Research Article

Forecasting The Price of Research Octane Number 97 (RON97) In Malaysia: A Box-Jenkins Approach

Nurul Syakinah Mohd Zaki[1], Noreha Mohamed Yusof[2,*]

[1] Faculty of Computer and Mathematical Sciences, University Teknologi MARA Negeri Sembilan, Seremban Campus, 70300 Negeri Sembilan; sykinhzack@gmail.com ; 0009-0006-3384-6629

[2] Faculty of Computer and Mathematical Sciences, University Teknologi MARA Negeri Sembilan, Seremban Campus, 70300 Negeri Sembilan; noreh144@uitm.edu.my; 0000-0002-4427-1684

* Correspondence: noreh144@uitm.edu.my; 60135306535.

Abstract: *In Malaysia, there are three types of gasoline which are Diesel, RON95 and RON97. In this study, the focus will be RON97. RON97 is one of the factors that can significantly impact Malaysia's economy. This fuel also has many factors that influence the price of it, which in turn affect both consumer and user. Consumers, businesses, policymakers, and energy industry players in Malaysia rely on accurate RON97 price forecasts for effective budgeting, cost management, economic planning, and strategic decision-making. The volatile nature of global crude oil prices, together with the influence of taxes, refining costs, and currency exchange rates, makes RON97 price prediction a complex challenge. This study addresses the need to determine the best forecasting model for monthly fuel price of RON97 between Box-Jenkins autoregressive integrated moving average (ARIMA) models. The data is taken from Malaysia's Official Open Data Portal, which consists of the price of Petroleum and Diesel in Malaysia. This data consists of monthly prices of RON97 fuel in Malaysia since January 2020 until January 2024. The ARIMA (1,1,1) model was selected as the best forecasting approach. The forecasted range of the price of RON97 generated by this model was between RM 3.37 and RM 3.62. Future research could investigate the use of external factors or seasonal components to improve model performance.*

Keywords: RON97, forecasting, ARIMA Model, AIC, BIC, 24 months

DOI: 10.5281/zenodo.15789499

1. INTRODUCTION

Oil, coal, and natural gas still hold a significant stake in the global energy portfolio, serving as pivotal industries within energy stock markets (Alam et al., 2023). Gasoline, regarded nearly as indispensable as food and clothing (Colladon et al., 2024), stands as a product of petroleum refining characterized by a diverse composition of volatile hydrocarbons and aromatic compounds. Gasoline and petrol are interchangeable terms used to describe a volatile and flammable fuel derived from petroleum refining, primarily utilized in spark-ignition engines like those in automobiles and certain aircraft. There are three types of gasoline used in Malaysia; Diesel, RON95 and RON97.

Figure 1 shows the differences prices of fuel in Malaysia. It indicates that in May 2024, the price of RON95, RON97 and Diesel are RM 2.05, RM2.15 and RM3.47 respectively. The price of RON95 and Diesel is subsidized by government. While global crude oil prices served as a primary driver, the price of RON97 in Malaysia between January 2020 and January 2024 was also subject to the influence of taxes, refining costs, and currency exchange rates, particularly given the absence of subsidies during this

period, unlike RON95 and Diesel. This study attempts to identify underlying trends and potential future price movements by analysing historical data on RON97 pricing.

