ADA PROJECT REPORT

ENERGY EFFICIENCY

Course Faculty: Dr. Mainak Thakur Sir

Team Members :

NAME	ROLL NO.	EMAIL
BHARGHAV SAI PENDYALA	S20190010017	bharghavsai.p19@iiits.in
YASH GUPTA	S20190010197	yash.g19@iiits.in

ABSTRACT

When designing and building a new energy-efficient house which is challenging. However, recent technological improvements in building elements and construction techniques also allow most modern energy-saving ideas to be seamlessly integrated into house designs while improving comfort, health, or aesthetics. And even though some energy-efficient features are expensive, there are others that many homebuyers can afford. While design costs, options, and styles vary, most energy-efficient homes have some basic elements in common: a well-constructed and tightly sealed thermal envelope; controlled ventilation; properly sized, high-efficiency heating and cooling systems; and energy-efficient doors, windows, and appliance. Our Energy Efficiency project helps to predict the values of these heating and cooling load.

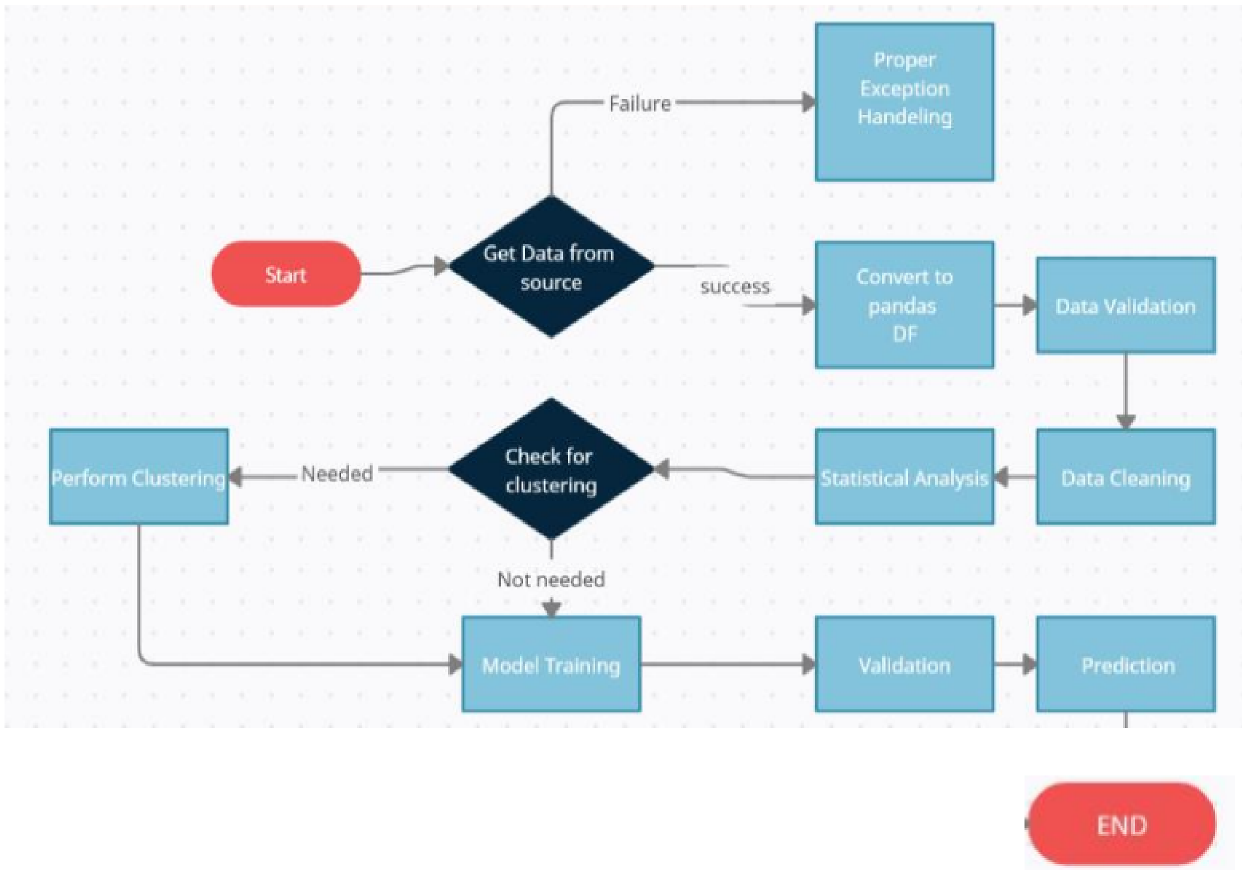
OBJECTIVE

