Capstone Project 2

Team Member's Name, Email and Contribution:

Pratik D. Jori
Email: pratikjori.jori@gmail.com
Contribution:

1. Cleaning and exploring the data, checking missing values and null values, convert Dtype of date column
2. Exploratory Data Analysis – outlier, feature modification, feature distribution, correlation.
3. Fitting Model – splitting the data, applying algorithm, evaluating model.
4. Ppt making, technical document.

Pritesh Tambat
Email: pritesh.tambat1@gmail.com
Contribution:

1. Cleaning and exploring the data, checking null and missing values
2. Exploratory Data Analysis – Feature modification, distribution, heatmaps, correlation.
3. Fitting model – train test split the model, apply algorithm and evaluate models.
4. Ppt making, technical document.

Please paste the GitHub Repo link.

Pratik's Github link:- https://github.com/pratikjori20/Appliance-Energy-Prediction

Pritesh's Github link:- https://github.com/pritesh2292/Appliance-Energy-Prediction

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Data-driven prediction of energy use of appliances the data set is at 10 min for about 4.5 months. The house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. Each wireless node transmitted the temperature and humidity conditions around 3.3 min. Then, the wireless data was averaged for 10 minutes periods. The energy data was logged every 10 minutes with m-bus energy meters. Weather from the nearest airport weather station (Chievres Airport, Belgium) was downloaded from a public data set from Reliable Prognosis (rp5.ru) and merged together with the experimental data sets using the date and time column. Two random variables have been included in the data set for testing the regression models and to filter out non-predictive attributes (parameters).

The problem statement is to build the machine learning model that could predict the energy consumption of the Appliances

The first step is to import the all-necessary library to execute and import the data set given by the Almabetter then we check the null and missing values in the data, clean the data further we convert the date column in the Dtype as datetime64, then we check the basic statistical description of the data.

The second step exploratory data analysis where we dig into the data as mean energy consumption, trends of energy consumption, target distribution, feature distribution and plot the heatmaps, correlation matrix of all attributes. Further we check outliers and also treated outliers.

The third step we have try some machine learning algorithms on our train test splitted and standardize data we tried Linear regression, Ridge regression, Lasso regression, Polynomial regression, Decision tree regression. We done model implementation and parameter tunning, and also, we implement the feature importance technique to understand the which is important feature

The best performance has given by the Decision Tree regression model where RMSE is low as compare to other algorithms.
Top three features are humidity attributes which means the humidity affect the more power consumption than the temperature.

We can do hyperparameter tunning to deploy better algorithm such as support vector regressor, gradient boosting, neural networks etc.
So, the controlling humidity inside the house may save the energy consumption.
The dataset is mainly focus on datetime so we can approach the 'time series analysis' problem to forecast the consumption of energy.