Dec 12, 2024

CO and NOx Emission Prediction Based on Gas Turbine

Ma Tieding 121090407

Chen zeyi 121090077

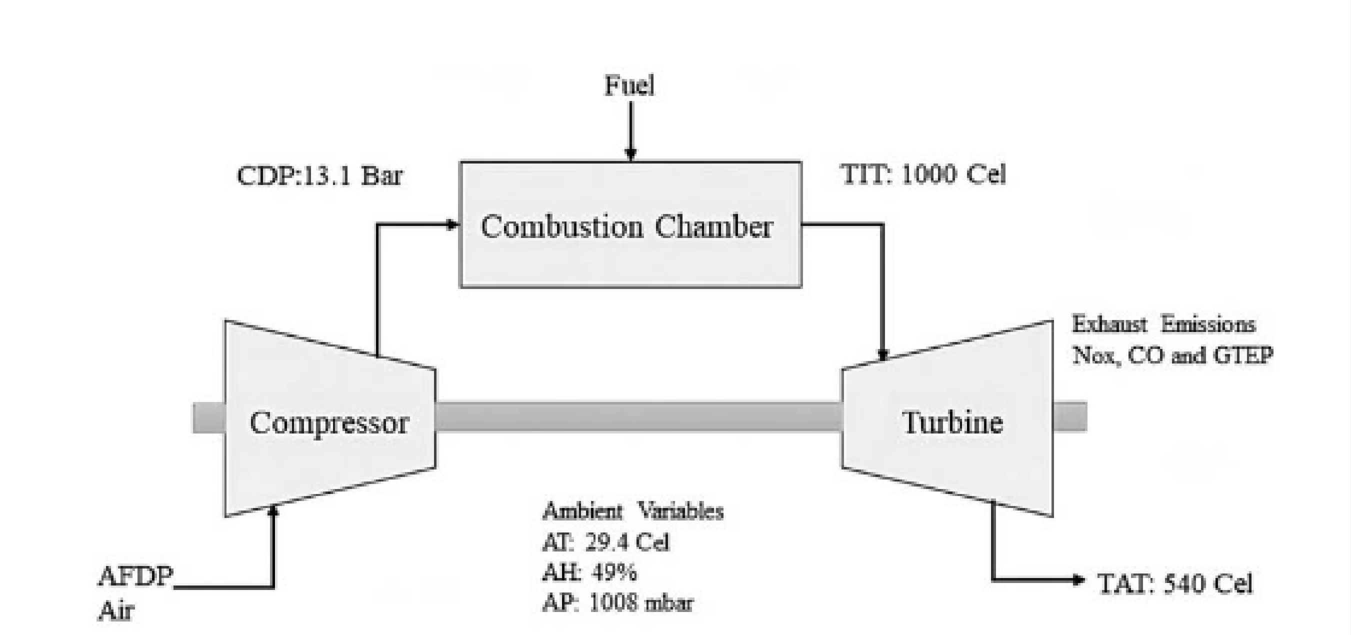
Dec 12, 2024

Today's Agenda

1. Introduction to the Research
2. Data Describe and Method Discussion
3. Result and Analysis
4. Conclusion

Looking Back....

CO and NOx Emission Prediction Based on Gas Turbine



Question 1

Why Focus on Gas Turbines (GT) ?

Question 2

Why Focus on CO and NOx ?

Question3

How to monitor CO and NOX now?

Question 4

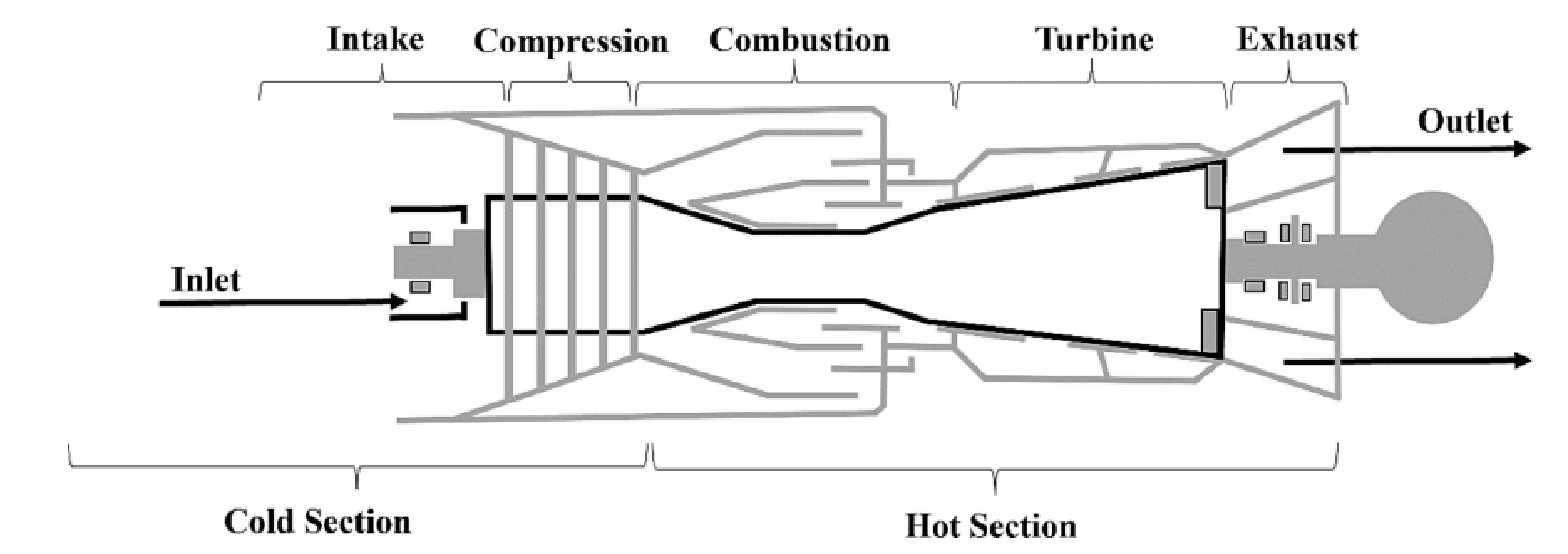
What have I done with this project?

