XG BOOST – IMPLEMENTATION AND RESULTS

# Introduction:

**XGBoost** is an optimized distributed gradient boosting library designed to be highly **efficient, flexible**and**portable**. It implements machine learning algorithms under the [**Gradient Boosting**](https://en.wikipedia.org/wiki/Gradient_boosting) framework. Execution Speed and Model performance are two advantages of this algorithm. XGBoost algorithm library implements **gradient boosting** decision tree algorithm.

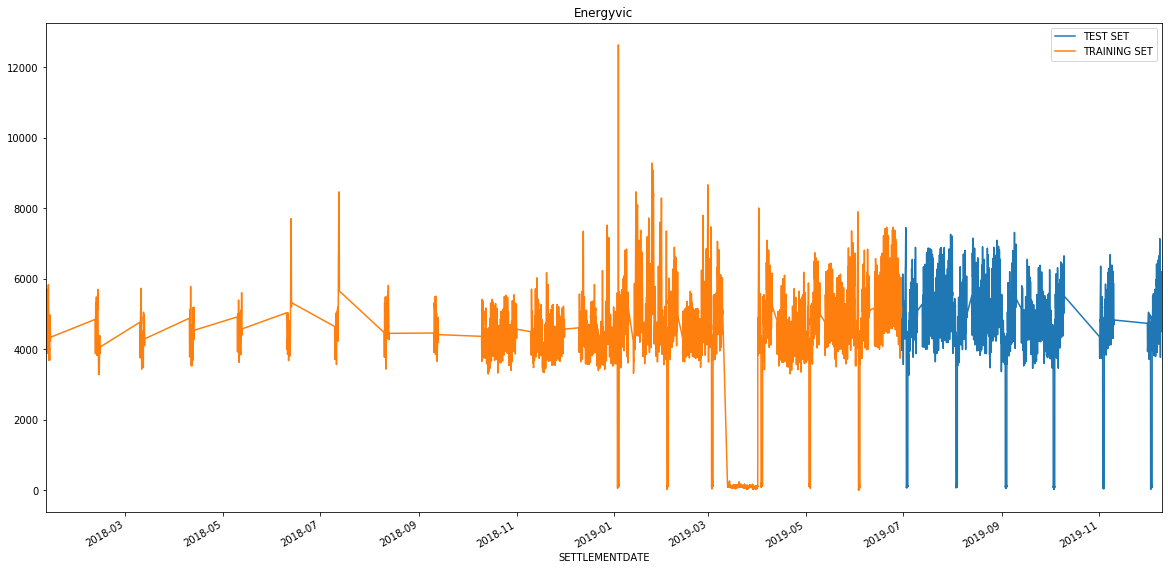
**Gradient boosting** is an approach where new models are created that predict the residuals or errors of prior models and then added together to make the final prediction. It is called gradient boosting because it uses a gradient descent algorithm to minimize the loss when adding new models.

# Dataset:

I have used 12 months data from October 2018 – September 2019 of Victoria electricity Demand. The data set contains 2 variables, Settlement date and Energy demand. Data Shape (17520 , 1).

