

Forecasting the System Marginal Price in the Malaysian Electricity Market Using Time-Series Models

Intan Azmira Wan Abdul Razak^{1*}, Mohamad Fani Sulaima¹, Farah Anishah Zaini¹,
Siti Noor Hidayah Abdul Rahman², Wan Syakirah Wan Abdullah³

¹Department of Electrical Engineering, Faculty of Electrical Engineering and Technology, Universiti Teknikal Malaysia Melaka (UTeM), Durian Tunggal, Melaka, Malaysia

²Vencast Sdn. Bhd., Jalan Spring 34/23, Off Jalan Bukit Kemuning, Seksyen, Shah Alam, Malaysia

³Department of Business Assessment and Engineering 2, TNB Renewables Sdn. Bhd., Jalan Bukit Pantai, Kuala Lumpur, Malaysia
Email: *intan.azmira@utem.edu.my

How to cite this paper: Wan Abdul Razak, I.A., Sulaima, M.F., Zaini, F.A., Abdul Rahman, S.N.H. and Abdullah, W.S.W. (2025) Forecasting the System Marginal Price in the Malaysian Electricity Market Using Time-Series Models. *Journal of Power and Energy Engineering*, 13, 61-80.
<https://doi.org/10.4236/jpee.2025.138005>

Received: July 30, 2025

Accepted: August 18, 2025

Published: August 21, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).

<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

Accurate forecasting of the system marginal price (SMP) is crucial to improve demand-side management and optimize power generation scheduling. However, predicting the SMP is challenging due to the high volatility of electricity prices, which are influenced by fuel prices, fluctuations in energy demand, and generation capacity. Hence, the aim of this study is to improve the accuracy of SMP forecasting for the Malaysian electricity market using time-series models. Correlation analysis was first conducted to identify the most significant variables affecting SMP, followed by the development of simple linear regression (SLR), multiple linear regression (MLR), and autoregressive integrated moving average (ARIMA) models to predict the SMP, and assessment of the models using actual market data from a Single Buyer (SB) website. The accuracy of the SLR, MLR, and ARIMA models was assessed using mean absolute error (MAE) and mean absolute percentage error (MAPE). The prediction accuracy of the models was significantly improved by tailoring the model parameters and selecting the relevant independent variables. Based on the results, the SLR, MLR, and ARIMA models outperformed other models published in the literature including the model used by the SB. The regression and ARIMA models appear promising for forecasting electricity prices accurately, offering valuable insights to power utilities and stakeholders in supporting informed decision-making in a dynamic electricity market.

Keywords

ARIMA, Correlation Analysis, Linear Regression, System Marginal Price

1. Introduction

The structure and operation of the electricity market in Malaysia have undergone significant changes over the years [1]. In the past, the electricity market was characterized by a single-buyer paradigm, where Tenaga Nasional Berhad (TNB) was the sole purchaser of electricity. However, changes in the electricity market, which included more competitive features, have increased the volatility of electricity prices. The interplay between fuel prices, fluctuations in energy demand, and generation capacity has increased the complexity and volatility of electricity pricing. In Malaysia, the system marginal price (SMP) is the price at which electricity is traded in the electricity market and is determined by the most expensive marginal generator required to supply electricity in order to fulfil the energy demand in a 30-min period. The SMP directly affects the electricity prices for generators and consumers, and therefore, it is essential to predict the SMP accurately.

The SMP is influenced by both internal factors (local energy demand trends, generation capacity, and grid reliability) and external factors (global fuel prices and economic conditions), all of which promote the volatility of the SMP. It is necessary to accurately predict the SMP in such a dynamic environment to facilitate the key stakeholders of the electricity market (generators, retailers, and regulators) in making wise decisions regarding pricing strategies, production scheduling, and policy development. At present, the existing forecasting models are incapable of capturing the internal and external factors that influence the SMP in the Malaysian electricity market, resulting in less accurate forecasts, which in turn, affect decision-making. Maniatis and Milonas examined the relationship between the volatility of wholesale power prices in Greece and the production of solar and wind energies [2]. They found that the increase in the use of renewable energy systems resulted in the volatility of electricity prices, which made SMP forecasting inherently complex. Hence, there is a need to develop advanced forecasting models that can account for the volatility in electricity prices associated with the use of renewable energy systems.

Time-series models have been used for SMP forecasting for many years as these models are capable of predicting time-dependent variables. The autoregressive integrated moving average (ARIMA) model is among the time-series models used in the power sector where significant amounts of historical data and the temporal characteristics of the data (*i.e.*, changes in the data with respect to time) are critical. The ARIMA model is suitable for cases where the process that creates the data itself exhibits a predictable pattern as this model is capable of capturing linear relationships in time-series data.

Regression models are alternative time-series models used to identify and quantify the relationships between the SMP and factors that influence the SMP owing

to their interpretability and adaptability. Regression models can account for many variables (e.g., conditions of the electricity market, economic indicators, and meteorological data), and therefore, these models can adapt to various forecasting scenarios. Ulgen and Poyrazoglu predicted power costs using the multiple linear regression model and they highlighted that it is important to select the relevant variables in order to enhance the prediction accuracy [3]. Huang and Liu performed regression analysis using MATLAB software and they highlighted the benefits of using this software for handling large amounts of data and performing complex computations required for accurate predictions [4].

Ahmad *et al.* predicted the load profiles of solar rooftop and integrated car park at a public university in Malaysia using the multiple regression time-series model [5]. They developed the multiple regression time-series model using MATLAB software, where the time-series data were differenced to eliminate trends and seasonal patterns from the dataset and the maximum likelihood estimation was used to estimate the parameters of the regression model with ARIMA time-series errors. Their results demonstrated the capability of the ARIMA model in producing short-term forecasts, which were crucial to optimise power consumption. This study is among the studies that demonstrate the versatility of the ARIMA model for energy management and energy forecasting beyond SMP.

