JOURNAL OF ENVIRONMENTAL AND ENERGY SCIENCE

ARTICLE
vol. 1-2 (2023), 4-12
DOI:10.5281/zenodo.8164764

Enhanced Power Demand Forecasting Accuracy in Heavy Industries Using Regression Learner – based Approched Machine Learning Model

Jayme de Porcento Y. Mendaros [1], Renante M. Moral [1], Edgar P. Cortes [1,*]

[1] College of Engineering; Cebu Technological University, Cebu City, Philippines

**** Corresponding author: cortesed2004@gmail.com***

OPEN ACCESS ARTICLE

FULL LENGTH ARTICLE

Received: 24 May 2023

Revised: 30 June 2023

Accepted: 9 July 2023

Published online: 16 July 2023

Published in issue: 16 July 2023

Abstract

For effective management of power systems in heavy industries, accurate power demand forecasting is essential. Traditional statistical models have been tried for this goal, but they frequently struggle to capture the intricate patterns and connections in the data. This paper proposes a method for predicting these power demand, it involves preparing the baseline data, training a surrogate model using a machine learning algorithm, and performing cross-validation to evaluate the performance of the model. To address the diversity in load behaviors and demand spike patterns, a statistical analysis-based machine learning algorithm selection approach is proposed to guide the accurate development of the surrogate model. This study provides a comprehensive framework for predicting the power demand, selecting appropriate machine learning algorithms, and avoiding overestimation. Results enable management to make better decisions, optimize energy usage, and reduce costs and avoid penalties, and surcharges.

Keywords: power demand, forecasting, heavy industries; load; power generation; load; forecasted demand

1. Introduction

Power is essential to the economy, in the most electrical grid that spans large geographical areas. It is typically made up of a few large networks and within those electrical networks electricity flows freely among secondary distribution system of many electrical systems directly to various customers [6].

In the process of bringing power from generation delivery points to the customers and transforming the same to a lower voltage that supplies all the loads, particularly in heavy industries, it is important to identify the exact contacted demand to avoid penalties and charges for instances of exceeding actual consumption versus actual and overspending for extremely high contracted demand under Philippines energy framework system which both scenarios are expenses to the organization.

Accurate load forecasting is crucial for efficient energy management since heavy businesses use a lot of energy to run their activities [4][7]. In load forecasting, the future power demand is predicted using historical data and other pertinent variables. Complex temporal correlations and patterns in the data are difficult to capture using conventional forecasting techniques, such as statistical models. Machine learning models, such as the Regression – learner neural network, have demonstrated tremendous potential in recent years for enhancing load forecasting accuracy in a variety of domains.

In this study, we propose an innovative Regression – learner neural network model-based method for load forecasting in heavy industries. To increase the precision of load forecasting, we offer a model that considers pertinent factors such as kilowatts, kVA, kVAR, power factor, and the maximum demand spike each day. The proposed Regression Learner model has the capacity to capture long-term dependencies and patterns in the data, which is significant in the context of heavy industries where the energy consumption is influenced by various factors such as production schedules, weather conditions, and modifications in industrial processes. Below is the load profile of the plant as baseline data.

