Vol 8 No 2 (2026) *e*-ISSN: 2682-8383

Journal of Engineering Technology and Applied Physics

Electricity Demand Forecasting using A Supervised Machine Learning Random Forest Algorithm

Abdul Halim Ikram Mohamed*, Ruzlaini Ghoni and Mohd Tarmizi Ibrahim
Faculty of Engineering Technology, University College Tati, 24100 Kemaman, Terengganu, Malaysia.
**Corresponding author*: 24D01001@siswa.uctati.edu.my, *ORCiD*: 0009-0003-0897-7780
https://doi.org/10.33093/jetap.2026.8.2.14
Manuscript Received: 8 January 2026, Revised: 13 February 2026, Accepted: 17 March 2026, Published: 15 September 2026

***Abstract*—The design and development of national power grids are critical to delivering a reliable electricity supply in Malaysia, where accurate load forecasting in low-voltage distribution networks is crucial in ensuring grid efficiency and stability. Traditional forecasting methods often lack of precision and adaptability required for modern energy management systems, particularly in dynamic environments such as the oil and gas industry. This study leverages supervised machine learning, specifically Random Forest, to enhance load-prediction accuracy in low-voltage networks, with a focus on Gas and Condensate Receiving facilities in the oil and gas sector in Kerteh, Terengganu, Malaysia. Based on analyses of variables such as equipment rating (kW), coincident factor, equipment efficiency (%), and equipment load duty, a robust forecasting model that demonstrates significant improvements over conventional approaches is developed. The result is reported as a best value of 80:20% training and testing split (scenario 4) after optimization with MAE of 1.01, MAPE of 0.49%, RMSE of 1.43, and R^2 of 99.07%. This show that the proposed method not only achieves higher forecasting accuracy but will also improves operational efficiency, reduces energy waste and enhances grid reliability. These findings highlight the transformative potential of machine learning for data-driven in low voltage distribution network, facilitating the integration of renewable energy sources and supporting Malaysia's grid modernization efforts.**

Keywords*—*Supervised Machine Learning, Random Forest, Regression analysis, Low voltage distribution network and electricity demand.

I. INTRODUCTION

Accurate load demand forecasting is of utmost importance for ensuring the secure, efficient, and economically viable operation of electrical power systems, particularly in the low voltage distribution networks industrial sector, where the stakes are considerably high. This level of precision is not merely beneficial but fundamentally vital for optimizing the distribution of energy resources, significantly minimizing waste, and strengthening the stability of the power grid, especially in dynamic environments characterized by frequent fluctuations in electricity demand [1]. Balancing electricity supply and demand in developing nations like Malaysia is a complex challenge that requires innovative strategies to prevent both resource wastage from oversupply and grid strain from undersupply. The increasing electricity demand, driven by rapid economic growth, urbanization, and technological advancements, underscores the urgent necessity for the adoption of advanced techniques, including Demand Side Management (DSM), the optimization of energy management systems, and the refinement of existing load forecasting methodologies [2, 3]. Besides that, traditional forecasting methods frequently fall short in these complex scenarios, highlighting the urgent need for more sophisticated, adaptive, and precise predictive models to manage evolving energy landscapes [4].

Although ongoing research is abundant in this field, the existing literature highlights significant shortcomings that currently restrict its implementation to industrial activities. While many studies have demonstrated the model's effectiveness using a comprehensive dataset [5], there remains a constraint and a lack of discussion regarding the inherent limitations of the dataset. These constraints involve crucial factors such as industry type, temporal coverage, and the potential impact of operational contingencies, which could adversely

Published by MMU PRESS. URL: https://journals.mmupress.com/index.php/jetap/index

affect the model's generalizability and applicability [6]. Furthermore, there is a need to explore the potential impact of other factors, such as equipment efficiency, coincident factors, and equipment duty cycle whether continuous, intermittent, or stand by, over a specified period of time, which could also contribute to inaccuracies in the predictions made by the model [7].

Therefore, the use of the supervised machine learning Random Forest (RF) algorithm for predicting electricity demand has emerged as a highly effective and innovative approach within the expansive field of energy management, underscoring its relevance and applicability. This advanced machine learning technique is particularly well-equipped to handle the inherently complex, non-linear relationships that characterize electricity consumption data [6]. The remarkable capability of the RF algorithm to process extensive datasets while adeptly incorporating multiple influencing factors, including environmental conditions and socioeconomic variables, positions it as a powerful and indispensable tool for achieving accurate and reliable demand forecasting [8].

Leveraging historical electricity consumption data, weather conditions, and operational events, the RF can effectively forecast demand and identify anomalies in low-voltage distribution networks [9]. This approach facilitates a more robust understanding of power consumption patterns, enabling proactive adjustments to energy distribution and improved operational efficiency [9]. The integration of supervised machine learning for predicting electricity demand not only optimizes real-time load adjustments but also plays a critical role in reducing energy waste, lowering operational costs, and ultimately enhancing the reliability and stability of the electrical grid.

This article presents the development and rigorous evaluation of a supervised machine learning–based electricity demand forecasting framework employing the RF algorithm for low voltage industrial distribution networks, with particular emphasis on large-scale oil and gas operations in Peninsular Malaysia. The work addresses key limitations reported in existing studies by systematically assessing the model's scalability, robustness, and generalizability across heterogeneous industrial datasets that capture realistic operational conditions, including temporal variability, sector-specific load characteristics, and operational contingencies. By integrating historical electricity consumption data with contextual operational factors, the study improves forecasting electricity demand accuracy. It provides empirically grounded insights to support reliable demand prediction, enhanced energy management strategies, and informed decision-making for secure and efficient power system operation in developing industrial environments.