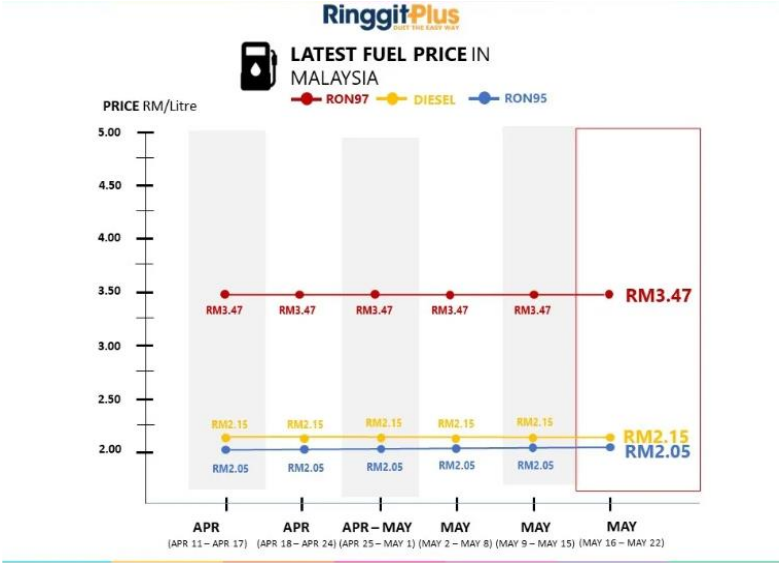

Figure 1. The differences price of RON95, RON97 and Diesel between April to May 2024 (Source: Ringgit Plus)

A study that has been done by Sokkalingam et al. (2021) used Autoregressive Integrated Moving Average (ARIMA) to model and forecast fuel price in Malaysia. The researcher used the historical data of RON97 price when the objective of this study was to assess the ability to forecast RON97 fuel price using ARIMA model and the outcome it is possible to forecast accurately fuel price using the ARIMA model but limited to a shorter period month. Other researchers also tried to examine the ARIMA model in various context such as Sarpong-Streetor et al. (2021) and Ers¨oz et al. (2022).

Additionally, Mabrouki et al. (2023) discussed on using ARIMA models to project the future prices of diesel and gasoline in Morocco over a period of eight years. The authors presented the findings that shows the ARIMA Model are the most effective when it come to forecasting the fuel price. ARIMA (3, 1, 2) was the better model produce the forecast values for diesel price.

Therefore, forecasting RON97 fuel prices has emerged as a crucial area of research, driven by the need to mitigate financial risks, inform strategic decision-making, and promote energy security and sustainability. This study aims to develop knowledge about RON97 fuel price forecasting by examining the historical data. Additionally, this study explores the effectiveness of various forecasting methodologies, including time series analysis, which is Box-Jenkins models in predicting RON97 fuel prices over different time horizons.

Consumers, businesses, policymakers, and energy industry players in Malaysia rely on accurate RON97 price forecasts for effective budgeting, cost management, economic planning, and strategic decision-making. The volatile nature of global crude oil prices, coupled with the influence of taxes, refining costs, and currency exchange rates, makes RON97 price prediction a complex challenge. This study addresses the need to determine the best forecasting model for monthly fuel price of RON97 between the ARIMA models. Thus, the need for a reliable and robust forecasting model for RON97 fuel prices that can predict the forecast in the next two years more accurately. As for the lack of accurate price forecasts affect consumers, business and others.

In this research, the objectives are to determine the best forecasting model for monthly fuel price of RON97 using ARIMA Model. Furthermore, this research also aim to forecast RON97 petrol fuel for the next 24 months ahead.

2. METHOD & MATERIAL

The data is taken from Malaysia's Official Open Data Portal (2025) which consist of price for Petroleum and Diesel in Malaysia. The target sample of this study is RON97 fuel price in Malaysia. This data consists of monthly price of RON95, Diesel and RON97 fuel in Malaysia since January 2020 until January 2024. The data that is used is the monthly data of RON97 fuel price. In this study, the method used is Box-Jenkins autoregressive integrated moving average (ARIMA).

2.1 ARIMA MODEL

Forecasting, control, and time-series analysis are popular applications of ARIMA modeling. The acronym ARIMA refers to the combination of Autoregressive, Integrated, and Moving Average models. Walker (1931) expanded on Yule's (1926) introduction of the Autoregressive (AR) model. Slutzky established the Moving Average (MA) model in 1937, and Wold offered the combination of these two models in 1983 (Lazim, 2021).