The main objective of the project is 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 to predict the value of heating load and cooling load. An Energy Efficiency contains information, such as • Relative Compactness • Surface Area • Wall Area • Roof Area • Overall Height • Orientation • Glazing Area • Glazing Area Distribution • Heating Load • Cooling Load

DATASET OVERVIEW

We get data from this link. The dataset is on Energy efficiency for the building.

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.


Proper Exception Handeling
Failure
Start
Get Data from source
success
Convert to pandas DF
Data Validation
Perform Clustering
Needed
Check for clustering
Statistical Analysis
Data Cleaning
Not needed
Model Training
Validation
Prediction
END

```
data
```

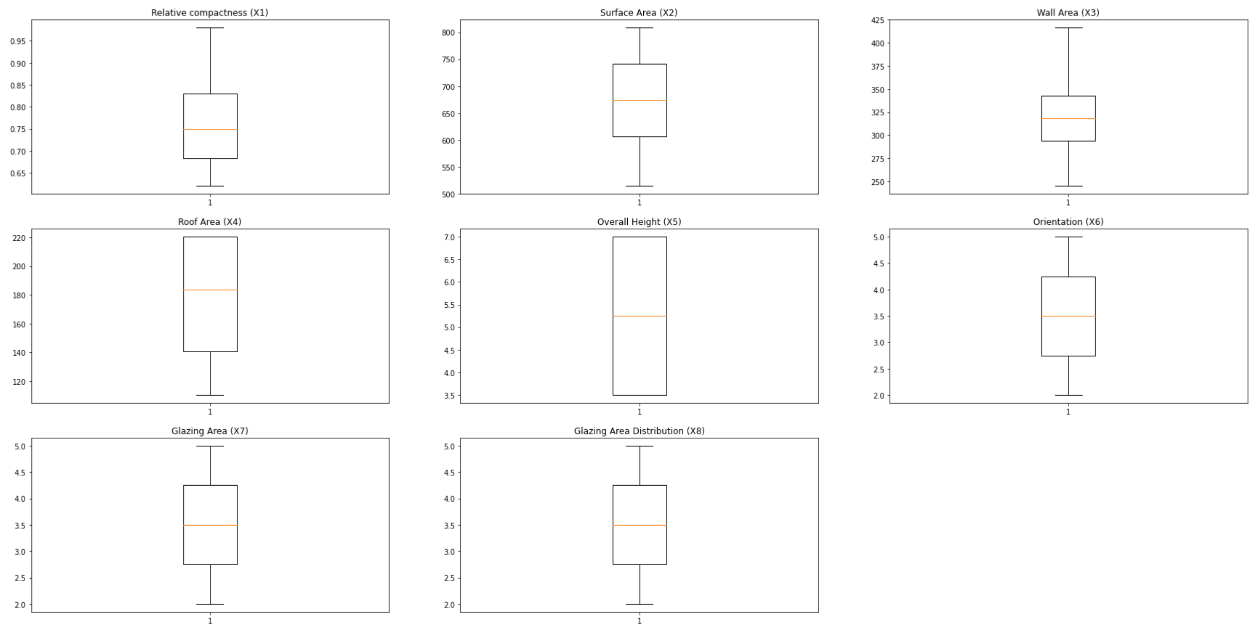
	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2
0	0.98	514.5	294.0	110.25	7.0	2	0.0	0	15.55	21.33
1	0.98	514.5	294.0	110.25	7.0	3	0.0	0	15.55	21.33
2	0.98	514.5	294.0	110.25	7.0	4	0.0	0	15.55	21.33
3	0.98	514.5	294.0	110.25	7.0	5	0.0	0	15.55	21.33
4	0.90	563.5	318.5	122.50	7.0	2	0.0	0	20.84	28.28
...	...	...	...	...	...	...	...	...	...	...
763	0.64	784.0	343.0	220.50	3.5	5	0.4	5	17.88	21.40
764	0.62	808.5	367.5	220.50	3.5	2	0.4	5	16.54	16.88
765	0.62	808.5	367.5	220.50	3.5	3	0.4	5	16.44	17.11
766	0.62	808.5	367.5	220.50	3.5	4	0.4	5	16.48	16.61
767	0.62	808.5	367.5	220.50	3.5	5	0.4	5	16.64	16.03

