



Predicting Air Quality Index with Data Science and Machine Learning

GROLLEAU Florian

2A–Dual Diploma

Sciences Po Saint-Germain-en-Laye / CY TECH Engineering School

July 2023

Abstract

Since urbanization and high-density human activities such as industrialization and overconsumption of services and goods, the global Air Quality Index (AQI) has regressed. However, the global AQI does not reflect the very localized impact and changes on a rather small geographic area. For this reason, there is a need for more accurate metrics and qualitative analysis of the AQI over smaller regions than a worldwide analysis. Thanks to the emergence of several Data Science tools and Machine Learning models, we aim for better monitoring and prediction of local AQI.

Contents

1	What is AQI?	4
1.1	Definition and Purpose	4
1.2	How AQI is Calculated	4
1.3	Global vs Local AQI	4
2	Data Collection and Methodology	4
2.1	Data Sources	4
2.2	Preprocessing and Cleaning	4
2.3	Feature Selection	4
2.4	Machine Learning Models Used	4
3	Results	4
3.1	Model Performance	4
3.2	Comparison with Existing AQI Predictions	4
3.3	Case Studies on Specific Regions	4

4	Discussion	4
4.1	Interpretation of Results	4
4.2	Limitations of the Study	4
4.3	Potential Improvements	4
5	Conclusion and Potential Improvements	4
5.1	Summary of Findings	4
5.2	Applications in Policy and Urban Planning	4
5.3	Next Steps in Research	4



Predicting Air Quality Index with Data Science and Machine Learning

GROLLEAU Florian

2A–Dual Diploma

Sciences Po Saint-Germain-en-Laye / CY TECH Engineering School

July 2023

1 What is AQI?

1.1 Definition and Purpose

1.2 How AQI is Calculated

1.3 Global vs Local AQI

2 Data Collection and Methodology

2.1 Data Sources

2.2 Preprocessing and Cleaning

2.3 Feature Selection

2.4 Machine Learning Models Used

3 Results

3.1 Model Performance

3.2 Comparison with Existing AQI Predictions

3.3 Case Studies on Specific Regions

4 Discussion

4.1 Interpretation of Results

4.2 Limitations of the Study

4.3 Potential Improvements

5 Conclusion and Potential Improvements

5.1 Summary of Findings

5.2 Applications in Policy and Urban Planning

5.3 Next Steps in Research