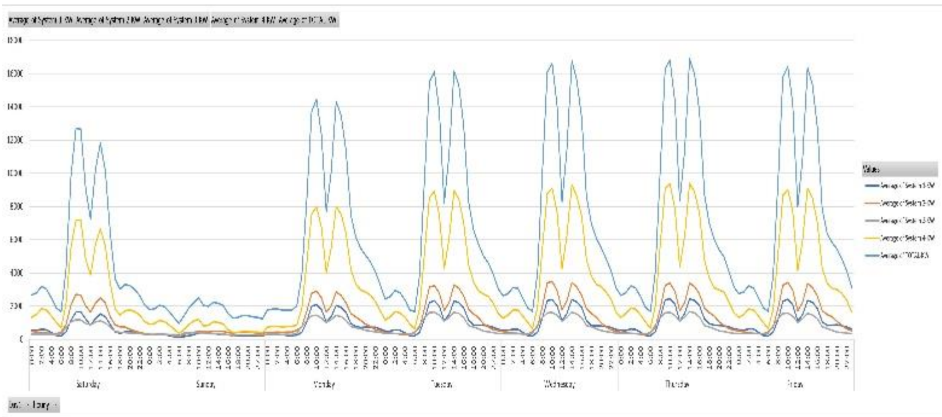


Fig. 1. – Plant historical data

ISSN: 2956-7416

The primary goal of this study is to show how well the suggested Regression Learner neural network model performs load forecasting in heavy industries. The historical load data was initially preprocessed with the pertinent parameters included. The Regression Learner model was then trained and tested using the preprocessed data, and its performance was compared to that of conventional forecasting techniques.

2. Regression Learner Model

Regression models used to predict data can be easily created with the help of the Regression Learner app. Its user-friendly interface enables simple data exploration, feature selection, specification of validation schemes, model training, and model evaluation [2]. Additionally, it offers automated training options for choosing the best regression model type, including support vector machines, regression trees, Gaussian process regression, linear regression, kernel approximation models, ensembles of regression trees, and neural network regression models [2],[5].

The application uses supervised machine learning to train models, which entails using a known set of observations of input data (predictors) and known responses to teach the model how to produce expected responses for new input data.

If the predictions are done with only one single variable, then it is treated as simple linear regression whose expression is given below which is also called as Hypothesis equation and the same analysis is presented in this paper [10].

$$Y=a+bX+\varepsilon \qquad (1)$$

Where, Y is the response or output or dependent variable, a is the intercept, b is slope of linear regression line, X is independent variable and ε is the error or residual of model.

Similarly, if the same predictions are carried out for more than one variable, then it is referred as multiple linear regressions and the expression goes as follows [10]:

$$Y=a+bX1+cX2+dX3+\ldots\ldots\ldots.+\varepsilon \qquad (2)$$

Where, Y is the response, X1, X2, X3 and b, c, d are independent variables and their slope respectively as it has multiple regression lines, a is the intercept and ε is sum of residual errors calculated for all regression lines. The most common factor in both simple and multiple linear regression lines is error 'ε'. The error should be minimum as such as it can, as it may result to better accurate model [10].

Certain mathematical methods are adopted to reduce the error. Some of the techniques include Root Mean Squared Error (RMSE), Minimum Squared Error (MSE), Minimum Absolute Error (MAE), R squared, Ordinary least squares method, Sum of absolute errors, Gradient descent method. Out of all these methods, the most common and comfortable method is RMSE method which is the root of squares of difference between predicted and true values of a model and the same technique has been carried over in this paper. However, the equations for calculating errors in different methods and their formulae is as listed below [1],[3].

$$RMSE = \sqrt{\frac{1}{n}\ \sum_{i=1}^{N}(Ypred - Ytrue)^2} \qquad (3)$$

Where, N is the number of observations or iterations to calculate error, Ypred is the predicted values of dependent values and Ytrue or actual

ISSN: 2956-7416

value. True values are the values which are fed as input to trained model and predicted values are the values obtained after performing LR analysis. Mean Absolute Error (MAE) is expressed as the difference between predicted and true responses and Mean Squared Error (MSE) is defined as the squares of difference between predicted and true responses which are given below [3].

$$MAE = \frac{1}{n}\sum_{i=1}^{N}(Ypred - Ytrue) \quad (4)$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{N}(Ypred - Ytrue)^2 \quad (5)$$

Regression Learner is a powerful tool that can be used to create regression models, such as linear regression models, regression trees, Gaussian process regression models, support vector machines, kernel approximation, ensembles of regression trees, and neural network regression models. It not only allows you to train models but also provides the ability to explore your data, select features, specify validation schemes, and evaluate results. You can export a model to the workspace to use it with new data or generate MATLAB® code for programmatic regression [5],[8].