The remainder of this article is structured as follows. Section I reviews prior studies on RF–based electricity demand forecasting and identifies existing research gaps. Section II describes the proposed methodology, including model development, data preprocessing, feature selection, hyperparameter tuning, and evaluation metrics. Section III to VII presents the industrial case study and simulation results, with comparative analysis of model performance before and after optimization. Section VIII concludes the paper with key findings and future research directions.

II. METHODOLOGY

The RF algorithm is an advanced ensemble learning method that enhances predictive accuracy and stability by constructing multiple independent decision trees [10]. This approach effectively addresses the limitations of a single decision tree, which often exhibits high variance and a tendency to overfit the training data. The RF algorithm can be visualized as shown in Fig. 1.

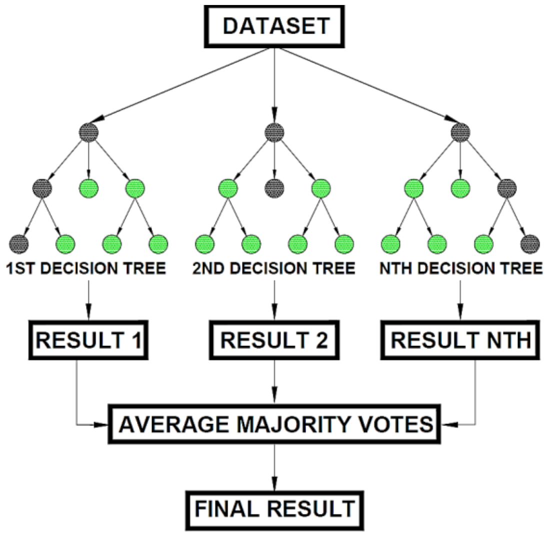


Fig. 1. RF algorithm.

RF achieves robust generalization capacity through the application of two core randomization techniques mainly bagging (bootstrap aggregating) and the random subspace method (or feature randomness). Given a training dataset comprising feature vectors($X = x_1, x_2, x_3 \dots x_n$) and their corresponding output values ($Y = y_1, y_2, y_3 \dots y_n$), the process begins with bagging. Multiple distinct training subsets, known as bootstrap samples, are repeatedly drawn with replacement from the original data. A dedicated and independent regression tree is then fitted to each of these unique samples. This sampling strategy ensures that each tree sees a slightly different version of the data, which is fundamental to introducing necessary diversity across the forest.

Simultaneously, as each individual tree undergoes recursive partitioning to determine optimal data splits, the algorithm employs the random subspace method. At every decision node, the tree is constrained to consider only a small random subset of the available features rather than evaluating all possible predictors. This feature constraint is particularly crucial in datasets where one or a few features might otherwise dominate all the trees leading to correlated outputs. By forcing the trees to rely on different combinations of predictors, the random subspace method ensures that the resulting trees are highly decorrelated. The partitioning within each tree continues until a predefined stopping criterion is met, such as reaching a minimum number of samples in a leaf node which further acts as a regularization mechanism.

This dual-randomization strategy is the cornerstone of the RF achievement. The resulting low correlation among the individual trees significantly reduces the overall variance of the ensemble model. This dual-randomization strategy is the cornerstone of the RF's success. The resulting low correlation among the individual trees significantly reduces the overall variance of the ensemble model. For a regression task like predicting electricity demand, the forest makes its final prediction by aggregating the outputs of all N trees. The mathematical expression for RF to predict electricity demand is given as follows.

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \quad (1)$$

, where

$f(x)$ is the prediction output for a given input x;

N is the total number of trees in the forest;

$T_i(x)$ is the prediction from a single tree.

A. Data Preprocessing

A comprehensive data preprocessing is conducted prior to model training to ensure the dataset quality and integrity. This involves standard procedures such as data cleaning to handle missing or anomalous entries and normalization or scaling of continuous variables. Although RF is generally robust to variations in feature scale, normalization ensures that the feature importance metrics, particularly those used in the subsequent selection process, are not biased by features with disparate ranges. Features with larger numerical ranges can arbitrarily dominate the model's distance calculations. Therefore, applying a StandardScaler to normalize all features to have zero mean and unit variance is a mandatory preprocessing step to ensure valid model performance [11].

B. Feature Selection

After preprocessing the dataset, a feature selection of Recursive Feature Elimination (RFE) is employed to the RF model. The objective of this step is to identify and retain only the most significant variables that demonstrated a substantial influence on electricity consumption, thereby reducing model complexity and improving predictive accuracy. RFE is an iterative wrapper-style feature selection algorithm that systematically removes the least important features from a dataset. A wrapper method iteratively removes features with the lowest importance, effectively reducing model complexity, mitigating overfitting, and ensuring only the most relevant parameters contribute to the prediction.

C. Hyperparameter Tuning

In order to enhance the overall performance metrics of the model, the RF algorithm is subjected to an optimization process which is realized through the employment of an advanced Grid Search Cross-Validation (CV) technique [12]. This method systematically explores a predefined grid of hyperparameter combinations to identify the optimal set that yields the highest cross-validated performance. For each possible parameter combination, the algorithm trains and evaluates the RF model using K-fold cross-validation. This comprehensive and iterative process significantly contributes to ensuring that the model performance on any given training and testing datasets are both robust and reliable as it effectively mitigates the risks associated with underfitting and overfitting, thereby leading to a substantial reduction in the generalization error that adversely affect the predictive capabilities of the model. The input hyperparameters tuning are summarized in Table I.

Table I. Hyperparameter tuning of RF algorithm.

