ENERGY EFFICIENCY ESTIMATION
High Level Design

Table of Contents

Abstract

The effect of eight input variables (relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution) on two output variables, namely heating load (HL) and cooling load (CL), of residential buildings is investigated using a statistical machine learning framework. We have to use a number of classical and non-parametric statistical analytic tools to carefully analyze the strength of each input variable's correlation with each of the output variables in order to discover the most strongly associated input variables. We need to estimate HL and CL, we can compare a traditional linear regression approach to a sophisticated state-of-the-art nonlinear non-parametric method, random forests.

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.

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.

1.3 Definitions

Term	Definitions
EE	Energy Efficiency
Database	Collection of all the information monitored by this system
IDE	Integrated Development Environment
Heroku	Heroku Cloud for Deployment

1.4. Overview

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

1.5 Uses

- This document is designed to help in operational requirement and can be used as a reference manual for how the modules interact.
- HLD briefly describes about the platforms/products/services/processes, flow of traffic that it depends on and includes any important changes that need to be made to them.

- HLD is the input for creating the LLD (Low Level Design) since the key communication items are displayed in HLD which are then converted to detailed communication in LLD, showing connectivity and physical level.

1.6 Application

- Website Development
- Application Development
- Data Science Project

2. GENERAL DESCRIPTION

2.1 Product Perspective

The EE (Energy Efficiency) solution system is machine learning model which helps to find out the co-relation between the input variable and output variable of residential building. Each row of dataset represents the one residential building parameter. So our task is to predict the output on the basis of given input variable.

2.2 Problem Statement

The effect of eight input variables (relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution) on two output variables, namely heating load (HL) and cooling load (CL), of residential buildings is investigated using a statistical machine learning framework. We have to use a number of classical and non-parametric statistical analytic tools to carefully analyze the strength of each input variable's correlation with each of the output variables in order to discover the most strongly associated input variables. We need to estimate HL and CL; we can compare a traditional linear regression approach to a sophisticated state-of-the-art nonlinear non-parametric method, random forests.

2.3 Proposed Solution

The heating load is the amount of heat energy that would need to be added to a space to maintain the temperature in an acceptable range. The cooling load is the amount of heat energy that would need to be removed from a space (cooling) to maintain the temperature in an acceptable range. So from the above problem statement we can find out that HL & CL from the given input attribute. The HL & CL will help you to find out the different parameter that is required for the building estimation in terms of the electricity. In the given model we take data from the client whatever we required for the predicting the HL & CL.

2.4 Technical Requirement

In this Project the requirements to get energy efficiency through various platform. For that, in this project we are going to use different technologies. Here are some requirements for this project.

- Model should be exposed through API or User Interface, so that anyone can test model.
- Model should be deployed on cloud (Azure, AWS, GCP or Heroku) for the public use.
- Cassandra database should be integrated in this project for any kind of user input.

2.5 Data Requirements

- Data Requirement completely depends on our problem.
- For training and testing the model, we are using energy efficiency prediction dataset.
- From user we are taking following input :
    - Relative Compactness :- Float Value By User
    - Surface Area :- Float Value By User
    - Wall Area :- Float Value By User
    - Roof Area :- Float Value By User
    - Overall Height :- Float Value By User
    - Orientation :- 2,3,4,5
    - Glazing Area :- Float Value By User
    - Glazing Area Distribution :- 0,1,2,3,4,5

2.6 Tools Used

Python programming language and frameworks such as NumPy, Pandas and Scikit-learn are used to build the whole model.

- PyCharm is used as IDE
- For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
- AWS used for deployment of the model
- Cassandra is used to insert, delete, retrieve and update the database.
- Front end development is done using HTML/CSS
- Python flask is used for backend development
- GitHub is used as version control system.

Figure No. 1 Entire Technology That Used in Project

2.7 Constraints

The energy efficiency prediction system must be user friendly, errors free and users should not be required to know any of the back-end working.

