A project report on

Predicting Energy Loads in Buildings

Submitted by

G VISHNU VARDHAN REDDY (18MCMB06)

D SAI SREEHARI PRASAD (18MCMI08)

Students of

M.TECH

SCHOOL OF COMPUTER & INFORMATION SCIENCE

UNIVERSITY OF HYDERABAD

2018-2020

Submitted to

Prof. Dr. V. RAVI

Professor & Head of Center of Excellence in Analytics, IDRBT

Predicting Energy Loads in Buildings

Table of Contents

Table of Contents **2**

1.Introduction **3**
1.1.Tools Used : 3

2.Problem Definition and Data Mining Task **4**

3.Techniques Used and Description **5**
3.1.Techniques Description: 5
3.2. Activation Functions: 7
3.3. Metrics: 9

4.Dataset Description **12**
4.1.Feature Description 12

5.Descriptive Analytics **14**
5.1.Univariate Analysis: 14
5.1.1 Box Plots for each feature with heating load and cooling load: 14
5.1.2.PDF’s of heating load and cooling load: 18
5.3. Correlation Matrix : 19

6.Data Preprocessing **20**
6.1 Data Cleansing: 20
6.1.1 Handling Missing Values: 20
6.1.2 Data Preparation and Transformations : 20
6.2. Data Normalization: 21
6.3. Feature Selection: 21
6.3.1.1 Domain Knowledge: 21
6.3.1.2 Filter Based : 21
6.4 Data Partitioning Methods : 22
6.4.1 Hold Out: 22
6.4.2 5 Fold CV: 22

7. Results **23**

5. Table **31**

8.Conclusions **47**

9. References **48**

1.Introduction

The basic principle of building energy efficiency is to use less energy for operations including heating, cooling, lighting and other appliances, without affecting the health and comfort of its occupants. Improving the energy efficiency of functional buildings brings many environmental and economic benefits such as reduced greenhouse gas emissions and operational cost savings. In many developed and developing countries, energy efficiency has become the main way to meet a rising energy demand.

Although building orientation and layout have been shown to be highly important in reducing building energy consumption in cold and hot climates, the design can be often constrained by the specific characteristics of the building planned and the size, shape, and orientation of the building plot. Energy-efficient buildings with special designs such as orientation, insulation and windows are being appropriately adapted to withstand severe weather conditions. Natural ventilation and natural light also play an important role in energy saving. Additionally, one can have buildings with walls composed by different materials and the consideration of daylight when evaluating buildings regarding energy performance.

1.1.Tools Used :

1. Python (3.7)
2. Tensor Flow
3. Google Colaboratory
4. Jupyter Notebook
5. Keras
6. Numpy
7. Scipy
8. Matlotlib
9. Pandas
10. Seaborn
11. Scikit Learn
12. Plotly

2.Problem Definition and Data Mining Task

Form the given dataset the task is to forecast the heating load and cooling load of buildings based on 8 characteristics of building Relative Compactness, Surface Area, Wall Area, Roof Area, Overall Height, Orientation, Glazing Area, Glazing Area Distribution.

This falls under the category of regression task. The dataset contain Input variables are 8 features and two output variables(heating load and cooling load). This falls under Multi Input and Multi Output (MIMO) problem. We can use multi layer perceptron with 2 output nodes to forecast two output variables.

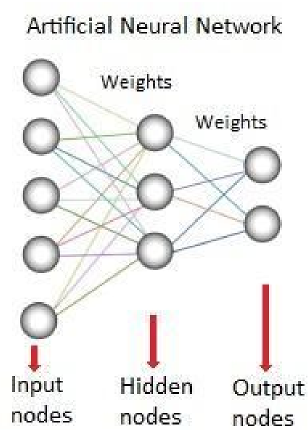
3.Techniques Used and Description

3.1.Techniques Description:

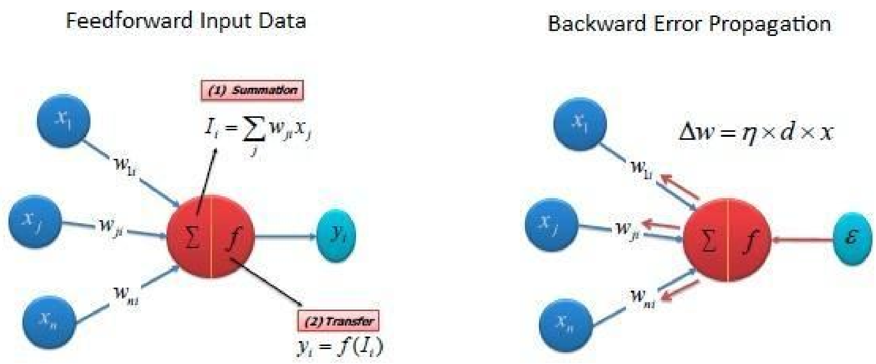
1. MLP:

A Multi Layer Perceptron(MLP) or Artificial Neural Network(ANN) is a system that is based on the biological neural network, such as the brain. The brain has approximately 100 billion neurons, which communicate through electrochemical signals. The neurons are connected through junctions called synapses. Each neuron receives thousands of connections with other neurons, constantly receiving incoming signals to reach the cell body. If the resulting sum of the signals surpasses a certain threshold, a response is sent through the axon. The ANN attempts to recreate the computational mirror of the biological neural network, although it is not comparable since the number and complexity of neurons and the used in a biological neural network is many times more than those in an artificial neural network.

Artificial Neural Network



An ANN is comprised of a network of artificial neurons (also known as "nodes"). These nodes are connected to each other, and the strength of their connections to one another is assigned a value based on their strength: inhibition (maximum being -1.0) or excitation (maximum being +1.0). If the value of the connection is high, then it indicates that there is a strong connection. Within each node's design, a transfer function is built in. There are three types of neurons in an ANN, input nodes, hidden nodes, and output nodes.



The input nodes take in information, in the form which can be numerically expressed. The information is presented as activation values, where each node is given a number, the higher the number, the greater the activation. This information is then passed throughout the network. Based on the connection strengths (weights), inhibition or excitation, and transfer functions, the activation value is passed from node to node. Each of the nodes sums the activation values it receives; it then modifies the value based on its transfer function. The activation flows through the network, through hidden layers, until it reaches the output nodes. The output nodes then reflect the input in a meaningful way to the outside world. The difference between predicted value and actual value (error) will be propagated backward by apportioning them to each node's weights according to the amount of this error the node is responsible for.

3.2. Activation Functions:

Activation functions are really important for a Artificial Neural Network to learn. Non-linear complex functional mappings between the inputs and response variable. They introduce non-linear properties to our Network. Their main purpose is to convert a input signal of a node in a neural network to an output signal. That output signal now is used as a input in the next layer in the stack.

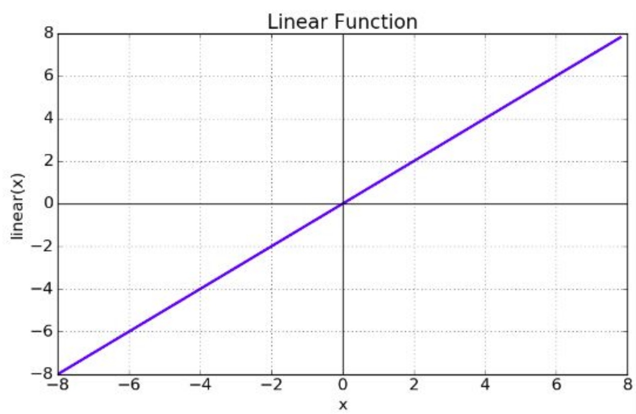
The Fundamental properties of any activation functions is:

Differential: Change in y-axis w.r.t. change in x-axis.It is also known as slope.

Monotonic function: A function which is either entirely non-increasing or non-decreasing.

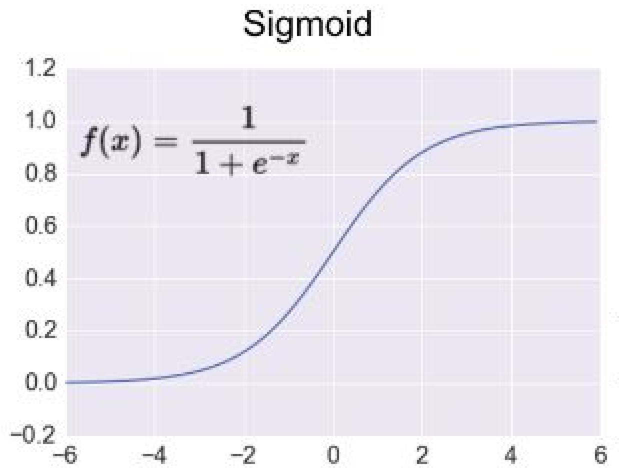
Some important activation functions are:

1. Linear:
A straight line function where activation is proportional to input.



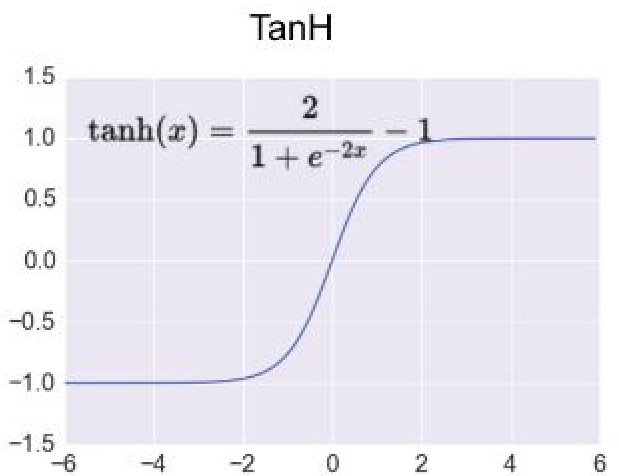
2. Sigmoid:

It is a activation function of form f(x) = 1 / 1 + exp(-x) . Its Range is between 0 and 1. It is a S — shaped curve. It is easy to understand and apply.



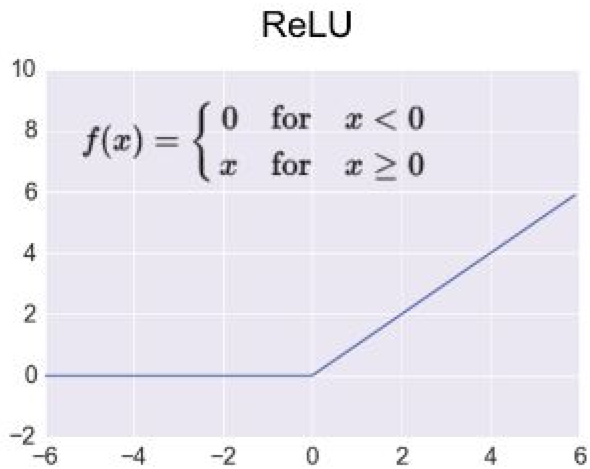
3. tanh:

tanh is also like logistic sigmoid but better. The range of the tanh function is from (-1 to 1). tanh is also sigmoidal (s - shaped).



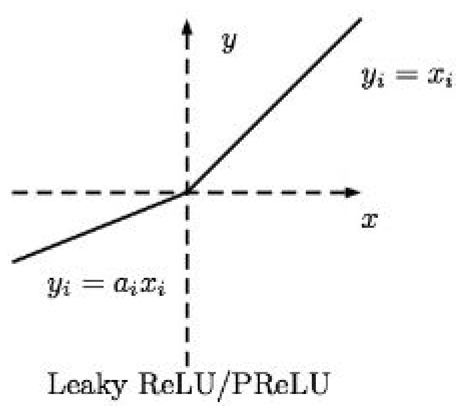
4. <u>REctifier Linear Unit (RELU):</u>

The Rectified Linear Unit has become very popular in the last few years. It computes the function *f*(*x*)=max(0,*x*)f(x)=max(0,x). In other words, the activation is simply thresholded at zero.



5. <u>Leaky RELU:</u>

Leaky ReLUs are one attempt to fix the "dying ReLU" problem. Instead of the function being zero when x < 0, a leaky ReLU will instead have a small negative slope (of 0.01, or so). Where a_i is a small constant.



3.3. Metrics:

To measure the performance of the regression model we have some metrics. Here are some of the metrics

1. Mean Square Error (MSE): Mean Square Error refers to the mean of the squared difference between the predicted parameter and the observed parameter.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

2. Root Mean Square Error (RMSE): Root Mean Square Error is the square root of mean square error.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

3. Normalized Root Mean Square Error: NRMSE measures the average magnitude of the error. It's the square root of the average of differences between maximum and minimum of actual observations.

$$NRMSE = \frac{RMSE}{\max_i y_i - \min_i y_i}$$

4. Mean Absolute Percentage Error: The MAPE (Mean Absolute Percent Error) measures the size of the error in percentage terms. It is calculated as the average of the unsigned percentage error.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| * 100$$

5. Symmetric Mean Absolute Percentage Error: The SMAPE (Symmetric Mean Absolute Percentage Error) is a variation on the MAPE that is calculated using the average of the absolute value of the actual and the absolute value of the forecast in the denominator.

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i - \hat{y}_i}{\frac{y_i + \hat{y}_i}{2}}\right] * 100$$

6. : Median Absolute Error: The median absolute deviation(MAD) is a robust measure of how spread out a set of data is.

$$MAD = \sum_{i=1}^{n}\frac{|y_i - \bar{y}|}{n}$$

7. <u>R^2 value</u>: The R^2 (or R Squared) metric provides an indication of the goodness of fit of a set of predictions to the actual values. In statistical literature, this measure is called the coefficient of determination.