No	Parameter	Description	Values
a.	n_estimators	The total number of decision trees in the forest.	[10, 12, 35]
b.	max_depth	The maximum depth of each individual tree, which controls model complexity and helps prevent overfitting.	[12, 25, 40]
c.	min_samples_leaf	Number of features considered for the best split at each node, which controls the diversity of the trees.	[1, 4, 7]

III. MODEL SIMULATION

A comprehensive simulation procedure is implemented using Python's open-source libraries to evaluate the predictive performance of the RF model and assess its robustness across various data partitioning schemes. The core of the simulation involved dividing the available dataset into four distinct scenarios, each comprising a specific combination of training and testing sets as summarized in Table II.

The simulation progressed through two phases for each of the four scenarios. In Phase 1, the RF model is trained and tested without feature selection and hyperparameter tuning. The objective of this phase is to establish a baseline accuracy and gauge the RF algorithm's inherent capability to predict electricity demand based on the specified input parameters. After that, in Phase 2, the RF algorithm is subsequently trained and tested on the same four scenarios, but this time incorporating RFE for feature selection and Grid Search CV for hyperparameter tuning. Finally, the baseline model before optimization and the optimized model after optimization are compared for each scenario. The results are also visualized in graphs and then discussed.

Table II. Four scenarios of training and testing set.

Scenario	Training (%)	Testing (%)	Set
1	20	80	20:80
2	50	50	50:50
3	70	30	70:30
4	80	20	80:20

IV. MODEL PERFORMANCE EVALUATION

Four performance metrics that are Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) are employed to evaluate the accuracy of the RF model in predicting electricity demand. These metrics provide a comprehensive understanding of the model performance from different perspectives.

A. Mean Absolute Error (MAE)

The MAE quantifies the average of the absolute differences between actual and predicted electricity demand values. This metric illustrates the extent to which the predictions made by the model diverge from the actual values. The lower the MAE value, the better the RF model performance in predicting electricity demand. The mathematical expression of MAE can be written as

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|actual_i - predict_i| \qquad (2)$$

B. Mean Absolute Percentage Error (MAPE)

The MAPE measures the average of the absolute percentage difference between actual and predicted electricity demand values. This metric enables an assessment of the significant discrepancies between the predicted and actual values of electricity demand. A lower MAPE percentage indicates a higher level of forecasting precision. The mathematical expression of MAPE can be written as

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{actual_i - predict_i}{actual_i}\right| \times 100\% \qquad (3)$$

C. Root Mean Square Error (RMSE)

The RMSE is defined as the square root of the average of the squared differences between actual and predicted electricity demand values. This metric quantifies the deviation exhibited by the model in predicting electricity demand compared to actual values. A lower value of RMSE denotes that a superior model performance and indicates that the predicted values are closely aligned with the actual. The mathematical expression of RMSE can be written as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(actual_i - predict_i)^2} \qquad (4)$$

D. Coefficient of Determination (R^2)

The R^2 is a statistical measure that represents the proportion of the variance in electricity demand that can be attributed to the input parameters. A high percentage of R^2 approaching 100% signifies an exceptional fit, indicating that the model effectively captures the underlying patterns in the dataset. Conversely, a lower percentage of R^2 indicates that the model explains a limited portion of the variability in the dataset. The mathematical expression of R^2 can be written as

$$R^2 = \left[1 - \frac{\sum_{a=1}^{n}(actual_i - predict_i)^2}{\sum_{a=1}^{n}(actual_i - average\ actual_i)^2}\right] \times 100\% \qquad (5)$$

V. CASE STUDY

The Kerteh Industrial Complex (KIC) in Terengganu, which has been operating since 1993, is a crucial Malaysian hub integrating crude oil refining, gas processing, and petrochemical manufacturing. As a high-volume, heavy industry centre, its processes are highly systematic and integrated, with the output of one unit often feeding into the next. This continuous operation necessitates advanced technology and massive equipment, resulting in exceptionally high electricity demand. Maintaining a stable power supply is paramount, as outages are costly, dangerous, and require time-consuming restarts. Therefore, the complex mandates systematic electricity demand planning, supported by sophisticated load forecasting and real-time monitoring, to ensure power grid stability. To improve the prediction of electricity demand and operational resilience, this study applied an RF model to a three-year dataset of historical consumption records spanning January 2022 to December 2024.

There are many processes, plants, and receiving facilities available in KIC. This research focuses on one of the Gas and Condensate Receiving (GCR) facilities. Currently, GCR facilities are fed by a redundant 11 kV supply from public utilities. The 11 kV supply then stepped down via 11/0.433 kV power transformers to supply power to the 415 V main control centre (MCC) panel, namely SB100E1-A and SB100E1-B. The outgoing supply of MCC SB100E1-A is distributed to an existing 415 V MCC SB100E4. Meanwhile, the outgoing MCC SB100E1-B supplied power to a 415 V MCC SB100E3. Figure 2 illustrates the electrical power system configuration of GCR facilities.

415 V MCC SB100E1-A/B, SB100E3, and SB100E4 supplies power to 85 essential loads and equipment. Due to the MCC panel's intolerance to outages, a power supply disruption to this MCC panel could result in severe operational consequences, including potential catastrophic

failure and substantial financial losses. The peak electricity demand for all three 415 V MCC panels is influenced by four main input parameters, as summarized in Table III.