Introduction

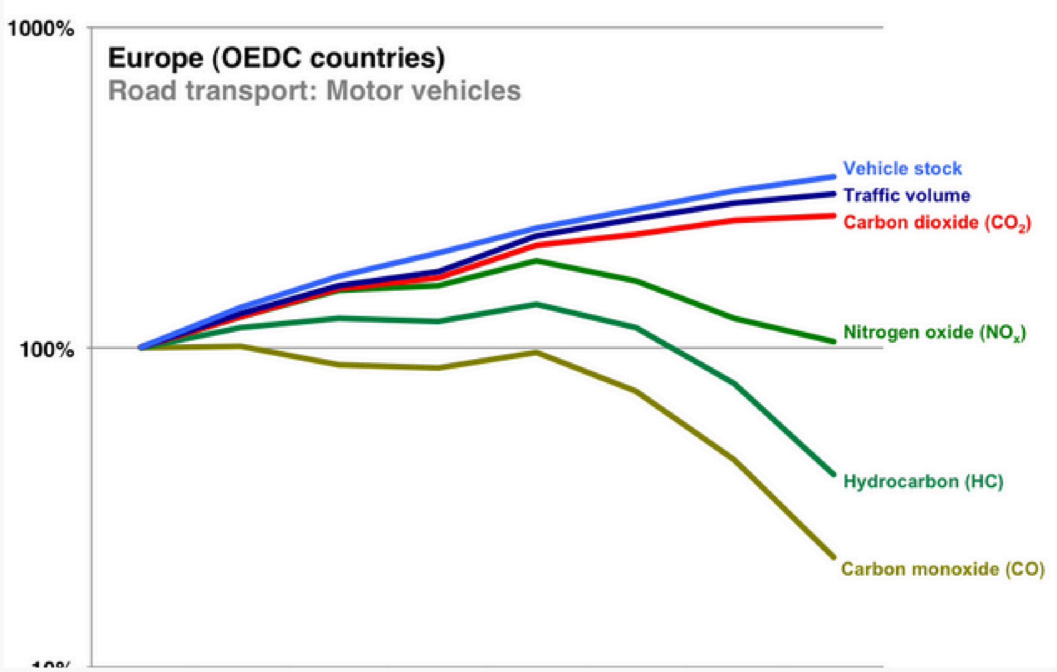
Dec 12, 2024

Why Focus on Gas Turbines (GT) ?

- Addressing Climate Change : Gas turbines have significant potential in reducing greenhouse gas emissions, supporting countries in achieving low-energy economy goals [1]
- Regulatory Pressure: As countries implement stringent environmental regulations, gas turbines are seen as a key technology for reducing harmful emissions. [2]



Why Focus on CO and NOx ?



- Environmental and Health Impact: CO and NOx are major air pollutants that can lead to acid rain and have negative effects on human health. [2] [3]
- Regulatory Compliance: Many countries have strict legal requirements for CO and NOx emissions, and companies must adhere to these regulations to avoid fines and maintain public image [3]
- Sustainable Development Goals: Reducing CO and NOx emissions is an important step toward achieving zero pollution, decarbonization, and circular economy objectives [4]

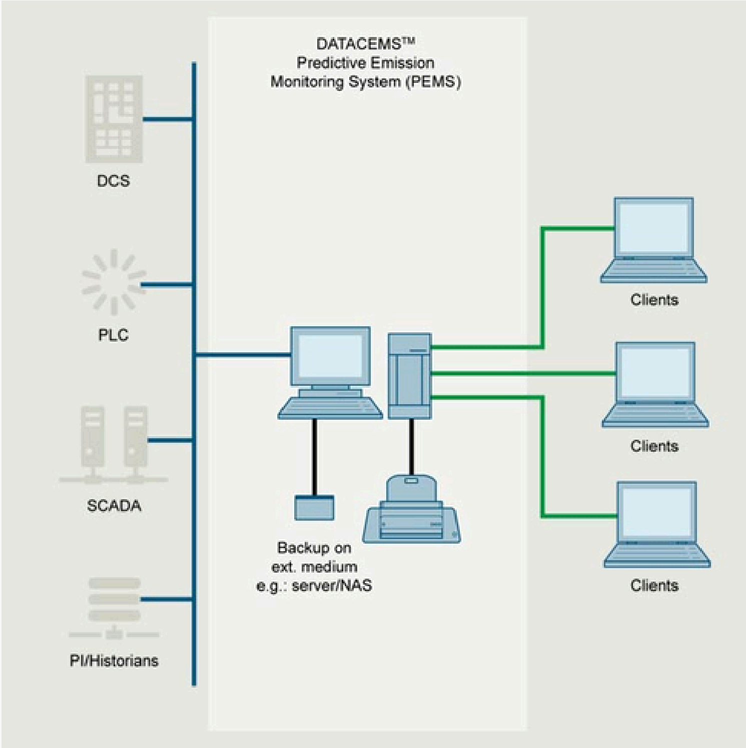
Introduction

Dec 12, 2024

How to monitor CO and NOX now ? [6][7][8]

Continuous emission monitoring systems (CEMS)

- Needs significant capital to be invested
- Maintenance due to weather conditions and interference of other gases during measurements



Application of predictive emission monitoring systems (PEMS)

- Computa-tional fluid dynamic modeling (CFD)
- Chemical network reactor
- Machine learning

- Needs personnel with technical skills in the technology area [5]
- The computational complexity and high computational time make the practical application of CFD models Fuel challenging

Introduction

Dec 12, 2024

What have I done with this project?

Authors (year)	Model/Technique	Study/ Injection	Gas Emissions
Casella and Maffezzoni (2003)	ANN, Semi-empirical	Modeling	CO, NO_x
Lipperheide et al. (2017)	Surrogate model	Modeling	NO_x
Kaya et al. (2019)	ANN, RF	Modeling	CO, NO_x
Kochueva and Nikolskii (2021)	Fuzzy system	Modeling	CO, NO_x
Rezazadeh (2021)	PCA, *k*NN	Modeling air	NO_x
Huang et al. (2022)	ANN	Modeling	NO_x

Some papers on the emissions in GT and diesel engines are summarized in Table 1, in ascending order of year of publication.[9][10][11][12]

- The model is complex
- Requires more computing power

The objective of the present project is the construction of a novel modeling pipeline and evaluation of the fusion of feature engineering methods with machine learning models applied to the prediction of CO and NOx emissions of a GT plant.

Dec 12, 2024

Today's Agenda

1. Introduction to the Research
2. Data Describe and Method Discussion
3. Result and Analysis
4. Conclusion

Data Describe

Dec 12, 2024

Measured values from the GT power plant.

Variable	Symbol	Unit	Min	Max	Mean	Std
Ambient temperature	AT	°C	-6.23	37.10	17.71	7.45
Ambient pressure	AP	mbar	985.85	1036.56	1013.07	6.46
Ambient humidity	AH	%	24.08	100.20	77.87	14.46
Air filter difference pressure	AFDP	mbar	2.09	7.61	3.93	0.77
Gas turbine exhaust pressure	GTEP	mbar	17.70	40.72	25.56	4.20
Turbine inlet temperature	TIT	°C	1000.85	1100.89	1081.43	17.54
Turbine after temperature	TAT	°C	511.04	550.61	546.16	6.84
Compressor discharge pressure	CDP	mbar	9.85	15.16	12.06	1.09
Turbine energy yield	TEY	MWH	100.02	179.50	133.51	15.62
Carbon oxides	CO	mg/m^3	0	44.10	2.37	2.26
Nitrogen oxides	NO_X	mg/m^3	25.90	119.91	65.29	11.68

Note: Max: Maximum; Min: Minimum; Std: Standard Deviation.

1. **Data Source** : This Project used a dataset from a GT located in Turkey's Northwestern region is used for evaluating flue gas emissions. The dataset contains samples that indicate the level of CO and NOx, which were used for building regression models.

