

Supervised Machine Learning - Regression

Topic - Appliance Energy Prediction



By:Ajit kumar Toppo

Benefits of energy prediction





Dataset(19735, 29)



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

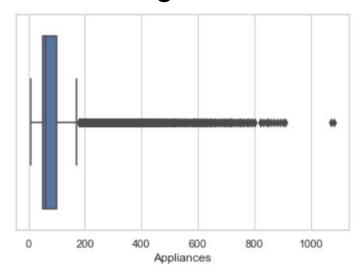
Tdewpoint (from Chievres weather station), °C.

rv1 Random variable 1, nondimensional.

rv2 Random variable 2, nondimensional.



Removing outliers

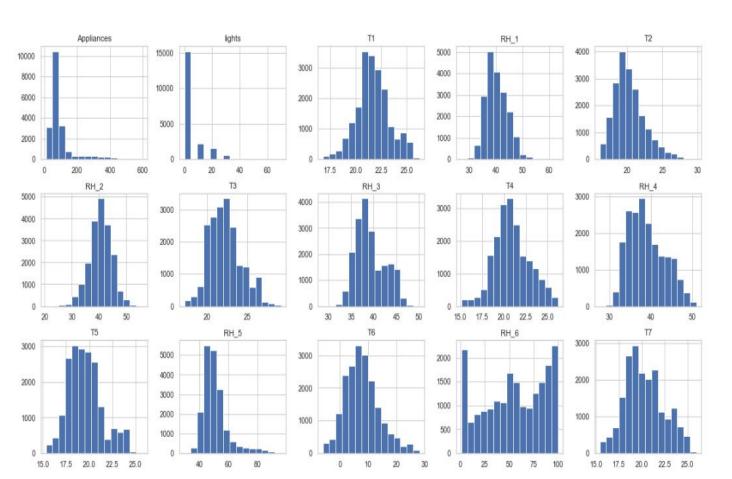


	count	mean	std	min	25%	50%	75%	max
Appliances	19735.0	97.694958	102.524891	10.000000	50.000000	60.000000	100.000000	1080.000000

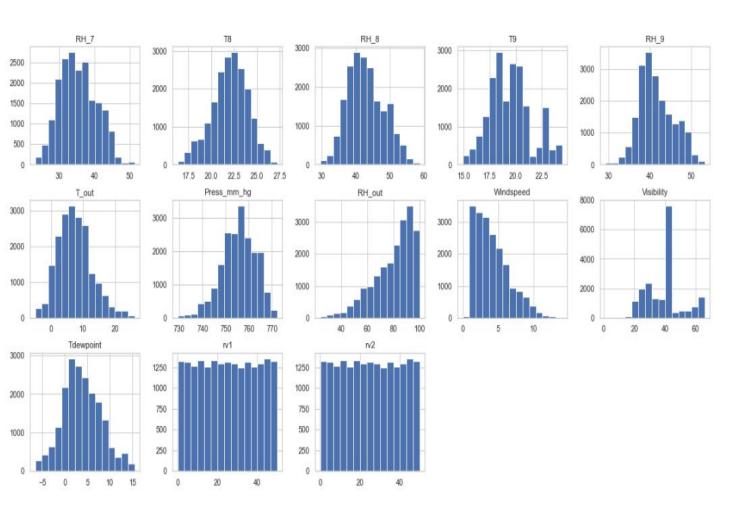
```
df = df.dropna()
df = df.drop(df[(df.Appliances>600)|(df.Appliances<0)].index)</pre>
```