3. Methodology

The process of training a model in Regression Learner can be divided into two parts: Validated Model and Full Model. The Validated Model trains a model with a validation scheme that protects against overfitting by applying cross-validation. Alternatively, you can choose holdout validation. The validated model is visible in the app. The Full Model trains a model on the entire dataset, excluding the test data. The app trains this model simultaneously with the validated model. However, the model trained on full data is not visible in the app. When you choose a regression model to export to the workspace, Regression Learner exports the full model [3].

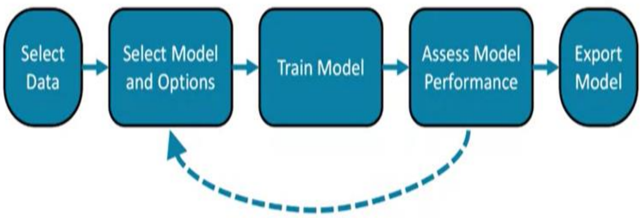
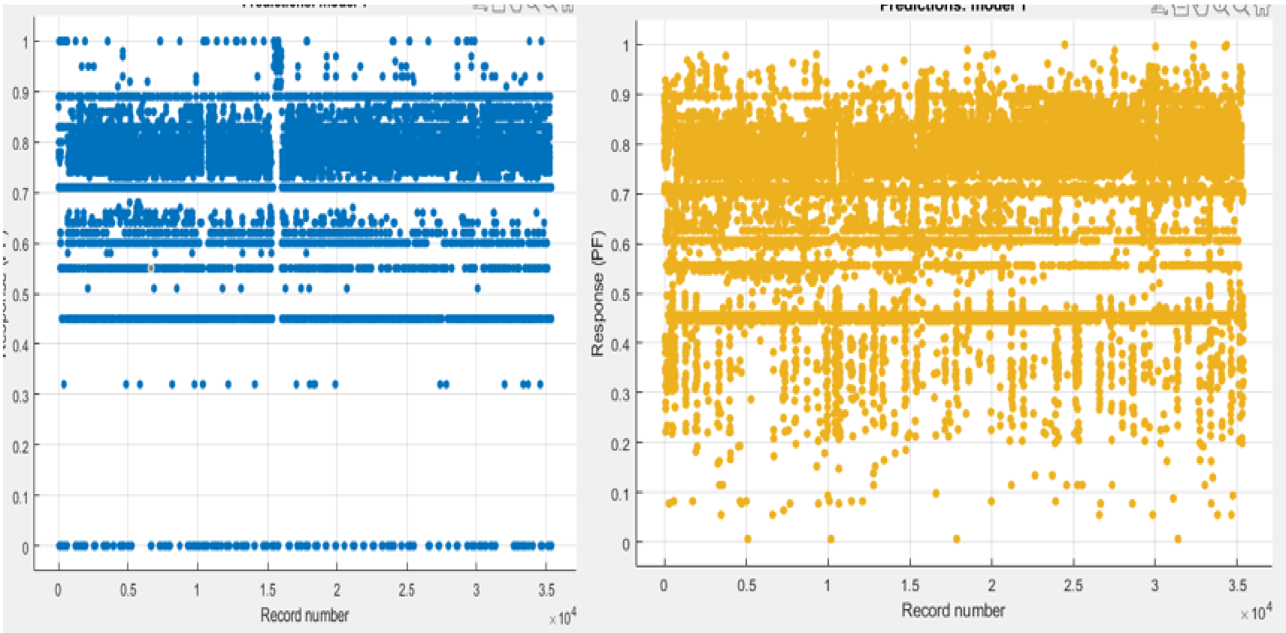


Fig. 2. Regression learner work flow

Collect historical power demand data, including kilowatts (kW), reactive power, apparent power, power factor, and maximum demand spike in a day. Additional relevant data such as weather data, economic indicators, and industrial activity data may also be collected. Second, is to clean and preprocess the data by handling missing values, outliers, and other data quality issues. Normalize the data to improve model performance. Third, select the relevant features for the Regression Learner model based on their importance in predicting power demand. Fourth, train the model using the preprocessed data. The model may be trained using various hyperparameters such as the number of layers, the number of neurons per layer, the learning rate, and the number of epochs. Fifth, validate the model using a holdout set of data or cross-validation. Evaluate the model's performance using various metrics such as mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), including fine-tune the model by adjusting the hyperparameters based on the validation results [2], [3].