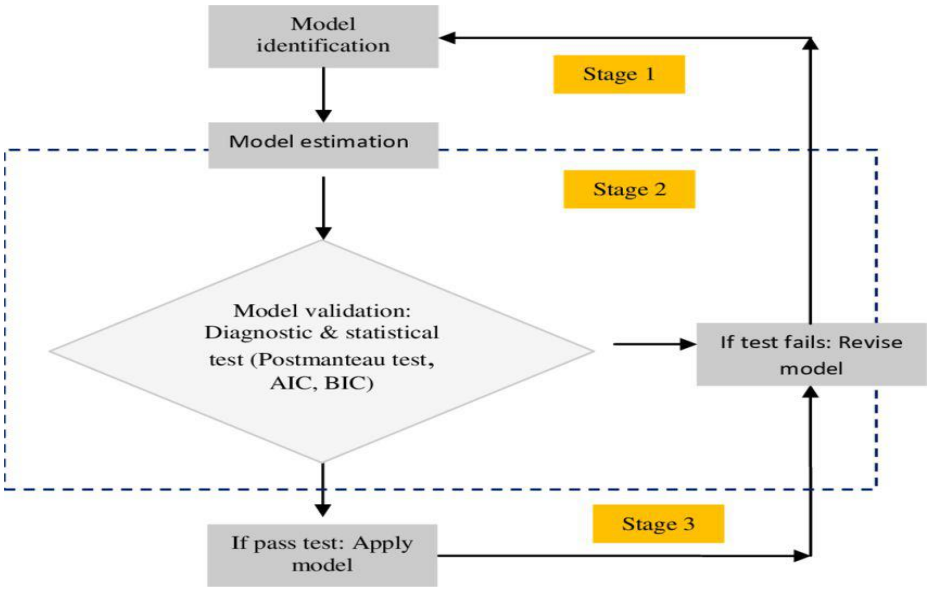

Figure 2. Stages in ARIMA Modelling

Figures 2 shows the stages in ARIMA Modelling which consists of Model Identification, Model Estimation, Model Validation and lastly Model Application.

The Augmented Dickey-Fuller (ADF) test represents an extension of the Dickey Fuller test, enabling the analysis of a larger set of time series models that exhibit greater complexity. The Augmented Dickey-Fuller test was used to determine stationary in time series data.

2.2 The Akaike's Information Criteria (AIC)

The AIC was commonly used metric of an ARIMA model's fitness. It calculated the penalty on likelihood for each new term introduced into the model. As a result, if the additional term does not increase the chance more than the penalty amount, it was not worth include in the model.

$$AIC = e^{\frac{2k}{T}} \left(\frac{\sum_{t=1}^{T} e_i^2}{T} \right)$$

Where k=p + q + P + Q denotes the number of parameters estimated in the model, p and q are the standard terms of the AR and MA sections, P and Q are the seasonality components of the ARIMA Model (if they exist), and T is the total number of observations in the data series. In fact $e^{2k/T}$ constitutes a penalty function whose goal is to prevent model's overfitting.

2.3 Bayesian Information Criterion (BIC)

Schwarz (1978) devised the Bayesian Information Criterion (BIC), also known as the Schwarz Criterion (SBC), to identify models that generated the most accurate out-of-sample forecasts while effectively balancing model complexity and goodness-of-fit. Unlike AIC and SBC, the BIC imposes a ignificantly larger penalty on degrees of freedom, as evidenced by the factor Tk/T in the following equation. Nevertheless, its similarity to AIC lies from the criterion utilized, selecting the best modelnwith the lowest SBC value.

$$BIC = T^{\frac{k}{T}} \left(\frac{\sum_{t=1}^{T} e_i^2}{T} \right)$$

Where k is the estimated model's parameter count, including the constant, and T is the number of observations in the series.

3. FINDINGS

3.1 Model Identification

3.1.1 Checking Stationary Condition

In order to check whether the data is stationary or not, ADF test is being used to prove the condition. The ADF test statistic shows that the value is -1.6191 with a p-value of 0.7357. Table 1 shows that p-value =0.7357 is larger than 0.05. Fail to reject the null hypothesis as the p-value is larger than alpha=0.05, which indicates that the time series is non-stationary.