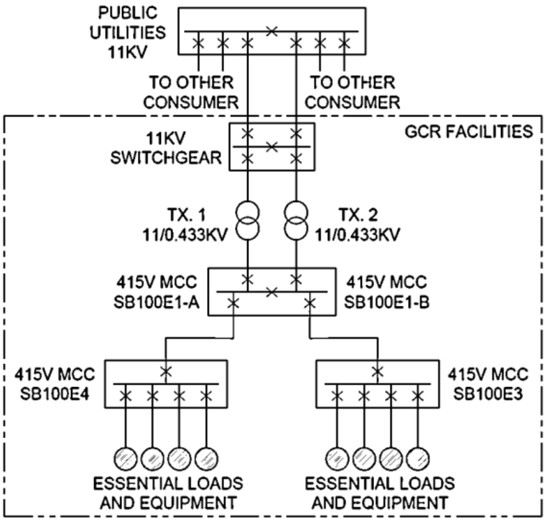


Fig. 2. Electrical power system configuration of GCR facilities.

Table III. Input parameters influencing electricity consumption of GCR facilities.

Items	Description			Unit
Equipment Rating	Nominal power ratings of each connected equipment connected to MCC panel			kW
Coincident Factor	A ratio that quantifies the probability of a group of loads operating simultaneously			Varies
Efficiency	Equipment efficiency when in operation			%
Equipment Load Duty	a	Continuous	Operates at a constant load for a long period	100%
	b	Intermittent	Operates at a constant load for sometimes	30%
	c	Standby	Operates during emergency	10%

VI. EQUIPMENT INFORMATION

A total of 85 distinct loads and equipment connected to the dedicated panel are identified. The loads and equipment can be categorized into four primary types, each with its associated electrical parameters, as specified in Table IV.

Table IV. Types of electrical equipment & loads of GCR facilities.

Type of Loads	No. of Equipment	Rating (kW)	Coincident Factor	Efficiency (%)
Lighting & Motorised Operated Valve Panel	47	2 to 32	0.80 to 0.95	85 to 98
Pumps and Motors	21	2.2 to 110	0.80 to 0.86	83 to 95
Fan and Blowers	12	0.5 to 5.6	0.80	68 to 87
Heating, Ventilation, Airconditioning (HVAC)	5	58 to 120	0.85	85
Total	85			

Each equipment power output rating (kW) is compiled through an on-site collection of the connected load list from the respective 415 V MCC panel. The operational efficiencies (%) for each piece of equipment are sourced directly from the respective manufacturer and provided in the electrical equipment datasheets. The coincident factors and equipment load duty applied to each equipment are determined in accordance with the system design philosophy. The historical record of electrical load consumption profiles spanning three years for all 415 V MCC panels is presented in Fig. 3.

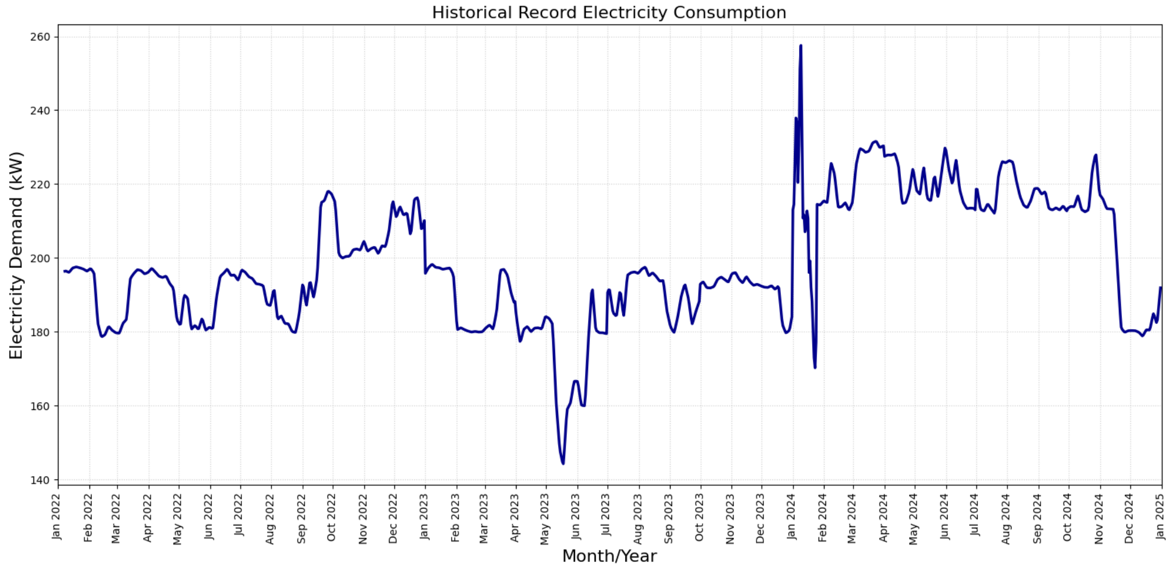


Fig. 3. Historical record electricity peak demand of GCR facilities.

VII. RESULT AND DISCUSSION

This section presents a quantitative assessment and discussion of the RF model performance in predicting electricity demand, utilizing the four specified input parameters. The analysis focuses on a comparative evaluation before and after the application of optimization techniques, specifically RFE for feature selection and Grid Search CV for hyperparameter tuning. This study also uses a historical record of the MCC panel spanning 2022 to 2024 to monitor electricity consumption. Table V summarizes the model performance before and after the application of optimization techniques.

Table V. Comparison of RF model performance before and after optimized.