This is a value between 0 and 1 for no-fit and perfect fit respectively.

$$\hat{R}^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

4.Dataset Description

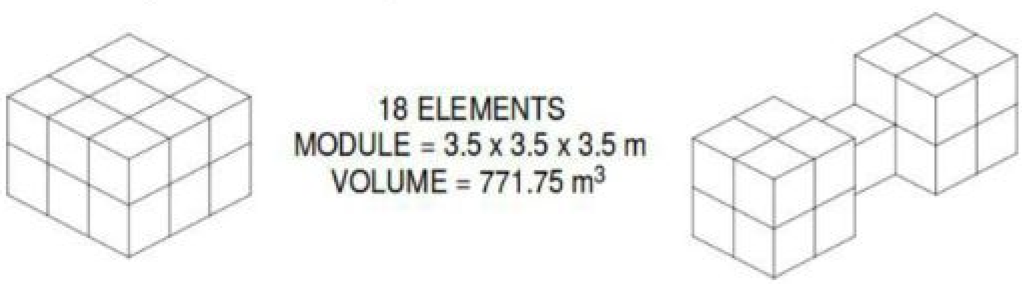
- Dataset is downloaded from below link as a CSV file.

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

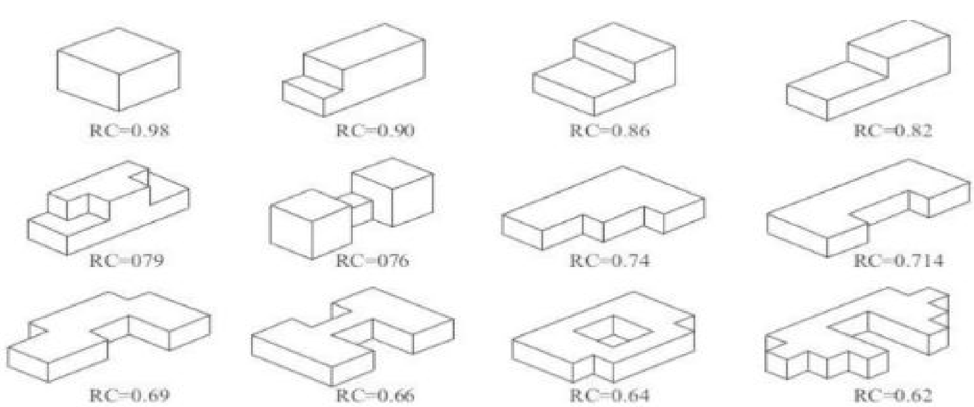
- Each row represents a structure of building, each column contains building attributes described on the column Metadata.
- The raw data contains 768 rows and 10 columns (features).
- We have two target functions "Heating Load" and "Cooling Load".

4.1.Feature Description

A modular geometry system was derived based on an elementary cube (3.5 × 3.5 × 3.5m). In order to generate different building shapes, eighteen such elements were used according to *Figure*.



A subset of twelve shapes with distinct relative compactness values was selected for the simulations, as shown in *figure*.



The buildings differ with respect to attributes such as glazing area, the glazing area distribution, and the orientation, amongst other parameters. The data set simulates various settings as functions of 8 attributes (x1 to x8) resulting in 768 samples and two real valued responses which are heating and cooling load (y1 and y2).

X1 - Relative Compactness,

X2 - Surface Area,

X3 - Wall Area,

X4 - Roof Area,

X5 - Overall Height,

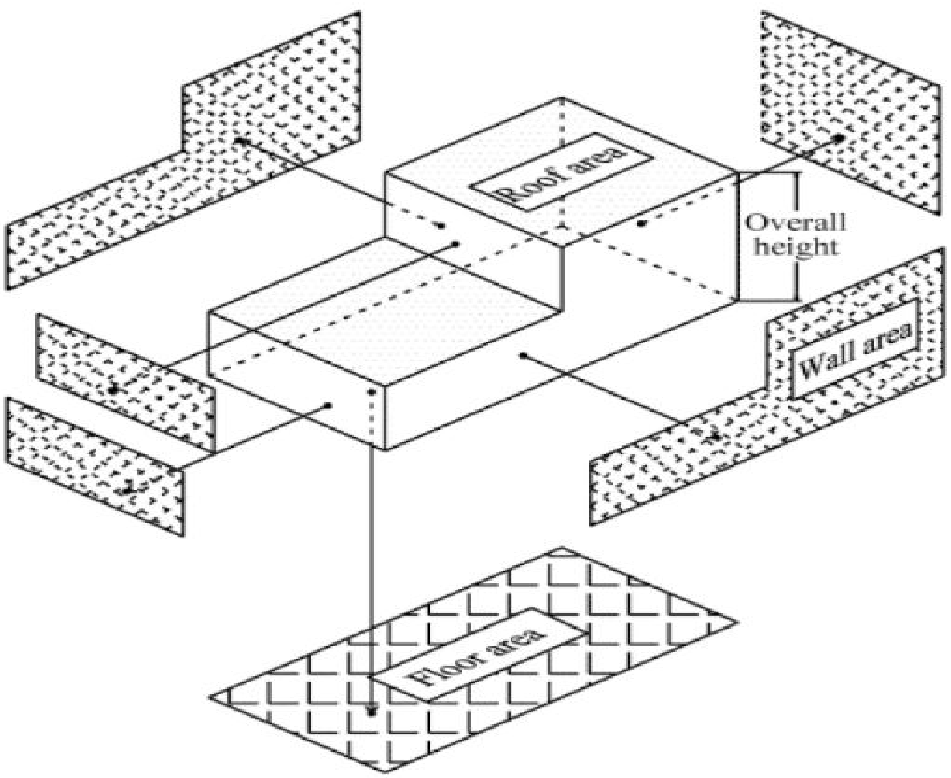
X6 - Orientation,

X7 - Glazing Area,

X8 - Glazing Area Distribution,

y1 - Heating Load,

y2 - Cooling Load



5.Descriptive Analytics

Distribution Plots : These plots helps to visualize the underlying probability distribution function of the data. All Numerical variables can be plotted with distribution plots.

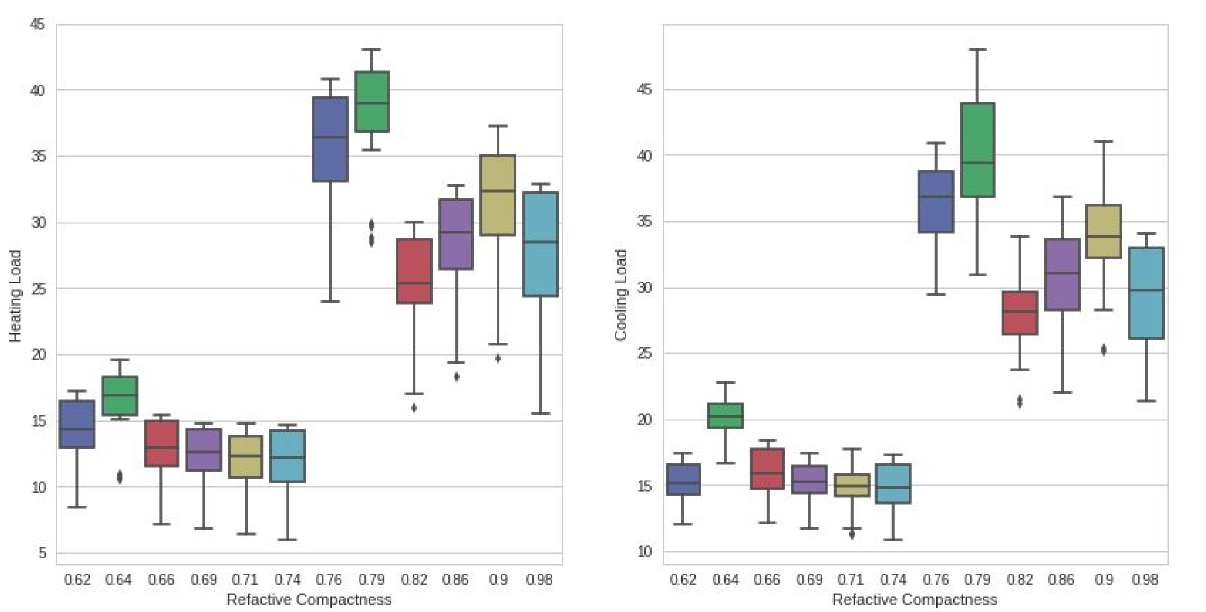
Violin Plots : A violin plot is a method of plotting numeric data. It is similar to a box plot with a rotated kernel density plot on each side.

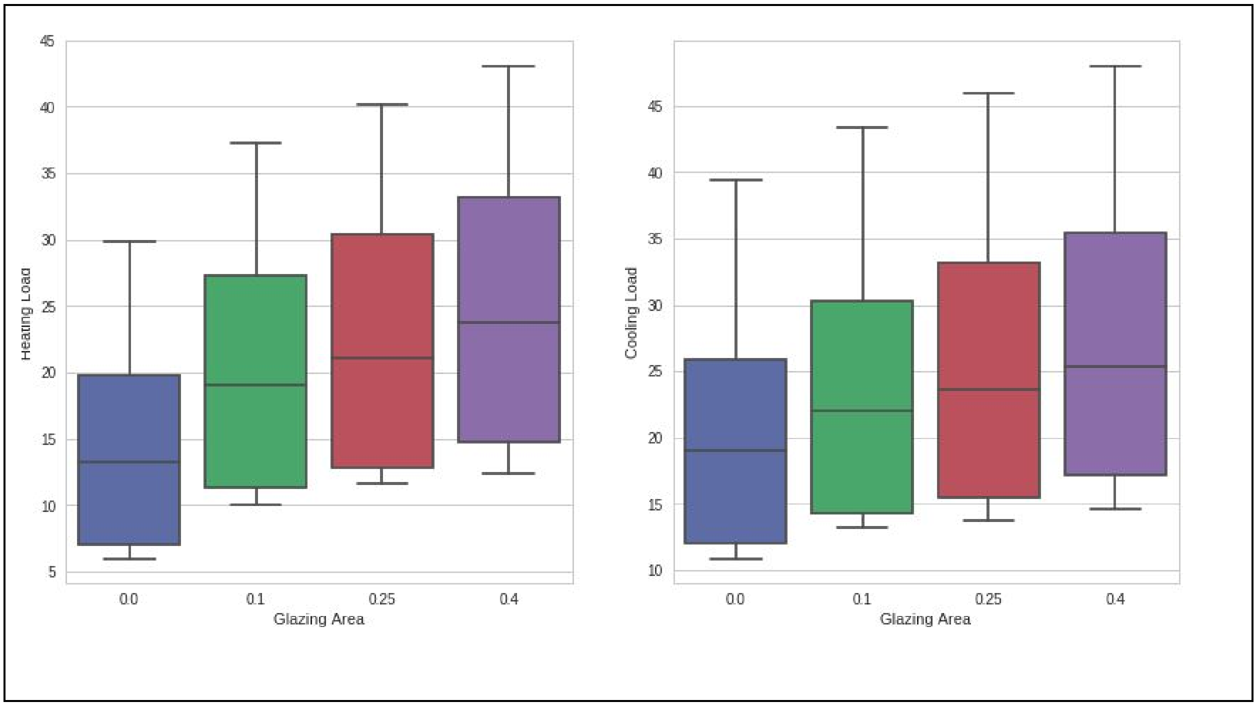
Pair Plots : Pair plots give clear understanding about 3 or more numerical variables.

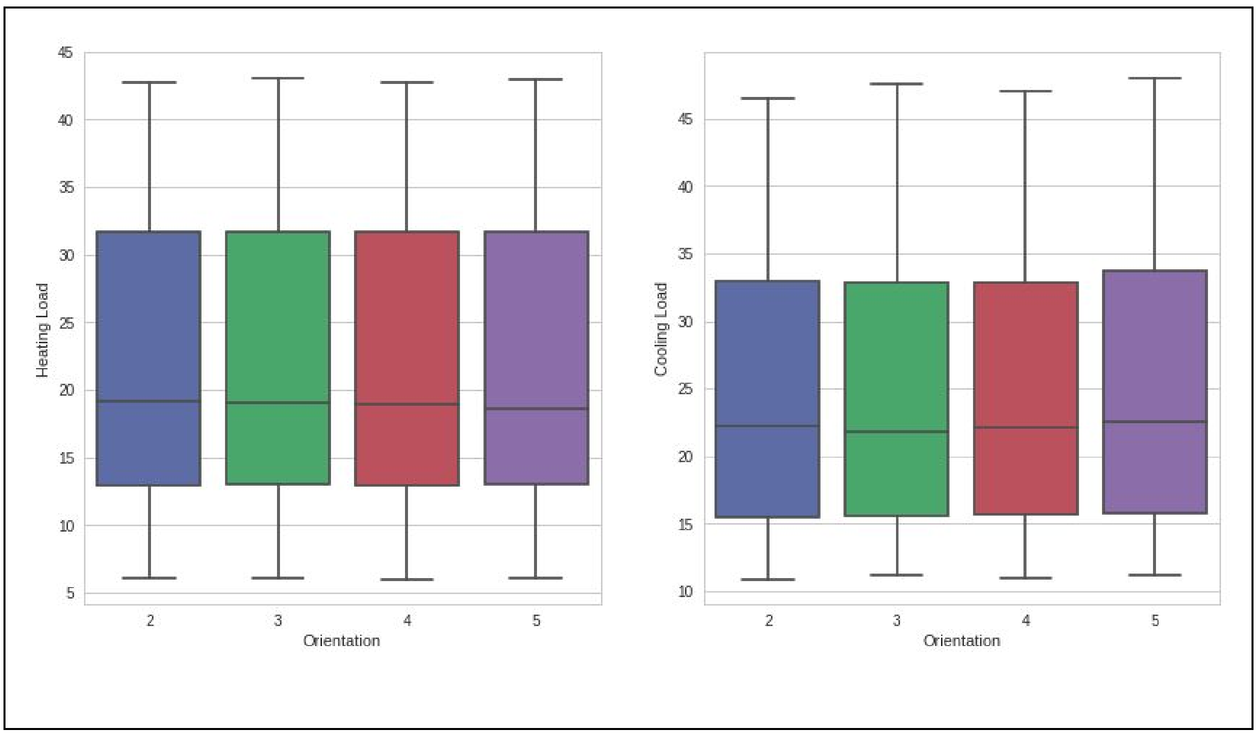
Correlation Matrix : A correlation matrix is a table showing correlation coefficients(Pearson Correlation Coefficient) between sets of variables. Each random variable(Xi) in the table is correlated with each of the other values in the table (Xj). This allows you to see which pairs have the highest correlation.

