CHEMICAL ENGINEERING TRANSACTIONS
VOL. 103, 2023
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Petar S. Varbanov, Panos Seferlis, Yee Van Fan, Athanasios Papadopoulos

**ISBN** 979-12-81206-02-1; **ISSN** 2283-9216
DOI: 10.3303/CET23103045

# Machine Learning Approach for Pump Price Prediction for the Philippines Post COVID-19 Pandemic and Amidst Russia-Ukraine Conflict

Sophia Bernadette R. Lunor[a], Jan Goran T. Tomacruz[a], Miguel Francisco M. Remolona[b], Joey D. Ocon[a,*]

[a]Laboratory of Electrochemical Engineering, Department of Chemical Engineering, University of the Philippines Diliman, Quezon City 1101, Philippines
[b]Chemical Engineering Intelligence Learning Laboratory, Department of Chemical Engineering, University of the Philippines Diliman, Quezon City 1101, Philippines
jdocon@up.edu.ph

The continued increase in national energy demand pushes oil and petroleum price prediction efforts for the net oil-importing Philippines to ensure adequate supply. These prices are commonly modeled by data-driven Machine Learning (ML) methods to encompass their extrinsic and volatile nature. However, recent studies have found that new features specific to COVID-19 and the Russia-Ukraine geopolitical conflict have significantly contributed to overseas oil price ML prediction. This work investigated the impact of these factors on seven pump prices in Manila, Philippines, from December 2019 until July 2022. Three ML regression models were chosen to extend the existing literature and ensure price model accuracy: Multiple Linear Regression (MLR), Support Vector Regression (SVR), and Random Forest Regression (RFR). Features were listed based on related literature and underwent data preprocessing using p-value testing and Principal Component Analysis. Models were then trained, tested, and optimized using nested splits and hyperparameter tuning. Mean Absolute Percent Error (MAPE) was used to evaluate accuracy. Generated models had MAPE values within the range 3.13 % - 12.67 %, which is within the range of MAPE values in oil and petroleum price prediction literature, 0.131 % - 19.2 %. MLR and SVR models generally exhibited the highest accuracy for each pump price. This study proves that period-specific features may be used for local pump price modeling. Future works may explore other ML models and geographic location effects and investigate newly identified period-specific features.

## 1. Introduction

Since all countries are consumers of oil and its derivatives, it is a major commodity with high volatility (An et al., 2019). Oil commodity prices are driven by supply from exporters and demand from industrialized countries (Gao et al., 2022). Continuous population rise, industrial development, and economic growth have tremendously increased Philippine energy demand. By 2040, the Philippine national energy demand is projected to reach 99.3 Mtoe, following an estimated 5.8 % annual increase in demand. Much of this demand is expected to be fulfilled by oil which held a major energy share of 49.4 % in 2020 (Department of Energy, 2022). As a small open economy and net oil-importing country, the Philippines is an oil price taker and is affected by not only economics but extrinsic elements such as social, political, and environmental factors (An et al., 2019). Recent global events, particularly the COVID-19 Pandemic (Gao et al., 2022) and the Russia-Ukraine War have been shown to affect oil commodity prices, with drastic changes in the economic activity affecting the market (Liadze et al., 2022). With changes in the economic landscape due to global crises, this study aims to establish how these events affect Philippine oil and gas products to understand its economic drivers and meet its energy needs.

Limited studies incorporate extrinsic features for net oil price takers like the Philippines. However, various methods and strategies are used to model global benchmark crude oil prices. In the last decade, machine learning methods have been commonly used due to their advantageous prediction accuracy and ability to incorporate factors (social, political, environmental, and economic) as inputs in their predictive models (Patel

