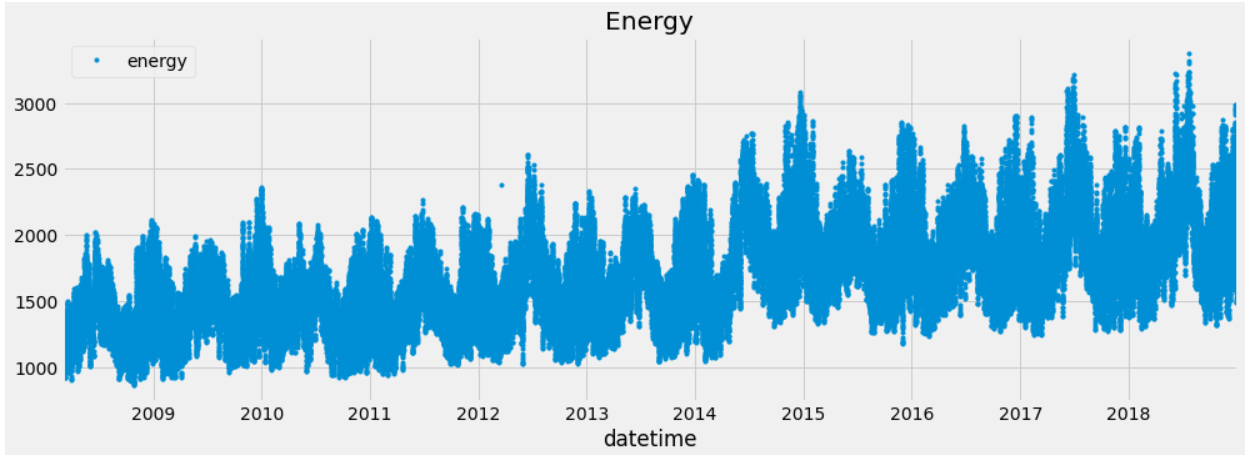
**Approach**

Participant Name: Zafar Mahmood Waris

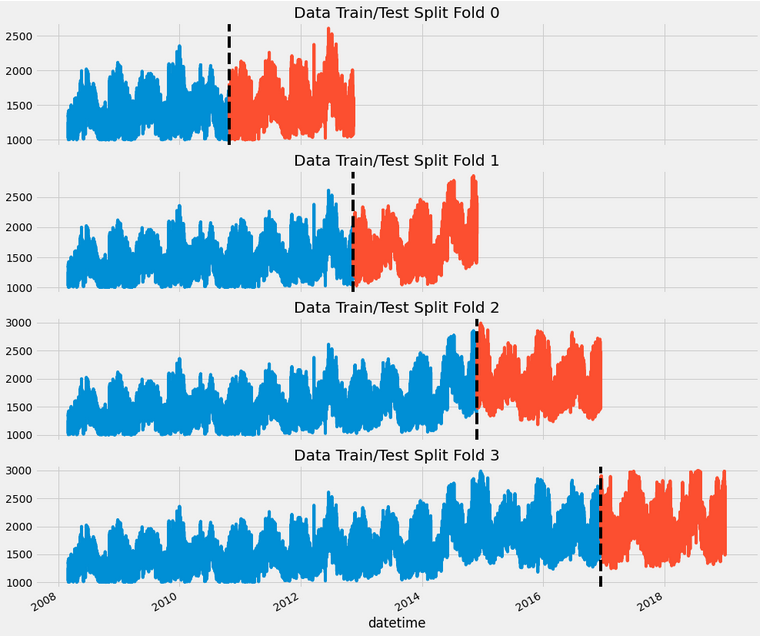
Xgboost is the one I used!!!

* This problem statement deals with future forecasting of energy requirement for a time period of 3 years with 9 years of available data.
* As there are no null values so we can start working on the data without the hassle of missing values handling.
* This time series data has hourly values present from 2008 to 2018.



**Feature Engineering:**

* To predict the value of energy requirement in the future for a 3 years horizon we need enough independent features.
* So for that use pandas to create features from the ‘datetime’ column and added 8 lag values in the features.
* After adding these columns we split the dataset into features and target.
* For splitting data we used TimeSeriesSplit from sci-kit learn which is used to split a time series data as it is a sequential data.



**Modelling:**

* For modelling I used:
  + Random Forest
  + XGBoost
* The best result was derived using XGBoost.
* As the data is time series XGBoost best predicted the future values with less error.

**Could have tried:**

* Ensemble of different models.

Link to the notebook:

<https://github.com/zmwaris1/Analytics_Vidhya_Job_A_Thon>

Github ID: Portfolio:

<https://github.com/zmwaris1> <https://zmwaris1.github.io/>