Study on factors affecting power output of a combined cycle power plant and prediction based on data.

Tannistha Chakraborty
Mousumi Dasgupta
Subham Dey
Sayak Acharjee
Prateek Majumder

Guide:
Prof. Madhumita Pal &
Prof. Amartya Mukherjee

Combined Cycle Power Plant

- The Combined Cycle Power Plant or ***combined cycle gas turbine***, a gas turbine generator generates electricity and waste heat is used to make steam to generate additional electricity via a steam turbine.

- A Combined Cycle Power Plant produces high power outputs at high efficiencies (up to 55%) and with **low emissions**. In a Conventional power plant we are getting ***33% electricity only*** and remaining 67% ***as waste***.
- By using combined cycle power plant we are getting ***68% electricity***.

Inner Workings of a Combined-Cycle Power Plant

- A combined-cycle power plant uses both a gas and a steam turbine together to produce up to 50 percent more electricity from the same fuel than a traditional simple-cycle plant. The waste heat from the gas turbine is routed to the nearby steam turbine, which generates extra power.

How a Combined-Cycle Power Plant Produces Electricity

- **Gas turbine burns fuel:**

  The fast-spinning turbine drives a generator that converts a portion of the spinning energy into electricity.

- **Heat recovery system captures exhaust:**

  The HRSG creates steam from the gas turbine exhaust heat and delivers it to the steam turbine.

➡ **Steam turbine delivers additional electricity:**

The steam turbine sends its energy to the generator drive shaft, where it is converted into additional electricity.

Advantages of Combined Cycle Power Plant

- The efficiency of the combined cycle plant is better or higher than the turbine cycle or steam cycle plant. The efficiency of combined cycle power plant will be of the order of about 45 to 50%.
- fewer moving parts and less vibration than a reciprocating engine
- very low toxic emissions
- runs on a wide variety of fuels
- high operating speeds

Disadvantages of Combined Cycle Power Plant

- Higher cost
- longer start-up
- less responsive to power demands
- shrill whining noise.

Data Set Information:

- The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the power plant was set to work with full load. Features consist of hourly average ambient variables Temperature (T), Ambient Pressure (AP), Relative Humidity (RH) and Exhaust Vacuum (V) to predict the net hourly electrical energy output (EP) of the plant.

- A combined cycle power plant (CCPP) is composed of gas turbines (GT), steam turbines (ST) and heat recovery steam generators. In a CCPP, the electricity is generated by gas and steam turbines, which are combined in one cycle, and is transferred from one turbine to another. While the Vacuum is collected from and has effect on the Steam Turbine, the other three of the ambient variables effect the GT performance.

Attribute Information:

Features consist of hourly average ambient variables

- Temperature (T) in the range 1.81°C and 37.11°C,

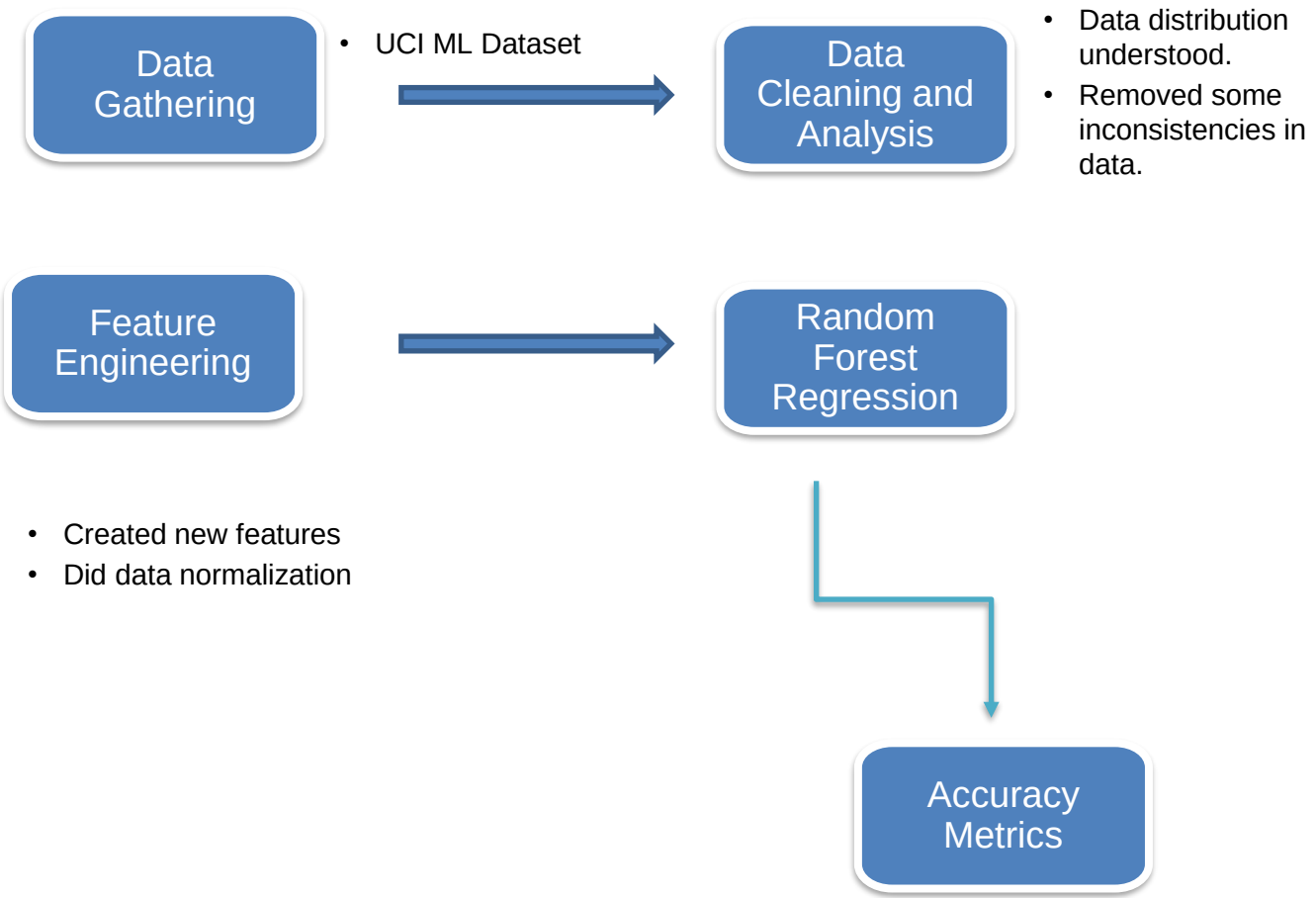
- Ambient Pressure (AP) in the range 992.89-1033.30 millibar,

- Relative Humidity (RH) in the range 25.56% to 100.16%

- Exhaust Vacuum (V) in the range 25.36-81.56 cm Hg

- Net hourly electrical energy output (EP) 420.26-495.76 MW

The averages are taken from various sensors located around the plant that record the ambient variables every second. The variables are given without normalization.


Data Gathering
UCI ML Dataset
Data Cleaning and Analysis
Data distribution understood.
Removed some inconsistencies in data.
Feature Engineering
Random Forest Regression
Created new features
Did data normalization
Accuracy Metrics

	AT	V	AP	RH	PE
0	14.96	41.76	1024.07	73.17	463.26
1	25.18	62.96	1020.04	59.08	444.37
2	5.11	39.4	1012.16	92.14	488.56
3	20.86	57.32	1010.24	76.64	446.48
4	10.82	37.5	1009.23	96.62	473.9

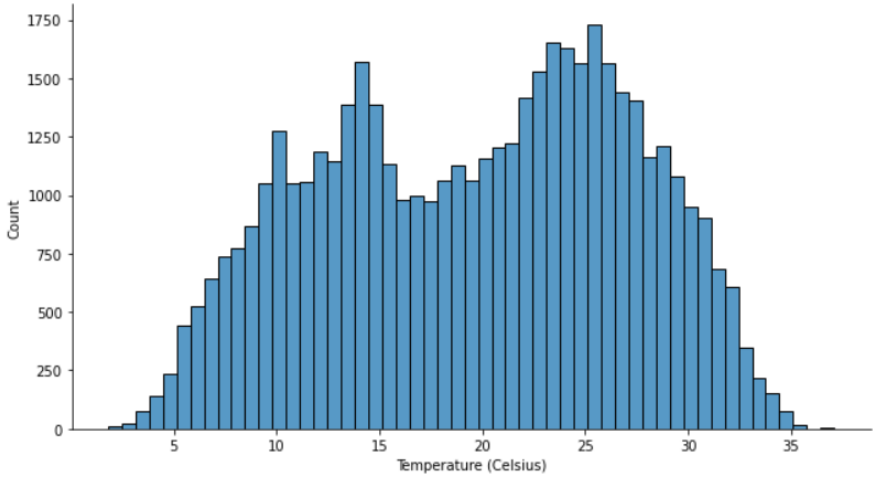
Attribute Information:

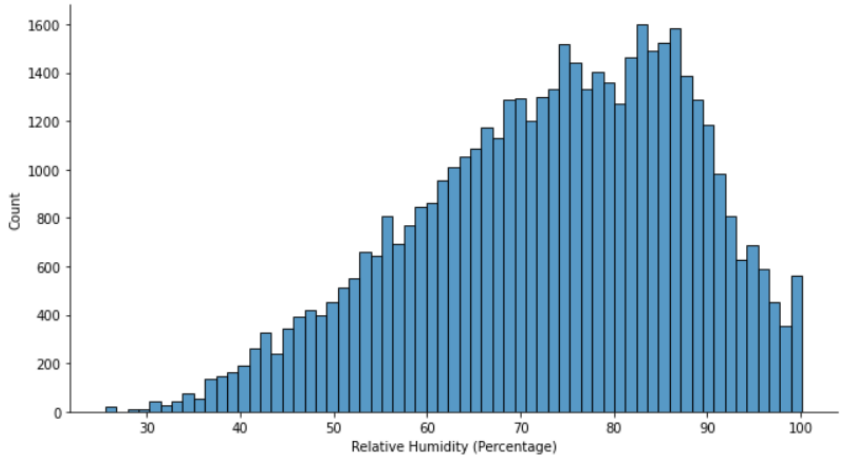
Features consist of hourly average ambient variables
- Temperature (T) in the range 1.81°C and 37.11°C,
- Ambient Pressure (AP) in the range 992.89-1033.30 milibar,
- Relative Humidity (RH) in the range 25.56% to 100.16%
- Exhaust Vacuum (V) in teh range 25.36-81.56 cm Hg
- Net hourly electrical energy output (EP) 420.26-495.76 MW

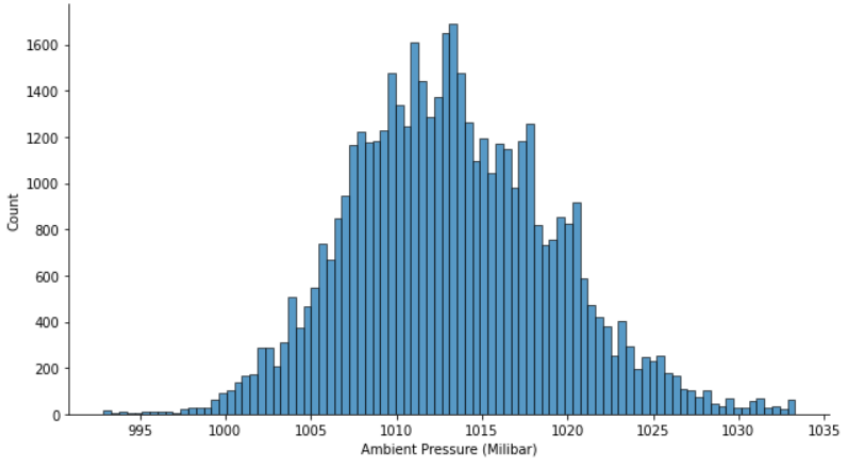
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 47844 entries, 0 to 47843
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   AT      47844 non-null  object
 1   V       47844 non-null  object
 2   AP      47844 non-null  object
 3   RH      47844 non-null  object
 4   PE      47844 non-null  object
dtypes: object(5)
memory usage: 2.2+ MB
```

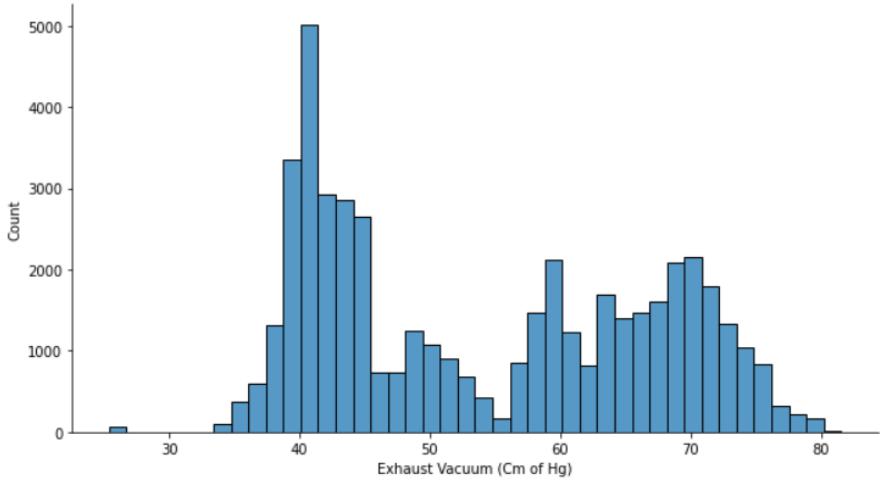
```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import preprocessing
from sklearn.model_selection import train_test_split

