

# Supervised Machine Learning - Regression

## Topic - Appliance Energy Prediction



By:-  
Ajit kumar Toppo

# Benefits of energy prediction



# Dataset(19735, 29)



#	Column	Non-Null Count	Dtype
0	date	19735 non-null	object
1	Appliances	19735 non-null	int64
2	lights	19735 non-null	int64
3	T1	19735 non-null	float64
4	RH_1	19735 non-null	float64
5	T2	19735 non-null	float64
6	RH_2	19735 non-null	float64
7	T3	19735 non-null	float64
8	RH_3	19735 non-null	float64
9	T4	19735 non-null	float64
10	RH_4	19735 non-null	float64
11	T5	19735 non-null	float64
12	RH_5	19735 non-null	float64
13	T6	19735 non-null	float64
14	RH_6	19735 non-null	float64
15	T7	19735 non-null	float64
16	RH_7	19735 non-null	float64
17	T8	19735 non-null	float64
18	RH_8	19735 non-null	float64
19	T9	19735 non-null	float64
20	RH_9	19735 non-null	float64
21	T_out	19735 non-null	float64
22	Press_mm_hg	19735 non-null	float64
23	RH_out	19735 non-null	float64
24	Windspeed	19735 non-null	float64
25	Visibility	19735 non-null	float64
26	Tdewpoint	19735 non-null	float64
27	rv1	19735 non-null	float64
28	rv2	19735 non-null	float64

**Date** year-month-day hour:minute:second.

**Appliances** energy use in Wh (Dependent variable).

**lights** energy use of light fixtures in the house in Wh.

**T1** Temperature in kitchen area, in Celsius.

**RH1** Humidity in kitchen area, in %.

**T2** Temperature in living room area, in Celsius.

**RH2** Humidity in living room area, in %.

**T3** Temperature in the laundry room area.

**RH3** Humidity in laundry room area, in %.

**T4** Temperature in office room, in Celsius.

**RH4** Humidity in the office room, in %.

**T5** Temperature in bathroom, in Celsius.

**RH5** Humidity in bathroom, in %.

**T6** Temperature outside the building (north side), in Celsius.

**RH6** Humidity outside the building (north side), in %.

**T7** Temperature in ironing room, in Celsius.

**RH7** Humidity in the ironing room, in %.

**T8** Temperature in teenager room 2, in Celsius.

**RH8** Humidity in teenager room 2, in %.

**T9** Temperature in parents room, in Celsius.

**RH9** Humidity in parents room, in %.

**T\_out** Temperature outside (from Chievres weather station), in Celsius.

**Press\_mm\_hg** Pressure (from Chievres weather station), in mm Hg.

**RHout** Humidity outside (from Chievres weather station), in %.

**Windspeed** (from Chievres weather station), in m/s.

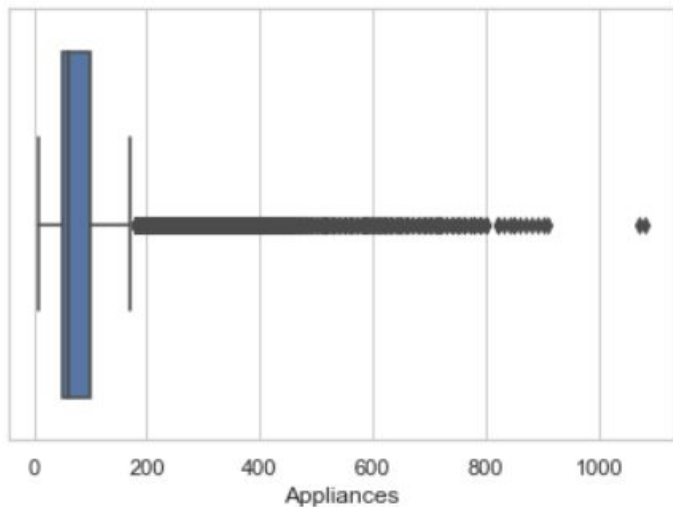
**Visibility** (from Chievres weather station), in km.

