

Capstone Project Appliances Energy Prediction (Supervised ML - Regression)



Appliances Energy Prediction



Make predictions

Onsumption by the Appliances of a household based on factors like temperature, humidity and pressure.

Model Development

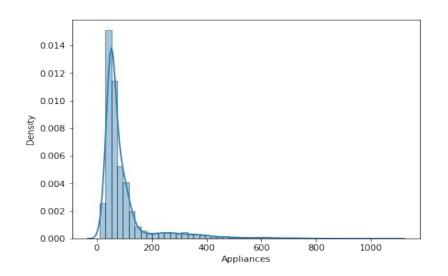
Develop a Supervised MachineLearning Model using regression.

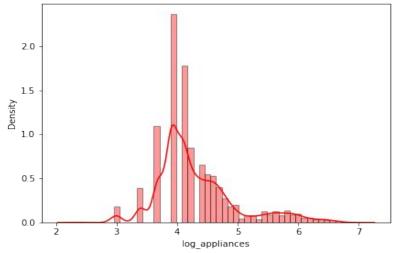
Enaturne

ΔΙ	

01	date	RH7 - Humidity in ironing roon	1 16
02	Appliances - Energy in Watt	T8 - Temperature in teenager roor	
03	T1 - temperature in kitchen area	,	
04	RH1 - humidity in kitchen area	RH8 -Humidity in teenager roor	n 18
05	T2 - temperature in living room	T9 - Temperature in parents room	19
06	RH2 -humidity in living room	RH9 - Humidity in parents roo	m 20
07	T3-temperature in laundry	Tout - Temperature outside	de 21
08	RH3- humidity in laundry	RHout - Humidity outside	de 22
09	T4- temperature in office room	Press_mm_l	ng 23
10	RH4-humidity in office room	Visibili	ty 24
11	T5-temperature in bathroom	Windspec	ed 25
12	RH5- humidity in bathroom	Tdewpoi	nt 26
13	T6-temperature outside the building	r	v1 27
14	RH6-humidity outside the building		v2 28
15	T7-temperature in ironing room	ligh	ts 29







