Gas Turbine Analysis Report

Suraj Meshram

Topic- the Emissions/process data from Gas Turbine (GT) system

---

Goals

Our goal is to analyse a dataset containing ambient, process, and emission variables from a gas turbine in Turkey to discover what relationships might exist between Carbon Monoxide emission and NOX.

Abstract

Predictive emission monitoring systems are important tools for validation and backing up of costly continuous emission monitoring systems used in gas-turbine-based power plants. Their implementation relies on the availability of appropriate and ecologically valid data. In this paper, we introduce a dataset collected over five years from a gas turbine for the predictive modelling of the CO and NOx emissions. We analyze the data using a recent machine learning algorithm, and present useful insights about emission predictions

Objective

• To develop a model or Emissions (CO, NOx) and estimate the important contributing factors

• Development of the what-if tool box to create the scenarios of the input parameters and see the corresponding values of CO, NOx)

• Dashboard using python for time-series trending of the parameters

Analysis On Datasets

In this analysis performed EDA, pre-processing, build different models, visualized feature importance and did prediction.

The code is implemented in Python and ML supervised Models are applied and identified the best model using the most common evaluation metrics for regression like R2, MSE and RMSE score best on conclusion draw.

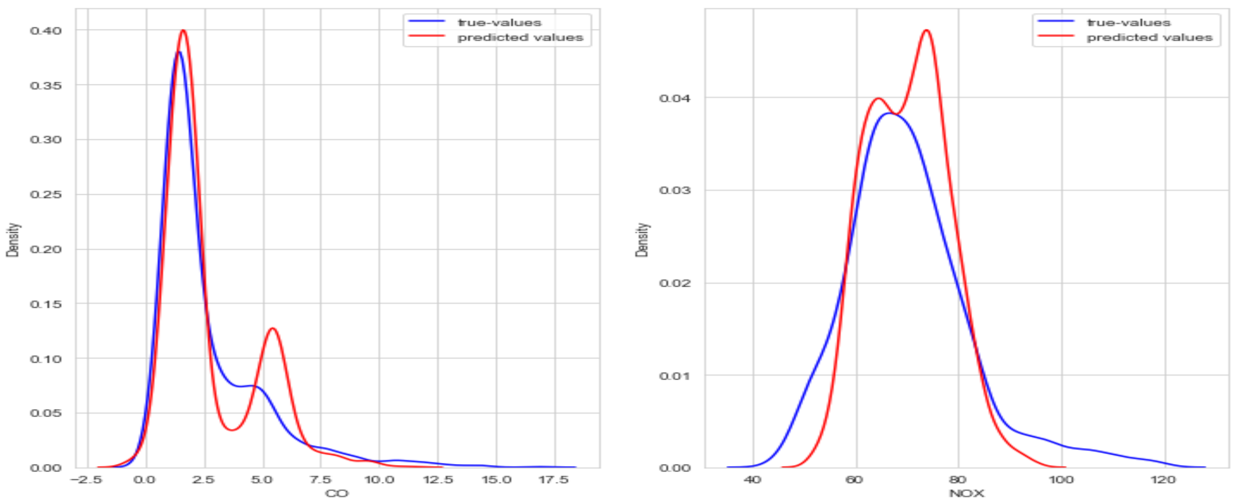
When I analysis dataset, its observed that dataset contain 11 variables so I considered 9 predictor variable and 2 response variables. The response variable in continuous so I use Multi-Output Regression models.

Approach 1: Multiple Linear Regression Model

Now that we have the important features, we can perform Multioutput Regression on that data. We will use train_test_split method to split our data in 75% training data and 25% test data. But before that we will create a new data frame which contains only the features x and Y which contains two the target variable.