In recent years, researchers have combined conventional time-series models with machine learning models. For example, Kim *et al.* developed a new online machine learning model to forecast the SMP in South Korea based on economic variables [6]. Unlike static models, their online machine learning model continuously updated its parameters in real time, improving the capability of the model to adapt to fluctuating conditions in the electricity market. To determine long-term dependencies in the time-series data, Li and Becker developed a hybrid model that combined the ARIMA model with long short-term memory (LSTM) neural network [7]. Their results showed that the hybrid model outperformed other models in predicting electricity prices, particularly in electricity markets with high unpredictability. Their hybrid model increased the accuracy of electricity price predictions by exploiting the benefits of the ARIMA model in capturing linear correlations as well as the benefits of LSTM neural network in handling non-linear trends. Kızıldağ *et al.* proposed a detailed forecasting framework for the electricity market in Türkiye, in which innovative machine learning tools such as support vector machine (SVM), extreme gradient boosting (XGBoost), and extreme learning machine (ELM) were combined with statistical models such as ARIMA and seasonal autoregressive integrated moving average (SARIMA) models [8]. The performance of the models was significantly improved by applying advanced feature selection techniques such as maximum likelihood feature selector (MLFS) and minimum redundancy maximum relevance (mRMR). Nowadays, regression models are often used in conjunction with machine learning models in order to improve prediction accuracy. For example, Polat *et al.* developed a bidding strategy for renewable electricity generation companies in the electricity

market in Türkiye by incorporating regression into their newsvendor-based neural network model [9].

Neural networks have garnered considerable attention from the scientific community for SMP forecasting owing to their capability to express complex, non-linear correlations in datasets. For example, Kim *et al.* developed a hybrid model combining one-dimensional neural network (1D-CNN) with (BiLSTM) neural network to forecast the peak electricity demand and SMP [10]. Their method exploits the benefits of CNN (which excels at feature selection) and the benefits of the BiLSTM neural network (which excels at extracting temporal relationships). Their hybrid model exhibited superior performance over conventional time-series models, indicating the possibility of combining sophisticated neural networks with conventional time-series models.

In a similar vein, Gaber *et al.* developed a hybrid deep learning model to predict short-term energy demands, in which the convolutional neural network–long short-term memory neural network (CNN-LSTM) model was combined with the bidirectional gated recurrent unit neural network (Bi-GRU) model [11]. In their hybrid model, CNN was used to elicit spatial trends, Bi-GRU was used to extract temporal dependencies, and LSTM was used to boost memory retention. The accuracy and robustness of the hybrid model for load forecasting was evaluated based on the root mean square error (RMSE) and mean absolute error (MAE), where the model was tested on the American Electric Power Company (AEP) dataset. The performance of the hybrid model was found to be superior to other conventional models. Based on their study, it can be deduced that deep learning models can enhance the accuracy of time-series forecasting, which may be beneficial to predict fluctuations in the SMP in dynamic electricity markets such as Malaysia. Hernandez-Matheus *et al.* conducted a detailed review of machine learning models for regional energy communities, and they highlighted that there is burgeoning interest in the development of hybrid models that combine statistical and machine learning models to enhance prediction accuracy [12].

Carnevale *et al.* developed an artificial intelligence-based forecasting model, and the model was proven to promote the accuracy of the energy imbalance predictions for the day-ahead electricity market [13]. Their results showed that ensemble forecasting (*i.e.*, the combination of predictions from multiple models) appreciably enhanced the prediction accuracy by taking into account the volatility and uncertainty in electricity prices. Based on their results, it can be deduced that ensemble forecasting is beneficial for electricity markets characterized by high volatility, where a single model is incapable of capturing all of the pertinent information. Ensemble forecasting produces predictions with higher accuracy and reliability by leveraging the advantages of multiple models.

Despite the advances in forecasting models, it is still challenging to predict the SMP in the Malaysian electricity market with high accuracy. Neural networks for SMP forecasting are intrinsically complex, and they require advanced knowledge to tailor the multitude of hyperparameters, which is particularly the case for deep

learning models [13]. Owing to the complexity of neural networks, overfitting may occur, where the model produces a very small error on the training dataset, but produces a large error on a new dataset [4]. It is crucial to prevent overfitting in electricity price forecasting because the model will perform poorly on new, unseen data (such as unanticipated or unobserved variables in the electricity market), resulting in inaccurate predictions. In contrast, time-series models are simpler, clearer, and easier to comprehend. For example, the ARIMA model (which is a time-series model) is based on a specific set of parameters that represent the inherent temporal characteristics of the data, and thus, it is less susceptible to overfitting. Owing to the lucid nature of time-series models, these models are easier to interpret and validate, which are crucial for the electricity market in Malaysia, where transparency and adherence to regulations are of utmost importance.

Additionally, the volume of historical SMP data available was a practical constraint in model selection. Deep learning models typically require large-scale datasets to generalize effectively and avoid underfitting. Given the moderate size of our dataset, traditional time-series models such as ARIMA were more appropriate, offering reliable performance without the need for extensive data preprocessing or augmentation.

Based on the literature review, there is a critical need to develop simple, straightforward models that predict the SMP in the Malaysian electricity market with high accuracy and account for the distinctive characteristics of the electricity market. Hence, in this study, simple linear regression, multiple linear regression, and ARIMA models are used to enhance the accuracy of SMP predictions. The scope of this study is outlined as follows:

- 1) First, correlation analysis is performed to examine the effects of the independent variables (input variables) on the SMP (dependent or target variable), where the data are retrieved from the Single Buyer (SB) website. Simulations with six distinct inputs are carried out to identify the independent variables that have the most pronounced effect on the short-term SMP predictions. The independent variables constitute the energy demand forecast for the upcoming day, day-ahead forecast for the monthly generation mix, and past SMP values for the preceding 1, 2, 3, and 7 days.

- 2) Next, the simple linear regression, multiple linear regression, and ARIMA models are developed to predict the SMP in the Malaysian electricity market with high accuracy.

- 3) Finally, the performance of the simple linear regression, multiple linear regression, and ARIMA models is evaluated and compared with the performance of the model used by the SB and other models published in the literature. The findings indicate that the regression and ARIMA models outperform the other models in predicting the SMP.

This remainder of this paper is organised as follows. The methodology adopted in this study to develop the SMP forecasting models is elaborated in detail in Section 2, along with a theoretical background of the regression and ARIMA models.

The results of the SMP forecasting models are presented in Section 3, and the performance of the regression and ARIMA models is discussed and compared with the performance of other models published in the literature. The conclusions drawn based on the key findings of this study are presented in Section 4, along with recommendations for future work.

2. Methodology

The time-series models used to predict the SMP in the Malaysian electricity market are elaborated in detail in this section. This section begins with a theoretical background of the regression and ARIMA models, followed by correlation analysis, development of the regression and ARIMA models, and finally, evaluation of the accuracy and reliability of the SMP forecasting models based on performance metrics. **Figure 1** shows the overall methodology adopted in this study.