2. **Data review** : The dataset contains 36,733 samples of eleven sensor measurements structured over one hour, through average or sum. The data comes from the GT power plant and are collected in a data range from 01/01/2011 to 12/31/2015, i.e., five-year or 1,825 days, including GT parameters, such as turbine inlet temperature and compressor discharge pressure, as well as ambient variables

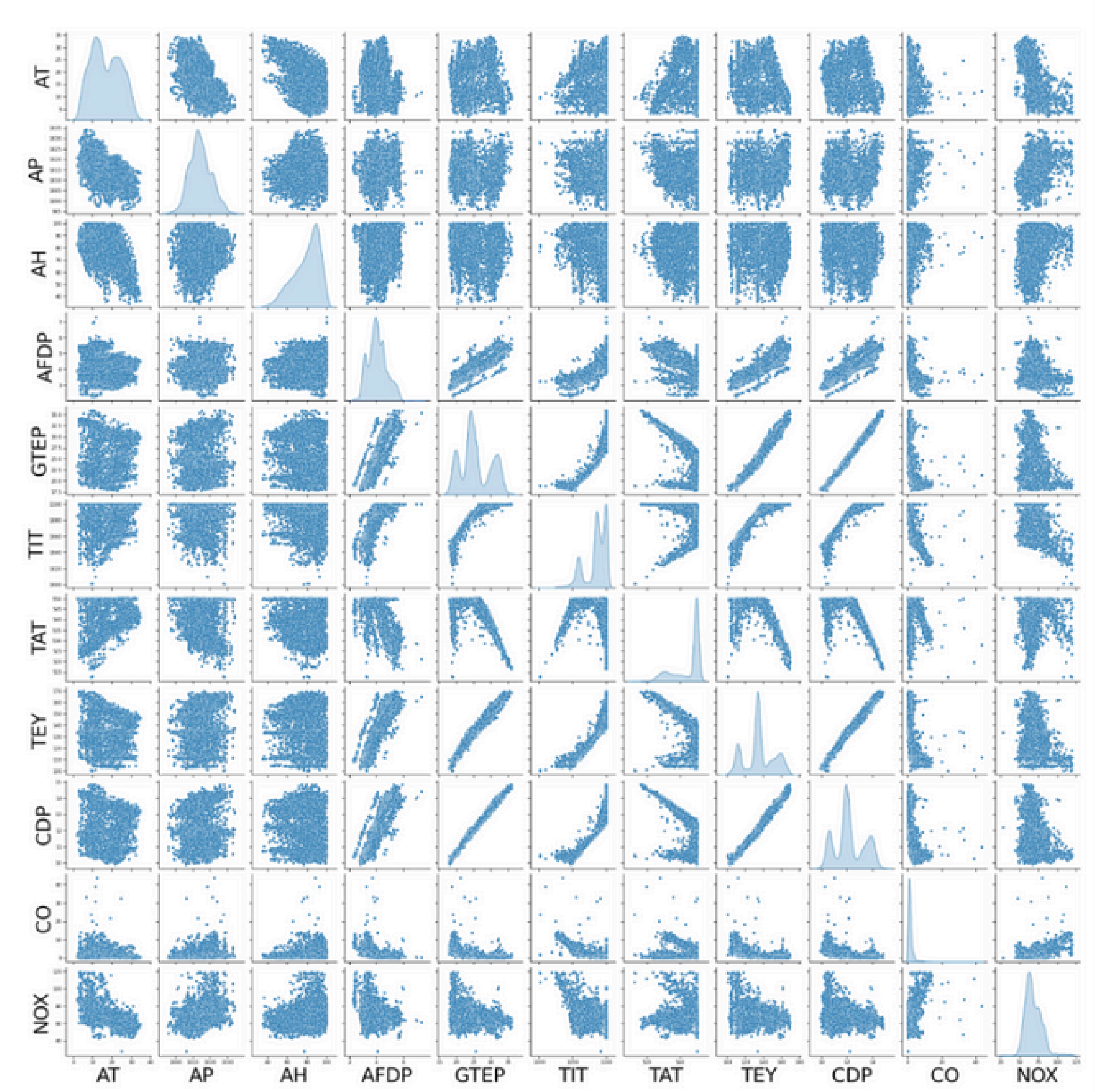
Data Describe

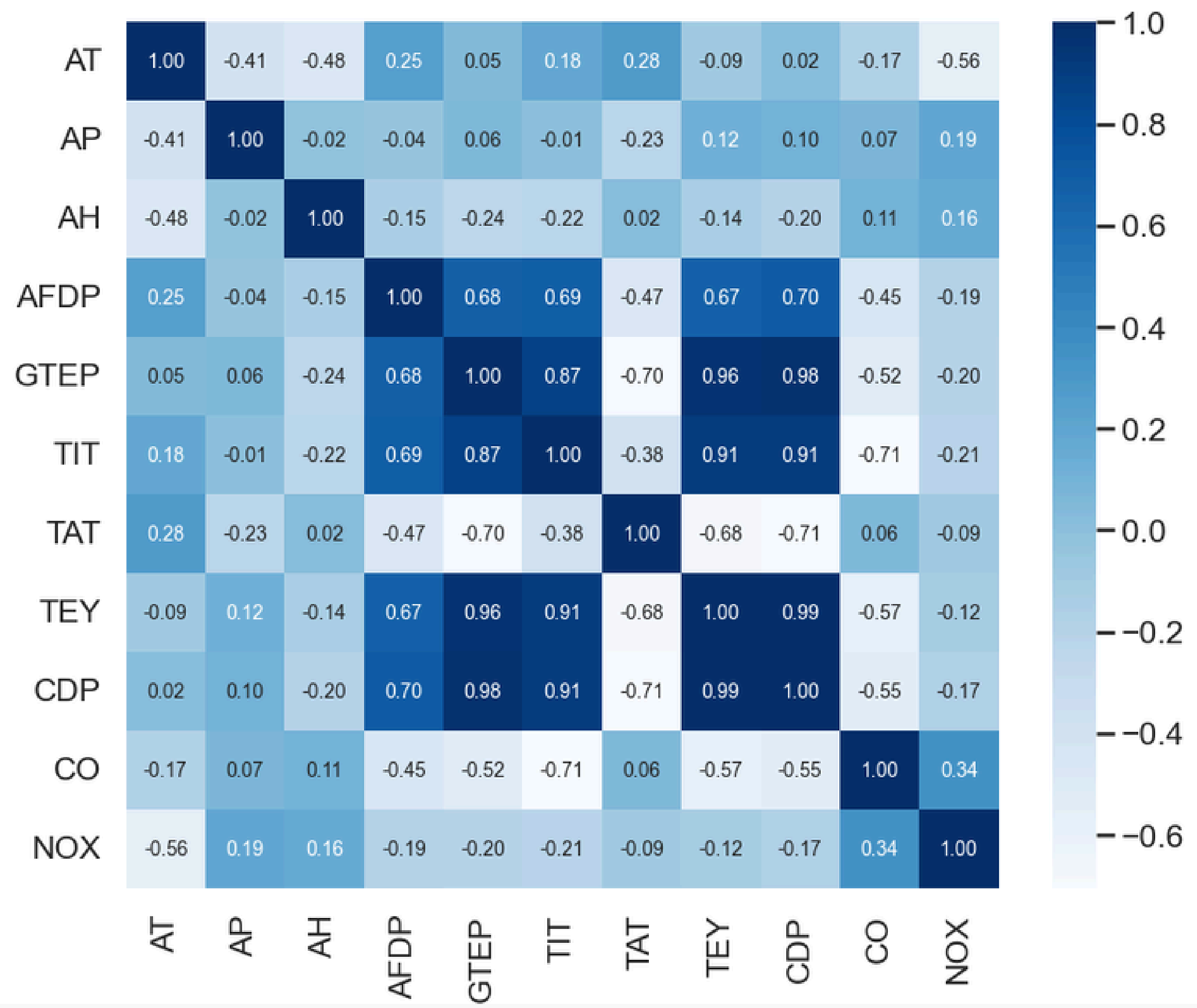
Dec 12, 2024

1 Pair Plot

2 Heatmap

3 Pairwise comparisons between different variables in a dataset are commonly adopted for exploratory data analysis. Pairwise plots and kernel density estimates of the input and output variables from a gas turbine are given in Fig.





Method- Evaluation and Data Processing

Higher Criteria A

Mean Absolute Error (MAE)

$$MAE=\sum 1/n \, | y_i - \hat{y}_i |$$

- Measures average error in predictions.
- Good for understanding typical prediction error.

Root Mean Squared Error (RMSE)

$$RMSE=\sum 1/n(y_i-\hat{y}_i)^2$$

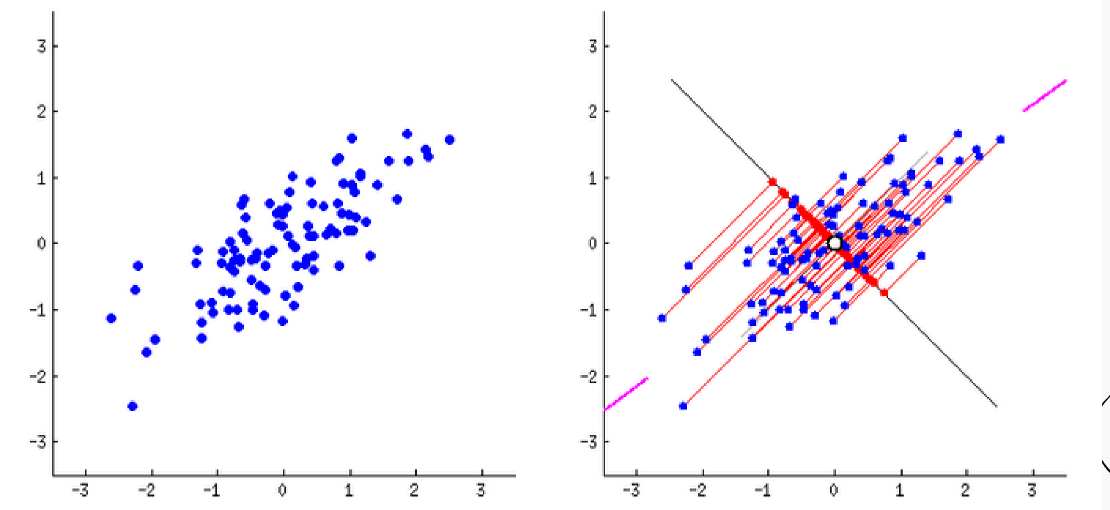
- Sensitive to large errors due to squaring.
- Useful for comparing model accuracy.

Lower Criteria B

Higher Criteria B

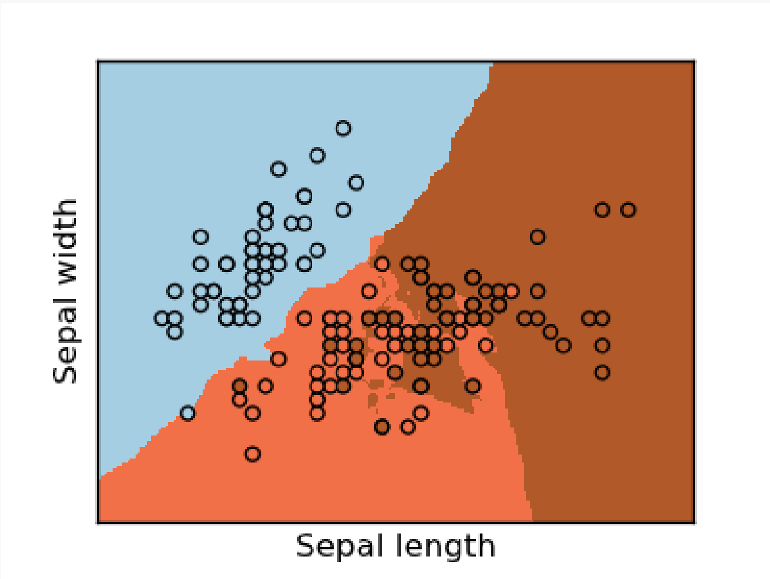