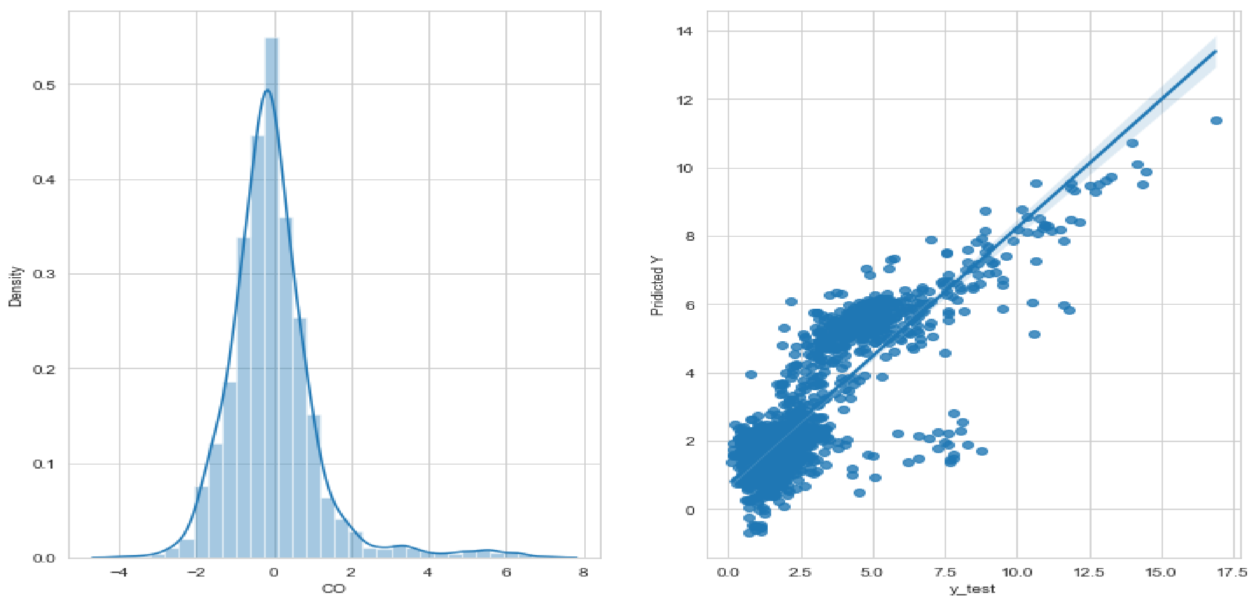
Then I performed standard scaling because our measured values in unit and magnitude its effect the model performance so the result of **standardization** that the features will be rescaled to ensure the mean and the standard deviation to be 0 and 1, respectively and its follow gaussian distribution.

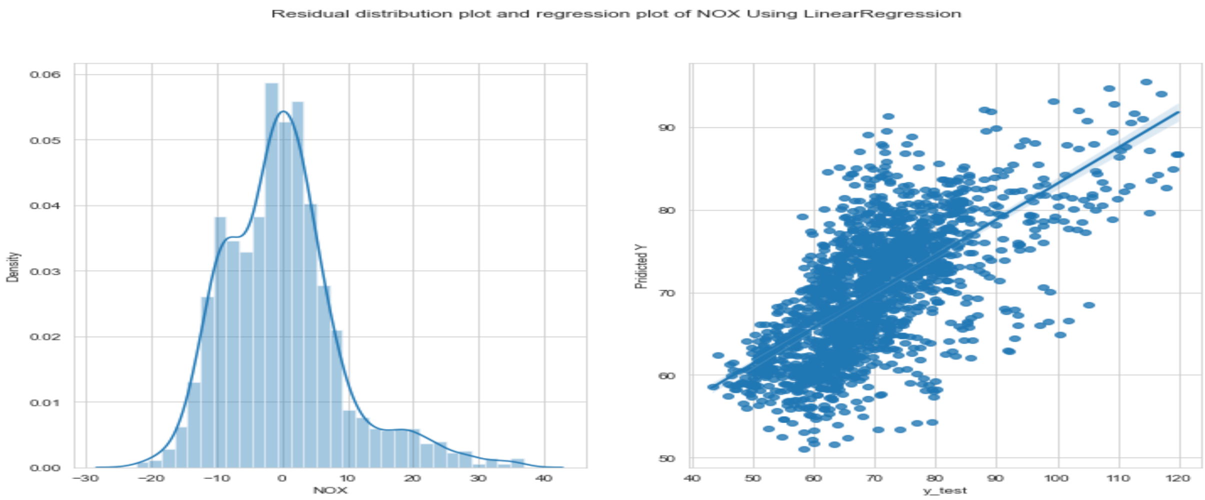
True Value Vs Predicted value of CO and NOX Using LinearRegression



