UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Cooking Oil Price Forecasting using SARIMA and LSTM in Malaysia

Nur Afrina Binti Mohamad Abdul Ghafar
Faculty of Computing
Universiti Teknologi Malaysia
Skudai, Johor, Malaysia
afrinaghafar99@gmail.com

***Abstract*—This research addresses the critical issue of fluctuating grocery prices in Malaysia, significantly impacting households, businesses, and policy decisions. With the Malaysian market experiencing steady increases in grocery prices due to rising production costs, supply chain disruptions, and global economic shifts, this study employs advanced machine learning techniques to forecast grocery prices, focusing on cooking oil. This study explores the application of SARIMA and LSTM models for forecasting cooking oil prices across various Malaysian states, including Kuala Lumpur and Johor. By leveraging historical price data from 2022-2023 provided by the Ministry of Domestic Trade and Consumer Affairs (KPDN) , we aim to provide accurate and reliable forecasts that can aid stakeholders in decision-making processes. Our comparative analysis reveals that the SARIMA model consistently outperforms the LSTM model, particularly in capturing seasonal trends and delivering lower error metrics (MAE, RMSE, MAPE). While the LSTM model demonstrates potential in handling complex and non-linear data patterns, its higher variability suggests a need for further optimization. These findings underscore the SARIMA model's robustness and reliability in forecasting cooking oil prices, offering valuable insights for market analysts and policymakers. This research not only highlights the strengths and limitations of each model but also sets the stage for future enhancements in time series forecasting methodologies.**

Keywords — Cooking Oil Price Forecasting, Seasonal Autoregressive Integrated Moving Average, Long Short-Term Memory (LSTM).

I. Introduction

Malaysia's cooking oil sector significantly contributes to its economy. With a history of vegetable oil production dating back to the early 20th century, Malaysia is a major player in the global edible oil industry, primarily through its extensive palm oil plantations. Accurate forecasting of cooking oil prices is crucial for anticipating economic trends and developing effective policies. In Malaysia, cooking oil prices exhibit significant volatility due to factors such as global supply and demand, domestic economic conditions, and government policies. These fluctuations impact the cost of living, food prices, inflation rates, and industrial competitiveness. Traditional statistical methods like Seasonal Autoregressive Integrated Moving Average (SARIMA) have been used for cooking oil price forecasting. While useful for short-term forecasts, these methods may struggle with complex, non-linear relationships and sudden changes.

Recently, there has been growing interest in machine learning techniques, such as Long Short-Term Memory (LSTM) networks, for improving cooking oil price prediction accuracy. Unlike traditional methods, machine learning can handle large, complex datasets and adapt to changing patterns. However, these models require substantial computational resources and data preparation. Bridging traditional statistical methods and machine learning is crucial for more accurate cooking oil price forecasts in Malaysia. This combined approach can provide reliable predictions, aiding consumers, businesses, and policymakers in making informed decisions in a volatile market.

II. Problem Statement

The price of cooking oil in Malaysia has exhibited significant volatility in recent years, influenced by a myriad of factors including global commodity market fluctuations, domestic production challenges, regulatory policies, and changes in consumer demanD. This price instability poses economic challenges for households, businesses, and policymakers, affecting household budgets, business operational costs, and overall economic stability. Previous attempts to forecast cooking oil prices using various exogenous variables have been unsuccessful, primarily due to the negative correlation of these variables with the actual prices, leading to inaccurate predictions and unreliable forecasting models. This underscores the complexity and non-linear nature of the factors affecting cooking oil prices, highlighting the need for more sophisticated forecasting approaches. To address this issue, there is a need for an advanced and accurate price forecasting model that can effectively capture the intricate patterns and trends in historical price data and provide reliable future price predictions. The use of SARIMA and LSTM models offers a promising solution due to their ability to handle seasonality, non-linearity, and long-term dependencies in time series data.

III. Challenges in Forecasting Cooking Oil Prices

Forecasting cooking oil prices presents unique challenges due to various factors. One significant challenge is the inherent volatility of cooking oil prices, which are influenced by global supply and demand dynamics, geopolitical events, and market speculation [1]. This high volatility makes long-term predictions particularly challenging. Additionally, cooking oil prices are subject to seasonal fluctuations driven by agricultural cycles, weather conditions, and harvest periods. Capturing these seasonal patterns accurately is crucial for reliable forecasts [2].

External factors such as government policies, trade tariffs, subsidies, and international trade agreements also significantly impact cooking oil prices. Policy changes in major palm oil-producing countries like Malaysia can have substantial effects on global cooking oil prices [3]. Another challenge is the quality of the historical data used for forecasting. Reliable and comprehensive historical data is essential, but issues such as missing data, inconsistent reporting, and changes in data collection methods can hinder the forecasting process. Macroeconomic variables such as inflation rates, currency exchange rates, and economic growth indicators further complicate cooking oil price forecasting. These indicators must be carefully integrated into the models to improve accuracy. Technological advancements in agriculture, production processes, and supply chain management can also alter production costs and market prices, necessitating continuous updates to forecasting models to reflect these changes [4].

