OmniaScience

Journal of Industrial Engineering and Management
JIEM, 2024 – 17(2): 321-343 – Online ISSN: 2013-0953 – Print ISSN: 2013-8423
https://doi.org/10.3926/jiem.7032

Materials Inventory Optimization Using Various Forecasting Techniques and Purchasing Quantity in Packaging Industry

Melissa Christian Dinata*, Suharjito Suharjito

Industrial Engineering Department, BINUS Graduate Program - Master of Industrial Engineering, Bina Nusantara University (Indonesia)

*Corresponding author: melissa.dinata@binus.ac.id
suharjito@binus.ac.id

Received: December 2023
Accepted: March 2024

Abstract:

Purpose: This paper studies the problem that occurs on material purchase quantity in price uncertainty situation. Larger buying quantity when the price at high will increase the purchase amount while smaller buying quantity could risk the inventory level. The decision on the purchase quantity of a cycle takes future price as input from price prediction output.

Design/methodology/approach: This paper examines five price prediction models, Classification and Regression Tree (CART), Random Forest Regressor (RFR), Support Vector Regressor (SVR), Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short Term Memory (LSTM) to predict four Petrochemical products, Linear Low Density Polyethylene (LLDPE), Low Density Polyethylene (LDPE), Biaxially Oriented Polypropylene (BOPP) and Cast Polypropylene (CPP) using dataset built from weekly datapoints from January 2020 to June 2023. The most performing model is validated with data from July 2023 to September 2023 where the prediction result is fed into Linear Programming Simplex method to minimize the amount of purchase by making advanced or postpone orders.

Findings: Result that RFR performs higher at most products tested, while SVR performs higher in LDPE product. The fitting of RFR and SVR models prediction, as predicted price to Linear Programming that decides optimum purchase quantity, delivers a total 2.2% of purchase amount reduction compared to original purchase quantity reflecting base scenario issued by the planner.

Research limitations/implications: This study does not include additional prediction factors such as freight cost and the hyperparameters tuning studies on the existing factors.

Originality/value: The novelty of this paper is prediction value is followed up by an optimization model that would guide the Procurement team decisions for future anticipation because imported raw materials should be purchased ahead of time. This research will provide a scientific approach input that would counterbalance or strengthen decision making that is typically made by individuals owning years of experience. This combined approach is rarely researched and has not been done to polymer products.

Keywords: petrochemical, prediction, optimization, inventory, linear programming

To cite this article:

Christian-Dinata, M., & Suharjito, S. (2024). Materials inventory optimization using various forecasting techniques and purchasing quantity in packaging industry. *Journal of Industrial Engineering and Management*, 17(2), 321-343. https://doi.org/10.3926/jiem.7032

1. Introduction

Plastics, derived from petrochemicals like ethylene and propylene, are widely used for packaging food and non-food products in both rigid and flexible forms. Polyethylene (PE) and polypropylene (PP) are examples of plastic resins that are considered commodities due to their abundant availability and dependence on petrochemical production processes. The prices of PE and PP fluctuate based on market demand and supply, influenced by various factors such as the pandemic, inflation, geopolitics, production capacity, shipping, and natural disasters in the Southeast Asia region.

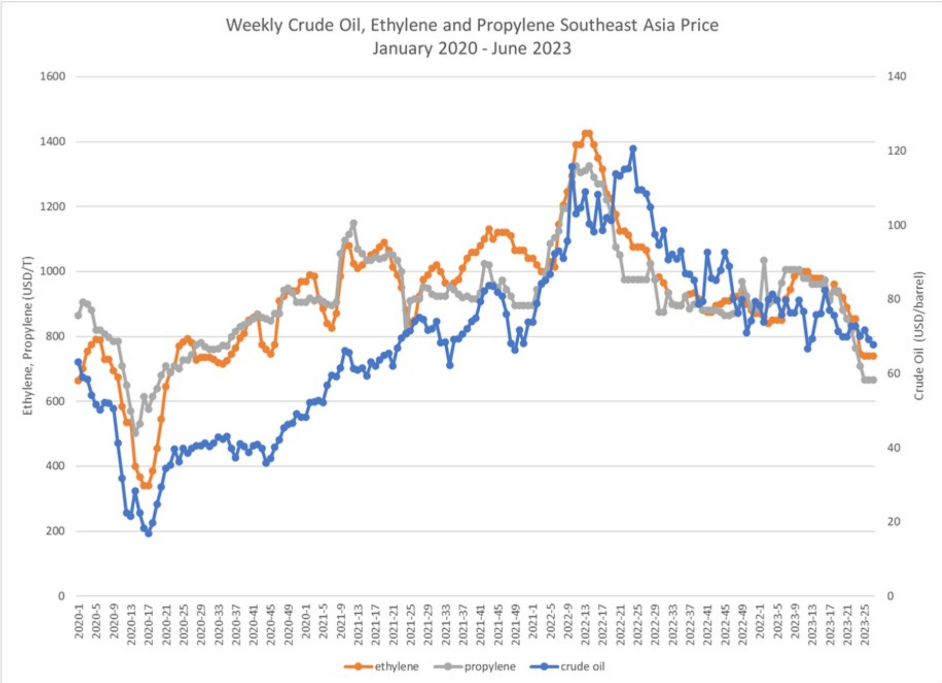


Figure 1. Weekly Price of Crude Oil, Ethylene and Propylene in Southeast Asia, January 2020-June 2023

Figure 1 illustrates the fluctuation in Crude Oil, Ethylene (C2), and Propylene (C3) prices during the pandemic, with a subsequent steady rise. Additionally, geopolitical tensions between Russia and Ukraine in early 2022 resulted in a rapid increase in PE and PP prices, even as the impact of the COVID-19 pandemic gradually subsided. For Procurement team, maintaining competitive prices at lowest possible level is crucial, as high material prices can significantly affect company profitability, given that 60-70% of the cost of sales is attributed to material expenses. Wrong decision to purchase could cause negative implication to business operation. High inventory value and quantity will reduce the competitiveness in the market. Therefore, in the era of Industry 4.0, data science plays a crucial role in Procurement by enhancing knowledge, improving performance, and facilitating better decision-making capabilities (Hamed, Richa & Dmitry, 2023). Through data science, organization gains understanding on the correlation between upstream market price indexes that enables accurate prediction of PE and PP resin or film prices as valuable inputs for Procurement decisions. Data published on wits.worldbank.org reveals that the import trade values of PE with HS code 390110, from all countries to Indonesia experienced an increase from $673.72 million in 2021 to $713.02 million in 2022, indicating a growth of 5.83%. Similarly, PP with HS code 390210 also witnessed a rise from $1,095.13 million in 2021 to $1,220 million in 2022, marking an increase of 11.4%. With an increasing amount of transaction of imported PE and PP to Indonesia, In increasing demand and value of transaction year on year, it is evident that PE and PP has increasing importance to the industry and price predictability and inventory value optimization will support business competitiveness for each company. This paper presents a solution for predicting the price of PE and PP products as target variable by using price indices as input for future inventory and value optimization. The novelty of this paper is prediction value is followed up by an optimization model that would guide the Procurement team decisions for future anticipation because imported raw materials should be purchased ahead of time. This research will provide a scientific approach input that would counterbalance or strengthen decision making that is typically made by individuals owning years of experience. This combined approach is rarely researched and has not been done to polymer products.

1.1. Price Forecasting Studies

Commodity price forecasting has undergone significant evolution. There is a lot of literature about price forecasting using statistical or Machine Learning (ML) and Deep Learning (DL) methods depending on the data and the needs. These forecasting methods study the relationship between various factors and the impact to the predicted value.

The summary of price prediction related studies is presented in Table 1. Most price prediction studies discuss techniques to accurately develop performing prediction models in commodities, agriculture and mostly Crude Oil. The successful prediction on these studies have put important foundation to this research to choose which the prediction models to for PE and PP.

Author	Object	Time Series			Regression						Sim	Optimization
		ARIMA	PROPHET	LSTM	MLR	CART	SVR	RFR	XGBOOST	SD	
Gargano and Timmermann, 2014	Commodity				X						
Chen, Goh, Sin, Lim, Chung and Liew, 2021	Agriculture	X	X	X				X			
Sen, Choudhury and Datta, 2023	Crude Oil			X							
Kwon, Do and Kim, 2020	Naphtha									X	MILP
Chen and He, 2019	Crude Oil	X			X	X		X			
Urolagin, Sharma and Datta, 2021	Crude Oil			X							
Jahanshahi, Uzun, Kaçar, Yao and Alassafi, 2022	Crude Oil			X	X		X	X			
Ensafi, Amin, Zhang and Shah, 2022	Furniture	X		X							
Kamdem, Essomba and Berinyuy, 2020	Commodity	X		X							

Table 1. Related studies

1.2. Classical Methods

Classical terms refer to the implementation of statistical approach for price prediction using regression and time-series methods. Multi-linear Regression (MLR) is used in a study to predict commodity prices using long-term financial and economy data from 1947 to 2010. That study showed that financial indicators could predict the commodity price movement in month and quarter range (Gargano & Timmermann, 2014). MLR is used in research to predict 10 months data of Naphtha price based on dataset that contains supply-demand, margin, and availability of substitution material from 2008 to 2010 (Sung, Kwon, Lee, Yoon & Moon, 2012). As the dataset is limited to 30 data points, it achieves R^2 at 0.651 with 50% accuracy. Another crude oil prediction using MLR and Random Forest is also applied (Chen & He, 2019). The study shows that MLR achieves lower performance than Random Forest with dataset built with 312 records of monthly crude oil price from 1992 to 2017.

Time-series prediction refers to a sequence of data meticulously recorded and subsequently analyzed in a chronological sequence with uniform time intervals (such as yearly, monthly, daily, or hourly). Autoregressive Integrated Moving Average (ARIMA) it the classical method for time-series prediction however ARIMA was unable to capture the recurring seasonal patterns in the data, but it did manage to identify an upward sales trend towards the end of each year (Ensafi et al., 2022). ARIMA shows best performance to predict stock for financial and fast-moving consumer goods (FMCG) sector but lacking accuracy for banking and automobile (Mondal, Shit & Goswami, 2014), crude oil (Chen & He, 2019) and agricultural commodity (Ouyang, Wei & Wu, 2019). To overcome the need of seasonal analysis, the Seasonal ARIMA (SARIMA) is model that includes seasonal terms of ARIMA. SARIMA is denoted as by $ARIMA(p,d,q)(P,D,Q)_m$. The p, d and q are non-seasonal while P, D and Q are seasonal factors. With p refers to order of the autoregressive part, d is degree of first differencing, q as order of the moving average, P is seasonal autoregressive order, D as seasonal difference order, Q as seasonal moving average order and m points to number of observations per year.