Paper Received: 05 April 2023; Revised: 05 July 2023; Accepted: 27 July 2023
Please cite this article as: Lunor S.B.R., Tomacruz J.G.T., Remolona M.F.M., Ocon J.D., 2023, Machine Learning Approach for Pump Price Prediction for the Philippines Post COVID-19 Pandemic and Amidst Russia-Ukraine Conflict, Chemical Engineering Transactions, 103, 265-270 DOI:10.3303/CET23103045

and Shah, 2021). These ML techniques generate more comprehensive models that are representative of the turbulent nature of the oil market (Gao et al., 2022). Tuna and Tuna (2022) found that the number of new COVID-19 cases worldwide and the infectious disease equity market volatility index affected the global benchmark crude oil price. Kamdem et al. (2020) quantitatively determined that the total number of local COVID-19 active cases and deaths impacted global crude oil price volatility. In addition, Firouzjaee and Khaliliyan (2022) observed that the correlation between the Russian Ruble and the Ukrainian Hryvnia currency to the global crude oil price index had increased during the conflict through a comparison of the correlations pre- and mid-crisis. The currencies were normalized to the Chinese yuan, which showed no significant movement over the conflict. The change in correlation was linked to Russia's increased interest rate growth policy and shifts in food and energy export movement in Russia and Ukraine during the crisis, which affected both the exchange rates and the oil price index. In this work, ML models were trained to model seven pump prices in Manila, Philippines using economic and crisis-based features for ML model fitting to determine how consistent these relationships are with local pump prices.

## 2. Methodology

The methodology of this work is summarized in Figure 1 below. Features were identified, and data was collected from online databases to compose the dataset for each pump price. Next, datasets were statistically tested and transformed to reduce feature intercorrelation. Then, ML regression techniques (MLR, SVR, RFR) were used to model pump prices. Finally, model accuracies were evaluated and compared to existing studies.

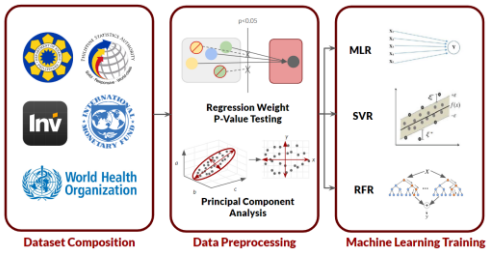


*Figure 1: Schematic of methodology flow in this work*

### 2.1 Dataset composition

*Table 1: Proposed features for the pump price dataset*

| Property Group | Feature | Description | Data Source | Literature Reference |
|---|---|---|---|---|
| Economic | DCO | Representative global crude oil market price | [a] | [j] |
| | CPI | Change in the average retail prices of a fixed household goods and services relative to a base year | [b] | [j] |
| | INF | Year-on-year change of the CPI relative to a base year | [c] | [j] |
| | PPP | Measure of the real Philippine peso value relative to a base year | [c] | [j] |
| | PPI | Change in average producer prices of fixed goods relative to a base year | [b] | [j] |
| | USD | Average value of the Philippine peso relative to a US dollar | [e] | [j] |
| Pandemic | NWC | End of interval value of newly reported worldwide cases | [f] | [k] |
| | EMV | Impact of infectious diseases on equity market volatility and risk | [g] | [k] |
| | PHC | End of interval value of total reported active Philippine cases | [f] | [l] |
| | PHD | End of interval value of total reported Philippine deaths from COVID-19 | [f] | [l] |
| Conflict | RUB | Average value of the Russian ruble relative to the Chinese yuan | [h] | [m] |
| | UAH | Average value of the Ukrainian hryvnia relative to the Chinese yuan | [i] | [m] |

[a] (Federal Reserve Bank of St. Louis, 2022b), [b] (Philippine Statistics Authority, 2022a), [c] (Philippine Statistics Authority, 2022b), [e] (Investing.com, 2022b), [f] (World Health Organization, 2022), [g] (Federal Reserve Bank of St. Louis, 2022a), [h] (Investing.com, 2022c), [i] (Investing.com, 2022a), [j] (Urrutia et al., 2018), [k] (Tuna and Tuna, 2022), [l] (Kamdem et al., 2020), [m] (Firouzjaee and Khaliliyan, 2022)