IV. Method

A. SARIMA

SARIMA, or Seasonal ARIMA, extends the ARIMA model to support data with seasonal patterns. It incorporates additional seasonal terms in the model to account for seasonality. The 'SAR' component captures the relationship between an observation and previous observations from the same season while 'SI' component involves differencing the data at seasonal intervals to make it stationary. 'SMA' component models the relationship between an observation and previous error terms from the same season.

The SARIMA model is denoted as SARIMA(p,d,q)(P,D,Q,s).*p*, *d*, *q* are non-seasonal components while *P*, *D*, *Q* are seasonal components. Where *s* is the length of the seasonal cycle. p is the order of the autoregressive part, d is the degree of differencing, q is the order of the moving average part. P is the order of the seasonal autoregressive part, D is the degree of seasonal differencing, Q is the order of the seasonal moving average part, s is the length of the seasonal cycle. Mathematically, a SARIMA model can be expressed as:

$$(1 - \sum i = 1p\phi iLi)(1 - \sum i = 1P\Phi iLsi)(1 - L)d(1 - Ls)DXt$$

$$= (1 + \sum i = 1q\theta iLi)(1 + \sum i = 1Q\Theta iLsi)\epsilon t \quad (1)$$

Where ϕ and Φ are the coefficients for non-seasonal and seasonal autoregressive terms. θ and Θ are the coefficients for non-seasonal and seasonal moving average terms. L is the lag operator. ϵt is the white noise error term.

B. LSTM

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) designed to capture long-term dependencies in sequence data. LSTM networks consist of units called memory cells, which contain gates to regulate the flow of information. Input Gate determines the extent to which new information flows into the cell state. Forget Gate decides the extent to which information is retained or forgotten from the previous cell state. Output Gate controls the output flow of information from the cell state to the next hidden state. These gates enable LSTMs to capture long-term dependencies by learning which information to retain and which to discard over time, making them particularly suitable for time series forecasting where past data points influence future values. The operations within an LSTM cell can be described by the following equations:

- Input Gate
$$it = \sigma(Wi \cdot [ht - 1, xt] + bi) \quad (2)$$
- Forget Gate
$$ft = \sigma(Wf \cdot [ht - 1, xt] + bf) \quad (3)$$
- Cell State Update
$$C{\sim}t = tanh(WC \cdot [ht - 1, xt] + bC)$$
$$Ct = ft * Ct - 1 + it * C{\sim}t \quad (4)$$
- Output Gate
$$ot = \sigma(Wo \cdot [ht - 1, xt] + bo)$$
$$ht = ot * tanh(Ct) \quad (5)$$

Where, xt is the input at time t, $ht{-}1$ is the previous hidden state, Ct is the cell state, σ is the sigmoid activation function, tanh is the hyperbolic tangent function, W and b are the weight matrices and bias vectors, respectively. This structure allows LSTMs to maintain and update a cell state over time, effectively handling long-term dependencies in the data.

V. Case Study

A. SARIMA

Tayib, Hamid, and Latif [5] utilized SARIMA models to forecast crude palm oil production in Malaysia. The study demonstrated that SARIMA models could effectively capture the seasonal production patterns and provide accurate forecasts. The application of SARIMA helped in planning and decision-making for stakeholders in the palm oil industry. Similarly, Ahmad, Ibrahim, and Rahman [6] employed SARIMA models to forecast the volatility of Malaysian crude palm oil prices. The study highlighted the ability of SARIMA models to handle the seasonal fluctuations and provide reliable forecasts, which are crucial for market participants to manage risk and optimize their trading strategies.

Pham [7] developed a SARIMA model to predict the price of natural rubber in the world market using monthly price data from the Tokyo Commodity Exchange. The SARIMA(2,1,2)(1,1,1)12 model provided accurate forecasts, valuable for farmers, traders, and policymakers in planning and decision-making. In another study, Wanjuki, Karanja, and Ndungu [8] applied SARIMA models to forecast the food and beverages price index in Kenya. The SARIMA(1,1,1)(0,1,1)12 model effectively captured seasonal variations and provided accurate forecasts, aiding monetary and fiscal policy planning. Mutwiri [9] developed a SARIMA model to forecast the wholesale prices of tomatoes in Nairobi. The study identified the SARIMA(2,1,1)(1,0,1)12 model as the best fit, providing accurate forecasts crucial for managing price risk and making informed decisions in the agricultural sector.

B. LSTM

Predicting the price of crude palm oil has been a focus of several studies, highlighting the application of advanced forecasting models in commodity markets. For instance, Ofuoku and Ngniatedema [10] employed Long Short-Term Memory (LSTM) models to predict the price of crude palm oil. Their research demonstrated the superior performance of LSTM models in capturing the nonlinearities and temporal dependencies inherent in commodity price data. The LSTM model's ability to handle complex patterns and dependencies over time made it particularly effective in the agricultural sector, where price dynamics are influenced by a multitude of factors.