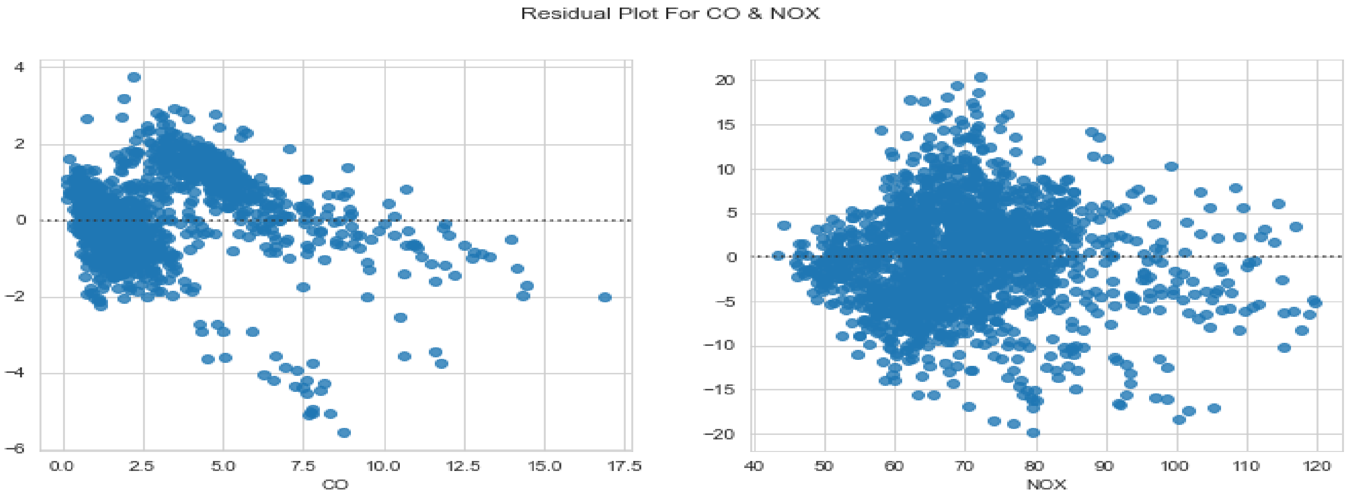
As we can see in the left plot, the predictions made by our modal on the test data of CO is not accurate around the range 1-2 and 4-6 as it varies by a greater extent. similarly, right plot the test data of NOX also not accurate around 60-70.

Residual distribution plot and regression plot of CO Using LinearRegression





On Left they are symmetrically distributed. The model for the chart on the far right is not strong relationships, the model's predictions are not very good at all.



In this residual plot, the data points are above the residual=0 line so we conclude that a linear model is not a right fit for the data.

