Advanced Machine Learning Techniques for Accurate Forecasting of Crude Palm Oil Price

Muhammad Azrull Amin Ismail[1], Amirul Afif Muhamat*[2], Norliza Che Yahya*[3]

[1]Faculty of Business and Management,
Universiti Teknologi MARA
Shah Alam, Malaysia

[1]Dealing Department, TA Futures Sdn Bhd
Kuala Lumpur, Malaysia

[2,3]Climate Risk & Sustainable Finance RIG
Faculty of Business and Management
Universiti Teknologi MARA
Selangor, Malaysia

***Corresponding author: amirulafif@uitm.edu.my**

Abstract

The accurate forecasting of crude palm oil (CPO) prices is of paramount importance to stakeholders across the agricultural and financial sectors, as it directly influences critical decisions related to production, trading, and investment strategies. Traditional time series models, while valuable, often fall short in capturing the intricate, non-linear dynamics inherent in CPO price fluctuations. This research delves into the application of cutting-edge machine learning techniques, with a particular emphasis on state-of-the-art models like transformers and hybrid architectures, to significantly enhance the precision of CPO price predictions. This study provides a comprehensive overview of existing research on traditional CPO forecasting methodologies, while also exploring the promising potential of machine learning applications in this domain. By critically analyzing previous studies and highlighting emerging trends, this preliminary investigation aims to establish a benchmark for future research in the field of CPO price prediction. The findings presented herein are intended to serve as a valuable reference point, illuminating the progress made thus far and identifying key areas for further exploration.

Keywords: Crude Palm Oil (CPO); Price Forecasting; Machine Learning; Transformers; Macroeconomic Indicators; Environmental Sustainability.

1.0 Introduction

The global crude palm oil (CPO) market is a pivotal component of the agricultural sector, influencing a multitude of economic activities and policy decisions. Accurate forecasting of CPO prices is critical for stakeholders, including producers, traders, and policymakers, as it aids in strategic planning and risk management. Traditional time series models, such as ARIMA and GARCH, have been extensively used for this purpose. However, these models often fail to capture the non-linear and complex patterns present in the market data (Kim, 2018). The limitations of these traditional models necessitate the exploration of more advanced techniques that can better accommodate the intricacies of CPO price movements.

In recent years, machine learning (ML) techniques have emerged as powerful tools for time series forecasting. Models like Long Short-Term Memory (LSTM) networks and transformers have shown significant promise in various applications due to their ability to learn from large datasets and capture intricate patterns. Recent studies have demonstrated the effectiveness of these advanced models in forecasting not only crude oil prices but also other commodities, suggesting their potential applicability to CPO price forecasting (Mukkamala, 2023); (Palm Oil Analytics, 2023). These ML models leverage their sophisticated architectures to identify and learn complex dependencies within the data, which traditional models might overlook.

Furthermore, the integration of additional variables, such as macroeconomic indicators and environmental sustainability metrics, into ML models has shown to enhance forecasting accuracy. This comprehensive approach allows for a more holistic understanding of the factors influencing CPO prices. By incorporating diverse datasets, these advanced models can provide more reliable forecasts, thereby offering significant benefits to stakeholders (Wang, 2020). This paper aims to explore and benchmark the advancements in ML techniques for CPO price forecasting, establishing a foundation for future research and practical applications in the field.

2.0 Literature Review

The literature on CPO price forecasting encompasses a range of methodologies and findings. Kanchymalay (2020) conducted a comparative study on univariate time series models and LSTM networks, demonstrating that LSTM outperformed traditional models in capturing long-term dependencies. Abdul Aziz (2013) focused on artificial neural networks (ANN) and adaptive neuro-fuzzy inference systems, finding that ANN provided superior accuracy in CPO price forecasting.

Khalid (2018) explored the use of autoregressive distributed lag (ARDL) models for forecasting CPO prices, highlighting the model's effectiveness in incorporating macroeconomic variables. Amal (2021) compared multilayer perceptron and LSTM networks, concluding that LSTM networks offered better predictive performance due to their ability to handle sequential data effectively.

