ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 6 No. 1 (2026) 554-565
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Analysing Electricity Consumption in Malaysia via Linear Programming, Time Series Analysis and Clustering Analysis

Tey Sin Ying[1], Wan Nor Munirah Ariffin[1*]

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA*

*Corresponding Author: normunirah@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2026.06.01.061

Article Info

Received: 4 January 2026
Accepted: 6 June 2026
Available online: 9 July 2026

Keywords

Electricity Consumption, Linear Programming, Time Series Analysis, Clustering Analysis

Abstract

Malaysia faces increasing electricity consumption that later creates challenges in long-term sustainability. The lack of accurate forecasting and clear consumption pattern classification limits effective energy planning. This study aims to minimize electricity transmission losses using linear programming, forecast future electricity consumption using time series analysis and categorize consumption levels using clustering analysis. Monthly electricity consumption data from January 2018 to May 2025 were obtained from the Department of Statistics Malaysia (DOSM) and analysed using Python. In the first phase, linear programming was applied to optimize electricity distribution among industrial, domestic and export sectors while satisfying supply and demand constraints. The results showed that optimized allocation reduced annual electricity losses by 60 to 80 MKWh that proves the model's effectiveness. In the second phase, time series analysis such as Seasonal Autoregressive Integrated Moving Average (SARIMA) model was developed to forecast future monthly electricity consumption. Mean Absolute Percentage Error (MAPE) showed the high predictive accuracy below 2% which indicates a stable increasing consumption trend. In the third phase, clustering analysis such as K-means clustering and Principal Component Analysis (PCA) were used to categorize monthly consumption levels. Results showed upward shift toward higher consumption after 2023 because of industrial growth and rising domestic demand. Overall, linear programming, time series analysis and clustering analysis successfully provide a comprehensive data-driven framework for sustainable electricity management. The findings offer valuable insights for policymakers and the Energy Commission in formulating strategies to reduce losses, enhance forecasting accuracy and improve energy distribution efficiency across Malaysia's electricity sector.

1. Introduction

The current electricity consumption and the challenges faced in balancing supply and demand due to urbanization, industrialization and residential reliance. Therefore, it needs to raise the concerns about efficient electricity management to solve issues such as transmission losses and underutilized electricity export. Besides, the research background emphasizes the importance of optimization, forecasting and consumption pattern analysis in electricity field. The methods that apply linear programming, time series analysis and clustering techniques are reviewed to prove the usefulness and effectiveness to address the challenges. Moreover, the problem statements identify some issues that impacted the Malaysian electricity sector such as rising electricity consumption across all sectors, forecasting limitations and the necessity of consumption characteristic. After that, the purpose of the research is determined by the research objectives and scope. Lastly, the significance of the research regarding the economic, environmental and sectoral benefits are explained. It can contribute more sustainable and efficient electricity consumption framework in Malaysia.

Electricity consumption for both the domestic and commercial sectors in Malaysia which provides electricity usage varies between different sectors and offers a foundation for optimizing energy consumption in Malaysia [1]. Residential consumers represent the largest portion of electricity users accounting for approximately 80% of total consumption as reported by the Energy Commission Malaysia in 2019 [2]. To address these challenges, some optimization techniques were used to improve electricity consumption such as linear programming has been widely used to minimize transmission losses. The effectiveness of Mixed Integer Linear Programming (MILP) was explored in optimizing power systems [3]. It emphasizes the role of optimization methods to overcome electricity management challenges to ensure efficient resource utilization. It can combine renewable energy sources into existing grids. In additionally, accurate modelling and forecasting of electricity consumption supported the generation and distribution of electricity [4]. In general, power system engineers apply forecasting method to predict electricity consumption trends which can avoid overestimating or underestimating future electricity demand. The use of Autoregressive Integrated Moving Average (ARIMA) models for forecasting electricity consumption empowering the model effectiveness in time series analysis related to energy demand [5]. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots used to check data stationarity. Not only that, the accuracy of the time series model is evaluated using performance metrics such as Mean Absolute Percentage Error (MAPE). Furthermore, electricity consumption behaviour was analyzed by applying cluster analysis techniques to identify patterns in user power usage [6]. Clustering techniques can identify consumption patterns such as apply K-means, hierarchical clustering and fuzzy clustering algorithms. Improved k-means clustering algorithm was applied to analyze electricity consumption data from 1000 residential users [7]. Cluster analysis uncovers hidden patterns and segment consumers based on the consumption characteristics.

Malaysia's electricity sector faces several challenges. Different amounts of energy are lost through the electricity transmission and distribution process. In total, the average power loss around 8% to 15% from the power generation to the consumers [8]. Although initiatives exist to solve the problem but the current strategies not enough to minimize losses. Besides, development of an accurate forecasting model was used for electricity demand in Malaysia since electricity demand is highly volatile and exhibits strong seasonal characteristics [9]. However, traditional forecasting models only solved with shorter time intervals such as weekly data. Besides, the lack of effective forecasting model for Malaysia's quarterly electricity consumption limits strategic power system planning and energy policy. Moreover, traditional clustering techniques such as k-means are limited in the ability to capture temporal variability in load profiles due to inadequate similarity metrics like the Euclidean distance [10]. Despite clustering methods used to uncover consumption patterns, traditional clustering approaches still face limitations in accurately capturing the temporal dynamics of electricity load profiles.

