PREPOCESSING

Steps used in Preprocessing (Ref)

Primary Data source: https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset

- Visualized the Data with Table

- Found that our Primary Dataset from Kaggle had 3 Independent variables

  Namely: Wind Speed, Wind Direction, Theoretical Power

- The Theoretical power was the least important logically

- Wind Speed the most affecting as it seems it is directly proportional.

- There were 3 data ranges missing form the data set

  MISSING DATA POINTS

  # 2018-01-26 06:20:00 to 2018-01-30 14:40:00

  # 2018-09-28 21:20:00 to 2018-10-02 16:30:00

  # 2018-11-10 21:10:00 to 2018-11-14 12:00:00

- We had data point values at an interval of 10 Minutes

- Training had to be done on the hourly based data points.

- Filled the NA values with avg