Gas Turbine Nitrogen Oxide Emission Reduction

STAT 443 Group 3

Overview

Dataset Introduction

- Objective
- Data Preparation

Data Exploration

- Data Splitting
- Outliers
- Correlations

Our Models

- Linear Regression
- Ridge Regression
- KNN
- Random Forest

Results

- R squared & MSPE

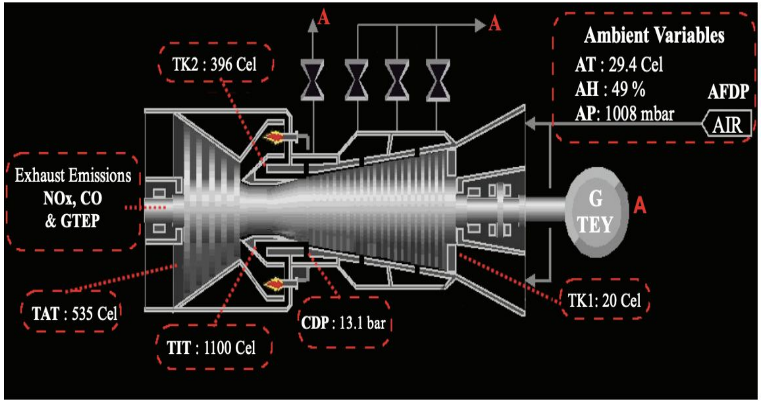
Conclusion

- Steps Moving Forward
- Suggestions

Introduction

Introduction

- Air pollution is responsible for around 11.65% of death in our population according to Our World In Data.
- Our Dataset
    - The UCI Machine Learning Repository dataset contains information on Turkish gas turbine carbon monoxide and nitrogen oxide emission.
    - Dataset contains measures related to the turbine and the surrounding environment.
    - 7158 observations, 11 variables



Heysem Kaya, PÄ±nar TÃ¼fekci and ErdinÃ§ Uzun. 'Predicting CO and NOx emissions from gas turbines: novel data and a benchmark PEMS', Turkish Journal of Electrical Engineering & Computer Sciences, vol. 27, 2019, pp. 4783-4796

Objective

- Create a model that aims to identify ways to reduce nitrogen oxide emission in relation to the turbine and the environment.
- Through analysis, we will find significant factors related to emission of Nitrogen Oxides (NOx)

Table 1. Basic statistical information of data used in the study.

Variable	Abbr.	Unit	Min	Max	Mean
Ambient temperature	AT	°C	–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/m^3	0.00	44.10	2.37
Nitrogen oxides	NO_x	mg/m^3	25.90	119.91	65.29