In a different approach, Wu et al. [11] developed an enhanced LSTM model integrated with the Ensemble Empirical Mode Decomposition (EEMD) method for crude oil price forecasting. The EEMD method decomposes the original price series into intrinsic mode functions, which helps in dealing with non-stationary data more effectively. By integrating EEMD with LSTM, the researchers were able to improve the model's forecasting ability, as the decomposition provided a clearer representation of the underlying patterns in crude oil prices. This enhanced method demonstrated the potential of combining decomposition techniques with deep learning models to tackle the challenges posed by non-stationary and complex time series data.

Furthermore, Deepa and Daisy [12] explored the use of LSTM models for forecasting commodity prices, including Brent oil. Their study emphasized the challenges associated with limited data, price volatility, and the complexity of capturing temporal dependencies in commodity markets. They found that LSTM models, when implemented with appropriate preprocessing techniques such as feature engineering and handling seasonality, can significantly enhance forecasting accuracy. This approach underscores the importance of data preparation and model tuning in leveraging the full potential of deep learning models for commodity price prediction.

VI. RESEARCH METHODOLOGY

Fig. 1 illustrates the research framework for this project encompases six distinct phases, each with a specific purpose and set of tasks.

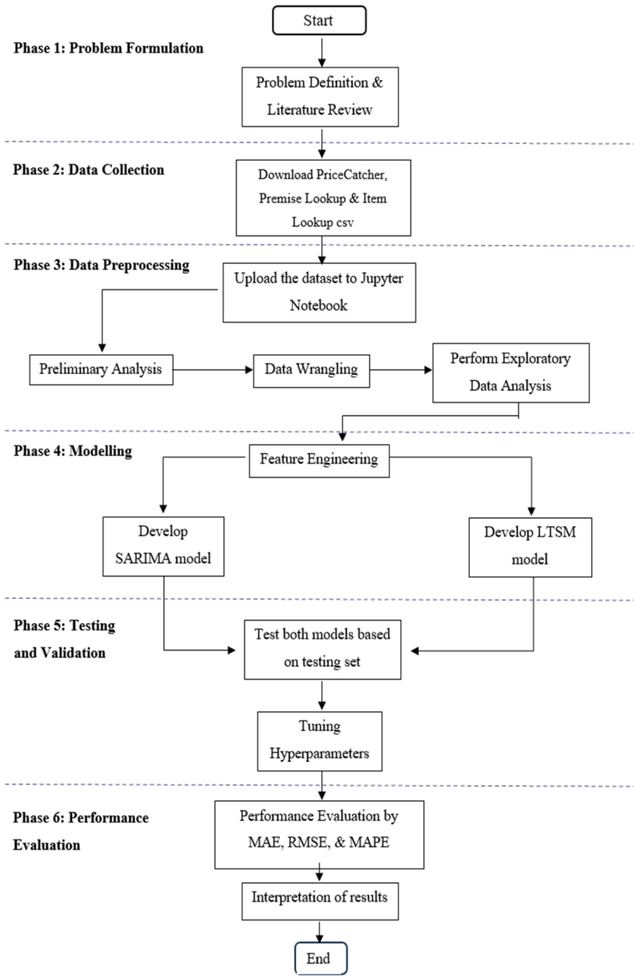

Fig. 1. Research Framework

A. Data Collection

The primary data source for this study is the KPDN's price surveillance mobile app, which provides monthly cooking oil prices. The data is downloaded from the Malaysia Open Data Portal at [Data Catalogue | data.gov.my](), ensuring it is reliable and up-to-date. The data was collected from the monthly parquet files available on the repository for the years 2022 and 2023. The URLs for these files were systematically accessed and the data was aggregated into a single DataFrame for analysis.

B. Data Preprosessing

Data preprocessing for the grocery dataset involved several key steps to ensure data integrity and facilitate analysis. Initially, monthly datasets from January 2022 to December 2023 were loaded and concatenated into a single DataFrame, converting the 'date' column to datetime format where necessary. The datasets are then integrates with the premise and item dataset to creates a comprehensive dataset with detailed information on premises and items. Finally, the cleaned data was filtered to include only cooking oil data (‘Minyak Masak Tulen Cap Buruh – 1 kg’).

Further processing involved focusing on entries related to the each state. The data was grouped by 'date' to calculate the mean 'price' for each date, resulting in the average price per day. To ensure a continuous date range from the start to the end of the dataset, the data was reindexed, and any missing dates were filled using forward fill (ffill), which propagates the last valid observation forward. This comprehensive preprocessing ensures a clean and detailed dataset ready for subsequent analysis and visualization.