Appliances

Positively Skewed Distribution

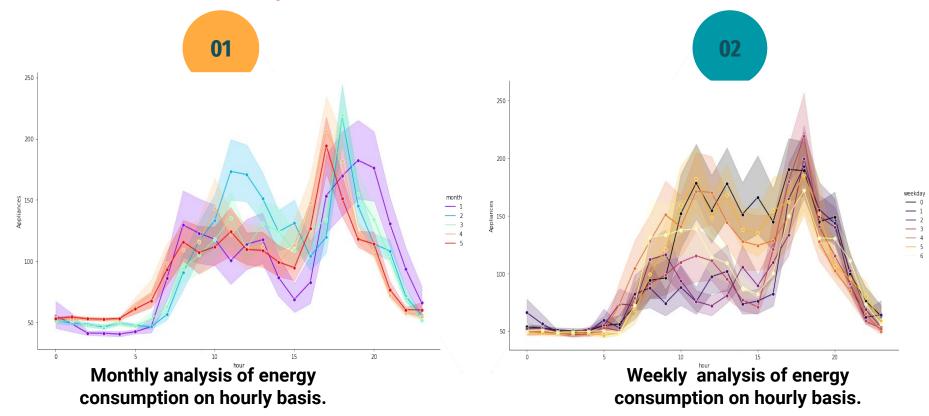
Vs

Log Transformed Appliances

Normal Distribution

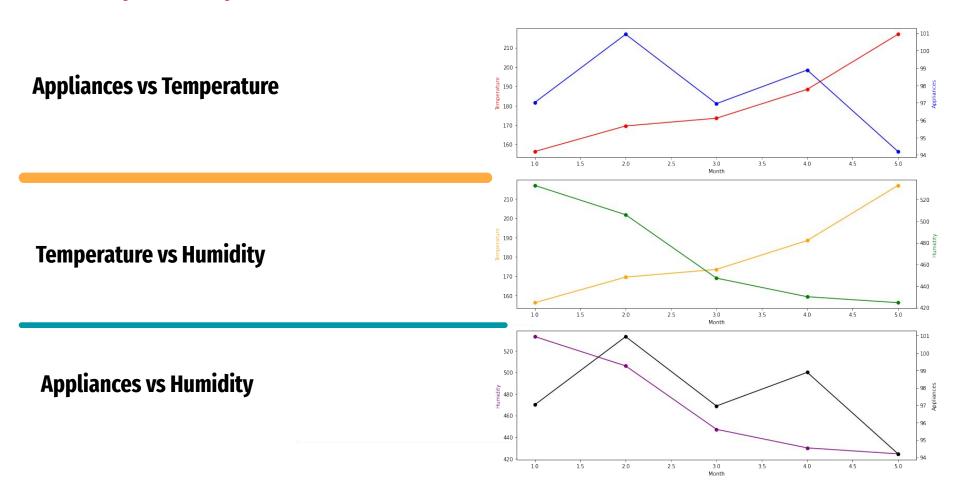


Multivariate analysis with respect to Appliances



Monthly Analysis of different variables





HEAT MAP



Appliances	- 1	0.2	0.055											-0.083				-0.094									-0.011 -0.011
lights	0.2	1	-0.024		-0.0056	0.051	-0.097	0.13	-0.0089		-0.079		-0.079	0.15	-0.14	0.035	-0.071		-0.16	-0.0088	-0.074				0.02	-0.036	0.000520.00052
т1 -	0.055		1	0.16	0.84	-0.0025	0.89	-0.029	0.88	0.098	0.89		0.65	-0.62	0.84	0.14	0.83	-0.0064	0.84	0.072	0.68		-0.35	-0.088	-0.076		-0.0062-0.0062
RH_1	0.086		0.16	1	0.27	0.8	0.25	0.84	0.11	0.88	0.21		0.32	0.25	0.021	0.8	-0.03	0.74	0.12	0.76	0.34		0.27		-0.021	0.64	-0.0007-0.0007
T2 -	0.12		0.84	0.27	1	-0.17	0.74	0.12	0.76	0.23	0.72		0.8	-0.58	0.66	0.23	0.58	0.069	0.68	0.16	0.79		-0.51		-0.07		-0.011 -0.011
RH_2	-0.06		-0.0025	0.8	-0.17	1	0.14	0.68	-0.047	0.72	0.11		-0.0097	0.39	-0.051	0.69	-0.041	0.68	0.055	0.68	0.034	-0.26	0.58		-0.0054		0.0063 0.0063
ТЗ -	0.085		0.89	0.25	0.74	0.14	1	-0.011	0.85	0.12	0.89	-0.066	0.69	-0.65	0.85	0.17	0.8	0.044	0.9	0.13	0.7		-0.28		-0.1	0.65	-0.0052-0.0052
RH_3	0.036		-0.029	0.84	0.12	0.68	-0.011	1	-0.14	0.9	-0.05	0.38	0.077	0.51	-0.25	0.83	-0.28	0.83	-0.2	0.83	0.12		0.36		0.017		0.000480.00048
T4	0.04	-0.0089	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	0.044	0.8	-0.095	0.89	-0.026	0.66		-0.39	-0.19	-0.1		-0.0018-0.0018
RH_4	0.017		0.098	0.88	0.23	0.72	0.12	0.9	-0.049	1	0.092		0.26	0.39	-0.13	0.89	-0.17	0.85	-0.045	0.86	0.29		0.34		0.0026	0.62	-0.0018-0.0018
T5 -	0.02		0.89		0.72	0.11	0.89	-0.05	0.87	0.092	1	0.033	0.63	-0.63	0.87	0.15	0.82	0.016	0.91	0.072	0.65				-0.084	0.59	-0.0055-0.0055
RH_5	0.00000				0.03	0.25	-0.066	0.38	-0.076		0.033	1	-0.078		-0.14		-0.087		-0.14		-0.053		10000		100000000000000000000000000000000000000	A CONTRACTOR	-0.011 -0.011
Т6	0.12		0.65		0.8	-0.0097	0.69	0.077	0.65	0.26	0.63		1	-0.67	0.62	0.26	0.48	0.074	0.67	0.18	0.97	-0.14	-0.57		-0.081	0.76	-0.015 -0.015
RH_6	-0.083		-0.62		-0.58	0.39	-0.65	0.51	-0.7	0.39	-0.63		-0.67	1	-0.75	0.36	-0.67	0.49	-0.74	0.39	-0.64	-0.066	0.72	0.098	0.11	-0.26	0.012 0.012
Т7 -		-0.14	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		-0.21	0.94	-0.078	0.63	-0.098			-0.11		-0.0039-0.0039
RH_7		0.035		0.8	0.23	0.69	0.17	0.83	0.044	0.89	0.15			0.36	-0.034	1	-0.12	0.88	0.028		0.29	-0.27	0.38		-0.0072	200000	0.0018 0.0018
Т8				-0.03	0.58	-0.041	0.8	-0.28	0.8	-0.17	0.82	-0.087	0.48	-0.67	0.88	-0.12	1	-0.21	0.87	-0.16	0.5	-0.16	-0.3		-0.06		-0.0032-0.0032
_		0.013			0.069	0.68	0.044	0.83	-0.095	0.85	0.016	0.36	0.074	0.49	-0.21	0.88	-0.21	1	-0.11	0.86	0.12	-0.23	0.49		0.046		0.0045 0.0045
T9 -		-0.16	0.84	0.12	0.68	0.055	0.9	-0.2	0.89	-0.045	0.91	-0.14		-0.74	0.94	0.028	0.87	-0.11	1	-0.0087	0.67	-0.16		-0.18	-0.1		-0.0012-0.0012
RH_9		-0.0088		0.76	0.16	0.68	0.13	0.83	-0.026	0.86	0.072		0.18	0.39	-0.078	0.86	-0.16	0.86	-0.0087		0.22		0.36		0.0087		-0.003 -0.003