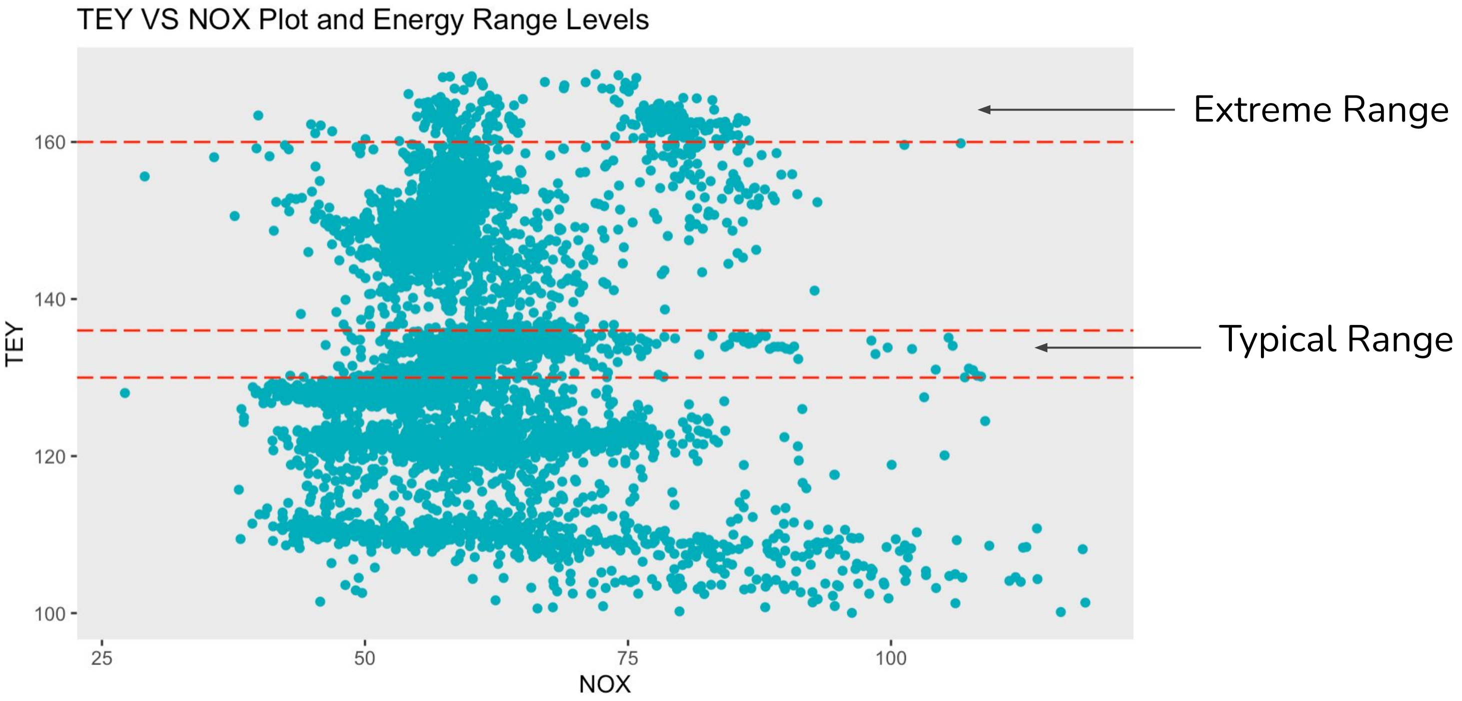
Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science

Data Preparation & Exploration

Data Preparation

- Removed CO to address the following research objectives:
  - Analyze the relationship between NOx and other predictors to order to reduce pollution
  - Identify relationships between turbine energy yield ranges and pollution
    - low energy [0 - 130 megawatts/hour]
    - middle lower energy [130 - 136 megawatts/hour]
    - middle upper energy [137 - 160 megawatts/hour]
    - high energy [160+ megawatts/hour]
- Split data into different ranges of TEY
- Split data into Train and Test groups

Energy Ranges for Analysis



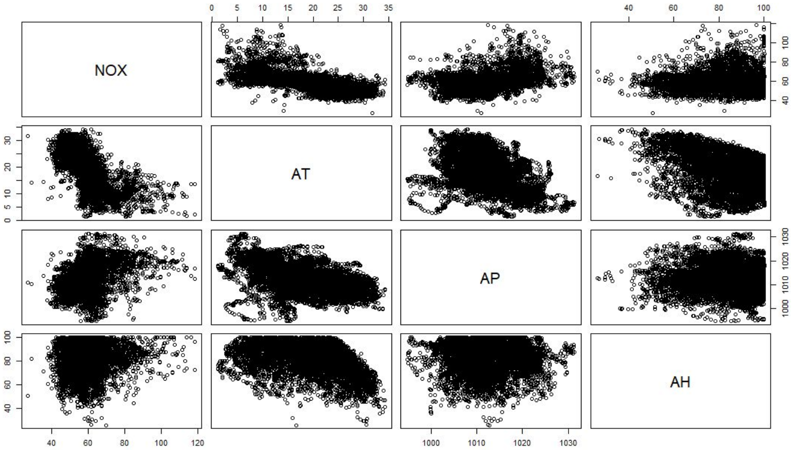
Data Exploration

Multicollinearity:

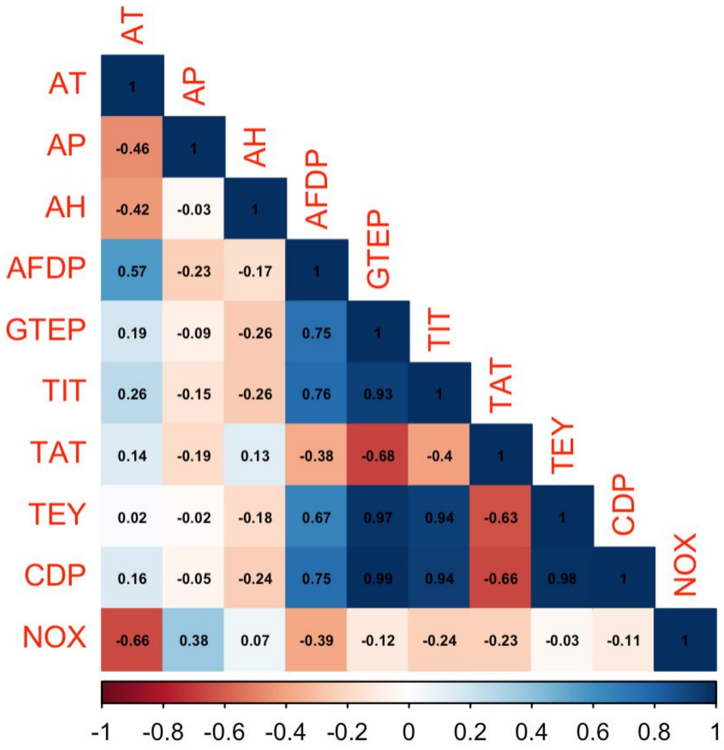
- Checked the correlation between the ambient variables and NOX
- Including the ambient variables, it accounted for the multicollinearity issue using VIF analysis.

Outliers:

- Outliers for all 4 ranges were observed through cook's distance and diagnostic plots
- Outliers were not removed prior to modeling