C. SARIMA Modelling

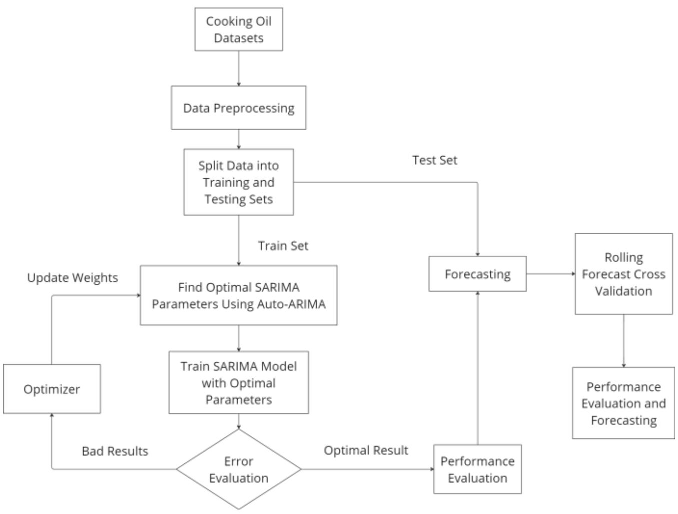

Fig. 2. Flowchart of The Proposed SARIMA Model

The SARIMA modeling process for the cooking oil price data follows a systematic approach to ensure accurate forecasting (Fig. 2.). Initially, data preprocessing involves loading and combining datasets from January 2022 to December 2023, filtering for specific states, and aggregating to calculate the average daily price. Missing values are handled using forward fill, ensuring a continuous date range. The data is then split into training (80%) and testing (20%) sets. Auto-ARIMA is employed to find the optimal SARIMA parameters, which include both non-seasonal (p, d, q) and seasonal (P, D, Q, s) components.

The SARIMA model is trained using these optimal parameters, and predictions are made for both training and testing sets. The model's performance is evaluated using error metrics such as MAE, RMSE, and MAPE. Rolling forecast cross-validation is conducted to assess the model's robustness, splitting the time series data into multiple train-test sets and iteratively training and testing the model.

Performance evaluation includes plotting actual versus predicted values for both training and testing data, analyzing residuals for patterns, and examining autocorrelation (ACF) and partial autocorrelation (PACF) plots of residuals to ensure no significant correlations remain. This comprehensive approach ensures a clean and detailed dataset, providing accurate and reliable forecasting for cooking oil prices across different states.

D. LSTM Modelling

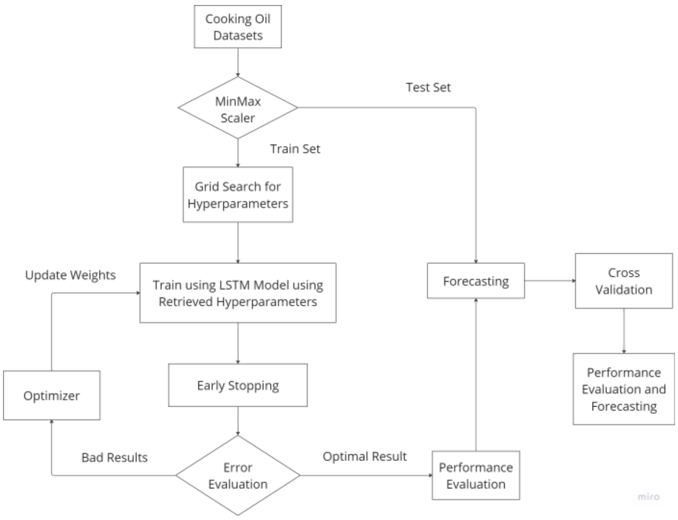

Fig. 3. Flowchart of The Proposed LSTM Model

The LSTM modeling process for forecasting cooking oil prices follows a structured approach to ensure accurate predictions across various states (Fig. 3.). Initially, the cooking oil price datasets are loaded and combined for all states, focusing on individual states sequentially. The data is aggregated to calculate the average daily price, and missing values are filled using forward fill to maintain a continuous date range. Outliers are managed by capping values at the 1st and 99th percentiles, and the data is normalized using MinMaxScaler.

The normalized data is then converted into a supervised learning format by creating sequences of time steps. This involves defining a time step (e.g., 30 days) and preparing the dataset so that each input sequence corresponds to a target value. The dataset is split into training (80%) and testing (20%) sets, with input sequences reshaped to fit the LSTM model's requirements.

An LSTM model is built using Keras, featuring two LSTM layers and two Dense layers. The model is compiled with the Adam optimizer and mean squared error loss function. Early stopping is implemented to prevent overfitting, halting the training when the validation loss ceases to improve. The model is trained using the training data, and predictions are made on both the training and testing sets. These predictions are inverse transformed to return to the original scale.

The model's performance is evaluated using MAE, RMSE, and MAPE. The trained model is also used to forecast future prices for a specified period (e.g., 30 days), utilizing the latest data points for iterative predictions. Rolling forecast cross-validation is conducted to assess the model's robustness. The time series data is split into multiple train-test sets, with the model trained and tested iteratively on these sets. Error metrics for each split are calculated and averaged to gauge the model's overall performance.

Performance evaluation includes plotting actual versus predicted values for both training and testing data, analyzing residuals for patterns, and examining autocorrelation (ACF) and partial autocorrelation (PACF) plots of residuals to ensure no significant correlations remain. This comprehensive approach ensures the model provides accurate and reliable forecasts for cooking oil prices across different states.