4. Power Load Forecasting Based on Regression Learner Model

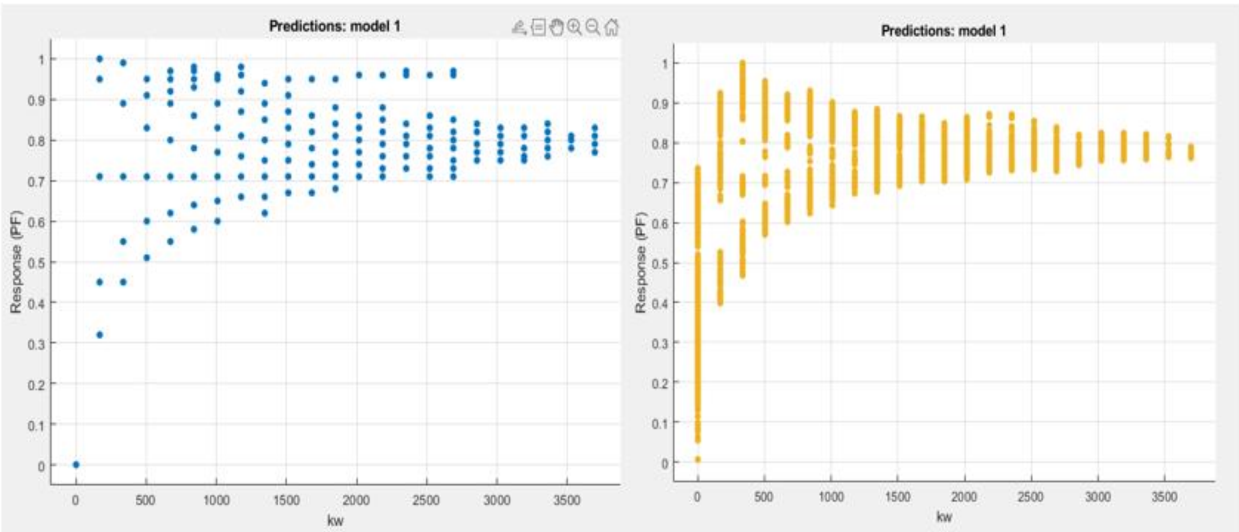
The baseline data are extracted in the smart metering from the previous year, after the extraction and validation, important parameters are set to be considered to a trained variable. These are actual kilowatt (kW), kilowatt – hour (kWh), the reactive power in kVAR, reactive power - hours in kVARh, apparent power in kVA and apparent power – hours in kVAh.



Baseline data Trained data

Fig. 3. Combined graph of the baseline & trained data

Individual parameters considered in this study graphically presented to provide an overview of the baseline data with the trained data. This graphs comparison presents the similarities and differences between these two sets of data into how the model predict the future load profile of the plant [9].