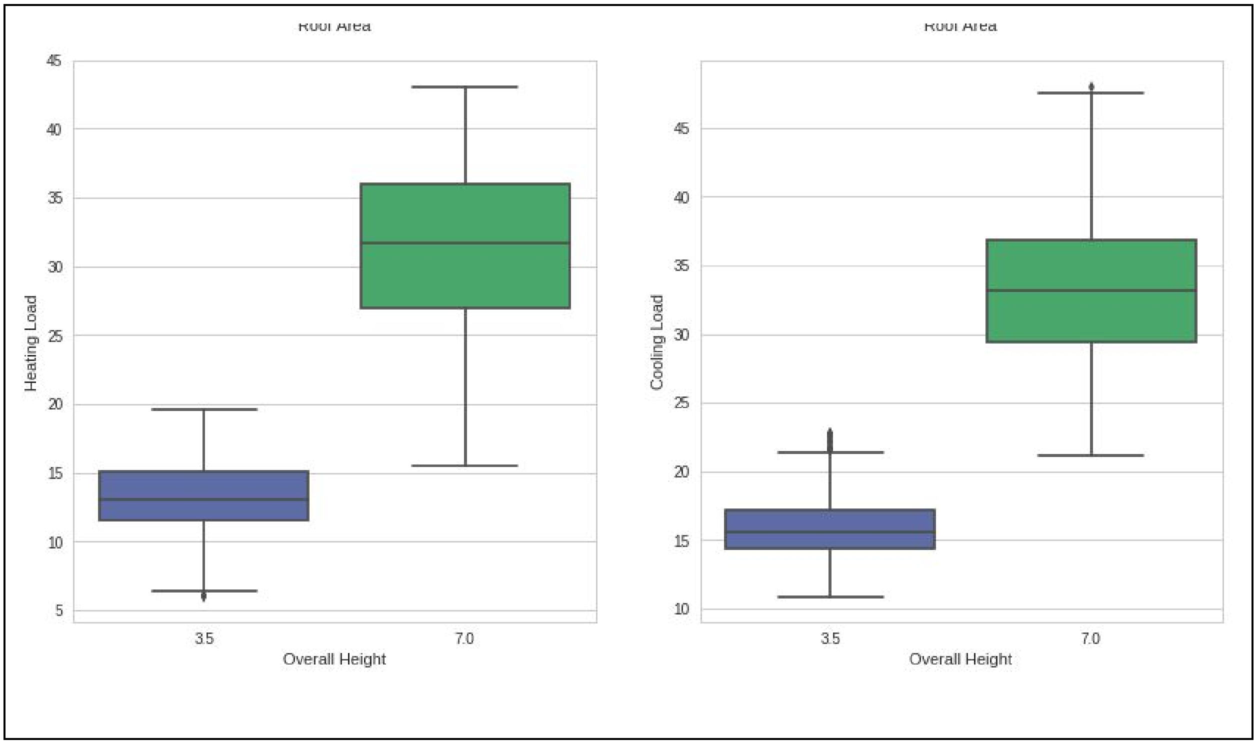
5.1.Univariate Analysis:

5.1.1 Box Plots for each feature with heating load and cooling load:

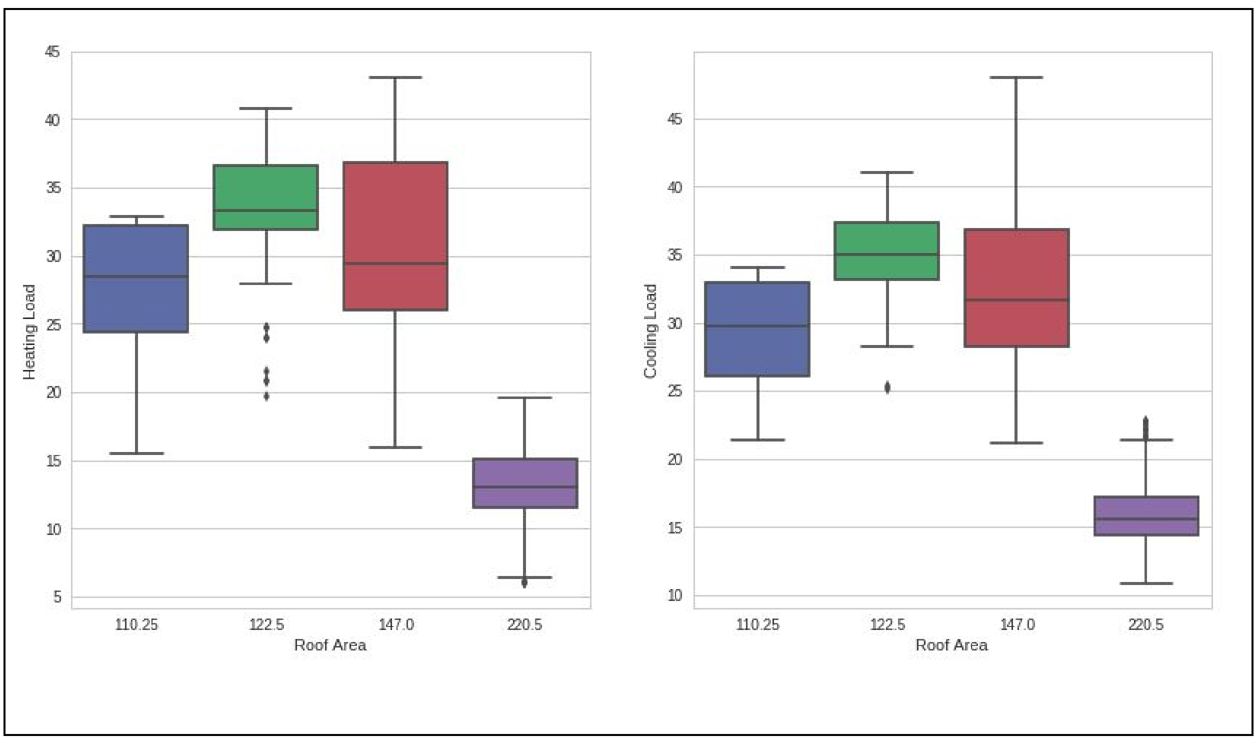



Heating Load
Glazing Area
0.0
0.1
0.25
0.4
Cooling Load
Glazing Area

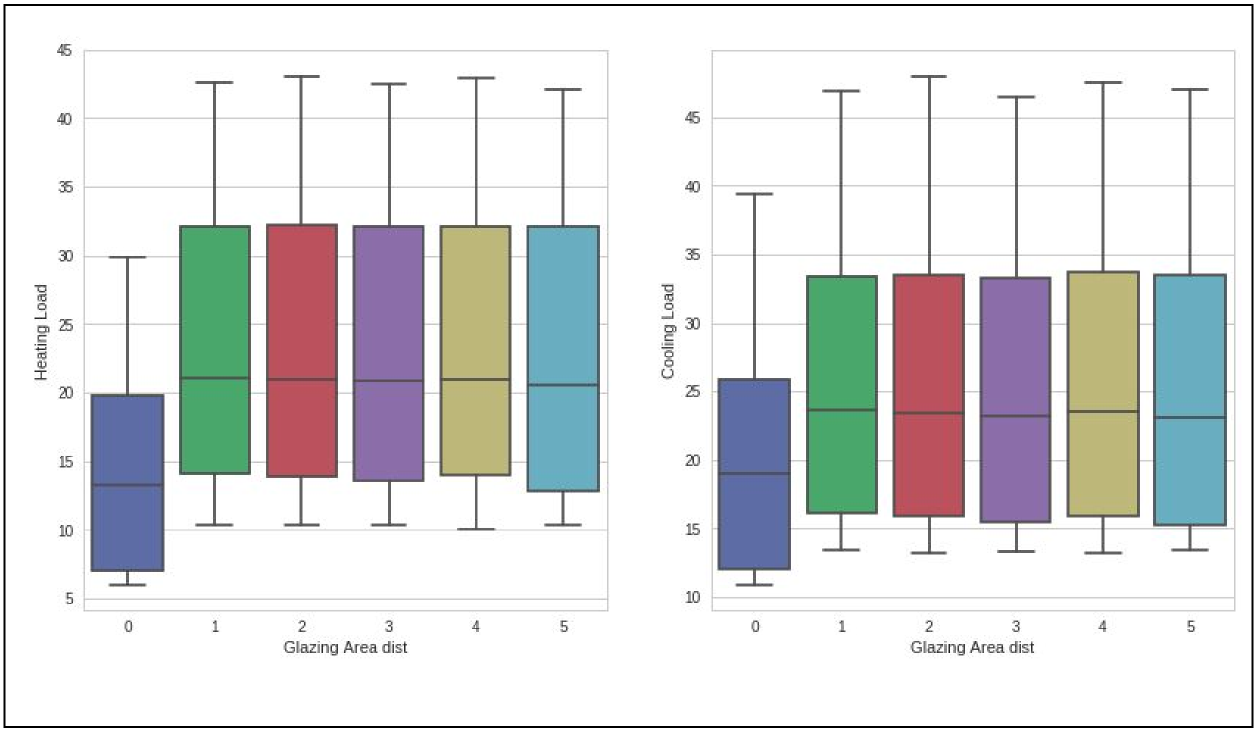

Heating Load
Orientation
2
3
4
5
Cooling Load
Orientation

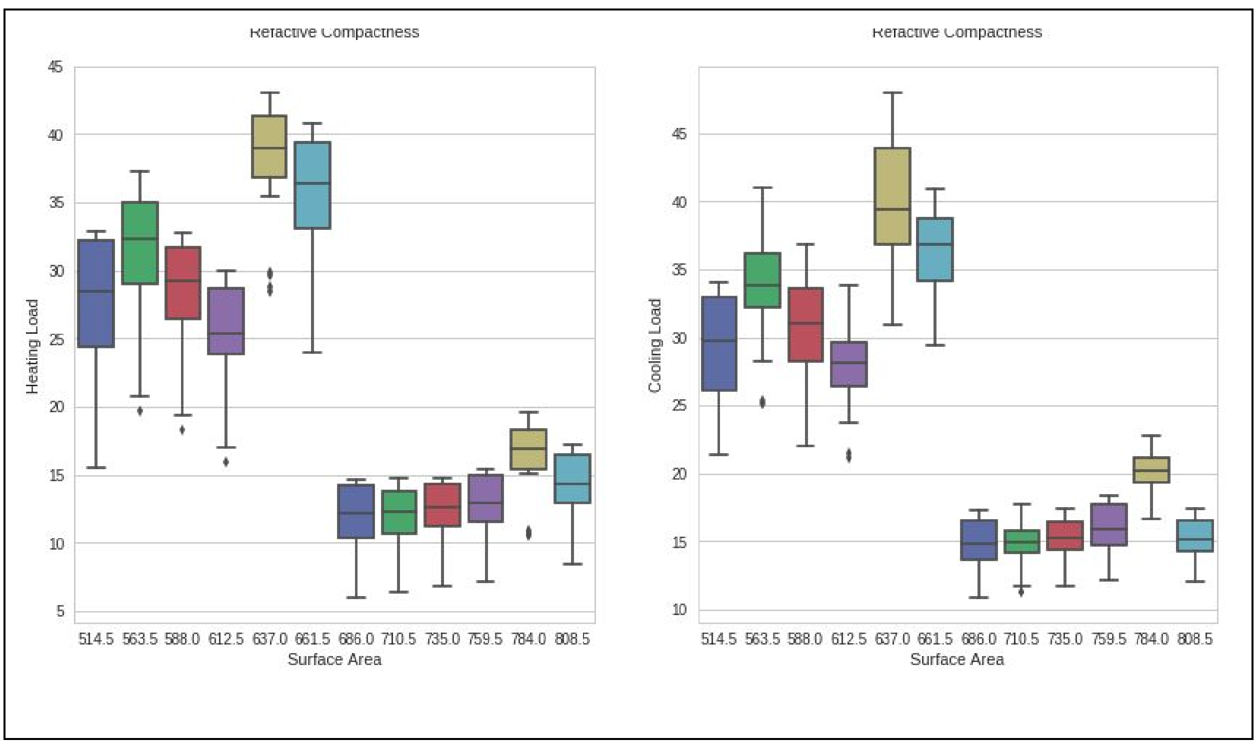


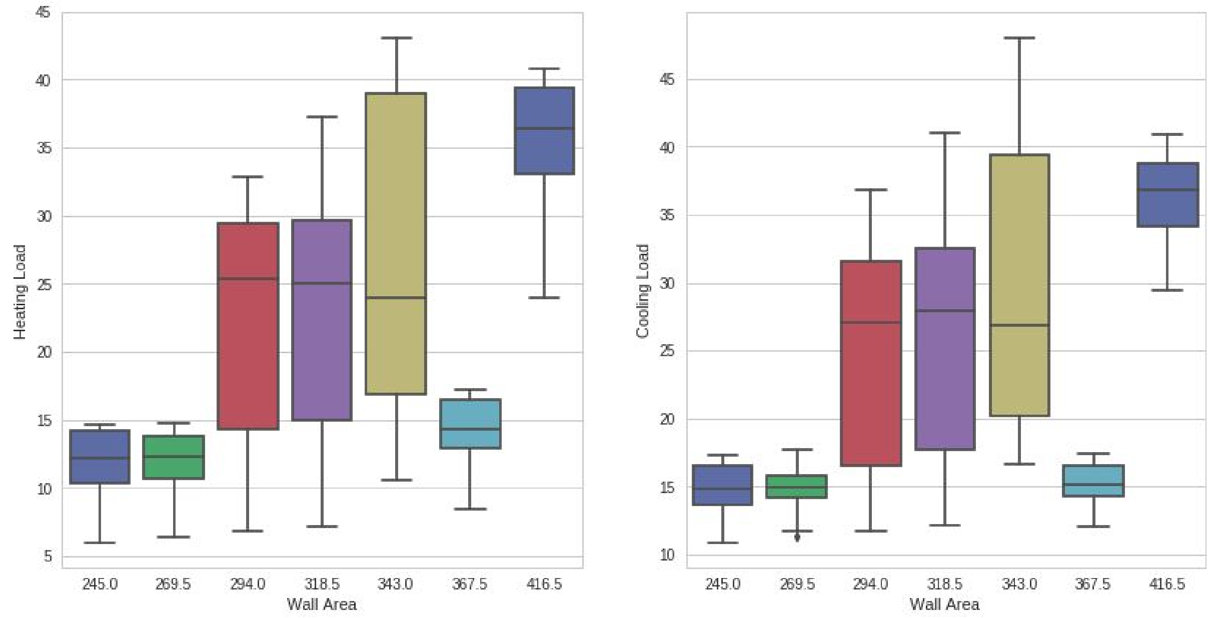
Roof Area
Heating Load
Overall Height
3.5
7.0
Roof Area
Cooling Load
Overall Height
3.5
7.0



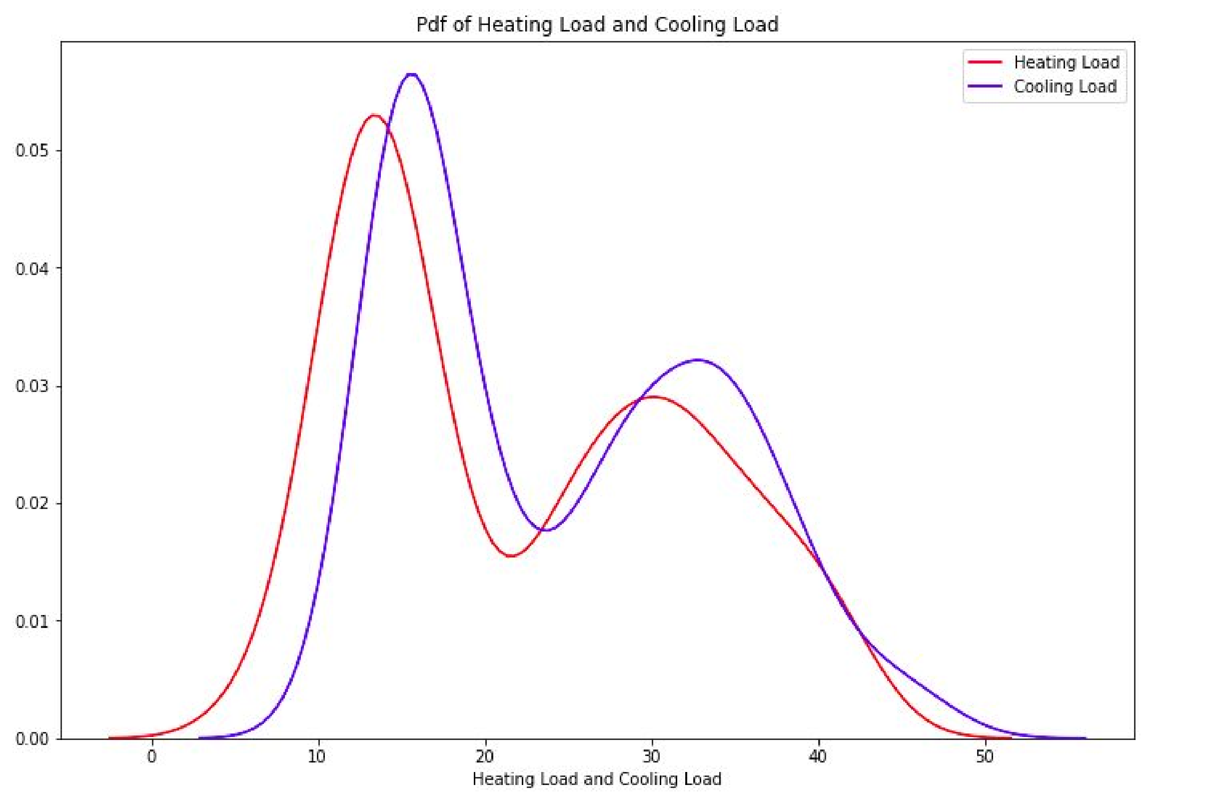
Heating Load
Roof Area
110.25
122.5
147.0
220.5
Cooling Load
Roof Area
110.25
122.5
147.0
220.5


Heating Load
Glazing Area dist
Cooling Load
Glazing Area dist

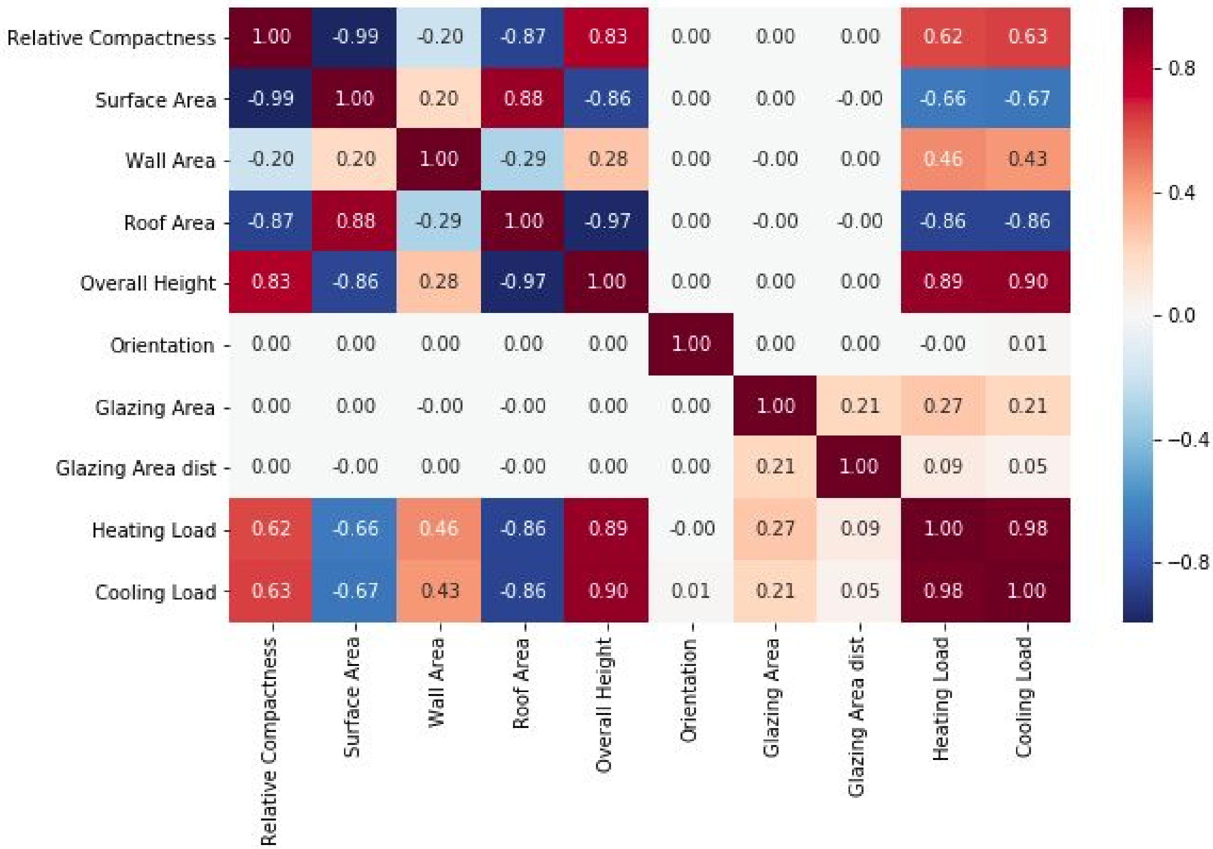

Refactive Compactness
Heating Load
Surface Area
Refactive Compactness
Cooling Load
Surface Area



5.1.2.PDF's of heating load and cooling load:



5.3. Correlation Matrix :



Conclusions from Descriptive analysis:

- From above correlation matrix we can say that Relative Compactness and Surface Area have a correlation of -0.99.
- If the Overall Height of a structure is 3.5 then Cooling Load is less than 22.
- If the Relative Compactness of a structure is less than 0.75 then cooling load is always less than 22.

6.Data Preprocessing

6.1 Data Cleansing:

6.1.1 Handling Missing Values:

There are no missing values in the dataset.

6.1.2 Data Preparation and Transformations :

Dataset has 10 features out of which 8 features are input variables and 2 output variable.

Feature	Number of Unique Values
Relative Compactness	12
Surface Area	12
Wall Area	7
Roof Area	4
Overall Height	2
Orientation	4
Glazing Area	4
Glazing Area dist	6
Heating Load	586
Cooling Load	636

Output variables heating load and cooling load are continuous variables, so this is a regression task.
Output variables:
1. Heating Load
2. Cooling Load

From above table we can conclude that we got all input variables are categorical features. But they have got some ordinal nature.

6.2. <u>Data Normalization:</u>

All the input and output variables are numerical features so it is good to normalize data on both input and output features. We performed data normalization using min-max scaler.

6.3. <u>Feature Selection:</u>

Selecting Important Features is very important step in data mining task. Feature importance reduces the number of input variables. Reduced features can perform better than all variables together.

In this project we performed feature selection with

1. Domain Knowledge
2. Filter Based Feature Selection

6.3.1.1 <u>Domain Knowledge:</u>

From the domain knowledge Relative Compactness of a structure is calculated by given formula.

$$RC = \frac{(V / A)_{building}}{(V / A)_{ref}}$$

Where -
RC = Relative Compactness
V = Voluma of the Structure
A = Surface Area of the Structure

All the given structures in dataset have same reference structure.Also volume of each structure in the given dataset is same volume, which is 771.75 m3,. So we can say that Relative Compactness(RC) of a shape is inversely directly proportional to surface area. Both Relative Compactness(RC) and surface area are perfectly negatively correlated to each other. So we can remove either RC or surface area. Now we have 7 input features.

6.3.1.2 <u>Filter Based :</u>

After removing relative compactness feature we passed dataset with 7 input features to forward feature selection with linear regression we got 4 most important features.

6.4 Data Partitioning Methods :

We used two data partition methods to split data into Train and Test. For both partition techniques we used stratified random sampling.

6.4.1 Hold Out:

Partitioned the original dataset into train and test.

- Train - 80%
- Test - 20%

6.4.2 5 Fold CV:

Since the dataset is very small we applied 5 fold cross validation on dataset where 4 folds are involved in training and 1 fold for testing.