Scenario/ Training & Testing Set		Before Optimized (Phase 1)			
		MAE	MAPE (%)	RMSE	R^2 (%)
1	20:80	10.61	5.02	15.29	26.70
2	50:50	5.54	2.57	7.76	77.20
3	70:30	2.48	1.13	3.86	92.19
4	80:20	1.18	0.57	1.64	98.78
Scenario/ Training & Testing Set		After Optimized (Phase 2)			
		MAE	MAPE (%)	RMSE	R^2 (%)
1	20:80	10.48	4.95	15.19	27.70
2	50:50	4.94	2.29	7.30	79.82
3	70:30	2.47	1.13	3.61	93.18
4	80:20	1.01	0.49	1.43	99.07

The evaluation successfully validated the RF model performance and robustness in predicting the critical electricity demand of all 415 V MCC panels within the GCR facilities. Each scenario exhibits different performance. The study also empirically confirmed that a fundamental principle of predictive modelling accuracy is intrinsically linked to the volume of training data. Higher proportions of data allocated to the training set consistently led to improve generalization and minimized error metrics, thereby establishing the need for a sufficiently large dataset to achieve high-fidelity forecasts. Meanwhile, the optimization technique also plays an important role, as it gives an impact and enhances the RF model performance.

Specifically, as detailed in Table V, Scenario 1 with 20% training and 80% testing yielded the worst performance for the RF model's prediction of electricity demand, with a MAE of 10.61, a MAPE of 5.02%, a RMSE of 15.29, and an R^2 of 26.70% prior to optimization. The limited training data, 20% of the total dataset, hindered the model's learning process, directly contributing to these suboptimal results. Conversely, model performance systematically improved as the training proportion increased, as demonstrated by scenarios 2 (50:50 split), 3 (70:30 split), and 4 (80:20 split).

After optimization, scenario 1 maintained the lowest performance, with an MAE of 10.48, an MAPE of 4.95%, an RMSE of 15.19, and an R^2 of 27.70%. However, the optimized model consistently demonstrated superior performance metrics compared to its unoptimized counterpart across all scenarios. This outcome underscores the critical importance of rigorous feature selection and hyperparameter tuning to maximize the predictive accuracy and generalization of the RF model.

The deployment of systematic optimization techniques is pivotal in maximizing the model's predictive precision, which involved two primary methods that are feature selection of RFE and hyperparameter tuning of Grid Search CV. The combined application of these techniques resulted in the optimal model configuration, with a 80:20% data split and optimization, yielding an outstanding MAE of 1.01, MAPE of 0.49, RMSE of 1.43, and R^2 of 99.07%. This result strongly validates that strategic optimization significantly enhances the RF model generalization capability and accuracy. The result is also shown in Fig. 4, which compares the actual and predicted accuracies of the RF model before optimization for scenario 4. Meanwhile, Fig. 5 compares the actual and predicted accuracies of the RF model after optimization for scenario 4.

However, the analysis also underscored the need for careful model handling concerning outliers, which pose a significant challenge in industrial load forecasting. Anomalies resulting from a typical operational incident, including abrupt equipment malfunctions or the implementation of emergency protocols have the potential to manipulate the decision boundaries of the model and undermine its capacity for effective generalization [13]. If these anomalies are unaddressed, it can lead to model overfitting to rare events potentially degrading the accuracy of routine electricity demand predictions. Therefore, a robust preprocessing strategy involving outlier detection and handling is deemed necessary to maintain the high predictive performance required for reliable electricity demand planning in this critical industrial environment [14]. This study also has demonstrated that the RF approach particularly when integrated with an optimization process and careful outlier management provides a highly accurate and dependable solution for critical industrial load forecasting. The comparison of the RF model performance before and after optimization for each scenario is also comprehensively illustrated in graphical representations, as summarized in Tables VI for scenario 1 and scenario 2 and Table VII for scenario 3 and scenario 4.

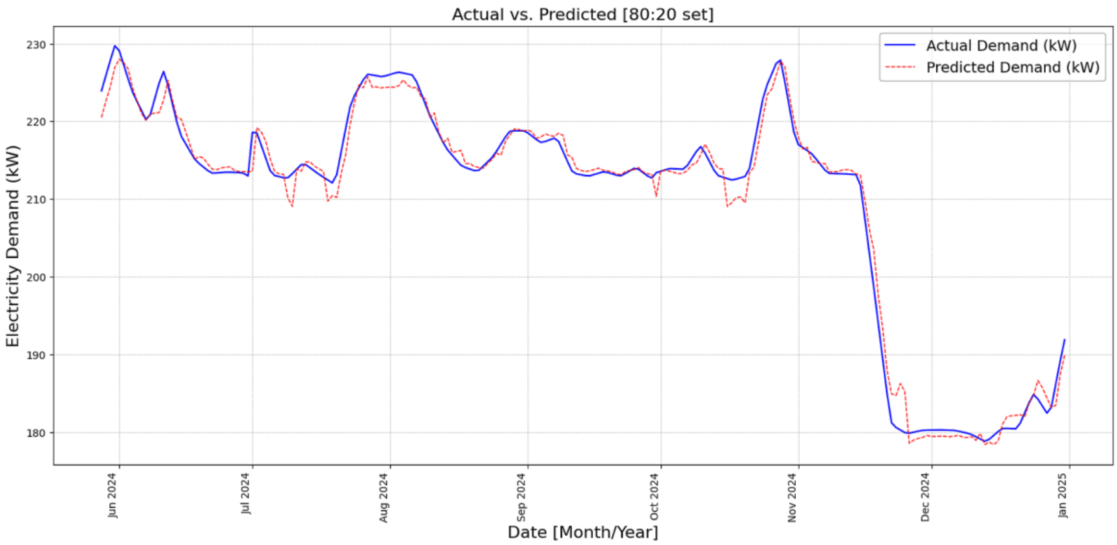


Fig. 4. Comparison between actual vs predicted of forecasting electricity demand before optimized for scenario 4.

