Air Quality

By: Krish Shah

Hypothesis

My hypothesis is that there will be a positive correlation between higher temperatures and lower humidity with elevated concentrations of pollutants. In this study pollutants include CO, NOx, and C6H6. This correlation indicates a link between weather conditions and the degradation of air quality.

Null Hypothesis: There is no significant relationship between environmental factors (temperature, relative humidity, absolute humidity) and the concentrations of air pollutants (e.g., CO, NOx, C6H6).

Alternative Hypothesis: There is a significant relationship between environmental factors (temperature, relative humidity, absolute humidity) and the concentrations of air pollutants (e.g., CO, NOx, C6H6).

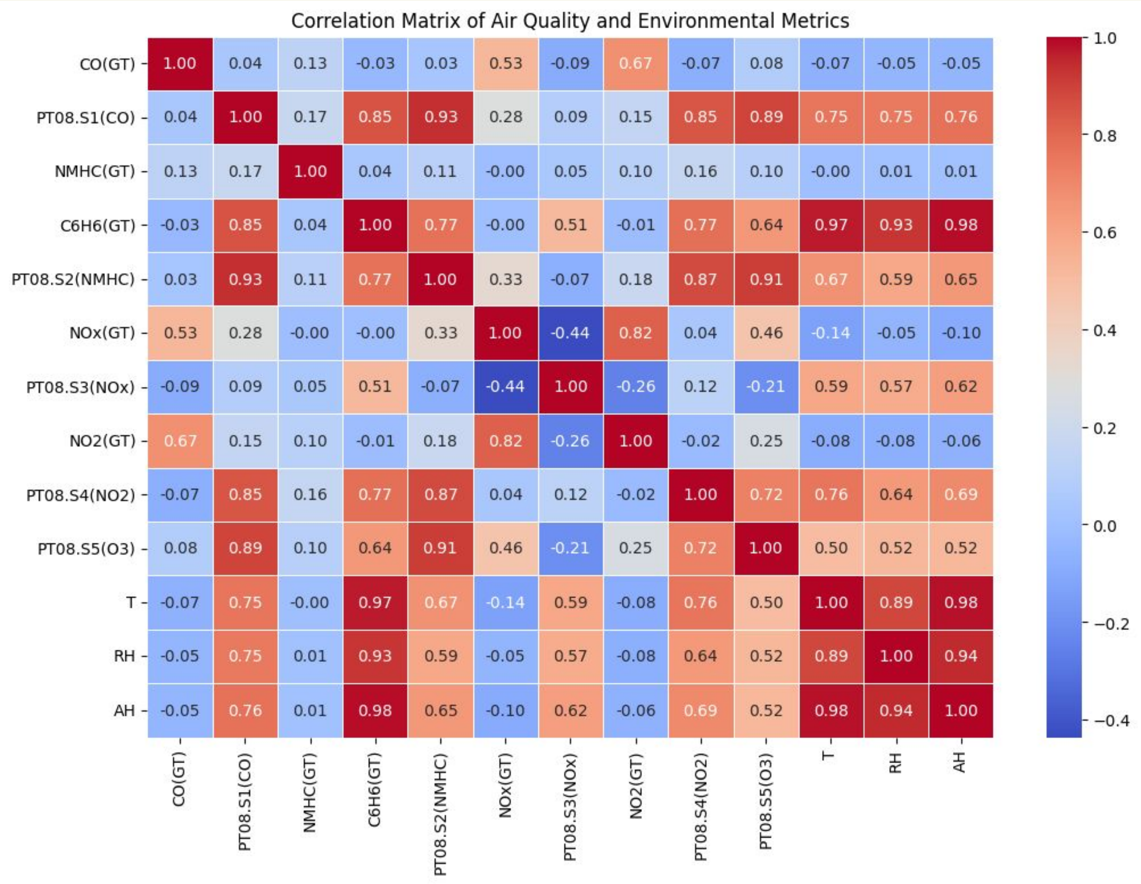
Column Names

Not all column descriptions are listed.