The objectives of this research are to minimize an electricity loss by optimizing consumption and exports using linear programming. Besides, it aims to forecast future monthly the total electricity consumption in Malaysia using Seasonal Autoregressive Integrated Moving Average (SARIMA) time series model. Moreover, the purpose is to categorize months into low and high electricity consumption based on total, industrial and domestic consumption using cluster analysis techniques.

The dataset that has been used in this study focuses on analysing Malaysia's monthly electricity consumption from January 2018 to May 2025 that includes local consumption across industrial, commercial and domestic sectors as well as electricity exports and transmission losses. This study has applied three primary analytical methods which are linear programming, time series analysis and clustering analysis. Linear programming optimizes electricity distribution to minimize losses, meanwhile meeting sectoral demands. Besides, time series analysis such as SARIMA model forecast monthly electricity consumption for the next 12 months, identify trends, seasonal patterns and potential fluctuations in demand. Additionally, clustering analysis such as K-means clustering will apply to segment months into low and high level based on total, industrial and domestic electricity usage by using cluster analysis techniques.

The significance of the study is to address challenges in Malaysia's electricity sector which faces increasing demand due to urbanization, industrialization and residential reliance. The study focuses on minimization of the electricity losses contributes to help energy suppliers and energy regulators reduce technical losses and increase income from Cross-Border Electricity trade. Furthermore, time series forecasting predicts the future of electricity consumption can help energy planners and government such as the Energy Commission and Tenaga Nasional Berhad (TNB) to make decisions in infrastructure planning, investment and generation capacity. Lastly, the study will adopt cluster analysis which can support policymakers to design some efficient energy programs and tariff adjustments.

2. Literature Review

Linear programming techniques design an optimal electrical system that integrates high-voltage direct current (HVDC) transmission and energy storage [11]. Linear programming capable to handle multiple variables and linear constraints so it suitable for electricity supply and demand. Besides, linear programming combined with time series forecasting make better decision in hybrid system to enhance long-term planning. An approach to improve the optimization of electric power systems which combined the discrete-event simulation model with mixed-integer linear programming (MILP) model [12]. The modelling of power generation and transmission systems has attracted research interest especially to optimize operations and improve resilience.

Moreover, various forecasting approaches estimate electricity demand because accurate forecasting analysis and effective electricity management are important [13]. The use of ARIMA, ANN and hybrid models to forecast electricity consumption [14]. In times series forecasting, Box-Jenkins autoregressive used to predict peak load electricity demand. Various multivariate techniques and time series methods which is autoregressive (AR), moving average (MA), exponential smoothing, ARMA and ARIMA algorithms also used to forecast electricity consumption. An innovative approach known as the Uncertainty Quantified SARIMA Neural Network (UQ-SNN) combined with artificial intelligence and uncertainty quantification techniques. The results showed the UQ-SNN model compared to traditional SARIMA models more performed well [15].

In addition, the *k*-medoids clustering clustered a historical building energy dataset which enhanced by Principal Component Analysis (PCA) for dimensionality reduction and Dynamic Time Warping (DTW) for temporal distance measurement [16]. Besides, Electrical Load Profiles (ELPs) by using clustering algorithms presented two main clustering approaches which focused on clustering with a specified number of clusters and clustering without determining the number of clusters [17]. Clustering methods such as *k*-means are useful in categorizing users into distinct behavioural groups [18].

3. Research Methodology

The research methodology to achieve the objectives about electricity consumption in Malaysia. There are three core analytical approaches which is optimization through linear programming, forecasting using time series analysis and characteristic segmentation via clustering techniques.

3.1 Data Collection and Sources

The dataset was obtained from the OpenDOSM portal which is Malaysia's official open data platform managed by the [19]. It draws on operational data compiled from the country's major electricity providers including Tenaga Nasional Berhad (TNB) for Peninsular Malaysia, Sabah Electricity Sdn. Bhd. (SESB) for Sabah and Sarawak Energy Berhad (SEB) for Sarawak. The dataset presents electricity consumption values in millions of kilowatt-hours (MKWh) which recorded at a monthly frequency. The column in dataset contains total electricity consumption, local consumption divided into commercial and domestic sectors, electricity exports and recorded transmission or distribution losses.

