Project Title	**Energy Efficiency**
Technologies	**Machine Learning**
Domain	**Industrial Automation**

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.

Dataset:
The Dataset you can get through this link: Dataset
The dataset contains eight attributes (or features, denoted by X1...X8) and two responses
(or outcomes, denoted by y1 and y2).

Project Evaluation metrics:
Code:
• You are supposed to write a code in a modular fashion (**in functional blocks**)
• Maintainable: It can be maintained, even as your codebase grows.
• Portable: It works the same in every environment (operating system)
• You have to maintain your code on **GitHub**.
• You have to keep your **GitHub** repo public so that anyone can check your code.
• Proper readme file you have to maintain for any project development.
• You should include basic workflow and execution of the entire project in the readme file on **GitHub**
• Follow the coding standards: https://www.python.org/dev/peps/pep-0008/

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 - Target**

y2 **Cooling Load - Target**

Note: After completion of all the task you need to create a **PowerPoint presentation** That should contain the :

1. Problem Statement
2. Tools Used
3. Approaches
4. EDA Insights
5. Best ML Model
6. Evaluation Metrics of Model
7. Future Development