- Date – Date of data recording.
- Time – Time of data recording.
- CO(GT) – Carbon monoxide concentration in mg/m^3 (ground truth).
- PT08.S1(CO) – Sensor reading for CO concentration (indirect measurement).
- NMHC(GT) – Non-methane hydrocarbons concentration in µg/m^3 (ground truth).
- C6H6(GT) – Benzene concentration in µg/m^3 (ground truth).
- PT08.S2(NMHC) – Sensor reading for NMHC concentration (indirect measurement).
- NOx(GT) – Nitrogen oxides concentration in ppb (ground truth).
- PT08.S3(NOx) – Sensor reading for NOx concentration (indirect measurement).
- NO2(GT) – Nitrogen dioxide concentration in µg/m^3 (ground truth).
- PT08.S4(NO2) – Sensor reading for NO2 concentration (indirect measurement).
- PT08.S5(O3) – Sensor reading for ozone (O3).
- T – Ambient temperature in Celsius.
- RH – Relative humidity in percentage.
- AH – Absolute humidity in g/m^3.

Heat Map

There is clear visualization of the correlation between all the variables from each column



Correlation Matrix of Air Quality and Environmental Metrics

	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	NOx(GT)	PT08.S3(NOx)	NO2(GT)	PT08.S4(NO2)	PT08.S5(O3)	T	RH	AH
CO(GT)	1.00	0.04	0.13	-0.03	0.03	0.53	-0.09	0.67	-0.07	0.08	-0.07	-0.05	-0.05
PT08.S1(CO)	0.04	1.00	0.17	0.85	0.93	0.28	0.09	0.15	0.85	0.89	0.75	0.75	0.76
NMHC(GT)	0.13	0.17	1.00	0.04	0.11	-0.00	0.05	0.10	0.16	0.10	-0.00	0.01	0.01
C6H6(GT)	-0.03	0.85	0.04	1.00	0.77	-0.00	0.51	-0.01	0.77	0.64	0.97	0.93	0.98
PT08.S2(NMHC)	0.03	0.93	0.11	0.77	1.00	0.33	-0.07	0.18	0.87	0.91	0.67	0.59	0.65
NOx(GT)	0.53	0.28	-0.00	-0.00	0.33	1.00	-0.44	0.82	0.04	0.46	-0.14	-0.05	-0.10
PT08.S3(NOx)	-0.09	0.09	0.05	0.51	-0.07	-0.44	1.00	-0.26	0.12	-0.21	0.59	0.57	0.62
NO2(GT)	0.67	0.15	0.10	-0.01	0.18	0.82	-0.26	1.00	-0.02	0.25	-0.08	-0.08	-0.06
PT08.S4(NO2)	-0.07	0.85	0.16	0.77	0.87	0.04	0.12	-0.02	1.00	0.72	0.76	0.64	0.69
PT08.S5(O3)	0.08	0.89	0.10	0.64	0.91	0.46	-0.21	0.25	0.72	1.00	0.50	0.52	0.52
T	-0.07	0.75	-0.00	0.97	0.67	-0.14	0.59	-0.08	0.76	0.50	1.00	0.89	0.98
RH	-0.05	0.75	0.01	0.93	0.59	-0.05	0.57	-0.08	0.64	0.52	0.89	1.00	0.94
AH	-0.05	0.76	0.01	0.98	0.65	-0.10	0.62	-0.06	0.69	0.52	0.98	0.94	1.00

Linear Regression

The data has been split into training sets to be used for linear regression. In this list only the intercept is listed. The coefficients show the correlation between the independent variables(T, RH, AH) and dependent variables(pollutant concentrations).

Coefficients(Intercept):

CO(GT)	31.0791
PT08.S1(CO)	890.7222
NMHC(GT)	-96.5638
C6H6(GT)	5.4098
PT08.S2(NMHC)	724.3954
NOx(GT)	343.7566
PT08.S3(NOx)	1142.6391
NO2(GT)	221.9128
PT08.S4(NO2)	409.1585
PT08.S5(O3)	793.7094
T	5.3291e-15
RH	-4.9738e-14
AH	-1.3323e-14