**Tdewpoint** (from Chievres weather station),  $^{\circ}\text{C}$ .

**rv1** Random variable 1, nondimensional.

**rv2** Random variable 2, nondimensional.

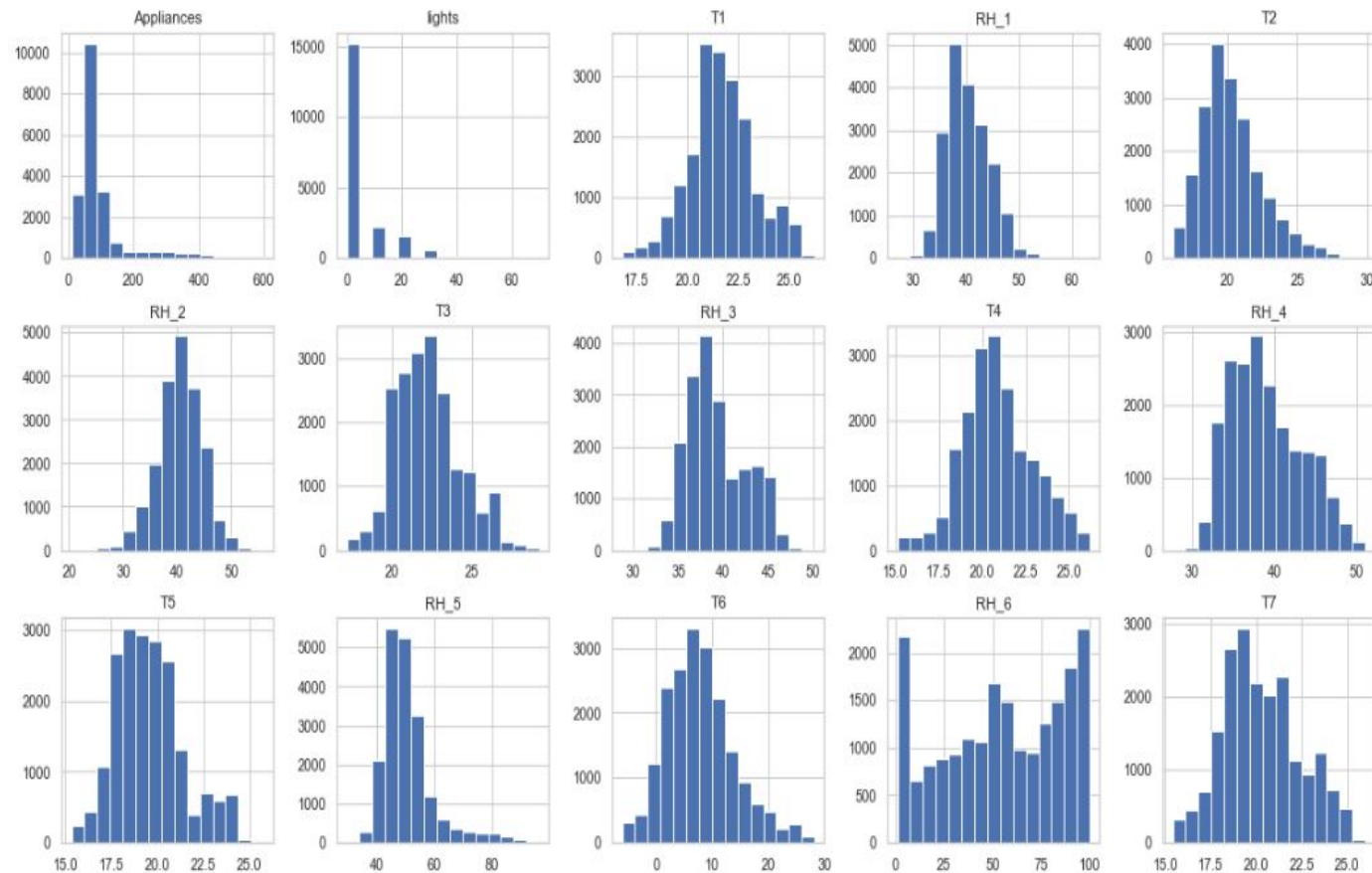