Seasonal plot of PE and PP products shown in Figure 2 is performed with Python statsmodels library (Vishwas & Patel, 2020). The chart showcases a seasonal pattern of upstream price indices and target variables in dataset containing data from January 2020 to June 2023. The data visualize increase at around beginning of the second quarter every year that is known as the peak season. Then data gradually decreases into the third quarter where it is known as the low season. As end of year approaches, data trend starts to indicate a movement to the higher price when the peak season occurs typically on the first quarter each year.

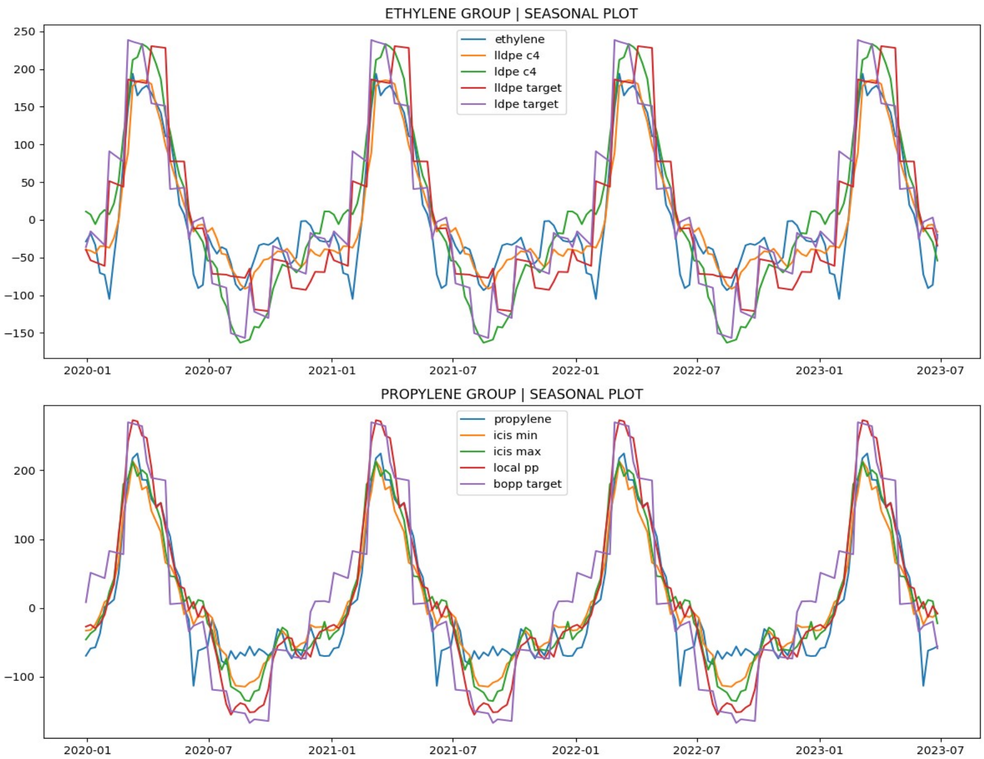


Figure 2. Seasonal plot of Polyethylene and Polypropylene products from January 2020-June 2023

1.3. Machine Learning and Deep Learning Methods

Besides the statistical method, price prediction studies have employed ML and DL models. The discussion encompasses Long Short-Term Memory (LSTM) and Support Vector Regression (SVR) approaches because of their capacity to handle non-linear patterns, comprehend intricate cause-and-effect relationships, and learn from extensive historical datasets (Chen et al., 2021). LSTM is frequently employed for classification and prediction tasks based on time-series data (Jahanshahi et al., 2022). LSTM is a type of Recurrent Neural Network (RNN) that stands out due to its distinctive memory cell configuration which effectively substituting the traditional hidden layer in RNN. Within the LSTM model, data is filtered using a gate structure such as input, forget, and output components. A comprehensive comparison between time-series forecasting concludes LSTM outperforms classical methods (ARIMA dan SARIMA)

(Ensafi et al., 2022). High LSTM performance at 98.2% was also shown in study of covid-19 spread implication to commodity price such as Crude Oil WTI, Brent, Wheat and Silver (Kamdem et al., 2020).

LSTM is used in the prediction of crude oil variation from 2011 to 2022 with 2218 data points. Several factors are included to determine crude oil variance. Granger causality test screens that S&P Index, gold variance, previous day crude oil variance and dollar index variance cause to price oil variance (Sen et al., 2023) which on optimized hyperparameter, the LSTM yielded R^2 0.8074.

Another research employs multivariate LSTM to forecast crude oil with an impressive R^2 value at 0.954 based on features variable such as Gold Prices, S&P Index, Dollar Index, US 10 Year Bond yield (Urolagin et al., 2021). The high performance is achieved by reducing the data nominally by filtering 19.4% data out by Mahalanobis distance and 9.62% by Z-score outlier. The Bidirectional LSTM was employed to study features from both forward and backward directions of daily crude oil price data (Vo, Nguyen & Le, 2020). This approach yielded the best performance compared to other methods like LSTM and CNN, achieving an RMSE of 1.55 and an MAE of 1.2.

Classification and regression tree (CART) is a non-parametric ML method for classification and regression tasks (Choubin,, Moradi, Golshan, Adamowski, Sajedi-Hosseini & Mosavi, 2019). Being a non-parametric model, CART does not depend on specific data distribution, thus the presence of outlier data does not affect performance (Ghiasi, Zendehboudi & Mohsenipour, 2020). Random Forest is a popular machine learning procedure to develop prediction models. Random forest uses binary split on the predictor to determine outcome predictions. Random forest consists of many classification and regression trees that are constructed by randomly selected training data subsets. The result of each tree is aggregated to give prediction of each observation. This way, Random Forest results in higher accuracy compared to a single decision tree (Speiser, Miller, Tooze & Ip, 2019).

Random Forest models outperformed MLR and ARIMA by demonstrating significantly higher forecasting accuracy with an impressive R^2 value of 0.996, RMSE 3.64, MAE 2.56. This analysis was conducted using 312 monthly data points of 8 predictors spanning from 1992 to 2017 to predict spot crude oil price (Chen & He, 2019). Random Forest is utilized to predict the stock price from May 2009 to May 2019 of the next closing day based on feature variables such as Open, High, Low, Close prices and moving average of seven, fourteen, twenty-one days and standard deviation of past seven days. In that study Random Forest RMSE, MAPE and MAE have similar performance compared to ANN for Pfizer, Johnson & Johnson, and Goldman Sachs stocks, while ANN outperform Random Forest for Nike and JP Morgan (Vijh, Chandola, Tikkiwal & Kumar, 2020). Random Forest Regressor (RFR) showed predictive capability for water prices (Xu, Lian, Bin, Hua, Xu & Chan, 2019) to offer solutions to the water market price uncertainty problem. In this study, RFR demonstrated potential in capturing complex and non-linear relationship between water price as target with many factors that traditional regression often fails (Xu et al., 2019) and RFR results R^2 value at 0.692 compared to decision tree at 0.541.

LSTM and RFR models developed in this study have great reputation to predict, yet there is still lack of past research that studies beyond the prediction specially to use the prediction output as input for decision making. One research example predicts time series Naphtha price that could influence the production of petrochemical products (Kwon et al., 2020) such as Ethylene, Propylene, Butadiene, Benzene, Toluene and Xylene to optimize the profit. The prediction factors are historical prices, quantity of demand and supply, price profiles, crude oil and financial statistics taken sourced from May 2009-December 2010 to train three prediction models multi-linear regression, Artificial Neural Network (ANN) and System Dynamic (SD) which result accuracy of 84% to predict January 2011-December 2011 data. Profit optimization is performed with Mixed Integer Linear Programming (MILP) at each decision stage.

1.4. Optimization Methods

In a situation of high commodity product price volatility, price minimization delivers significant profitability to the company. However, there is not much literature that concerns price optimization recently. Linear programming is a mathematical equation that seeks for optimum solution whether to maximize or minimize output in given constraint. This study employs Linear Programming as inventory value increases or decreases as a function of price and quantity of the purchase. Linear Programming is used to analyze network of revenue of airline industry to maximize revenue (An, Mikhaylov & Jung, 2021) and optimize inventory (Perez, Hubbs, Li & Grossmann, 2021). Mixed Integer Linear Programming (MILP) used in research to increase the profitability of Liquid Natural Gas

(LNG) (Sangaiah, Tirkolaee, Goli & Dehnavi-Arani, 2020) that focuses on under uncertainty of the constraints such as supply situation, sales strategy, planning period and price discount possibility. A combination of prediction and integer programming to optimize R&D budget was studied in (Jang, 2019) research to maximize output on allocated budget.

2. Methodology

To conduct price prediction research a dataset is constructed for model training and testing purposes. This dataset consists of 183 data points of Ethylene (C2), LLDPE, LDPE, Propylene (C3) and Polypropylene (PP) price which the source is taken from Independent Commodity Intelligence Services (ICIS) weekly price list consolidated from the 1st week of January 2020 to the 30th week 2023 or end June 2023. Crude oil price refers to WTI grade source that is weekly data point collected from investing.com and the Naphtha price source is from weekly data from tradingeconomics.com.

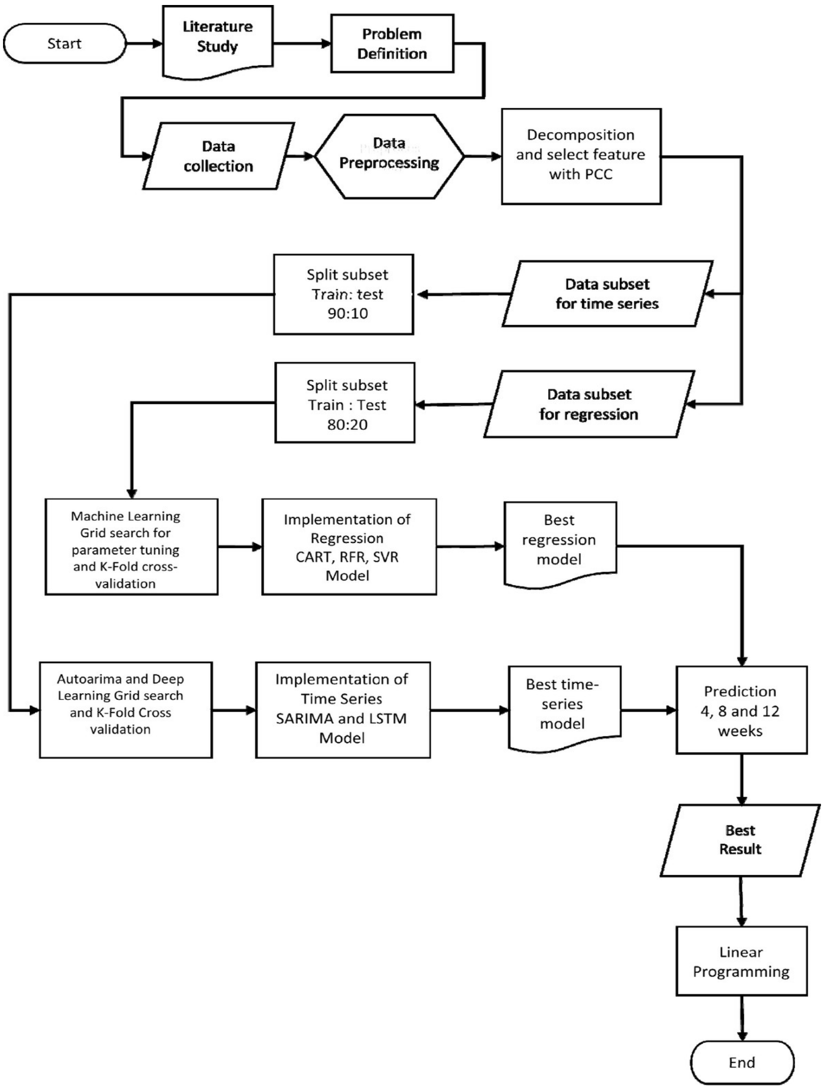


Figure 3. Research Framework

Target material prices such as LLDPE, LDPE, BOPP and CPP are actual data obtained from supplier price list. Dataset's the most important information for time-series forecasting is that date of each datapoint (Ensafi et al., 2022) therefore the dataset index is set according to date format to generate a discrete time-series dataset. All models will be trained and tested with this dataset as pictured in Figure 3. Model prediction performance is measured with new data records from the first week of July 2023 to end September 2023 is appended to the dataset.