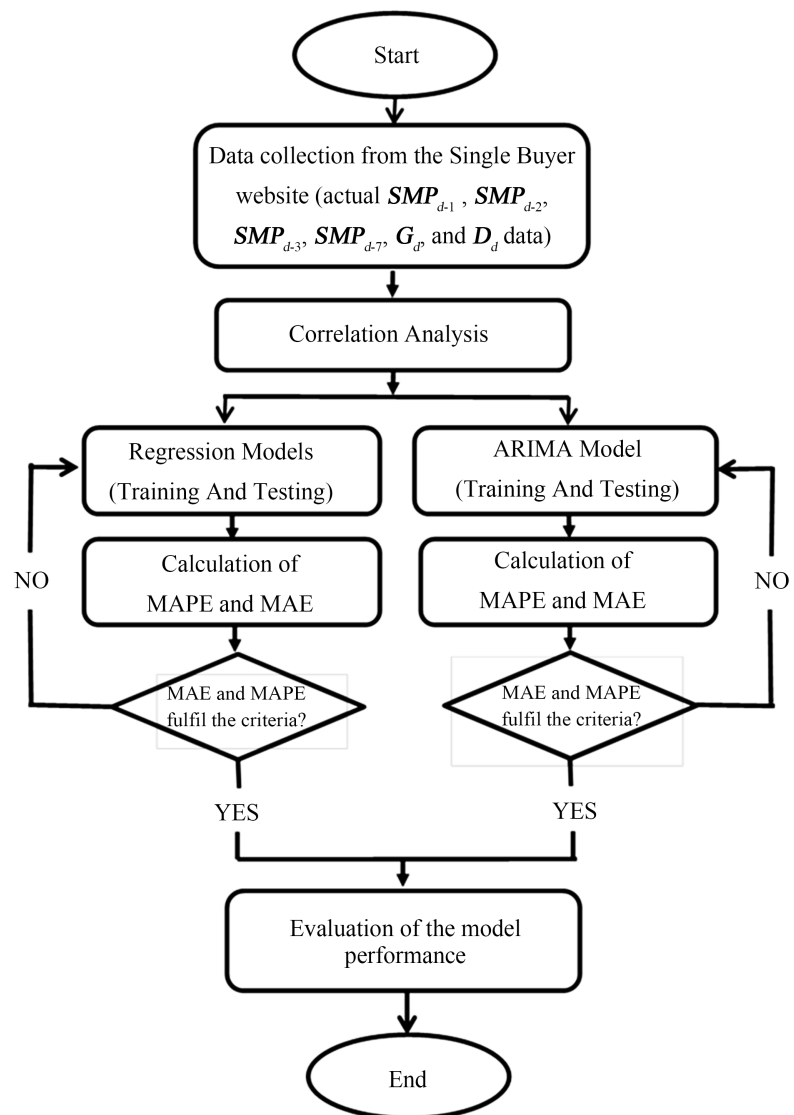


Figure 1. Flow chart of the methodology adopted in this study.

2.1. Theoretical Background

Time-series forecasting involves the use of mathematical models to predict future values based on historical data. Time-series forecasting is imperative for electricity markets, in which operational and financial decisions are influenced by electricity price fluctuations. Regression and ARIMA models are commonly used to predict the SMP, where the former model is used to establish the relationships between the independent and dependent variables, and the latter model is used to capture the temporal characteristics of the data. A brief description of these models and their role in enhancing the prediction accuracy of SMP is given in the following sub-sections.

2.1.1. Simple Linear Regression and Multiple Linear Regression Models

Regression model is a statistical model that establishes the relationship between one or multiple independent variables and a dependent variable. Regression model facilitates SMP forecasting by providing insight into how the independent variables (e.g., fuel prices, fluctuations in energy demand, and generation capacity) influence the SMP. The most basic type of regression model is the simple linear regression model, where the relationship between the dependent variable and a single independent variable is represented by a linear equation. In contrast, the multiple linear regression model can be used to gain insight into the effects of multiple independent variables on the dependent variable [4].

In general, the simple linear regression and multiple linear regression models for SMP forecasting can be expressed as Equation (1) and Equation (2), respectively:

$$SMP_t = \beta_0 + \beta_1 X_{1t} + \varepsilon_t \quad (1)$$

$$SMP_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_n X_{nt} + \varepsilon_t \quad (2)$$

where SMP_t is the SMP at time t , $X_{1t}, X_{2t}, \dots, X_{nt}$ are the independent variables, β_0 is a constant (y -intercept of the regression line), $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients of the independent variables, and ε_t is the error term.

The following statistical measures are used to ensure that the multiple linear regression model is appropriate, dependable, and interpretable [14] [15]:

1) The coefficient of determination (R^2 value) needs to be within a range of 0 - 1. The R^2 value reflects the proportion of the variability of the dependent variable that is explained by the independent variables of the model. An R^2 value of 1 indicates that all of the variability in the dependent variable is accounted for by the independent variables, and therefore, it indicates a perfect fit between the model and data. In contrast, an R^2 value of 0 indicates that none of the variability in the dependent variable is accounted for by the independent variables, and therefore, the model does not approximate the data at all. The R^2 value is a statistical measure of the goodness of fit, and thus, it is crucial to achieve a high R^2 value.

2) The t -value (also known as t -statistic) must be greater than 2 as this indicates that the coefficient of the independent variable deviates appreciably from 0, which in turn, highlights that the independent variable has a significant influence on the

dependent variable. A high t -value indicates that the independent variable is likely to be a useful variable in the multiple linear regression model.

3) A p -value of less than 0.05 signifies that the independent variable is statistically significant and significantly affects the dependent variable. The p -value is used to test whether the null hypothesis is true (*i.e.*, there is no relationship between the independent variable and dependent variable, where the coefficient of the independent variable is 0). A p -value of less than 0.05 indicates that the relationship between the independent variable and dependent variable is statistically significant.

4) The F -value (also known as F -statistic) is determined by comparing the computed F -statistic to a critical value from the F -distribution table at a chosen significance level (α), typically 0.05. This test evaluates the null hypothesis that all regression coefficients are simultaneously equal to zero, where none of the independent variables have a statistically significant effect on the dependent variable. If the F -value exceeds the critical value or the corresponding p -value is below the significance level, the null hypothesis is rejected, indicating that the multiple linear regression model is statistically significant, meaning at least one independent variable has a meaningful relationship with the dependent variable. Conversely, if the F -value is lower than the critical value or the p -value is higher than the significance level, the null hypothesis cannot be rejected, suggesting that the model does not offer a significantly better fit than a model with no predictors.