Principal Component Analysis (PCA)

$$C=1/(n-1)X^TX$$



K-Nearest Neighbors (KNN)

$$d(x,x_i)=\text{sqrrt}(\sum(x_j-x_{ij})^2)$$



Lower Criteria A

1. Mean Absolute Error (MAE)
2. Root Mean Squared Error (RMSE)
3. Principal Component Analysis (PCA)
4. K-Nearest Neighbors (KNN)

Method-Models

Linear Regression Class

Linear Regression

$y=\beta 0+\beta 1x1+\beta 2x2+\ldots+\beta nxn+\epsilon$

- Simple and interpretable.
- Sensitive to outliers

Support Vector Regression

Support Vector Regression

$y=\sum N(\alpha i\cdot K(xi,x)+b)$

- Uses kernel functions to handle non-linear relationships.
- Robust to outliers but can be sensitive to parameter settings.

Tree-Based Models Class

Decision Tree Regression

- Models data using a tree structure.
- Easy to interpret but prone to overfitting.

Random Forest Regression

$y=1/T\sum yt$

- T is the number of trees,
- yt is the prediction value of the t-th tree.

XGBoost Regression

$y=\sum fk(x)$

- Gradient boosting framework optimized for speed and performance

LightGBM Regression

$y=\sum fk(x)$

- Handles large datasets well and supports parallel learning

Extra Trees Regression

$y=1/T\sum yt$

- Similar to Random Forest but uses a more random approach for tree splitting.

