

# Combined cycle power plant performance prediction



Presented by  
Artem Ramus

# Background

## **Abstract**

The data set contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the plant was set to work with full load.

## **Source**

Pınar Tüfekci, Çorlu Faculty of Engineering, Namık Kemal University, TR-59860 Çorlu, Tekirdağ, Turkey Email: [ptufekci@nku.edu.tr](mailto:ptufekci@nku.edu.tr)

Heysem Kaya, Department of Computer Engineering, Boğaziçi University, TR-34342, Beşiktaş, İstanbul, Turkey Email: [heysem@boun.edu.tr](mailto:heysem@boun.edu.tr)

<https://archive.ics.uci.edu/ml/datasets/Combined+Cycle+Power+Plant>

# Introduction

The goal of this project is to predict combined power plant output from ambient properties and exhaust vacuum.

## **Features consist of hourly average ambient variables**

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

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

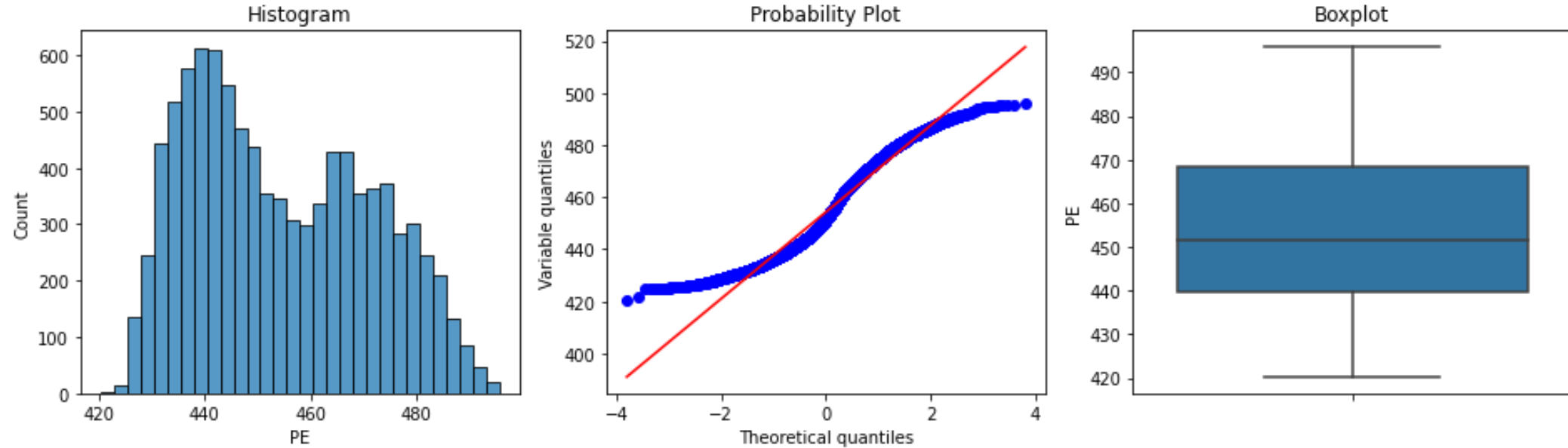
Relative Humidity (RH) in range 25.56% to 100.16%

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

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

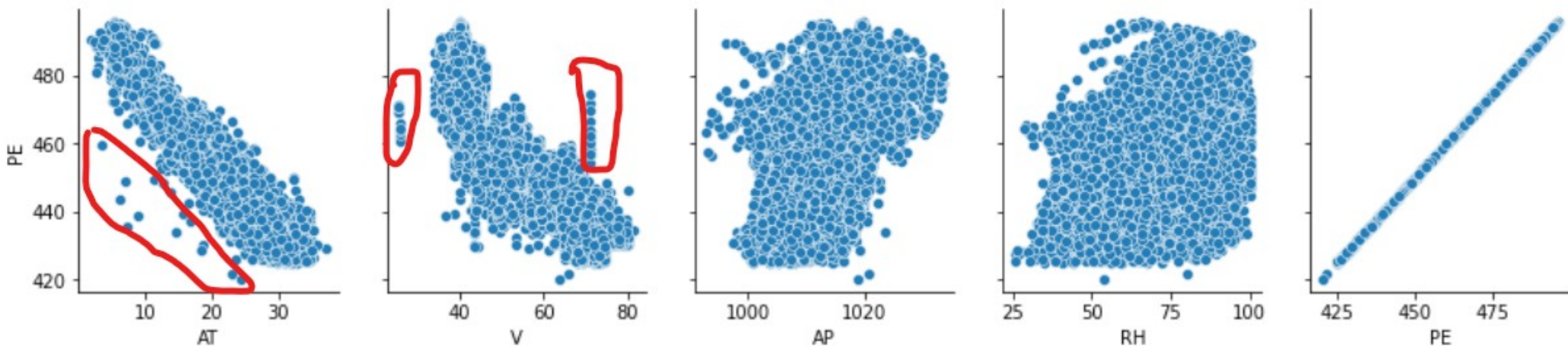
The averages are taken from various sensors located around the plant that record the ambient variables every second.

# Exploratory Data Analysis



The target's distribution seems to be combined from two distributions.  
Variables' distributions are close to normal distribution.

# Exploratory Data Analysis



AT and V have great influence on the target PE.

AT-PE and scatter show possible outliers in left low area. V-PE scatter show possible outliers: a line near V=20 and and a line above the main 'swarm' near V=70. This atypical distribution might be related to rare weather condition, to special operation mode, sampling strategy or error in data.

# Feature Engineering and Selection

Original distribution of all the variables selected.

Based on single feature shuffling method, variable 'RH' showed least influence of about 0.2%.

Selected variables: 'AT', 'V' and 'AP'.

# Summary and Conclusions

Energy output was predicted using XGBoost regression.

RMSE\_train is 12.77 and RMSE\_test - 15.68 MW.

R2\_train is 0.96 and R2\_test - 0.95.

The end

Thank you for your attention!