Variable	Count	Mean	SD	Min	25%	50%	75%	Max
Crude Oil	183	68	23	17	51	71	82	121
Naphtha	183	596	189	139	471	625	699	1,078
Ethylene	183	922	194	340	803	935	1,028	1,425
LLCPE C4	183	1,084	190	690	948	1,060	1,245	1,495
LDPE C4	183	1,283	281	840	1,048	1,220	1,550	1,768
LLDPE Target	183	1,329	199	1,030	1,160	1,260	1,480	1,760
LDPE Target	183	1,529	321	945	1,330	1,580	1,820	2,090
Propylene	183	915	148	503	855	910	975	1,325
ICIS MIN	183	1,107	178	770	963	1,070	1,243	1,480
ICIS MAX	183	1,200	196	830	1,040	1,170	1,358	1,590
Local PP	183	19,399	3,141	14,140	16,930	19,180	21,590	26,450
BOPP Target	183	24,808	3,635	19,562	21,727	23,842	28,211	32,818
CPP Target	183	26,680	4,747	20,980	22,300	25,976	30,955	35,714

Table 2. Dataset descriptive statistics

The statistical data provided in Table 2 pertains to the dataset intended for both training and testing stages in constructing various prediction models for polyethylene (PE) and polypropylene (PP) materials, which are petrochemical products. Models developed in this study, including CART, RFR, SVR, SARIMA, and LSTM, were selected due to their proven effectiveness in predicting commodity materials derived from crude oil, as indicated in previous research.

2.1. Data Preprocessing

Data outlier filtration is not implemented (Kwon et al., 2020) because each data point represents the price volatility of commodity grade material (Sen et al., 2023), that naturally occurs due to supply and demand. Crude oil price unit is US dollar per barrel, meanwhile Naphtha and the rest of variables are in US dollar per ton except Local PP, BOPP target and CPP target are in local currency. LSTM prediction procedure follows three main steps such as data normalization, data training and parameter tuning (Kamdem et al., 2020). Data normalization is LSTM pre-processing step that scales all input features to common range because each feature has different scales. MinMax scaling is performed with Scikit-Learn library in Python that is in accordance with Equation (1). MinMax data normalization is performed separately to the train and test dataset after the split.

$$X_{scaled} = \frac{(X - X_{\min})}{(X_{\max} - X_{\min})} \tag{1}$$

2.2. Correlation Check

Pearson Correlation Coefficient (PCC) is required to evaluate the relationship between variables (Kotu & Deshpande, 2019). It is important to indicate the strength of feature variables that influence the target variable. PCC is calculated according to formula in Equation (2) below.

$$r = \frac{\sum (X_i - \overline{X}) \times (Y_i - \overline{Y})}{\sqrt{(X_i - \overline{X})^2 \times (Y_i - \overline{Y})^2}} \tag{2}$$

with r as Pearson Correlation Coefficient, X_i and Y_i are individual data, $\overline{X}$ and $\overline{Y}$ are the means. PCC value greater than 0.5 indicates a correlation between the variables.

2.3. Dataset Split

Splitting the data into training and testing subsets is a crucial step in time series analysis and variables for the purpose of model evaluation and performance assessment. The primary reasons for splitting the data are to simulate real-world scenarios (Ensafi et al., 2022) and to avoid introducing data leakage, which can lead to overly optimistic performance metrics (Kamdem et al., 2020). This research implemented train and test data split at 80:20 that will be used in the regression algorithm such as Classification and Regression Tree (CART), Random Forest Regression (RFR) and Support Vector Regression (SVR). Quantitatively, out of a total of 183 data records, 142 will serve as training data, while 36 will be reserved for testing purposes to validate the model. Meanwhile, dataset split on time-series prediction LSTM and SARIMAX is 90:10 to achieve best performance. The 10% time-series test data covers 5 months prediction from February 2023 to June 2023.

2.4. Prediction Performance Measurement

Model performance is assessed by four measures as used in literature study. Mean Squared Error (MSE) as in Equation (3) determines the average of squared predicted error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - F_i\right)^2 \tag{3}$$

Root Mean Squared Error (RMSE) in Equation (4) determines the standard deviation of MSE.

$$RMSE = \sqrt{MSE} \tag{4}$$

Mean Absolute Error (MAE) is an error measure between paired observation in regression analysis. It quantifies the closeness between the predicted value against the actual observed values expressed in Equation (5).

$$MAE = \frac{\sum_{i=1}^{n}\left|Y_i - F_i\right|}{n} \tag{5}$$

with n as number of data points, Y_i, is actual value and F_i, as predicted value.

Mean Absolute Percentage Error (MAPE) in Equation (6) determines the absolute percentage error function for the prediction and actual.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - F_i}{Y_i}\right| \times 100 \tag{6}$$

Coefficient of determination R^2 as in Equation (7) determines how well the predictor explains the dependent variable.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - F_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{F}\right)^2} \tag{7}$$

With n as number of data points, Y_i, is actual value of the dependent variable for the i^{th} data point, F_i, is predicted value of the dependent variable for i^{th} data point and $\overline{F}$ is mean of predicted value.

To complement the prediction performance, residual analysis will provide insight where the prediction performs well and where it may struggle. The residual of i^{th} data point is difference between actual Y_i, and predicted data F_i, as formulated in Equation (8) below.

$$e_i = Y_i - F_i \tag{8}$$

2.5. K-Fold Cross Validation and Hyperparameter Tuning

K-Fold cross-validation is a method in machine learning and statistical modeling for evaluating a model's performance and its ability to generalize. This approach divides a dataset into K subsets or folds and subsequently conducts K rounds of training and testing. In each round, one of the folds is designated as the test set, while the remaining K-1 folds serve as the training set. The process is repeated K times, with each fold acting as the test set once. The results are then averaged to provide an overall assessment of the model's performance (Han, Kamber & Pei, 2012). The result of cross-validation depends on model's hyperparameters tuning. Hyperparameters are model's variables that control the behavior of the model to fit with the data that is part of experimental efforts to determine a configuration that delivers best and stable performance (Passos & Mishra, 2022). Therefore, it is important to perform cross validation and parameter tuning simultaneously as the number of iterations of each hyperparameter combination could be overwhelming. To overcome the situation, this paper uses GridSearchCV in Python sklearn library.

Dataset train and test split for LSTM follows the same proportion as in SARIMA that is 90:10 for unbiased comparison and analysis of both method's performance. This research uses Python tensorflow.keras library to transform time-series data to supervised learning using previous observation as input and actual observation as output. LSTM model can be further tuned to determine the model at desired level of performance. Hyperparameter tuning of LSTM in Table 3 is performed on three parameters such as LSTM units (neurons), learning rate (steps) and batch size. Higher LSTM unit number affect to ability for LSTM to capture complex models at the cost of computational resources. Learning rate or step size controls the gradient descent optimization. Step size contributes to convergence stability. Batch size regulates the number of samples during the training stage. Large batch size may generate stable convergence and may be less accurate because of the effect of generalization.

Neuron	Learning Rate	Batch Size	Layer	Optimizer	Activation	Epoch	Dropout
60	2	8	1	Adam	ReLu	100	0
100	5	16	1	Adam	ReLu	100	0
150	10	32	1	Adam	ReLu	100	0

Table 3. LSTM Hyperparameters Search Combination

2.6. Seasonal ARIMA Model Optimization

Seasonal ARIMA is an extended model from ARIMA to predict seasonal time-series which is denoted by $ARIMA(p,d,q)(P,D,Q)_m$. The p, d and q are non-seasonal while P, D and Q are seasonal factors. This research set m as 12 to cover the season in the yearly horizon. To identify the best parameters for SARIMA, this research adopted a stepwise approach using autoarima as used in (Ensafi et al., 2022) paper using Python pdmarima library to find the lowest Akaike Information Criterion (AIC) value. AIC is a statistical test and comparative measure for time-series model. AIC provides estimation of information lost when a specific model is used, lower AIC indicates the best model. Mathematically, AIC is calculated by Equation (9) below.

$$AIC = -2 \times \frac{l}{n} + 2 \times \frac{k}{n} \tag{9}$$

with n as number of data or observation, k as number of estimated parameters (regressor and intercept) and l as the log of likelihood. When evaluating numerous alternative models, selecting the one with the lowest AIC value guarantees a well-rounded equilibrium between the goodness of fit and completeness (Profillidis & Botzoris, 2019).

2.7. Optimization Problem and Mathematical Model

Imagine a buyer engaged in price negotiations for imported material m during month i. This purchaser possesses information about the inventory value at the start of month i since the warehouse retains the material from the preceding purchase quantity q_{i-1} at the prior order price P_{i-1}. When the supplier presents a new pricing offer, the purchaser faces the task of determining the purchase quantity qp_i which may align with, be lesser than, or exceed the base quantity qb_i initially suggested by the planner in the purchase request. The planner has released the purchase quantity qb_i three times, considering the assumed consistent usage of each material $U_{j,i}$.

This approach enables the purchaser to deliberate whether to postpone a portion of the quantity to the subsequent purchase cycle or expedite a portion ahead. To guide the decision the total quantity qp_i of each month should be equal to three times qb_i and the purchase should adhere to the month-end-stock ratio range. To provide guidance for the decision-making process, it is important that the total quantity qp_i for each month corresponds to three times the qb_i value as formulated in Equation (13), while also ensuring that the purchases adhere to the range of month-end stock ratios as in Equation (14).