3.2 Linear Programming Method

Linear programming problem is a type of optimization problem characterized by three main components which are objective function, decision variables and constraint [20]. The variables x and y are known as decision variables and the linear function that must be optimized with the specified condition to obtain the final solution is known as the objective function which is denoted as Z. Constraints are the limitations placed on choice variables that affect the values. The objective function is as shown in Equation (1).

$$Z = ax + by \tag{1}$$

Here, Z is the total electricity loss, x and y are electricity consumption from each sector, a and b are coefficient representing the impact of the decision variables. Constraints in the form of linear inequalities are Equations (2) and (3).

$$cx + dy \geq e \tag{2}$$

$$px + qy \leq r \tag{3}$$

Where c, d, p and q are coefficients in the constraint equation. e and r are constant of the constraint which is baseline requirements. Meanwhile, the non-negative restrictions are shown in Equations (4) and (5).

$$x \geq 0 \tag{4}$$

$$y \geq 0 \tag{5}$$

3.2.1 Objective Function Formulation

The decision variables can set x_1 as electricity industrial sectors, x_2 as electricity domestic sectors while x_3 as electricity export before develop the objective function. Losses can be distributed proportionally based on sectoral usage which is pro-rata concept. The loss per unit for each sector can be computed as loss coefficient equals to sector's share of loss divided sector consumption [21]. The objective function is to minimize electricity losses. Electricity losses are directly related to consumption and export through the coefficient. Usually, the flows within the network generate losses prove the causal relationship between the levels of electricity consumption and export affect electricity losses on the system. As the demand is met, the current flows through the wires of transmission lines can generate heat which known as losses.

3.2.2 Constraint Formulation

The model includes several constraints which the electricity allocated to the industrial and domestic sectors must meet minimum demand levels. For the yearly optimization situation, constraints are imposed whereby reductions in electricity allocation for each sector cannot exceed 5% of the original demand and also the total monthly electricity usage must not surpass 105% of the baseline demand to prevent over allocation.

Factor of electricity allocation for each sector cannot exceed 5% of the original demand and also the total monthly electricity usage must not surpass 105% of the baseline demand due to the threshold based on international policy benchmarks such as Egypt's National Energy Efficiency Action Plan (NEEAP) which targeted a 5% reduction in national energy consumption. 105% ceiling on total monthly electricity consumption is applied to represent a 5% allowable increase from the baseline consumption [22].

3.2.3 Optimization Model

The linear programming model consists of three key components which are decision variables, objective functions and constraints. The objective function for the linear programming model is minimizing electricity losses. Besides, the total electricity supply must not exceed the total available energy generated within the system. The constraint states that the optimization method does not suggest using more electricity than its available. To solve this optimization problem, an optimization solver in Python is used. The PuLP Python library was used to solve the model for every month separately.

3.3 Time Series Analysis

Time series analysis is a statistical technique that enables the identification of patterns, trends and seasonal variations in electricity usage. Autocorrelation function (ACF) and partial autocorrelation function (PACF) in the time series forecasting generally determine the level of differencing required to achieve stationarity and identify the potential AR and MA terms. Once a tentative model is identified, the parameters are estimated using a nonlinear optimization procedure to minimize an overall measure of error such as the sum of squared residuals. Furthermore, Ljung–Box test verifies that the white noise residuals and specify satisfaction of model assumptions.

3.3.1 Exploratory Data Analysis

Exploratory data analysis was conducted to visualize the trends and seasonal patterns which are present in the data. Line of the monthly consumption was plotted to observe long-term trends whether is repeating seasonal fluctuations or irregular variations. The visualization provided useful insights into the structure of the time series.

3.3.2 Stationarity Check

Augmented Dickey-Fuller (ADF) test was applied to the original series to assess stationarity. As the p-value of the test exceeded the 0.05 threshold, differencing was applied to the series until stationarity was achieved. The presence of non-stationarity may result in misleading inferences and incorrect regression outcomes. The regression equation used in the ADF test shown as Equation (6).

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-1} + \varepsilon_t \tag{6}$$

Where Δy_t is the first difference of the series, t is the time trend, α is a constant term, β captures a time trend, γ is the coefficient on the lagged level of the series, p is the number of lagged difference terms and ε_t is the error terms.

3.3.3 Model Identification

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the stationary series able to identify suitable parameters for the time series model. The plots helped in determining the potential orders of the autoregressive (AR) and moving average (MA) components of the model as well as the degree of differencing (d) required.

3.3.4 Model Training

Determined parameters were used to train the time series model. In some cases that having strong seasonal patterns, a Seasonal ARIMA (SARIMA) model was fitted to better analyse both the non-seasonal and seasonal dynamics of the time series [23]. The training set consisted of historical data up to a specified cutoff point to ensure the rest of the data become model evaluation.

3.3.5 Model Evaluation

Forecast accuracy using error metrics such as MAE, RMSE and MAPE is essential for guiding improvements in forecasting methods [24], [25]. MAE measures the average magnitude of the absolute differences between forecasted values, $\hat{y}_t$ and the actual observed values, y_t. The formula is given by Equation (7).

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | y_t - \hat{y}_t | \quad (7)$$

Lower MAE indicates a better-performing model. Additionally, MAPE measures the average absolute percentage difference between the forecasted values and the actual observed values. It is calculated as Equation (8).