# Removing outliers

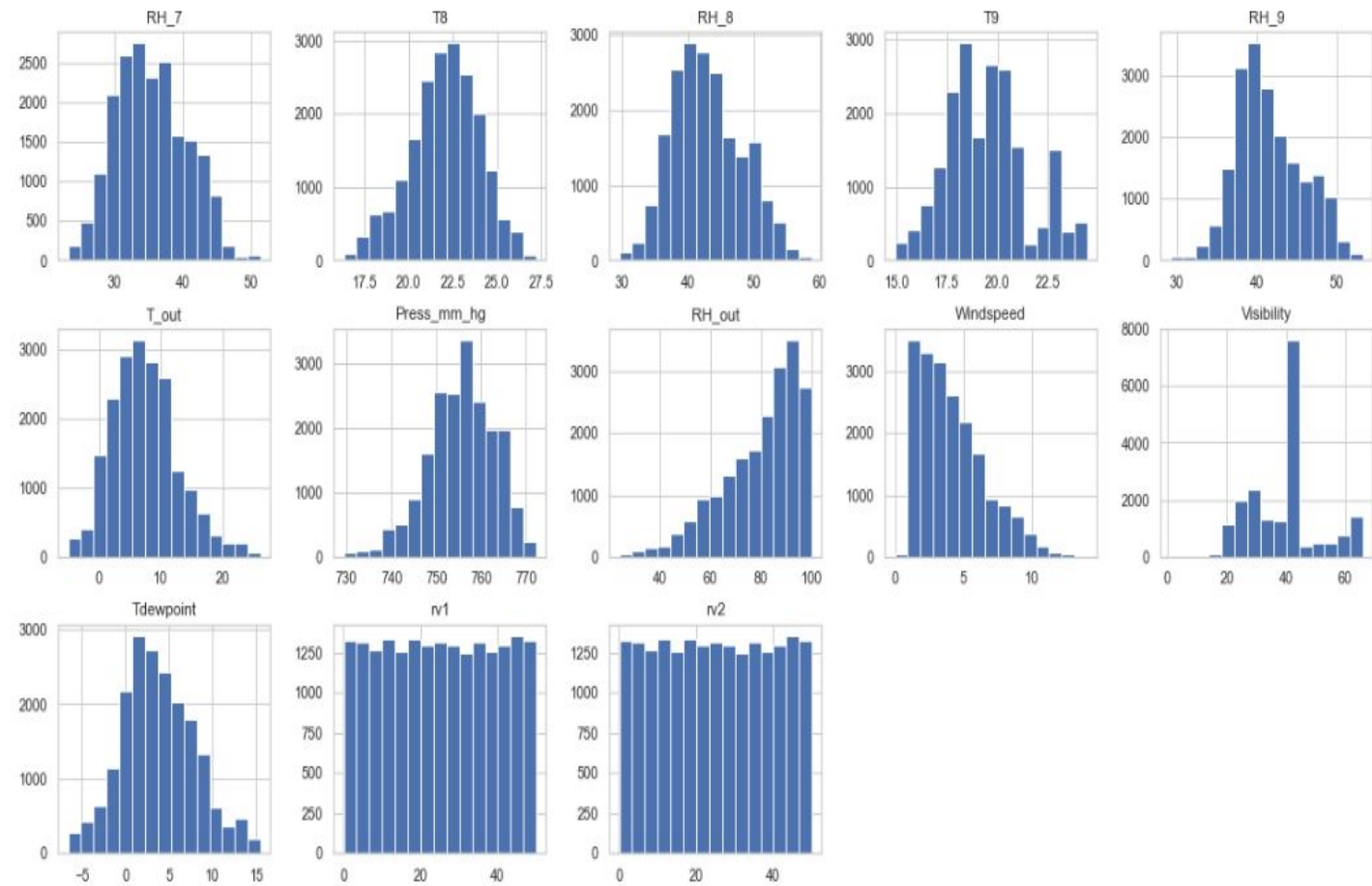


	count	mean	std	min	25%	50%	75%	max
<b>Appliances</b>	19735.0	97.694958	102.524891	10.000000	50.000000	60.000000	100.000000	1080.000000