<u>Coefficient Table:</u>

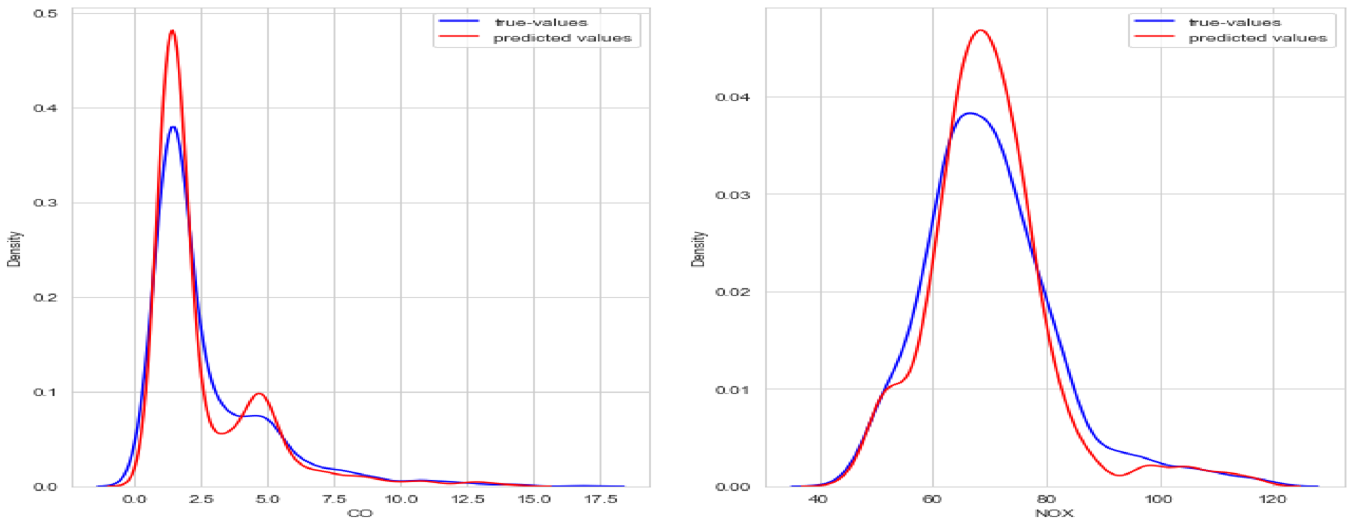
X_i	Coefficient Of CO	Coefficient Of NOX
AT	0.474438	-14.827805
AP	-0.098817	-1.599383
AH	0.258317	-3.724036
AFDP	-0.024943	-0.725617
GTEP	-5.952363	-3.461453
TIT	0.278104	14.711388
TAT	-2.518002	-3.984927
TEY	-0.249301	-43.553020
CDP	2.586461	29.286270

In this table, A positive coefficient indicates that as the value of the independent variable increases, the mean of the dependent variable also tends to increase. A negative coefficient suggests that as the independent variable increases, the dependent variable tends to decrease.