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (8)$$

3.3.6 Forecasting

Forecasts for future monthly electricity consumption had been calculated using the final model after it had been successfully validated. 95% confidence intervals were included with the forecasts to provide a measure of the amount of uncertainty related to the point predictions.

3.4 Clustering Analysis

Clustering analysis is a method that can be used to group into many clusters based on similar electricity usage patterns. Clustering analysis identifies distinct electricity usage profiles across different months. Clustering enables the grouping of observations based on similarity that showing the hidden patterns within multidimensional datasets.

3.4.1 Data Preprocessing

Min Max Scaler transforms features by scaling each one individually to a given range between 0 and 1. The formula applied is Equation (9).

$$X_{scaled} = \left(\frac{X - X_{\min}}{X_{\max} - X_{\min}}\right) \times (\max - \min) + \min \quad (9)$$

where X is the original feature value, $X_{\max}$ and $X_{\min}$ are the minimum and maximum values of the feature and the range (min, max) is defined by the user typically as (0, 1). Besides, Principal Component Analysis (PCA) can address the challenges of visualizing and analysing datasets with high dimensionality. PCA transforms from original set of features into a new set of uncorrelated variables called principal components.

3.4.2 Clustering Model Selection

K-means clustering to cluster the dataset into distinct groups based on characteristic similarity [25]. K-means minimize the within-cluster sum of squares (WCSS) to ensure each data point is as close as possible to its respective cluster centroid. It can be calculated as Equation (10).

$$WCSS = \sum_{i=1}^{K} \sum_{x \in C_i} (x - \mu_i)^2 \quad (10)$$

Where K is the number of clusters, C_i is the ith cluster, x is a data point. μ_i is the centroid of cluster C_i.

3.4.3 Cluster Profiling

After choosing the optimal number of clusters, the mean values of each feature within each cluster were computed and analysed. It can better understand the distinct traits and characteristics of each cluster. [18] stated that mean cluster analysis helps identify distinct consumer segmentation which is very useful to distinguish average of electricity usage levels, peak demand times and variability.

Radar chart was used to visualize summarise and compare cluster profiles based on multiple variables [26]. Principal Component Analysis (PCA) can be applied in clustering to better understanding multivariate data. Generally, PCA reduces the dimensional into 2 or 3 principal components and the principal components are visualize in pattern scatterplot [27].

4. Results and Discussion

Each section provides a detailed explanation of the results from the models which are supported by tables, figures and interpretations. The findings are aligned with the research objectives which aim to minimize electricity losses, forecast future consumption and categorize electricity usage patterns among sectors.

4.1 Linear Programming

The objective 1 is to minimize electricity losses by optimizing electricity consumption between the industrial, x_1, domestic, x_2 sectors and export, x_3 while satisfying sectoral demand and overall supply constraints. The optimization method applied the monthly data from January 2018 to May 2025 with PuLP in Python.

The electricity loss allocation between sectors follows a pro-rata distribution method where each sector's share of losses is proportional to its consumption [21]. The formula as shown in Equations (11) to (18).

$$Loss\ Share = \frac{Sector\ Consumption}{Total\ Consumption} \times Total\ Losses \tag{11}$$

Where

$$\begin{aligned} Total\ Consumption &= 9390.08 + 2455.82 + 109.82 \\ &= 11955.72 \end{aligned} \tag{12}$$

$$\begin{aligned} Loss\ Share\ of\ Industrial &= \left(\frac{9390.08}{11955.72}\right) \times 1200.38 \\ &= 942.78 \end{aligned} \tag{13}$$

$$\begin{aligned} Loss\ Share\ of\ Domestic &= \left(\frac{2455.82}{11955.72}\right) \times 1200.38 \\ &= 246.57 \end{aligned} \tag{14}$$

$$\begin{aligned} Loss\ Share\ of\ Export &= \left(\frac{109.82}{11955.72}\right) \times 1200.38 \\ &= 11.03 \end{aligned} \tag{15}$$

Divide by each sector's usage to get loss per unit:

$$\begin{aligned} Industrial\ loss\ per\ unit &= \frac{942.78}{9390.08} \\ &= 0.1004 \end{aligned} \tag{16}$$

$$\begin{aligned} Domestic\ loss\ per\ unit &= \frac{246.57}{2455.82} \\ &= 0.1004 \end{aligned} \tag{17}$$

$$\begin{aligned} Export\ loss\ per\ unit &= \frac{11.03}{109.82} \\ &= 0.1004 \end{aligned} \tag{18}$$

The sector’s usage to get loss per unit are loss coefficients in objective function. It is forming the objective function Equation (19).

$$Minimize\ z = 0.1004x_1 + 0.1004x_2 + 0.1004x_3 \tag{19}$$

For each month, industrial, domestic and consumption reductions remained within 5% ceiling of original consumption and total available electricity consumption not more than 105% of the baseline consumption shown as in Equations (20) to (23).

