**Project on Predicting the Turbine\_Energy\_Yield Using TensorFlow Neural Network**

**Project Overview**

**Dataset Description:**

The dataset contains 36733 instances of 11 sensor measures aggregated over one hour (by means of average or sum) from a gas turbine. The Dataset includes gas turbine parameters (such as Turbine Inlet Temperature and Compressor Discharge pressure) in addition to the ambient variables.

**Problem statement: predicting turbine energy yield (TEY) using ambient variables as features.**

Attribute Information:

The explanations of sensor measurements and their brief statistics are given below.

Variable (Abbr.) Unit Min Max Mean

Ambient temperature (AT) C â€“6.23 37.10 17.71

Ambient pressure (AP) mbar 985.85 1036.56 1013.07

Ambient humidity (AH) (%) 24.08 100.20 77.87

Air filter difference pressure (AFDP) mbar 2.09 7.61 3.93

Gas turbine exhaust pressure (GTEP) mbar 17.70 40.72 25.56

Turbine inlet temperature (TIT) C 1000.85 1100.89 1081.43

Turbine after temperature (TAT) C 511.04 550.61 546.16

Compressor discharge pressure (CDP) mbar 9.85 15.16 12.06

Turbine energy yield (TEY) MWH 100.02 179.50 133.51

Carbon monoxide (CO) mg/m3 0.00 44.10 2.37

Nitrogen oxides (NOx) mg/m3 25.90 119.91 65.29

**Methodology**

**Data Exploration and Preprocessing:**

* Import necessary Packages like tensorflow, pandas, numpy, seaborn, sklearn, keras in the Jupyter Environment.
* Checked Data Information (no null values).
* Explored basic statistics of the data using describe.

**Exploratory Data Analysis:**

Heatmap : Plotted a heatmap to find the correlation between the columns. We found the dependent column Turbine Energy Yield (TEY) is more dependent on Air filter difference pressure, Gas turbine exhaust pressure, Turbine inlet temperature, Compressor discharge pressure.

**Data Preparation:**

TEY is assigned as the dependent variable(y). All the other columns are assigned as the independent variables(X).

**Data Scaling, Fitting and Transform:**

Data scaling is a recommended pre-processing step when working with **deep learning neural networks**. Data scaling can be achieved by normalizing or standardizing real-valued input and output variables.

**Splitting the Data for Training and Testing:**

This process involves splitting the data for training and testing. The training data is splitted into 75% and testing data into 25%.

**Model using tensorflow:**

TensorFlow is used to build and train deep learning models as it facilitates the creation of computational graphs and efficient execution on various hardware platforms.

* Here the Regression model is created using TensorFlow.
* The dense layer is a fundamental building block in neural networks.
* The dense layer is created with 64 neurons or units.
* It’s also known as a fully connected layer.
* The compilation (coverting code to binary form) is from high level to machine code is done with the following parameters:
* Optimizer = adam
* Loss =mean\_squared\_error
* Metrics = mean\_absolute\_error and mean\_squared\_error

**Fitting the model:**

The model is fitted with the training data with epochs = 1000. Epochs is the number of iterations that passes through the entire data.

**Prediction:**

The data is tested with the training data and the values are predicted for testing data.

**Evaluation of the model:**

* The model predicted an accuracy of **100%.**
* The mean\_squared\_error is **0.53.**
* The mean\_squared\_log\_error is **0.0**.
* The above metrics conclude this is a **great model**.