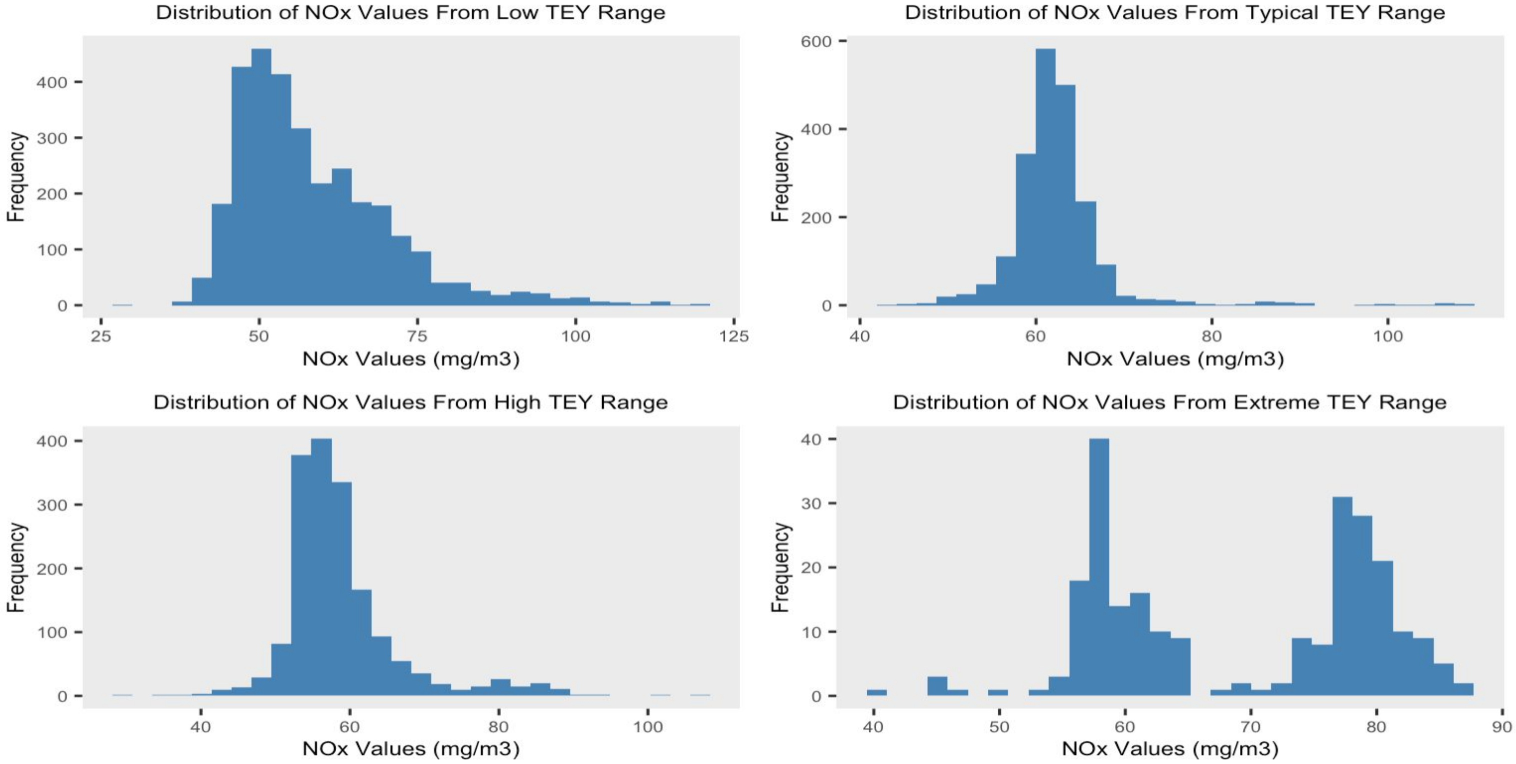


Correlations



Variable Name	Correlation with NOx
AFDP	-0.3852846
GTEP	-0.1174178
CDP	-0.1098818
TAT	-0.2273953
TIT	-0.2366259
TEY	-0.03187826
AT	-0.6567065
AP	0.3784665
AH	0.07064628

NOx Distribution for All Ranges



Model Building Process

Linear Regression Models

Typical Range Linear Regression								
Variable	Intercept	AFDP	CDP	TIT	TEY	AT	AP	AH
Estimate	153.86	-3.23	-5.34	-0.17	0.85	-0.19	0.07	-0.12

Extreme Range Linear Regression							
Variable	Intercept	AFDP	TIT	TAT	TEY	AT	AP
Estimate	696.92	11.59	-3.06	3.46	4.43	2.29	0.11

Ridge Regression Models

Typical Range Ridge Regression										
Variable	Intercept	AFDP	GTEP	CDP	TIT	TAT	TEY	AT	AP	AH
Estimate	383.48	-3.36	-0.74	-4.55	-0.14	-0.47	0.89	-0.15	0.06	-0.11

Extreme Range Ridge Regression										
Variable	Intercept	AFDP	GTEP	CDP	TIT	TAT	TEY	AT	AP	AH
Estimate	-1253.53	11.48	0.08	-2.73	-0.54	2.48	2.19	1.14	0.21	0.11

Model Comparison / Shortcomings

(Typical Energy Range)	MSPE	R Squared
Linear Regression	18.66805	0.3158
Ridge Regression	18.67567	0.31526
KNN Regression	27.71	N/A

(Extreme Energy Range)	MSPE	R Squared
Linear Regression	40.21	0.653
Ridge Regression	30.96	0.733
KNN Regression	119.95	N/A

Linear

- Multicollinearity
- Normality Issues
- High-Leverage Points
- Poor Model Performance