Baseline power in kW Predicted power in kW

Fig. 4. Power in kW graph of the baseline & trained data

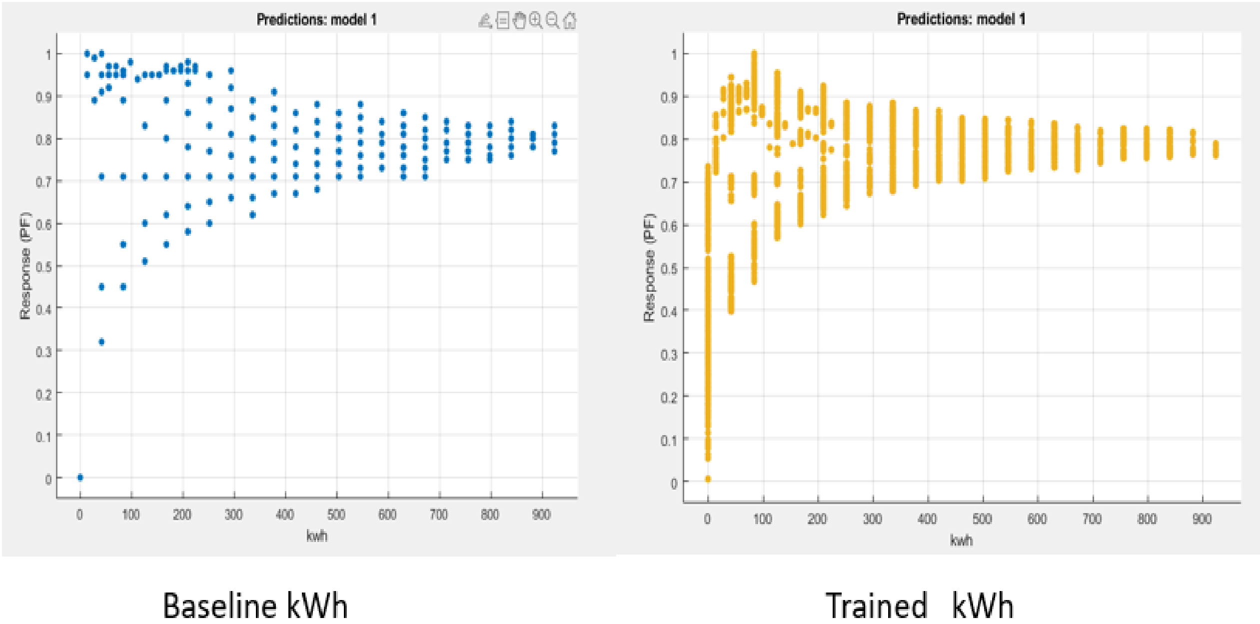