Subject to

$$x_1 \geq 8920.58 (\textit{minimum industrial needs}) \quad (20)$$

$$x_2 \geq 2333.03 (\textit{minimum} \text{ domestic } \textit{needs}) \quad (21)$$

$$x_3 \geq 104.31 (\textit{minimum} \text{ export } \textit{needs}) \quad (22)$$

$$x_1 + x_2 + x_3 \leq 12553.51 (\textit{total available electricity}) \quad (23)$$

The linear programming model was implemented in Python using a monthly loop structure to optimize electricity consumption from 2018 to 2025.

Table 1 Annual Performance Comparison Before and After Optimization

Year	After Optimization (Average Loss, MKWh)	Before Optimization (Average Loss, MKWh)	Reduction
2018	1329.29	1399.26	69.97
2019	1344.85	1415.63	70.78
2020	1315.60	1384.84	69.24
2021	1414.65	1489.10	74.45
2022	1564.16	1646.49	82.33
2023	1090.07	1147.45	57.38
2024	1082.38	1139.34	56.96
2025	1182.11	1244.33	62.22

Based on Table 1, most significant electrical losses reduction in 2022 which is 82.33 MKWh and most efficient performance in 2024 with the lowest optimization losses which is 1082.38 MKWh.

4.2 Time Series Analysis

Based on Fig. 1, the time series plot of total electricity consumption from January 2018 until May 2025 shows a gradual upward trend from approximately 12000 to 16000 MKWh. The plot shows an upward trend over time which means that Malaysia has a significant increase in electricity consumption.

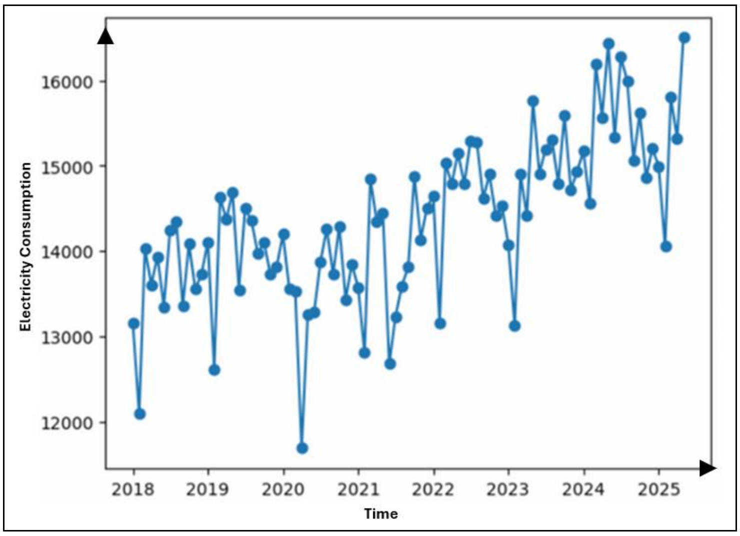
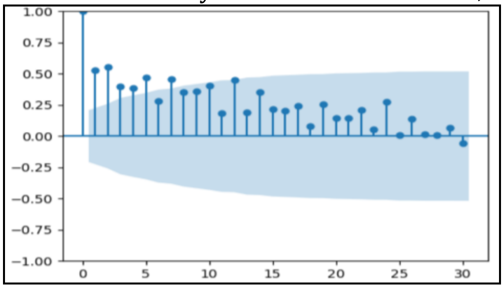


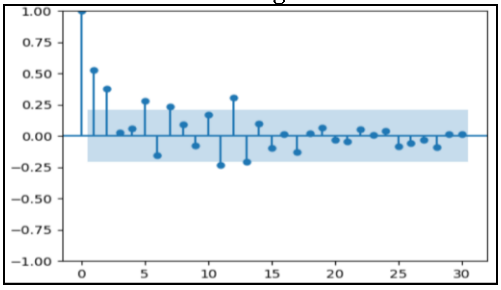
Fig. 1 Time Series Plot on monthly electricity consumption (2018 to 2025)

Furthermore, the time series plot shows predictable fluctuations within each year called as seasonal pattern. Besides, the spread of the data points around the trend means that the variance occurs stable over time. Thus, no logarithmic transformation needs to stabilize the variance before proceeding with further analysis.

In Fig. 2(a), the ACF plot shows a slow decay of autocorrelation across some lags. The autocorrelation is high at lower lags and decay towards zero. The PACF plot of the original series is shown in Fig. 2(b). The plot shows a sharp cut-off after the first few lags where the autocorrelations quickly approach to zero. The pattern means the presence of non-stationarity in the trend. Therefore, the time series need do differencing to achieve stationarity.



(a)



(b)

Fig. 2 (a) Autocorrelation Plot; (b) Partial Autocorrelation Plot

The initial ADF statistic of -0.638 with a *p*-value of 0.862 which greater than 0.05 means that the series is non-stationary. Thus, regular differencing was applied to achieve stationarity. The ADF statistic improved to -4.307 with a *p*-value of 0.00043 which conclude that the series became stationary after the first difference. The dataset split into training and testing sets. The model selection is SARIMA (0,1,2) (1,0,2,12). The model performance is evaluated using standard error metrics. The mean absolute error (MAE) is 310.98, the root mean squared error (RMSE) is 382.88 and the mean absolute percentage error (MAPE) is 1.99%. In addition, the Ljung-Box test can test whether the residuals behave such as white noise. The Ljung-Box statistic of 0.0525 with a *p*-value of 1 which greater than the 0.05 significance level. It concludes that the residuals are independent.