Minimizing inventory value becomes crucial in the face of fluctuating commodity prices, as it underpins a sustainable approach for customer contracts, consistent operational costs, and predictable cash flow, therefore the objective of the minimization process is formulated in Equation (10) as follows:

$$Min\, Z = \sum_{j=1}^{m}\sum_{i=1}^{n} U_{j,i} \times \left(Pb_{j,i} - Pp_{j,i}\right) \tag{10}$$

$$Pb_i = \frac{\left(P_{i-1} \times q_{i-1}\right) + \left(P_i \times qb_i\right)}{\left(q_{i-1} + qb_i\right)} \tag{11}$$

$$Pp_i = \frac{\left(P_{i-1} \times q_{i-1}\right) + \left(Pp_i \times qp_i\right)}{\left(q_{i-1} + qp_i\right)} \tag{12}$$

subject to:

$$\sum_{i=1}^{n} qp_i - \left(n \times U_i\right) \leq 0 \tag{13}$$

$$0.3 \leq \frac{\left(q_{i-1} + qp_i\right) - U_i}{U_i} \leq 1.0 \tag{14}$$

Decision variable:

qp_i: Purchased quantity after adjustment decision arriving at month i

where:

m is total material group,
j is material group (i.e., LLDPE, LDPE, BOPP, CPP),
n is total months in prediction range,
i is the month number,
$U_{j,i}$ is usage quantity of material j in month i,
$Pb_{j,i}$ is base price of material j in month i,
$Pp_{j,i}$ is predicted price of material j in month i
P_{i-1} Price of previous month, stock price of month i,
q_{i-1} Balance quantity of previous month, beginning quantity of month i,
P_i Price current month from purchase arriving at month i, from the prediction,
qb_i Purchased quantity according to base requirement arriving at month i

3. Result

3.1. Pearson Correlation Check

The correlation between upstream price indices and target is presented in Figure 4. In general, the price of petrochemical downstream products such as LLDPE, LDPE, BOPP and CPP have strong correlation to the upper stream products such as PE and PP which also strongly correlates to crude oil price.

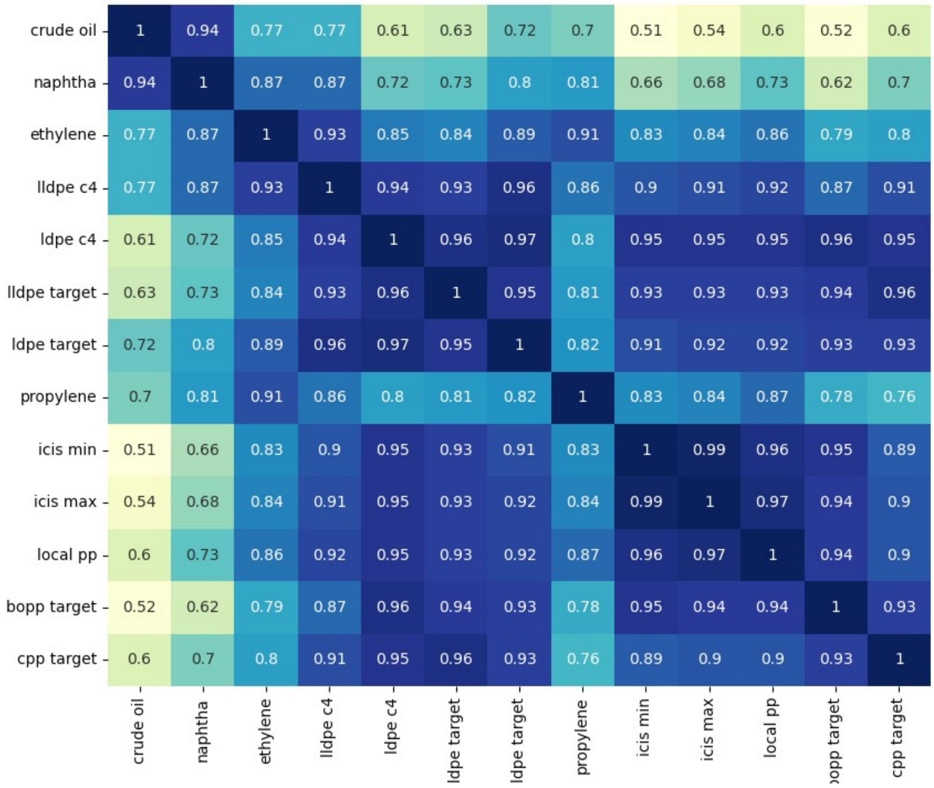


Figure 4. Pearson Correlation Check Result

This general finding implies that the fluctuation of Crude Oil could influence the price of lower stream products including the target. Correlation between upstream price indices as dataset feature and downstream price as dataset target shows clear direction therefore the dataset does not require data reduction to reveal the correlation (Salem & Hussein, 2019). Further, the data shows Crude Oil has a relatively stronger correlation to Naphtha (0.94) than to Ethylene (0.77) and Propylene (0.70). Naphtha price influence Ethylene (0.87) and Propylene (0.81) but may not be as strong as Crude Oil to Naphtha. This means Naphtha price fluctuation has more influence to lift or push Ethylene and Propylene price higher. Ethylene has a stronger correlation to LLDPE C4 than LDPE C4 and LLDPE Target. While LDPE target has stronger correlation to LLDPE target. To this correlation check, main factors that influence LLDPE target and LDPE target are Ethylene, LLDPE C4 and LDPE C4. While at the other end, Propylene, minimum ICIS, maximum ICIS, and local PP price have stronger correlations to BOPP and CPP target. Crude oil is not included as dataset feature as it has weaker correlation to the targets LLDPE (0.63), LDPE (0.72), BOPP (0.52), CPP (0.60). Naphtha also is not included as a dataset feature as its weaker correlation to the targets as well. The result is in line with Figure 2 where the data display seasonal pattern. Therefore time-series analysis to adopt is SARIMA instead of ARIMA.

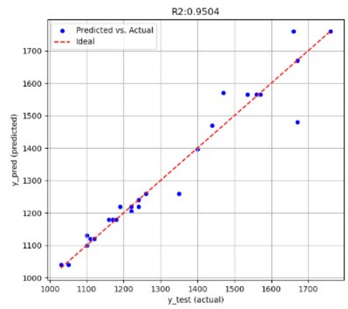
3.2. Regression Model Performance

The efficacy of each regression model is assessed using evaluation criteria like MSERMSE, MAE, MAPE, and R^2. The accuracy of these predictions is contingent upon the hyperparameter value selected during the training phase. This involves the simultaneous execution of the training process and 5-fold cross-validation, employing the Python scikit-learn GridSearchCV library to minimize the MSE. The outcome of GridSearchCV is best performing hyperparameter configuration summarized in Table 4.

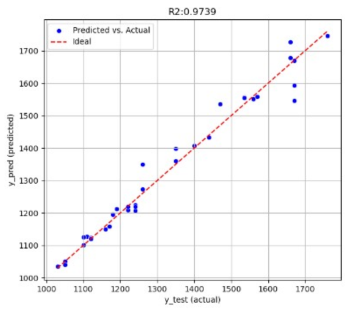
CART, RFR and SVR demonstrated strong linearity to predict LLDPE value as indicated by high R^2 value in Figure 5a, Figure 5b and Figure 5c with RFR being the strongest model to predict LLDPE. High performance of RFR could be further explained from residual plot Figure 5f. Compared to CART and SVR RFR low MSE score is indicated by residual value that converge around the ideal line while high RFR R^2 score is indicated by the proximity of the residuals to zero line across value range compared to other models.

Target	Model	Kernel	C	Epsilon	n Estimator	Criterion	Max Depth	Min Split	Min Leaf	MSE	RMSE	MAE	MAPE	RSQ
LLDPE	CART					Absolute error	10	10	1	2402.8716	49.0191	26.2838	1.8303	0.9504
	RFR				200		10	2	1	1263.1907	35.5414	22.5192	1.6063	0.9739
	SVR	linear	0.1	0.5						2872.9371	53.5998	36.3398	2.67	0.9406
LDPE	CART					poisson	20	5	4	3087.6737	55.5668	36.4739	2.5326	0.9744
	RFR				50		10	2	2	2320.3025	48.1695	29.0736	2.0007	0.9807
	SVR	linear	10	0.5						2418.4998	49.1782	35.6594	2.5656	0.9799
BOPP	CART					Absolute error	10	1	2	1113716.081	1055.3275	547.2162	2.1422	0.9259
	RFR				200		10	2	1	839360.6427	916.1663	592.3425	2.2795	0.9441
	SVR	linear	0.1	0.1						1404730.82	1185.2134	1002.3049	4.0907	0.9065
CPP	CART					friedman mse	None	2	2	3363973.703	1834.1139	862.1712	3.1588	0.8597
	RFR				50		20	5	2	2730226.426	1652.3397	874.1079	3.2967	0.8861
	SVR	linear	1	0.5						4273603.17	2067.2695	1648.5414	6.4245	0.8217

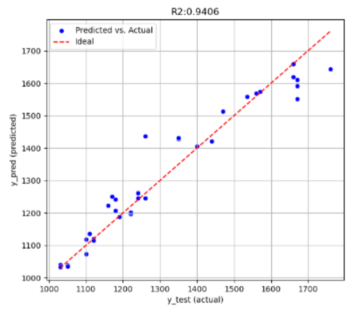
Table 4. Test dataset prediction performance in cross-validation run



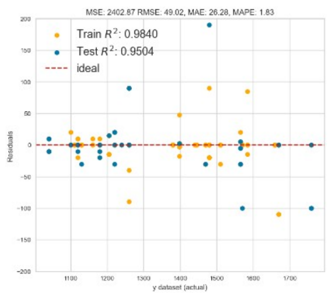
(a) LLDPE: CART R^2



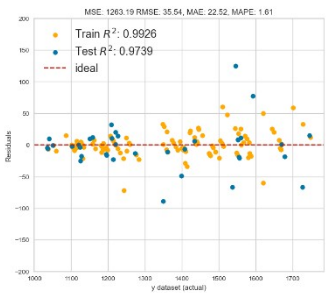
(b) LLDPE: RFR R^2



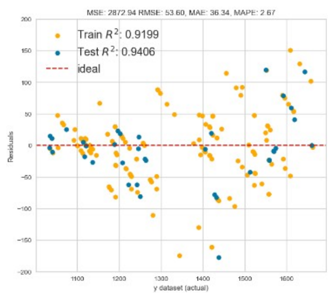
(c) LLDPE: SVR R^2



(d) LLDPE: CART Residual



(e) LLDPE RFR Residual

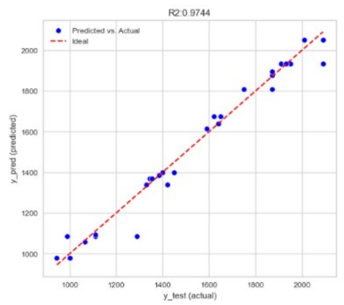


(f) LLDPE: SVR Residual

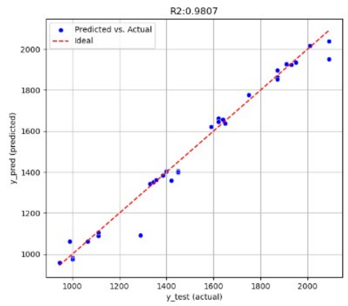
Figure 5. R^2 and residual of models to predict LLDPE test dataset

Meanwhile, SVR lowest performance to predict LLDPE is clearly plot in Figure 5f where the residual datapoints on both training and test dataset are scattered in wider spread. This spread indicates that all models struggle to predict upper range value, however adding more data in the space of higher value range to train and test dataset might potentially improve prediction performance of RFR.

