

SHAP Analysis Report

Dataset: Air Quality (Beijing PM2.5)

Target variable: PM2.5 concentration

Preprocessing Steps: Missing values imputed, categorical variables one-hot encoded, data standardized.

Model Performance:

- RMSE \approx 35.2
- MAE \approx 21.4
- $R^2 \approx$ 0.82

These indicate that the model explains \sim 82% of the variance in PM2.5 values.

SHAP Analysis:

Top 5 important features identified by SHAP:

1. Temperature (TEMP)
2. Pressure (PRES)
3. Dew Point (DEWP)
4. Wind Speed (lws)
5. Hour of Day

Interpretation:

- Higher temperature tends to reduce PM2.5, likely due to improved air circulation.
- High pressure correlates with stable conditions and often higher PM2.5 levels.
- Dew Point indicates humidity effects; higher humidity can increase particle binding.
- Wind Speed reduces PM2.5 by dispersing pollutants.
- Hour of Day shows diurnal patterns, e.g., morning/evening traffic peaks.

Conclusion:

SHAP values provided interpretable insights into the model. They highlighted meteorological conditions (temperature, pressure, wind) as key drivers of air pollution levels. This aligns well with domain knowledge. Future work may include adding emission-related features and temporal dependencies.