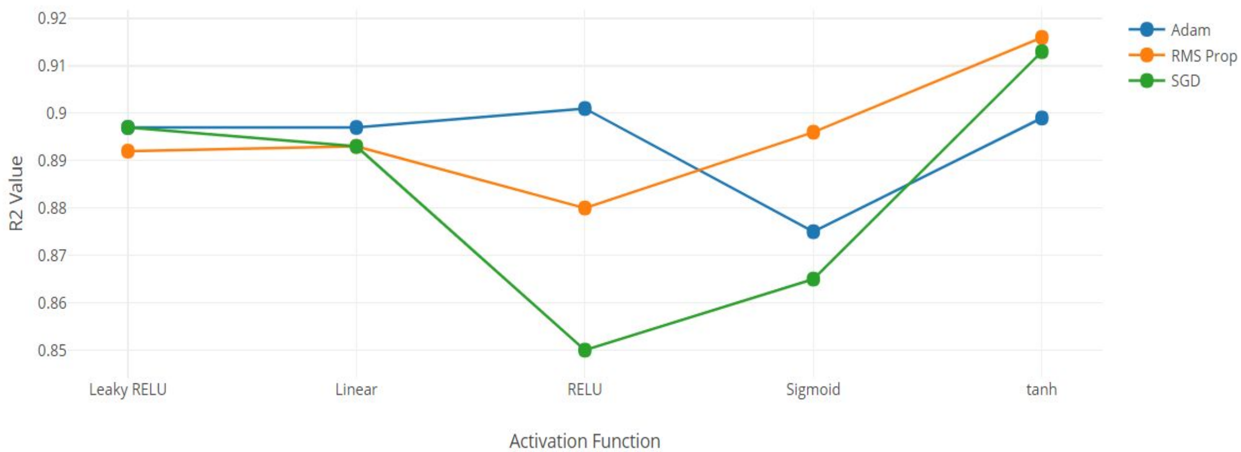
7. Results

1. Table

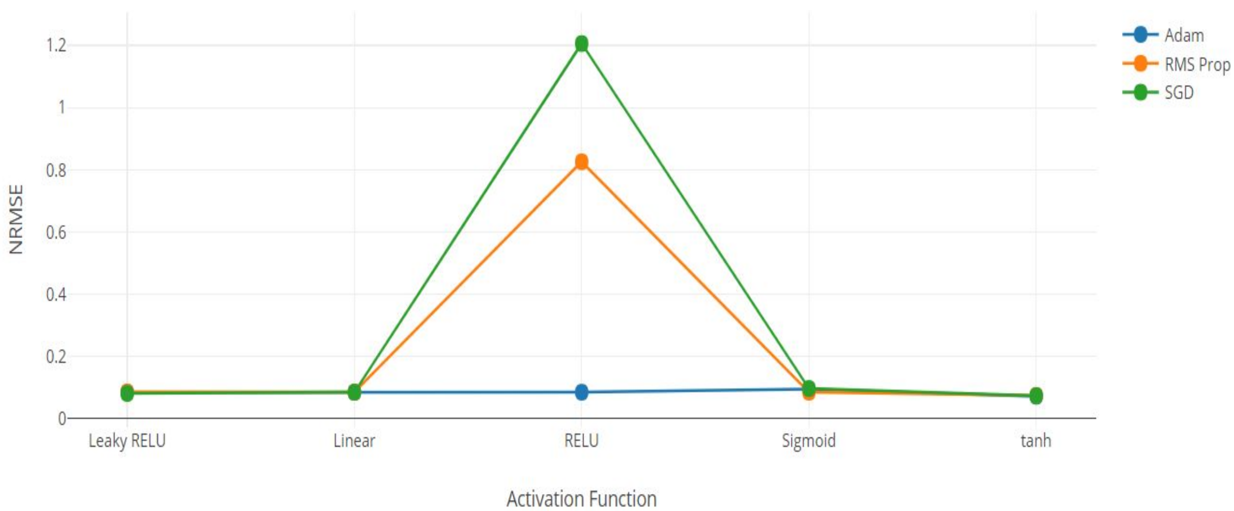
- Learning Rate is 0.001
- With 1 Hidden layer
- Hold Out with Train and Test Split (80-20)

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	1.998	0.093	8.43	0.084	0.897	2.903	0.092
Linear	Adam	2.084	0.103	8.626	0.085	0.897	2.937	0.1
RELU	Adam	1.976	0.096	8.005	0.085	0.901	2.829	0.094
Sigmoid	Adam	2.327	0.111	10.259	0.095	0.875	3.203	0.109
tanh	Adam	2.121	0.119	8.407	0.072	0.899	2.9	0.106
Leaky RELU	RMS Prop	2.062	0.094	8.747	0.086	0.892	2.958	0.094
Linear	RMS Prop	2.181	0.098	8.981	0.085	0.893	2.997	0.1
RELU	RMS Prop	8.909	0.375	94.671	0.826	0.88	9.73	0.405
Sigmoid	RMS Prop	1.993	0.092	8.41	0.085	0.896	2.9	0.091
tanh	RMS Prop	1.72	0.078	6.746	0.075	0.916	2.597	0.076
Leaky RELU	SGD	2.004	0.091	8.373	0.081	0.897	2.894	0.09
Linear	SGD	2.101	0.1	8.799	0.085	0.893	2.966	0.098
RELU	SGD	8.922	0.377	94.776	1.206	0.85	9.735	0.406
Sigmoid	SGD	2.393	0.11	11.2	0.097	0.865	3.347	0.109
tanh	SGD	1.805	0.084	7.038	0.074	0.913	2.653	0.082

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**

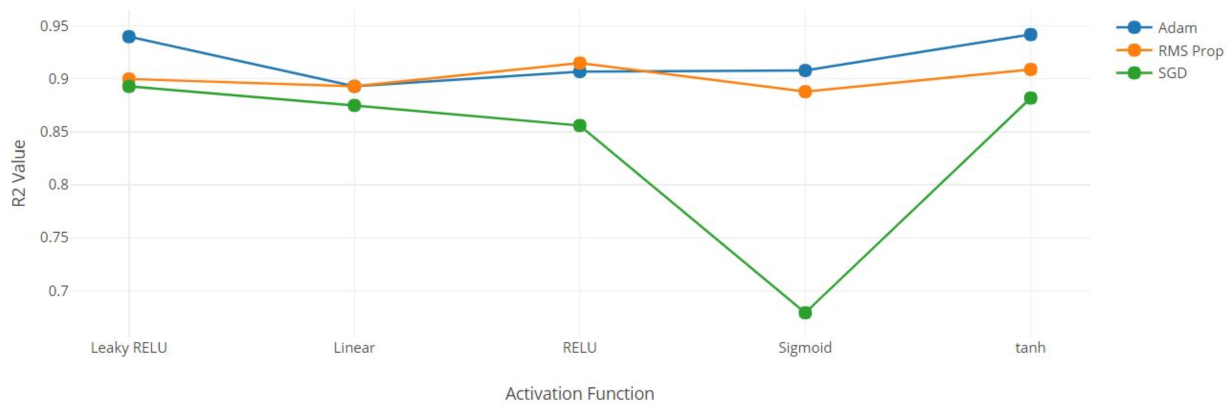


2. Table

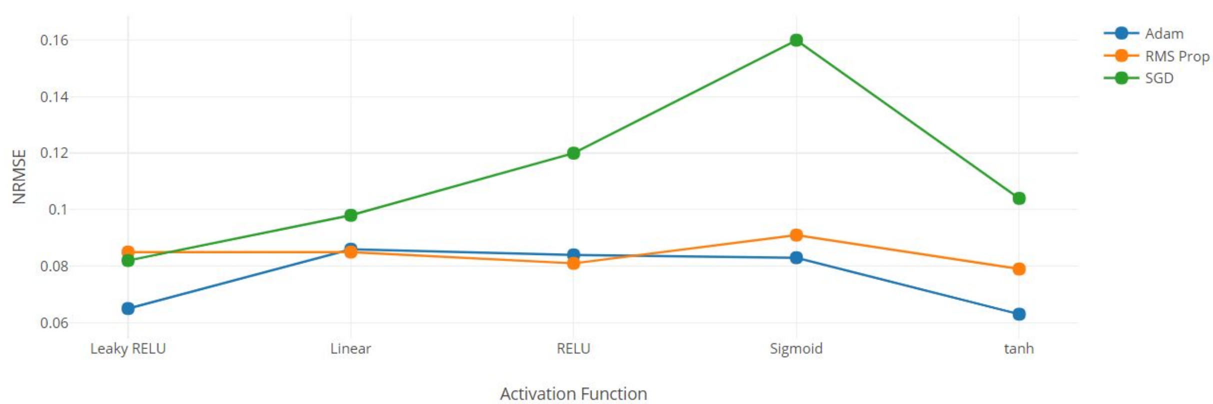
- Learning Rate is 0.001
- With two Hidden layers
- Hold Out with Train and Test Split (80-20)

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	1.652	0.079	5.08	0.065	0.94	2.254	0.079
Linear	Adam	2.103	0.103	8.693	0.086	0.893	2.948	0.1
RELU	Adam	1.823	0.083	7.412	0.084	0.907	2.722	0.082
Sigmoid	Adam	1.855	0.084	7.271	0.083	0.908	2.696	0.084
tanh	Adam	1.566	0.071	4.808	0.063	0.942	2.193	0.074
Leaky RELU	RMS Prop	1.974	0.088	8.08	0.085	0.9	2.843	0.089
Linear	RMS Prop	2.087	0.102	8.685	0.085	0.893	2.947	0.1
RELU	RMS Prop	1.885	0.086	7.101	0.081	0.915	2.665	0.086
Sigmoid	RMS Prop	2.143	0.102	9.161	0.091	0.888	3.027	0.099
tanh	RMS Prop	1.772	0.081	7.34	0.079	0.909	2.709	0.079
Leaky RELU	SGD	2.099	0.103	8.774	0.082	0.893	2.962	0.1
Linear	SGD	2.304	0.107	10.217	0.098	0.875	3.196	0.107
RELU	SGD	2.546	0.113	12.151	0.12	0.856	3.486	0.115
Sigmoid	SGD	3.256	0.147	19.055	0.16	0.679	4.365	0.151
tanh	SGD	2.12	0.098	9.338	0.104	0.882	3.056	0.098

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**

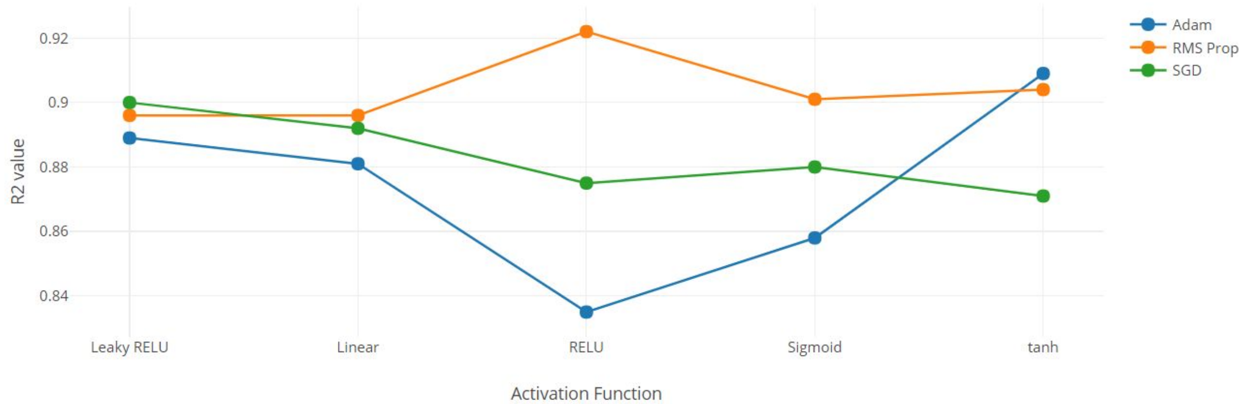


3. Table

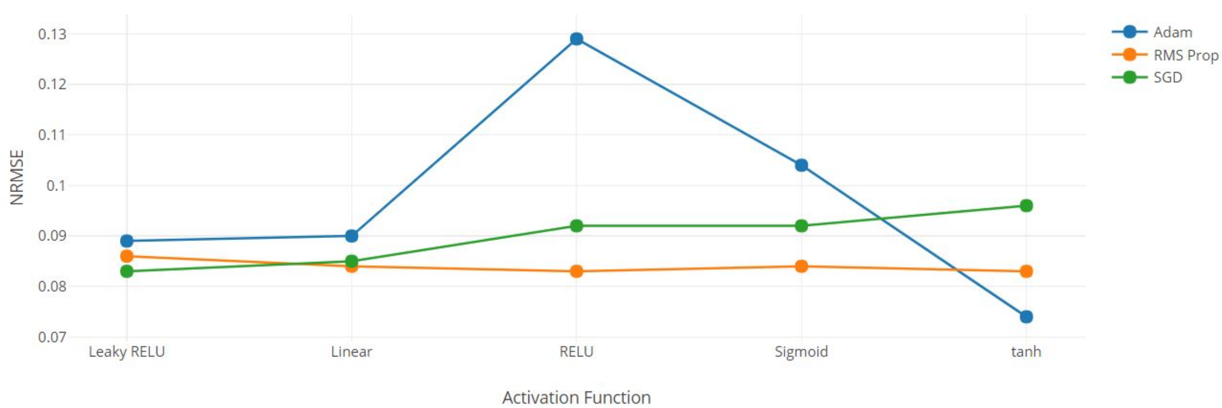
- Learning Rate is 0.01
- With 1 Hidden Layer
- Hold Out with Train and Test Split (80-20)

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	2.172	0.109	9.123	0.089	0.889	3.02	0.106
Linear	Adam	2.211	0.103	9.781	0.09	0.881	3.127	0.101
RELU	Adam	2.57	0.119	12.222	0.129	0.835	3.496	0.122
Sigmoid	Adam	2.442	0.109	11.805	0.104	0.858	3.436	0.108
tanh	Adam	1.819	0.082	7.332	0.074	0.909	2.708	0.081
Leaky RELU	RMS Prop	2.152	0.102	8.803	0.086	0.896	2.967	0.102
Linear	RMS Prop	2.076	0.102	8.555	0.084	0.896	2.925	0.099
RELU	RMS Prop	1.769	0.087	6.215	0.083	0.922	2.493	0.087
Sigmoid	RMS Prop	1.988	0.088	8.058	0.084	0.901	2.839	0.088
tanh	RMS Prop	1.826	0.085	7.653	0.083	0.904	2.766	0.084
Leaky RELU	SGD	1.886	0.087	8.012	0.083	0.9	2.831	0.086
Linear	SGD	2.112	0.1	8.769	0.085	0.892	2.961	0.099
RELU	SGD	2.155	0.092	10.309	0.092	0.875	3.211	0.091
Sigmoid	SGD	2.277	0.11	9.747	0.092	0.88	3.122	0.108
tanh	SGD	2.313	0.105	10.594	0.096	0.871	3.255	0.104