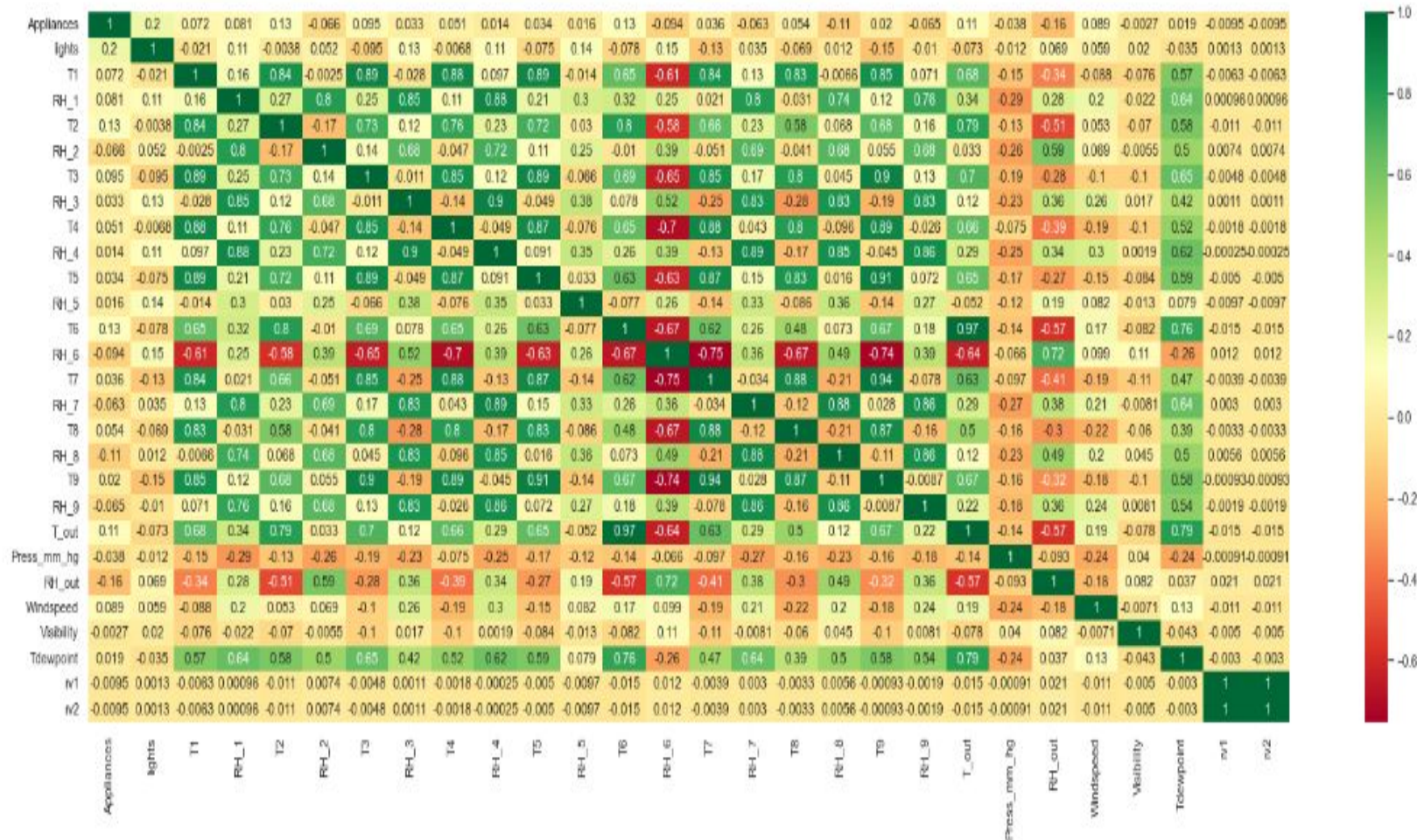
```
df = df.dropna()  
df = df.drop(df[(df.Appliances>600)|(df.Appliances<0)].index)
```