Neural Network Class

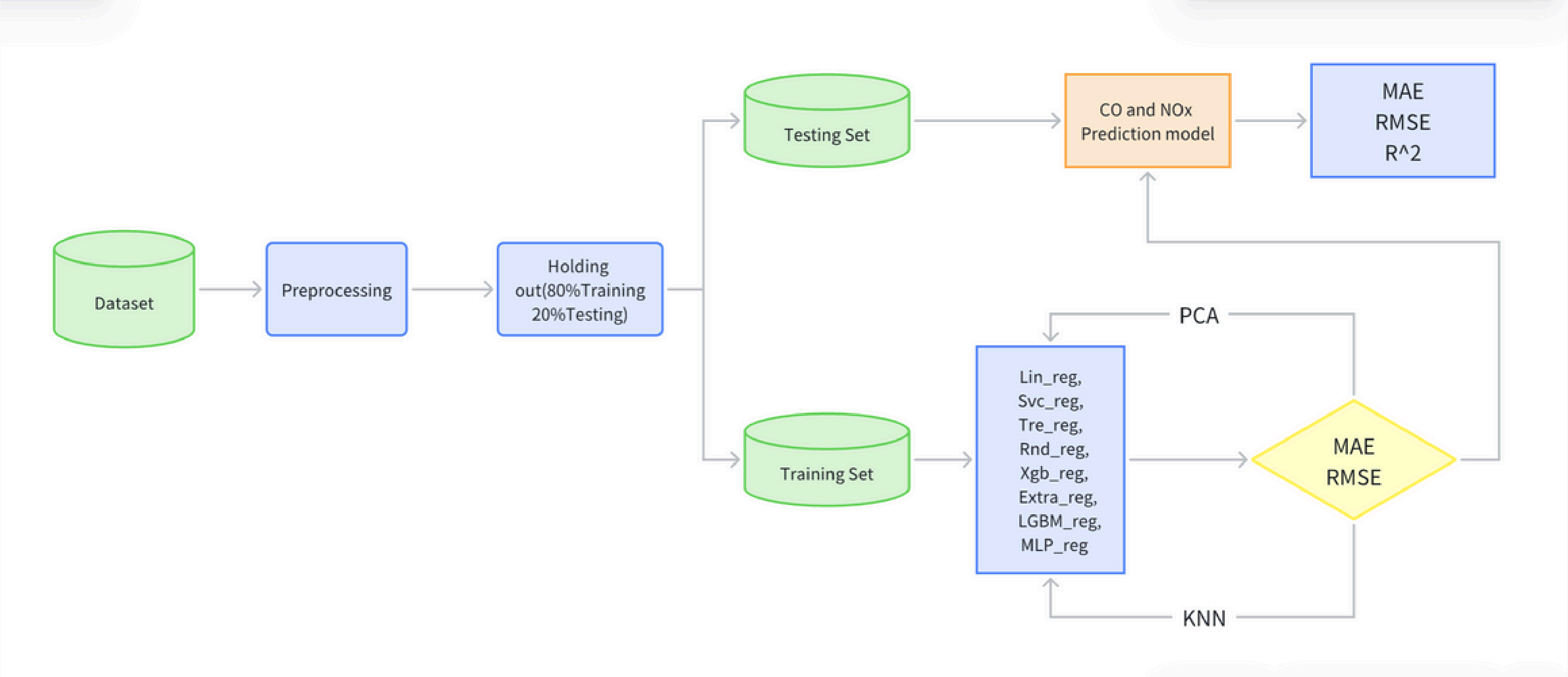
MLP Regression

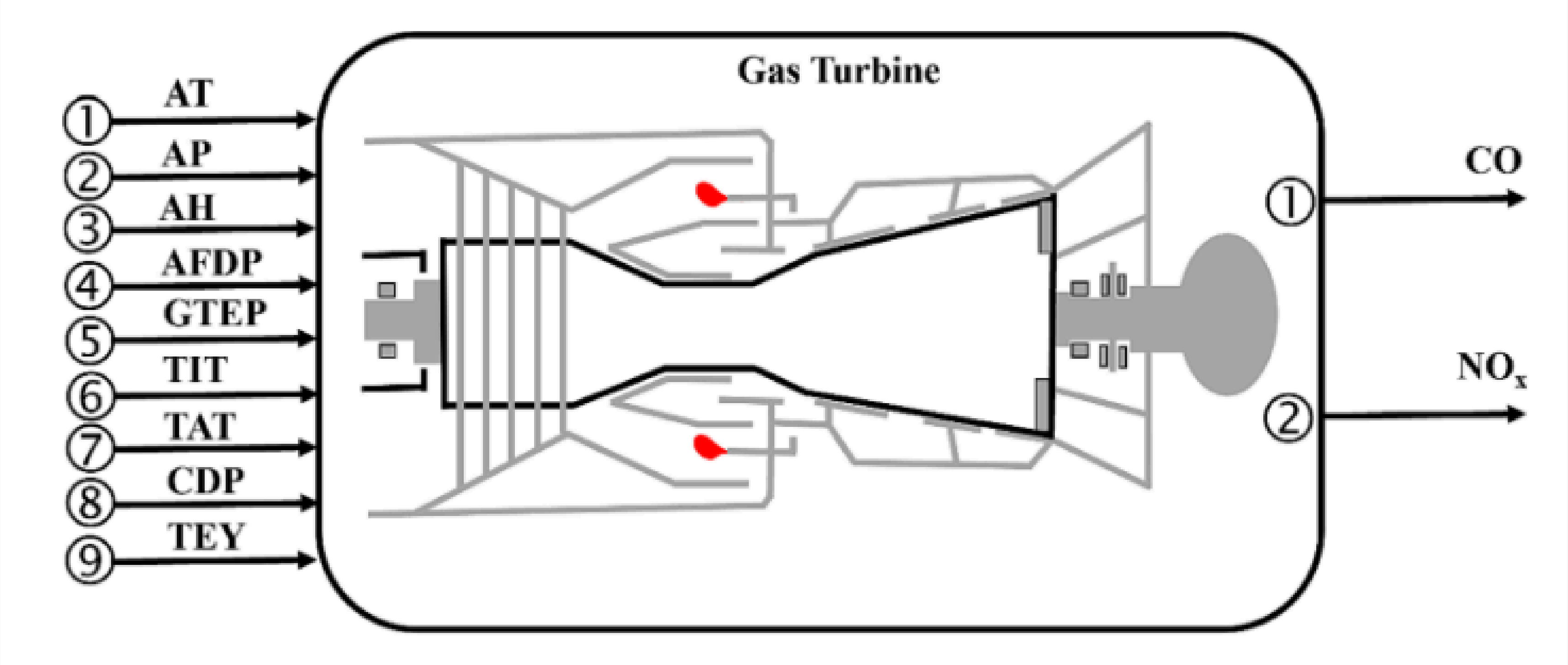
$y=f(W\cdot x+b)$

- A type of neural network with one or more hidden layers.
- Capable of modeling complex non-linear relationships.

Process Overview

These pictures present the workflow for developing a CO and NOx prediction model using a dataset. It includes data preprocessing, splitting into training and testing sets, model evaluation with various regression techniques, and dimensionality reduction with PCA, alongside KNN analysis.