Fig. 5. kWh graph of the baseline & trained data

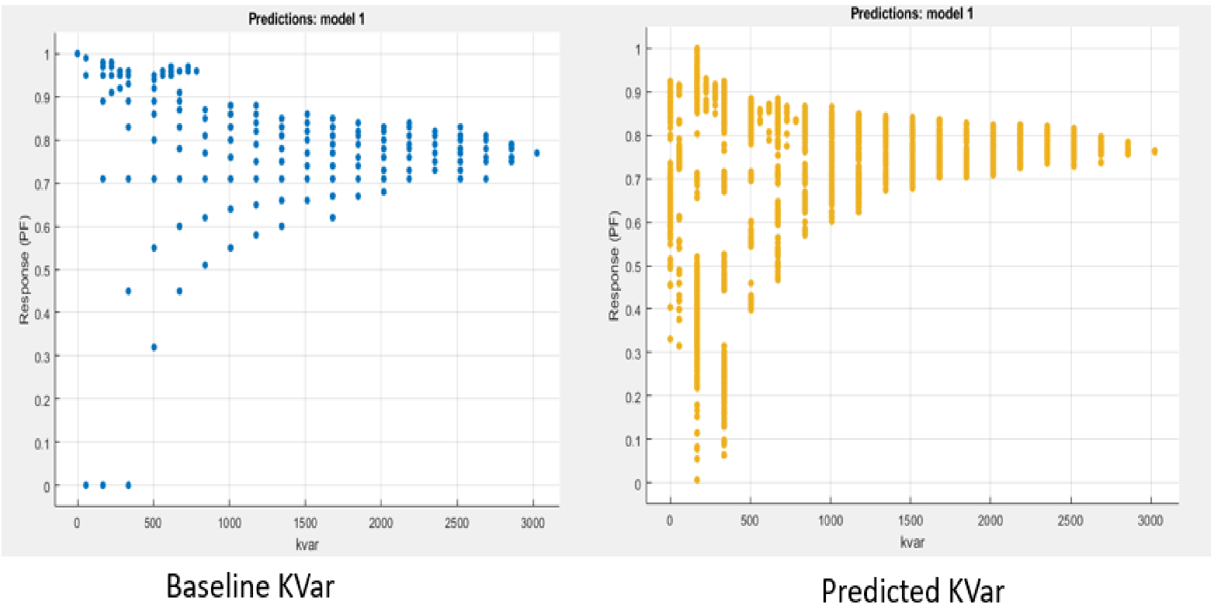
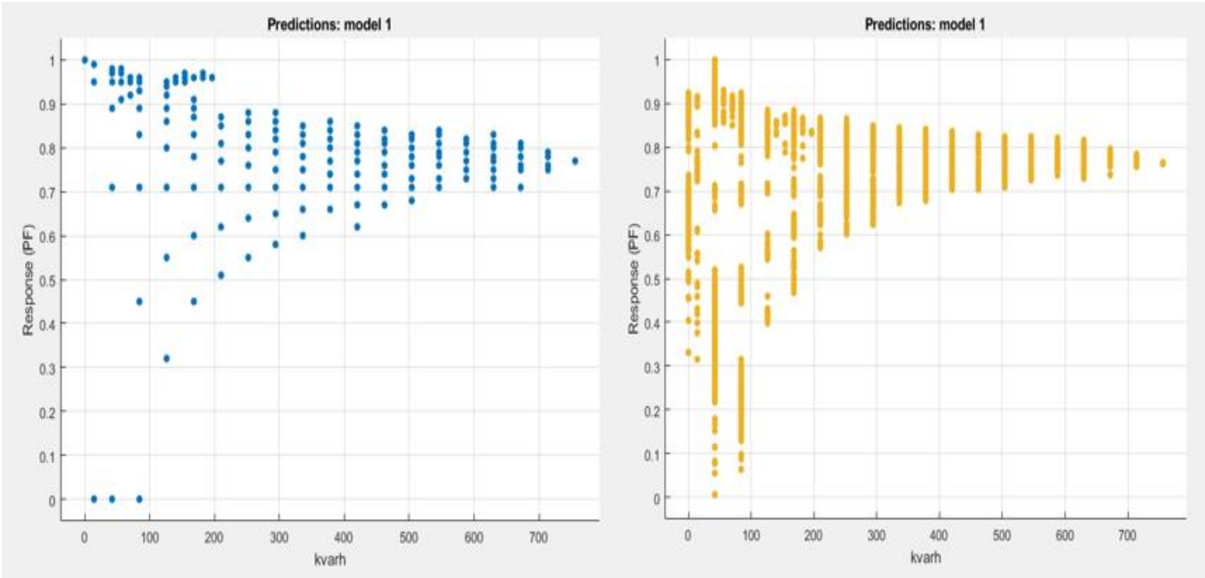