Ridge

- Lack of Feature Selection
- Homoscedasticity
- Poor Model Performance

KNN

- Poor Model Performance

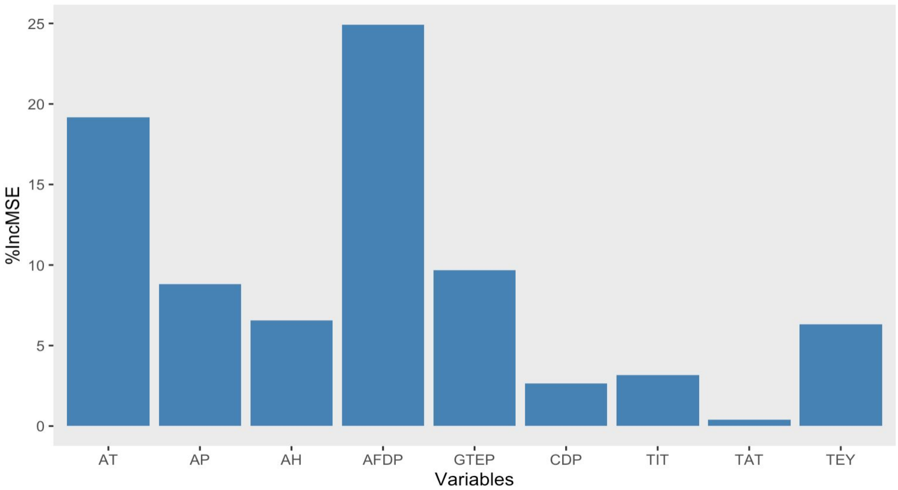
RandomForest / Final Model

Why Random Forest

- Resolves multicollinearity issues
- Versatile to unconventional data distributions (2 clusters, bimodal peaks, etc ...)
- Leverages bootstrapping to create better predictions that avoids overfitting and underfitting problems
- Easy to measure the relative importance of each feature for prediction

Random Forest Typical Energy Range

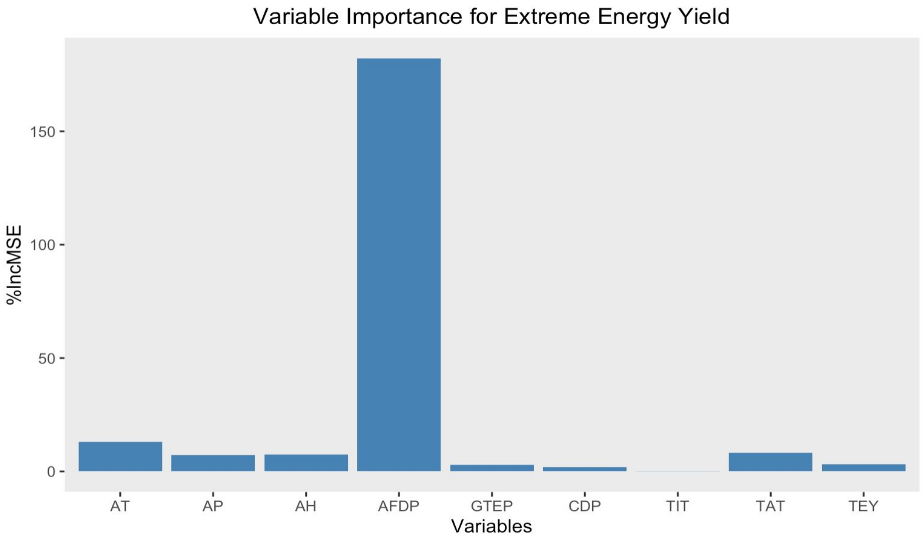
Estimator	
MSPE	**R Squared**
10.36214	0.620075



	AFDP	AT	GTEP	AP	AH	TEY	TIT	CDP	TAT
%IncMSE	24.92	19.81	9.69	8.82	6.56	6.32	3.16	2.65	0.38
Node Purity	12634.81	8597.78	3765.17	7155.25	4811.96	4482.14	3107.24	2277.10	1675.36