T_out		-0.074		0.34	0.79	0.034	0.7	0.12	0.66	0.29	0.65	-0.053		-0.64	0.63	0.29	0.5	0.12	0.67	0.22	1	-0.14	-0.57		-0.077	0.79	-0.015 -0.015
Press_mm_hg		-0.011				-0.26		-0.23	200000000000000000000000000000000000000	0.000				NAME OF TAXABLE PARTY.	340000000	-0.27			-0.16		-0.14	1		-0.24			0.0007 0.0007
RH_out					-0.51	0.58	-0.28		-0.39		-0.27		-0.57	0.72	-0.41	0.38	-0.3	0.49	-0.32		-0.57	-0.092	1				
Windspeed						0.069	-0.1		-0.19				0.17				-0.22		-0.18		0.19		-0.18				-0.011 -0.011
Visibility					-0.07															0.0087							-0.0059-0.0059
Tdewpoint				0.64	0.58	0.5	0.65	0.41	0.52	0.62	0.59		0.76		0.47	0.64	0.39	0.5	0.58	0.54	0.79					1	-0.0039-0.0039
rv1																									-0.0059		
rv2	-	0.00052	2-0.0062	-0.0007	-0.011	0.0063	-0.0052 	.0.0004	80.0018 2	-0.0018	-0.0055 	-0.011	-0.015	0.012	-0.0039	0.0018	-0.0032 pe	0.0045	-0.0012 e	-0.003	-	0.0007 E	100	-0.011	-0.0059	0.0039	1 1
	Appliances	light	<u> </u>	퓬	-	포	<u> </u>	Æ,	-	퍒'	—	퓬	<u> </u>	돌'	_	Æ	-	Æ,	<u>-</u>	돌	Tout	Press mm ho	RH_out	Windspee	Visibilit	Tdewpoin	N1 N2



```
# Binning of hour column into different buckets
                                                Feature Engineering
     def hourly basis(hr):
      if 0 <= hr <= 5:
        return 1
                                                                            Binning
      elif 6 <= hr <= 11:
                                                                                                                        Hour
       return 2
      elif 12<= hr <=17:
        return 3
                                                                                                                       Hourly
      else:
                                                                     One Hot Encoding
                                                                                                                        Month
        return 4
                                                                                                                      Weekday
# creating instance of one-hot-encoder
enc = OneHotEncoder(handle unknown='ignore')
                                                                                                                     RH_5,
# passing bridge-types-cat column (label encoded values of bridge types)
                                                                          Skewness
                                                                                                                RH_6, RH_out,
enc df = pd.DataFrame(enc.fit transform(df1[['hourly','month','weekday']]).toarray())
                                                                                                                  Appliances
# merge with main df1 on key values
df1 = df1.join(enc df)
df1.head()
                                                                       Log Transform
                                                                                                               RH_5,RH_6,RH_out
                                                                                                                    Appliances
```