3. DESIGN DETAILS

3.1 Process Flow

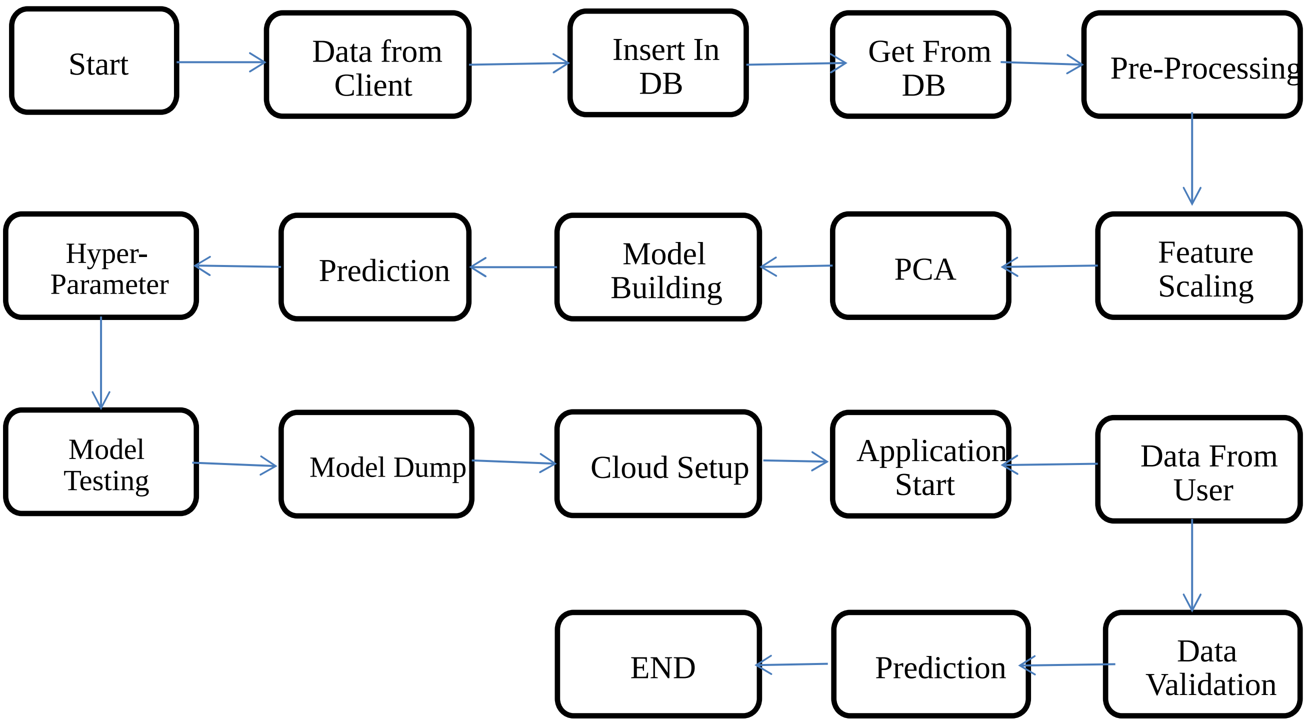


Figure No. 2 The Total Model Process

3.2 Deployment Process

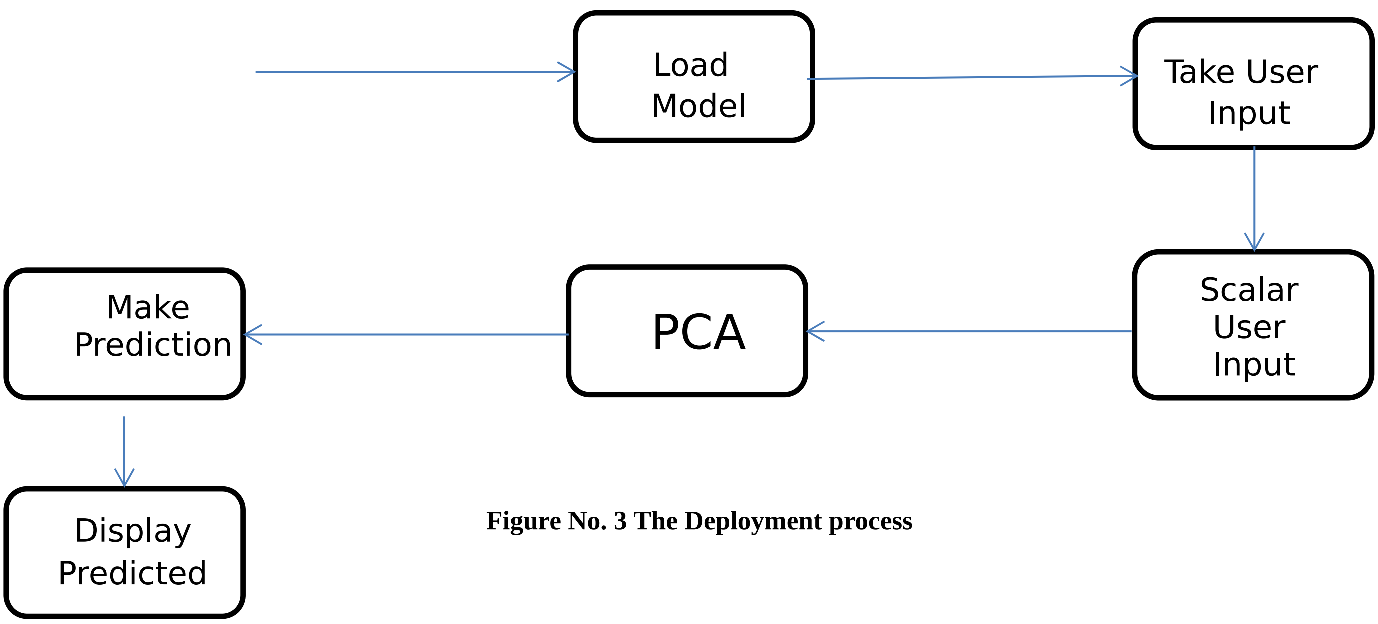


Figure No. 3 The Deployment process

3.3 Event Log

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

Internal Step-By-Step Description :-

- In this Project we defined logging for every function, class.
- By logging we can monitor every insertion, every flow of data in database.
- By logging we are monitor every step which may create problem or every step which is important in file system.
- We have designed logging in such a way that system should not hang even after so many logging's, so that we can easily debug issues which may arise during process flow.

3.4 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 usage.

4 . Performance

11Solution of energy efficiency prediction is used to predict the heating & cooling load in advance, so it should be as accurate as possible so that it should give as much as possible accurate prediction. That's why before building this model we followed complete process of Machine Learning. Here are summary of complete process:

1. First we cleaned our dataset properly by removing all null value and duplicate value present in dataset and if not present leave the dataset as it is.
2. Second we have to handle the outlier form the dataset so that it can't effect on the accuracy. So in our dataset there is no outlier present.
3. Then after that we scale down the whole dataset in between 0 & 1 with help of Min Max scalar.
4. After the scaling our data we given that data as input to the PCA technique for dimensionality reduction and remove the multi-colinearity between the column.
5. Then I split the whole data set into train-test split with test size will be 33% of the whole dataset.
6. After performing above step we ready for model training. In this step, I trained my dataset on different Regression algorithm such as Linear, Lasso, Ridge, Elastic Net, and Random Forest etc.
7. After that I applied hyper-parameter tuning on all models which I have described above.
8. After that I saved my model in pickle file format for model deployment.
9. After that my model was ready to deploy. I deployed this model on Heroku cloud.

4.1 Re-usability

We have done programming of this project in such way that it should be reusable so that anyone can add and contribute without facing problem.