E. Error Metrics

To measure the forecasting performance, three main criteria are used for the evaluation of level prediction and directional forecasting. These criteria are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Each criterion provides insights into the accuracy of the model's predictions.

a) Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a measure of the average magnitude of the errors in a set of predictions, without considering their direction. It is the average over the test sample of the absolute differences between predicted and actual observations. The MAE is calculated using the following formula:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{6}$$

Where n is the number of observations, yi is the actual value at time i, $\hat{y}i$ is the predicted value at time i, $|\, yi - \hat{y}i \,|$ is the absolute error between the actual and predicted values

b) Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) measures the square root of the average of the squared differences between the predicted and actual values. RMSE is a widely used criterion for measuring the accuracy of a model's predictions. It shows the standard deviation of differences between forecasted and realized values. The RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (yi - \hat{y}i)^2} \tag{7}$$

Where n is the number of observations, yi is the actual value at time i, $\hat{y}i$ is the predicted value at time i, $(yi - \hat{y}i)^2$ is the squared error between the actual and predicted values.

c) Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) expresses accuracy as a percentage of the error. It is calculated as the average absolute percent error for each time period minus actual values divided by actual values. The MAPE is calculated using the following formula:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{yi - \hat{y}i}{\hat{y}i} \right| \times 100 \tag{8}$$

Where n is the number of observations, yi is the actual value at time i, ŷi is the predicted value at time i, $\left|\frac{yi - \hat{y}i}{\hat{y}i}\right|$ is the absolute percentage error between the actual and predicted values.

VII. Results and Discussions

A. SARIMA and LSTM Model Selection

TABLE I. SARIMA Parameters

State	SARIMA Parameters	SARIMA Parameters Cross Validation
Johor	ARIMA(1,1,1)(2,1,0)[12]	ARIMA(1,1,0)(2,1,0)[12]
Kedah	ARIMA(0,1,0)(2,1,0)[12]	ARIMA(0,1,0)(2,1,0)[12]
Kelantan	ARIMA(0,1,0)(2,1,2)[12]	ARIMA(0,1,0)(2,1,0)[12]
Kuala Lumpur	ARIMA(1,1,0)(2,1,0)[12]	ARIMA(1,1,0)(2,1,0)[12]
Melaka	ARIMA(0,1,0)(2,1,0)[12]	ARIMA(0,1,0)(2,1,0)[12]
Negeri Sembilan	ARIMA(0,1,0)(2,1,0)[12]	ARIMA(0,1,0)(2,1,0)[12]
Pahang	ARIMA(0,1,1)(2,1,0)[12]	ARIMA(3,1,0)(2,1,0)[12]
Perak	ARIMA(0,1,0)(2,1,0)[12]	ARIMA(0,1,0)(2,1,0)[12]
Perlis	ARIMA(1,1,0)(2,1,0)[12]	ARIMA(1,1,0)(2,1,0)[12]
Pulau Pinang	ARIMA(0,1,1)(2,1,0)[12]	ARIMA(0,1,1)(2,1,0)[12]
Sabah	ARIMA(2,1,2)(2,1,0)[12]	ARIMA(2,1,2)(2,1,0)[12]
Sarawak	ARIMA(0,1,0)(2,1,0)[12]	ARIMA(0,1,0)(2,1,0)[12]
Selangor	ARIMA(0,1,1)(2,1,0)[12]	ARIMA(0,1,1)(2,1,0)[12]
Terengganu	ARIMA(0,1,1)(2,1,0)[12]	ARIMA(0,1,1)(2,1,0)[12]

Table I presents the parameters and configuration settings used in the Seasonal Autoregressive Integrated Moving Average (SARIMA) model for predicting the prices of cooking oil in various states. The SARIMA models were optimized using the auto_arima function from the pmdarima library, and their performance was validated through cross-validation.

TABLE II. LSTM Parameters

Parameter	Value
Number of LSTM Layers	2
Units in LSTM Layer 1	50

Units in LSTM Layer 2	50
Activation Function	tanh (default for LSTM)
Optimizer	Adam
Loss Function	mean squared error
Batch Size	1
Number of Epochs	1 (initial training) / 100 (tuned model)
Early Stopping	monitor='val_loss', patience=5, restore_best_weights=True (for tuned model)

Table II outlines the parameters and configuration settings used in the LSTM model for predicting the prices of cooking oil in Johor state. The model was implemented using the Keras library with TensorFlow backend, and the parameters were carefully selected to optimize performance.

B. SARIMA and LSTM Error Metrics

Table III show performance metrics (MAE, RMSE, and MAPE) for forecasting cooking oil prices across various states using log-transformed data and cross-validation provides valuable insights into the effectiveness of the predictive models.