2.1.2. Arima Model

The ARIMA model works well for short-term forecasts based on historical data with temporal dependencies [14]. The ARIMA model is composed of three components: autoregressive (AR), integrated (I), and moving average (MA). The autoregressive component models the relationship between a current observation and its own lagged observations (past observations), whereas the moving average component models the relationship between a current observation and the residual errors obtained from applying the moving average model to the lagged observations. The integrated component addresses the non-stationarity of the time-series data by means of differencing.

The ARIMA model is represented by the notation $ARIMA(p, d, q)$, where p denotes the order of the autoregressive model (which indicates the number of lagged observations that are used by the model to predict the current observation), d denotes the degree of differencing (which is the number of the times the data are differenced to make the time series stationary), and q denotes the order of the moving average model (which is the number of lagged forecast errors used in the model). In general, the prediction equation for the ARIMA model can be expressed as:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

where Y_t is the actual value at time t , c is a constant term (which is optional), $\phi_1, \phi_2, \dots, \phi_p$ are the autoregressive coefficients (to account for the effect of past

observations on the current observation in the time-series data), $\theta_1, \theta_2, \dots, \theta_q$ are the moving average coefficients (to account for the effect of past residual errors on the current observation in the time-series data), and ε_t is the white noise or error term at time t .

In this study, the ARIMA model was implemented through the following steps. Firstly, the relevant time-series data were collected, and the data were scrubbed to address missing values and outliers. Secondly, the time-series data were checked for stationarity by identifying trends and seasonal patterns. The time-series data are said to be stationary if the statistical parameters such as mean, and variance do not vary with respect to time. Differencing was carried out to make the time-series data stationary [16]. Thirdly, the autocorrelation function (ACF) and partial autocorrelation function (PACF) were plotted to identify trends and relationships within the time-series data and determine the optimal values for the order of the autoregressive model (p) and order of the moving average model (q). The degree of differencing (d) was established based on the stationarity check. Fourthly, the ARIMA model was trained on historical SMP data to capture temporal dependencies. Finally, the accuracy and reliability of the ARIMA model for SMP forecasting were evaluated based on the following performance metrics: mean absolute error (MAE) and mean absolute percentage error (MAPE).

2.2. Correlation Analysis

Correlation analysis was conducted to determine the direction and strength of the correlations between the dependent variable (SMP) and different independent variables. The SB website [17] was used to acquire the historical SMP data in order to evaluate the accuracy of the simple linear regression, multiple linear regression, and ARIMA models in this study. The historical SMP data were examined meticulously and sedulously to ensure that the data were accurate and reliable, and that the duration was sufficiently long to capture the changes and trends in the Malaysian electricity market. The effects of independent variables on the SMP can be assessed based on the correlation coefficients.

The correlation coefficient falls within a range of -1 and 1 , which indicates the strength of the linear correlation between the dependent and independent variables. A correlation coefficient of 1 indicates a perfect positive correlation, where a change in the independent variable results in a change in the dependent variable in the same direction. For example, if the independent variable increases, the dependent variable increases by the same proportion. In contrast, a coefficient correlation of -1 indicates a perfect negative correlation, where a change in the independent variable results in a change in the dependent variable in the opposite direction. For example, if the independent variable increases, the dependent variable decreases by the same proportion. In this study, the correlations between the SMP on the targeted day and the following independent variables were examined:

- 1) Past SMP values for the preceding 1, 2, 3, and 7 days (SMP_{d-1} , SMP_{d-2} ,

SMP_{d-3} , and SMP_{d-7}).

2) Energy demand forecast for the upcoming day (D_d).

3) Day-ahead forecast for the monthly generation mix, which is a combination of electricity generation from fossil fuels (coal and natural gas) and renewable energy sources (solar and hydro energies) (G_d).

2.3. Development of the Simple Linear Regression, Multiple Linear Regression, and Arima Models

In this study, the pre-processing stage involved selecting the datasets for training and testing. The training dataset was used to facilitate the regression and ARIMA models in learning the relationships and trends between the independent variables and dependent variable (SMP). All the observations were half-hourly data, where the daily forecast generated 48 points of half-hourly SMP for the day ahead. The training dataset spanned from 1 February 2021 to 30 October 2021, whereas the testing dataset spanned from 1 February 2022 to 24 April 2022.

Both the simple linear regression and multiple linear regression models were developed to determine the effects of different independent variables on the SMP. The independent variables that had the most pronounced effects on the SMP were identified using these models.

The ARIMA model was developed by preparing the time-series data (SMP data), ensuring that the data were stationary. The parameters of the ARIMA model (p and q) were then determined by analysing the ACF and PACF plots, whereas the degree of differencing (d) was determined based on the number of times the SMP data needed to be differenced in order to attain stationarity. The stationary SMP data were then fitted with the ARIMA model to identify the trends and temporal dependencies in SMP. The ARIMA model was trained using the training dataset after the parameters p , d , and q were determined.

The accuracy and reliability of the regression and ARIMA models were evaluated based on the MAE and MAPE, which will be described succinctly in the following sub-section.

2.4. Evaluation of the Simple Linear Regression, Multiple Linear Regression, and Arima Models

The prediction accuracy of the simple linear regression, multiple linear regression and ARIMA models was evaluated based on two performance metrics: MAE and MAPE. The MAE represents the mean of the absolute errors (*i.e.*, the sum of the absolute errors (differences between the actual and predicted SMP values) divided by the total number of hours), as given by Equation (4). As the name implies, the MAPE represents the mean of the absolute percentage errors (*i.e.*, the sum of the relative errors (ratios of the absolute error to the actual SMP value) divided by the total number of hours and multiplied by 100), as given by Equation (5). In these equations, $SMP_{actual,t}$ denotes the actual SMP value at hour t , $SMP_{forecast,t}$ denotes the predicted SMP value at hour t , and N denotes for the total number

of hours.

$$\text{MAE} = \frac{1}{N} \times \sum_{t=1}^N |SMP_{actual_t} - SMP_{forecast_t}| \quad (4)$$

$$\text{MAPE} = \frac{100}{N} \times \sum_{t=1}^N \frac{|SMP_{actual_t} - SMP_{forecast_t}|}{SMP_{actual_t}} \quad (5)$$

3. Results and Discussion

The results of the correlation analysis and performance of the simple linear regression, multiple linear regression, and ARIMA models were analysed and presented and discussed in this section. The results reveal the strengths and limitations of the regression and ARIMA models, which offer insight into the applicability of these models in predicting the SMP for the Malaysian electricity market. The most appropriate SMP forecasting model for the Malaysian electricity market was then selected based on the model that yielded the lowest MAE and MAPE.