Visualising the distribution of values in each variable









Appliances	1	0.2	0.072	0.081	0.13	-0.066	0.095	0.033	0.051	0.014	0.034	0.016	0.13	-0.094	0.036	-0.063	0.054	-0.11	0.02	-0.065	0.11	-0.038	-0.16	0.089	-0.0027	0.019	-0.0095	-0.0095
lights	0.2	1	-0.021	0.11	-0.0038	0.052	-0.095	0.13	-0.0068	0.11	-0.075	0.14	-0:078	0.15	-0.13	0.035	-0.069	0.012	-0.15	-0.01	0.073	-0.012	0.069	0.059	0.02	-0.035	0.0013	0.0013
T1	0.072	-0.021	1	0.16	0.84	-0.0025	0.89	-0.028	0.88	0.097	0.89	-0.014	0.65	-0.61	0.84	0.13	0.83	-0.0066	0.85	0.071	0.68	-0.15	-0.34	-0.088	-0.076	0.57	-0.0063	-0.0063
RH_1	0.081	0.11	0.16	1	0.27	8.0	0.25	0.85	0.11	0.88	0.21	0.3	0.32	0.25	0.021	0.8	0.031	074	0.12	078	0.34	0.29	0.28	0.2	0.022		0.00098	0.00096
T2	0.13	-0.0038	0.84	0.27	1	0.17	0.73	0.12	0.76	0.23	0.72	0.03	0.8	-0.58	0.68	0.23	0.58	0.068	0.68	0.16	0.79	-0.13	-0.51	0.053	-0.07	0.58	-0.011	-0.011
RH_2	0.066	0.052	-0.0025	0.8	0.17	1	0.14	0.68	-0.047	0.72	0.11	0.25	-0.01	0.39	-0.051	0.69	-0.041	0.68	0.055	0.68	0.033	0.26	0.59	0.089	0.0055	0.5	0.0074	0.0074
T3	0.095	0.095	0.89	0.25	0.73	0.14	1	0.011	0.85	0.12	0.89	0.086	0.69	0.65	0.85	0.17	0.8	0.045	0.9	0.13	0.7	0.19	0.28	0.1	0.1		0.0048	-0.0048
RH_3	0.033	0.13	-0.028	0.85	0.12	89.0	-0.011	1	-0.14	0.9	-0.049	0.38	0.078	0.52	-0.25	0.83	-0.28	0.83	-0.19	0.83	0.12	-0.23	0.36	0.26	0.017	0.42	0.0011	0.0011
T4	0.051	-0.0068	0.88	0.11	0.76	-0.047	0.85	-0.14	1	-0.049	0.87	-0.076	0.65	-0.7	0.88	0.043	0.8	-0.096	0.89	-0.026	0.66	-0.075	0.39	-0.19	-0.1	0.52	-0.0018	-0.0018
RH_4	0.014	0.11	0.097	88.0	0.23	0.72	0.12	0.9	-0.049	1	0.091	0.35	0.26	0.39	-0.13	0.89	-0.17	0.85	-0.045	0.86	0.29	-0.25	0.34	0.3	0.0019	0.62	0.00025	10.00025
15	0.034	-0.075	0.89	0.21	0.72	0.11	0.89	-0.049	0.87	0.091	. 1	0.033	0.63	-0.63	0.87	0.15	0.83	0.016	0.91	0.072	0.65	-0.17	40.27	-0.15	-0.084	0.59	-0.005	-0.005
F8H_5	0.016	0.14	-0.014	0.3	0.03	0.25	-0.066	0.38	-0.076	0.35	0.033	1	-0.077	0.26	-0.14	0.33	-0.086	0.36	-0.14	0.27	-0.052	-0.12	0.19	0.082	-0.013	0.079	-0.0097	
T6	0.13	-0.078	0.65	0.32	8.0	-0.01	0.69	0.078	0.65	0.26	0.63	-0.077	1	-0.67	0.62	0.26	0.48	0.073	0.67	0.18	0.97	-0.14	-0.57	0.17	-0.082	0.76	100000	-0.015
RH_6	-0.094	0.15	-0.61	0.25	-0.58	0.39	-0.65	0.52	-0.7	0.39	-0.63	0.26	-0.67		-0.75	0.36	-0.67	0.49	-0.74	0.39	-0.64	-0.066	0.72	0.099	0.11	-0.26	0.012	
17	0.036	-0 13	0.84	0.021	0.66	-0.051	0.85	-0.25	0.88	-0.13	0.87	-0.14	0.62	-0.75	1	-0.034	0.88	-0.21	0.94	-0.078	0.63	-0.097	4) 41	-0.19	-0.11	0.47	-0.0039	
RH_7	-0.063	0.035	0.13	0.8	0.23	0.69	0.17	0.83	0.043	0.89	0.15	0.33	0.26	0.36	-0.034		-0.12	0.88	0.028	0.88	0.29	-0.27	0.38	0.21	0.0081	0.64		0.003
T8	0.054	0.069	AND DESCRIPTION	0.031	0.58	-0.041	0.8	-0.28	8.0	-0.17	0.83	-0.086	0.48	-0.67	0.88	-0.12	0.01	-0.21	0.87	-0.18	0.5	-0.16	-0.3	-0.22	0.06		-0.0033	200
RH_8 T9	0.11	-0.15	-0.0068 0.85	0.74	0.068	0.68	0.045	-0.19	0.096	40.045	0.016	0.36	0.073	0.49	0.21	0.88	0.21	-0.11	-0.11	0.86 -0.0087	0.12	-0.16	0.49	-0.18	0.045		0.0056 0.00093	
	-0.065	-0.01	0.071	0.76	0.16	0.68	0.13	0.83	-0.026	0.86	0.072	0.14	0.18	0.39	-0.078	0.86	-0.16	-	-0.0087	-0.0001	0.22	-0.18	0.36	0.24	0.0081		-0.0019	
RH_9 T_out		0.073		0.34	0.79	0.033	0.7	0.12	0.66	0.29	0.65	-0.052	0.10	-0.64	0.63	0.29	0.10	0.12	0.0001	0.22	0.22	-0.14	-0.57	0.19	-0.078	0.79	40.015	
Press_mm_hg	-0.038	-0.012	-0.15	0.29	-0.13	-0.26	-0.19	-0.23	-0.075	-0.25	-0.17	-0.12	-0.14	-0.066	-0.097	-0.27	-0.16	40.23	-0.16	-0.18	JD 14	1	-0.093	-0.24	0.04		0.00091	
FMI out		0.069	-0.34	0.28	40.51	0.59	-0.28	0.36	-0.39	0.34	-0.27	0.19	-0.57	0.72	-0.41	0.38	-0.3	0.49	0.32	0.36	-0.57	-0.093	1	-0.18	0.082	0.037	0.021	11500
Windspeed	0.089	0.059	-0.088	0.2	0.053	0.069	-0.1	0.26	-0.19	0.3	-0.15	0.082	0.17	0.099	-0.19	0.21	-0.22	0.2	-0.18	0.24	0.19	-0.24	-0.18	1	-0.0071	0.13		-0.011
Visibility	-0.0027	0.02	-0.076	-0.022	-0.07	-0.0055	-0.1	0.017	-0.1	0.0019	-0.084	-0.013	-0.082	0.11	-0.11	-0.0081	-0.06	0.045	-0.1	0.0081	-0.078	0.04	0.082	-0.0071	1	-0.043	-0.005	-0.005
Tdewpoint	0.019	-0.035	0.57	0.64	0.58	0.5	0.65	0.42	0.52	0.62	0.59	0.079	0.76	-0.26	0.47	0.64	0.39	0.5	0.58	0.54	0.79	-0.24	0.037	0.13	-0.043	1	-0.003	-0.003
rv1	-0.0095	0.0013	-0.0083	0.00096	0.011	0.0074	-0.0048	0.0011	-0.0018	-0.00025	-0.005	-0.0097	-0.015	0.012	-0.0039	0.003	-0.0033	0.0056	0.00093	0.0019	0.015	0.00091	0.021	-0.011	0.005	-0.003	1	1
1/2	-0.0095	0.0013	-0.0063	0.00098	-0.011	0.0074	-0.0048	0.0011	-0.0018	-0.00025	-0.005	-0.0097	-0.015	0.012	-0.0039	0.003	-0.0033	0.0058	0.00093	-0.0019	-0.015	0.00091	0.021	-0.011	-0.005	-0.003		
	Appliances	Aghts	F	H	Ŋ	FF.	P	RH_3	7	<u> </u>	13	E H	91	BH 6	4	F	18	RET_8	£	Ret 9	T_out	by www ssa	RH_out	Windspeed	Visibility	Tdewpoint	L/v	n2