Dec 12, 2024

Today's Agenda

1. Introduction to the Research
2. Data Describe and Method Discussion
3. Result and Analysis
4. Conclusion

Dec 12, 2024

Result Using Basic Data

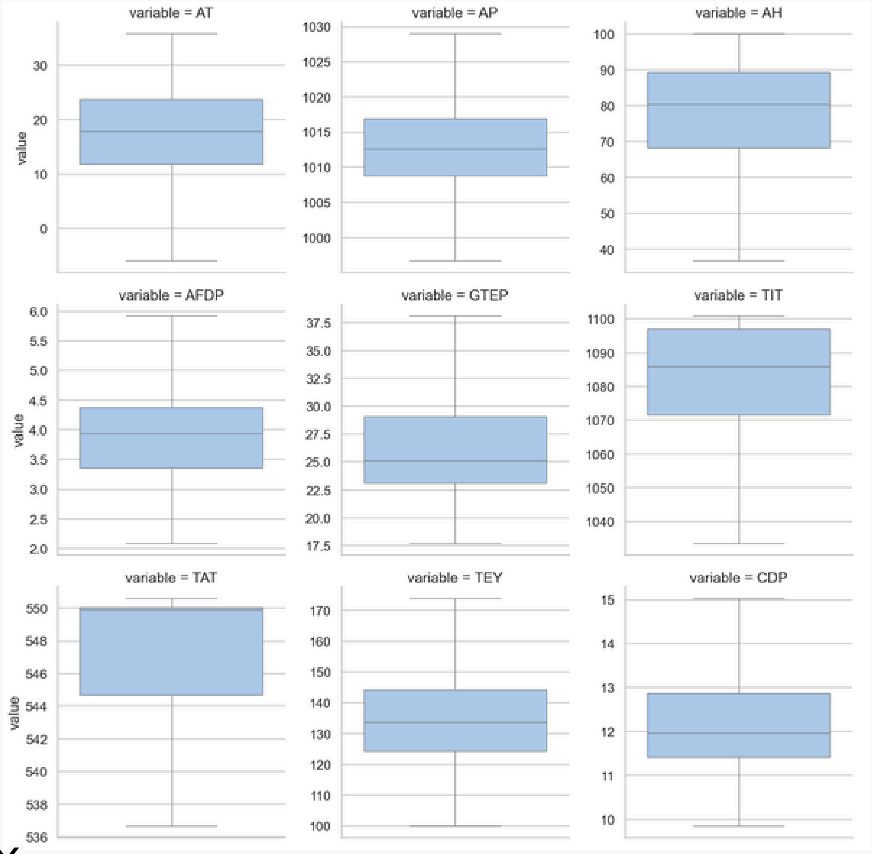
1.Code

```
for feature in features:
    for reg in (
        lin_reg,
        svc_reg,
        tre_reg,
        rnd_reg,
        xgb_reg,
        extra_reg,
        LGBM_reg,
        MLP_reg
        ):
        scores = cross_validate(reg, X_train_std[idx], y_train[feature].iloc[idx], scoring=scorers, n_jobs=-1)
        metrics2[reg.__class__.__name__] = {
            **metrics2[reg.__class__.__name__],
            **{
                f'{feature}_rmse': np.sqrt(-scores['test_neg_mean_squared_error']).mean(),
                f'{feature}_rmse_std': np.sqrt(-scores['test_neg_mean_squared_error']).std(),
                f'{feature}_mae': -scores['test_neg_mean_absolute_error'].mean(),
                f'{feature}_mae_std': scores['test_neg_mean_absolute_error'].std()
            }}
```

3.Data processing

```
def calculate_iqr_boundaries(series):
    q25 = series.quantile(0.25)
    q75 = series.quantile(0.75)
    iqr = q75 - q25
    boundaries = (q25 - 1.5 * iqr, q75 + 1.5 * iqr)
    return boundaries
```

In the first dataset we only did the basic filling of vacancies and so on, and in the second dataset we removed the special quartile values



2. Metrics Matrix(Basic)

	CO_rmse	CO_rmse_std	CO_mae	CO_mae_std	NOX_rmse	NOX_rmse_std	NOX_mae	NOX_mae_std
LinearRegression	1.475	0.140	0.850	0.020	8.094	0.132	5.768	0.060
SVR	1.242	0.164	0.572	0.019	5.771	0.119	3.816	0.059
DecisionTreeRegressor	1.526	0.104	0.670	0.010	6.211	0.118	3.794	0.063
RandomForestRegressor	1.093	0.143	0.488	0.011	4.290	0.107	2.724	0.053
XGBRegressor	1.146	0.149	0.539	0.013	4.451	0.100	2.989	0.059
ExtraTreesRegressor	1.067	0.161	0.476	0.015	4.000	0.090	2.547	0.058
LGBMRegressor	1.136	0.156	0.553	0.014	4.772	0.089	3.234	0.039
MLPRegressor	1.151	0.157	0.578	0.017	4.706	0.099	3.218	0.047