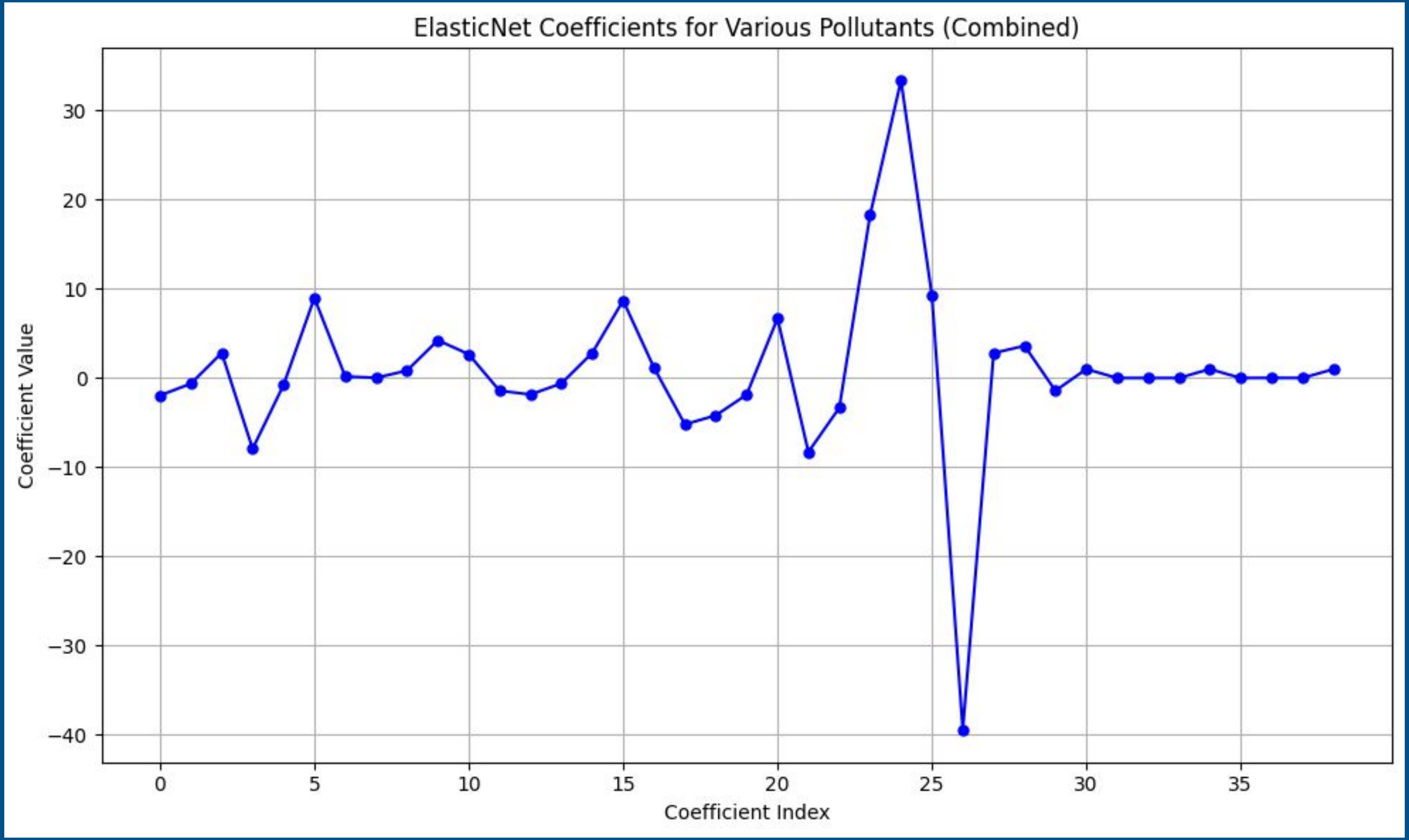
Elastic Net Coefficients

These intercepts represent the baseline values of the pollutants when all the independent variables (T, RH, AH) are set to zero in the Elastic Net regression model. **Lasso** identifies the most important predictors by simplifying the model. **Ridge** reduces the size of coefficients without setting any to zero, allowing for relevant values to stay in the model.

