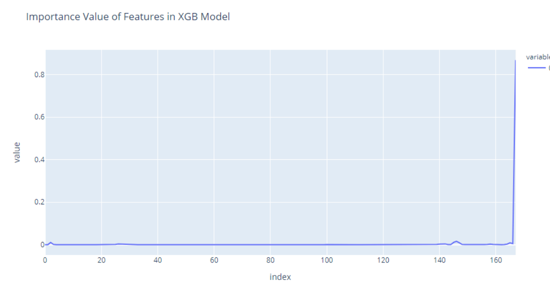
MISM6203 Project 2B

Michael Conan

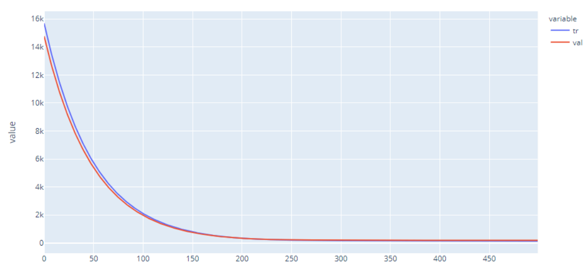
Energy Usage Modeling Analysis Part 2

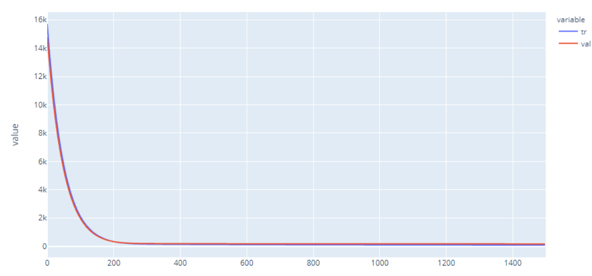
Similar to the penalized linear regression developed for the previous forecasting model, the extreme gradient boosting (XGBoost) model trained uses weekly intervals as the time series to train the predictive model. This was selected based on the comparison of daily, weekly, and seasonal fluctuations observed in the energy consumption data.

The XGBoost predicts energy usage for the future period by applying a set of decision tree regression models to the input data, which in this case is a series of 167 datapoints of hourly sequential energy usage (7 days of 24 hours). The model is created by optimizing a number of decision trees and mathematically averaging the results to generate the most accurate prediction. Based on a review of the most important features in the trained model, the optimized model relies very heavily on the most recent recording in order to predict the next value.

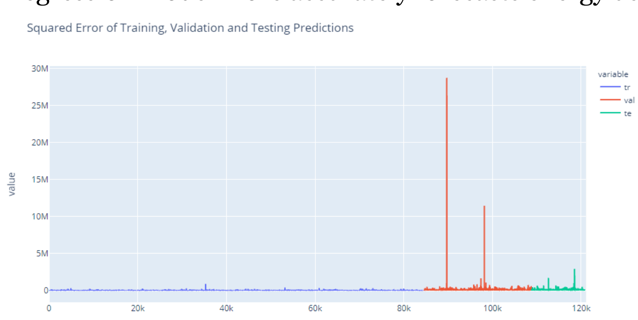


The final model was trained using decision trees with a depth of 8 layers, trained at a learning rate of 2% and incorporating an alpha regularization parameter of 2. After testing multiple potential parameters with 500 iterations, the final model was trained with 1500 to ensure optimal results, which reduced the RMSE from 185 to 171 on the validation data.





Based on the r-squared value, a common measure of the accuracy of the model, the Penalized Linear Regression model more accurately forecasts energy usage compared to the XGBoost. The optimal XGBoost had r-squared values of 0.9953 and 0.9958 on the validation and test data, respectively, while the PLR had 0.9963 and 0.9968, respectively. Interestingly, the PLR r-squared actually improved from the training data to the validation and testing, while the XGBoost accuracy decreased notably. This suggests that the XGBoost model may have overfit the training data and the PLR was more generalizable.



While the PLR appears more generalizable, the model appears to hold up over time compared to the PLR, which clearly broke down when the model was applied for a year of forecasting. Based on this observation, the XGBoost should be used for any forecasting performed for longer periods of time.