# Visualising the distribution of values in each variable





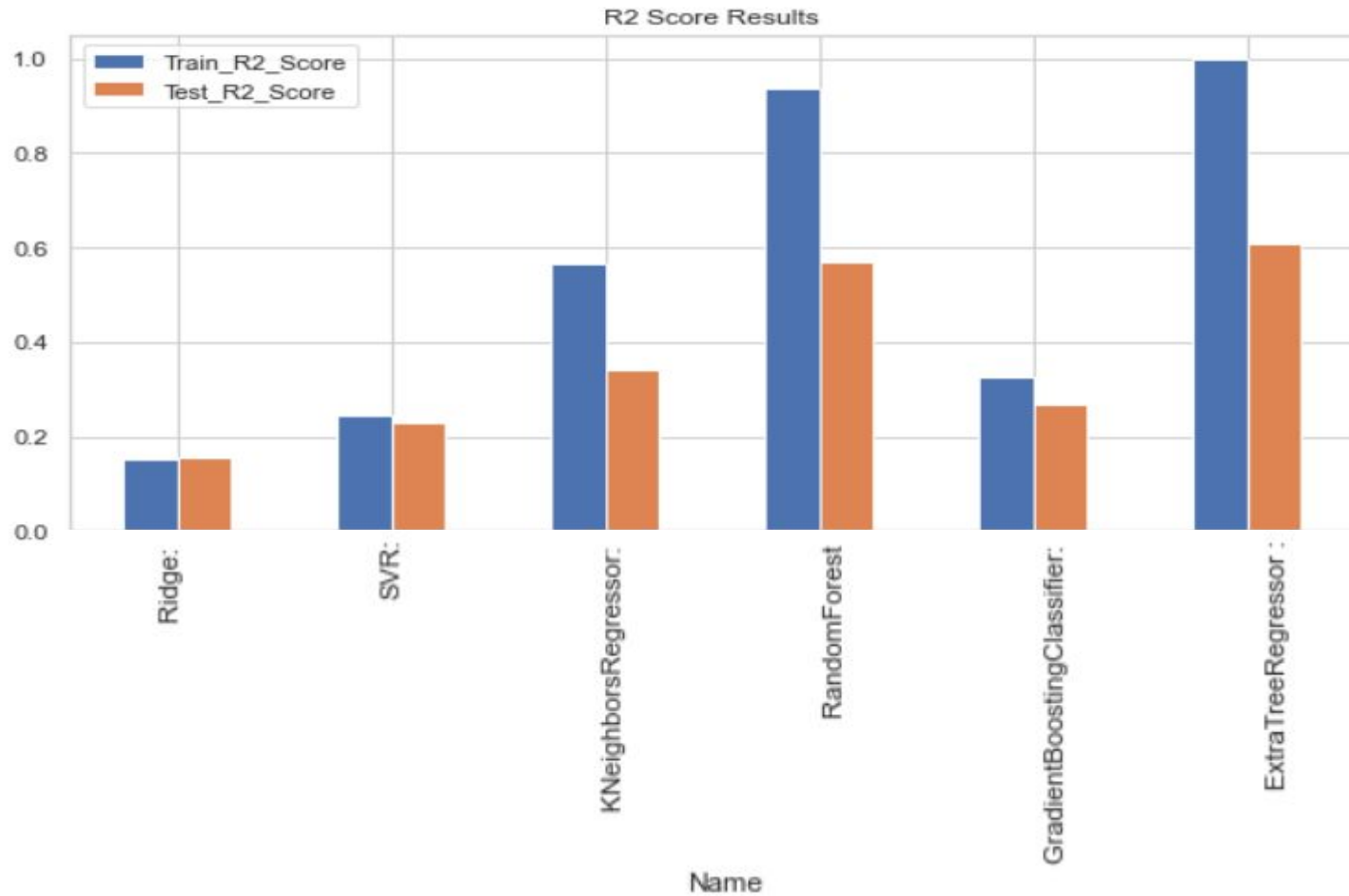




- **Drop the columns light and date.**
- **Applying StandardScaler on dataset.**
- **Splitting the data into features and target variables, using train\_test\_split on our dataset.**
- **We used regression model like ridge regression, Support Vector Regression (SVR), K-nearest Neighbours Regression, Random Forest Regressor, Gradient Boosting Regressor, Extra Trees Regressor**

## Model Performance

	Name	Train_Time	Train_R2_Score	Test_R2_Score	Test_RMSE_Score
0	Ridge:	0.025209	0.153706	0.158032	0.914342
1	SVR:	23.218666	0.244216	0.229533	0.874657
2	KNeighborsRegressor:	0.000000	0.567563	0.342654	0.807900
3	RandomForest	67.298949	0.937116	0.567892	0.655023
4	GradientBoostingClassifier:	15.859000	0.325181	0.270085	0.851328
5	ExtraTreeRegressor :	18.905303	1.000000	0.606715	0.624906



<b>Extra tree regressor</b>	<b>Before using grid search cv</b>	<b>After using grid search cv</b>
<b>Train_R2_Score</b>	<b>1.000000</b>	<b>1.0</b>
<b>Test_R2_Score</b>	<b>0.606715</b>	<b>0.6108375137430209</b>
<b>Test_RMSE_Score</b>	<b>0.624906</b>	<b>0.6216216623016521</b>

