

Power Plant Energy Output - ANN Regression

<https://drive.google.com/drive/folders/1EagV-DlhkkEDCydho2ynxBThPbv7FG40?usp=sharing>

Context :

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

Data Dictionary:

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 the range 25.36-81.56 cm Hg

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

This is the regressand (dependent variable) for the given task.

Criteria :

- Defining problem statement, importing the data and data structure analysis
- Null value handling
- Data cleaning
- Visualisation for various features (Univariate and Bivariate)
- Outlier Treatment
- Data splitting
- Data scaling
- Defining NN architecture

- Trying different combinations and hyperparameters
- Model training
- Model testing (RMSE, MSE, R2 score, Adjusted R2 score)
- Plotting the trend of validation loss and training loss over the epochs

Note : Appropriate business insights and comments must be added.

Follow-up Questions :

- Why do we need to check for outliers in our data?
- Why is scaling required for neural networks?
- Briefly explain your choice of optimizer.
- Which activation function did you use and why?