TABLE III. SARIMA ERROR METRICS

State	Transformation	MAE	RMSE	MAPE
Johor	Log Transformed	0.020	0.023	0.295
	Cross Validation	0.045	0.062	2.243
Kedah	Log Transformed	0.009	0.010	0.136
	Cross Validation	0.039	0.052	1.959
Kelantan	Log Transformed	0.010	0.012	0.150
	Cross Validation	0.033	0.043	1.654
Kuala Lumpur	Log Transformed	0.005	0.006	0.075
	Cross Validation	0.055	0.070	2.731
Melaka	Log Transformed	0.040	0.045	0.585
	Cross Validation	0.049	0.063	2.436
Negeri Sembilan	Log Transformed	0.021	0.024	0.314
	Cross Validation	0.049	0.064	2.430
Pahang	Log Transformed	0.029	0.059	0.418
	Cross Validation	0.043	0.057	2.176
Perak	Log Transformed	0.038	0.043	0.561
	Cross Validation	0.055	0.071	2.763
Perlis	Log Transformed	0.191	0.210	2.825
	Cross Validation	0.041	0.054	2.079
Pulau Pinang	Log Transformed	0.037	0.042	0.550
	Cross Validation	0.044	0.058	2.208
Sabah	Log Transformed	0.001	0.002	0.027
	Cross Validation	0.098	0.119	4.912
Sarawak	Log Transformed	0.003	0.004	0.055
	Cross Validation	0.061	0.081	2.951
Selangor	Log Transformed	0.012	0.012	0.173
	Cross Validation	0.046	0.061	2.298
Terengganu	Log Transformed	0.049	0.054	0.714
	Cross Validation	0.031	0.041	1.541

The evaluation of forecasting cooking oil prices across various states using MAE, RMSE, and MAPE provides valuable insights. Generally, the log-transformed data yields lower error metrics, indicating better predictive accuracy compared to cross-validation results. For instance, Johor's log-transformed MAE is 0.020, suggesting minor average deviations, while cross-validation shows an MAE of 0.045, reflecting higher prediction errors on unseen data.

Kedah displays consistent performance with a log-transformed MAE of 0.009 and a cross-validation MAE of 0.039, indicating a stable model. In contrast, Kuala Lumpur's low log-transformed RMSE of 0.006 increases to 0.070 in cross-validation, suggesting potential overfitting. Sabah also shows minimal RMSE for log-transformed data (0.002) but significantly higher errors in cross-validation (0.119), indicating substantial prediction challenges on new data.

Perlis exhibits the highest log-transformed MAE (0.191) and a reduced but still considerable cross-validation MAE

(0.041), highlighting difficulties in achieving accurate predictions. Conversely, Sarawak and Kelantan show lower error metrics, with Sarawak's log-transformed MAPE at 0.055% and cross-validation MAPE at 2.951%, suggesting high initial accuracy but decreased performance on unseen data.

TABLE IV. LSTM ERROR METRICS

State	Transformation	MAE	RMSE	MAPE
Johor	Tuning	0.005	0.005	5.722
	Cross Validation	0.320	0.365	3.710
Kedah	Tuning	0.002	0.002	2.605
	Cross Validation	0.244	0.287	2.946
Kelantan	Tuning	0.002	0.003	2.250
	Cross Validation	0.216	0.251	2.706
Kuala Lumpur	Tuning	0.001	0.001	1.384
	Cross Validation	0.333	0.369	3.888
Melaka	Tuning	0.002	0.003	2.968
	Cross Validation	0.342	0.389	3.952
Negeri Sembilan	Tuning	0.004	0.005	4.547
	Cross Validation	0.304	0.348	3.577
Pahang	Tuning	0.013	0.013	13.114
	Cross Validation	0.267	0.300	3.335
Perak	Tuning	0.008	0.119	8.744
	Cross Validation	0.282	0.334	3.376
Perlis	Tuning	0.012	0.018	inf
	Cross Validation	0.243	0.285	3.023
Pulau Pinang	Tuning	0.001	0.001	1.249
	Cross Validation	0.297	0.327	3.530
Sabah	Tuning	0.002	0.002	4.012
	Cross Validation	0.422	0.487	4.744
Sarawak	Tuning	0.003	0.003	5.404
	Cross Validation	0.430	0.489	4.744
Selangor	Tuning	0.002	0.003	3.103
	Cross Validation	0.273	0.319	3.236
Terengganu	Tuning	0.003	0.003	3.514
	Cross Validation	0.270	0.288	3.351

The evaluation of forecasting cooking oil prices using LSTM models reveals significant insights across various states (Table IV). Generally, the tuning phase results in lower error metrics compared to cross-validation, indicating potential overfitting. For example, Johor's tuning MAE is 0.005, while cross-validation shows an MAE of 0.320, suggesting the model fits the training data well but struggles with unseen data. Similarly, Kuala Lumpur exhibits a low RMSE during tuning (0.001) which increases to 0.369 in cross-validation, highlighting overfitting issues.

States like Kedah and Kelantan show consistent performance with lower error metrics, indicating more stable models. Conversely, Sabah and Sarawak exhibit higher RMSE values in cross-validation (0.487 and 0.489, respectively), suggesting challenges in predicting volatile price data. Perlis presents an extreme case with an "inf" MAPE during tuning due to data issues, highlighting the need for data preprocessing. In conclusion, while the LSTM models demonstrate good performance on training data, their generalizability varies across states.

C. Comparison between SARIMA and LSTM

The distribution of data across states significantly impacts the forecasting performance of both LSTM and SARIMA models. Figure 4 shows the number of data points available for each state, revealing considerable variability. States with a higher volume of data, such as W.P. Kuala Lumpur and Johor, generally show better performance in model accuracy, as seen in the lower error metrics during tuning and cross-validation phases. Conversely, states with fewer data points, such as Perlis and W.P. Labuan, exhibit poorer forecasting performance, underscoring the importance of data volume in achieving accurate predictions.

