Combined Cycle Power Plant Performance

Sandra Poza

Abstract

This project studies the relationship between weather conditions and performance of a combined cycle power plant (CCPP). We used a dataset containing CCPP performance to train a regression model that predict the power output of the plant.

Then with data from US weather datasets we predicted the performance it would have in different US locations.

It was found that it would perform better in coldest weather in northern areas.

Motivation

Energy efficiency has become an essential matter in todays society. For climate, economic and sustainability reasons, every one gets benefit from a more effective use of energy.

In this project we try to find out where would a Combined Cycle Power Plant (CCPP) have the highest power output (which would mean a highest efficiency), based on the weather conditions.

Dataset(s)

Combined Cycle Power Plant Data Set: The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011). Obtained from UCI's Machine learning repository.

US Weather datasets: Obtained from NOAA National Centers for Environmental Information

- **Temperatures dataset:** Contains monthly and yearly average temperatures for different US weather stations.
- **Relative Humidity dataset:** Contains monthly and yearly average relative humidity for different US weather stations.
- **Weather station locations dataset:** Contains weather stations information with elevation, used for ambient pressure calculation.

US cities dataset: Contains US cities information, including the fips codes. Obtained from https://simplemaps.com

Methods

The Combined Cycle Power Plant Dataset contains data of a CCPP working at full load over 6 years. This include the following features:

Temperature (T) [°C]
Ambient Pressure (AP) [milibar]
Relative Humidity (RH) [%]
Energy output (EP) [MW] of the plant

With this data we will train a model that predict **Energy output** based on the 3 first. We will use scikit-learn, and train the model as a multiple linear regression.

Then, from the weather datasets we are going to extract information about the average Temperature, Relative Humidity and Ambient Pressure in different US locations.

Then we will use the model previously trained with the CCPP data, and feed the weather average data from different locations to predict the expected power output of a CCPP in this locations.

We will use Matplotlib and Plotly for plotting results.

Findings

Multiple Linear Regression Model Accuracy

The mean squared error for the model with test data is 4.65. Taking into consideration that the output of the model is in the order of 400-500 it is a good number.

```
RMSE = sqrt(mean_squared_error(y_true = Ytest, y_pred = Ypred))
RMSE
```

```
4.657571391074255
```

A better stimator is the R2 score. This is a statistical measure of how well the regression predictions approximate the real data points. It ranges from 0-1 being 1 the best possible result.
Our R2 score is 0.92, what means that **the model is highly accurate** on test data.

```
from sklearn.metrics import r2_score

r2 = r2_score(Ytest, Ypred)
r2
```

```
0.9266765701220129
```

Findings

Performance prediction at US locations

We can appreciate how the energy output is higher in the coldest areas like Alaska, and lower in the warmer ones like Florida.

	Wban	City	State	Temperature	Humidity	Pressure	Power_out
6	27502	BARROW	AK	-11.222222	81.0	1012.164777	526.034436
147	14755	MT. WASHINGTON	NH	-2.555556	0.0	812.822902	516.454667
8	26533	BETTLES	AK	-4.722222	67.0	991.362149	512.871904
17	26616	KOTZEBUE	AK	-5.055556	78.0	1012.210452	512.005739
11	26411	FAIRBANKS	AK	-2.388889	65.0	998.288099	507.921088
18	26510	MCGRATH	AK	-2.555556	68.5	1001.703277	507.703816
19	26617	NOME	AK	-2.555556	76.0	1012.792951	506.490944
12	26425	GULKANA	AK	-2.111111	66.0	959.996103	506.043838
9	26415	BIG DELTA	AK	-1.666667	62.5	969.545252	505.946333
7	26615	BETHEL	AK	-0.722222	78.5	1009.700784	501.556013

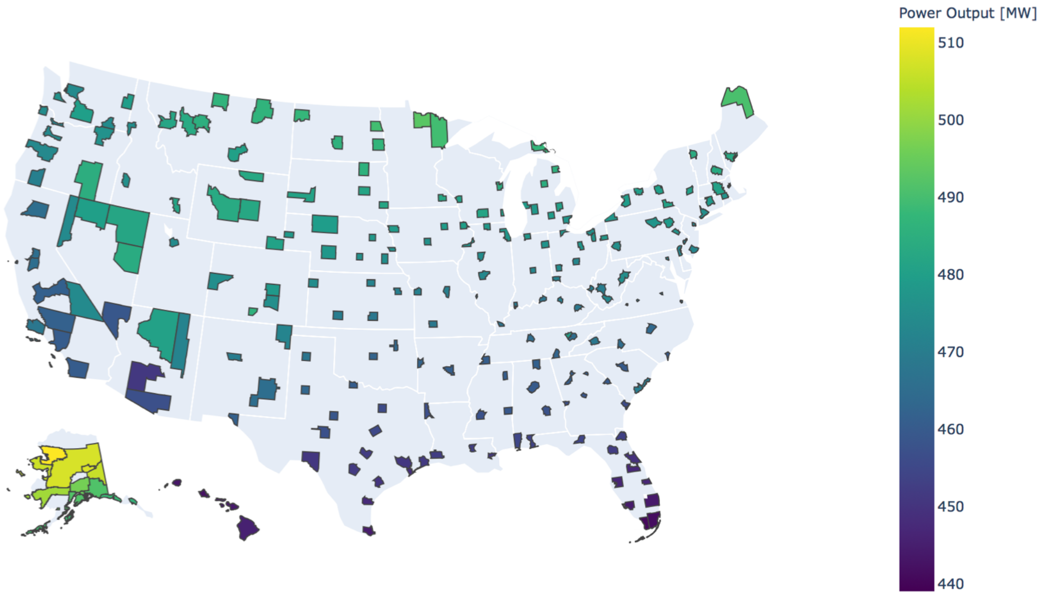
10 Highest Energy Output

	Wban	City	State	Temperature	Humidity	Pressure	Power_out
61	12842	TAMPA	FL	23.000000	72.0	1012.587332	446.648664
70	21504	HILO	HI	23.277778	74.0	1011.925011	445.569949
211	12919	BROWNSVILLE	TX	23.611111	75.0	1012.416009	444.591033
53	12835	FORT MYERS	FL	23.944444	71.5	1012.724408	444.511729
72	22516	KAHULUI	HI	24.388889	67.5	1011.479853	444.228015
63	12844	WEST PALM BEACH	FL	24.111111	71.5	1012.587332	444.112604
73	22536	LIHUE	HI	24.333333	72.0	1009.769163	443.409941
71	22521	HONOLULU	HI	25.388889	65.0	1013.010029	442.398239
57	12839	MIAMI	FL	25.111111	71.5	1012.244709	441.730603
56	12836	KEY WEST	FL	25.444444	73.5	1013.215716	440.563445

10 Lowest Energy Output

Findings

Performance prediction at US locations



This map shows the power output that it would be obtained by a CCPP situated in different US counties.

We can observe the relationship between southern, warmer, and more humid areas with lower power output

Also the northern, colder and drier areas whit highest power output.

It should be mentioned that Alaska is represented on the bottom left, but is in fact the northernmost state, with the lowest temperatures, so we can appreciate how that region is the most yellowish.

Limitations

We did not have enough weather data to study the whole US map.

The weather data is average, so it doesn’t reflect seasonal changes.

The CCPP dataset measures was taken in a fixed location, so the pressure changes were not well represented.

Conclusions

We can conclude, that the performance of a Combined Cycle Power Plant would be the best in the coldest and driest possible weathers.

A CCPP situated in Alaska would have the highest efficiency, and one located in Florida the lowest. That means that the same power plant would consume less gas in Alaska to produce the same power output.