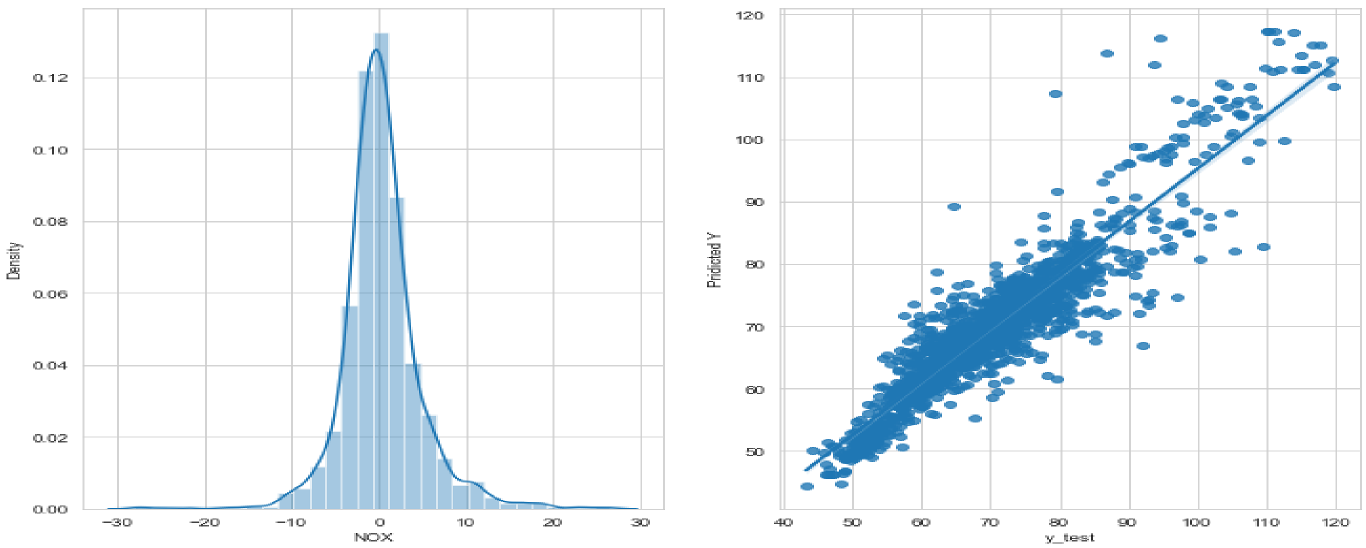
Approach 2: Random Forest Regressor

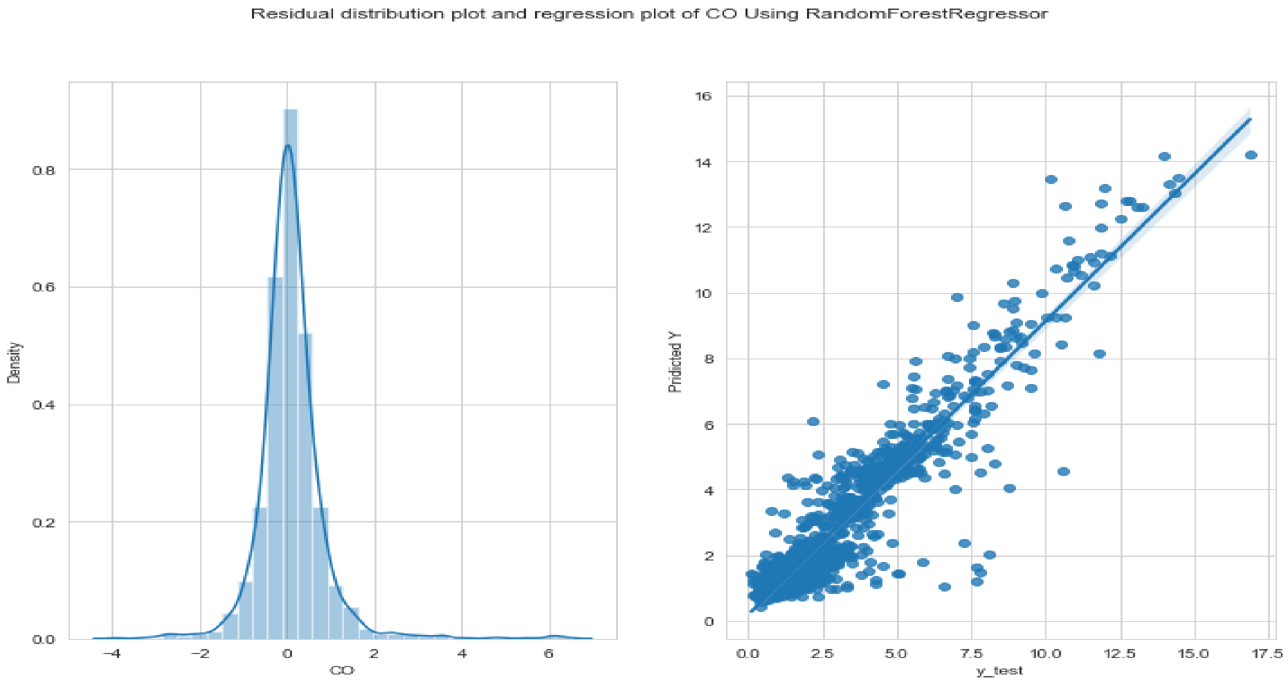
When Random Forest Regressor model give more accuracy as compared to Linear Regression As we can see they are much lower this time. It fits better than our baseline. The gap between the two lines has reduced!!