Filter Method

Information Gain Chi-Square test Fischer's score



Embedded Method

Regularization Random Forest Importance





Wrapper Method

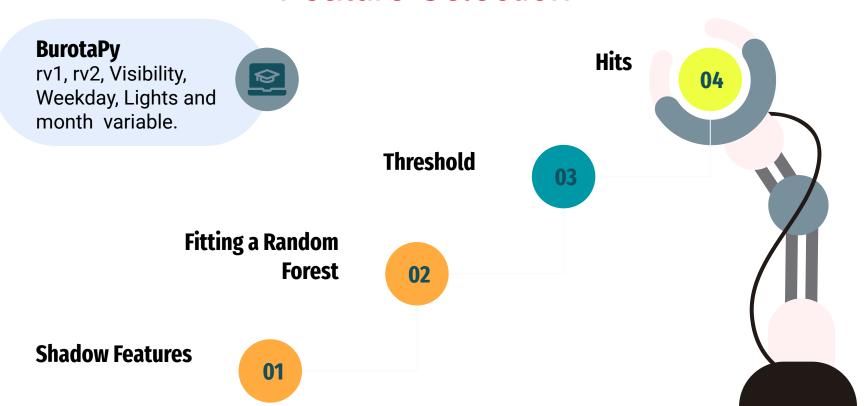
Forward selection Backward Elimination Exhaustive



BurotaPy



Feature Selection







Feature Scaling and Its need



Feature Scaling techniques

Vs



Standerdization

- Values are centred around mean with unit standard deviation
- Values are not restricted to a particular range
- Used when data follows gaussian distribution

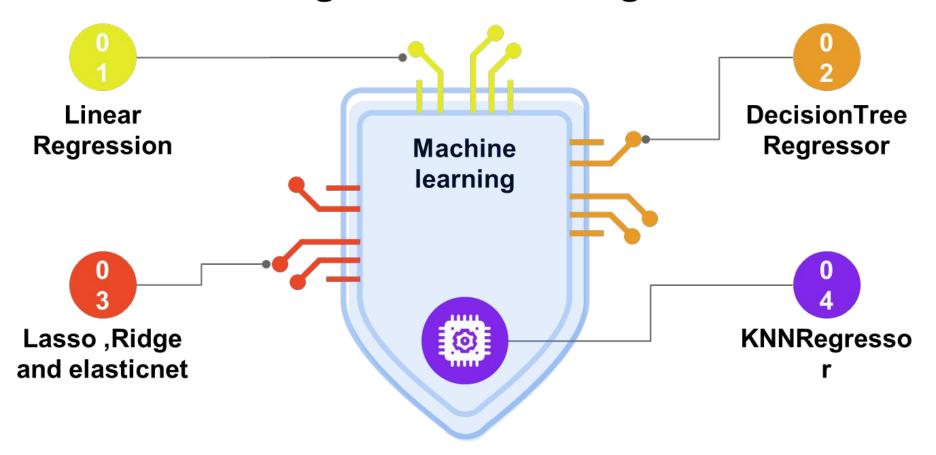


Normalization

- Rescale the values between [0,1]
- More helpful in knn or NN type of algorithms
- Is good to use when data is not following gaussian distribution

