**Artificial Intelligence for Engineering**

**COS 40007**

Mohammed Barsat Zulkarnine

103810626

Studio 1 – 5

# Introduction

# Dataset Selection

The selected dataset contains 9568 data points collected from a Combined Cycle Power Plant over six years (2006-2011). The features consist of hourly average ambient variables:

* Temperature (T) in the range 1.81°C to 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 the range 25.36-81.56 cm Hg
* Net hourly electrical energy output (PE) in the range 420.26-495.76 MW

As a software engineering student all the available datasets are suitable, however as someone who has a lot of interest in energy, I chose the Combined Cycle Power Plant dataset.

Understanding the raw data

The data was explo  
It was found that there were 41 duplicates. AP and RH had outliers and there were no missing data in the dataset. The figure below shows the data with the outliers  
Cleaning the data

41 rows were dropped. The data which didn’t fit the IQR were removed. The figure below shows the data with the outlikers removed.

EDA analysis

Target variable:

Predictors (Input varibles):

Univariate Analysis

Numerical columns – Temperature, Vacuum, Pressure, Humidity, Energy Output

A screenshot of a computer screen

Description automatically generated Figure shows univariate analysis

Temperature (AT) Column:  
Distribution: Bimodal distribution with two peaks.

Exhaust Vacuum (V) Column:  
Distribution: Bimodal distribution with peaks at lower and higher values.

Ambient Pressure (AP) Column:  
Distribution: Approximately normal distribution.

Relative Humidity (RH) Column:  
Distribution: Right skewed distribution.

Net hourly electrical energy output (PE) Column:  
Distribution: Bimodal distribution with peaks around 435 and 465.

Multivariate Analysis

Class labeling

The target variable (PE) was converted into categorical classes to simplify the model development. The class labels were defined based on the range of PE values:

* Class 1: 420.26-435 MW
* Class 2: 435-450 MW
* Class 3: 450-465 MW
* Class 4: 465-480 MW
* Class 5: 480-495.76 MW

This ensured a nearly balanced distribution of classes.

Featured engineering

Feature selection

Model development