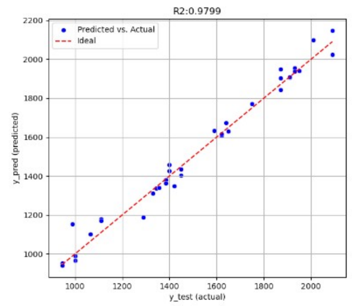
LDPE prediction performance shown in Figure 6. RFRin Figure 6b, demonstrated best performance than CART and SVR from R^2 score. For the CART, residual plot in Figure 6d shows that residual spread runs wider when the value shift to higher range. Meanwhile, RFR residual plot in Figure 6d has narrower spread across the value range, especially on the high value range. Lastly, SVR lower performance to estimate LDPE value is due to wider residual spread than other two models. However, SVR generates narrow residual spread between 1300-1600 value range is observed in Figure 6f.



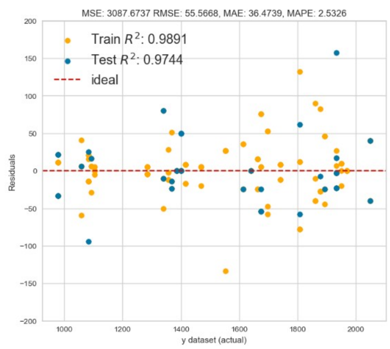
(a) LDPE: CART R^2



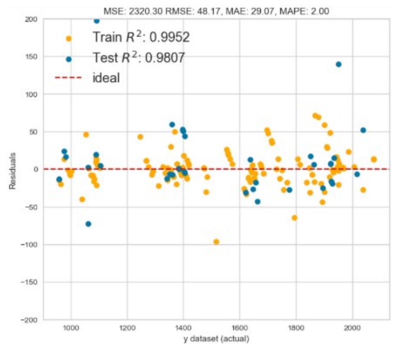
(b) LDPE: RFR R^2



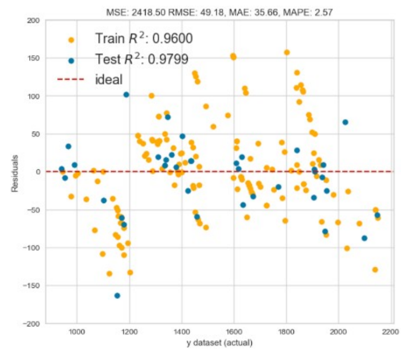
(c) LDPE: SVR R^2



(d) LDPE: CART Residual

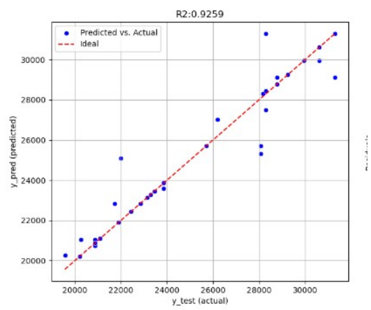


(e) LDPE RFR Residual

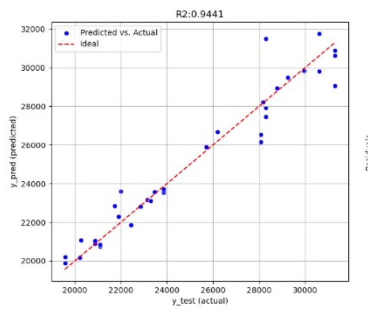


(f) LDPE: SVR Residual

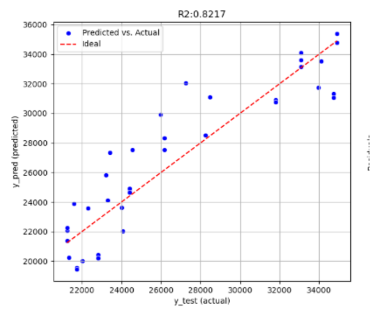
Figure 6. R^2 and residual of models to predict LDPE test dataset



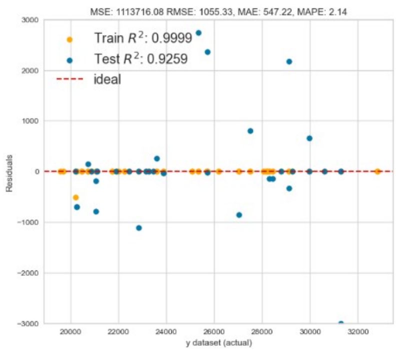
(a) BOPP: CART R^2



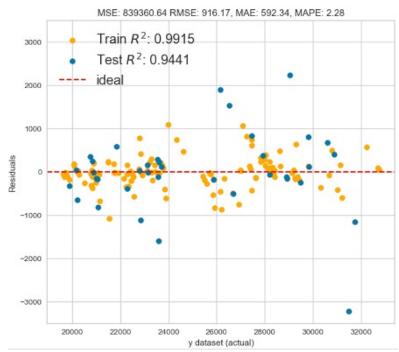
(b) BOPP: RFR R^2



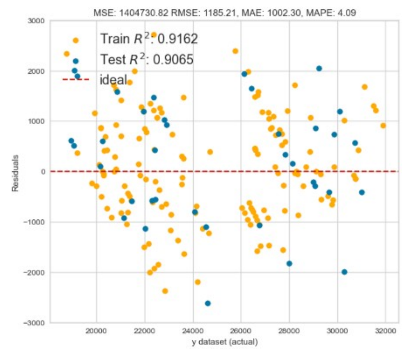
(c) BOPP: SVR R^2



(d) BOPP: CART Residual



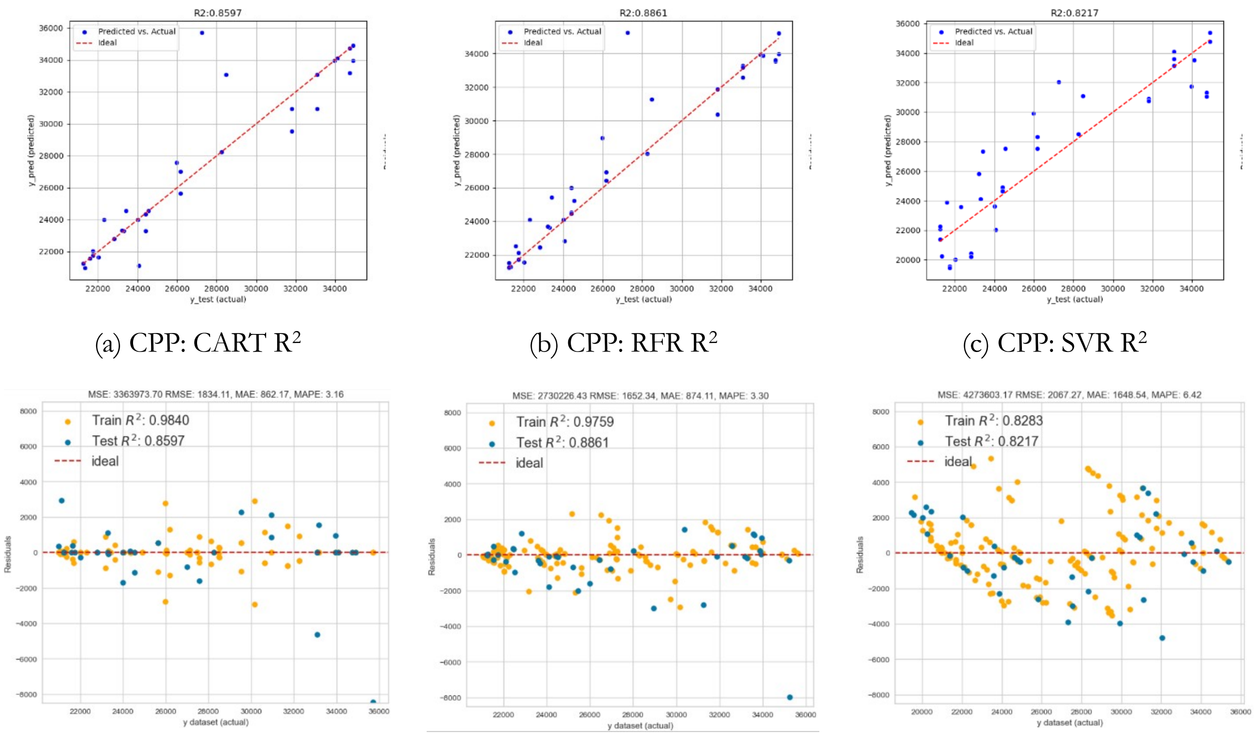
(e) BOPP: RFR Residual



(f) BOPP: SVR Residual

Figure 7. R^2 and residual of models to predict BOPP test dataset

The comparison of BOPP best prediction model is shown in Figure 7. All three models result in a high R^2 value. The residual plot of BOPP in Figure 7a, b and c have consistent observable behavior with previous material. RFR shows better performance than CART and SVR model. RFR shows narrower spread of training and test dataset at value range 20000-24000 but wider on higher value range. BOPP SVR training data residual plot in Figure 7f demonstrates weaker performance. The widespread of residual in training and test dataset is observed across the value range. Thus, adding more BOPP data to training dataset may not be able to influence better prediction performance with SVR.



(a) CPP: CART R^2 (b) CPP: RFR R^2 (c) CPP: SVR R^2

(d) CPP: CART Residual (e) CPP: RFR Residual (f) CPP: SVR Residual

Figure 8. R2 and residual of models to predict CPP test dataset

The performance of CART, RFR and SVR to predict CPP price could be seen in Figure 8. Compared to LLDPE, LDPE and BOPP, the performance of CPP prediction model has lower R^2 and higher MAPE value. The residual plot of CART in Figure 8a and RFR in Figure 8b shows that the both models have potential to perform equivalently as indicated by close MAE score of both in Table 4 and specifically in when prediction value falls between 20000-24000 range. SVR residual plot in Figure 8f has poor performance. The widespread residual of training and test dataset is observed across value range. Therefore, like BOPP, adding more CPP data points to training dataset may not influence better CPP prediction performance using SVR.