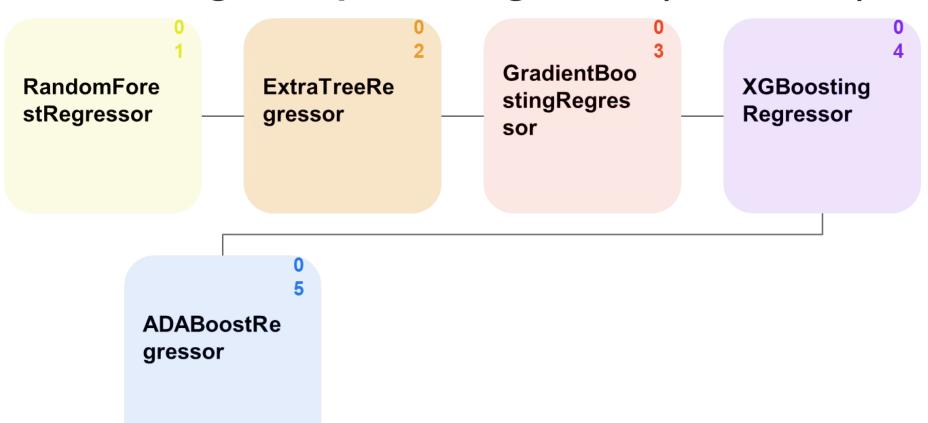


Selecting Best Perfoming Model



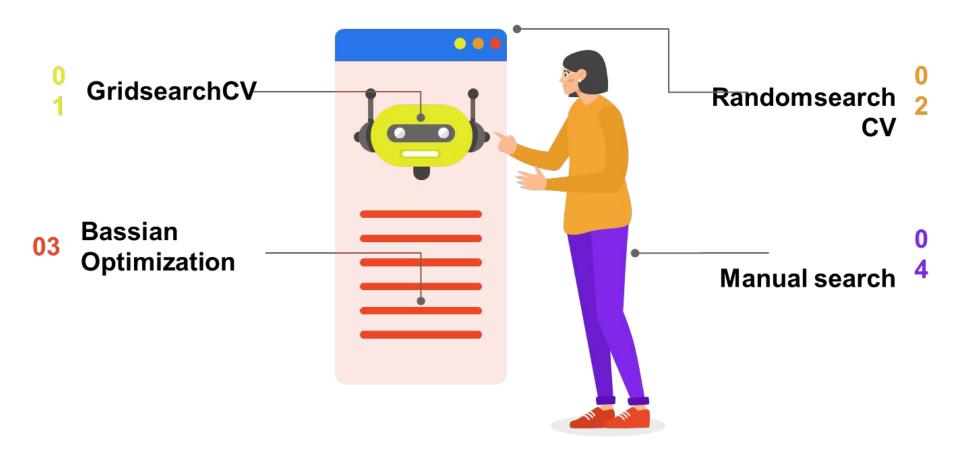


Selecting best perfoming model(Ensamble)

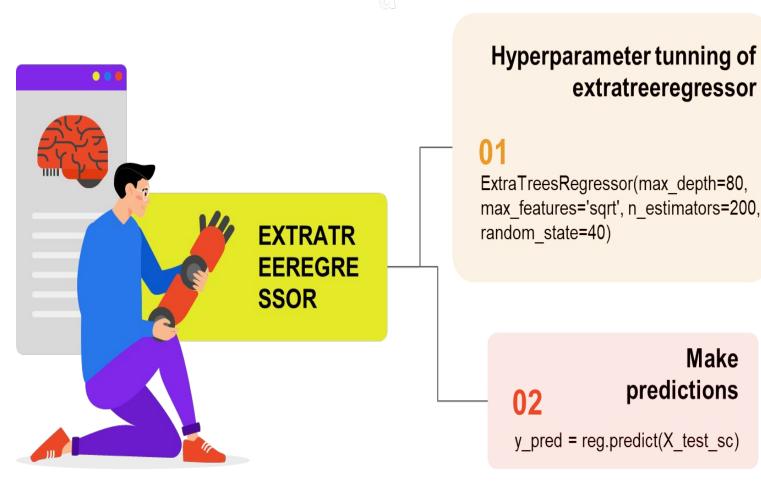




Hyperparameter Tunning

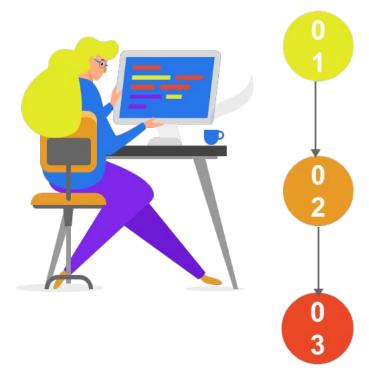








Model Evaluation Metrics



Mean Absolute Error(MAE)

MAE is a very simple metric which calculates the absolute difference between actual and predicted values.(0.65) we aim to get a minimum MAE because this is a loss

Mean Squared Error(MSE)

It represents the squared distance between actual and predicted values. (1.3220) we perform squared to avoid the cancellation of negative terms and it is the benefit of MSE.