Table 1. Table of Augmented Dickey-Fuller Test

Dickey-Fuller	P-value	Alpha
-1.6191	0.7357	0.05

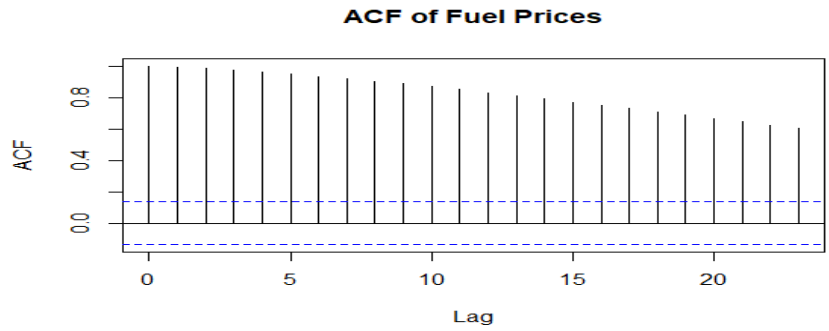

Figure 3. ACF plot of the price

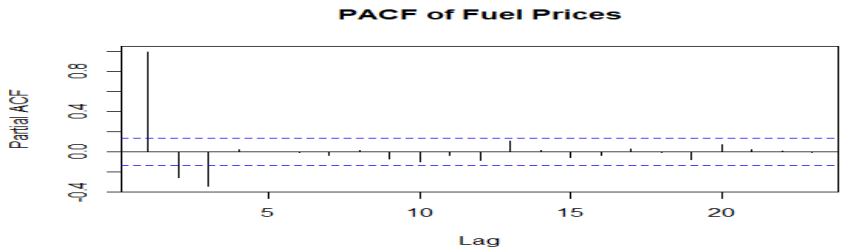

Figure 4. PACF plot of price

This is supported by Figure 3 and Figure 4. Based on the ACF and PACF plots, the data appears to be non-stationary. The ACF shows a slow decay, with several lags exhibiting significant autocorrelation, indicating the presence of a trend or persistent dependencies. In contrast, the PACF shows a significant spike at lag 1, which may suggest an autoregressive structure but does not confirm stationarity.

3.1.2 First Differencing

Figure 5 represents the first-differenced time series, where the y-axis indicates the changes (returns) between consecutive values in the original series. The fluctuations oscillate around zero, suggesting that the original series does not exhibit a strong trend but contains inherent variability. Noticeable spikes in the middle of the plot indicate periods of heightened volatility or abrupt changes in the underlying data. These spikes may correspond to external events or shifts in behaviour within the original series. Toward the end of the graph, the fluctuations appear to stabilize, implying a period of lower volatility or steadier changes. Overall, the graph reflects a time series with periodic variability and moments of significant change which could be relevant for understanding dynamics in the original dataset or preparing for further statistical modelling, such as stationarity testing or ARIMA forecasting.

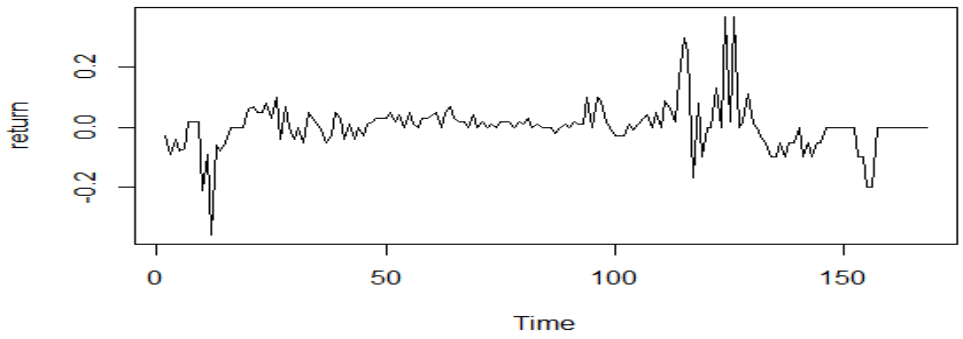

Figure 5. The first differencing plot of the RON97 price data

To check whether the data is stationary or not. ADF test is being used to prove the data. The ADF test statistic shows the value is -3.8417 with a p-value of 0.01865. From the information, p-value =0.01865 is smaller than 0.05. Reject the null hypothesis as the pvalue is smaller than alpha=0.05, which indicates that the time series is stationary.