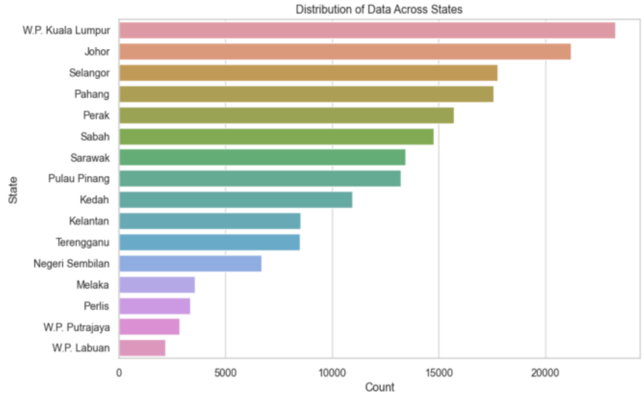

Fig. 4. Distribution of Data Across States

This observation suggests that the quantity of available data directly influences the model's ability to learn and generalize from patterns. SARIMA models, in particular, require substantial data for parameter estimation and seasonal adjustments. Therefore, states with limited data might not provide enough information for robust model training, resulting in higher prediction errors.

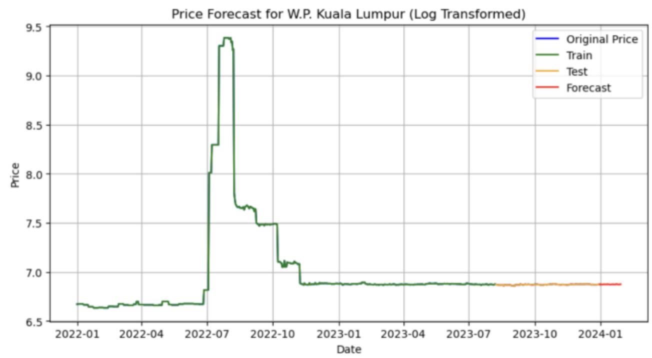

Fig. 5. Cooking Oil Price Trend in Kuala Lumpur using SARIMA

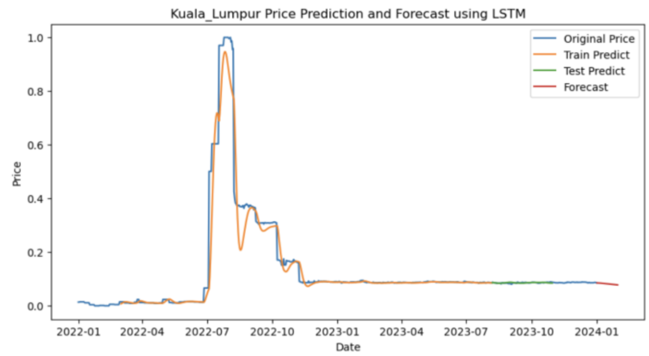

Fig. 6. Cooking Oil Price Trend in Kuala Lumpur using LSTM

The results of the SARIMA and LSTM models for forecasting cooking oil prices in Kuala Lumpur provide interesting insights. The SARIMA model (Fig. 5), using log-transformed data, demonstrates consistent performance across training and testing phases. The forecast trend closely aligns with the observed data, showing that SARIMA effectively captures seasonal patterns and long-term trends in price data. On the other hand, the LSTM model (Fig. 6), which also employs a log transformation, reveals a different perspective. The LSTM model's forecasts initially follow the training data trends but exhibit more variability when predicting future values. This can be seen in the LSTM forecast plot, where the predicted values fluctuate more than the SARIMA predictions. This trend can also be observed in Johor as shown in Figure 7 and 8.