Fig. 6. kVar graph of the baseline & trained data

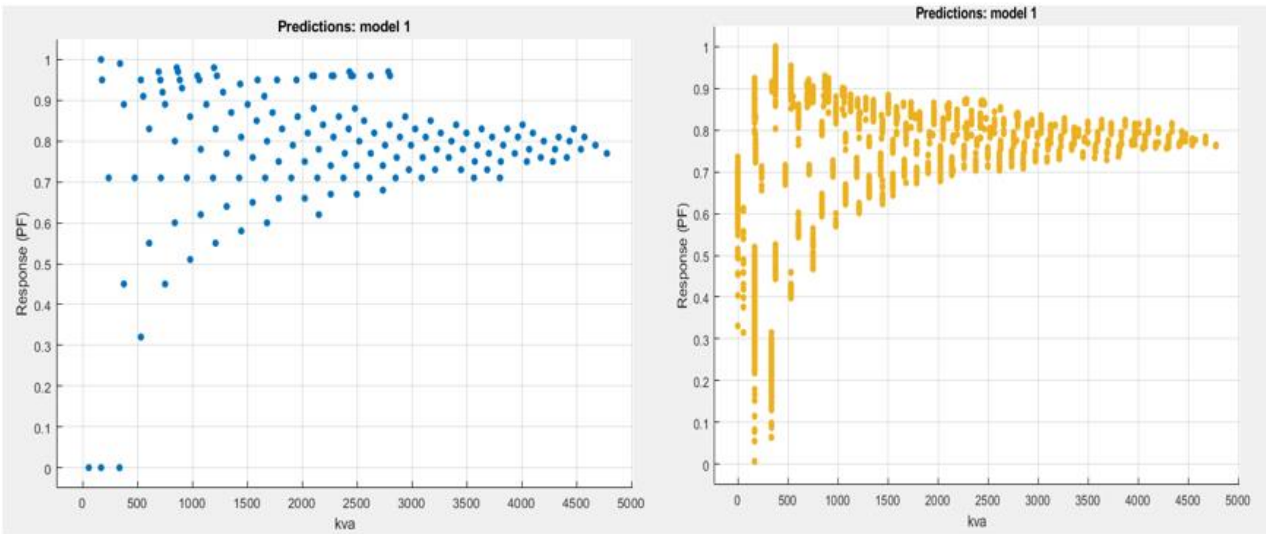
ISSN: 2956-7416



Baseline KVarh

Predicted KVarh

Fig. 7. kVar-h graph of the baseline & trained data



Baseline kVA

Predicted KVA

Fig. 8. kVA graph of the baseline & trained data

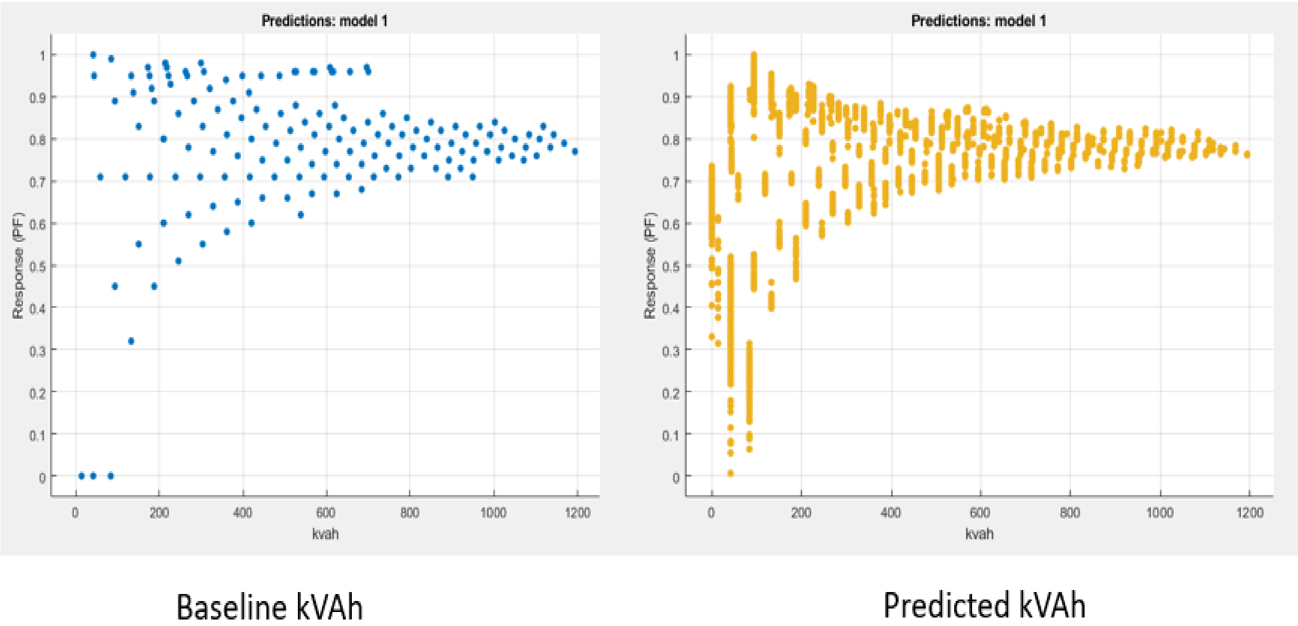


Fig. 9. kVAh graph of the baseline & trained data

5. Analysis and Experimental Results Summary and conclusion

A measure of how well a model fits the data is R-squared. The target variable's variance is 84% explained by the model, according to the R-squared value of 0.84 found in this experiment. MSE is a different way to quantify the discrepancies between the target variable's predicted and actual values. In this experiment, the mean of the squared errors, or MSE, was calculated to be 0.0044997. The target variable's expected and actual values are separated by an absolute amount known as the MAE. The average absolute error of the model's predictions from the actual values is 0.018981 units, according to the MAE value this experiment yielded.

Overall, the model appears to be performing reasonably well with an R-squared value of 0.84 indicating that the model is able to explain most of the variance in the data. The RMSE score of 0.067079, however, indicates that there is still some room for the model's predictions to be improved.

6. Conclusion

The model with an automated box constraint, an epsilon value, and a kernel value of 0.66 performed reasonably well in terms of predicting the target variable. The RMSE value of 0.067079 and the MAE value of 0.018981 show that there is still space for improvement in the model's predictions, but the R-squared value of 0.84 indicates that the model is able to explain the majority of the variance in the data. The experiment's findings show that the model is a viable strategy for power demand forecasting with comparable predictions and better than the traditional methods. With the help of the results, management is able to improve judgments, use energy more efficiently, cut costs, avoid fines and surcharges, and reduce costs through energy usage optimization.

Funding

This research received no external funding.

All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declares no conflict of interest

Permissions and rights