True Value Vs Predicted value of CO and NOX Using RandomForestRegressor

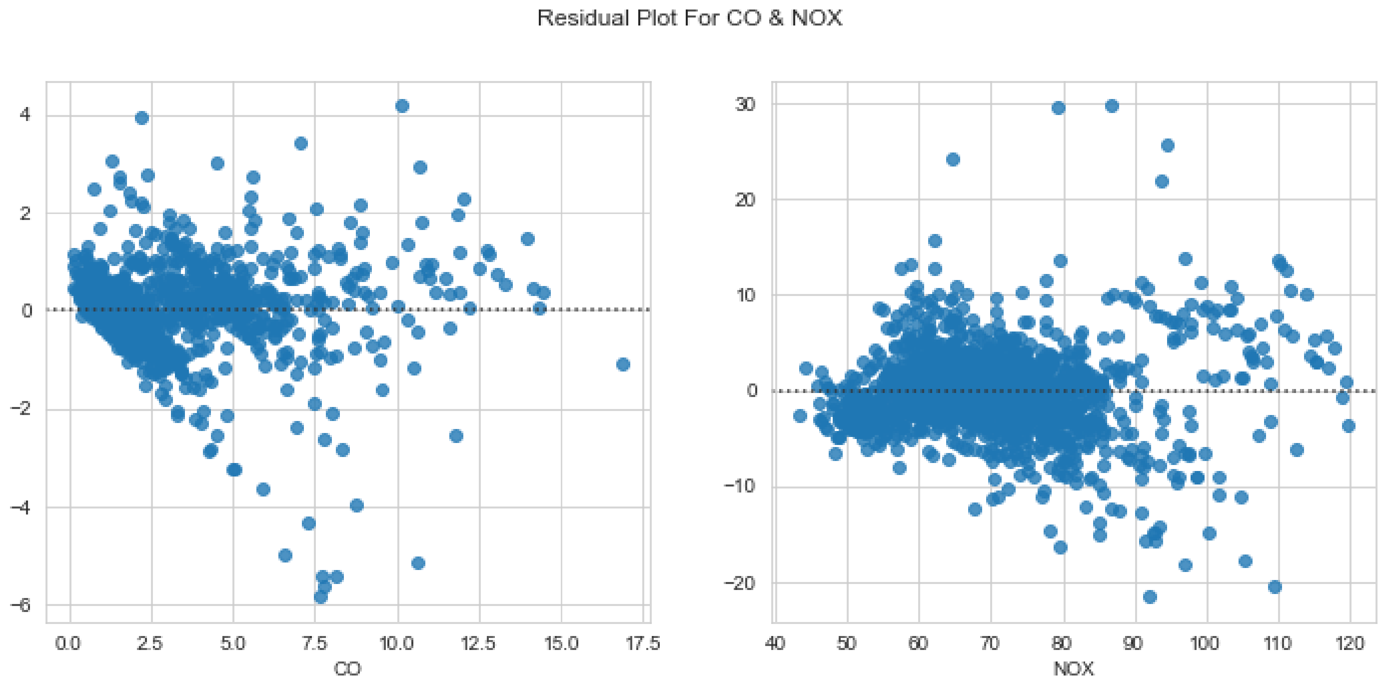


Residual distribution plot and regression plot of NOX Using RandomForestRegressor





Left plot show follows Normal Distribution and they are symmetrically distributed. The model for the chart on the left is very accurate, there is a strong correlation between the model's predictions and its actual results.