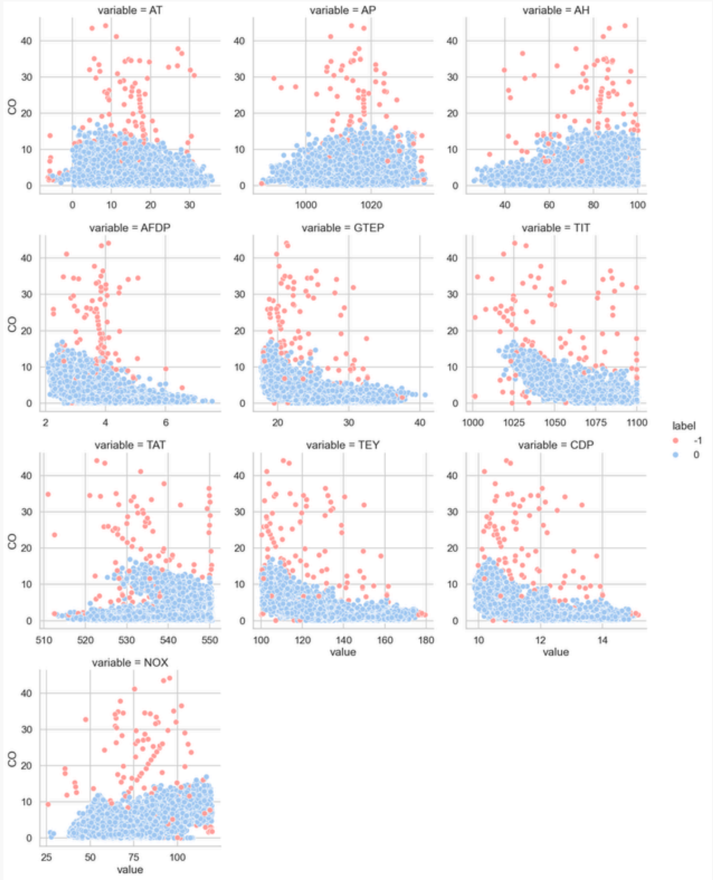
4.Metrics Comparison Matrix

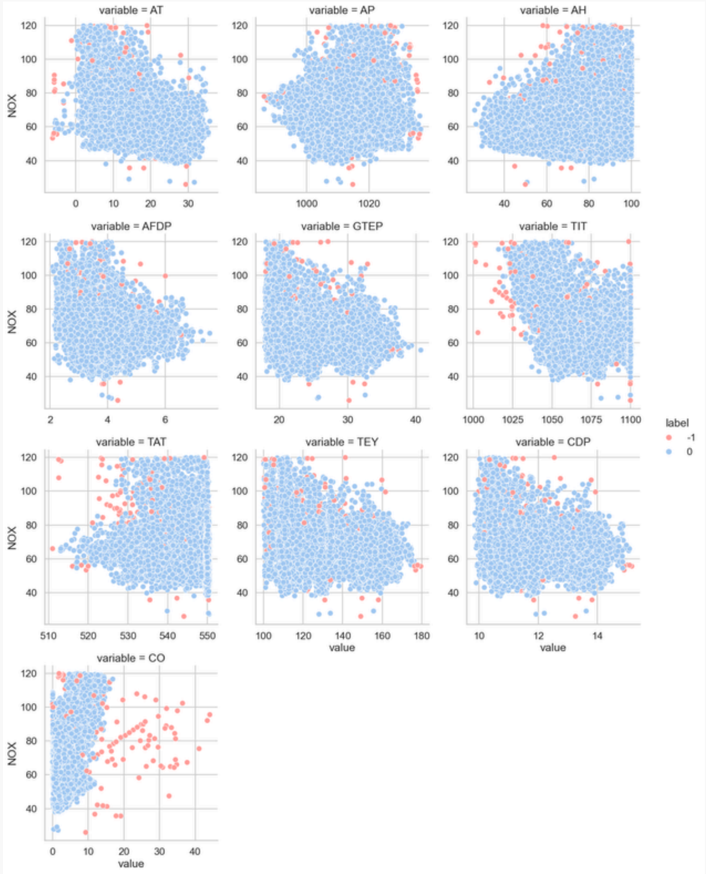
CO_rmse	CO_rmse_std	CO_mae	CO_mae_std	NOX_rmse	NOX_rmse_std	NOX_mae	NOX_mae_std
-1.38	-3.65	-2.43	-9.92	-6.87	-3.89	-5.36	19.76
0.74	1.14	1.01	2.70	0.19	4.81	1.32	15.78
1.19	-1.22	1.35	-7.74	0.51	6.01	0.97	-16.49
1.17	-10.51	0.59	2.81	0.88	-1.60	1.17	-3.20
0.23	-17.72	0.85	-38.14	1.57	-1.33	1.63	-11.80
0.98	-4.30	0.70	-2.86	1.08	12.98	1.27	-5.67
0.37	-12.48	0.51	-17.59	0.98	12.09	1.29	30.70
0.76	-12.52	0.93	-43.52	3.42	8.63	3.60	91.23

Dec 12, 2024

Result Using KNN

1.Knn Scatter Plot





```
db = DBSCAN(eps=1.9, min_samples=40, n_jobs=-1)
db.fit(df_train_std)

unique, counts = np.unique(db.labels_, return_counts=True)
np.c_[unique, counts]
```

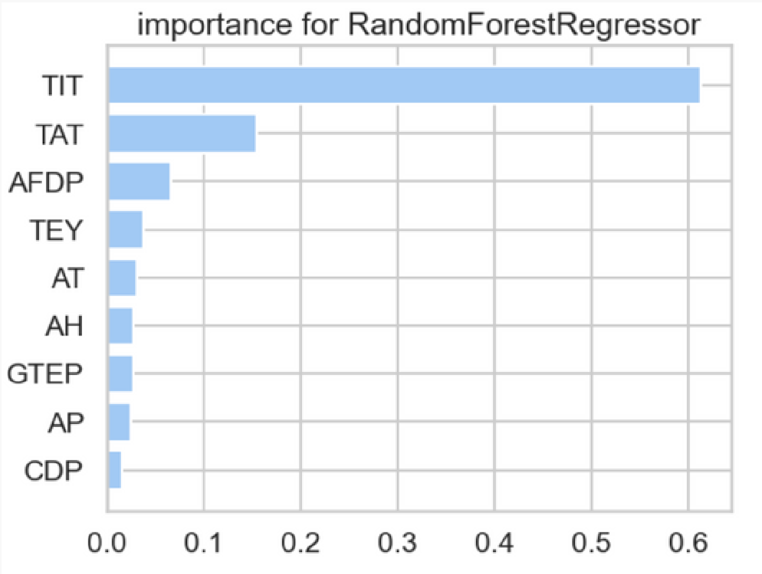
Marginal points with poor clustering values were excluded by knn.

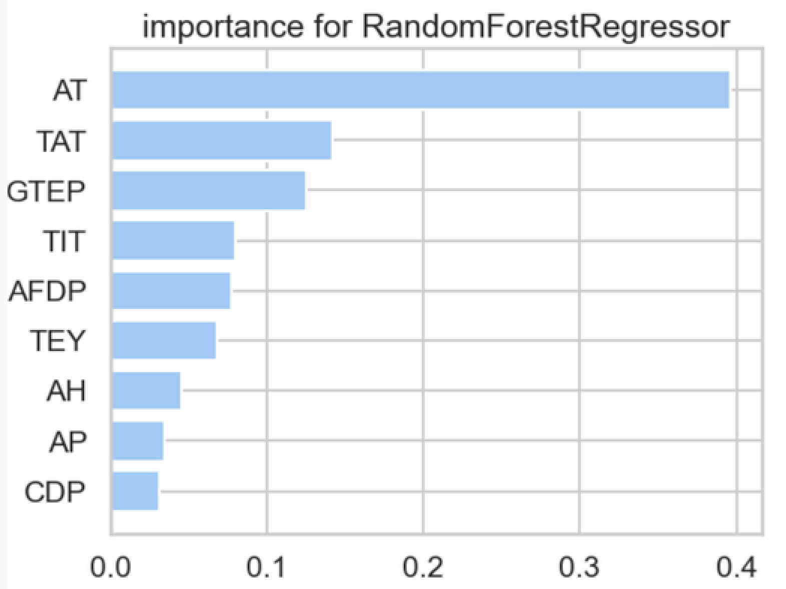
The index comparison matrix is obtained, and the feature important values are viewed according to the random forest

2.Metrics Comparison Matrix

CO_rmse	CO_rmse_std	CO_mae	CO_mae_std	NOX_rmse	NOX_rmse_std	NOX_mae	NOX_mae_std
-27.11	-86.04	-11.02	-59.54	-7.05	-38.42	-7.12	-10.16
-36.11	-87.86	-10.43	-62.11	-7.32	-18.25	-4.56	-9.90
-37.47	-74.29	-12.35	21.33	-7.00	14.89	-4.09	-7.20
-38.74	-80.02	-12.79	-24.61	-7.49	-10.22	-4.28	-12.03
-38.57	-85.32	-13.95	-37.49	-7.00	-9.15	-4.42	-13.01
-39.34	-83.86	-13.12	-43.71	-7.20	-7.24	-3.92	-14.10
-37.09	-83.65	-14.25	-31.59	-7.88	-5.23	-5.61	0.72
-37.64	-88.79	-18.53	-62.37	-6.12	-61.74	-3.39	-29.85