The authors declare that they have all rights and permissions to the information contained in the publication.

References

[1] Y. -J. Lee and H. -J. Choi, "Forecasting Building Electricity Power Consumption Using Deep Learning Approach," 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Korea (South), 2020, pp. 542-544, doi: 10.1109/BigComp48618.2020.000-8.

[2] Elamary, A. S., & Taha, I. B. (2021). Determining the shear capacity of steel beams with corrugated webs by using optimised regression learner techniques. Materials, 14(9), 2364.

[3] Abd Rahaman, N. S., & Jumadi, N. A. (2021). Comparison of Regression Analysis using Origin, Matlab, and Regression Learner App in Formulating the Linear Equation for Leg Flexibility Index. Evolution in Electrical and Electronic Engineering, 2(2), 164-172.

[4] Ekici, S., Unal, F., & Ozleyen, U. (2019). Comparison of different regression models to estimate fault location on hybrid power systems. IET Generation, Transmission & Distribution, 13(20), 4756-4765.

[5] Karamdel, S., Liang, X., Faried, S. O., & Shabbir, M. N. S. K. (2022, December). A Regression Model-Based Short-Term PV Power Generation Forecasting. In 2022 IEEE Electrical Power and Energy Conference (EPEC) (pp. 261-266). IEEE.

[6] Wang, Y., Wang, D., & Tang, Y. (2020). Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods. IEEE Access, 8, 17071-17079.

[7] E. Yuniarti, N. Nurmaini, B. Y. Suprapto and M. Naufal Rachmatullah, "Short Term Electrical Energy Consumption Forecasting using RNN-LSTM," 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), Batam, Indonesia, 2019, pp. 287-292,

[8] P. Virdi, Y. Narayan, P. Kumari and L. Mathew, "Discrete Wavelet Packet based Elbow Movement classification using Fine Gaussian SVM," 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India, 2016, pp. 1-5, doi: 10.1109/ICPEICES.2016.7853657.

[9] P. Kumar and M. Biswas, "SVM with Gaussian kernel-based image spam detection on textual features," 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, India, 2017, pp. 1-6, doi: 10.1109/CIACT.2017.7977283.

[10] Narayan, V., & Daniel, A. K. (2022). Energy Efficient Protocol for Lifetime Prediction of Wireless Sensor Network using Multivariate Polynomial Regression Model. Journal of Scientific & Industrial Research, 81(12), 1297-1309.