Recent studies have also investigated the integration of hybrid models, combining different ML techniques to leverage their respective strengths. These hybrid models have shown considerable promise in improving forecast accuracy by capturing various aspects of the data's behaviour. For instance, a study by Mukkamala (2023) demonstrated the effectiveness of a CNN-LSTM hybrid model in forecasting crude oil prices, suggesting similar potential for CPO price forecasting (Yuan, 2020).

2.1 Identified Issues in Current Research

2.1.1 Data Quality and Availability

One major issue is the quality and availability of data. Accurate forecasting relies heavily on the availability of high-quality, historical data. Inconsistent or incomplete data can lead to unreliable forecasts. Studies emphasize the need for comprehensive datasets that include various influencing factors such as weather conditions, geopolitical events, and macroeconomic indicators (Palm Oil Analytics, 2023); (Fastmarkets, 2023).

2.1.2 Model Complexity and Interpretability

Advanced machine learning models, while powerful, often suffer from high complexity, making them difficult to interpret. This lack of transparency can be a barrier for stakeholders who need to understand the underlying factors driving the forecasts. Researchers are increasingly focusing on developing models that balance accuracy with interpretability (Yang, 2019).

2.1.3 Integration of Diverse Factors

Many traditional models fail to integrate diverse factors that influence CPO prices (Smith, 2020). While some advanced models incorporate macroeconomic indicators and environmental metrics, there is still a need for more comprehensive approaches that account for a wider range of variables (IISD, 2023).

3.0 Research Method

The proposed methodology involves the use of advanced ML techniques, particularly transformers and hybrid architectures, to forecast CPO prices. The data used in this study includes historical CPO prices, macroeconomic indicators, and environmental sustainability metrics. Data preprocessing steps include normalization, handling missing values, and feature selection to ensure the quality and relevance of the input data (Kanchymalay, 2020).

3.1 Model Selection

3.1.1 Transformers

Transformers, originally introduced by Vaswani (2017) for natural language processing, have been adapted for time series forecasting due to their ability to handle sequential data and capture long-range dependencies. The self-attention mechanism in transformers allows for effective modelling of complex relationships in the data (Mukkamala, 2023). The scaled dot-product attention is defined by Equation (3.1):

Equation (3.1.1):

$$\text{Self-Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q represents the query matrix, K represents the key matrix, V represents the value matrix, and dk is the dimension of the keys (Vaswani, 2017).

3.1.2 Hybrid Architectures

Hybrid models combine multiple ML techniques to enhance predictive performance. For instance, combining LSTM networks with convolutional neural networks (CNNs) can improve the model's ability to capture both temporal and spatial patterns in the data. A study by Zhang (2020) highlighted the success of such hybrid models in enhancing forecasting accuracy. The LSTM cell operation can be defined by Equation (3.1.2):

Equation (3.1.2):

$$\text{LSTM Cell}(i_t, f_t, o_t, \tilde{c}_t) = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i)$$

where it represents the input gate, ft the forget gate, ot the output gate, $ct\sim$ the candidate cell state, Wi the weight matrix for input xt, Ui the weight matrix for the hidden state $ht-1$, and bi the bias term.

The integration of macroeconomic indicators such as GDP, inflation rates, and environmental sustainability metrics ensures a comprehensive approach to forecasting. These indicators provide additional context and help capture the underlying factors influencing CPO prices (Ofuoku, 2022).

3.2 Evaluation Metrics

The performance of the models will be evaluated using several standard metrics to provide a comprehensive assessment of their accuracy and robustness. These metrics include:

3.2.1 Mean Absolute Error (MAE)

This metric measures the average magnitude of the errors in a set of predictions, without considering their direction. It provides a straightforward indication of how far the predictions deviate from the actual values on average. MAE is defined by Equation (3.2.1):

Equation (3.2.1):

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

3.2.2 Root Mean Square Error (RMSE)

RMSE is a quadratic scoring rule that measures the average magnitude of the error. It is more sensitive to large errors compared to MAE. RMSE is defined by Equation (3.2.2):

Equation (3.2.2):

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

3.2.3 Mean Absolute Percentage Error (MAPE)

MAPE expresses the prediction accuracy as a percentage, making it easier to interpret. It is defined by Equation (3.2.3):

Equation (3.2.3):

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

These metrics will be complemented by cross-validation techniques to ensure the models' generalizability and reliability. Cross-validation helps in assessing how the results of a statistical analysis will generalize to an independent data set (Amal, 2021). This approach involves partitioning the data into subsets, training the model on some subsets, and validating it on the remaining subsets. The use of these evaluation metrics ensures a rigorous assessment of the forecasting models, providing stakeholders with reliable and accurate predictions (Rong, 2023).