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**

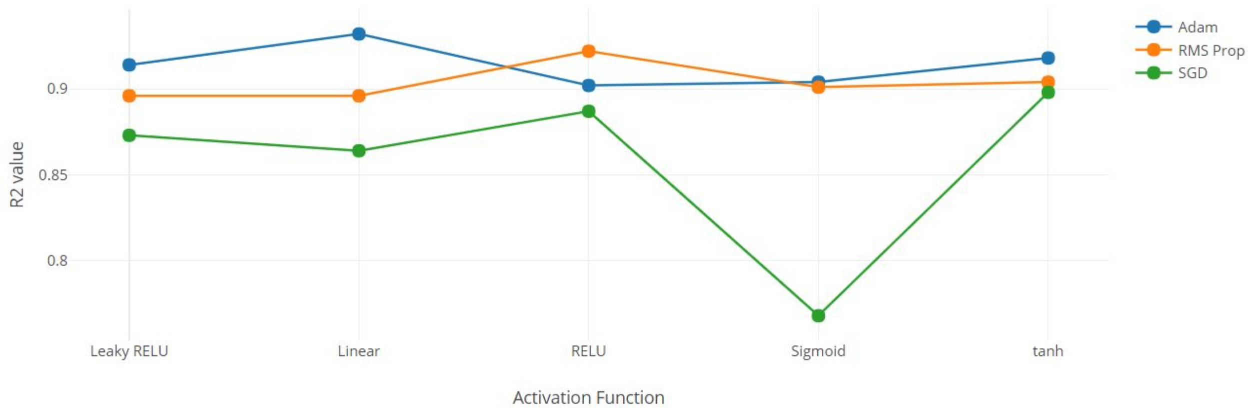


4. Table

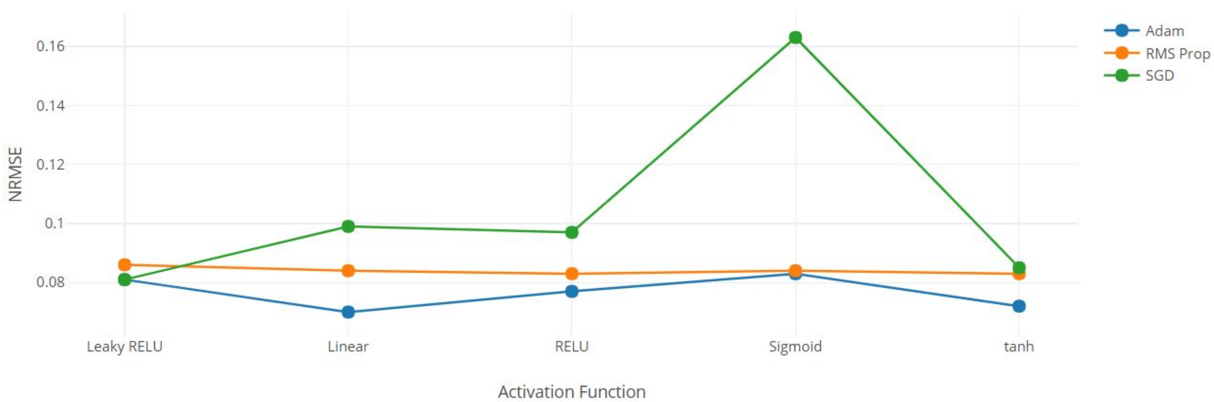
- Learning Rate is 0.01
- With 2 Hidden Layers
- Hold Out with Train and Test Split (80-20)

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	1.849	0.083	7.148	0.081	0.914	2.674	0.083
Linear	Adam	1.727	0.082	5.345	0.07	0.932	2.312	0.083
RELU	Adam	1.782	0.08	7.789	0.077	0.902	2.791	0.077
Sigmoid	Adam	2.097	0.095	7.707	0.083	0.904	2.776	0.099
tanh	Adam	1.666	0.078	6.575	0.072	0.918	2.564	0.075
Leaky RELU	RMS Prop	2.152	0.102	8.803	0.086	0.896	2.967	0.102
Linear	RMS Prop	2.076	0.102	8.555	0.084	0.896	2.925	0.099
RELU	RMS Prop	1.769	0.087	6.215	0.083	0.922	2.493	0.087
Sigmoid	RMS Prop	1.988	0.088	8.058	0.084	0.901	2.839	0.088
tanh	RMS Prop	1.826	0.085	7.653	0.083	0.904	2.766	0.084
Leaky RELU	SGD	2.291	0.105	10.164	0.081	0.873	3.188	0.104
Linear	SGD	2.414	0.113	11.044	0.099	0.864	3.323	0.115
RELU	SGD	2.1	0.097	9.149	0.097	0.887	3.025	0.097
Sigmoid	SGD	3.035	0.137	17.189	0.163	0.768	4.146	0.139
tanh	SGD	2.084	0.099	8.505	0.085	0.898	2.916	0.097

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**

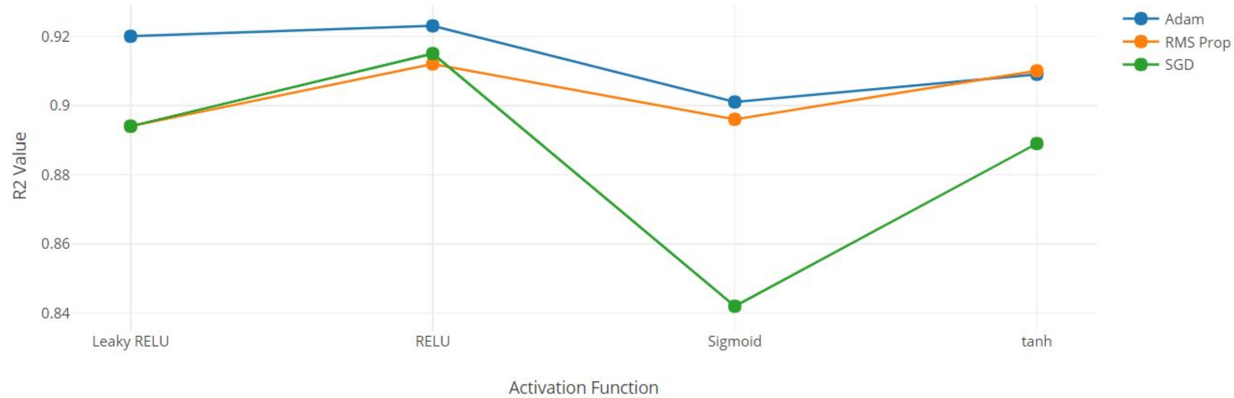


5. Table

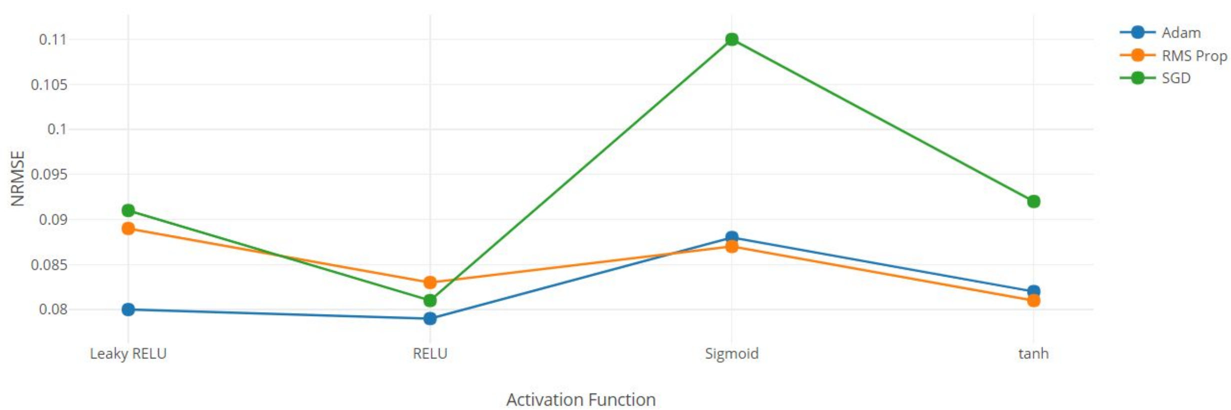
- Learning Rate is 0.001
- With 1 hidden layer
- 5 fold CV with 7 input features

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	1.775	0.075	6.831	0.08	0.92	2.606	0.074
RELU	**Adam**	**1.832**	**0.08**	**6.662**	**0.079**	**0.923**	**2.576**	**0.081**
Sigmoid	Adam	2.016	0.084	8.407	0.088	0.901	2.888	0.084
tanh	Adam	1.907	0.081	7.625	0.082	0.909	2.753	0.081
Leaky RELU	RMS Prop	2.065	0.086	8.994	0.089	0.894	2.99	0.086
RELU	RMS Prop	1.954	0.086	7.548	0.083	0.912	2.742	0.085
Sigmoid	RMS Prop	2.152	0.09	8.951	0.087	0.896	2.987	0.091
tanh	RMS Prop	1.968	0.083	7.801	0.081	0.91	2.784	0.083
Leaky RELU	SGD	2.109	0.09	9.071	0.091	0.894	3.006	0.09
RELU	SGD	1.896	0.082	7.29	0.081	0.915	2.698	0.082
Sigmoid	SGD	2.561	0.111	12.881	0.11	0.842	3.576	0.11
tanh	SGD	2.158	0.091	9.484	0.092	0.889	3.075	0.091

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**

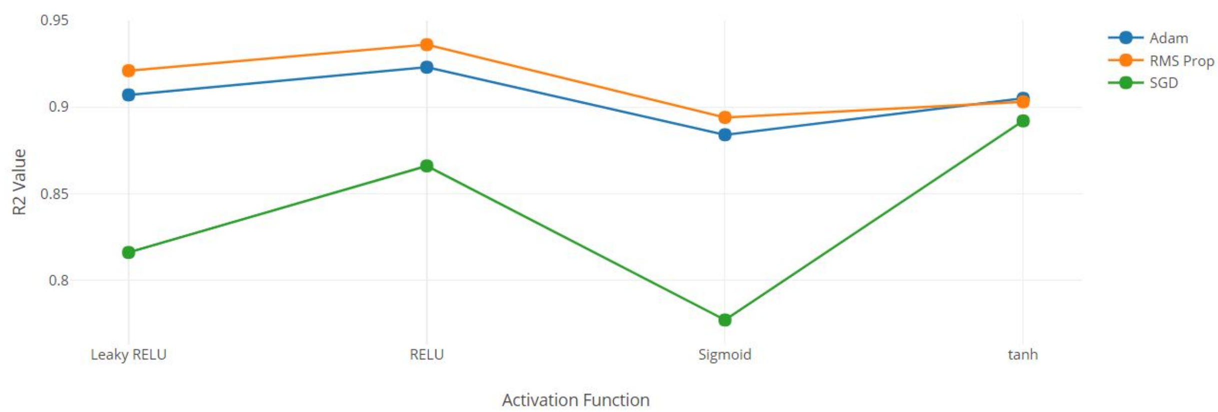


6. Table

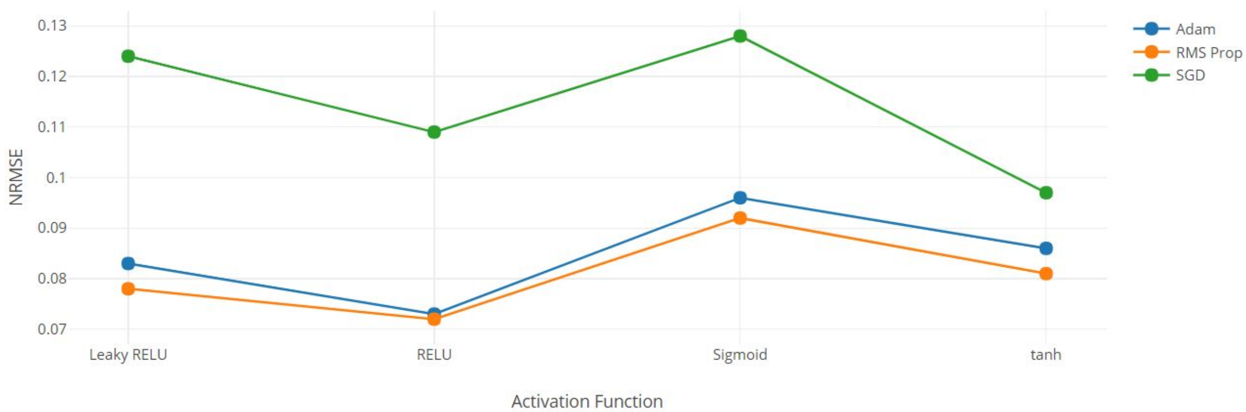
- Learning Rate is 0.001
- With 2 hidden layer
- 5 fold CV with 7 input features

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	1.917	0.083	7.893	0.083	0.907	2.796	0.082
RELU	Adam	1.824	0.08	6.565	0.073	0.923	2.555	0.08
Sigmoid	Adam	2.203	0.097	9.756	0.096	0.884	3.116	0.096
tanh	Adam	1.963	0.086	8.109	0.086	0.905	2.831	0.084
Leaky RELU	RMS Prop	1.821	0.079	6.69	0.078	0.921	2.58	0.079
RELU	RMS Prop	**1.658**	**0.074**	**5.629**	**0.072**	**0.936**	**2.345**	**0.075**
Sigmoid	RMS Prop	2.149	0.091	8.99	0.092	0.894	2.989	0.092
tanh	RMS Prop	1.898	0.08	7.842	0.081	0.903	2.779	0.079
Leaky RELU	SGD	2.752	0.118	14.915	0.124	0.816	3.836	0.119
RELU	SGD	2.375	0.103	11.114	0.109	0.866	3.323	0.103
Sigmoid	SGD	2.825	0.121	15.304	0.128	0.777	3.868	0.122
tanh	SGD	2.171	0.095	9.185	0.097	0.892	3.023	0.095