3.1. Correlation Analysis

Table 1 shows the strength of the correlations between the independent variables (past SMP values for the preceding 1, 2, 3, and 7 days (SMP_{d-1} , SMP_{d-2} , SMP_{d-3} , and SMP_{d-7}), day-ahead forecast for the monthly generation mix (G_d), and energy demand forecast for the upcoming day (D_d) and the dependent variable (current day's SMP) for January 2021. A correlation coefficient of 0.5 - 1.0 indicates that there is a strong correlation between the independent and dependent variables, whereas a correlation coefficient of 0.30 - 0.49 indicates that there is a moderate correlation between the independent and dependent variables [18].

Based on the results, there was a stronger correlation between SMP_{d-1} and SMP_d (as indicated by the correlation coefficient of 0.22), compared with the correlations between SMP_{d-2} , SMP_{d-3} , SMP_{d-7} , and SMP_d . This is indeed expected since electricity prices frequently follow short-term trends and are influenced by market factors, and thus, the electricity prices from the previous day can have a more significant effect on the current day's SMP. The correlation between SMP_{d-2} and SMP_d was weaker due to the increase in the lag time. The past SMP value for the preceding 2 days (SMP_{d-2}) can still reflect ongoing market trends. However, external market factors such as sudden changes in energy demand or supply, fuel price changes, the integration of renewable energy, and regulatory shifts can diminish the correlation between the past and current day's SMP values over time. This is also reflected by the correlation coefficients for SMP_{d-3} and SMP_{d-7} , which indicate that there is a weak correlation between these independent variables and SMP_d . The conditions of the Malaysian electricity market vary rapidly, and even though there is still a correlation between the past and current day's SMP values, the strength of the correlation decreases with an increase in the lag time (*i.e.*, the number of days in which the past SMP influences the current day's SMP).

Table 1. Results of the correlation analysis.

Target variable (dependent variable)	Input variable (independent variable)	Correlation coefficient
SMP_d	SMP_{d-1}	0.22
SMP_d	SMP_{d-2}	0.10
SMP_d	SMP_{d-3}	0.14
SMP_d	SMP_{d-7}	0.14
SMP_d	G_d	0.42
SMP_d	D_d	0.39

Meanwhile, the day-ahead forecast for the monthly generation mix (G_d) appeared to have a moderate correlation with the current day's SMP, which is expected because the cost of electricity generation varies depending on the energy source. Likewise, the energy demand forecast for the upcoming day (D_d) had a moderate correlation with the current day's SMP. In general, electricity prices increase with an increase in the energy demand due to the higher need for electricity generation, which may require bringing more expensive generation sources online. Hence, it is likely that changes in D_d will result in changes in SMP_d .

3.2. Performance of the Simple Linear Regression, Multiple Linear Regression, and Arima Models

As mentioned previously in Section 2.4, the accuracy and reliability of the simple linear regression, multiple linear regression, and ARIMA models were assessed based on two performance metrics: MAE and MAPE. An MAE approaching 0 signifies a highly accurate model, whereas a MAPE below 10% is typically regarded as an indicator of a reliable forecasting model. The performance of the regression model was assessed in response to six different inputs (comprising combinations of independent variables as well as single independent variable) and the results are summarised in **Table 2**. In contrast, the performance of the ARIMA model was assessed in response to four different inputs since the ARIMA model could only handle a single independent variable. The testing dataset was acquired from 1 February 2022 to 24 April 2022, whereas the training dataset was acquired from 1 February 2021 to 30 October 2021.

Table 2. Performance of the simple linear regression, multiple linear regression, and ARIMA models in predicting the SMP using different independent variables as the model inputs (1 February 2022 to 24 April 2022).

Input variables (independent variables)	Regression model		ARIMA model	
	MAE	MAPE	MAE	MAPE
SMP_{d-1} , D_d , and G_d	0.0188	5.98%	-	-

Continued

SMP_{d-7} and D_d	0.0249	8.07%	-	-
SMP_{d-1}	0.0298	10.70%	0.0221	8.21%
SMP_{d-7}	0.0333	11.77%	0.0259	9.44%
D_d	0.0966	35.34%	1.2443	48.07%
G_d	0.0984	36.08%	1.2329	47.61%

The multiple linear regression model was developed by determining if the combination of independent variables resulted in a model that fulfilled the statistical criteria (R^2 value, t -value, p -value, and F -value) described in Section 2.1.1. The combination of SMP_{d-1} , D_d , and G_d as well as the combination of SMP_{d-7} and D_d resulted in a multiple linear regression model that fulfilled the statistical criteria. Based on the results, the lowest MAE and MAPE were achieved for the multiple linear regression model with three independent variables: past SMP value for the preceding day (SMP_{d-1}), energy demand forecast for the upcoming day (D_d), and day-ahead forecast for the monthly generation mix (G_d). This result indicates the capability of the multiple linear regression model to capture complex relationships by incorporating a diverse set of independent variables.

In contrast, the ARIMA model was more accurate than the simple linear regression model in predicting the SMP when the past SMP value for the preceding 1 and 7 days (SMP_{d-1} and SMP_{d-7}) were used as the inputs of the model, as indicated by the lower MAE and MAPE. This is indeed expected since the ARIMA model typically performs well in univariate forecasting (*i.e.*, the prediction of SMP based on a single independent variable), particularly when the SMP follows identifiable trends, cycles, or seasonal patterns. The ARIMA model is especially useful for short-term forecasts where the past SMP values are strong predictors of future SMP values.

In addition, the simple linear regression and ARIMA models yielded results that were significantly different from those of the correlation analysis (**Table 1**). Although the energy demand forecast for the upcoming day (D_d) and day-ahead forecast the monthly generation mix (G_d) had a moderate correlation with the current day's SMP, using these independent variables as the model inputs resulted in a higher MAPE. This indicates that demand and generation alone may not be adequate for accurately forecasting SMP, potentially due to external market dynamics, price volatility, or the impact of other unaccounted factors, such as fuel prices, regulatory policies, and fluctuations in renewable energy sources. The past SMP value for the preceding 7 days (SMP_{d-7}) appeared to be a valuable independent variable, despite its lower correlation coefficient. This suggests the presence of cyclical trends in the Malaysian electricity market, as the SMP tends to follow weekly trends, emphasising the importance of using historical data from 7 days prior to make accurate predictions.