Table 2 presents the monthly electricity consumption forecasts for the period of June 2025 to December 2026. Higher consumption values are during certain months such as March, August, October and December 2026 while lower values appear in months such as July 2025 and February 2026. It linked to increased industrial activities, higher cooling demand due to warmer weather or festival day.

Table 2 Forecast Result

Forecast	Value
1/6/2025	15394.36
1/7/2025	14212.99
1/8/2025	15885.40
1/9/2025	15105.37
1/10/2025	15800.70
1/11/2025	14985.19
1/12/2025	15635.01
1/1/2026	15876.25
1/2/2026	15499.54
1/3/2026	16142.49
1/4/2026	15469.40
1/5/2026	15721.34
1/6/2026	15695.65
1/7/2026	14587.29
1/8/2026	16317.34
1/9/2026	15598.52
1/10/2026	16402.65
1/11/2026	15494.69
1/12/2026	16194.63

4.3 Clustering Analysis

In clustering analysis, the values of electricity consumption were normalized using the Min-Max scaler. Table 3 presents the explained variance ratio for the three principal components derived from the dataset. The results highlight that almost all variability in electricity consumption is captured by the first two principal components that explaining 100% of the variance. PC3 can be excluded from analysis as it does not enhance the explanation of variance in electricity consumption data.

Table 3 Variance Ratio

Principal Component	Explained Variance Ratio	Percentage
PC1	0.7812	78.12%
PC2	0.2188	21.88%
PC3	0.0000	0.00%

Moreover, Table 4 displays the loadings of the three variables on each of the principal components. For the PC1, all three variables show strong positive contributions with domestic consumption of 0.6525 and total consumption of 0.6011 having the highest weights followed by industrial consumption of 0.4615. PC1 represents an overall consumption pattern where increases in total, industrial and domestic electricity consumption move in the same direction.

Table 4 PCA Loadings

Variable	PC1	PC2	PC3
Total Electricity Consumption	0.6011	-0.2995	0.7409
Industrial Electricity Consumption	0.4615	-0.6268	-0.6278
Domestic Electricity Consumption	0.6525	0.7193	-0.2385

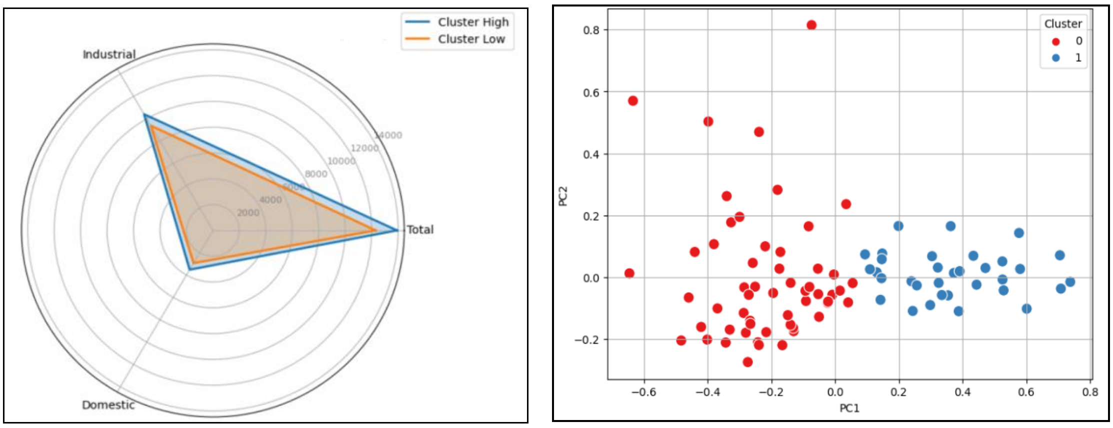
As shown in Table 5, the highest coefficient is obtained at K=2 which is 0.5155, after which the values decline gradually as the number of clusters increases. It provides strong evidence that the dataset is best represented by two clusters so K=2 is the optimal clustering solution.

Table 5 Silhoutte Coefficient

Number of Clusters (K)	Silhouette Coefficient
2	0.5155
3	0.4852
4	0.4279
5	0.4356
6	0.4083
7	0.3879
8	0.3881
9	0.3897
10	0.3946

In addition, the clustered results by month that show the total electricity consumption, industrial and domestic electricity consumption. From 2018 up to early 2022, the months were categorized under the low consumption cluster. A shift in consumption pattern is observed beginning in March 2022 as several months changed into the high consumption cluster. The trend became more consistent throughout 2023 and 2024 where almost all months were classified as high consumption.