Al

-08

-06

-0.4

-02

-00

--0.2

0.4

--0.6

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.

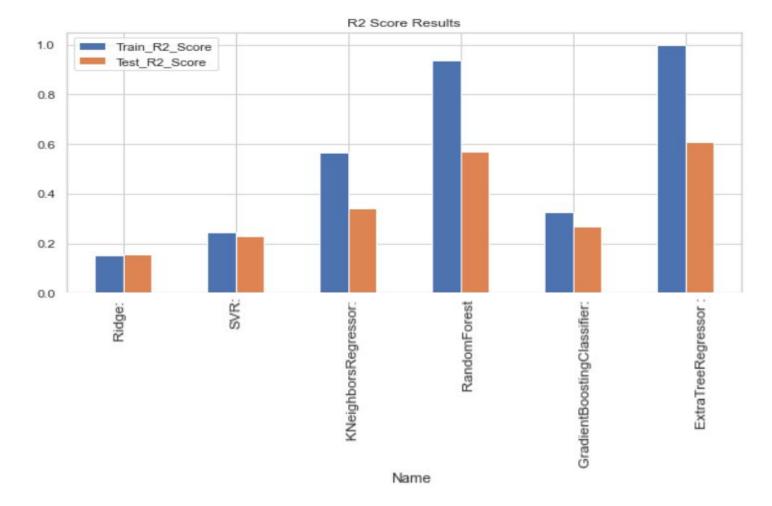
 We used regression model like ridge regression, Support Vector Regression (SVR), K-nearest Neighbours Regression, Random Forest Regressor, Gradient Boosting Regressor, Extra Trees Regressor



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







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