The desired model outputs were pump prices, while the significant features were set as the model inputs. Prices of pump products monitored by the Department of Energy in the Philippines during the crises, from December

2019 to July 2022, were studied: RON 91, RON 95, RON 97, RON 100, Diesel, Diesel Plus, and Kerosene. The input features were listed from studies that demonstrated statistical causality or significant relation between the feature and oil prices (summarized in Table 1). The economic features reflect the national economic state and global price dynamics, while the pandemic and conflict features depict their respective global crises.
Economic feature data sources provided data at monthly intervals, which limited the dataset size. As such, two hypotheses were explored to expand the dataset to weekly intervals. First, the hypothesis that these economic data are constant within the month, and second, the hypothesis that these economic data progress at a linear basis between its two monthly values, where the weekly values could then be interpolated. These hypotheses investigate model dataset size requirements and their effect on model accuracy. Three datasets of different intervals were then constructed: monthly intervals, weekly intervals with constant economic data assumption, and weekly intervals with linearly regressed economic data assumption. The monthly intervals had 32 data samples, and the weekly intervals had 140 data samples.

### 2.2 Data preprocessing

The datasets underwent an initial multiple linear regression weight p-value test to determine individual feature significance. A p-value of 0.05 was used, following the methodology of Urrutia et al. (2018) to increase ML model accuracy. The simplified datasets comprising of significant features per price model were processed through Principal Component Analysis (PCA) to reduce multicollinearity and increase model interpretability (Alredany, 2018). PCA produces a set of new linearly independent variables through orthogonal linear transformations on the original datasets (Jolliffe and Cadima, 2016). This study retained 95 % cumulative variability during PCA to reduce data noise and variable intercorrelation.

### 2.3 ML prediction models

Regression analysis has been used to relate crises to economic developments (Hoke and Tomaštík, 2022). Three regression models were then used for pump price prediction (described in Table 2): Multiple Linear Regression (MLR), Support Vector Regression (SVR), and Random Forest Regression (RFR). These ML methods were chosen to extend the scope of a previous study on Philippine pump price prediction using MLR (Urrutia et al., 2018) and to use accurate oil price ML models based on literature using MLR, SVR, and RFR (Jahanshahi et al., 2022). These three ML models were tested for each of the dataset intervals and per pump price.

*Table 2: Comparison of the Three Regression Models and Their Hyperparameter Tunings in This Study*

| Model Name | MLR | SVR | RFR |
|---|---|---|---|
| Working principle [a] | Models a response variable as a function of linearly related predictor variables | Fits a decision boundary within a threshold of values | Constructs multiple decision trees and outputs the mean/ mode of the trees' predictions |
| Hyperparameter tuning technique | - | RandomizedSearchCV (Five nested splits, Four months testing size, 300 iterations in total) | RandomizedSearchCV (Five nested splits, Four months testing size, 600 iterations in total) |
| Optimized hyperparameters | - | kernel, C, gamma | bootstrap, max_depth, min_samples_split, min_samples_leaf n_estimators, random_state |

[a] (Paper, 2020)

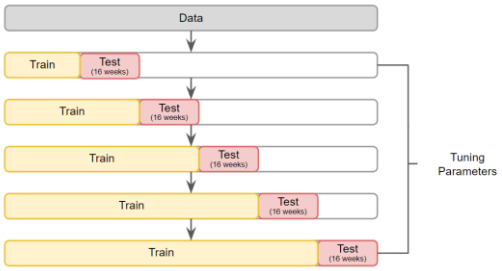


*Figure 2: Schematic for Five-fold Nested Temporal Cross Validation*

In training, only a portion of the data called the training set was used to fit the pump price model. The resulting model's accuracy was based on its estimation of the testing data across all training-testing splits. Nested Temporal Cross-Validation (schematic in Figure 2) generated five different training-testing data splits, with the test set size set at four months. A chronological data split was done to the price time series model to avoid data leaking between the training and testing sets. Regression model parameters were optimized by across the five data spits. However, the hyperparameter tuning technique performed did not consider all possible parameter combinations and only used parameter sets generated at random. Three rounds of hyperparameter tuning were then performed to generate the optimized pump price models. This validation method reduced model overfitting and ensured that the training-testing splits were not arbitrarily determined.

