



AIR POLLUTION

Team No: 10

20WH1A12A9-S.Niharika

20WH1A12B0-G.Krishna Prathibha

20WH1A12B3-CH.Dharani

20WH1A12B4-B.Sowmya

20WH1A12B5-G.Sneha

Problem Statement



- We are predicting the values of air pollution measurements over time, based on basic weather information like temperature, humidity and the input values of 5 sensors.
- The target values we are predicting are: Carbon Monoxide, Benzene, Nitrogen Oxide.



Python Packages Used

- Numpy:- Used for working with arrays, matrices, and numerical computations in general.
- Pandas:- It is open-source data Analysis and data manipulation for python.
- Scikit learn:- Used for to implement machine learning models and statistical modelling
- Matplotlib:- Used for data plotting.

Types of Algorithms used



- Randomforest Regression Algorithm
- Linear Regression Algorithm
- XGBoost

Random Forest Regression



- Random Forest Regression Algorithm:-
 - Random forest regression is a machine learning algorithm that uses a group of decision trees to predict a numerical value. It takes the average to improve the predictive accuracy of the dataset.

Linear Regression



- Linear Regression Algorithm:-
 - It is a popular algorithm used in statistics to model the relationship between a dependent variable and one or more independent variables. The goal of linear regression is to find the minimized sum of the squared errors between the predicted values and the actual values.

XGBoost




- XGBoost(Extreme Gradient Boosting):-
 - The XGBoost algorithm is an ensemble learning method that uses a combination of weak models (typically decision trees) to create a strong model that can make accurate predictions. Efficient for large datasets. XGBoost can also handle missing values.

Output



submission.csv ✕

...

1 to 10 of 2247 entries Filter 

date_time	target_carbon_monoxide	target_benzene	target_nitrogen_oxides
2011-01-01 00:00:00	1.4080000000000021	4.4250000000000006	206.08900000000003
2011-01-01 01:00:00	2.0399999999999996	7.1869999999999999	308.62100000000001
2011-01-01 02:00:00	1.8909999999999993	7.64	326.14900000000002
2011-01-01 03:00:00	2.046	8.643999999999998	341.30500000000006
2011-01-01 04:00:00	1.4179999999999995	7.2210000000000001	276.529
2011-01-01 05:00:00	1.1530000000000005	4.8390000000000002	209.09099999999995
2011-01-01 06:00:00	1.3590000000000004	4.796	249.65800000000002
2011-01-01 07:00:00	1.2940000000000007	4.7899999999999999	240.86699999999985
2011-01-01 08:00:00	0.8080000000000002	3.4350000000000002	165.51299999999998
2011-01-01 09:00:00	0.8900000000000001	3.2049999999999983	152.409

Show per page

1 2 10 100 200 220 225



deg_C	CarbonMonoxide
<input type="text" value="13.1"/>	<input type="text" value="2.719"/>
relative_humidity	Benzene
<input type="text" value="46"/>	<input type="text" value="12.852"/>
absolute_humidity	NitrogenOxide
<input type="text" value="0.7578"/>	<input type="text" value="279.3210000000002"/>
sensor_1	<input type="button" value="Flag"/>
<input type="text" value="1387.2"/>	
sensor_2	
<input type="text" value="1087.8"/>	
sensor_3	
<input type="text" value="1056"/>	
sensor_4	

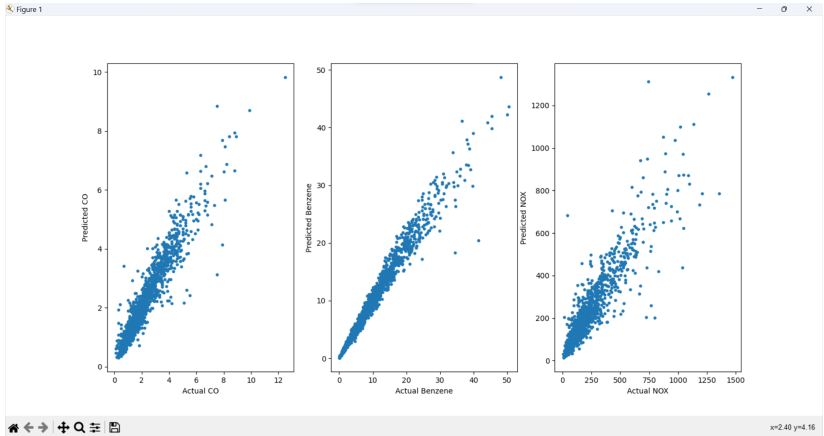


Air Pollution Measurements Prediction

deg_C	<input type="text" value="13.1"/>
relative_humidity	<input type="text" value="46"/>
absolute_humidity	<input type="text" value="0.7578"/>
sensor_1	<input type="text" value="1387.2"/>
sensor_2	<input type="text" value="1087.8"/>
sensor_3	<input type="text" value="1056"/>
sensor_4	<input type="text" value="1742.8"/>
sensor_5	<input type="text" value="1293.4"/>

Predicted Values: CO:2.97, Benzene: 14.11, NOX:275.65

Graph



Comparision Table



Random Forest

RMSLE :- 0.187

XGBoosting

RMSLE :- 0.336

Linear Regression

RMSLE :- 0.398



THANK YOU