlekar, A comparative assessment of holt winter exponential smoothing and autoregressive integrated moving average for inventory optimization in supply chains, *Supply Chain Analytics* **8** (2024), 100084.

[10] L. Kumar, The Role of Agriculture in Economic Development, *Journal of Environmental Agriculture and Agroecosystem Management* **1**(1) (2024), 62–68.

[11] L. Lara-Estrada, L. Rasche, L.E. Sucar and U.A. Schneider, Inferring missing climate data for agricultural planning using Bayesian networks, *Land* **7**(1) (2018), 4.

[12] H. Li, M. Li, X. Lin, F. He and Y. Wang, A spatiotemporal approach for traffic data imputation with complicated missing patterns, *Transportation research part C: emerging technologies* **119** (2020), 102730.

[13] P. Likhitpichitchai, J. Khermkhan and S. Kuhaswonvetch, The competitiveness of Thai durian export to China's market, *GMSARN International Journal. 17 (4): 371* **378** (2023).

[14] Y. Liu, T. Dillon, W. Yu, W. Rahayu and F. Mostafa, Missing value imputation for industrial IoT sensor data with large gaps, *IEEE Internet of Things Journal* **7**(8) (2020), 6855–6867.

[15] L. Ren, T. Wang, A.S. Seklouli, H. Zhang and A. Bouras, A review on missing values for main challenges and methods, *Information Systems* (2023), 102268.

[16] D.A. Rusdah and H. Murfi, XGBoost in handling missing values for life insurance risk prediction, *SN Applied Sciences* **2**(8) (2020), 1336.

[17] V. Saiz-Rubio and F. Rovira-Más, From smart farming towards agriculture 5.0: A review on crop data management, *Agronomy* **10**(2) (2020), 207.

[18] G.D. Santangelo, The impact of FDI in land in agriculture in developing countries on host country food security, *Journal of World Business* **53**(1) (2018), 75–84.

[19] F. Tang and H. Ishwaran, Random forest missing data algorithms, *Statistical Analysis and Data Mining: The ASA Data Science Journal* **10**(6) (2017), 363–377.

[20] T. Van Hau, D.H. Tien, N.M. Thuy, H. Ky, T.T.O. Yen, M. Van Tri, N. Van Hoa and T.S. Hieu, Durian, in: *Tropical and Subtropical Fruit Crops*, Apple Academic Press, 2023, pp. 161–200.

[21] X. Wang and C. Wang, Time series data cleaning: A survey, *Ieee Access* **8** (2019), 1866–1881.

[22] N. Wongsai, S. Wongsai and A.R. Huete, Annual seasonality extraction using the cubic spline function and decadal trend in temporal daytime MODIS LST data, *Remote Sensing* **9**(12) (2017), 1254.

[23] R. Wu, S.D. Hamshaw, L. Yang, D.W. Kincaid, R. Etheridge and A. Ghasemkhani, Data imputation for multivariate time series sensor data with large gaps of missing data, *IEEE Sensors Journal* **22**(11) (2022), 10671–10683.

[24] C. Yozgatligil, S. Aslan, C. Iyigun and I. Batmaz, Comparison of missing value imputation methods in time series: the case of Turkish meteorological data, *Theoretical and applied climatology* **112** (2013), 143–167.

[25] W. Zhai, X. Kuang, X. Cheng, J. Pan and C. Wu, Reconstruction of missing points in agricultural machinery trajectory based on bidirectional adjacent information, *Computers and Electronics in Agriculture* **220** (2024), 108920.