High Level Design (HLD)

Energy Efficiency

Revision Number: 1.0

Last date of revision: 03-07-2022

Document Version Control

Date Issued	Version	Description	Author
03-07-2022	1	Initial HLD – V1.0	Karan/Sunil

Contents

Abstract

In this project we looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters. This helps in load requirement prediction and thus Energy requirements can be optimized.

Data Set Information:

We perform energy analysis using 12 different building shapes simulated in Ecotect. The buildings differ with respect to the glazing area, the glazing area distribution, and the orientation, amongst other parameters. We simulate various settings as functions of the afore-mentioned characteristics to obtain 768 building shapes. The dataset comprises 768 samples and 8 features, aiming to predict two real valued responses. It can also be used as a multi-class classification problem if the response is rounded to the nearest integer.

Attribute Information:

The dataset contains eight attributes (or features, denoted by X1...X8) and two responses (or outcomes, denoted by y1 and y2). The aim is to use the eight features to predict each of the two responses.

Specifically:
X1 Relative Compactness
X2 Surface Area
X3 Wall Area
X4 Roof Area
X5 Overall Height
X6 Orientation
X7 Glazing Area
X8 Glazing Area Distribution
y1 Heating Load
y2 Cooling Load

.

1 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
  - Security
  - Reliability
  - Maintainability
  - Portability
  - Reusability
  - Application compatibility
  - Resource utilization
  - Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

2 General Description

2.1 Product Perspective

This study looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters.

2.2 Problem statement

To create an AI solution to predict heating load and cooling load based on the building parameters. This helps in predicting load requirements and Energy Efficiency.

2.3 PROPOSED SOLUTION

The solution proposed here is to develop and deploy an app which can predict the load and energy requirements based on below parameters.

X1 Relative Compactness
X2 Surface Area
X3 Wall Area
X4 Roof Area
X5 Overall Height
X6 Orientation
X7 Glazing Area
X8 Glazing Area Distribution

2.4 FURTHER IMPROVEMENTS

This implementation can be improved by including more parameters.

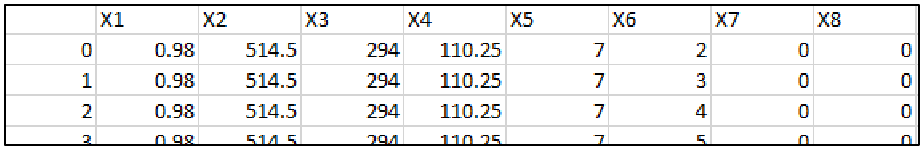
Model re-training pipeline integration can be implemented as a next phase for further improvements.

2.5 TOOLS USED

Python will be used for defining the Predictive Model as well as generating the predictions for user input. This project takes the input in the form of an excel spreadsheet. User's Web Browser will be done by Flask framework, scripted on Python. The System will be deployed on Heroku through CI/CD pipeline.

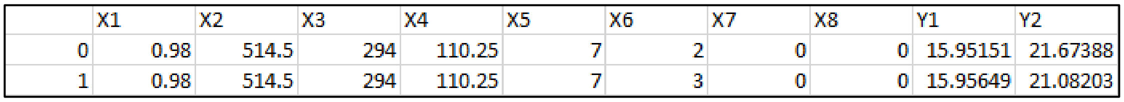
3 Assumptions

Assumption for our implementation is that user provides the inputs in the form of an excel spreadsheet ensuring all parameters are included as per agreed format (See below).