Figure 2 shows the comparison between the actual SMP and SMP predicted by the multiple linear regression model, where the past SMP value for the preceding day (SMP_{d-1}), energy demand forecast for the upcoming day (D_d), and day-ahead forecast of the monthly generation mix (G_d) were used as the model inputs. **Figure 3** shows the comparison between the actual SMP and the SMP predicted by the simple linear regression model, where the past SMP value for the preceding 1 and 7 days (SMP_{d-1} and SMP_{d-7}) were used as the model inputs. It can be observed that both the multiple linear regression and simple linear regression models predicted the SMP in the Malaysian electricity market to an acceptable degree except for some days when there were spikes in the actual SMP values.

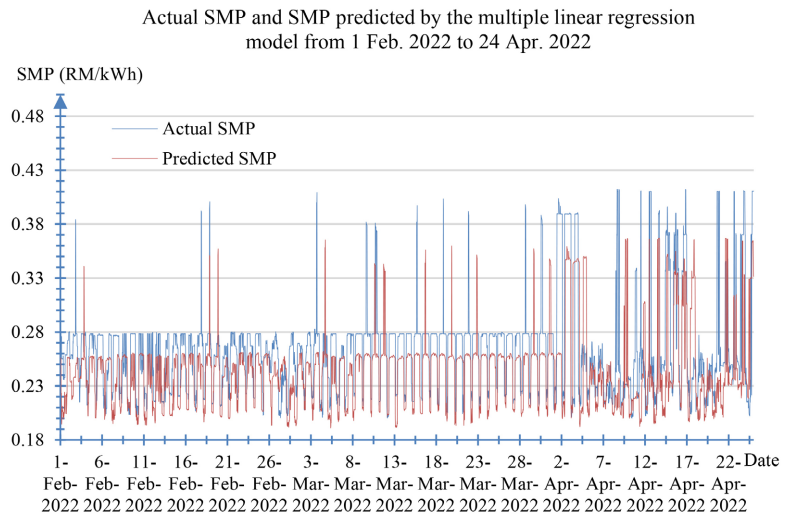


Figure 2. Comparison between the actual SMP and SMP predicted by the multiple linear regression model using SMP_{d-1} , D_d and G_d as the inputs.

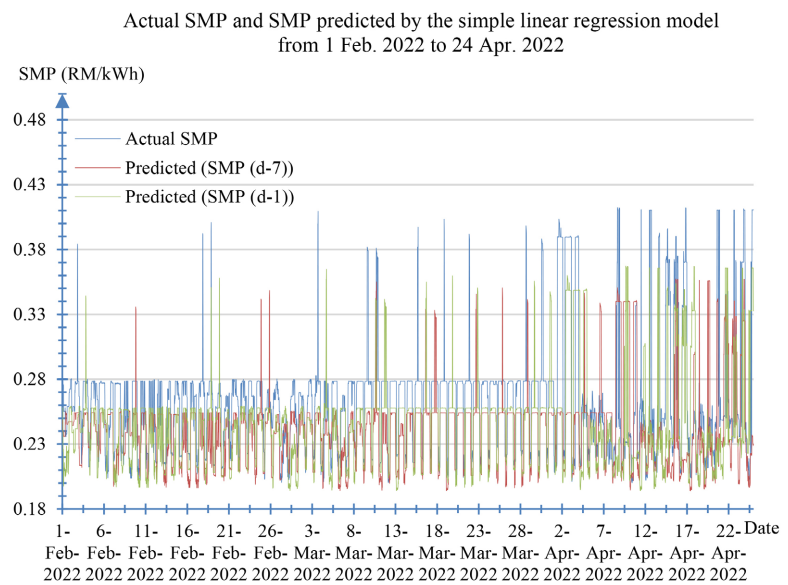


Figure 3. Comparison between the actual SMP and SMP predicted by the simple linear regression model using SMP_{d-1} and SMP_{d-7} as the inputs.

3.3. Comparison of the Performance of the Simple Linear Regression, Multiple Linear Regression, and Arima Models with Other Models

The performance of the simple linear regression, multiple linear regression, ARIMA, and other models published in the literature (least squares support vector machine (LSSVM) and least squares support vector machine-genetic algorithm (LSSVM-GA) in predicting the SMP using two testing datasets (1 February 2022-24 April 2022 and February 2022) were compared and the results are tabulated in **Table 3**. It can be observed that the ARIMA model performed better than the model used by the SB to predict the short-term variations in SMP, with a MAPE of less than 10%. The cyclical and temporal characteristics of the Malaysian electricity market were well captured by the ARIMA model, which was initially created to handle time-dependent data. The multiple linear regression and simple linear regression models also outperformed the model used by the SB, with a MAPE of ~6% - 7%. By inputting multiple independent variables (the past SMP values for the preceding day (SMP_{d-1}), energy demand forecast for the upcoming day (D_d), and day-ahead forecast for the monthly generation mix (G_d)) into the model, the multiple linear regression model demonstrated its capability in handling multiple linear relationships between the independent and dependent variables, encompassing various factors that influence the SMP.

Table 3. Comparison of the performance of the simple linear regression, multiple linear regression, ARIMA, and other models in predicting the SMP in the Malaysian electricity market using different independent variables and testing datasets.

Testing dataset	Forecasting model	Input variables (independent variables)	MAE	MAPE
1 February 2022-24 April 2022	ARIMA	SMP_{d-1}	0.0221	8.21%
	Multiple linear regression	SMP_{d-1} , D_d , and G_d	0.0188	5.98%
	LSSVM [19]	SMP_{d-7} , SMP_{d-14} , SMP_{d-21} , and SMP_{d-28}	0.0259	9.28%
	LSSVM-GA [19]	SMP_{d-7} , SMP_{d-14} , SMP_{d-21} , and SMP_{d-28}	0.0339	12.06%
	Single Buyer	Not known	0.0351	13.31%
February 2022	ARIMA	SMP_{d-1}	0.0156	6.27%
	Simple linear regression	SMP_{d-1}	0.0165	6.69%
	LSSVM-GA [20]	SMP_{d-1}	0.0180	7.09%
	Single Buyer	Not known	0.0254	10.63%

In addition, it is evident that the regression and ARIMA models in this study were superior to the machine learning models (LSSVM and LSSVM-GA) pub-

lished in the literature [19] [20] in predicting the SMP in the Malaysian electricity market, as indicated by the lower MAPE values. It can be deduced that time-series models (in particular, regression and ARIMA models) perform well in cases where the temporal characteristics of the data are essential to predict the SMP accurately. The performance of the regression and ARIMA models was comparable to the performance of the LSSVM and LSSVM-GA models; however, the effectiveness of the time-series models may differ depending on the independent variables used as the inputs.

