Group 2 – Day ahead load forecasting model

# Introduction

In the context of electricity networks, load forecasting emerges as a critical tool used to anticipate future electricity demand, enabling Distribution Network Operator, and consequently, Transmission monopolies to plan their network infrastructure and future connections. This is especially crucial given today’s circumstances with the global shift to green electricity, amplifying the strain on the grid [1]. Forecasting correctly ensures that an appropriate amount of electricity is generated and distributed to homes. Under-generation at substation sites could cause power outages, blackouts, and hence safety concerns.

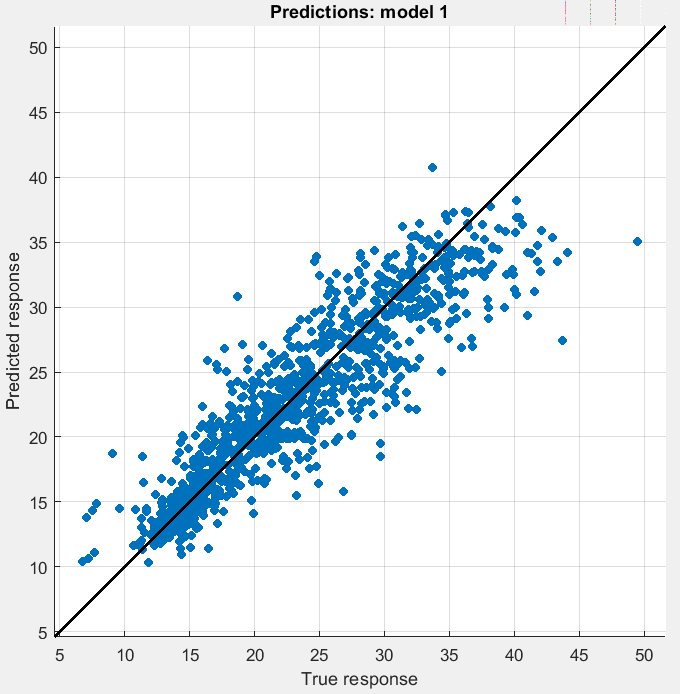
## Forecast Model Design

The proposed models follow a time-series approach and use half-hourly (HH) transformer data on feeder 2 in July, combined with lag data and weather data, to predict a load forecast in August. The actual August load data is used in comparison with the model’s forecast and generates error statistics used to gauge the success of the forecast.

A strategically selected lagged data profile and weather observations served as predictors for the models. The methodology involved analysis of Pearson’s correlation coefficient, mean, and median where appropriate variables were chosen as predictors for the model. Visual inspection of Matlab’s RRelief algorithm plotting tool gave further reassurance of this selection consisting of: Lag 48HH, 72HH, 96HH, 336HH, GSI, air temperature, and wet bulb. Since the model predicts a day-ahead forecast, then the horizon for the lagged data is 48 HH (1 day) as trends throughout the day will not be captured otherwise. This horizon states the smallest lag value that can be used.

Recognised for its simplicity and effectiveness, the team opted for the Linear Regression variation for all 3 models. Linear regression learns the best-fitting linear relationship between the input and target output data and forecasts data using this relationship.

## Results and Discussion

The performance of the load forecasting models were gauged by R-squared, RMSE, MSE, and MAE. R-squared gives an indication of how much an independent variable influences the change on the dependant variable. RMSE, MSE, and MAE, highlight the quantity in which the forecasted data deviates from the true data [2]. The aforementioned array of error metrics is used to adequately capture the performance of each model and is displayed in Table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | R-Squared | RMSE | MSE | MAE |
| 1 | 0.83 | 3.1602 | 9.9869 | 2.2596 |
| 2 | 0.81 | 0.1621 | 0.0263 | 0.1195 |
| 3 | 0.84 | 0.1616 | 0.0261 | 0.1216 |

**Table 1. Model performance metrics**

**Model 1:** The benchmark model was created by using the Linear Regression Learner and feeding the strategically selected lagged load and weather data. The results, depicted in Figure 1, exhibit a vague linear relationship between the forecasted and true response, given by a R-squared of 0.83. Large RMSE, MSE, and MAE values indicate poor predictive performance.

Figure 1: Benchmark model

**A graph with blue dots

Description automatically generatedModel 2:** To improve Model 1, pre-processing of the load data for the train and test data set was executed to create more normally distributed data. Preprocessing involved extracting a log of the data which normalises the mean for test and train data. This is done due to linear models performing better with normal distributed data [3]. This greatly decreased errors however furthered the problem of outliers as seen in Figure 2.

A graph with blue dots

Description automatically generated**Model 3:** Model 3 introduced the notion of removing outliers at the pre-processing phase. This was accomplished by evaluating the train and test data and omitting datapoints that exceeded 3 standard deviations of the mean resulting in cleaner, less noisy data sets. The outcome of this approach has further refined the model’s performance, improving R-squared, RMSE, and MSE metrics.

Figure 2: Pre-processed data approach

Feasibility of model 1 is hindered by its poor RMSE, MSE, and MAE values indicating the model predictions are far from the actual load values. However, its high R-squared value hints that most of the load fluctuations and trends are captured. Model 2 demonstrates opposite traits to 1 and hence is also deemed not feasible. Given their error margins, use of these models in the real world could result in poor forecasting, potentially resulting in unexpected peak loads which could cause unexpected blackouts and power outages due to the power network not having enough capacity. Model 3 strikes a balance between error metrics establishing itself as the most feasible. Figure 3 suggests that this model is great at predicting lower loads, given by the densely populated points around the line, but poor with higher loads since the points are scattered.

Figure 3: Removing outliers and then pre-processing data

## Conclusion

In conclusion, the escalating demand for effective load forecast models are more than ever given the current landscape of increasing demand and global shift towards green renewable energy. A forecast model design journey was explored where forecasting model parameters were ascertained by different statistical analysis methods. The advantages and consequences of poor and good forecasting were extensively examined in the analysis of the 3 models. Regression learner error metrics and their impact on all 3 forecasting models were also highlighted in the analysis before considering the feasibility of these models in the real world.

References

[1] IET, "Forecasting the load of electrical power systems in mid- and long-term horizons," 01 Decemeber 2016. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-gtd.2016.0340. [Accessed 22 11 2023].

[2] Y. Yoshikawa, "Gaussian Process Regression With Interpretable Sample-Wise Feature Weights", IEEE Explore, 10 December 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9646444. [Accessed 27 November 2023].