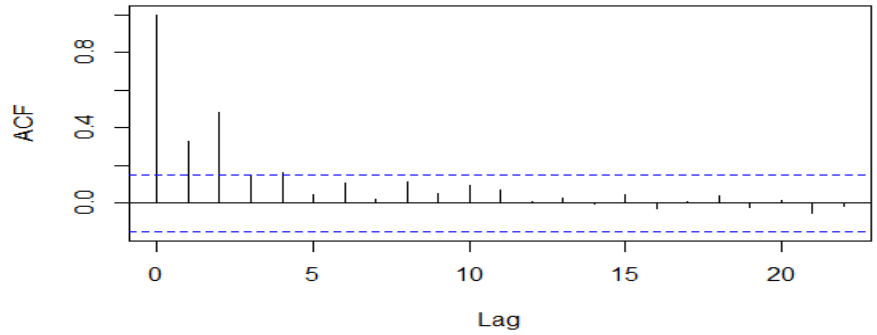

Figure 6. ACF plot after the first differencing

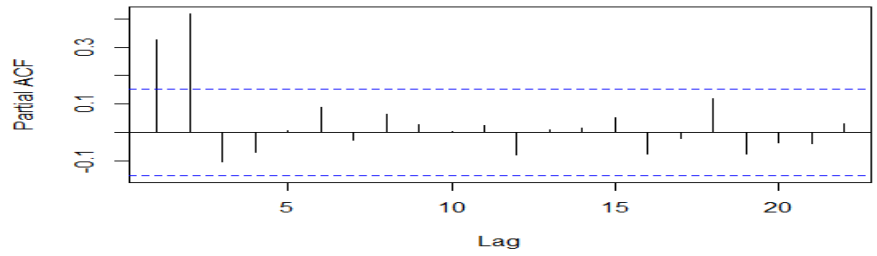

Figure 7. PACF plot after the first differencing

The ACF and PACF plots in Figure 6 and 7 suggest that the differenced time series is stationary, making it suitable for ARIMA modeling. The ACF plot shows a significant spike at lag 1 and 2, followed by a rapid decline within the confidence intervals, indicating the absence of a strong moving average

(MA) process. The PACF plot, on the other hand, exhibits a significant spike at lag 1 and 2, with subsequent lags tapering off, which is characteristic of an autoregressive (AR) process. Based on these patterns, an ARIMA (2, 1, 2) model is recommended. This model includes one autoregressive term (AR (2)), one differencing step (d = 1), and moving average terms (MA (2)). This structure reflects the series' reliance on its first lag and the fact that stationarity is achieved after differencing. Further diagnostics, such as residual analysis, should be performed to confirm the adequacy of the model and ensure the residuals exhibit white noise characteristics.

3.2 Model Estimation

The estimation of ARIMA models are developed. Four ARIMA models have been constructed to monitor the best ARIMA model for RON97 data set. The moduls are ARIMA(1,1,1), ARIMA (1,1,3), ARIMA(2,1,2) and ARIMA (2,1,3).

3.3 Model Validation

The adequacy of the models was assessed by utilizing Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) values. The criteria were chosen to minimize random errors and adhere to the principle of parsimony which sought to select models with the lowest unnecessary complexities.