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**

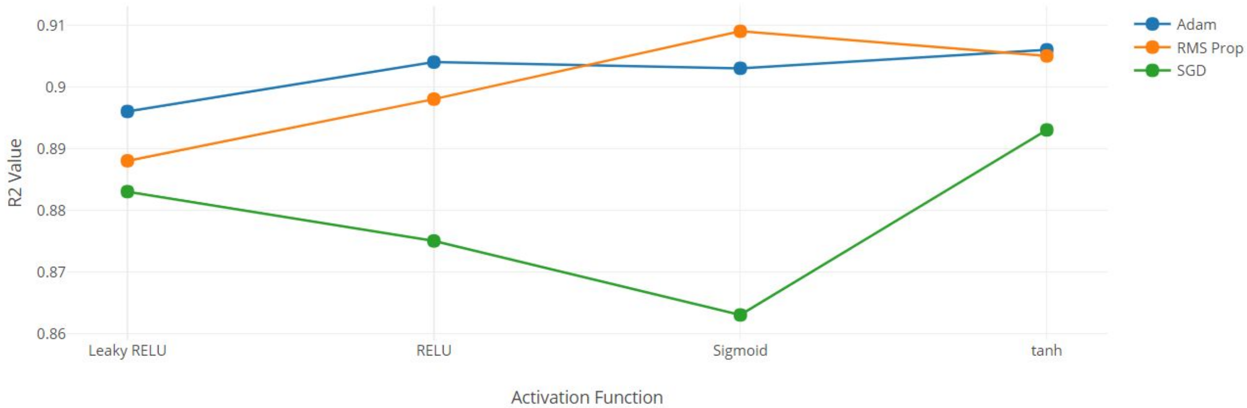


7. Table

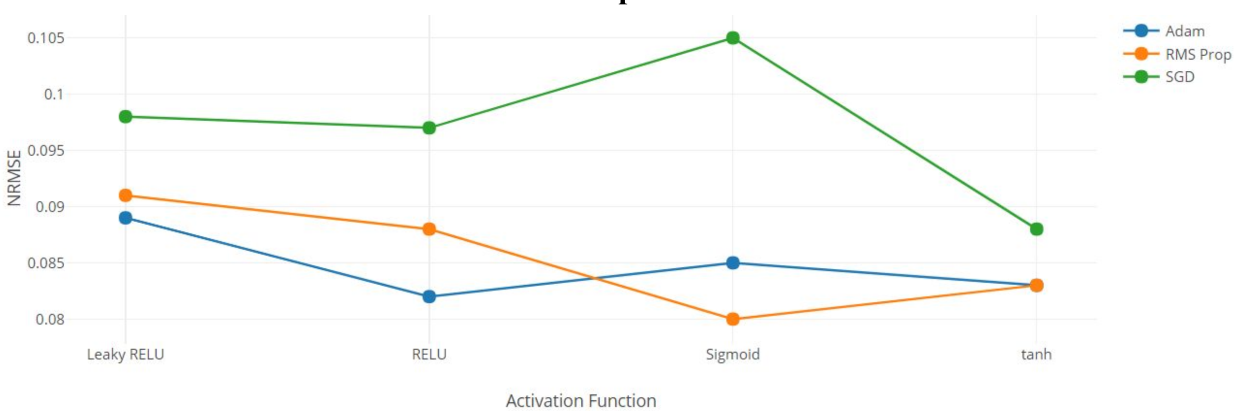
- Learning Rate is 0.01
- With 1 hidden layer
- 5 fold CV with 7 input features

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	2.053	0.085	8.877	0.089	0.896	2.971	0.085
RELU	Adam	1.983	0.081	8.312	0.082	0.904	2.872	0.08
Sigmoid	Adam	1.965	0.083	8.239	0.085	0.903	2.857	0.083
tanh	Adam	1.961	0.083	7.857	0.083	0.906	2.791	0.084
Leaky RELU	RMS Prop	2.138	0.092	9.517	0.091	0.888	3.077	0.09
RELU	RMS Prop	2.017	0.084	8.672	0.088	0.898	2.936	0.084
Sigmoid	**RMS Prop**	**1.929**	**0.083**	**7.774**	**0.08**	**0.909**	**2.777**	**0.082**
tanh	RMS Prop	1.945	0.083	8.11	0.083	0.905	2.835	0.083
Leaky RELU	SGD	2.23	0.096	9.875	0.098	0.883	3.135	0.096
RELU	SGD	2.334	0.105	10.438	0.097	0.875	3.226	0.104
Sigmoid	SGD	2.423	0.106	11.185	0.105	0.863	3.337	0.105
tanh	SGD	2.118	0.09	9.03	0.088	0.893	2.997	0.089

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**



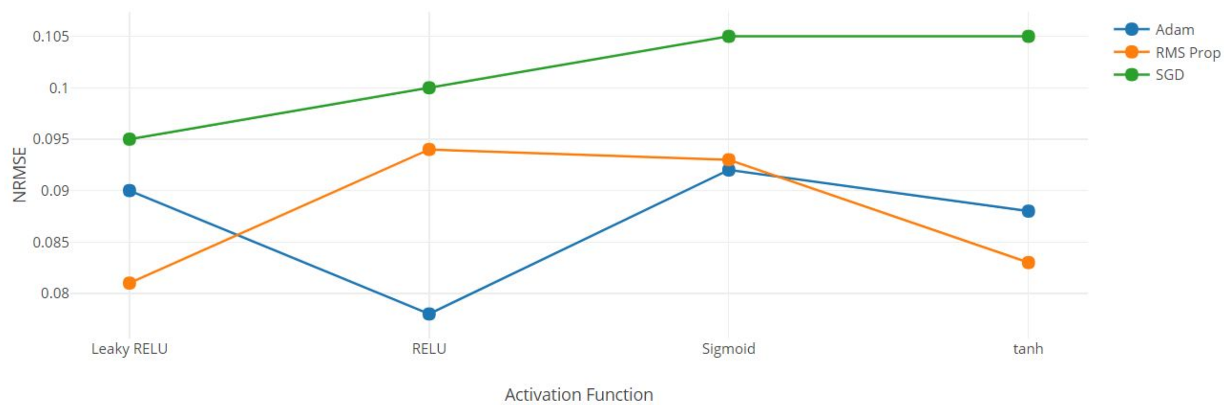
8. Table

- Learning Rate is 0.01
- With 2 hidden layer
- 5 fold CV 7 input features

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	2.113	0.088	8.98	0.09	0.894	2.991	0.088
RELU	Adam	**1.706**	**0.073**	**6.178**	**0.078**	**0.928**	**2.422**	**0.073**
Sigmoid	Adam	2.143	0.093	9.312	0.092	0.89	3.046	0.092
tanh	Adam	1.904	0.083	7.371	0.088	0.913	2.707	0.084
Leaky RELU	RMS Prop	1.815	0.079	6.667	0.081	0.923	2.577	0.08
RELU	RMS Prop	2.051	0.086	8.033	0.094	0.904	2.827	0.088
Sigmoid	RMS Prop	2.227	0.096	9.588	0.093	0.887	3.093	0.096
tanh	RMS Prop	1.865	0.082	7.215	0.083	0.917	2.675	0.082
Leaky RELU	SGD	2.335	0.101	10.678	0.095	0.872	3.257	0.101
RELU	SGD	2.346	0.104	10.264	0.1	0.876	3.198	0.104
Sigmoid	SGD	2.442	0.107	11.442	0.105	0.855	3.373	0.107
tanh	SGD	2.651	0.11	13.479	0.105	0.836	3.657	0.111

- **Plot for R^2 values with various Optimizers and Activation Functions**

- **Plot for NRMSE values with various Optimizers and Activation Functions**



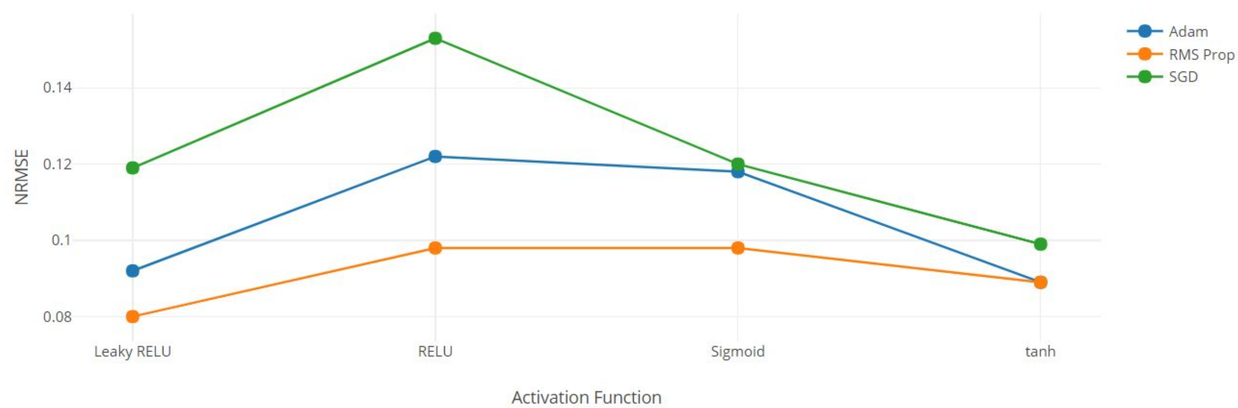
9. Table

- Learning Rate is 0.001
- With 1 hidden layers
- 5 fold CV with 4 input features

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	2.161	0.093	9.66	0.092	0.886	3.103	0.092
RELU	Adam	2.548	0.102	15.107	0.122	0.806	3.871	0.102
Sigmoid	Adam	2.689	0.114	14.853	0.118	0.811	3.839	0.113
tanh	Adam	2.028	0.087	8.464	0.089	0.9	2.903	0.086
Leaky RELU	RMS Prop	**1.836**	**0.08**	**6.776**	**0.08**	**0.921**	**2.597**	**0.081**
RELU	RMS Prop	2.194	0.092	10.301	0.098	0.88	3.204	0.091
Sigmoid	RMS Prop	2.276	0.096	10.758	0.098	0.872	3.263	0.096
tanh	RMS Prop	2.03	0.087	8.412	0.089	0.901	2.896	0.088
Leaky RELU	SGD	2.664	0.11	15.332	0.119	0.803	3.898	0.109
RELU	SGD	2.776	0.11	15.682	0.153	0.801	3.945	0.111
Sigmoid	SGD	2.698	0.114	15.405	0.12	0.802	3.909	0.113
tanh	SGD	2.251	0.095	10.559	0.099	0.872	3.238	0.095

- **Plot for R^2 values with various Optimizers and Activation Functions**

- **Plot for NRMSE values with various Optimizers and Activation Functions**

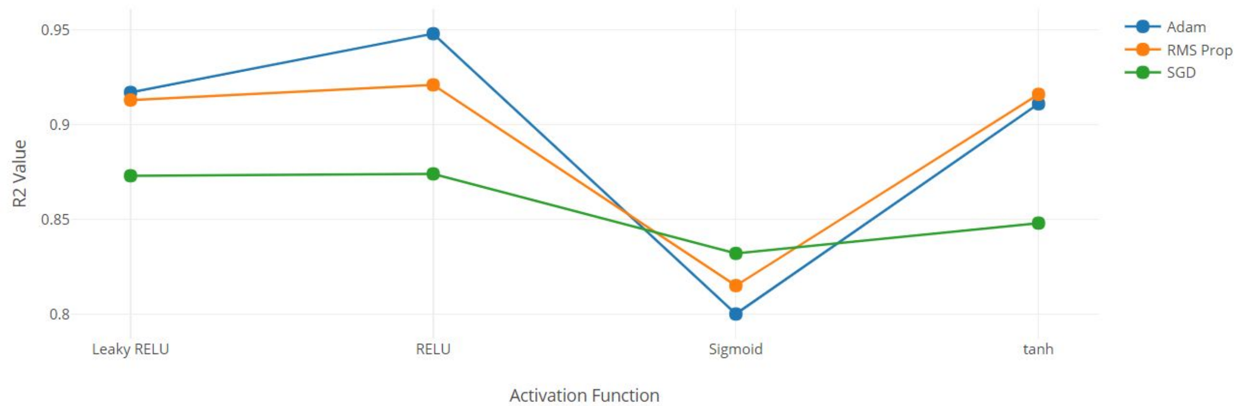


<u>10. Table</u>

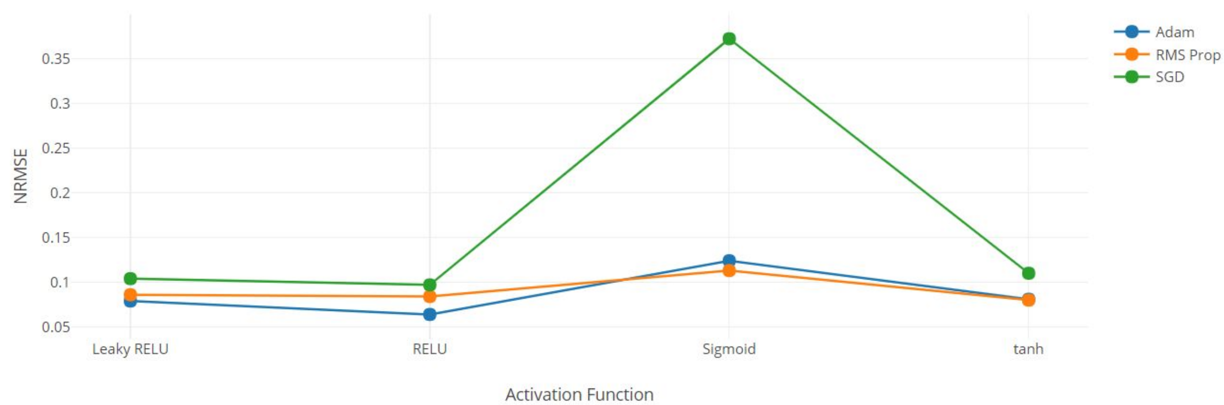
- Learning Rate is 0.001
- With 2 hidden layers
- 5 fold CV with 4 input features