3.3. SARIMA Model Optimization

Result of SARIMA best fit model using stepwise process is summarized in Table 5 with steps to achieve lowest AIC shown in Figure 9 plot a, e, i, and m. The performance of this model on test data subset is measured with the same metrics such as MSE, RMSE, MAE, MAP for data that spans from February 2023-June 2023. The saved model is then used to predict test data subset and compared against actual test data as in Figure 9 plot b, f, j and n. In overall, SARIMA shows poor prediction performance for all target variables. However, SARIMA shows promising qualitative results due to its ability to track the future trend of LLDPE and LDPE but struggles to track BOPP and CPP future trend.

Target	Model	MSE	RMSE	MAE	MAPE
LLDPE	ARIMA(0,1,0)(2,1,0)[12]	1896.299	43.5465	3.072	37.1046
LDPE	ARIMA(0,1,0)(2,1,0)[12]	23794.92	154.256	143.8585	10.5146
BOPP	ARIMA(0,1,0)(2,1,1)[12]	2706450	1645.129	1342.524	6.4163
CPP	ARIMA(1,1,0)(0,1,1)[12]	1444831	1202.011	837.6754	3.7576

Table 5. Sarima stepwise optimum result

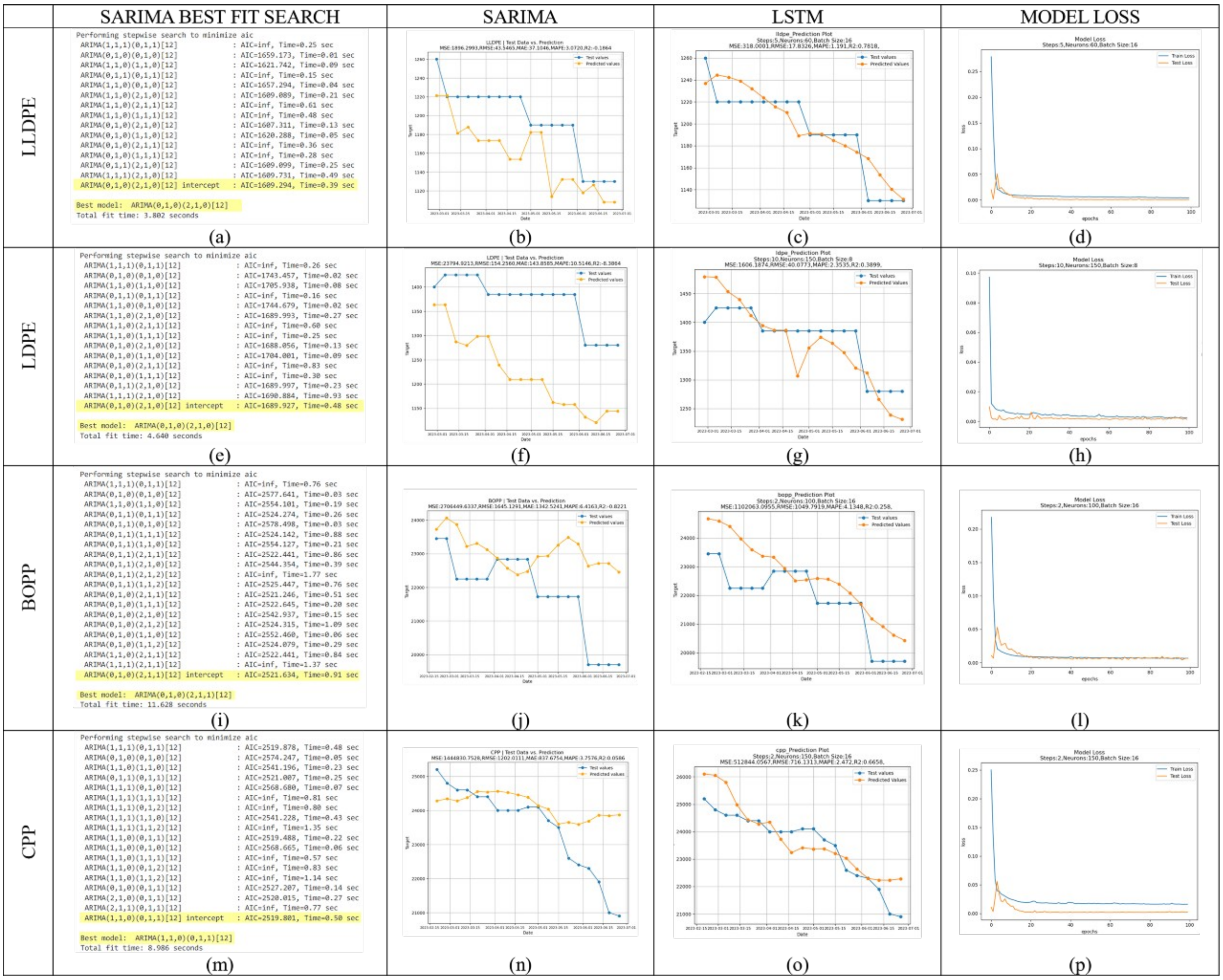


Figure 9. SARIMA and LSTM prediction result

3.4. LSTM Model Optimization

Hyperparameter tuning of LSTM is an experimental process, thus 27 fits involving number of neurons, learning rate and batch size as in Table 3 conducted to determine which hyperparameters that could deliver higher and stable prediction performance. The optimal LSTM hyperparameters are shown in Table 6. These parameters will be used for prediction. Other hyperparameters such as number of LSTM layers, optimizer, activation are not studied in this paper. The use of Epoch at 100 shown in Figure 9 plot d, h, l and p provided clear indication that the model loss stationarity could be achieved without needing of further extension. Dropout is not implemented due to limited dataset size.

Table 6 and Figure 9 plot c, g, k and o demonstrate capability single layer LSTM to track trend of test data subset of all target variables. All plots in Figure 9 indicates that the superiority of LSTM over SARIMA model with limited data the train data subset.

Material	Steps	Neurons	Batch Size	MSE	RMSE	MAPE	R2
LLDPE	5	60	16	349.0924	18.6645	1.222	0.7412
LDPE	10	150	8	2031.251	44.9234	2.6393	0.2284
BOPP	2	100	16	1004839	1001.725	3.9777	0.3235
CPP	2	150	16	741471	859.1224	3.2092	0.6611

Table 6. LSTM Optimal Hyperparameter Performance

4. Discussion

4.1. Prediction Performance

This section will focus on how trained regression and time-series model delivers prediction result of new dataset of weekly data points from July 2023 to end September 2023 displayed as solid red line in the following chart figures. This data is not part of the training and test data; therefore it is unseen from the model development process. Generated prediction result from the highest performing model will be employed in the inventory optimization with linear programming.

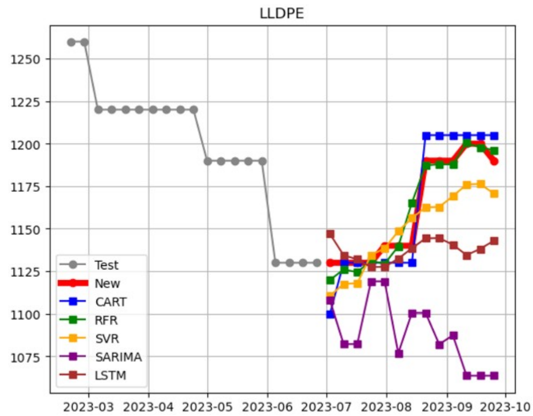


Figure 10. LLDPE actual and model prediction

Model	MSE	RMSE	MAE	MAPE
CART	165.385	12.86	10	0.8586
RFR	74.0712	8.607	5.618	0.4893
SVR	343.917	18.55	16.71	1.4255
SARIMA	7269.9	85.26	73.17	6.2131
LSTM	1343	36.65	27.87	2.35

Table 7. New LLDPE data prediction metrics

LLDPE actual data prediction and the performance is shown in Figure 10 and Table 7. RFR achieves the highest performance with the lowest MSE, RMSE, MAE and MAPE compared to other models. The R^2 value is not measured due to a small number of data composed from 13 weekly data from the first week of July 2023 to the last week of September 2023. RFR prediction value tracks the new data closely compared to CART, SVR, SARIMA and LSTM. LSTM and SARIMA are not able to predict the target, further SARIMA displays the opposite trend than all other models.

LDPE prediction shown in Figure 11 and Table 8 clearly show that SVR outperforms all other models. SVR prediction of new data is consistent with finding in residual plot. in training process. Where This performance is contributed by trained SVR model that has narrow residual range in the range 1300-1600. CART and RFR demonstrate the ability to track the price change, however both methods have a similar pattern that is higher than the actual. SARIMA shows different prediction meanwhile LSTM qualitatively able to indicate higher trend for the first 4 weeks.

BOPP prediction performance is displayed in Figure 12 and Table 9. RFR has the highest performance compared to the rest when comparing the MAPE value and the ability to trace the weekly data points. SVR shows the ability to trace the trend, which means the model could follow the feature data lead. SARIMA and LSTM are not able to indicate trend.

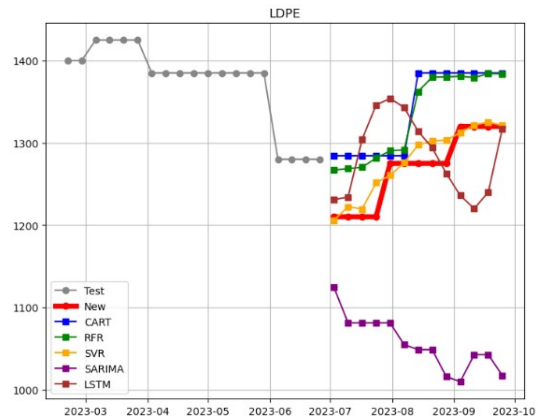


Figure 11. LDPE actual and model prediction

Model	MSE	RMSE	MAE	MAPE
CART	5803.532	76.181	69.67	5.5072
RFR	4701.341	68.566	63.485	5.0082
SVR	332.5969	18.237	13.569	1.0829
SARIMA	50439.13	224.59	212.95	16.615
LSTM	4977.308	70.55	58.385	4.6134

Table 8. New LDPE data prediction metrics

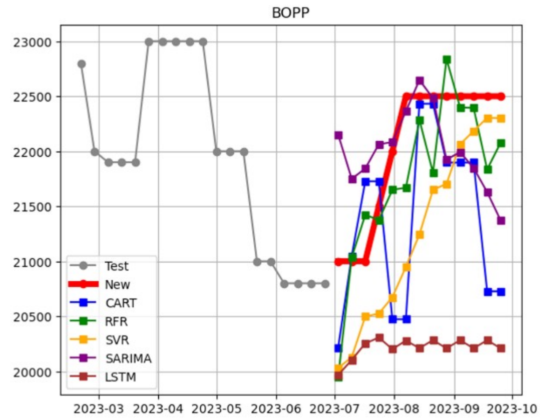


Figure 12. BOPP actual and model prediction

Model	MSE	RMSE	MAE	MAPE
CART	1155973.4	1075.2	833.231	3.7568
RFR	262266.04	512.12	412.615	1.8755
SVR	798379.31	893.52	788.508	3.5875
SARIMA	462603.61	224.59	571.562	2.6169
LSTM	3674381.1	70.55	1823.58	8.2031

Table 9. New BOPP data prediction metrics

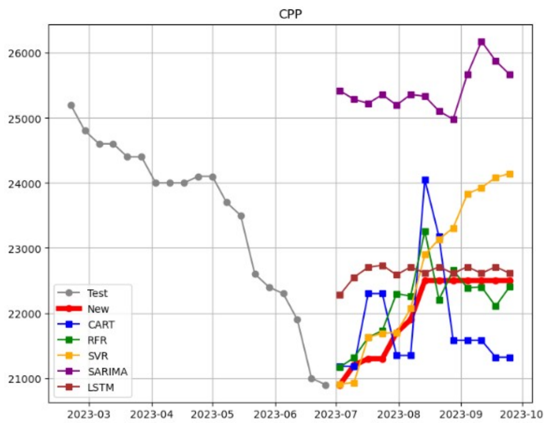


Figure 13. CPP actual and model prediction

Model	MSE	RMSE	MAE	MAPE
CART	820318.08	905.714	810.385	3.6575
RFR	133924.29	365.957	308.2	1.4047
SVR	814710.2	902.613	691.708	3.0955
SARIMA	12217522	3495.36	3446.67	15.752
LSTM	721606.04	849.474	642.023	2.9913

Table 10. New CPP data prediction metrics

Similar performance of models is observed in CPP in in Figure 13 and Table 10 where RFR outperforms other models and SVR can predict the trend. As for LSTM, although it results lower MAPE demonstrate rather flat prediction.