Table 2. The ARIMA Model Validation

Criteria		**ARIMA(1,1,1)**	**ARIMA(1,1,3)**	**ARIMA(2,1,2)**	**ARIMA(2,1,3)**
Estimation Part	AIC	-401.1923	-416.75	-417.01	-415.3060
	BIC	-391.8383	-401.16	-401.42	-396.5980
Evaluation Part	AIC	-217.11	-213.11	-213.11	-211.11
	BIC	-211.97	-204.54	-204.54	-200.83

The lower the values of AIC and BIC, it indicates a better fit of the model compared to another model. To summarize the values in the table had shown, all of the models have different values. Thus, the most noticeable ARIMA Model is ARIMA (1,1,1) that have the lowest value of AIC which is -217.11 and BIC which is -211.97 suggesting that this method is relatively superior among the other models.

3.4 *Forecasting Research Octane Number 97 (RON97) In Malaysia Using ARIMA(1,1,1)*

The Box-Jenkins method was utilized to predict the price of RON97 beyond Jan 2024 using ARIMA (1,1,1) model. The table presented projected limits of the fuel price of RON97 for a period of 24 months from February 2024 until Jan 2026. Using ARIMA (1,1,1) to forecast as it has the best forecast accuracy.

Table 3. The summary forecasted value with the interval within next 24 months

Year	**Month**	**Lower limit forecast (RM)**	**Upper Limit Forecast (RM)**
2024	Feb	3.4496	3.5012
2024	Mac	3.4411	3.5141
2024	April	3.4346	3.5241
2024	May	3.4292	3.5324
2024	June	3.4244	3.5398
2024	July	3.4200	3.5465
2024	August	3.4160	3.5526
2024	Sept	3.4123	3.5583
2024	Oct	3.4088	3.5636

2024	Nov	3.4055	3.5687
2024	Dec	3.4023	3.5735
2025	Jan	3.3993	3.5781
2025	Feb	3.3964	3.5826
2025	Mac	3.3936	3.5868
2025	Apr	3.3909	3.5909
2025	May	3.3884	3.5949
2025	June	3.3858	3.5987
2025	July	3.3834	3.6024
2025	Aug	3.3810	3.6061
2025	Sept	3.3787	3.6096
2025	Oct	3.3765	3.6130
2025	Nov	3.3743	3.6164
2025	Dec	3.3721	3.6197
2026	Jan	3.3700	3.6229

Based on Table 3, a recursive multi-step forecasting technique was employed for multiple forecasts for twenty-four months. The forecasting model utilized the ARIMA (1,1,1) approach to predict the price of RON97 one step ahead and recently used those prediction inputs to generate the forecasted values. The forecasted range of the price of RON97 generated by this model was between RM 3.37 and RM 3.62. Additionally, the Figure 4.9 provides a visual presentation to enhance the understanding of the forecast.

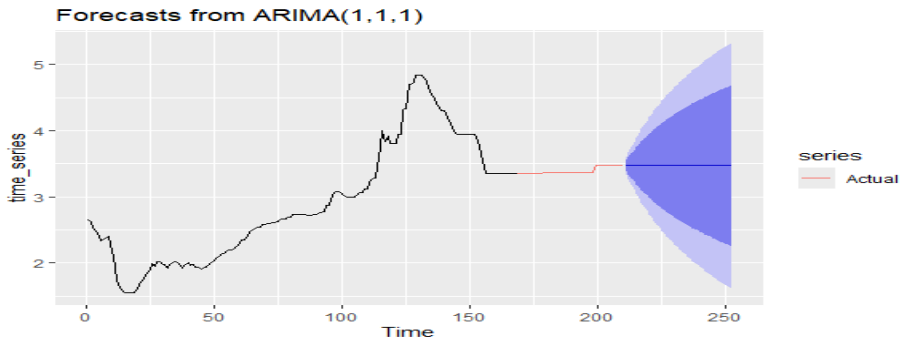

Figure 8. The forecasted plot of RON97 in the 24 months

The line graph depicted in Figure 8 shows the actual value of the price of RON97 the value of the model's predictions and the associated confidence levels, starting from January 2020 until January 2024. Starting from there, the black line shows the historical time series data that was used to train the model. The red line represents the predicted values for the future, while the blue-colored areas surrounding the forecasts reflect the prediction interval. The deeper blue region corresponds to a tighter confidence interval of 80%, while the lighter blue region shows a broader confidence interval of 95%. The forecasts indicate a consistent trend with increasing uncertainty over time, as evidenced by wider prediction intervals.