Fig. 3(a) presents the radar chart comparing the mean values of total, industrial and domestic electricity consumption across the two clusters. Cluster high is higher values across all three variables while cluster low is months with lower electricity consumption.



(a) (b)

Fig. 3 (a) Radar Chart; (b) Principal Component Analysis Scatterplot

Fig. 3(b) shows the scatterplot demonstrates a separation between the clusters with cluster 0 which is red colour concentrated the left-hand side of the plot and cluster 1 which is blue colour concentrated on the right. Cluster 0 is low electricity consumption and cluster 1 is high electricity consumption. Since PC1 captures the overall level of electricity consumption while PC2 which contrasts domestic and industrial consumption that provides additional variation within each cluster but does not override the dominant separation along PC1.

5. Conclusion and Recommendation

This study successfully achieved its objectives of minimizing electricity losses, forecasting electricity consumption and categorizing consumption levels through clustering analysis. The optimization model effectively reduced electricity losses while satisfying sectoral demands. Besides, the SARIMA model exhibited strong forecasting accuracy that showed increase in electricity demand and distinct seasonal fluctuations influenced by industrial cycles, weather variations and festive periods. Moreover, clustering analysis further distinguished between low and high consumption periods, revealing a transition toward higher electricity usage from 2023 onwards due to industrial expansion, domestic consumption growth and digital transformation.

Several recommendations are proposed to enhance electricity management and policy development. Electricity authorities should adopt the linear programming optimization model for monthly distribution planning to minimize transmission and distribution losses while maintaining sectoral balance. Furthermore, energy policies should be revised to reflect the transition toward higher consumption levels with initiatives that encourage industries to adopt energy-efficient technologies. For the domestic sector, promoting the use of energy-efficient appliances and rooftop solar installations. The clustering approach can be utilized periodically as a monitoring tool to track consumption behaviour and guide policy adjustments based on real-time data.

Acknowledgement

The author sincerely thanks to Universiti Tun Hussein Onn Malaysia (UTHM) for the facilities and academic support provided. Special thanks to the Department of Statistics Malaysia (DOSM) for supplying the electricity consumption dataset.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

*The authors confirm contribution to the paper as follows: **study conception and design:** Tey Sin Ying; **analysis and interpretation of results:** Tey Sin Ying; **draft manuscript preparation:** Tey Sin Ying, Wan Nor Munirah Ariffin. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Ponniran, A., Mamat, N. A., & Joret, A. (2012). Electricity profile study for domestic and commercial sectors. International Journal of Integrated Engineering, 4(3), 8–12. https://penerbit.uthm.edu.my/ojs/index.php/ijie/article/view/206

