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

By,
SIDHARTH

PROBLEM STATEMENT

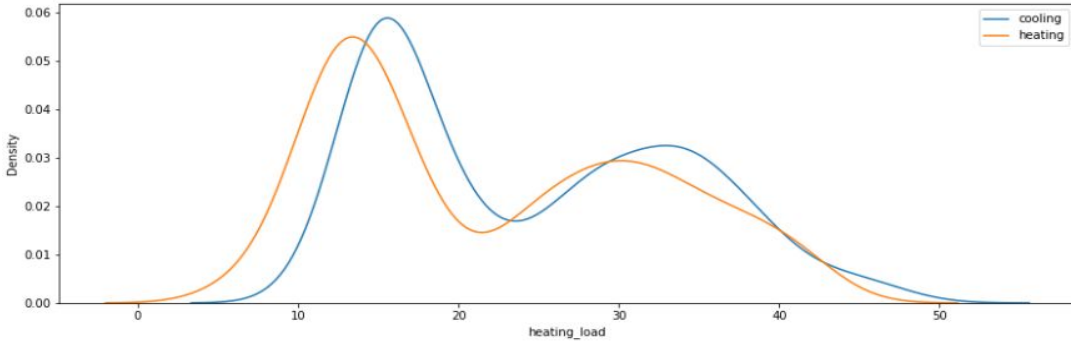
- The effect of **8 Input** variables (relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution) on **2 Output** variables, namely heating load (HL) and cooling load (CL), of residential buildings.
- Analyze the strength of each input variable's correlation with each of the output variables in order to discover the most **strongly associated input variables**.
- Estimate **Heating Load** and **Cooling Load**.

Tools used & Approach

- TOOLS USED:
  1. Machine learning - Scikit learn
  2. Data Visualization - Seaborn, Matplotlib
  3. Data Manipulation - Pandas, Numpy
- APPROACH:
  1. Splitting Targets from Features
  2. Exploratory Data Analysis
  3. Feature Selection
  4. Modelling
  5. Inferences

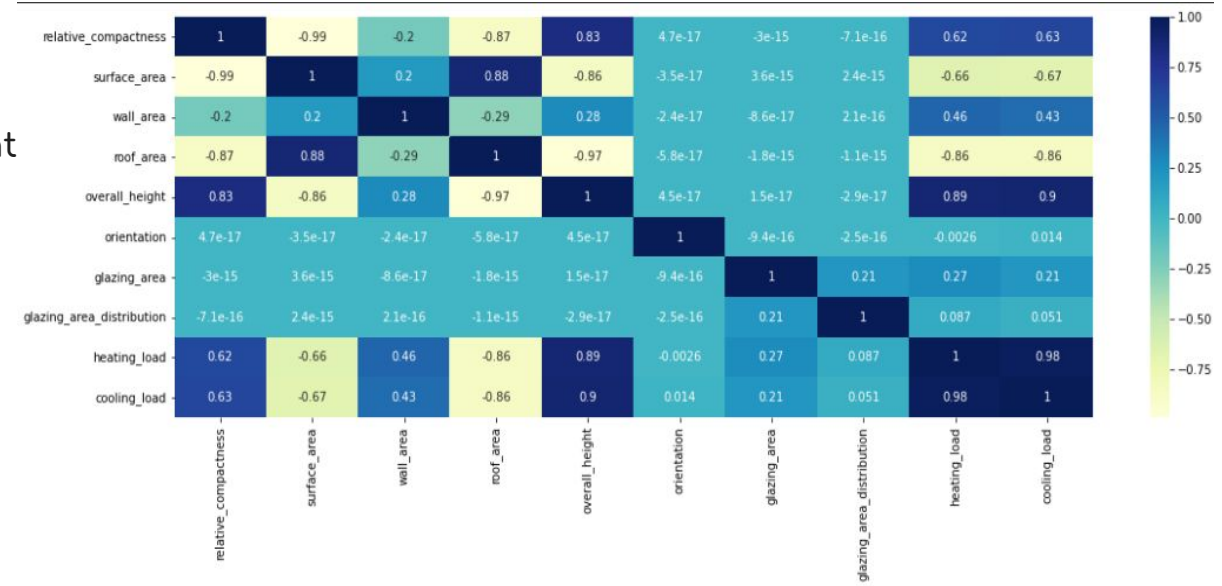
EDA Insights (Target variables)

- ➔ Most of the datapoints fall between 10 and 20 of both 'heating_load' and 'cooling_load' regressional output classes.
- ➔ It follows Bimodal distribution with a Positive skew.



EDA Insights (feature correlation)

- ➔ We can see that some features are <u>Positively</u> correlated (overall_height relative_compactness)
- ➔ some are <u>Negatively</u> correlated (roof_area, surface_area)
- ➔ some have <u>No-Linear</u> correlation (orientation, glazing area)



ML Models used

Linear Regression

Ridge & Lasso Regression

Knn - Regressor

Decision Tree Regressor

Random Forest Regressor

Evaluation Metrics
(cross validation using R2)

```
Linear reg:          CV-SCORE=0.89        (std=0.07)
Ridge reg:         CV-SCORE=0.89        (std=0.07)
Lasso reg:         CV-SCORE=0.74        (std=0.19)
RandomForest:          CV-SCORE=0.97        (std=0.08)
KNN reg:         CV-SCORE=0.88        (std=0.19)
Decision tree:         CV-SCORE=0.97        (std=0.08)
```

The best models for <u>heating_load</u>:

- Random forest (97%)
- Decision tree (97%)

```
Linear reg:          CV-SCORE=0.88        (std=0.03)
Ridge reg:         CV-SCORE=0.87        (std=0.04)
Lasso reg:         CV-SCORE=0.75        (std=0.10)
RandomForest:          CV-SCORE=0.97        (std=0.02)
KNN reg:         CV-SCORE=0.92        (std=0.09)
Decision tree:         CV-SCORE=0.96        (std=0.02)
```

The best models for <u>cooling_load</u>:

- Random forest (97%)
- Decision tree (96%)
- KNN (92%)

Future Developments

- Tuning the models using Hyperparameters
  - Using Grid Search cross validation
- Using other Algorithmns:
  - XGBoost
- Using Deep Learning techniques:
  - Neural Networks (Keras)

Thank You