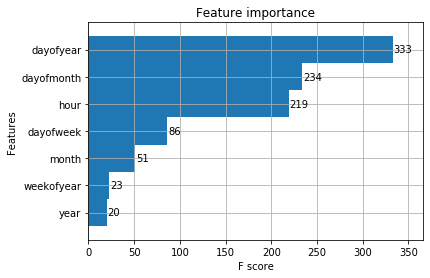
Train data – October 2018 to June 2019

Testing data – July 2019 to September 2019



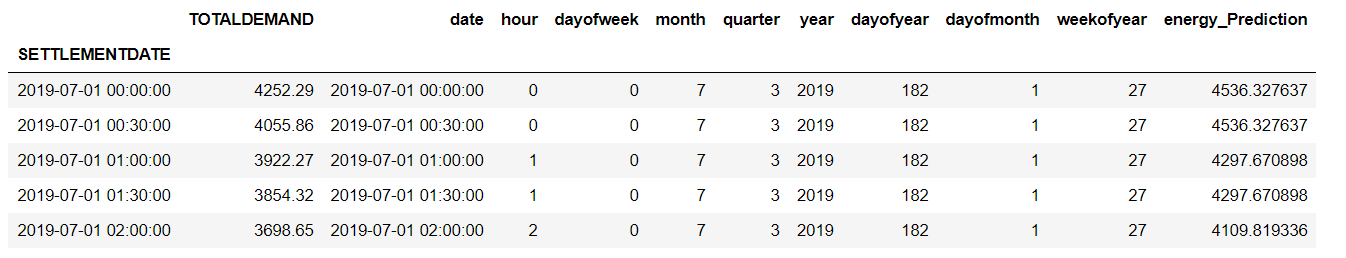
# **Feature Importances:**

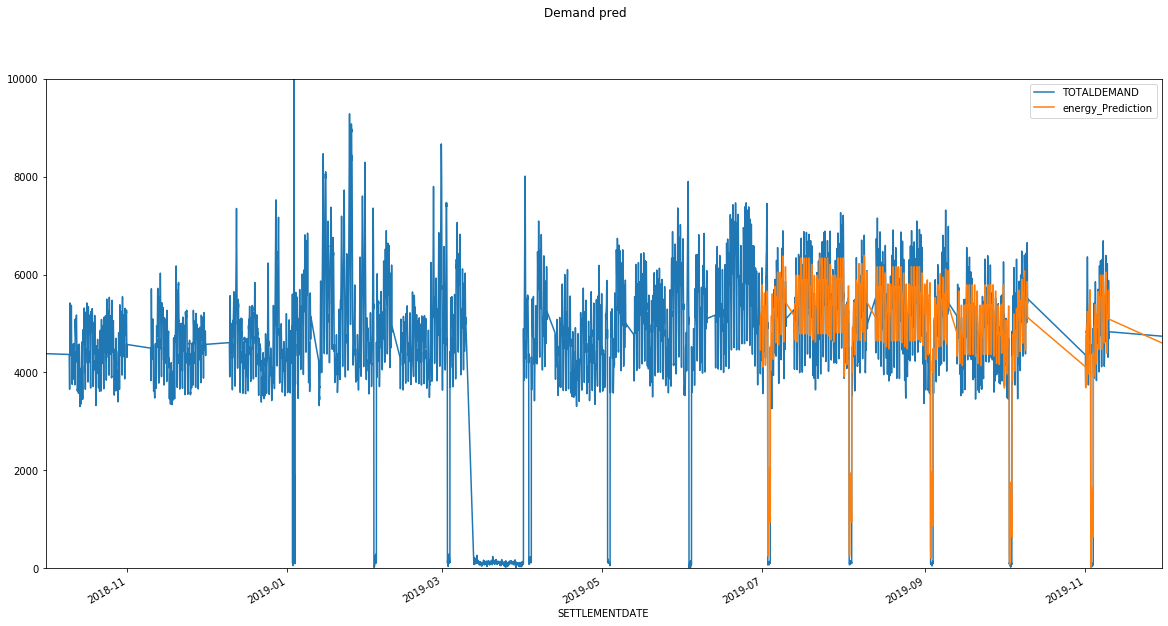
Feature importance is a great way to get a general idea about which features the model is relying on most to make the prediction. This is a metric that simply sums up how many times each feature is split on.



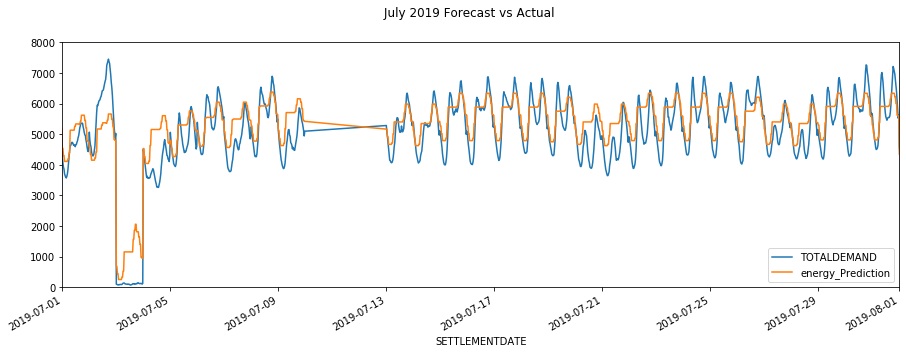
We can see that the day of year was most commonly used to split trees, while dayofmonth and hour came in next.