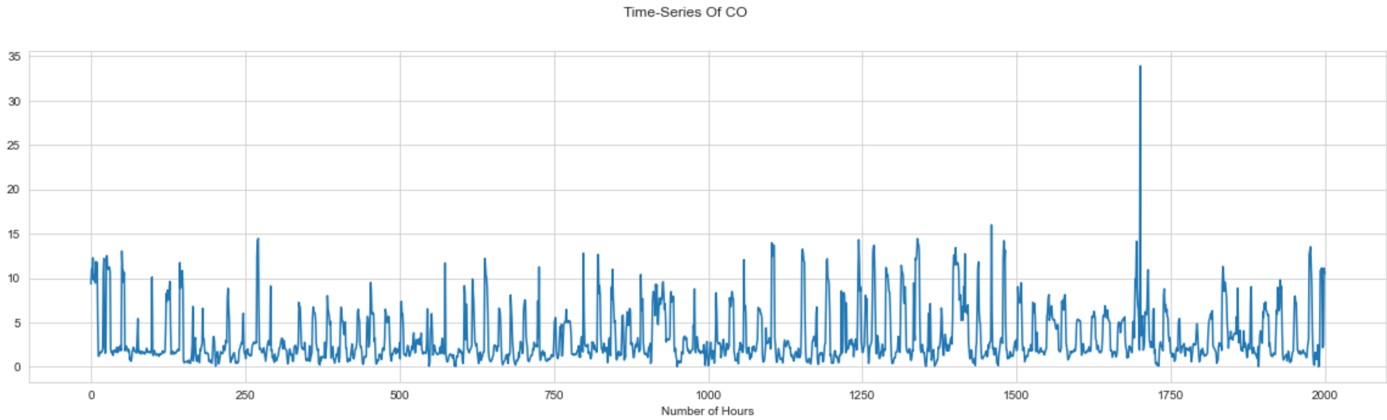


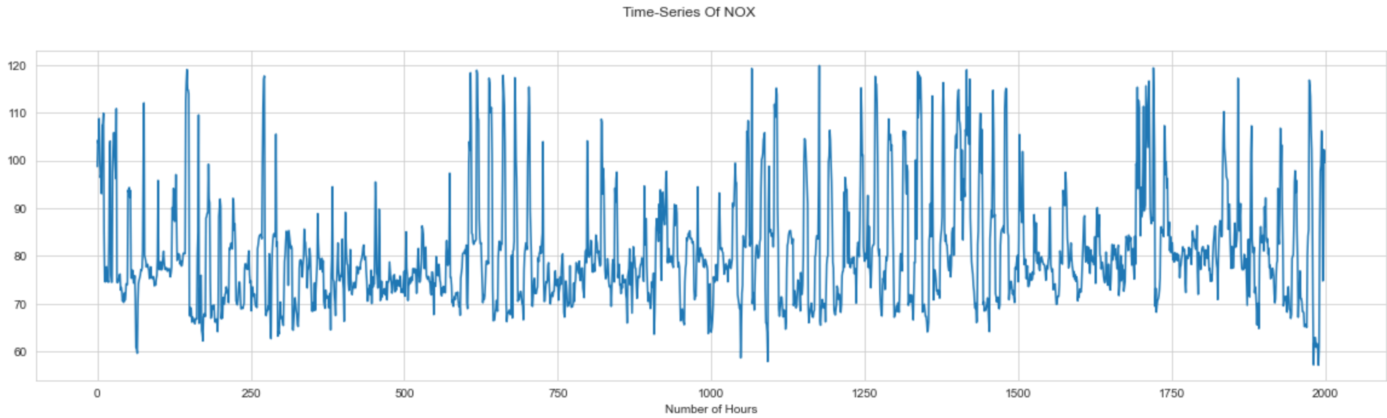
In the Plot, we see that residuals tend to concentrate around the x-axis, in this residual plot, the points are scattered randomly around the residual=0 line. We can conclude that a linear model is appropriate for modelling this data.

For time-series trending of the parameters

Approach :1

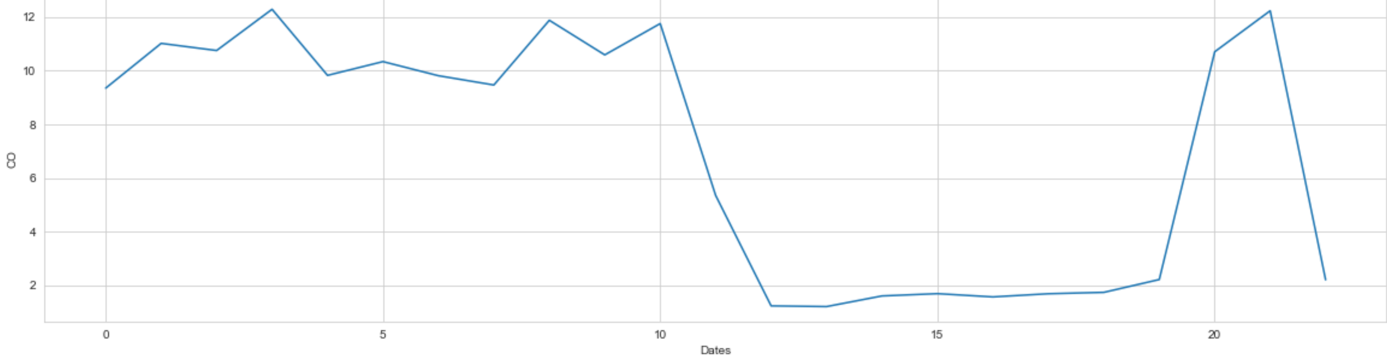
We know that in this datasets sensor measures aggregated over one hour (by means of average or sum) from a gas turbine & this data collected from range 01.01.2011 - 31.12.2015. so, I decided add another feature "Number of Hours " that is counts hours of each row.

