Appliance Energy Prediction

Project Description

Business Context

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. The energy data was logged every 10 minutes with m-bus energy metres. 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. You need to predict the energy use of appliances.

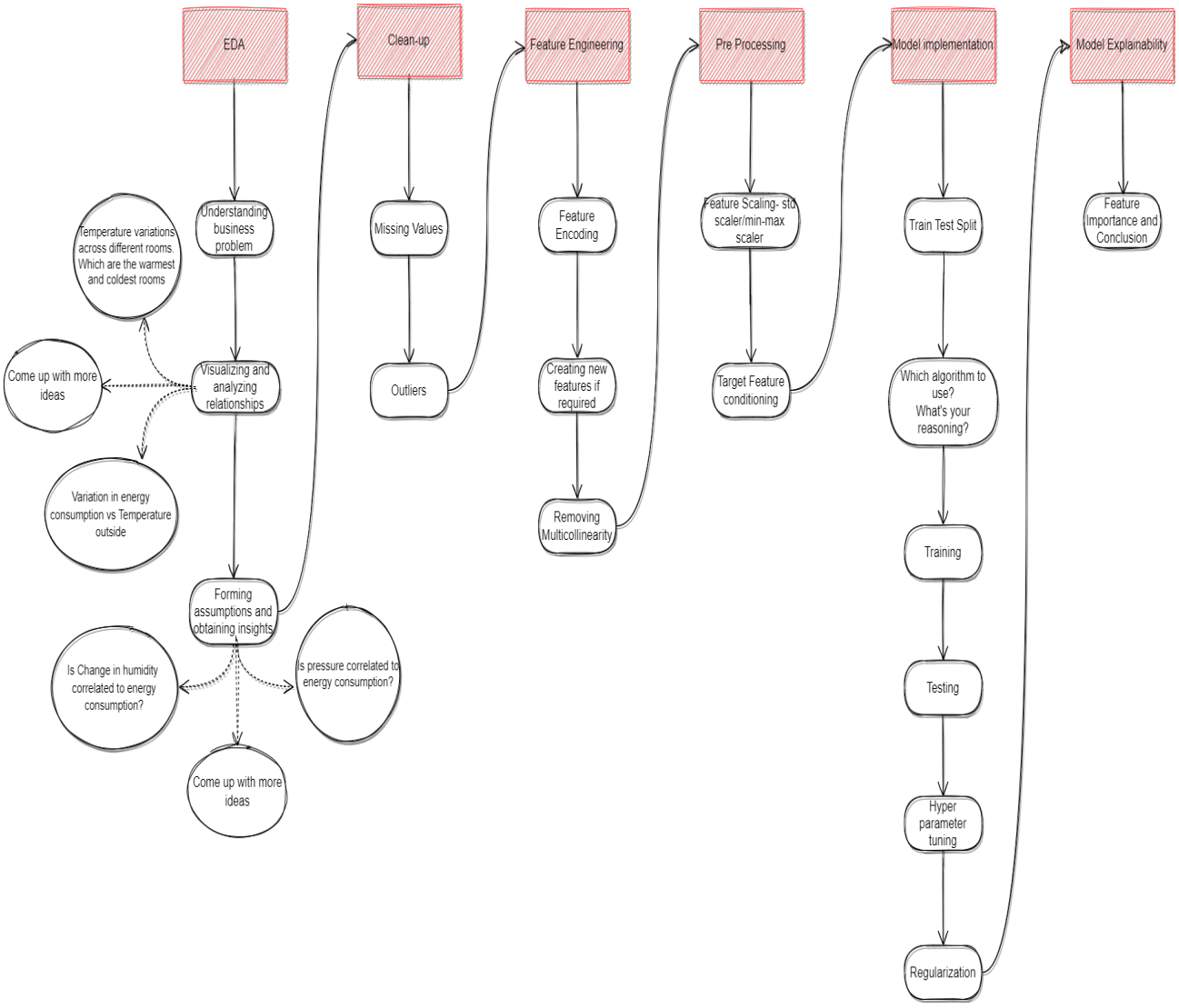
Data Description:

Fields	Description
T1	Temperature in kitchen area, in Celsius
T2	Temperature in living room area, in Celsius
T3	Temperature in laundry room area, in Celsius
T4	Temperature in office room, in Celsius
T5	Temperature in bathroom, in Celsius
T6	Temperature outside the building (north side), in Celsius
T7	Temperature in ironing room , in Celsius
T8	Temperature in teenager room 2, in Celsius
T9	Temperature in parents room, in Celsius
RH1	Humidity in kitchen area, in %
RH2	Humidity in living room area, in %
RH3	Humidity in laundry room area, in %
RH4	Humidity in office room, in %
RH5	Humidity in bathroom, in %
RH6	Humidity outside the building (north side), in %
RH7	Humidity in ironing room, in %
RH8	Humidity in teenager room 2, in %
RH9	Humidity in parents room, in %
To	Temperature outside (from Chievres weather station), in Celsius
Pressure	(from Chievres weather station), in mm Hg
Hg RHout	Humidity outside (from Chievres weather station), in %
Wind speed	(from Chievres weather station), in m/s
Visibility	(from Chievres weather station), in km
Tdewpoint	(from Chievres weather station), Â°C
Appliances, energy use in Wh	Dependent variable

Main Libraries to be Used:

- Pandas for data manipulation, aggregation
- Matplotlib and Seaborn for visualisation and behaviour with respect to the target variable
- NumPy for computationally efficient operations
- Scikit Learn for model training, model optimization, and metrics calculation

Project Architecture:



Project Evaluation Criteria

- Efficient EDA
- Encoding if necessary.
- Feature selection, new feature creation
- Dealing with multicollinearity if any
- Feature scaling
- Understanding the target feature and its distribution
- Modeling - use at least two algorithms
- Evaluation and improvement of model.
- Feature Importance and Conclusion
- Understanding of how your project is useful to stakeholders?

Rubrics

Rubrics	Weightage	Max Credits
GitHub Commits	5	100
Summary and Technical Documentation in Collab Notebook	10	100
EDA and Visualization	5	100
Looking for and Handling NaN/ Null/ Missing Values and Outliers	2.5	100
Finding Correlation in Variables (Both Dependent and Independent, Visualizations on Data)	10	100
Pick Appropriate Independent Variables, Test Train Split, Train Model	10	100
Prediction and Calculate Some Evaluation Metrics for Model	10	100
Number of Models Experimented (Atleast 2)	5	100
Hyperparameter Tuning	5	100
Final Summary of Conclusion	2.5	100
Commented Code	5	100
Proper Output Formatting	2.5	100
Modularity of Code	2.5	100
Video Presentation	20	100
Fluency and Grammatical Accuracy in Video	5	100

Data File

https://drive.google.com/file/d/1F8b3o67GpdIvEsHEVNZCD5tmQksW4vuJ/view?usp=sharing

Project Template
https://colab.research.google.com/drive/1pLHuk5ml8brfq76USc0n0suZU4FLxCqd?usp=sharing