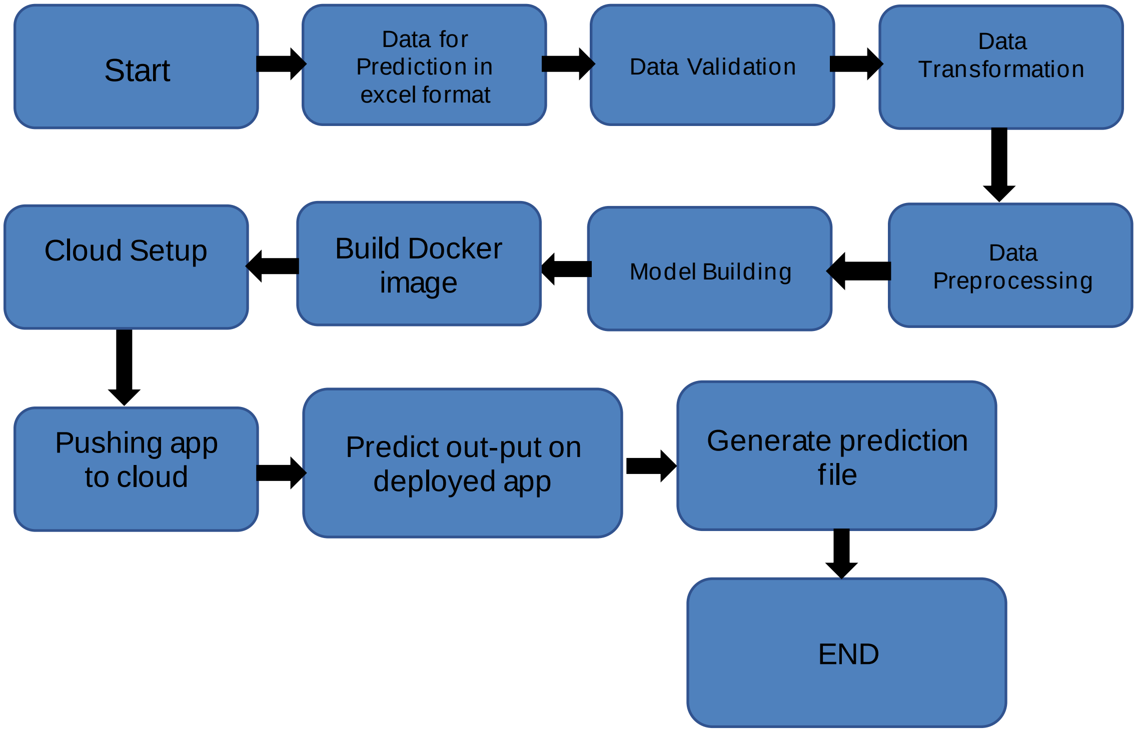
	X1	X2	X3	X4	X5	X6	X7	X8
0	0.98	514.5	294	110.25	7	2	0	0
1	0.98	514.5	294	110.25	7	3	0	0
2	0.98	514.5	294	110.25	7	4	0	0
3	0.98	514.5	294	110.25	7	5	0	0

Out put file will included predictions Y1(Heating Load) and Y2(Cooling Load) along with input columns (See below).



	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2
0	0.98	514.5	294	110.25	7	2	0	0	15.95151	21.67388
1	0.98	514.5	294	110.25	7	3	0	0	15.95649	21.08203

4 Application Architecture



4.1 Event log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description:

1. The System identifies at what step logging required
2. The System should be able to log each and every system flow.
3. Developer can choose logging method. You can choose database logging/ File logging as well.
4. System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

4.2 Error Handling

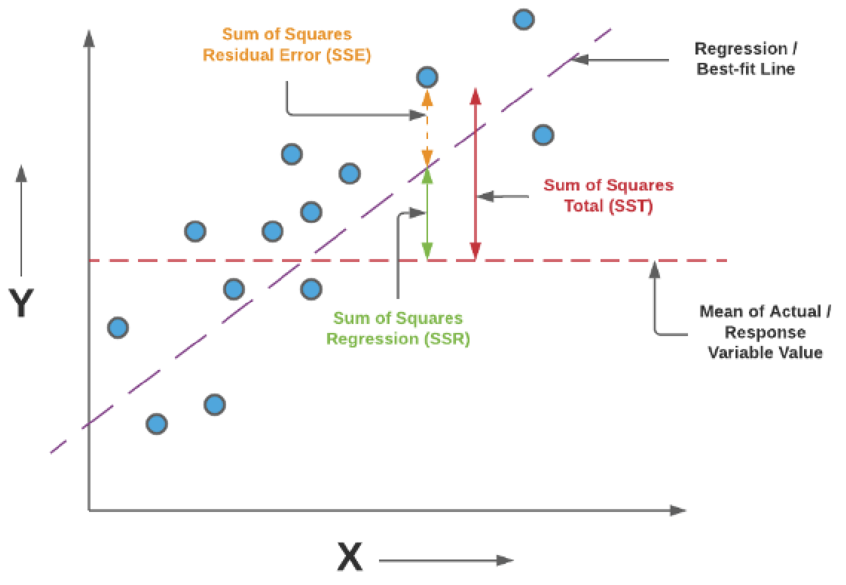
Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usag e.

5 Performance

Performance metric chosen for model training and implementation is R2 score,

- It is recommended to use R-Squared or rather adjusted R-Squared for evaluating the model performance of the regression models.
- This is primarily because R-Squared captures the fraction of variance of actual values captured by the regression model and tends to give a better picture of the quality of the regression model
- MSE values differ based on whether the values of the response variable are scaled or not. A better measure instead of MSE is the root mean squared error (RMSE) which takes care of the fact related to whether the values of the response variable are scaled or not.
- The RMSE tells us how well a regression model can predict the value of the response variable in absolute terms while R2 tells us how well a model can predict the value of the response variable in percentage terms.

