

Forecast Green Energy -Approach

Objective: Your task at hand is to build a machine learning/deep learning approach to forecast the total energy demand on an hourly basis for the next 3 years based on past trends.

Our Approach will be as follows:

Given Dataset is consists of Row_id, Date-time, energy columns.

As Energy variable varies according to date-time given so we came to a conclusion here that Energy is a dependent variable and Date-time & row_ids are independent variables.

We will first see a graph of energy consumption on hourly bases of given data and then will see the graph between predicted Energy values in test data on hourly(given) bases.

Exploratory analysis:

- Finding datatypes of given datasets
- Finding null values there were ≈ 1600 null values.
- Dropping / filling the places where null values exist in datasets: I chose to drop null values as rest of the data is quite huge and variables are also two.

Bivariate Analysis

- using lineplot, jointplot

Model Building:

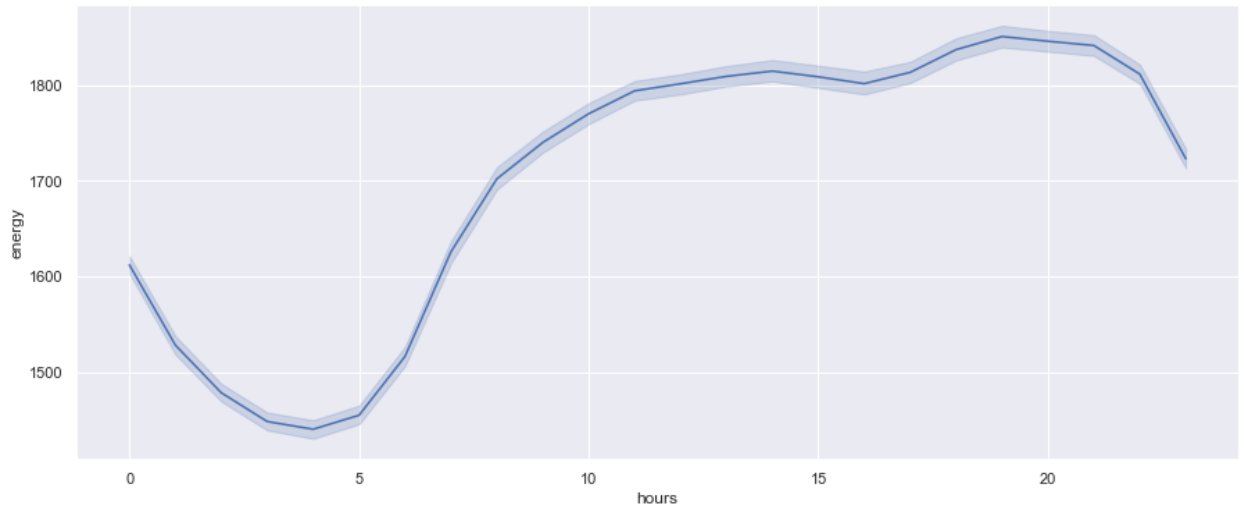
- To build a model or to train the model we need to fit large spread data into the model as we are using Linear regression algorithm so we will be importing `Linear_model.LinearRegression` and `mean_squared_error` metrics to evaluate the same.

Alter the given data:

- To do the model fitting we need to fix date-time, column's datatype from "object" to "datetime".
- Split datetime given values into date , day, month, year, hours, minutes, seconds.

Plot a graph:

- To see Energy consumption on hourly basis plot a line graph between Energy and Hours column.



Model Training:

- Define X, y where X is independent and y is dependent variables.
- Train the model with Linear Regression model.
- And prepare the model to predict the output for test data.

Test Data:

- To fit test data into the model to predict the output we first need to split “test.datetime” column into Date, day, month, year, hours, minutes, seconds. And then predict test data values and save them into “ypred”.

RMSE Metrics:

- Evaluate the RMSE using the mean_squared_error metrics.
- Received a value of 816 as RMSE.

Sample_Submission:

- As desired by the Examiner to store the results in Sample_submission.csv hence a file named the same is created.
- Desired data in terms of row_id , Energy is stored in sample_Submission.

Final Graph:

- A graph is plotted to see the 2019-2021 Energy consumption on hourly basis.