**Forecast on Test data:**

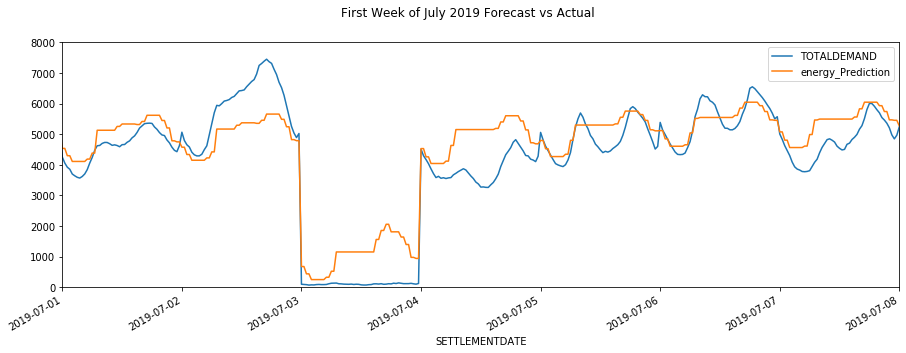




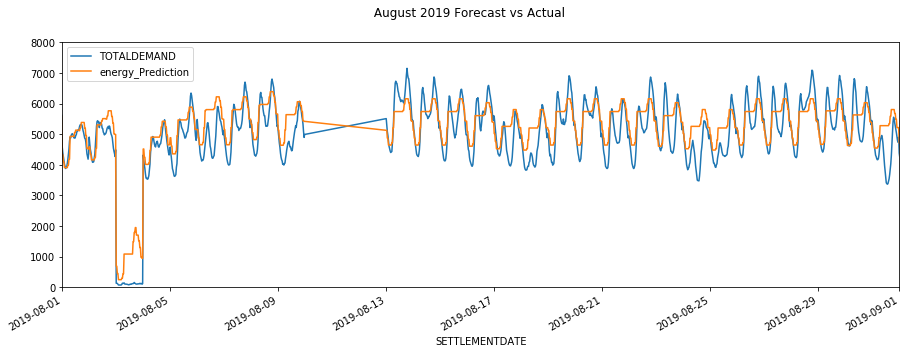
**Forecast on July month:**



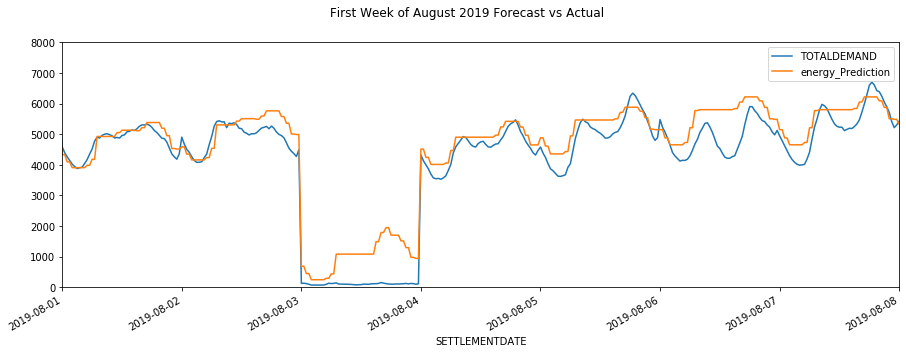
**Forecast 1st week of July:**



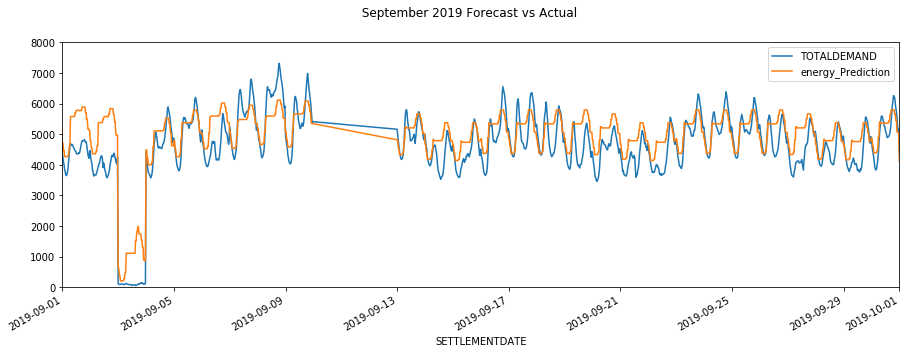
**Forecast on August Month:**



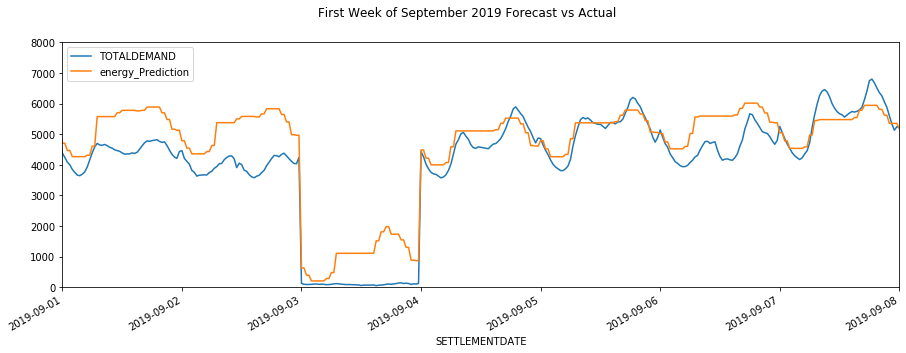
**Forecast on August 1st week:**



**Forecast on September month:**



**Forecast on September 1st week:**



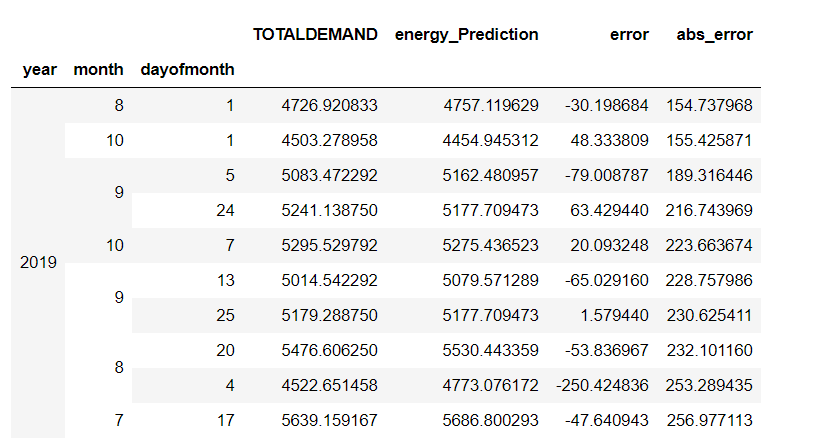