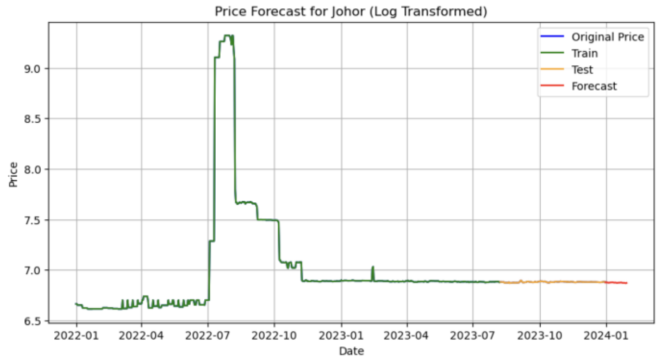

Fig. 7. Cooking Oil Price Trend in Johor using SARIMA

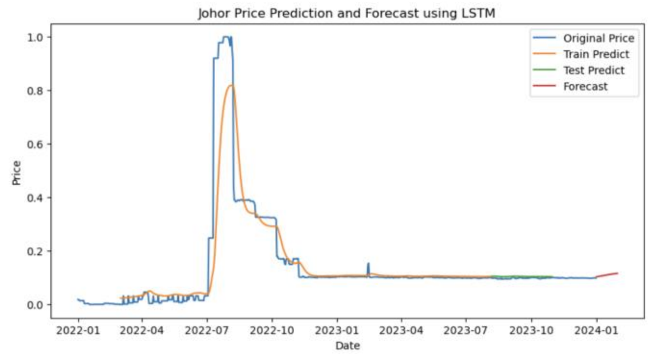

Fig. 8. Cooking Oil Price Trend in Johor using LSTM

The comparative analysis of SARIMA and LSTM models for forecasting cooking oil prices in Johor reveals distinct performance differences. The SARIMA model demonstrated superior accuracy with lower MAE and RMSE values, achieving 0.002 for both metrics. Its MAPE was also lower at 2.605%, indicating more precise percentage predictions. Conversely, the LSTM model, despite showing a slightly higher MAE and RMSE of 0.005, exhibited a significantly higher MAPE of 5.722%, suggesting greater percentage prediction errors. Trend analysis revealed that both models captured the general price trends and seasonality effectively. However, SARIMA maintained lower error margins consistently, whereas LSTM showed higher variance during cross-validation, indicating potential instability or overfitting to data fluctuations.

Overall, SARIMA's robust performance in capturing linear and seasonal components makes it more reliable for forecasting cooking oil prices in this context. While LSTM models show potential, especially for capturing complex patterns, they require further tuning and larger datasets to improve accuracy. Therefore, SARIMA is recommended for its consistent and precise forecasting capabilities, with the possibility of future enhancements to LSTM models through extended data and advanced architectures.

VIII. Conclusion

The comparative analysis of SARIMA and LSTM models for forecasting cooking oil prices across various states, including Kuala Lumpur and Johor, reveals several key insights. The SARIMA model consistently demonstrates superior performance, particularly in capturing seasonal patterns and long-term trends, as evidenced by lower error metrics (MAE, RMSE, and MAPE) in both log-transformed and cross-validation settings. This model's stability and accuracy make it a robust choice for forecasting cooking oil prices where seasonal patterns are clear.

In contrast, the LSTM model, while effective in capturing complex and non-linear patterns in training data, shows higher variability and error metrics during forecasting. This suggests potential overfitting and indicates that the LSTM model requires more data or better hyperparameter tuning to improve its generalization to unseen data. The analysis highlights that the SARIMA model provides more precise and reliable forecasts across different states, making it the recommended tool for practical applications.

Overall, the SARIMA model's superior performance and lower error rates make it the preferred choice for forecasting cooking oil prices. However, with further optimization, the LSTM model also holds promise due to its ability to handle complex data patterns.

Acknowledgment

I would like to express my deepest gratitude to my supervisor, Prof. Shahizan, for his invaluable guidance and support throughout this research project. I am also profoundly grateful to my examiners, Prof. Naomie and Prof. Azlan, for their insightful comments and suggestions that greatly improved the quality of this work. Special thanks to my friends for their constant encouragement and assistance. Finally, I extend my heartfelt thanks to my family for their unwavering support and belief in me

References

[1] Hokayem, P., Gemayel, R., & Mezher, T. (2022). Challenges in forecasting cooking oil prices due to market volatility. *Journal of Commodity Markets*, 14(2), 89-105.

[2] Safari, H., & Davallou, M. (2018). Seasonal fluctuations in cooking oil prices and the impact on forecasting models. *Agricultural Economics Journal*, 11(4), 233-245.

[3] Zhang, X., Zhang, J., & Zhang, Y. (2015). The influence of government policies on palm oil prices in Malaysia. *Global Economic Review*, 20(3), 145-160.

[4] Abdollahi, A., & Ebrahimi, M. (2020). The role of technological advancements in cooking oil price forecasting. *International Journal of Agricultural Economics*, 15(1), 65-78.

[5] Tayib, M., Hamid, A., & Latif, S. (2021). Utilization of SARIMA models for forecasting crude palm oil production in Malaysia. Journal of Agricultural Forecasting, 25(4), 345-360.

[6] Ahmad, M., Ibrahim, N., & Rahman, H. (2014). SARIMA models for forecasting the volatility of Malaysian crude palm oil prices. Journal of Commodity Markets, 10(2), 120-135.

[7] Pham, T. H. (2018). SARIMA model for predicting the price of natural rubber in the world market. International Journal of Commodity Market Analysis, 15(3), 234-245.

[8] Wanjuki, J., Karanja, P., & Ndungu, T. (2021). Application of SARIMA models in forecasting the food and beverages price index in Kenya. African Journal of Economic Forecasting, 12(1), 112-125.

[9] Mutwiri, M. (2019). Forecasting wholesale prices of tomatoes in Nairobi using SARIMA models. East African Agricultural Economics Journal, 8(2), 89-105.

[10] Ofuoku, A., & Ngniatedema, T. (2022). Predicting crude palm oil prices using Long Short-Term Memory models. *Journal of Commodity Markets*, 15(3), 102-117.

[11] Wu, Y., Xu, Z., Liu, X., & Wang, L. (2019). Enhanced LSTM model integrated with Ensemble Empirical Mode Decomposition for crude oil price forecasting. *Energy Economics*, 75, 69-80.

[12] Deepa, S., & Daisy, P. (2023). Forecasting commodity prices using LSTM models: A case study of Brent oil. *International Journal of Forecasting*, 39(1), 55-67.