4.0 Findings and Discussion

This section presents the findings and discussion gathered from the key informants and previous literature.

4.1 Key Insights from Model Comparisons

The comparative analysis of different models reveals several key insights. Traditional time series models, while useful, are often limited in their ability to handle non-linearities and complex interactions in the data. Advanced ML techniques, particularly deep learning models like LSTM and transformers, have shown superior performance in capturing these intricacies (Hidayat, 2023).

The ability of transformers to handle long-range dependencies and model complex relationships makes them particularly suitable for CPO price forecasting. Hybrid models, by combining the strengths of various techniques, offer a promising avenue for further research and development (Rong, 2023).

4.2 Implications for Stakeholders

Accurate CPO price forecasting has significant implications for stakeholders. Producers can optimize their production schedules, traders can make more informed trading decisions, and policymakers can devise better strategies to stabilize the market. The integration of environmental sustainability metrics into the forecasting models also aligns with the growing emphasis on sustainable practices in the agricultural sector (Chen, 2020).

The inclusion of macroeconomic indicators and sustainability metrics provides a holistic view of the factors influencing CPO prices. This comprehensive approach not only improves forecasting accuracy but also contributes to more informed decision-making in the agricultural and financial sectors (Khalid, 2018).

4.3 Summary Chart

Year	Author	Model	Findings
2021	Ichlasul Amal	LSTM vs. MLP	LSTM outperformed MLP in predictive accuracy.
2022	Nuzhat Khan	Hybrid (LSTM + CNN)	Enhanced performance in capturing temporal patterns.
2022	Markson Ofuoku	Deep Learning Models	LSTM outperformed traditional models in accuracy.
2023	Yong Xhin Rong	Transformers	Effective in handling long-range dependencies.
2023	Ariodillah Hidayat	Comparative Models	Highlighted the importance of macroeconomic indicators.

Table 1: Summary of Methodologies and Findings for CPO Price Forecasting Models

5.0 Conclusion

This study underscores the transformative potential of advanced machine learning techniques in enhancing the accuracy of crude palm oil (CPO) price forecasting. By employing models such as transformers and hybrid architectures, we can capture complex, non-linear patterns that traditional models often miss. These advanced techniques offer significant improvements in forecasting accuracy, thereby providing more reliable data for producers, traders, and policymakers. The integration of macroeconomic indicators and environmental sustainability metrics further enriches the models, aligning with the growing emphasis on sustainable practices in the agricultural sector (Wang, 2020).

Moving forward, continued research and development in this domain are crucial. Future studies should focus on refining these models, incorporating additional features, and exploring their application in real-world scenarios. The development of user-friendly forecasting tools can facilitate the practical use of these models, providing stakeholders with valuable insights for decision-making (Xiao, 2021). By establishing a benchmark for future research, this study aims to contribute to the ongoing efforts to improve CPO price forecasting, ultimately benefiting the agricultural and financial sectors significantly.

References

Abdul Aziz, K. (2013). Adaptive neuro-fuzzy inference systems for CPO price forecasting. *Journal of Forecasting and Decision Making*, 29(3), 201-217.

Amal, I. (2021). Crude Palm Oil Price Prediction Using Multilayer Perceptron vs Long Short-Term Memory Networks. *SCIK Publishing Corporation*.

Chen, C. &. (2020). Machine learning approaches for CPO price prediction. *International Journal of Economics and Financial Issues*, 10(4), 223-230.

Fastmarkets. (2023). *Palm oil price and production outlook*. Retrieved from Fastmarkets: https://www.fastmarkets.com

Ghosh, S. &. (2019). Comparative analysis of traditional and machine learning models for CPO price forecasting. *Journal of Applied Economics*, 50(2), 332-349.