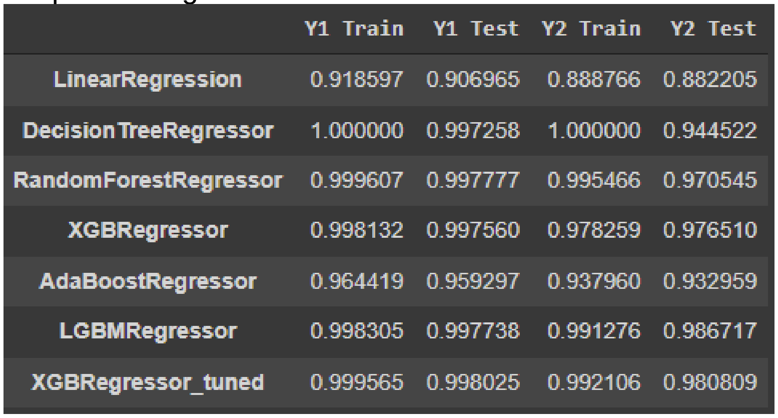
See below for R-Squared illustration



Diagrammatic representation for understanding R-Squared

$$R^2 = \frac{SSR}{SST} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

Below is the R2 scores for various ML models after calibration, Light GBM Regressor is chosen as it is best performing of all.



	Y1 Train	Y1 Test	Y2 Train	Y2 Test
LinearRegression	0.918597	0.906965	0.888766	0.882205
DecisionTreeRegressor	1.000000	0.997258	1.000000	0.944522
RandomForestRegressor	0.999607	0.997777	0.995466	0.970545
XGBRegressor	0.998132	0.997560	0.978259	0.976510
AdaBoostRegressor	0.964419	0.959297	0.937960	0.932959
LGBMRegressor	0.998305	0.997738	0.991276	0.986717
XGBRegressor_tuned	0.999565	0.998025	0.992106	0.980809

6 Reusability

The code written and the components used should have the ability to be reused with no problems.

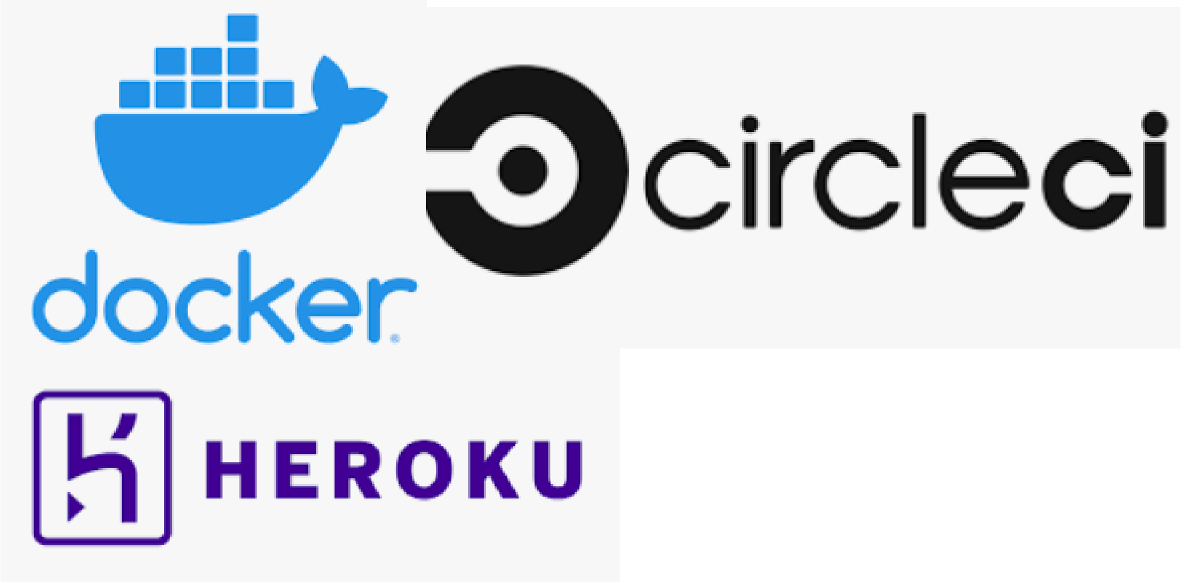
7 Application Compatibility

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

8 Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

9 Deployment



10 Conclusion

This study looked into assessing the heating load and cooling load requirements of buildings (that is, energy efficiency) as a function of building parameters. This helps in load requirement prediction and thus Energy requirements can be optimized.

11 References

https://archive.ics.uci.edu/ml/datasets/energy+efficiency

https://www.statology.org/rmse-vs-r-squared/

https://vitalflux.com/mean-square-error-r-squared-which-one-to-use/