%matplotlib inline

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

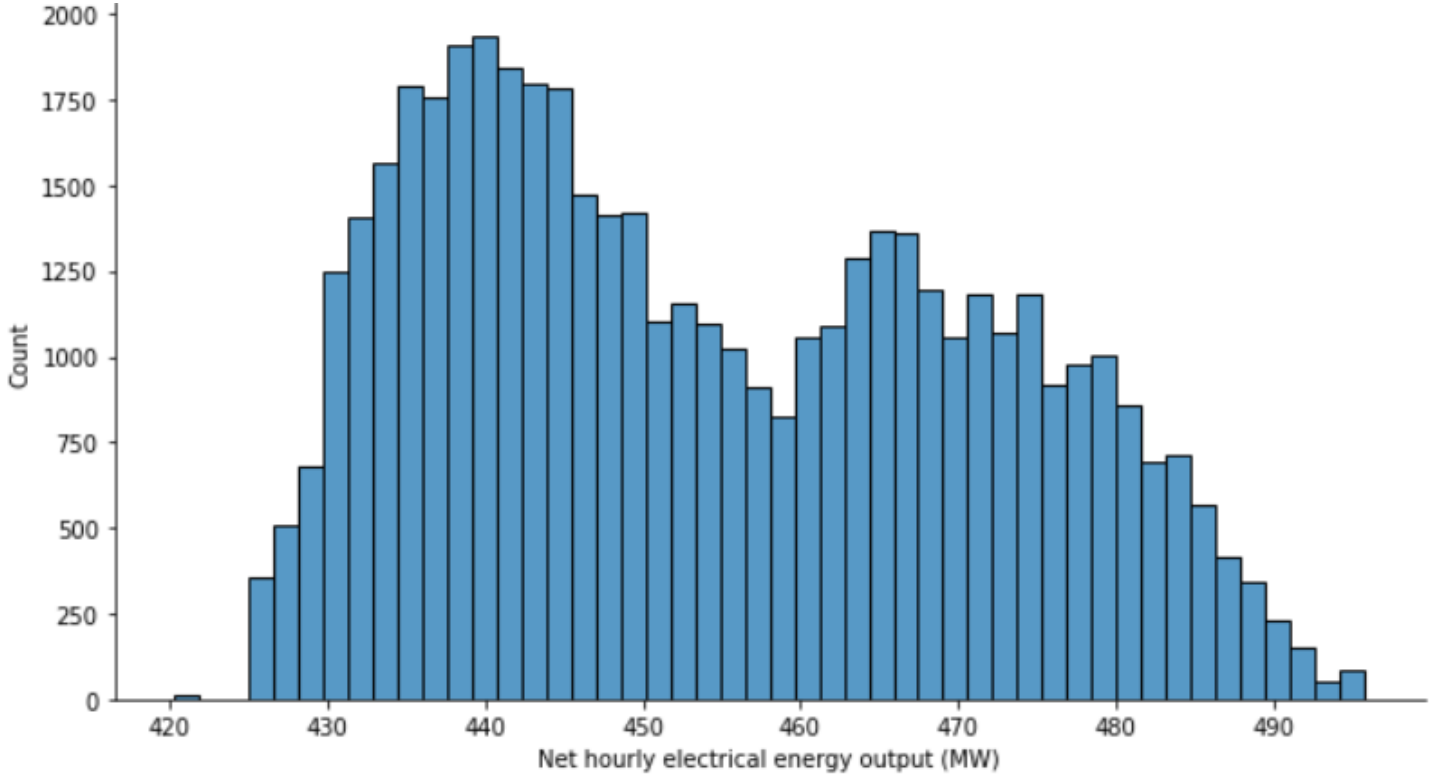
```
/kaggle/input/uci-ml-combined-cycle-power-plant-data-set/Folds5x2_pp.csv
```


Count
Temperature (Celsius)


Count
Relative Humidity (Percentage)


Count
Ambient Pressure (Milibar)


Count
Exhaust Vacuum (Cm of Hg)


2000
1750
1500
1250
1000
750
500
250
0
Count
420
430
440
450
460
470
480
490
Net hourly electrical energy output (MW)

```
from sklearn.metrics import r2_score

r2_score(y_test, y_pred)
```

```
0.9389030571854318
```

```
]:
from sklearn.metrics import mean_absolute_error

mean_absolute_error(y_test, y_pred)
```

```
']:
3.2532696176907088
```

```
from sklearn.metrics import mean_squared_error

mean_squared_error(y_test, y_pred)
```

```
17.72997182143494
```

```
import math

from sklearn.metrics import mean_squared_error

print (math.sqrt(mean_squared_error(y_test, y_pred)))
```

```
4.210697308218075
```