768 rows × 10 columns

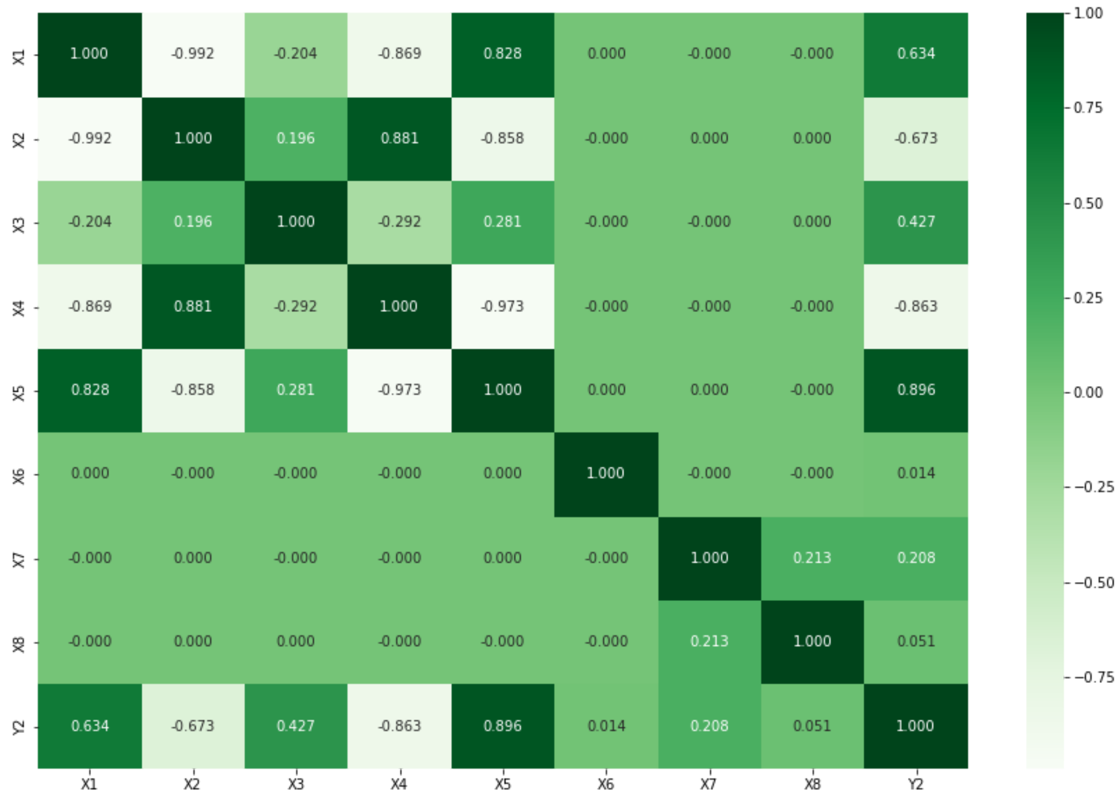
PROCEDURE

EDA on the data set

1. Finding the boxplot for individual xi’s



-> finding correlation matrix

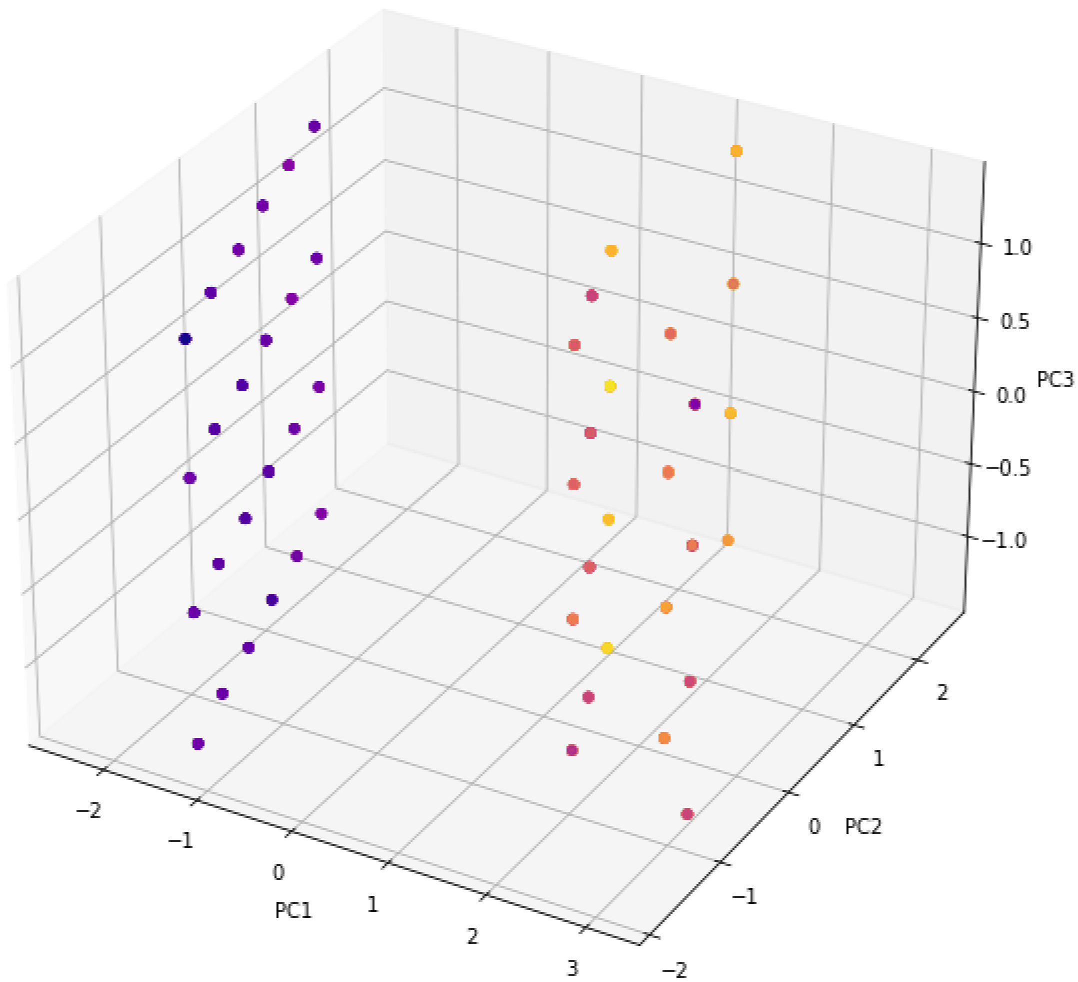


2. Check for Multicollinearity using VIF

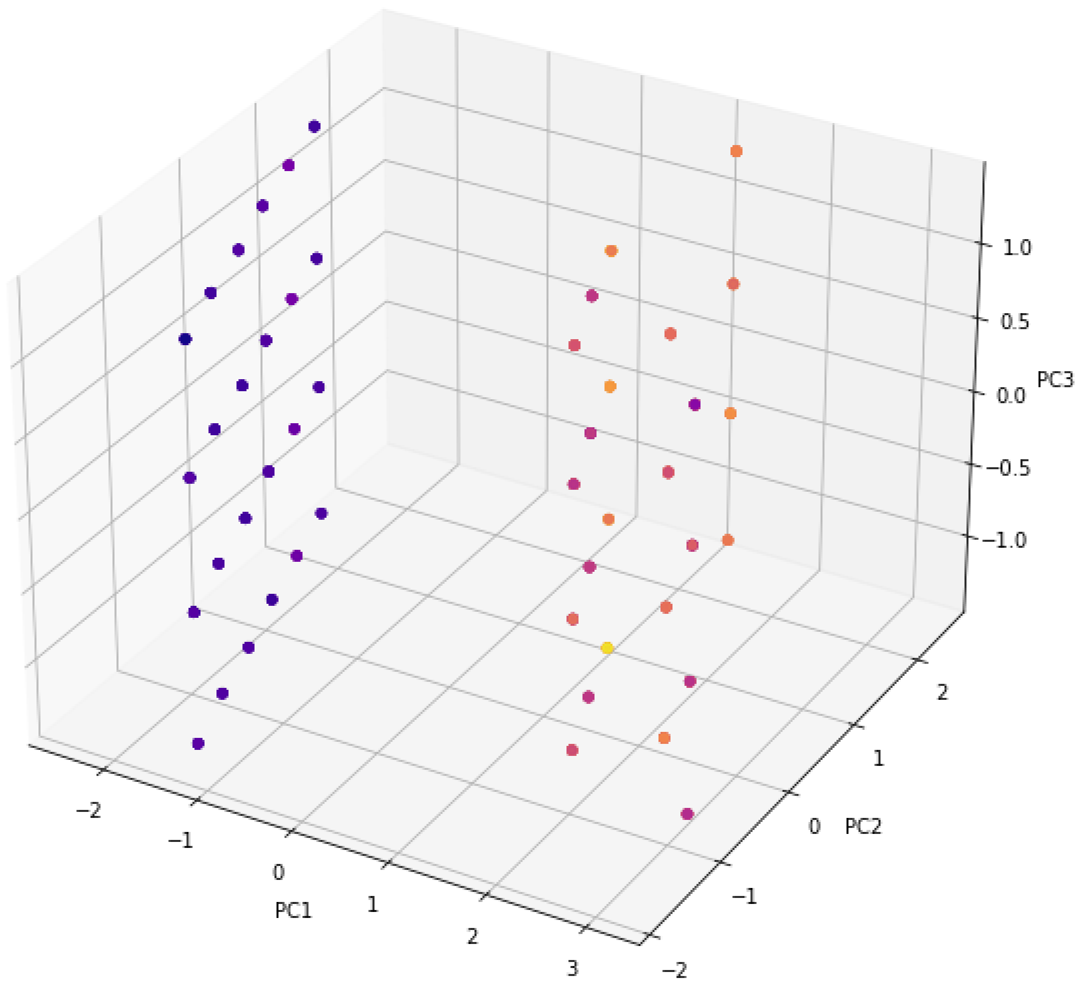
	VIF	Column
0	168.948751	X1
1	inf	X2
2	inf	X3
3	inf	X4
4	134.035782	X5
5	10.796725	X6
6	4.293656	X7
7	4.496320	X8

4. Here we can see that multicollinearity exists so dimensionality reduction required for X1 to X6 so we try to convert them to three principal components using PCA.

For Y1 :



For Y2 :



Principal Components :

```
array([[ 4.95951415e-01, -5.01733022e-01,  3.25142980e-02,
        -5.04962227e-01,  4.96237987e-01,  1.52492590e-17],
       [-2.44734692e-01,  2.31540671e-01,  8.94291660e-01,
        -2.06120761e-01,  2.10358046e-01,  4.68428397e-16],
       [-1.06398924e-16,  1.42287264e-16,  4.03033217e-16,
        -1.67367294e-16,  8.33237188e-17, -1.00000000e+00]])
```

Percentage of variance explained by principal components : (96% approx.)

```
[ ]  # check how much variance is explained by each principal component
     print(principal.explained_variance_ratio_)

     [0.61715655 0.20664189 0.16666667]
```

Now we are Combining PCAs to remaining attributes X7 and X8

3. Split the new data set into test and train

For Cooling Load (Y2) :