4. DISCUSSION

The ARIMA (1,1,1) model was determined to be the best match. The ARIMA model was created through a systematic identification process that included stationarity testing with the ADF test and evaluating ACF and PACF plots. The specified parameters (1,1,1) strike a balance between

modelcomplexity and accuracy. ARIMA's merits include its ability to handle datasets withtrends and autocorrelations. If seasonal impacts are detected, further investigation may include evaluating seasonal ARIMA (SARIMA).

The ARIMA model anticipated the range of RON97 prices for the following 24 months. The predicted price range is useful for stakeholders such as consumers, corporations, and regulators. The estimates can help with strategic decisions including pricing strategies, budgeting, and market strategy. However, the forecast's credibility is dependent on the stability of external factors including economic conditions, geopolitical events, and market dynamics.

5. CONCLUSION

The fuel price of RON97 is still a major subject that interest businesses and impacts the economy. Its fluctuations are subject to various determinants, which tend to either positively or negatively alter prices, the profitability of corporations, the people that use it, and the shareholders and state of the economy of several countries that may be affected. Forecast using ARIMA Model as ARIMA(1,1,1) had been chosen.

Acknowledgments: Thousands of appreciations to all the students and staffs of Faculty of Computer and Mathematical Sciences, University Teknologi MARA Negeri Sembilan, Seremban Campus especially to the Department of Statistics for their unwavering support, guidance, and dedication throughout the research.

References

Alam, M. S., Murshed, M., Manigandan, P., Pachiyappan, D., & Abduvaxitovna, S. Z. (2023). Forecasting oil, coal, and natural gas prices in the pre-and post-covid scenarios: Contextual evidence from india using time series forecasting tools. Resources Policy, 81, 103342.

Colladon, A. F., Verdoliva, G., Segneri, L., & Vitali, A. G. (2024). Analyzing gasoline prices in five italian cities: Insights from social network analysis. Cities, 150, 105075.

Ers¨oz, N. S¸ ., G¨uner, P., Akbas¸, A., & Bakir-gungor, B. (2022). Comparative performance analysis of arima, prophet and holt-winters forecasting methods on european covid- 19 data. International Journal of 3D Printing Technologies and Digital Industry, 6(3), 556-565.

Lazim, M. A. (2021). Introductory business forecasting : A practical approach. Pusat Penerbitan Universiti, Universiti Teknologi Mara.

Mabrouki, A., El Hafdaoui, H., &Khallaayoun, A. (2023). A forecast of pumped fuel prices in morocco using arimamodel. 2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), 1-5.

Malaysia' s Official Open Data Portal. (2024). Price of Petroleum & Diesel. [Accessed: 2025-01-14}. Price of Petroleum & Diesel | data.gov.my

Ringgit Plus. (2024). Petrol Price Malaysia Live Updates (RON95, RON97 & Diesel) [Accessed: 2025-01-14} Petrol Price Malaysia Live Updates (RON95, RON97 & Diesel)

Sarpong-Streetor, R. M., Sokkalingam, R., Othman, M., Daud, H., & Owusu, D. A. (2021). Arimax modelling of ron97 price with crude oil price as an exogenous variable in malaysian. Proceedings of the 6th International Conference on Fundamental and Applied Sciences: ICFAS 2020, 679-691.

Sokkalingam, R., Sarpong-Streetor, R. M., Othman, M., Daud, H., & Owusu, D. A. (2021). Forecasting petroleum fuel price in malaysia by arima model. Proceedings of the 6th International Conference on Fundamental and Applied Sciences: ICFAS 2020, 671-678.