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	1.866	0.081	6.933	0.079	0.917	2.629	0.082
RELU	Adam	**1.558**	**0.072**	**4.58**	**0.064**	**0.948**	**2.118**	**0.072**
Sigmoid	Adam	2.783	0.117	15.699	0.124	0.8	3.952	0.116
tanh	Adam	1.909	0.081	7.54	0.081	0.911	2.733	0.081
Leaky RELU	RMS Prop	1.924	0.086	7.436	0.086	0.913	2.723	0.086
RELU	RMS Prop	1.803	0.079	6.759	0.084	0.921	2.596	0.08
Sigmoid	RMS Prop	2.675	0.114	14.623	0.113	0.815	3.8	0.113
tanh	RMS Prop	1.882	0.082	7.334	0.08	0.916	2.703	0.081
Leaky RELU	SGD	2.371	0.103	10.677	0.104	0.873	3.253	0.103
RELU	SGD	2.318	0.101	10.427	0.097	0.874	3.214	0.101
Sigmoid	SGD	4.222	0.176	33.081	0.372	0.832	5.413	0.184
tanh	SGD	2.465	0.103	12.189	0.11	0.848	3.473	0.103

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**

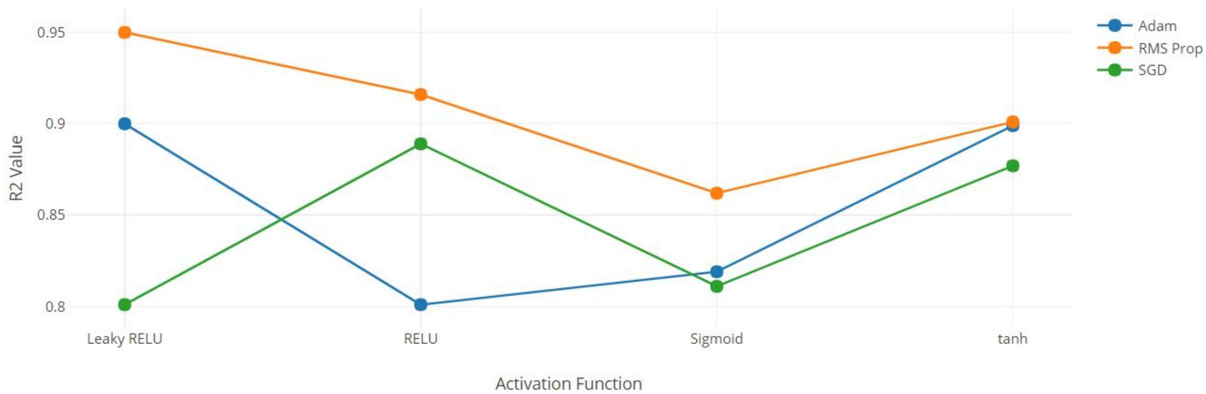


11. Table

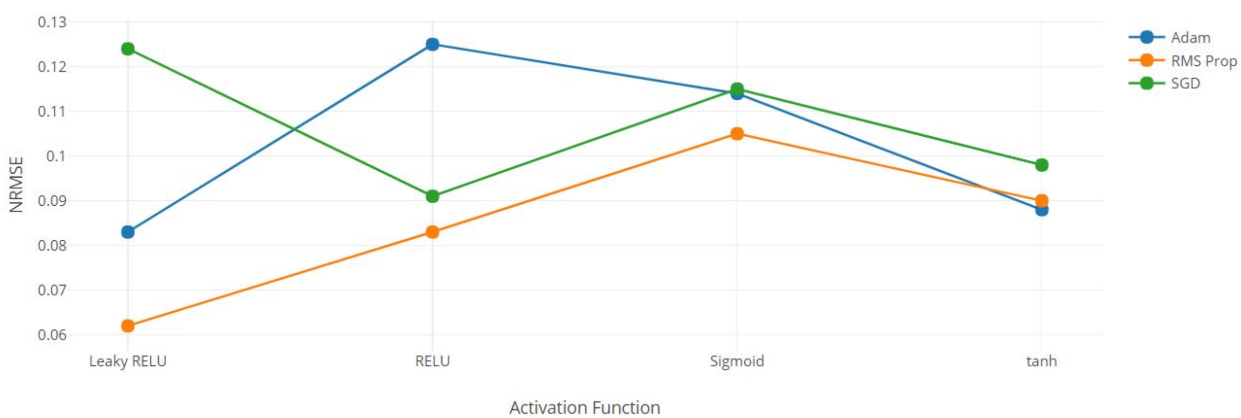
- Learning Rate is 0.01
- With 1 hidden layers
- 5 fold CV with 4 input features

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	2.1	0.089	8.474	0.083	0.9	2.901	0.089
RELU	Adam	2.716	0.114	15.456	0.125	0.801	3.915	0.113
Sigmoid	Adam	2.623	0.112	14.358	0.114	0.819	3.759	0.11
tanh	Adam	2.053	0.087	8.58	0.088	0.899	2.921	0.087
Leaky RELU	RMS Prop	**1.522**	**0.072**	**4.375**	**0.062**	**0.95**	**2.085**	**0.071**
RELU	RMS Prop	1.939	0.087	7.219	0.083	0.916	2.681	0.087
Sigmoid	RMS Prop	2.283	0.098	11.421	0.105	0.862	3.343	0.097
tanh	RMS Prop	1.969	0.084	8.194	0.09	0.901	2.85	0.084
Leaky RELU	SGD	2.703	0.113	15.489	0.124	0.801	3.919	0.113
RELU	SGD	2.121	0.088	9.307	0.091	0.889	3.044	0.088
Sigmoid	SGD	2.676	0.114	14.892	0.115	0.811	3.845	0.113
tanh	SGD	2.248	0.095	10.202	0.098	0.877	3.184	0.095

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**

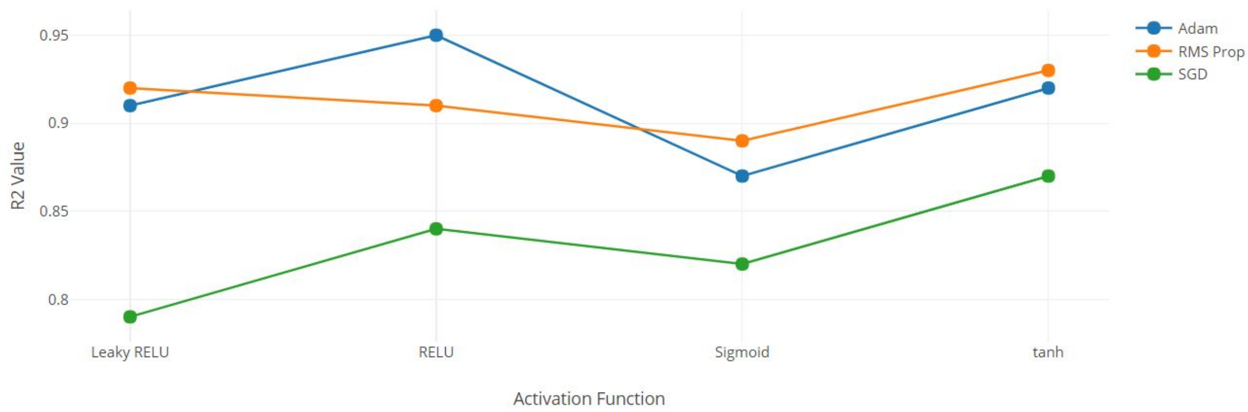


12. Table

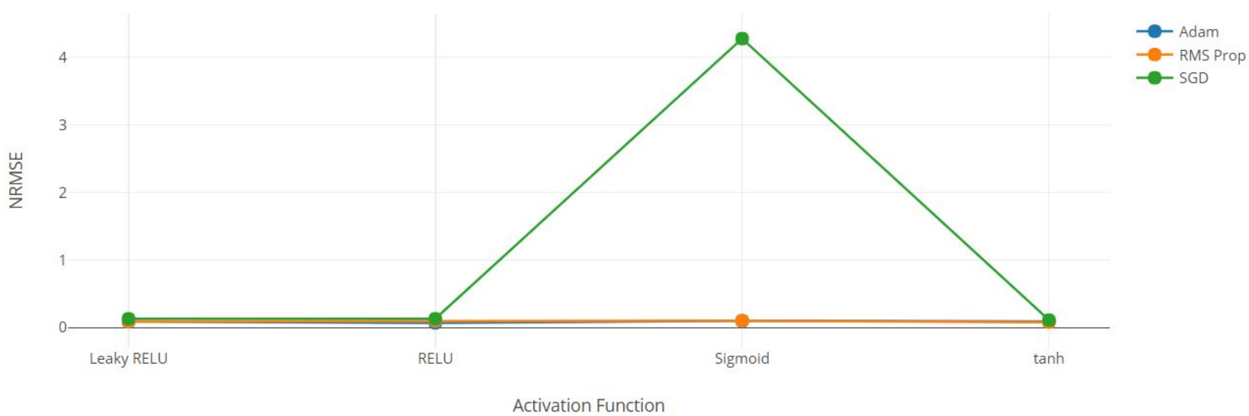
- Learning Rate is 0.01
- With 2 hidden layers
- 5 fold CV with 4 input features

Activation functions	Optimizer	mad	mape	mse	nrmse	R^2	rmse	smape
Leaky RELU	Adam	1.93	0.09	7.61	0.09	0.91	2.75	0.09
RELU	**Adam**	**1.51**	**0.07**	**4.41**	**0.07**	**0.95**	**2.08**	**0.07**
Sigmoid	Adam	2.26	0.1	10.55	0.1	0.87	3.23	0.1
tanh	Adam	1.89	0.09	7.12	0.09	0.92	2.66	0.09
Leaky RELU	RMS Prop	2.02	0.1	7.61	0.09	0.92	2.75	0.09
RELU	RMS Prop	2.01	0.09	8.03	0.1	0.91	2.83	0.09
Sigmoid	RMS Prop	2.18	0.09	9.71	0.1	0.89	3.11	0.09
tanh	RMS Prop	1.72	0.07	6.07	0.08	0.93	2.46	0.07
Leaky RELU	SGD	2.9	0.12	16.4	0.13	0.79	4.04	0.12
RELU	SGD	2.47	0.1	12.49	0.13	0.84	3.51	0.1
Sigmoid	SGD	8.89	0.38	96.67	4.28	0.82	9.83	0.39
tanh	SGD	2.31	0.1	10.86	0.11	0.87	3.28	0.1

- **Plot for R^2 values with various Optimizers and Activation Functions**



- **Plot for NRMSE values with various Optimizers and Activation Functions**



8.Conclusions

From the above hyper parameters with and without feature selection using MLP. we got these hyperparameters works good with no feature selection and feature selection. We compare MLP with different hyperparameters with and without feature selection using t-test value on R^2 value. This gives the statistical significance between different hyper parameters with and without feature selection. We calculate t-statistic with 1% level of significance and 5+5-2 = 8 degree of freedom. T - statistic value at 8 degree of freedom and 1% level of significance is 2.896. We reject the null hypothesis below 2.896 and accept if it is above 2.896.

- T - Statistic Test with MLP for different parameters with and without feature selection with R^2 Values

S. No	Features	Hidden Layers	Activation Function	Optimizer	Learning Rate	R^2 value	T-Statistic
1 (a)	7	1	ReLU	Adam	0.001	0.92	3.573
1(b)	4	1	ReLU	Adam	0.001	0.81	
2(a)	7	1	Leaky ReLU	RMS Prop	0.01	0.88	3.465
2(b)	4	1	Leaky ReLU	RMS Prop	0.01	0.95	
3(a)	7	2	ReLU	Adam	0.01	0.93	0.942
3(b)	4	2	ReLU	Adam	0.01	0.95	
4(a)	7	2	ReLU	RMS Prop	0.001	0.94	1.856
4(b)	4	2	ReLU	RMS Prop	0.001	0.92	

- From the above table 3 and 4 hyper parameters are having t-test score less than 2.896, so with these hyper parameters are having statistically good R^2 values.

- MLP with Hidden Layers = 2 and Activation Function = ReLU and Optimizer = Adam with Learning Rate = 0.01 with 4 features is good model and is giving 0.95 R^2 value.

9. References

- https://archive.ics.uci.edu/ml/datasets/Energy+efficiency
- http://people.maths.ox.ac.uk/tsanas/Preprints/ENB2012.pdf
- http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1678-86212017000300103#e01
- http://people.maths.ox.ac.uk/tsanas/Preprints/ENB2012.pdf
- https://suw.biblos.pk.edu.pl/downloadResource&mId=504384
- http://scikit-learn.org
- https://docs.scipy.org/doc/scipy-0.19.0/reference/index.html
- https://docs.scipy.org/doc/numpy-1.15.0/reference/generated/numpy.interp.html
- https://pandas.pydata.org/pandas-docs/stable/index.html