Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
V Rupesh Kumar Patro: (varnasipatro@gmail.com) Upload dataset to Google Colab · Data Cleanup. · Data cleaning. . EDA · Preprocessing · Model Implementation · Technical Write up . Power Point Presentation **Shashank Maindola: (shashank.ddun@gmail.com)** · Upload dataset to Google Colab · EDA · Correction of data types · Data Visualizations . Feature Engineering · Technical Write up . Model Explanation · PowerPoint presentation · Project summary
Please paste the GitHub Repo link.

Github Link:- https://github.com/rupeshpatro2001/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)

In this century, the world is being driven towards more cleaner means of energy. The electricity generated using the renewable sources plays a major role. But simply generating and utilizing the electrical energy is not viable, any extra generation will lead to non utilization and less generation will lead to power outages. Hence, we need demand side and supply side management so we can flatten the generation distribution curve. The residential load contributes to 27% of the total energy consumption (**Source:**:Eurostat). Residential household consumption is generally governed by weather conditions. Here we have a data of one of e residential building in Belgium and using the machine learning techniques we have to figure out which algorithm give the best output.
Few of the salient features of our dataset:-

- There are 29 columns and 19735 rows in our dataset.
- Max energy usage of appliance is 1080 and min is 10 watt
- light column having majority of the data 0 values
- Max pressure outside house is 772.3 mm_hg
- Except date column There is none categorical column in the dataset.
- Average temperature outside is about 7.5 degrees. While it ranges from -6 to 28 degrees.
- There is no null or missing values.
- Average humidity outside is higher than average humidity inside.
- Max wind speed is 14 m/s

We were provided with a dataset namely Data Application Energy. The following are the details of the data set

1. Data Application Energy contains 29 features:-

   'date', 'Appliances', 'lights', 'T1', 'RH_1', 'T2', 'RH_2', 'T3', 'RH_3', 'T4', 'RH_4', 'T5', 'RH_5', 'T6', 'RH_6', 'T7', 'RH_7', 'T8', 'RH_8', 'T9', 'RH_9', 'T_out', 'Press_mm_hg', 'RH_out', 'Windspeed', 'Visibility', 'Tdewpoint', 'rv1', 'rv2'

Firstly, we imported the library which were required to process our data, then we mounted the data from the drive link folder. We used following steps for appliance energy prediction using machine learning.

- Exploratory Data Analysis
- Clean Up
- Feature Engineering
- Pre Processing
- Model Implementation
- Model Explanation

Firstly, we conducted Exploratory Data analysis to get meaning full insight. We figured out few observations and based on insights we put them in following categories:

1. Temperature and Humidity
2. Seasonal and hourly consumption
3. Distribution
4. Target Variable

It was observed that more or less the humidity and temperature was same inside home hence we can take a mean of the temperature and humidity inside house. Secondly, we observed that energy consumption is directly related to the time of the day. Thirdly, we observed the distribution of the features and found that appliance is positively skewed , Temperature outside and windspeed are moderately positive skewed , outside humidity, visibility, Tdewpoint, rv1, rv2, Mean_house_temp, Mean_relative_humidity, months,hour are normally distributed and Relative humidity outside building is negatively skewed. Lastly, a lot of outliers present in our data set and there is positive correlation between temperature inside and outside with the appliance energy consumption while there is low correlation of appliance energy consumption with other variables

Then we did a cleanup of the data. We removed house parameters and replaced them with mean. We used calculated variance inflation factor and removed r1 and r2. We also dropped a few parameters like date which was not necessary for our model training

After that we applied standard scalar for standardization as part of pre processing.

Then we trained our model using various regression models and checked their performance. The using Grid Search we did hyperparameter tuning to have a more better model performance.

Based on the model performance we selected the best model and using SHAP we interpreted the model and we summarised our results.