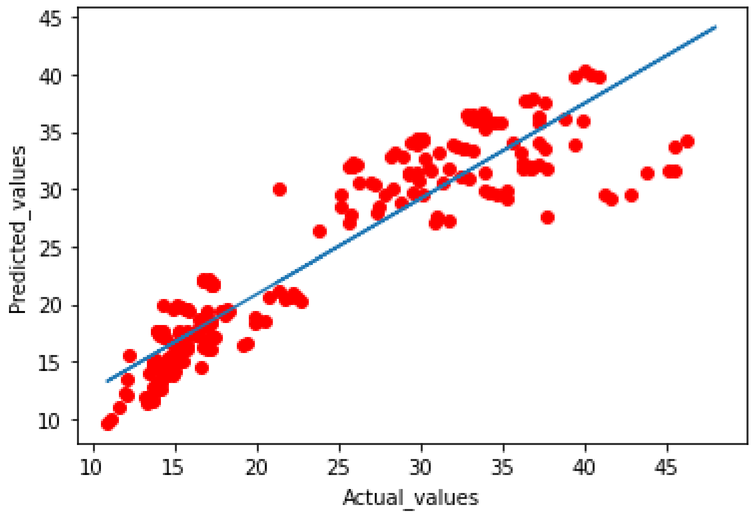
=> Applying Linear Regression

-> Training model with training data set

-> Predicting values using test data set

Scatter plot :

Actual vs Predicted :



->Calculating R2 score for predicted data.(0.848)

->Calculating RMSE for Cooling load of (3.729)

Similarly training the model using Decision Trees, Gradient Boosting, Random Forest, Support Vector Regressor

Comparison between them :

	Model	R2 Score	RMSE Value
0	Linear Regression	0.848165	3.729873
1	Decision Tree	0.972902	1.575712
2	Random Forest	0.971620	1.612541
3	Gradient Boosting	0.974838	1.518388
4	Support Vector Regressor	0.863519	3.536261

Conclusion: **So according to RMSE, we can say that Gradient Boosting is better to predict Cooling Load followed by Decision Trees and Random Forest**

For Heating Load (Y1) :

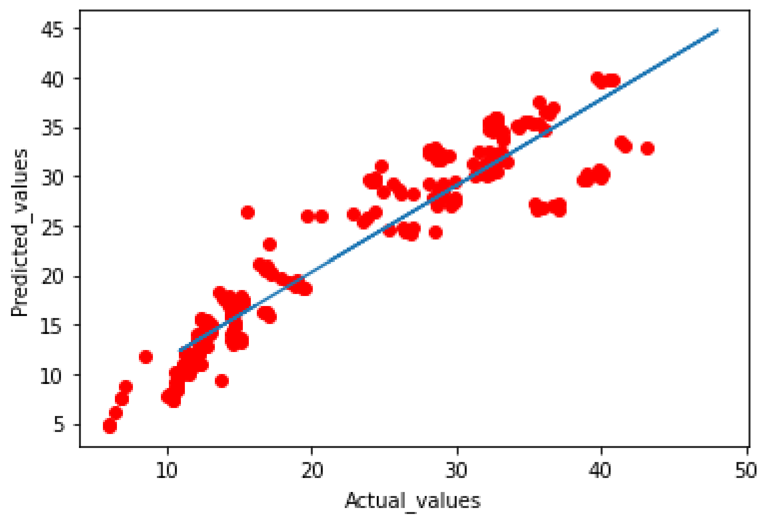
=> Applying Linear Regression

-> Training model with training data set

-> Predicting values using test data set

Scatter plot :

Actual vs Predicted :



->Calculating R2 score for predicted data.(0.879)

->Calculating RMSE for Cooling load of (3.499)

Similarly training the model using Decision Trees, Gradient Boosting, Random Forest, Support Vector Regressor

Comparison between them :

	Model	R2 Score	RMSE Value
0	Linear Regression	0.879976	3.499458
1	Decision Tree	0.996228	0.620414
2	Random Forest	0.997214	0.533129
3	Gradient Boosting	0.997240	0.530648
4	Support Vector Regressor	0.879973	3.499515

Conclusion: **So according to RMSE, we can say that Gradient Boosting is better to predict Cooling Load followed by Random Forest and Decision Trees.**