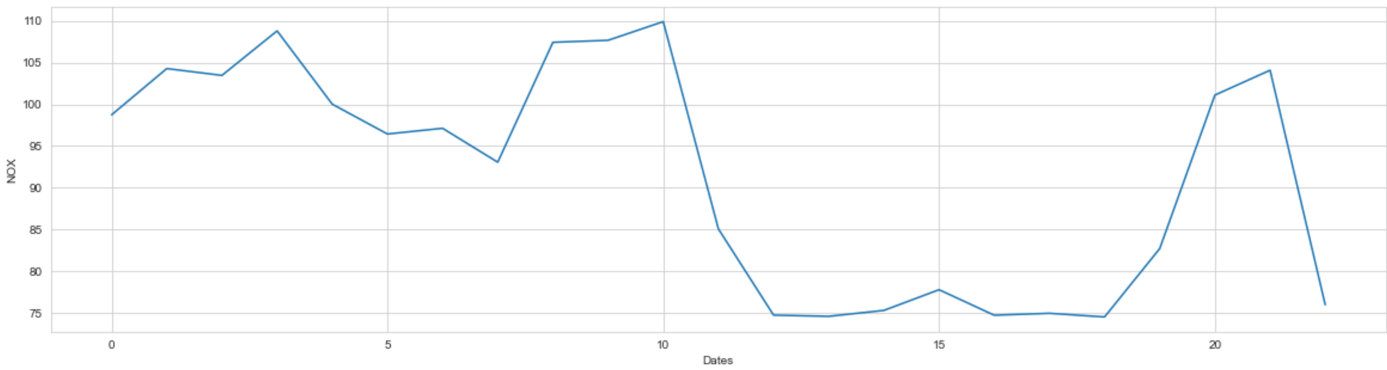




Approach :2

In our dataset we have 7153 row it is not fitting that range year that is way I considered 24 rows as 1 day then using pandas "groupby" function I plotted the Time- Series graph.





Above graph, Plot for 1 days that measures 24 hours values i.e., sensor measures aggregated over one hour So we can see that data of CO and NOX decrease around the range 10-20.

Conclusion

So Linear Regression will not give better Predication with maximum error rate between true value and Predicted values. that 'way I am try another model for better predication!!

```
Test Accuracy for CO:  0.7296689616158056
Test Accuracy for NOX:  0.43987463660957715
Mean Squared Error for CO:  1.3597625609196047
Mean Squared Error for NOX:  79.78780102497436
```

The AT, AH, TIT and CDP are most effecting most parameters which is causing the emissions for CO to increase dynamically similarly TIT and CDP are affect the emission for NOX.

Now use this feature on ensemble-based Random Forest Regressor for improve our model score!!!

The accuracy of our model is 88% for CO and 86% for NOX. its MSE has reduced to 0.5858 for CO and 19.91 for NOX so we can say that it performs better!!!

```
Test Accuracy for CO:  0.8816123330495216
Test Accuracy for NOX:  0.8557011496606973
Mean Squared Error for CO:  0.5954888427021697
Mean Squared Error for NOX:  20.55484131144362
```