Root Mean Squared Error(RMSE)

As RMSE is clear by the name itself, that it is a simple square root of mean squared error.(1.14)

Root Mean Squared Log Error(RMSLE)

It is the Root Mean Squared Error of the log-transformed predicted and log-transformed actual values.(0.1396) the log of the RMSE metric slows down the scale of error.

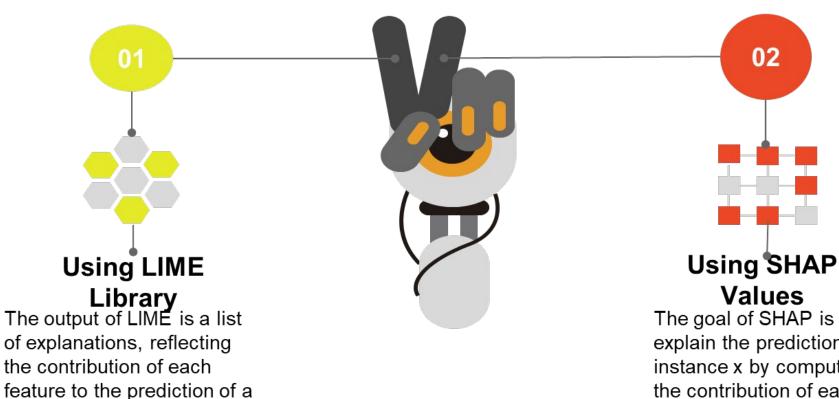
R Squared

0

R2 score is a metric that tells the performance of our model.(0.75) when the R2 score is between zero and one like 0.8 which means your model is capable of explaining 80 per cent of the variance of data.



Model Interpretation



of explanations, reflecting the contribution of each feature to the prediction of a data sample.

The goal of SHAP is to explain the prediction of an instance x by computing the contribution of each feature to the prediction.

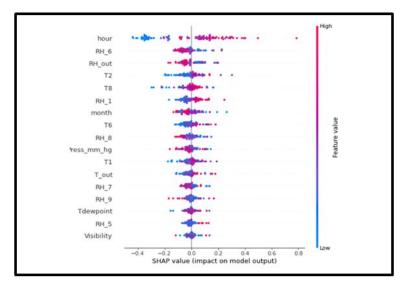
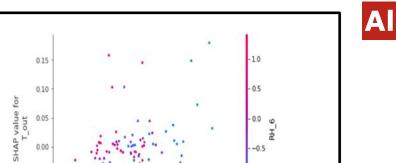


Fig:-SHAP Summary Plot

The summary plot combines feature importance with feature effects. Each point on the summary plot is a Shapley value of an instance per feature.



-1.0

Fig:-SHAP Dependence Plot

T out

A partial dependence plot can show whether the relationship between the target and a feature is linear, monotonic, or more complex.

Conclusion



- In data visualization, we used log transformation to remove the skewness. And on removal of skewness it is observed that it follows normal distribution.
- The energy consumption for each day of the week is at its highest during hours between 4 pm to 8 pm while it is at its lowest between 12 am to 5 am. For other times of the day it doesn't follow a general trend, it differs from day to day.
- As Temperature increases, energy consumption of the Appliances increases from Jan-Feb and March-April. And consumption decreases with increase in temperature for months Feb-March and April-May.

04

1)Filter Method 2) Wrapper Method 3) Embedded Method 4) Burota. We got the best results from the burota method. After performing feature selection using burota five features were dropped - lights, rv1,rv2,weekday, Visibility and month..

- According to best fit model, the 5 most and least important features. The top 3 important features are humidity attributes, which leads to the conclusion that humidity affects power consumption more than temperature. 'Wind Speed' is least important as the speed of wind doesn't affect power consumption inside the house. So controlling humidity inside the house may lead to energy savings.
- The ExtraTree Regressor model came out to be the best model with an R2 score of 75.10 Extra Trees Regressor, the features and splits are selected at random.



Thank You