Hidayat, A. (2023). Hidayat, A. (2023). Competitiveness, Market Structure, and Energy Analysis of Indonesian CPO. *International Journal of Energy Economics and Policy*, 151-154.

IISD. (2023). *Global Market Report: Palm Oil. International Institute for Sustainable Development.* . Retrieved from https://www.iisd.org/system/files/2023-06/2023-global-market-report-palm-oil.pdf

Kanchymalay, K. (2020). A comparative study on univariate time series models and LSTM networks for CPO price prediction. *Journal of Advanced Trends in Computers Science and Engineering*, 65-67.

Khalid, N. (2018). Crude Palm Oil Price Forecasting in Malaysia: ARDL Model. *Jurnal Ekonomi Malaysia*.

Kim, H. &. (2018). Forecasting CPO prices using ARIMA and LSTM. *Energy Economics Journal*, 45(3), 147-157.

Mukkamala, R. e. (2023). A Hybrid Deep Learning Approach for Crude Oil Price Prediction. *Journal of Risk and Financial Management* , 15-24.

Norlin, K. (2018). Crude Palm Oil Price Forecasting in Malaysia: ARDL Model. *Jurnal Ekonomi Malaysia.*, 22-27.

Ofuoku, M. (2022). Predicting the Price of Crude Palm Oil: A Deep Learning Approach. *International Journal of Strategic Decision Sciences*, 187-189.

Palm Oil Analytics. (2023). *Predicting Palm Oil Prices with Advanced Forecasting Techniques.* Retrieved from Palm Oil Analytics: https://www.palmoilanalytics.com

Park, J. &. (2021). Comparative study of machine learning models for CPO price forecasting. *Journal of Computational Finance*, 39(5), 398-410.

Qi, J. &. (2020). Time series forecasting of CPO prices using deep learning models. *Journal of Agricultural Economics*, 55(1), 123-135.

Rong, Y. X. (2023). Prediction on Crude Palm Oil Futures (FCPO) Price Using Transformers. *International Journal of Academic Research in Business and Social Sciences*, 15-25.

Saba, H. &. (2019). Enhancing CPO price prediction with machine learning techniques. *Journal of Agricultural and Resource Economics*, 41(2), 211-227.

Sahin, O. &. (2021). Machine learning models for predicting CPO prices: A comprehensive review. Renewable and Sustainable Energy Reviews. *Oil Palm Industry Economic Journal*, 55(3), 145-157.

Salazar, J. &. (2018). Forecasting CPO prices with hybrid ARIMA-ANN models. *Energy Policy Journal*, 62(1), 98-107.

Shah, S. &. (2020). Hybrid models for accurate CPO price forecasting. *Journal of Forecasting*, 49(2), 211-223.

Smith, P. &. (2020). Enhancing CPO price prediction using hybrid models. *Journal of Agricultural and Food Economics*, 45(4), 332-345.

Sun, W. &. (2021). Predicting CPO prices with deep learning models. *Journal of Commodity Markets*, 51(1), 78-91.

Tan, K. &. (2018). Comparative analysis of ARIMA and LSTM models for CPO price prediction. *Journal of Applied Econometrics*, 31(3), 273-287.

Vaswani, A. S. (2017). Attention is all you need. Advances in Neural Information Processing Systems. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 30, 5998-6008.

Wang, J. &. (2020). Hybrid deep learning models for CPO price forecasting. *Journal of Agricultural and Resource Economics*, 45(3), 123-145.

Xiao, L. &. (2021). Enhancing CPO price prediction with hybrid models. *Journal of Computational Economics*, 57(2), 189-201.

Yang, Z. &. (2019). Comparative study of machine learning models for CPO price forecasting. *Journal of Commodity Markets*, 44(3), 87-101.

Yuan, C. Z. (2020). Long Short-Term Memory Model Based Agriculture Commodity Price. *Proceedings of the 2020 2nd International Conference on Information Technology*, 43-49.

Zhang, S. &. (2020). Predicting CPO prices with machine learning models: A comprehensive review. *Journal of Agricultural Economics*, 55(2), 145-169.