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

Github Link:- https://github.com/Link/to/Repo

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

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).

Data used include measurements of temperature and humidity sensors from a wireless network, weather from a nearby airport station and recorded energy use of lighting fixtures. data filtering to remove non-predictive parameters and feature ranking plays an important role with this data. Different statistical models could be developed over this dataset.

The appliances energy consumption prediction in a low energy house is the dataset content Weather data from a nearby station was found to improve the prediction.

Pressure, air temperature and wind speed are important parameters in the prediction.

We have imported libraries to execute the given dataset. Then we checked the data, missing value and outliers, data wrangling.

Then we have done exploratory data analysis for better understanding of the given data such as mean energy consumption per day of the week and hours of the day, Feature distribution, heatmap and correlation matrix of all attributes.

Then we have done fitting models. We have done Linear regression, Ridge regression, Lasso regression, Polynomial regression, Decision tree regression and done feature importance from this we conclude the result as

Humidity attributes are more important than temperature attributes. The best performance achieved by Decision tree model.

We can use Principal Component Analysis that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables.
We can also do hyperparameter tuning deploy better algorithm such Support Vector Regressor, Decision Tree regressor, Gradient Boosting, Neural Nets etc., to get lower RMSE values. We can approach this as “Time Series Analysis” problem to forecast the energy usage based on the previous trend. We can use Arima or time series Decomposition methods to forecast the energy usage.