### 2.4 ML model accuracy

Models were evaluated using their Mean Absolute Percentage Error (MAPE) values, an indicator of accuracy based on the relative error between the actual and predicted values and is commonly used as a loss function for regression analysis (de Myttenaere et al., 2016). As such, it is used as a benchmark in oil price prediction (Lu et al., 2021). MAPE is an intuitive measurement, with zero as its best value. Its formula is given by Eq(1), where $\hat{y}_i$ is the actual value, $y_i$ is the predicted value, and $n$ is the number of data points.

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

## 3. Results and discussions

Insignificant features from the p-value test were dropped for each pump price model. For feature datasets in the same interval, Diesel and Diesel Plus had a different set of significant features than all other pump products, indicating varying market dynamics. Identified significant features were transformed into four to five principal components that accounted for 95 % cumulative variability in their respective datasets (Scree plot in Figure 3).

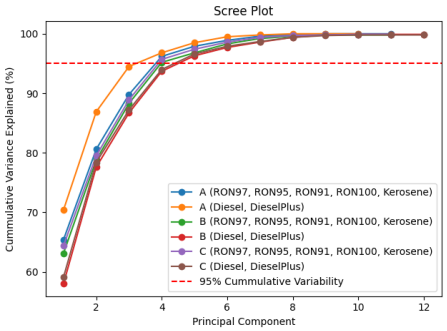


*Figure 3: PCA Scree Plot for All Datasets [A: Monthly Interval, B: Weekly Interval (Constant Assumption), C: Weekly Interval (Linearly Regressed Assumption)]*

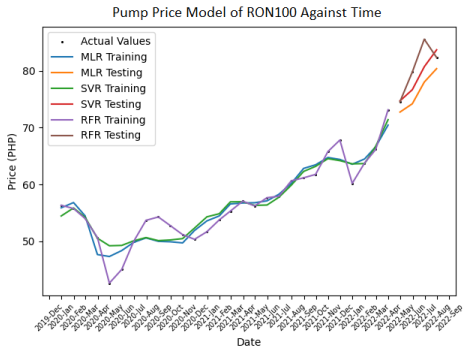


*Figure 4: Pump Price Model of RON100 With Monthly Interval Data Using MLR, SVR and RFR*

In the trained and optimized ML models, MLR and SVR models generally gave the best fit, and RFR was the least accurate (sample pump price model in Figure 4). A summary of the best and worst models per price, with their MAPE values, is shown in Table 3. No conclusive trends were observed between the different dataset time intervals investigated in the study.

*Table 3. Summary of Best and Worst ML Models per Pump Price and Their Testing Set MAPE Values*

| Pump Product | Model Information | | Model MAPE (%) | |
|---|---|---|---|---|
| | Best | Worst | Best | Worst |
| Diesel Plus | SVR, Weekly Interval (Constant Hypothesis) | RFR, Weekly Interval (Linearly Regressed Hypothesis) | 7.31 ± 2.57 | 11.66 ± 5.82 |
| Diesel | MLR, Monthly Interval | RFR, Weekly Interval (Linearly Regressed Hypothesis) | 5.57 ± 1.55 | 12.67 ± 6.23 |
| Kerosene | SVR, Weekly Interval (Constant Hypothesis) | RFR, Monthly Interval | 6.05 ± 4.77 | 9.75 ± 5.11 |
| RON 91 | SVR, Weekly Interval (Linearly Regressed Hypothesis) | RFR, Weekly Interval (Linearly Regressed Hypothesis) | 4.28 ± 1.38 | 7.40 ± 2.28 |
| RON 95 | MLR, Weekly Interval (Constant Hypothesis) | RFR, Weekly Interval (Constant Hypothesis) | 4.05 ± 2.51 | 7.22 ± 1.97 |
| RON 97 | SVR, Monthly Interval | RFR, Monthly Interval | 3.56 ± 1.66 | 6.94 ± 2.35 |
| RON 100 | SVR, Monthly Interval | RFR, Monthly Interval | 3.13 ± 1.83 | 6.50 ± 2.58 |

