WORLD
RESOURCES
INSTITUTE

Global powerplant database

Covering approximately 30,000 power plants from 164 countries and includes thermal plants (e.g. coal, gas, oil, nuclear, biomass, waste, geothermal) and renewables (e.g. hydro, wind, solar)

Michael Wenzelhumer

Objective:

WORLD RESOURCES INSTITUTE

To predict gigawatt per hour output using :

https://datasets.wri.org/dataset/globalpowerplantdatabase

capacity_mw

latitude

longitude

prior2

prior1

Current

Data cleaning:

WORLD RESOURCES INSTITUTE

-Dataset was approx 29000 rows with 24 columns

-Removed all categorical columns to get to final 6 for exploration and 3 concatenated datasets.

	capacity_mw	latitude	longitude	prior2	prior1	current
0	33.00	32.3220	65.1190	0.0	0.0	0.0
1	66.00	34.5560	69.4787	0.0	0.0	0.0
2	100.00	34.6410	69.7170	0.0	0.0	0.0
3	11.55	34.4847	70.3633	0.0	0.0	0.0
4	42.00	34.5638	69.1134	0.0	0.0	0.0
...	...	...	...	...	...	...
89725	50.00	-12.9667	28.6333	0.0	0.0	0.0
89726	20.00	-12.8167	28.2000	0.0	0.0	0.0
89727	108.00	-17.9167	25.8500	0.0	0.0	0.0
89728	920.00	-18.3835	26.4700	0.0	0.0	0.0
89729	750.00	-16.5222	28.7619	0.0	0.0	0.0

89730 rows × 6 columns

Concatenation:

WORLD RESOURCES INSTITUTE

-Applied Lambda Function to normalize data

Concatenated 3 seperate data sets into 1:

Giving new dataset of 89730 rows

-This was done to get output values for

2013 -2014-2015-2016-207 and convert to prior 2 prior 1 and current(2017)

Data collection

WORLD RESOURCES INSTITUTE

Values to be tested:

	capacity_mw	latitude	longitude	prior2	prior1	current
count	89730.000000	89730.000000	89730.000000	89730.000000	89730.000000	89730.000000
mean	186.294810	32.497558	-12.458680	169.183678	173.764769	178.141796
std	525.697713	23.654208	76.977407	1188.848308	1193.766551	1198.127832
min	1.000000	-77.847000	-179.977700	-989.619000	-989.619000	-934.944000
25%	4.772240	28.861200	-79.211900	0.000000	0.000000	0.000000
50%	18.900000	40.071250	-3.746700	0.000000	0.000000	0.000000
75%	100.000000	47.134900	24.933200	0.000000	0.000000	0.225000
max	22500.000000	71.292000	179.388700	59546.865000	59546.865000	59546.865000

Model results

WORLD RESOURCES INSTITUTE

Linear Regression:

0.8028505579010543
0.8292692491000222

Root mean Squared error:

mse is: 213561.5927061941
rmse is: 462.12724730986605

Decision Tree Regressor:

R squared is: 0.8065837936494216
mse is: 241938.09765189045
rmse is: 491.8720338176287

Random Forest Regressor:

R squared is: 0.876757516996191
mse is: 154160.04920390082
rmse is: 392.63220601970596

Results/findings

WORLD RESOURCES INSTITUTE

KNN Regressor:

R squared is: 0.8606902117234034
mse is: 174258.12343168532
rmse is: 417.442359412273

Linear regression after adding clusters:

0.5828643982315611
R squared is: 0.5741420296834572

Decision tree after clusters:

R squared is: 0.7808436596373483
mse is: 274135.60154098034
rmse is: 523.5796038244617

After train test split

WORLD RESOURCES INSTITUTE

RandomForestRegressor

R squared is: 0.8311492012368606
mse is: 211210.02117945813
rmse is: 459.5759144901505

KNeighborsRegressor

R squared is: 0.78081714636564
mse is: 274168.766137891
rmse is: 523.6112738834898

Conclusion:

WORLD RESOURCES INSTITUTE

-Dataset had many Nan values to be taken into consideration

-Need more practice cleaning and preprocessing

-Visualizations in improved model

-Include OLS and Kelbow

-Include secondary fuel sources into dataset