The ARIMA model was superior to the simple linear regression and LSSVM-GA models in forecasting the SMP based on a single independent variable. This indicates that the ARIMA model performs exceptionally well in predicting the short-term SMP based on only the past SMP values for the preceding day (SMP_{d-1}). Even though the simple linear regression model was slightly less accurate than the ARIMA model, the simple linear regression model could also be used to predict the SMP in the Malaysian electricity market based on a single independent variable since there was only a minor difference between the MAPE values of the simple linear regression and ARIMA models (6.69% and 6.27%, respectively).

The performance of the regression and ARIMA models was also compared with the performance of the ARIMA model developed by Kızıldağ *et al.* [8]. However, it shall be highlighted that the comparison is complicated by discrepancies in the datasets, including those related to historical periods, market situations, and geographical locations. Notably, both studies utilized hourly SMP data, so differences in time granularity do not account for the variation in forecasting performance. Hence, there is a need to exercise caution when interpreting the results. Kızıldağ *et al.* found that the ARIMA model consistently showed higher MAPE values (with an average MAPE of 15.26%), indicating that the model was not capable of capturing the characteristics of the highly volatile electricity market in Türkiye [8]. In contrast, the ARIMA model performed relatively well in this study, with a MAPE of less than 10%. This discrepancy can be attributed to differences in the data characteristics and market dynamics between Malaysia and Türkiye, as well as the configurations used for the ARIMA models. Even though Kızıldağ *et al.* discovered that the ARIMA model was incapable of capturing the rapid electricity price fluctuations and intricate interdependencies between the variables, this study shows that the ARIMA model can predict the SMP in the Malaysian electricity market with good accuracy by appropriately tailoring the model parameters and incorporating the relevant independent variables. Kızıldağ *et al.* also implemented advanced feature selection techniques such as minimum redundancy maximum relevance (mRMR) and maximum likelihood feature selector (MLFS) to remove redundant or less informative variables in order to improve the prediction accuracy. Even though these feature selection techniques were not implemented in this study, this study shows that it is possible to develop an ARIMA model with high prediction accuracy by judiciously selecting the model parameters and incorporating the relevant independent variables as inputs.

Based on the MAE and MAPE values, it can be deduced that both regression and ARIMA models can be used to predict the short-term SMP in the Malaysian electricity market with high accuracy, without the need to deal with the complex intricacies of machine learning models. Both the regression and ARIMA models are able to capture the temporal characteristics of electricity prices. The multiple linear regression model, in particular, proves to be useful in univariate forecasting with multiple independent variables as inputs, which gives a more realistic picture since the SMP is influenced by internal and external factors.

4. Conclusions

The electricity market in Malaysia is undergoing a paradigm shift, transitioning from a single-buyer model to a more competitive model. Accurate SMP forecasting is crucial for demand-side management and efficient power generation scheduling; however, this is made difficult by the transient nature of the SMP. Hence, time-series models (specifically simple linear regression, multiple linear regression, and ARIMA models) were developed in this study to predict the SMP in the Malaysian electricity market with high accuracy. Correlation analysis was first conducted to identify the primary factors (independent variables) that influence the SMP (dependent variable). Next, the simple linear regression, multiple linear regression, and ARIMA models were developed. The regression models were selected owing to their simplicity in interpreting the relationships between the SMP and independent variables compared with machine learning models, while the ARIMA model was selected because of its capability in capturing temporal dependencies in the data. The accuracy and reliability of these models were evaluated based on the MAE and MAPE. The key findings of this study are summarized as follows:

1) Comparison between the ARIMA model and model used by the SB showed that the ARIMA model performed better in predicting the short-term variations in SMP, with a MAPE of less than 10%. Both the multiple linear regression and simple linear regression models outperformed the model used by the SB, with a MAPE of ~6% - 7%.

2) The simple linear regression, multiple linear regression, and ARIMA models in this study outperformed the machine learning models (LSSVM and LSSVM-GA) published in the literature [19] [20] in predicting the SMP in the Malaysian electricity market.

3) The performance of the simple linear regression model was comparable to that of the ARIMA model, where there was only a marginal difference between the MAPE values (6.69% and 6.27%, respectively).

4) The multiple linear regression model performed exceptionally well in predicting the short-term SMP based on a combination of the following independent variables: past SMP value for the preceding day (SMP_{d-1}), energy demand forecast for the upcoming day (D_d), and day-ahead forecast for the monthly generation mix (G_d). The corresponding MAPE was 5.98%, indicating high prediction accuracy.

5) The ARIMA model performed relatively well in this study (MAPE: <10%) unlike the ARIMA model developed by Kızıldağ *et al.* [8] (average MAPE: 15.26%), indicating that the ARIMA model can be used to predict the SMP with good accuracy by fine-tuning the model parameters and using the relevant independent parameters as the model inputs. The SMP in the Malaysian electricity market can be predicted by the ARIMA model without the need of sophisticated feature selection techniques to refine the independent variables.

6) Both the regression and ARIMA models can be used to predict the short-term SMP in the Malaysian electricity market with high accuracy, which is an advantage because these models are not as inherently complex as machine learning models.

With the regression and ARIMA models, the accuracy of short-term SMP forecasting can be improved, which will facilitate the key stakeholders of the Malaysian electricity market in making well-informed decisions. For instance, generators can optimize their bidding strategies based on anticipated price movements, while industrial consumers can design demand-response programs to reduce costs during peak pricing periods. Accurate SMP forecasting will facilitate balancing electricity supply and demand, which will ensure grid stability. Accurate SMP forecasting is imperative for efficient energy management, and it will facilitate generators to incorporate renewable energy sources into the power grid as well as boost resource allocation efficiency.

Despite these promising outcomes, the study has several limitations. Traditional models such as ARIMA and regression, while interpretable and straightforward, may not fully capture nonlinear relationships or complex temporal dependencies inherent in electricity markets. Future research could explore hybrid modeling approaches that combine statistical techniques with machine learning algorithms such as ARIMA-LSTM or ensemble methods to further improve forecasting accuracy and adaptability in dynamic market environments.