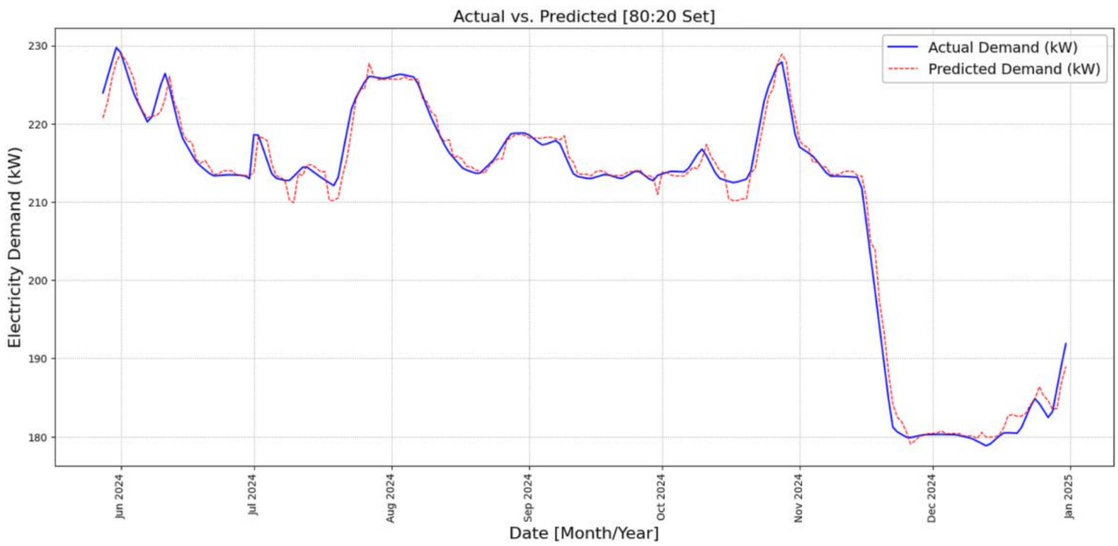


Fig. 5. Comparison between actual vs predicted of forecasting electricity demand after optimized for scenario 4.

Table VI. Actual vs predicted of the RF model performance for scenario 1 & 2.

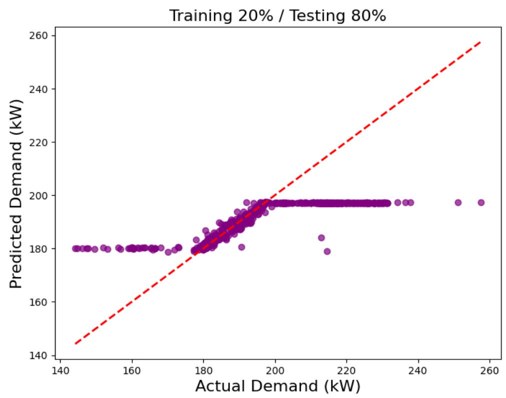
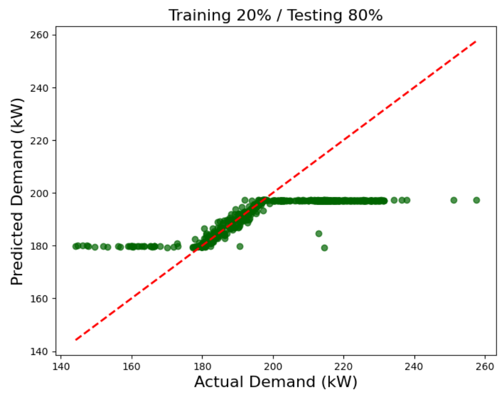
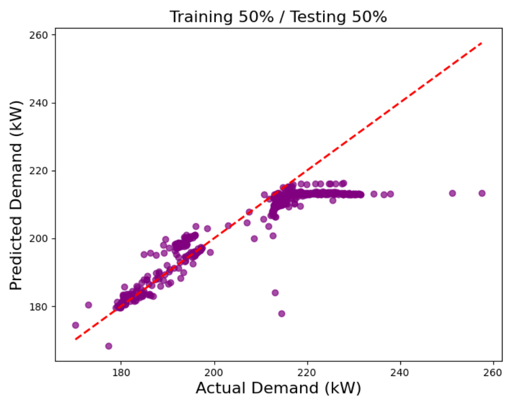
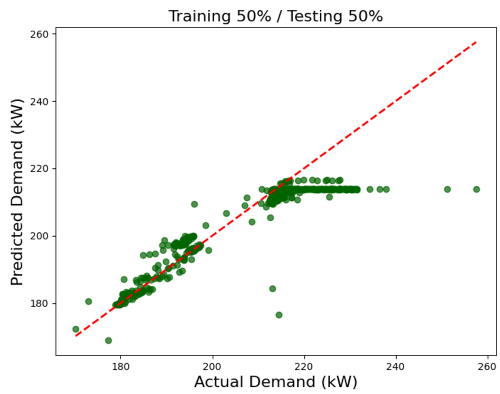
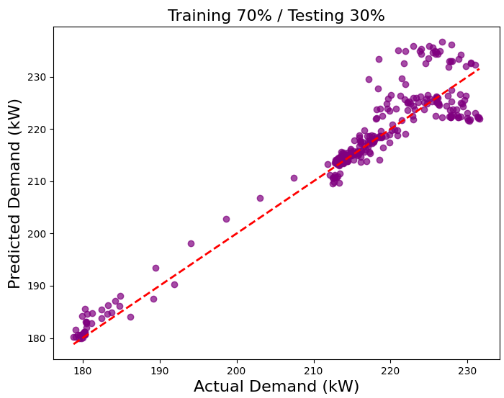
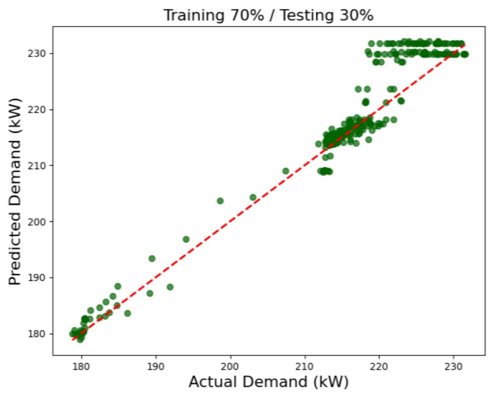
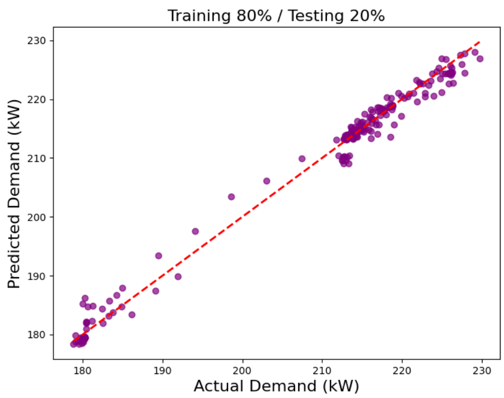
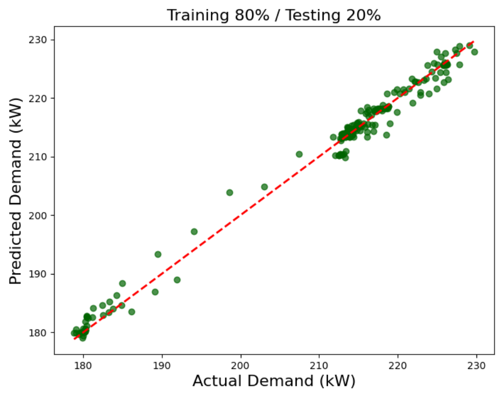
Before Optimized (Phase 1)		After Optimized (Phase 2)	
			
