

### Summary on Assignment 3

This part asks to use forecast model to predict RTLMP along with exploratory data analysis. Plots of dependent variable against different independent ones were created to provide a general sense of potential relationships. Some preprocessing methods including data imputation, outliers' removal were also applied to gain a better representation of the data. After that, a multivariate time series forecasting LSTM model was used to perform predictions. The data frame was divided into a train and test set so that the predictability of the model could be evaluated. Since RTLMP is the dependent variable and other features as the independent variables, the LSTM model aimed to use data from independent variable to first learn their relationships to RTLMP before predicting. The LSTM model requires 3-dimensional data (number of samples, lookback period, number of features). In this case, a 5-day or 120-hour look back period was used. The model is also capable of providing one day or multiple days prediction. In this case, a one-day prediction was used. In essence, the model looks at 120-hour observations from the independent variables and a one-day ahead observation of RTLMP to learn the relationships. The loss function for the model was `mean_absolute_error`. The model's prediction achieved such error of 11.9 after running on the test split.

While the model managed to follow part of the trends, there were jumps of prices the mode failed to predict, which was expected since it is hard to closely follow the exact movement of a time series. Further improvement of the model could be more feature engineering, hyperparameter tuning for the model, and adjustments of the lookback periods. However, the learning rate for the model can sometimes be too large that the gradient explodes, and the loss becomes nan in the training process as lookback period increases. More detailed implementations are in the Jupyter notebook.