Comparing the results to related literature, the MAPE values of all models (MAPE = 3.13 % - 12.67 %) fit within the benchmarked error range (MAPE = 0.131 % - 19.2 %) published by Lu et al. (2021) in their decade review for oil price prediction. The MAPE threshold summarizes the error indicator values of all published oil price models covered in the review. This implies that the models generated are accurate and within the set standards.

## 4. Conclusion

This work uses ML regression models to analyze the effect of the COVID-19 outbreak and the Russia-Ukraine war on Philippine pump prices. Weekly and monthly price prediction datasets were constructed from economic and crisis-based features from online databases. Out of the finely-tuned and trained models, MLR and SVR models generally had the lowest MAPE values, indicating the highest accuracy. The work showed that ML models encompassing pandemic and conflict features reflected the same accuracy as that found in the literature. Overall, this work is a successful proof of concept that connects local pump prices with global phenomena.

Future works may explore the use of other classes of Machine Learning models, feature importance analysis, and price forecasting. These works may also investigate how the effect of location on prices since the study was limited to Manila, Metro Manila, Philippines. Finally, both global crises are still ongoing, and the complete short-term and long-term consequences of these events are still undetermined.

**Nomenclature**

COVID-19 – Coronavirus disease 2019
CPI – Consumer Price Index
CV – Cross-validation
DCO – Dubai Crude Oil Prices
EMV - Infectious Disease Equity Market Volatility Index
INF – Inflation Rate
MAPE – Mean Absolute Percentage Error
ML – Machine Learning
MLR – Multiple Linear Regression
NWC – Number of new worldwide COVID-19 cases
PCA – Principal Component Analysis
PHC - Number of Philippine COVID-19 cases
PHD - Number of Philippine COVID-19 deaths
PPI – Producer Price Index
PPP – Purchasing Power of Peso
RFR – Random Forest Regression
RON – Research Octane Number
RUB – Russian ruble/ Chinese yuan Exchange Rate
SVR – Support Vector Regression
UAH – Ukrainian hryvnia/Chinese yuan Exchange Rate
USD – PHP-USD Exchange Rate

**Acknowledgments**

S.B.R.L would like to acknowledge the Department of Science and Technology – Science Education Institute (DOST-SEI). The authors would like to acknowledge the Commission on Higher Education – Philippine California Advanced Research Institutes (CHED-PCARI) through the CIPHER Project (IIID 2018-008). The authors would also like to thank the Computing and Archiving Research Environment (COARE) of the Department of Science and Technology – Advanced Science and Technology Institute (DOST-ASTI) for facilitating the computations required for this study.

**References**

Alredany W.H.D., 2018, A Regression Analysis of Determinants Affecting Crude Oil Price. International Journal of Energy Economics and Policy, 8(4), 110–119, <econjournals.com/index.php/ijeep/article/view/6621>, accessed 23.09.2022.

An J., Mikhaylov A., Moiseev N., 2019, Oil Price Predictors: Machine Learning Approach. International Journal of Energy Economics and Policy, 9(5), 1–6, DOI: 10.32479/ijeep.7597.

de Myttenaere A., Golden B., Le Grand B., Rossi F., 2016, Mean Absolute Percentage Error for Regression Model. Neurocomputing, 192, 38–48, DOI: 10.1016/j.neucom.2015.12.114.

Department of Energy, 2022, Philippine Energy Plan 2020-2040, Republic of the Philippines Department of Energy, <doe.gov.ph/sites/default/files/pdf/pep/PEP%202022-2040%20Final%20eCopy_20220819.pdf>, accessed 18.04.2023.