[2] Sulaima, M. F., Dahlan, N. Y., Hanapi, W. N. A. W., & Din, M. M. N. (2022). Malaysia residential load profile management based on Time of Use tariff using Ant Colony Optimization algorithm. Journal of Sustainability Science and Management, 17(3), 232–242. https://doi.org/10.46754/jssm.2022.03.018
[3] Superchi, F., Giovannini, N., Moustakis, A., Pechlivanoglou, G., & Bianchini, A. (2024). Optimization of the power output scheduling of a renewables-based hybrid power station using MILP approach: The case of Tilos island. Renewable Energy, 220, 119685. https://doi.org/10.1016/j.renene.2023.119685
[4] Yeoh, Y. P., & Kamarudin, A. N. (2022). Forecasting the electricity demand in Malaysia using ARIMA model. Proceedings of Science and Mathematics, 8, 278–288. Retrieved from https://science.utm.my/procscimath/wp-content/uploads/sites/605/2022/10/29.-YEOH-YEN-PING-A18SC0350.pdf
[5] Elsaraiti, M., Ali, G., Musbah, H., Merabet, A., & Little, T. (2021). Time series analysis of electricity consumption forecasting using ARIMA model. 2021 IEEE Green Technologies Conference (GreenTech), 259–262. https://doi.org/10.1109/GreenTech48523.2021.00049
[6] Zhu, L., Liu, J., Hu, C., Zhi, Y., & Liu, Y. (2024). Analysis of electricity consumption pattern clustering and electricity consumption behavior. Energy Engineering, 121(9), 2639–2653. https://doi.org/10.32604/ee.2024.041441
[7] Guan, Y., Zhang, H., Jin, Y., & Zhou, S. (2024). Classification and analysis of users' electricity consumption behavior using cluster analysis algorithm. Journal of Electrical Systems, 20(9s), 1396–1402.
[8] CHINT Global. (2021, August 29). How much power loss in transmission lines. https://www.chintglobal.com/global/en/about-us/news-center/blog/how-much-power-loss-in-transmission-lines.html
[9] Zaim, S., Wan Yusoff, W. N. S., Mohamad, N. N., Ahmad Radi, N. F., & Yaziz, S. R. (2023). Forecasting of electricity demand in Malaysia with seasonal highly volatile characteristics using SARIMA–GARCH model. MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics, 39(3), 293–313. https://doi.org/10.11113/matematika.v39.n3.1512
[10] Zhang, X., Ramírez-Mendiola, J. L., Li, M., & Guo, L. (2022). Electricity consumption pattern analysis beyond traditional clustering methods: A novel self-adapting semi-supervised clustering method and application case study. Applied Energy, 308, 118335. https://doi.org/10.1016/j.apenergy.2021.118335
[11] Clack, C. T. M., Xie, Y., & MacDonald, A. E. (2015). Linear programming techniques for developing an optimal electrical system including high-voltage direct-current transmission and storage. International Journal of Electrical Power & Energy Systems, 68, 103–114. https://doi.org/10.1016/j.ijepes.2014.12.049
[12] Blas, M., Alvarez, G., & Sarli, J. (2023). Improving the optimization of electric power systems through a discrete event based simulation model. Journal of Applied Research and Technology 21(1). https://doi.org/10.22201/icat.24486736e.2023.21.1.2167
[13] Ugbehe, P. O., Diemuodeke, O. E., & Aikhuele, D. O. (2025). Electricity demand forecasting methodologies and applications: A review. Sustainable Energy Research, 12(19). https://doi.org/10.1186/s40807-025-00149-z
[14] Izudin, N. E. M., Sokkalingam, R., Daud, H., Mardesci, H., & Husin, A. (2022). Forecasting electricity consumption in Malaysia by hybrid ARIMA-ANN. In S. A. Abdul Karim et al. (Eds.), Proceedings of the 6th International Conference on Fundamental and Applied Sciences (pp. 749–760). Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-16-4513-6_66
[15] Tai, V. C., Tan, Y. C., Abd Rahman, N. F., Che, H. X., Chia, C. M., Saw, L. H., & Ali, M. F. (2021). Long-term electricity demand forecasting for Malaysia using artificial neural networks in the presence of input and model uncertainties. Energy Engineering, 118(3), 791–809. https://doi.org/10.32604/EE.2021.014865
[16] Cui, Y., Zhu, Z., Zhao, X., & Li, Z. (2023). Energy schedule setting based on clustering algorithm and pattern recognition for non-residential buildings electricity energy consumption. Sustainability, 15(11), 8750. https://doi.org/10.3390/su15118750
[17] Kusuma, D. T., Ahmad, N., & Syed Ahmad, S. S. (2024). Clustering algorithm for electrical load profiling analysis: A systematic review of machine learning approaches for improved clustering algorithms. Journal of Theoretical and Applied Information Technology, 102(10), 4454–4463. https://www.jatit.org/volumes/Vol102No10/21Vol102No10.pdf
[18] Guan, Y., Zhang, H., Jin, Y., & Zhou, S. (2024). Classification and analysis of users' electricity consumption behavior using cluster analysis algorithm. Journal of Electrical Systems, 20(9s), 1396–1402.
[19] Department of Statistics Malaysia. (2024, September 1). Monthly electricity consumption. OpenDOSM. https://open.dosm.gov.my/data-catalogue/electricity_consumption
[20] Kimutai, I., Maina, P., & Makokha, A. (2019). Energy optimization model using linear programming for process industry: A case study of textile manufacturing plant in Kenya. International Journal of Energy Engineering, 9(2), 45-52. https://doi.org/10.5923/j.ijee.20190902.03
[21] LTSE Team. (2025). What is pro rata? Long-Term Stock Exchange. Retrieved April 14, 2025, from https://ltse.com/insights/what-is-pro-rata

[22] Elrefaei, H., & Khalifa, M. A. (2014). A critical review on the National Energy Efficiency Action Plan of Egypt. Journal of Natural Resources and Development, 4, 18–24. https://doi.org/10.5027/jnrd.v4i0.03
[23] iBForecast. (2024, January 20). Which of the following forecasting methodologies is considered a time series forecasting technique? https://ibforecast.com/which-of-the-following-forecasting-methodologies-is-considered-a-time-series-forecasting-technique/
[24] Jedox. (2021, July 6). Error metrics: How to evaluate your forecasts. Jedox. https://www.jedox.com/en/blog/error-metrics-how-to-evaluate-forecasts/
[25] ML Journey. (2024, June 3). Elbow method in Python: A comprehensive guide. https://mljourney.com/elbow-method-in-python-a-comprehensive-guide/
[26] Fajri, M., Rais, R., Gamayanti, N. F., Dg Mabaji, S. N., Jati, S. Y. R., & Arisandi, R. (2024). Mapping of village population profile with schistosomiasis cases using clustering large applications. Jurnal Varian, 7(2), 207–216. Retrieved from https://www.researchgate.net/publication/381985675_Mapping_of_Village_Population_Profile_with_Schistosomiasis_Cases_Using_Clustering_Large_Applications
[27] Aluja-Banet, T., Morineau, A., & Sanchez, G. (2018). PCA and clustering. Principal component analysis for data science. PCA4DS. https://pca4ds.github.io/pca-and-clustering.html