Scenario 1: 20:80% split		Scenario 1: 20:80% split	
MAE	10.61	MAE	10.48
MAPE (%)	5.02	MAPE (%)	4.95
RMSE	15.29	RMSE	15.19
R^2 (%)	26.70	R^2 (%)	27.70
			
Scenario 2: 50:50% split		Scenario 2: 50:50% split	
MAE	5.54	MAE	4.94
MAPE (%)	2.57	MAPE (%)	2.29
RMSE	7.76	RMSE	7.30
R^2 (%)	77.20	R^2 (%)	79.82

Table VII. Actual vs predicted of the RF model performance for scenario 3 & 4.

<table>
<tr><th colspan="2">Before Optimized (Phase 1)</th><th colspan="2">After Optimized (Phase 2)</th></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td colspan="2">Scenario 3: 70:30% split</td><td colspan="2">Scenario 3: 70:30% split</td></tr>
<tr><td>MAE</td><td>2.48</td><td>MAE</td><td>2.47</td></tr>
<tr><td>MAPE (%)</td><td>1.13</td><td>MAPE (%)</td><td>1.13</td></tr>
<tr><td>RMSE</td><td>3.86</td><td>RMSE</td><td>3.61</td></tr>
<tr><td>R^2 (%)</td><td>92.19</td><td>R^2 (%)</td><td>93.18</td></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><td colspan="2">Scenario 4: 80:20% split</td><td colspan="2">Scenario 4: 80:20% split</td></tr>
<tr><td>MAE</td><td>1.18</td><td>MAE</td><td>1.01</td></tr>
<tr><td>MAPE (%)</td><td>0.57</td><td>MAPE (%)</td><td>0.49</td></tr>
<tr><td>RMSE</td><td>1.64</td><td>RMSE</td><td>1.43</td></tr>
<tr><td>R^2 (%)</td><td>98.78</td><td>R^2 (%)</td><td>99.07</td></tr>
</table>

Furthermore, Fig. 6 presents the feature importance score, quantifying the relative contribution of the four primary input parameters to the RF model.

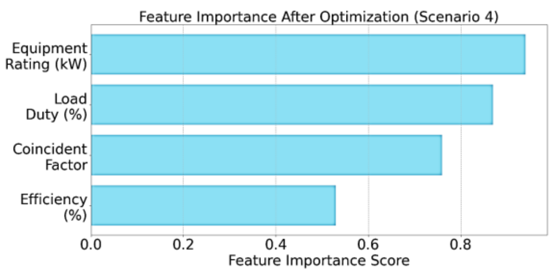


Fig. 6. Feature importance score after optimization.

The analysis identifies equipment rating and its load duty as the predominant drivers of electricity demand within the GCR facilities. This is aligned with industrial environment where the facility utilizes high power machinery with a continuous operation, where one process output serves as subsequent inputs to the next process. The significant contribution of the coincident factor further supports the model ability to capture the integrated nature of the KIC, ensuring that grid stability is maintained without wasting the resources. By applying RFE, the model has improved its ability to predict electricity demand, resulting in the high R^2 of 99.07% in Scenario 4.

VIII. CONCLUSION

The RF algorithm is developed to predict electricity demand for one of the GCR facilities in Kerteh Industrial Complex, Terengganu, Malaysia. This study considers four input parameters mainly equipment rating (kW), coincident factor, equipment efficiency (%), and equipment load duty (%), with a historical record of electricity consumption spanning 2022 to 2024. Four scenarios, each with its own training and testing set combination are presented.

