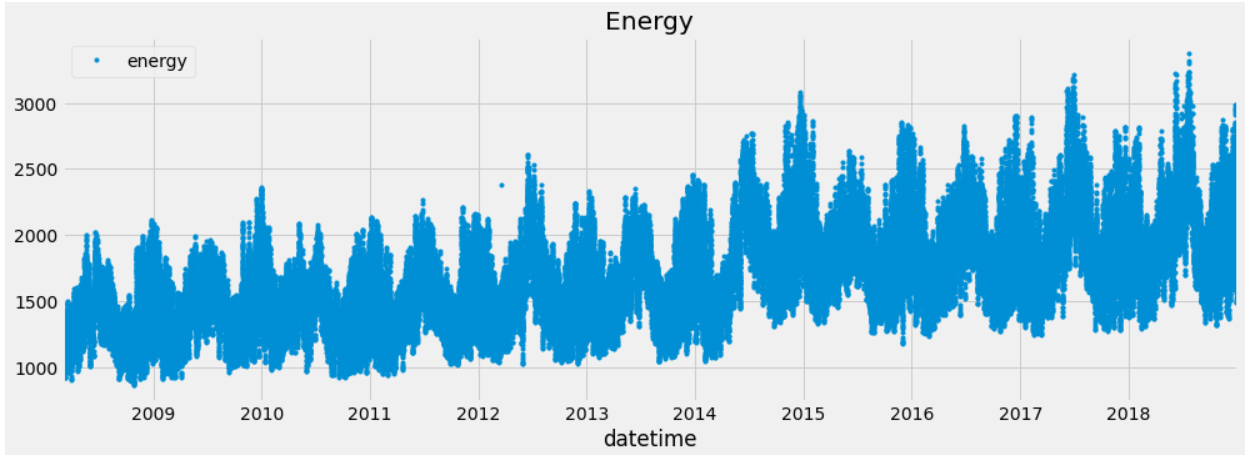
**Approach**

**Participant Name**: *Prashant Shukla*

* This problem statement deals with future forecasting of energy requirement for a time period of 3 years with 9 years of available data.
* 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.
* We have to **split** the **datetime** data type into separate features named as **date, month, year and hour.**

**Data Cleaning:-**

* As there are null values so we have to replace the null values by suitable techniques. I have tried **Iterative, Simple & KNN Imputer**. Out of these, **Iterative** Imputer has given **best** results.

**Train-Test Split:-**

* Generally we used **80:20** ratio for train-test split. I have tried more than 10 combinations, in which **test ratio** having value **0.155** has given the **best** results.

**Modelling:**

* For modelling I have performed:-
  + Linear Regression
  + Random Forest
  + CatBoost
  + XGBoost
* The best result was derived using XGBoost D-Matrix(used to give best iteration) having best **hyperparameters**, with same number of **iterations** with least **RMSE**.
* As the data is **time series**, & problem is **regression** type, XGBoost was the best to predict the future values with least error.

**Possible improvements:**

* Ensemble of different models.
* ANN
* LSTM

**Link to the notebook:** https://github.com/Prashant2091/energy\_future\_values\_prediction

*Github ID:*  *Kaggle:*

https://github.com/Prashant2091 [**https://www.kaggle.com/prashantshukla91**](https://www.kaggle.com/prashantshukla91)

*Personal UI’s link:*

1. <https://prashant2091-diff-disease-cat-ml-diff-cat-oys6dm.streamlit.app/>

2. <https://prashant2091-cdacproject-app-dkfex0.streamlit.app/>

3. <https://prashant2091-texttospeech-app-t9dlq6.streamlit.app/>