Random Forest Extreme Energy Range

Estimator	
MSPE	**R Squared**
6.364047	0.9450999



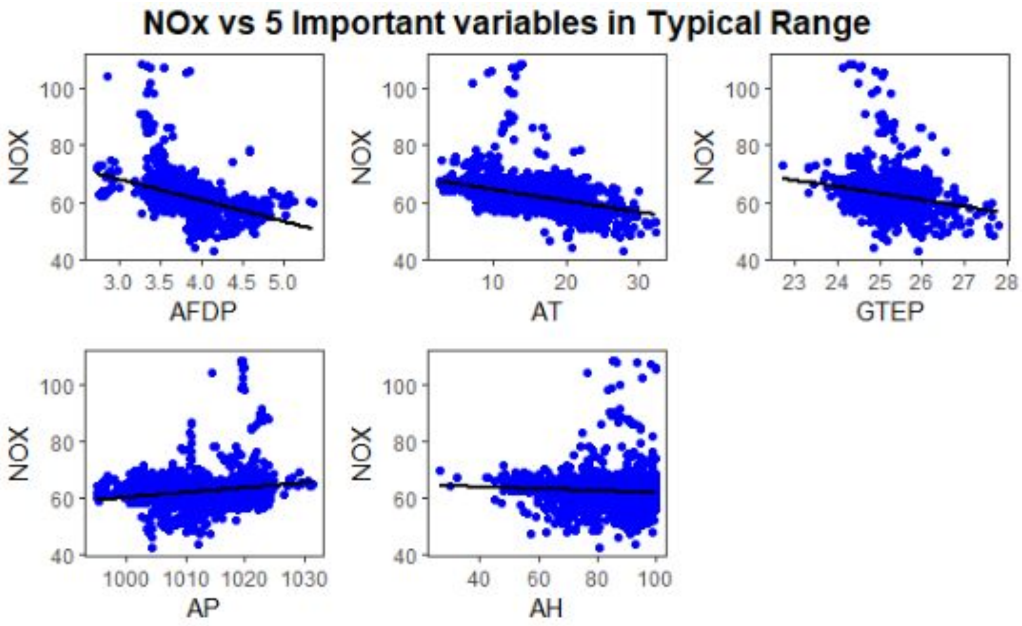
	AFDP	**AT**	**TAT**	**AH**	**AP**	**TEY**	**GTEP**	**CDP**	**TIT**
%IncMSE	182.16	12.98	8.28	7.81	7.04	3.27	2.86	1.94	0.06
Node Purity	13236.81	1580.02	1883.13	1346.26	638.38	269.36	497.39	292.61	55.06

Suggestions/Conclusion

Suggestions to Decrease NOx (Typical Range)

- Significant Variables
  - AFDP, AT, GTEP, AP, AH

- **Suggestions**
  - Increase: AFDP, GTEP,
  - Ambients: Prefer higher AT and AH, lower AP



Suggestions to Decrease NOx (Extreme Range)

- Suggestions on AFDP
  - Negative relationship between NOx and AFDP by clusters
  - Recommend to maintain AFDP lower than 4.25

Conclusion / Scientific Reasoning

- Typical Energy Range: Increase AFDP and GTEP
    - Filter pressure drop corresponds to a filter's air flow rate
    - Higher the pressure drop, the more restrictive the filter is to air flow. This could potentially decrease NOx levels as a result (airfilterusa)
- Extreme Energy Range: Reduce AFDP, as the cluster at level 4 shows lower NOx
    - High energy → high temperature (Particles would have a higher kinetic energy)
    - High temperature accelerates the rate of combustion and results in a high flue gas temperature (Li et al., 2012)
    - High flue gas temperature effects:
        - Drives reactions to the formation nitrogen rather than NOx

References

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science

Hannah Ritchie and Max Roser (2017) - "Air Pollution". Published online at OurWorldInData.org. Retrieved from:'https://ourworldindata.org/air-pollution' [Online Resource]

Heysem Kaya, PÄ±nar TÃ¼fekci and ErdinÃ§ Uzun. 'Predicting CO and NOx emissions from gas turbines: novel data and a benchmark PEMS', Turkish Journal of Electrical Engineering & Computer Sciences, vol. 27, 2019, pp. 4783-4796

Li, Zhengqi, and Yong Liu. “ACS Publications: Chemistry Journals, Books, and References Published ...” *Effect of the Air Temperature on Combustion Characteristics and NOx Emissions from a 0.5 MW Pulverized Coal-Fired Furnace with Deep AirStaging*, ACS Publications, 23 Mar. 2012, https://pubs.acs.org/doi/abs/10.1021/ef300233k.

Questions?