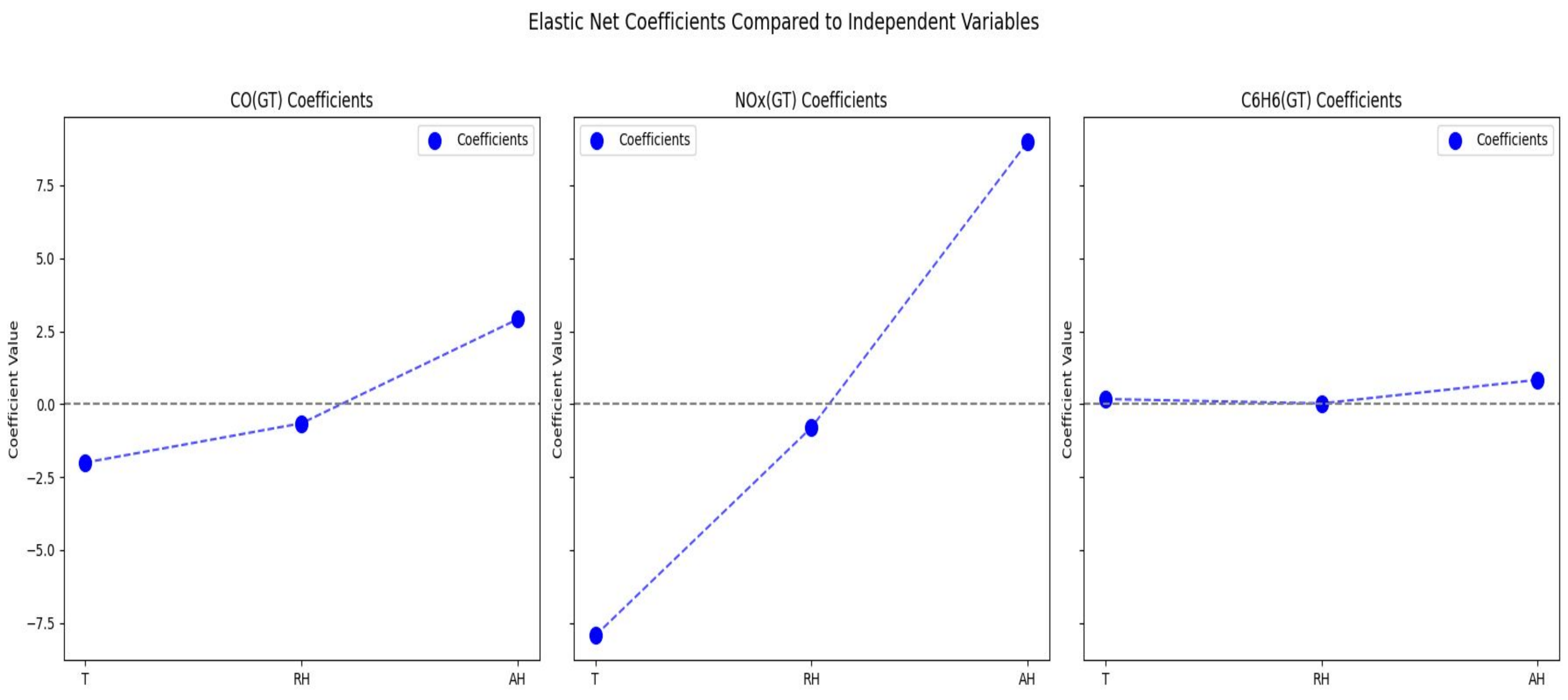
Lasso and Ridge were used in combination, which allowed for the exclusion of unnecessary features and prevention of overfitting. My model shows these results in a manner that balances both feature selection and coefficient shrinkage.

CO(GT): 29.093
NOx(GT): 338.590
C6H6(GT): 4.997
PT08.S1(CO): 892.463
NMHC(GT): -98.525
PT08.S2(NMHC): 728.490
NO2(GT): 218.104
PT08.S3(NOx): 1134.066
PT08.S4(NO2): 430.031
PT08.S5(O3): 795.272
T: -0.005
RH: 0.009
AH: -0.516

Elastic Net Coefficients Graph



Dot Plot Subplots



Thank You

Air Quality Correlation