4.2 Application Compatibility

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

4.3 Resource Utilization

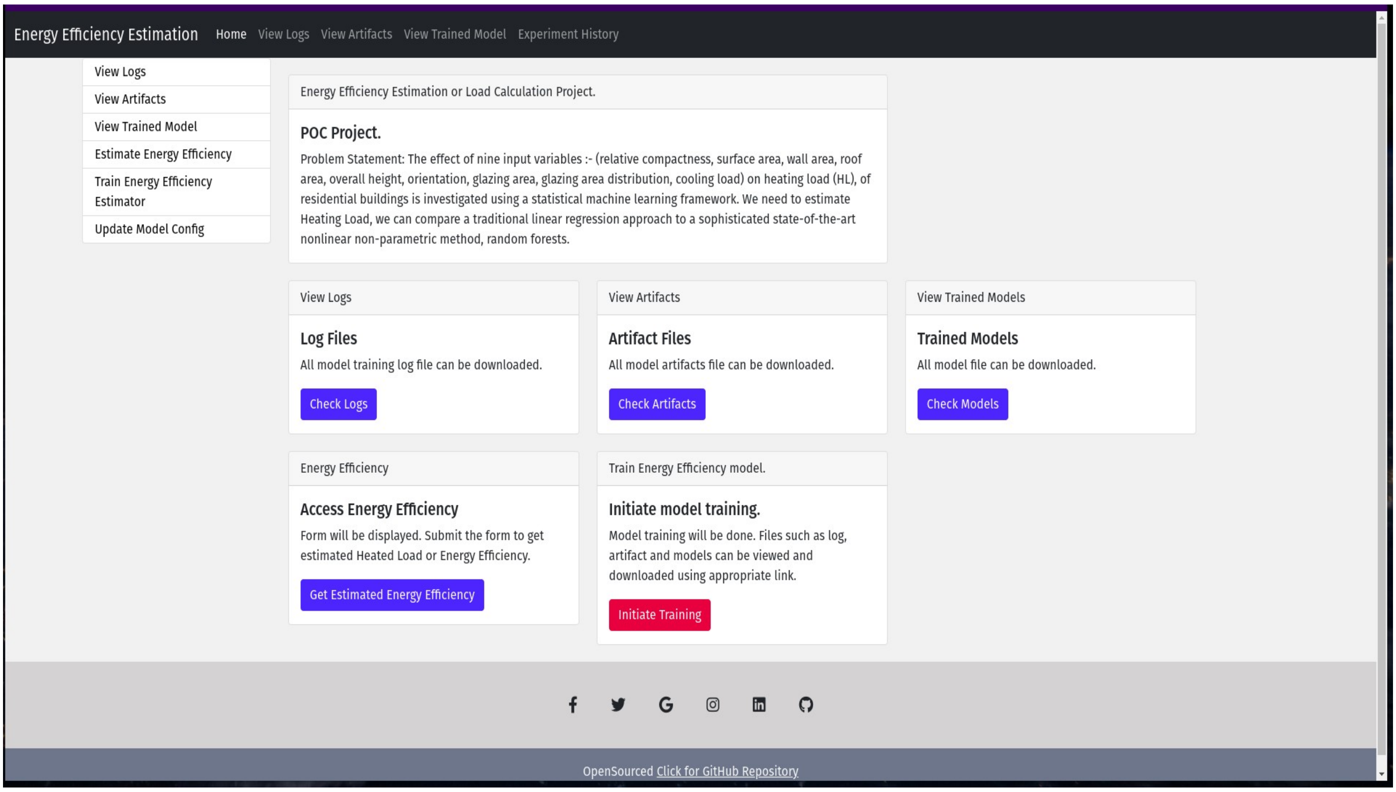
When any task is performed, it will likely use all the processing power available until that function is finished.

4.4 Deployment

I have deployed the project on Heroku cloud service only.

4.5 User Interface

UI for user has been created by using HTML and CSS.



5. Conclusion

The Energy Efficiency application will find out the Heating Load (HL) and Cooling Load (CL). From these we can find out how much load we required for building estimation.