Acknowledgements

The authors would like to express their most heartfelt appreciation to the Centre for Research and Innovation Management (CRIM), Universiti Teknikal Malaysia Melaka for their technical and administrative support throughout the course of this study.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Sibeperegasam, M., Ramachandramurthy, V.K., Walker, S. and Kanesan, J. (2021) Malaysia's Electricity Market Structure in Transition. *Utilities Policy*, **72**, Article ID: 101266. <https://doi.org/10.1016/j.jup.2021.101266>
- [2] Maniatis, G.I. and Milonas, N.T. (2022) The Impact of Wind and Solar Power Generation on the Level and Volatility of Wholesale Electricity Prices in Greece. *Energy*

- Policy*, **170**, Article ID: 113243. <https://doi.org/10.1016/j.enpol.2022.113243>
- [3] Ulgen, T. and Poyrazoglu, G. (2020) Predictor Analysis for Electricity Price Forecasting by Multiple Linear Regression. 2020 *International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*, Sorrento, 24-26 June 2020, 618-622. <https://doi.org/10.1109/speedam48782.2020.9161866>
 - [4] Huang, Y. and Liu, W. (2022) Regression Analysis Model Based on Data Processing and MATLAB Numerical Simulation. 2022 *IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, Dalian, 14-16 April 2022, 1115-1118. <https://doi.org/10.1109/ipsec54454.2022.9777319>
 - [5] Ahmad, N., Othman, J., Abd Aziz, M.A.S., Sivaraju, S.S., Borhan, N., Wan Mohtar, W.A.A.I. and Abdullah, M.F. (2024) Load Profile Forecasting Using a Time Series Model for Solar Rooftop and Integrated Carpark of a Public University in Malaysia. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, **111**, 86-98. <https://doi.org/10.37934/arfmts.111.2.8698>
 - [6] Kim, T., Ha, B. and Hwangbo, S. (2023) Online Machine Learning Approach for System Marginal Price Forecasting Using Multiple Economic Indicators: A Novel Model for Real-Time Decision Making. *Machine Learning with Applications*, **14**, Article ID: 100505. <https://doi.org/10.1016/j.mlwa.2023.100505>
 - [7] Li, W. and Becker, D.M. (2021) Day-Ahead Electricity Price Prediction Applying Hybrid Models of LSTM-Based Deep Learning Methods and Feature Selection Algorithms under Consideration of Market Coupling. *Energy*, **237**, Article ID: 121543. <https://doi.org/10.1016/j.energy.2021.121543>
 - [8] Kızıldağ, M., Abut, F. and Akay, M.F. (2024) Development of New Electricity System Marginal Price Forecasting Models Using Statistical and Artificial Intelligence Methods. *Applied Sciences*, **14**, Article 10011. <https://doi.org/10.3390/app142110011>
 - [9] Polat, E., Güler, M.G. and Ulukuş, M.Y. (2024) Renewable Genco Bidding Strategy Using Newsvendor-Based Neural Networks: An Example from Turkish Electricity Market. *Electric Power Systems Research*, **231**, Article ID: 110301. <https://doi.org/10.1016/j.epsr.2024.110301>
 - [10] Kim, J., Oh, S., Kim, H. and Choi, W. (2023) Tutorial on Time Series Prediction Using 1D-CNN and BILSTM: A Case Example of Peak Electricity Demand and System Marginal Price Prediction. *Engineering Applications of Artificial Intelligence*, **126**, Article ID: 106817. <https://doi.org/10.1016/j.engappai.2023.106817>
 - [11] Gaber, A., H. Aly, H., Ghatwary, N. and K. Abdelsalam, A. (2024) Electrical Load Forecasting Using a Novel BI-GRU Encoder Decoder Model. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, **51**, 1-14. <https://doi.org/10.37934/araset.51.1.114>
 - [12] Hernandez-Matheus, A., Löschenbrand, M., Berg, K., Fuchs, I., Aragüés-Peñalba, M., Bullich-Massagué, E., et al. (2022) A Systematic Review of Machine Learning Techniques Related to Local Energy Communities. *Renewable and Sustainable Energy Reviews*, **170**, Article ID: 112651. <https://doi.org/10.1016/j.rser.2022.112651>
 - [13] Carnevale, D., Cavaiola, M. and Mazzino, A. (2024) A Novel AI-Assisted Forecasting Strategy Reveals the Energy Imbalance Sign for the Day-Ahead Electricity Market. *Energy Reports*, **11**, 4115-4126. <https://doi.org/10.1016/j.egyrs.2024.03.058>
 - [14] Ottaviani, F.M. and Marco, A.D. (2022) Multiple Linear Regression Model for Improved Project Cost Forecasting. *Procedia Computer Science*, **196**, 808-815. <https://doi.org/10.1016/j.procs.2021.12.079>
 - [15] Piekutowska, M., Niedbała, G., Piskier, T., Lenartowicz, T., Pilarski, K., Wojciechowski, T., et al. (2021) The Application of Multiple Linear Regression and Artificial Neural

- Network Models for Yield Prediction of Very Early Potato Cultivars before Harvest. *Agronomy*, **11**, Article 885. <https://doi.org/10.3390/agronomy11050885>
- [16] Sina, L.B., Secco, C.A., Blazevic, M. and Nazemi, K. (2023) Hybrid Forecasting Methods—A Systematic Review. *Electronics*, **12**, Article 2019. <https://doi.org/10.3390/electronics12092019>
- [17] Single Buyer. <https://www.singlebuyer.com.my>
- [18] Dursun, B., Aydin, F., Zontul, M. and Sener, S. (2014) Modeling and Estimating of Load Demand of Electricity Generated from Hydroelectric Power Plants in Türkiye Using Machine Learning Methods. *Advances in Electrical and Computer Engineering*, **14**, 121-132. <https://doi.org/10.4316/aece.2014.01019>
- [19] MAMR and Wan Abdul Razak, I.A., (2024) Forecasting System Marginal Price (SMP) for Malaysian Power Market. Final Year Project, Universiti Teknikal Malaysia Melaka.
- [20] Wan Abdul Razak, I.A., Wan Abdullah, W.S. and Sulaima, M.F. (2023) Enhanced Short-Term System Marginal Price (SMP) Forecast Modelling Using a Hybrid Model Combining Least Squares Support Vector Machines and the Genetic Algorithm in Peninsula Malaysia. *International Journal of Intelligent Systems and Applications in Engineering*, **11**, 289-298. <https://ijisae.org/index.php/IJISAE/article/view/3525>