**An Analysis on Best and Worst days forecasted:**

**Best days forecasted:**

Number 1 best day, August 1st, 2019

Number 2 worst day, October 1st, 2019

Number 3 worst day, September 5th, 2019

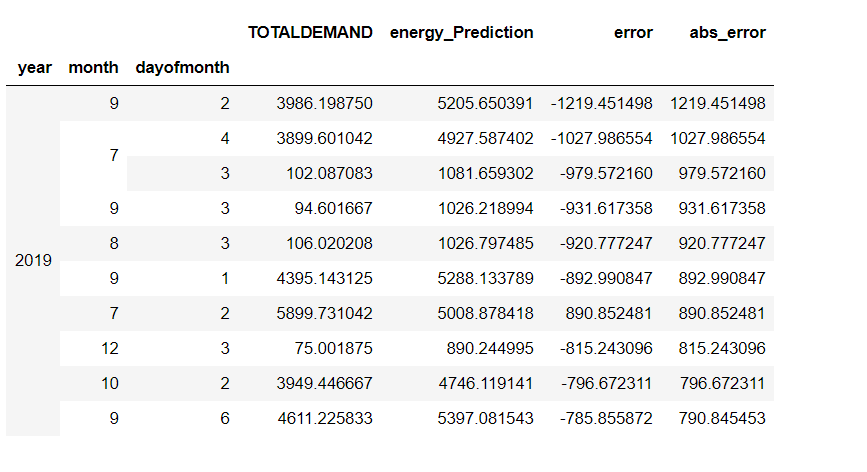


**Worst days forecasted:**

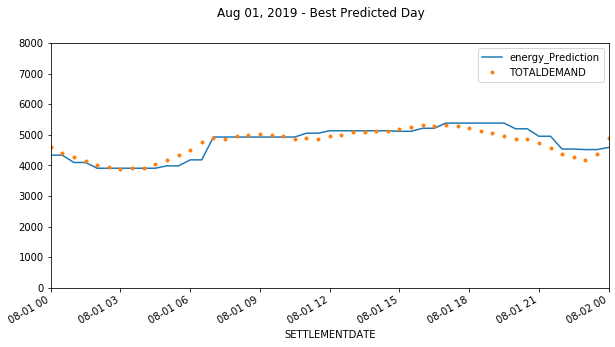
Number 1 worst day, September 2nd, 2019

Number 2 worst day, July 4th, 2019

Number 3 worst day, July 3rd, 2019



**Plotting best predicted day – August 1st, 2019:**



**Plotting worst predicted day – September 2nd, 2019**