Therefore, based on the actual data from July to September 2023, RFR is the best fit model used to predict LLDPE, BOPP and CPP, while SVR fits for LDPE. As upstream data indices are updated on a weekly basis, LLDPE, LDPE, BOPP and CPP prediction price could be used as input for spot price negotiation that could take place on a weekly basis with the local supplier. Meanwhile, for imported goods purchased from overseas suppliers, LSTM could be used for qualitative forecast whether the price will increase or decrease within 4 weeks. Lastly, SARIMA is found to be not fit for LLDPE, LDPE, BOPP and CPP prediction.

The limitation of this study that might contribute to the performance of regression and time-series model prediction might be caused by the number of total data and limited hyperparameters tuning conducted.

4.2. Inventory Optimization

One objective of this research is how to benefit machine learning and deep learning models for price prediction that provide input for developing purchasing strategy to optimize the inventory and its value. The predicted value is input for Linear Programming (LP) that considers the predicted price fluctuation to determine the quantity to purchase. The optimization problem brought in this paper is how to minimize cost and purchase inventory based on future price. Linear Programming optimization is selected for price and inventory optimization as there is linear correlation between the price, quantity, and the value of material. The optimization is performed using simplex method in Excel Solver.

In general, the organization have implemented strict control of inventory, therefore the planner maintains the stock within the allowable MES ratio as pointed in Equation (14) where the stock at end of month should be within 0.3-1.0 ratio. This optimization works to squeeze further opportunity to lower the purchase cost. Referring to Table 11, purchasing period is denoted with P1, P2 and P3 with P0 as current period. Beginning stock is the last inventory quantity accumulated at the end of previous period.

LLDPE	Base Scenario							Optimized Scenario					
PERIOD	Beginning Stock	Base Purchase Quantity	Usage Plan	Month End Stock	MES Ratio (%)	Predicted Price	Base Inventory Value	Beginning Stock	Optimized Purchase Quantity	Usage Plan	Optimized Month End Stock	Optimized MES Ratio (%)	Optimized Inventory Value
P0			300	200	0.7		-			300	200	0.7	-
P1	200	300	300	200	0.7	1,129.4	338,820	200	400	300	300	1.0	451,760
P2	200	250	300	150	0.5	1,187.9	296,975	300	300	300	300	1.0	356,370
P3	150	300	300	150	0.5	1,196.2	358,860	300	150	300	150	0.5	179,430
TOTAL		850					994,655		850				987,560
												GOAL	(7,095)

Table 11. Comparison between base and optimized scenario of LLDPE purchase

Inventory Planner release Purchase Requisition (PR) quantity identified in Base Purchase Quantity column with quantity slotted in P1, P2 and P3. The purchase request quantity in P1, P2 and P3 are regulated by the Month End Stock (MES) ratio as calculated in Equation (14) that reflects the organization inventory control policy. The MES ratio value is typically maintained at 0.5 for LLDPE and LDPE as these materials are mostly imported. For locally supplied material such as BOPP and CPP it is maintained at min 0.3 due to shorter lead time. Predicted value is prediction result of in the last week of each month that reflect the supplier's new offering time window. Base Inventory Value is the multiplication between Base Purchase Quantity and the Predicted value. Optimized Inventory Value is the multiplication between Optimized Purchased Quantity as decision variable and the predicted price. The Optimized Purchased Quantity must not exceed the Base Purchase Quantity as described as optimization constraint in Equation (13). The objective of optimization is determining Optimized Purchase Quantity at each period as decision variable to minimize inventory value.

In Table 11, LLDPE price is predicted to increase from P1 to P3. Therefore, to minimize the inventory value, LP set MES of P1 and P2 to maximum 1.0 which plot them Figure 14a as higher than base scenario. However, as more quantity has been purchased in P1 and P2, less quantity in P3 as remainder is released when the price reached the highest point. As result, total inventory value after optimization in Figure 14b is reduced by 0.71% compared to the base scenario.

Next product is LDPE. In Table 12, the base scenario has indicated inventory control that is in adherence with organizational policy since the planner set the base purchase quantity if P1, P2 and P3 according to MES ratio 0.5 from previously high in P0 at 1.2.

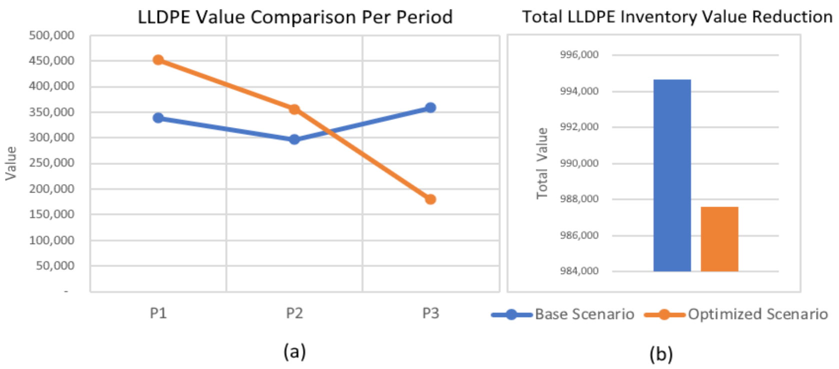


Figure 14. LLDPE value comparison per period and total inventory value reduction

LDPE PERIOD	Base Scenario: Beginning Stock	Base Purchase Quantity	Usage Plan	Month End Stock	MES Ratio (%)	Predicted Price	Base Inventory Value	Optimized Scenario: Beginning Stock	Optimized Purchase Quantity	Usage Plan	Optimized Month End Stock	Optimized MES Ratio (%)	Optimized Inventory Value
P0			300	350	1.2		-			300	350	1.2	-
P1	350	100	300	150	0.5	1260.4	126,040	350	250	300	300	1.0	315,100
P2	150	300	300	150	0.5	1303.1	390,930	300	300	300	300	1.0	390,930
P3	150	300	300	150	0.5	1321.1	396,330	300	150	300	150	0.5	198,165
TOTAL		700					913,300		700				904,195
												GOAL	(9,105)

Table 12. Comparison between base and optimized scenario of LDPE purchase

To optimize the inventory value on increasing predicted price in P1, P2 and P3, LP maximizes the allowable purchase quantity by advancing quantity to earlier period which plot P1 in Figure 15a higher than base. However, as more quantity has been purchased earlier in P1 and P2, less quantity is purchased in P3 when the predicted price is highest. As result, total LDPE inventory value is reduced by 1.0% as shown in Figure 15b.

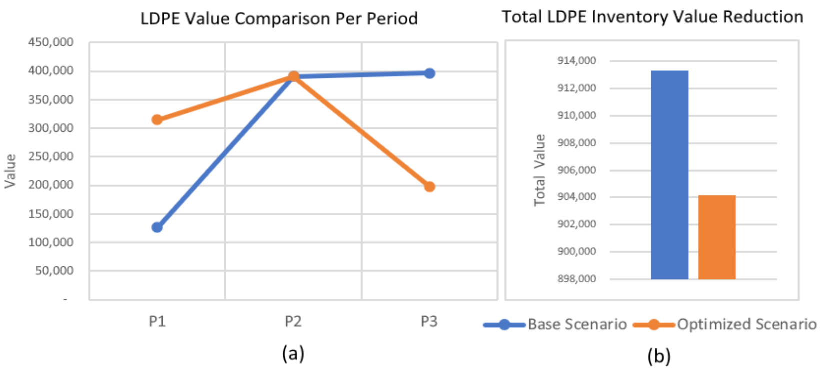


Figure 15. LDPE value comparison per period and total inventory value reduction

BOPP base scenario indicates a low inventory control with MES ratio 0.3 at P0 as in Table 13. Predicted price indicate increase on P1 and P2 then decline in P3.

BOPP PERIOD	Base Scenario: Month Beginning Stock	Base Purchase Quantity	Monthly Usage Plan	Month End Stock	MES Ratio (%)	Average Predicted Price	Base Inventory Value ('000)	Optimized Scenario: Month Beginning Stock	Optimized Purchase Quantity	Monthly Usage Plan	Optimized Month End Stock	Optimized MES Ratio (%)	Optimized Inventory Value ('000)
P0			150	50	0.3		-			150	50	0.3	-
P1	50	100	120	30	0.3	21,649.2	2,164,920	50	130	120	60	0.5	2,814,396
P2	30	100	100	30	0.3	22,839.2	2,283,920	60	90	100	50	0.5	2,055,528
P3	30	100	100	30	0.3	22,078.6	2,207,860	50	80	100	30	0.3	1,766,288
TOTAL		300					6,656,700		300				6,636,212
												GOAL	(20,488)

Table 13. Comparison between base an optimized scenario of BOPP purchase

Despite of already in a tight inventory control, the inventory value optimization could still be achieved because LP maximizes purchase on P1 and P2 with limited MES ratio at 0.5. Figure 16a shows progression of value reduction since P2 because larger quantity was processed P1 before price rise to the highest point in P2.

As result, total BOPP inventory value after optimization is reduced by 0.31% from base scenario as seen in Figure 16b.