Federal Reserve Bank of St. Louis, 2022a, Equity Market Volatility: Infectious Disease Tracker (INFECTDISEMVTRACKD), <fred.stlouisfed.org/series/INFECTDISEMVTRACKD>, accessed 01.12.2022.

Federal Reserve Bank of St. Louis, 2022b, Global price of Dubai Crude (POILDUBUSDQ), <fred.stlouisfed.org/series/POILDUBUSDQ>, accessed 01.12.2022.

Firouzjaee J.T., Khaliliyan P., 2022, Machine learning model to project the impact of Ukraine crisis, arXiv, DOI: 10.48550/arXiv.2203.01738.

Gao X., Wang J., Yang L., 2022, An Explainable Machine Learning Framework for Forecasting Crude Oil Price during the COVID-19 Pandemic. Axioms, 11(8), 374, DOI: 10.3390/axioms11080374.

Hoke E., Tomaštík M., 2022, Economic Impacts of the Covid-19 Pandemic on the National Economy of the Czech Republic. Chemical Engineering Transactions, 91, 85-90, DOI: 10.3303/CET2291015.

Investing.com, 2022a, CNY/UAH - Chinese Yuan Ukrainian Hryvnia, <investing.com/currencies/cny-uah-historical-data>, accessed 01.12.2022.

Investing.com, 2022b, PHP/USD - Philippine Peso US Dollar, <investing.com/currencies/php-usd>, accessed 01.12.2022.

Investing.com, 2022c, RUB/CNY - Russian Ruble Chinese Yuan, <investing.com/currencies/rub-cny-historical-data>, accessed 01.12.2022.

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(2), 4361, DOI: 10.3390/math10224361.

Jolliffe I.T., Cadima J., 2016, Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A, 374(2065), 20150202, DOI: 10.1098/rsta.2015.0202.

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 and Fractals, 140, 110215, DOI: 10.1016/j.chaos.2020.110215.

Liadze I., Macchiarelli C., Mortimer-Lee P., Juanino, P.S., 2022, The Economic Costs of the Russia-Ukraine Conflict, National Institute of Economic and Social Research, <niesr.ac.uk/wp-content/uploads/2022/03/PP32-Economic-Costs-Russia-Ukraine.pdf >, accessed 11.03.2023.

Lu H., Ma X., Ma M., Zhu S., 2021, Energy price prediction using data-driven models: A decade review, Computer Science Review, 39, 100356, DOI: 10.1016/j.cosrev.2020.100356.

Paper D., 2020, Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python. Apress, Berkeley, California, DOI: 10.1007/978-1-4842-5373-1.

Patel H., Shah M., 2021, Energy Consumption and Price Forecasting Through Data-Driven Analysis Methods: A Review. SN Computer Science, 2, 315(2021), DOI: 10.1007/s42979-021-00698-2.

Philippine Statistics Authority, 2022a, Price Indices. <openstat.psa.gov.ph/Database/Prices/Price-Indices> accessed 01.12.2022.

Philippine Statistics Authority, 2022b, Prices, <openstat.psa.gov.ph/PXWeb/pxweb/en/DB/DB__2M__PI__CPI__2018/?tablelist=true>, accessed 01.12.2022.

Tuna G., Tuna V.E., 2022, Are effects of COVID-19 pandemic on financial markets permanent or temporary? Evidence from gold, oil and stock markets. Resources Policy, 76, 102637, DOI: 10.1016/j.resourpol.2022.102637.

Urrutia J.D., Alair A.R., Iglesias S.A., Malvar R.J., Baccay E.B., Oliquino A.B., Gano L.A., 2018, Forecasting Petroleum Product Prices In the Philippines. Indian Journal of Science and Technology, 11(20), 1-7, DOI:10.17485/ijst/2018/v11i20/123340.

World Health Organization, 2022, COVID-19 Dashboard, <covid19.who.int/data>, accessed 01.12.2022.