In Phase 1, the model is trained and tested without feature selection and hyperparameter tuning to assess its performance in predicting electricity demand without optimization. In phase 2, using the same training and testing sets, the model is applied to feature selection using RFE to select the most relevant features for training. The model is also optimized using a grid search cross-validation technique. The model performance before and after optimization is then compared and presented in graphs. To evaluate the model performance, four performance matrices which are MAE, MAPE, RMSE, and R^2 are used for each training and testing set. The model performance is reported as a best value of 80:20% training and testing split (scenario 4) after optimization with MAE of 1.01, MAPE of 0.49%, RMSE of 1.43, and R^2 of 99.07%. This is a very effective prediction result, indicating that the proposed algorithm effectively addresses the problem of forecasting electricity demand using all four input parameters.

For the training and testing set without optimization, the evaluation model reported a 80:20% split with MAE of 1.18, MAPE of 0.57%, RMSE of 1.64, and R^2 of 98.78%. In this case, the results indicated that the complete model can predict demand with acceptable accuracy, especially given the variation in oil and gas industry demand. Finally, these techniques have been successfully applied to forecasting problems with complex state structures in the explanatory variables, making them valuable tools for dealing with uncertainty in electricity demand. Another line of future work is to enrich the studied models to generate mid- and long-term synthetic demand scenarios that preserve the statistical structure of historical data. These models are highly relevant for inclusion in planning and operational models based on new computational intelligence techniques, such as reinforcement learning or approximate dynamic programming. Furthermore, prediction results can be applied in practice to household energy planning through intelligent recommendation systems.

ACKNOWLEDGEMENT

The author gratefully acknowledges the Faculty of Engineering Technology, University College Tati for supporting this research and ensuring its successful completion.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Abdul Halim Ikram Mohamed: Conceptualized the study, established the research framework, and provided primary supervision and critical revisions.

Ruzlaini Ghoni: Managed manuscript restructuring for logical flow and performed an extensive synthesis of correlated literature.

Mohd Tarmizi Ibrahim: Executed technical formatting, proofreading, and a preliminary compliance review against journal guidelines.

CONFLICT OF INTERESTS

No conflict of interests was disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

DATA AVAILABILITY STATEMENT

No available

REFERENCES

[1] S. H. Khaleefah *et al.*, "Optimizing Smart Power Grid Stability Based on The Prediction of A Deep Learning Model," *Int. J. Informat. Visual.*, vol. 8, no. 3, pp. 1091-1098, 2024.

[2] N. Ahmad *et al.*, "Load Profile Forecasting using A Time Series Model for Solar Rooftop and Integrated Carpark of A Public University in Malaysia," *J. Adv. Res. Fluid Mechan. and Therm. Sci.*, vol. 111, no. 2, pp. 86–98, 2024.

[3] M. M. B. Tufail *et al.*, "Forecasting Impact of Demand Side Management on Malaysia's Power Generation using System Dynamic Approach," *Int. J. Ener. Econom. and Polic.*, vol. 11, no. 4, pp. 412–418, 2021.

[4] S. Narayanan, R. Kumar, S. Ramadass and J. Ramasamy, "Hybrid Forecasting Model Integrating RNN-LSTM for Renewable Energy Production," *Electr. Pow. Compon. and Syst.*, pp. 1–19, 2024.

[5] M. A. H. Abdul Haris and S. L. Lim, "Neural Network Facial Authentication for Public Electric Vehicle Charging Station," *J. Eng. Technol. and Appl. Phys.*, vol. 3, no. 1, pp. 17–21, 2021.

[6] Y. Liu, "Predicting Short-term Electric Power Demand using Weather Data," *IET Conf. Proc.*, vol. 2024, no. 19, pp. 101–108, 2024.

[7] K. Alhashemi and O. T. Altinoz, "Applied Time Series Regression by Using Random Forest Algorithm for Forecasting of Electricity Consumption on a Daily Basis," *Proc. 2022 Int. Symp. Ener. Manage. and Sustain.*, pp. 197-207, 2023.

[8] Y. A. Atalan, H. Şahin, A. Keskin and A. Atalan, "Strategic Forecasting of Renewable Energy Production for Sustainable Electricity Supply: A Machine Learning Approach Considering Environmental, Economic, and Oil Factors in Türkiye," *PLoS One*, vol. 20, no. 8, pp. e0328290, 2025.

[9] S. Levin, "Using Machine Learning for The Optimisation of Operations and Management in Electric Systems and Networks," in *E3S Web of Conf.*, vol. 524, pp. 01010, 2024.

[10] H. L. Tong, H. Ng and H. Arul Ananthan, "Predicting Diabetes Mellitus with Machine Learning Techniques," *J. Eng. Technol. and Appl. Phys.*, vol. 6, no. 1, pp. 91–99, 2024.

[11] K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia and A. J. Tallón-Ballesteros, "The Impact of Data

Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis," *Proc. Int. Conf. Soft Comput. Models in Industr. and Environ. Appl.*, pp. 344–353, 2023.

[12] C. M. M. Mansoor, S. K. Chettri and H. M. M. Naleer, "Tuning Hyper Parameters for Predicting Coronary Heart Disease Based on Grid Search Algorithm," in *2025 6th Int. Conf. for Emerg. Technol.*, Belgaum, India, pp. 1–6. 2025.

[13] M. Wadinger and M. Kvasnica, "Real-Time Outlier Detection with Dynamic Process Limits," in *2023 24th Int. Conf. Process Contr.*, pp. 138–143, 2023.

[14] M. W. Rodrigues and L. E. Zárate, "A Multivariate Method for Detecting and Characterizing The Changes in Responses of Sensors When Extreme Outliers Arise," *Eng. Appl. Artif. Intell.*, vol. 133, pp. 108424, 2024.