3.Importance Bar Chart





Dec 12, 2024

Result Using PCA

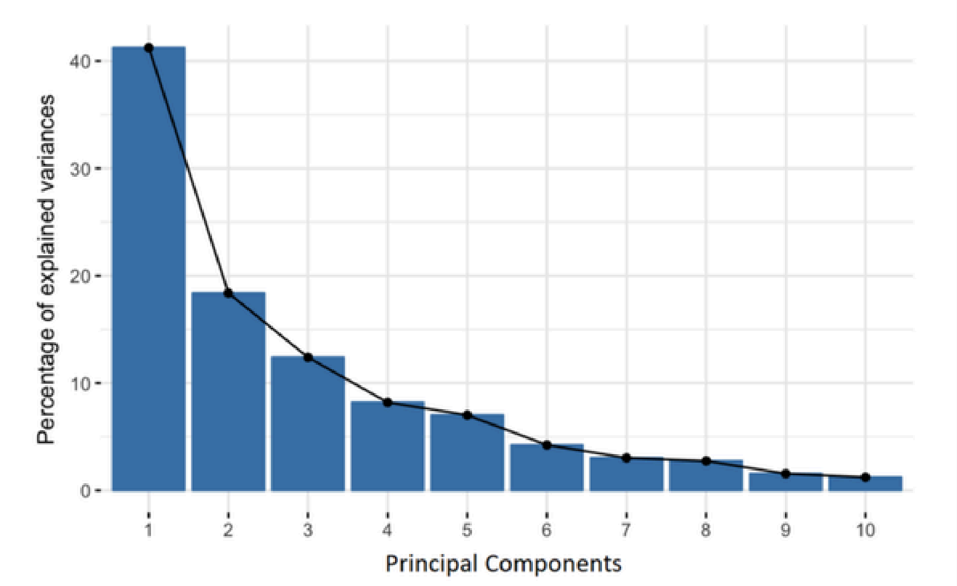
1. Metrics Matrix for Testing(KNN)

	CO_rmse	CO_mae	CO_r2	NOX_rmse	NOX_mae	NOX_r2
LinearRegression	1.578	0.828	0.544	8.230	5.760	0.490
SVR	1.408	0.585	0.637	5.759	3.834	0.750
DecisionTreeRegressor	1.482	0.651	0.598	5.798	3.629	0.747
RandomForestRegressor	1.308	0.493	0.687	4.214	2.651	0.866
XGBRegressor	1.335	0.534	0.674	4.355	2.937	0.857
ExtraTreesRegressor	1.294	0.482	0.693	3.978	2.507	0.881
MLPRegressor	0.944	0.542	0.764	5.609	3.928	0.755

2. Metrics Matrix for Testing(KNN +PCA)

	CO_rmse	CO_mae	CO_r2	NOX_rmse	NOX_mae	NOX_r2
LinearRegression	1.190	0.778	0.624	8.052	5.672	0.496
SVR	0.938	0.534	0.766	5.654	3.778	0.751
DecisionTreeRegressor	1.085	0.603	0.688	5.677	3.560	0.749
RandomForestRegressor	0.846	0.445	0.810	4.090	2.596	0.870
XGBRegressor	0.863	0.485	0.802	4.277	2.898	0.858
ExtraTreesRegressor	0.825	0.434	0.819	3.868	2.453	0.884
MLPRegressor	0.907	0.529	0.782	5.219	3.596	0.788

3.PCA Bar chart




```
from sklearn.decomposition import PCA
pca = PCA(n_components=0.95)  # 保留 95% 方差
x_train_pca = pca.fit_transform(X_train_std)
x_test_pca = pca.transform(X_test_std)
```

We bring in the test set data and add R-squared as an indicator. The influence of PCA on metrics was compared before and after

Conclusion

Dec 12, 2024

Increasing recognition of the adverse environmental and global health impacts of gas turbines has highlighted the urgent need to reduce nitrogen oxide and carbon monoxide emissions from gas turbines. In order to solve this problem, this study aims to investigate the prediction ability of traditional machine learning models for NOx and CO emissions from gas turbines, especially under very low computational conditions.

1. A thorough investigation of four conventional machine learning type of 8 machine learning models' prediction performance was conducted using the dataset of NOx and CO emission

2. EXTRA_TREE model became the best model using PCA and KNN, with a fit degree of 0.819 and 0.858 in the CO and NOX models, respectively, and the numerical performance of rsme and mae were the best

Future

1. At present, many models have commonly used default parameters, especially XGBOOST, which may perform better after adding parameter adjustments

2. With the further development of transformer, we can combine neural networks with traditional machine learning to take full advantage of all kinds of benefits

Dec 12, 2024

Thank you!

Reference

[1]Air pollutant emissions worldwide from 2015 to 2050, by compound, 2020. Statista. https://www.statista.com/statistics/1014091/air-pollutants-emissionsworldwide-by-compound/.
[2] IEA Global Energy Review. 2021. Paris: IEA; 2021. https://www.iea.org/reports/ global-energy-review-202.
[3] Assessment of options for the revision of the Industrial Emissions Directive, Final Report, Luxembourg: Publications Office of the European Union, 2021, http:// www.europa.eu.
[4] Mohamed H, Ticha HB, Mohamed S. Simulation of pollutant emissions from a gasturbine combustor. Combust Sci Technol 2004;176(5–6):819–34.
[5] Rizk NK, Mongia HC. Three-dimensional NOx model for rich/lean combustor, AIAA 31st Aerospace Sciences Meeting & Exhibit 1993;1 –14.
[6]Lin CY, Chen LW. Engine performance and emission characteristics of three-phase diesel emulsions prepared by an ultrasonic emulsification method. Fuel 2006;85 (5–6):593–600.
[7] Shakil M, Elshafei M, Habib MA, Maleki FA. Soft sensor for Nox and O2 using dynamic neural networks. Comput Electr Eng 2009;35(4):578–86.
[8] Kaya H, Tüfekci P, Uzun E. Predicting CO and Nox emissions from gas turbines: Novel data and a benchmark PEMS. Turk J Elect Eng Comput Sci 2019;27(6): 4783–96.
[9] Kaya H, Tüfekci P, Uzun E. Predicting CO and Nox emissions from gas turbines: Novel data and a benchmark PEMS. Turk J Elect Eng Comput Sci 2019;27(6): 4783–96.
[10] Lipperheide M, Weidner F, Wirsum M, Gassner M, Bernero S, Long-term Nox emission behavior of heavy-duty gas turbines: an approach for model-based monitoring and diagnostics, Turbo Expo: Power for Land, Sea, and Air, Charlotte, NC, USA, GT2017-63181;2017.

Reference

[11]Pachauri N. An emission predictive system for CO and NOx from gas turbine based on ensemble machine learning approach[J]. Fuel, 2024, 366: 131421.
[12]Rezazadeh A. Environmental pollution prediction of NOx by predictive modelling and process analysis in natural gas turbine power plants[J]. Pollution, 2021, 7(2): 481-494.