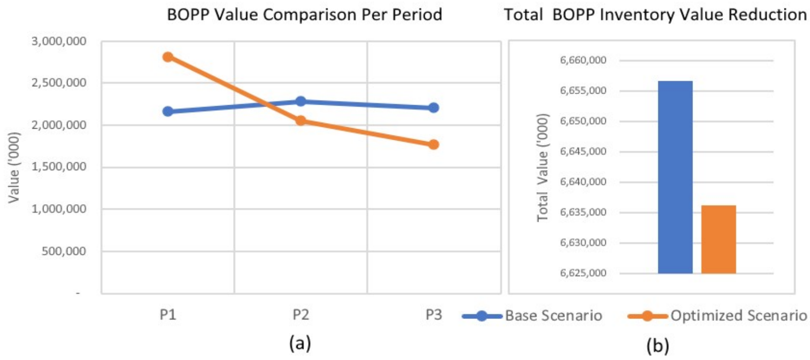


Figure 16. BOPP value comparison per period and total inventory value reduction

In regards of CPP, Table 14 shows that inventory position in P0 is at maximum level as indicated by MES ratio 1.0. Base quantity set by the planner is intended to restore the MES ratio at 0.5 in P3 while the price is predicted to exhibit an increase in P2 and then decrease in P3.

CPP	Base Scenario							Optimized Scenario					
PERIOD	Beginning Stock	Base Purchase Quantity	Usage Plan	Month End Stock	MES Ratio (%)	Predicted Price	Base Inventory Value ('000)	Beginning Stock	Optimized Purchase Quantity	Usage Plan	Optimized Month End Stock	Optimized MES Ratio (%)	Optimized Inventory Value ('000)
P0			75	75	1.0		-			75	75	1.0	-
P1	75	60	70	65	0.9	22299.4	1,337,964	75	65	70	70	1.0	1,449,461
P2	65	60	70	55	0.8	22658.3	1,359,498	70	35	70	35	0.5	793,040
P3	55	50	70	35	0.5	22408.6	1,120,430	35	70	70	35	0.5	1,568,602
TOTAL		170					3,817,892		170				3,811,104
												GOAL	(6,788)

Table 14. Comparison between base an optimized scenario of CPP purchase

LP works to maximize the quantity at P1 when the price is at the lowest then LP reduces the quantity at P2 when the price is the highest. Therefore, the value of P2 plot in Figure 17a drops. The remaining quantity is released as purchase order in P3. When the price declines, more quantity is placed to meet MES ratio at 0.5. As result, Figure 17b shows CPP inventory value reduction by 0.31 % compared to base scenario.

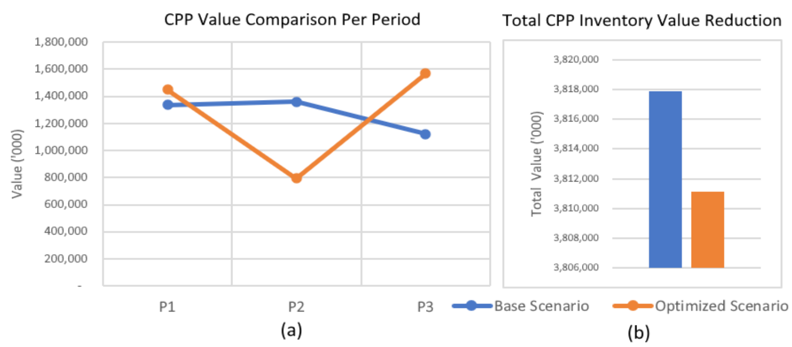


Figure 17. CPP value comparison per period and total inventory value reduction

5. Conclusion

Commodity markets are complex and subject to a wide range of factors that cause price fluctuations. Supply and demand are the most fundamental reason that could be influenced by other factors such as economic condition, governmental policies, geopolitical tensions, weather, and natural disasters. These factors cause high price volatility and speculation of what to come. In the era of Industry 4.0, the implementation of Machine Learning and Deep Learning approach in the Procurement department strengthen strategic role to predict material future price and optimize the inventory value. Regression models such as CART, RFR and SVR offer better prediction performance of LLDPE, LDPE, BOPP and CPP materials in comparison to time-series models such as classic SARIMA and Deep Learning model LSTM. Among the models, RFR is the best predictor for LLDPE, BOPP and CPP while SVR is best for LDPE. The follow up of prediction result to linear programming delivers inventory quantity and value optimization to deal with future price trends. Linear programming enables Procurement whether to purchase material earlier or postpone minimizing the inventory value. As result combined value reduction is 2.2 % which contributes to competitive advantage of company in price sensitive market and global economy situation that diminishes demand. Future research is needed to address other factors' role that is currently present in the commodity value chain of petrochemical products delivery from the producer to the user such as the freight cost.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

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

References

An, J., Mikhaylov, A., & Jung, S.U. (2021). A Linear Programming approach for robust network revenue management in the airline industry. *Journal of Air Transport Management,* 91, 101979. https://doi.org/10.1016/j.jairtraman.2020.101979

Chen, E., & He, X.J. (2019). Crude Oil Price Prediction with Decision Tree Based Regression Approach. *Journal of International Technology and Information Management,* 27(4), 2-16. https://doi.org/10.58729/1941-6679.1391

Chen, Z., Goh, H.S., Sin, K.L., Lim, K., Chung, N.K., & Liew, X.Y. (2021). Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques. *Advances in Science, Technology and Engineering Systems Journal,* 6(2).

Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., & Mosavi, A. (2019). An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of The Total Environment,* 651, 2087-2096. https://doi.org/10.1016/j.scitotenv.2018.10.064

Ensafi, Y., Amin, S.H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights,* 2(1), 100058. https://doi.org/10.1016/j.jjimei.2022.100058

Gargano, A., & Timmermann, A. (2014). Forecasting commodity price indexes using macroeconomic and financial predictors. *International Journal of Forecasting,* 30(3), 825-843. https://doi.org/10.1016/j.ijforecast.2013.09.003

Ghiasi, M.M., Zendehboudi, S., & Mohsenipour, A.A. (2020). Decision tree-based diagnosis of coronary artery disease: CART model. *Computer Methods and Programs in Biomedicine,* 192, 105400. https://doi.org/10.1016/j.cmpb.2020.105400

Hamed, J., Richa, J., & Dmitry, I. (2023). Data science and big data analytics: a systematic review of methodologies used in the supply chain and logistics research. *Annals of Operations Research*, 1-58. https://doi.org/10.1007/s10479-023-05390-7

Han, J., Kamber, M., & Pei, J. (2012). *Classification.* Elsevier. https://doi.org/10.1016/B978-0-12-381479-1.00008-3

Jahanshahi, H., Uzun, S., Kaçar, S., Yao, Q., & Alassafi, M.O. (2022). Artificial Intelligence-Based Prediction of Crude Oil Prices Using Multiple Features under the Effect of Russia–Ukraine War and COVID-19 Pandemic. *Mathematics,* 10(22), 4361. https://doi.org/10.3390/math10224361

Jang, H. (2019). A decision support framework for robust R&D budget allocation using machine learning and optimization. *Decision Support Systems,* 121, 1-12. https://doi.org/10.1016/j.dss.2019.03.010

Kamdem, J.S., Essomba, R.B., & Berinyuy, J.N. (2020). Deep learning models for forecasting and analyzing the implications of COVID-19 spread on some commodities markets volatilities. *Chaos, Solitons & Fractals,* 140, 110215. https://doi.org/10.1016/j.chaos.2020.110215

Kotu, V., & Deshpande, B. (2019). *Classification.* Elsevier. https://doi.org/10.1016/B978-0-12-814761-0.00004-6

Kwon, H., Do, T.N., & Kim, J. (2020). Comprehensive Decision Framework Combining Price Prediction and Production-Planning Models for Strategic Operation of a Petrochemical Industry. *Industrial & Engineering Chemistry Research,* 59, 11610-11620. https://doi.org/10.1021/acs.iecr.0c01957

Mondal, P., Shit, L., & Goswami, S. (2014). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications,* 4(2), 13-29. https://doi.org/10.5121/ijcsea.2014.4202

Ouyang, H., Wei, X., & Wu, Q. (2019). Agricultural commodity futures prices prediction via long- and short-term time series network. *Journal of Applied Economics,* 22(1), 468-483. https://doi.org/10.1080/15140326.2019.1668664

Passos, D., & Mishra, P. (2022). A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks. *Chemometrics and Intelligent Laboratory Systems,* 223, 104520. https://doi.org/10.1016/j.chemolab.2022.104520

Perez, H.D., Hubbs, C.D., Li, C., & Grossmann, I.E. (2021). Algorithmic Approaches to Inventory Management Optimization. *Processes,* 9. https://doi.org/10.3390/pr9010102

Profillidis, V.A., & Botzoris, G.N. (2019). Chapter 6 - Trend Projection and Time Series Methods. *Modeling of Transport Demand* (225-270). Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-12-811513-8.00006-6

Salem, N., & Hussein, S. (2019). Data dimensional reduction and principal components analysis. *Procedia Computer Science,* 163, 292-299. https://doi.org/10.1016/j.procs.2019.12.111

Sangaiah, A.K., Tirkolaee, E.B., Goli, A., & Dehnavi-Arani, S. (2020). Robust optimization and mixed-integer linear programming model for LNG supply chain planning problem. *Soft Computing,* 24(11), 7885-7905. https://doi.org/10.1007/s00500-019-04010-6

Sen, A., Choudhury, K.D., & Datta, T.K. (2023). An analysis of crude oil prices in the last decade (2011-2020): With deep learning approach. *PLOS ONE,* 18(3), e0268996. https://doi.org/10.1371/journal.pone.0268996

Speiser, J.L., Miller, M.E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications,* 134, 93-101. https://doi.org/10.1016/j.eswa.2019.05.028

Sung, C., Kwon, H., Lee, J., Yoon, H., & Moon, I. (2012). *Forecasting Naphtha Price Crack Using Multiple Regression Analysis.* Elsevier. https://doi.org/10.1016/B978-0-444-59507-2.50021-4

Urolagin, S., Sharma, N., & Datta, T.K. (2021). A combined architecture of multivariate LSTM with Mahalanobis and Z-Score transformations for oil price forecasting. *Energy,* 231, 120963. https://doi.org/10.1016/j.energy.2021.120963

Vijh, M., Chandola, D., Tikkiwal, V.A., & Kumar, A. (2020). Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science,* 167, 599-606. https://doi.org/10.1016/j.procs.2020.03.326

Vishwas, B.V., & Patel, A.S. (2020). *Hands-on Time Series Analysis with Python.* New York: Apress. https://doi.org/10.1007/978-1-4842-5992-4

Vo, A.H., Nguyen, T., & Le, T. (2020). Brent Oil Price Prediction Using Bi-LSTM Network. *Intelligent Automation & Soft Computing,* 26(4), 1307-1317. https://doi.org/10.32604/iasc.2020.013189

Xu, Z., Lian, J., Bin, L., Hua, K., Xu, K., & Chan, H.Y. (2019). Water Price Prediction for Increasing Market Efficiency Using Random Forest Regression: A Case Study in the Western United States. *Water,* 11(2), 228. https://doi.org/10.3390/w11020228