Applying Machine Learning To London’s Pollution Problem

By

Nayi Cai, Peter Clark, Tengfeng Peng, Seungmin Kim,

Sinan Bilir

Submitted to

The University of Liverpool

MASTER-OF-DATA SCIENCE AND AI PROGRAMME

*Machine Learning in Practice*

Word Count: 1400

07/10/2024

Applying Machine Learning To London’s Pollution Problem

Submitted to

The University of Liverpool

Word Count: 1400

07/10/2024

TABLE OF CONTENTS

Page

[LIST OF TABLES 2](#_Toc179202241)

[LIST OF FIGURES 3](#_Toc179202242)

[Chapter 1. Introduction 5](#_Toc179202243)

[Chapter 2. High Level Feature Selection 6](#_Toc179202244)

[Chapter 3. Data Inferences/ Cleaning 8](#_Toc179202245)

[3.1 Loading the Dataset 8](#_Toc179202246)

[3.2 Handling Missing Data 8](#_Toc179202247)

[3.3 Detecting and Removing Duplicates 9](#_Toc179202248)

[3.4 Adjusting Data Types 9](#_Toc179202249)

[3.5 Renaming Columns 9](#_Toc179202250)

[Chapter 4. Machine Learning Model Goals and Techniques 10](#_Toc179202251)

[4.1 Machine learning techniques used and their applications 10](#_Toc179202252)

[Chapter 5. Python Code Review 11](#_Toc179202253)

[5.1 Function use overview 11](#_Toc179202254)

[5.2 Data Sources 12](#_Toc179202255)

[5.3 Python Code 13](#_Toc179202256)

[5.3.1 Part One: Correlate vehicle metrics with emissions by TOID. 13](#_Toc179202257)

[5.3.2 Part Two: Correlate pollutant emissions with concentrations by grid coordinate. 15](#_Toc179202258)

[5.3.3 Part Three: Create final models predicting concentrations of pollutant from vehicle metrics. 17](#_Toc179202259)

[Chapter 6. Findings, Analysis and predictions 19](#_Toc179202260)

[6.1 Part One 19](#_Toc179202261)

[6.2 Part Two 21](#_Toc179202262)

[6.3 Part Three 23](#_Toc179202263)

[6.4 Predictions 24](#_Toc179202264)

[Chapter 7. Future Improvements 24](#_Toc179202265)

[Chapter 8. Conclusion 25](#_Toc179202266)

[REFERENCES 27](#_Toc179202267)

[APPENDICES 28](#_Toc179202268)

LIST OF TABLES

[Table 1: Machine Learning Approaches 11](#_Toc179224999)

[Table 2: Principal DataFrame Functions used 11](#_Toc179225000)

[Table 3: Model Function used 11](#_Toc179225001)

[Table 4: Data Sources 12](#_Toc179225002)

[Table 5: NOx LIR concentration performance metrics 19](#_Toc179225003)

[Table 6: NOx LIR concentration performance metrics 19](#_Toc179225004)

[Table 7: NOx LIR concentration performance metrics 21](#_Toc179225005)

[Table 8: pm25 LIR concentration performance metrics 21](#_Toc179225006)

[Table 9: NOx LIR concentration performance metrics 22](#_Toc179225007)

[Table 100: pm25 LIR concentration performance metrics 22](#_Toc179225008)

[Table 11: Predictions 22](#_Toc179225009)

[Table 12: Future Improvements 23](#_Toc179225010)

LIST OF FIGURES

Page

[Figure 1: NOx concentrations 2019 6](#_Toc179195810)

[Figure 2: pm25 Concentrations 2019 6](#_Toc179195811)

[Figure 3: Data source definition and library import 12](#_Toc179195812)

[Figure 4: Load Part One data 12](#_Toc179195813)

[Figure 5: List comprehension to copy relevant columns 12](#_Toc179195814)

[Figure 6: Sum emissions from vehicle classes 13](#_Toc179195815)

[Figure 7: Retrieve speed data, grouping by mean 13](#_Toc179195816)

[Figure 8: Use map to create a summary DataFrame correlated by TOID 13](#_Toc179195817)

[Figure 9: Load AADT by TOID. (VKM uses same method) 13](#_Toc179195818)

[Figure 10: Remove whitespace and drop non-numeric values 13](#_Toc179195819)

[Figure 11: Use merge when adding to summary\_df 13](#_Toc179195820)

[Figure 12: Training and use of RFR 13](#_Toc179195821)

[Figure 13: Preparation for and use of PCA 14](#_Toc179195822)

[Figure 14: Load Part Two data 14](#_Toc179195823)

[Figure 15: Use column filtering to reduce source data dimensions 14](#_Toc179195824)

[Figure 16: Use astype and concatenation to combine coordinates into single feature 15](#_Toc179195825)

[Figure 17: Use sum and lamda to calculate the percentage of total nox for a coordinate 15](#_Toc179195826)

[Figure 18: Filter on non-zero values and by road transport 15](#_Toc179195827)

[Figure 19: Combine grid coordinates in concentration data 15](#_Toc179195828)

[Figure 20: Use apply to calculate percentage concentration 15](#_Toc179195829)

[Figure 21: Create LIR models for each pollutant using a pipeline 16](#_Toc179195830)

[Figure 22: Apply LIR model to feature 16](#_Toc179195831)

[Figure 23: Create LOR for each pollutant based on key features 16](#_Toc179195832)

[Figure 24: Assess performance 17](#_Toc179195833)

[Figure 25: RFR for NOx Emissions 18](#_Toc179195834)

[Figure 26: RFR for pm25 Emissions 19](#_Toc179195835)

[Figure 27: PCA results for four principal components 20](#_Toc179195836)

[Figure 28: NOx concentration LIR model 21](#_Toc179195837)

[Figure 29: pm25 concentration LIR model 22](#_Toc179195838)

# Introduction

Atmospheric emissions have a significant impact on both public health and environmental sustainability. The project aims to develop a machine learning model that predicts emissions based on 2019 data supplied by the London Atmospheric Emissions Inventory (LAEI). Supervised models such as Linear Regression (LIR), Logistic Regression (LOR), Random Forest Regressor (RFR), and the unsupervised model Principal Component Analysis (PCA) are employed to offer predictive insights into emission levels across London boroughs. By delivering accurate emission forecasts, the project will help government bodies make well-informed decisions about environmental policies and regulations.

# High Level Feature Selection

Reviewing the summary document (LAEI 2019 Summary, 2021) it is apparent that NOx and pm25 are key pollutants and Road Transport the key contributor. Most sections focus on the two pollutants, including the population exposure section, which considers educational, health and care facilities, the dataset description itself and the World Health Organisation (WHO) limits. Later sections of the document highlights Road Transport’s contribution, examples include Figure 13 which shows the percentage of road kilometers meeting 30ug/m3 as well as Figure 17 and Figure 25 which show Road Transport as the greatest contributor to the chosen pollutants. This is visually obvious in *Figure 1* and *Figure 2* which are available separately at LAEI 2019 Concentrations (2019).

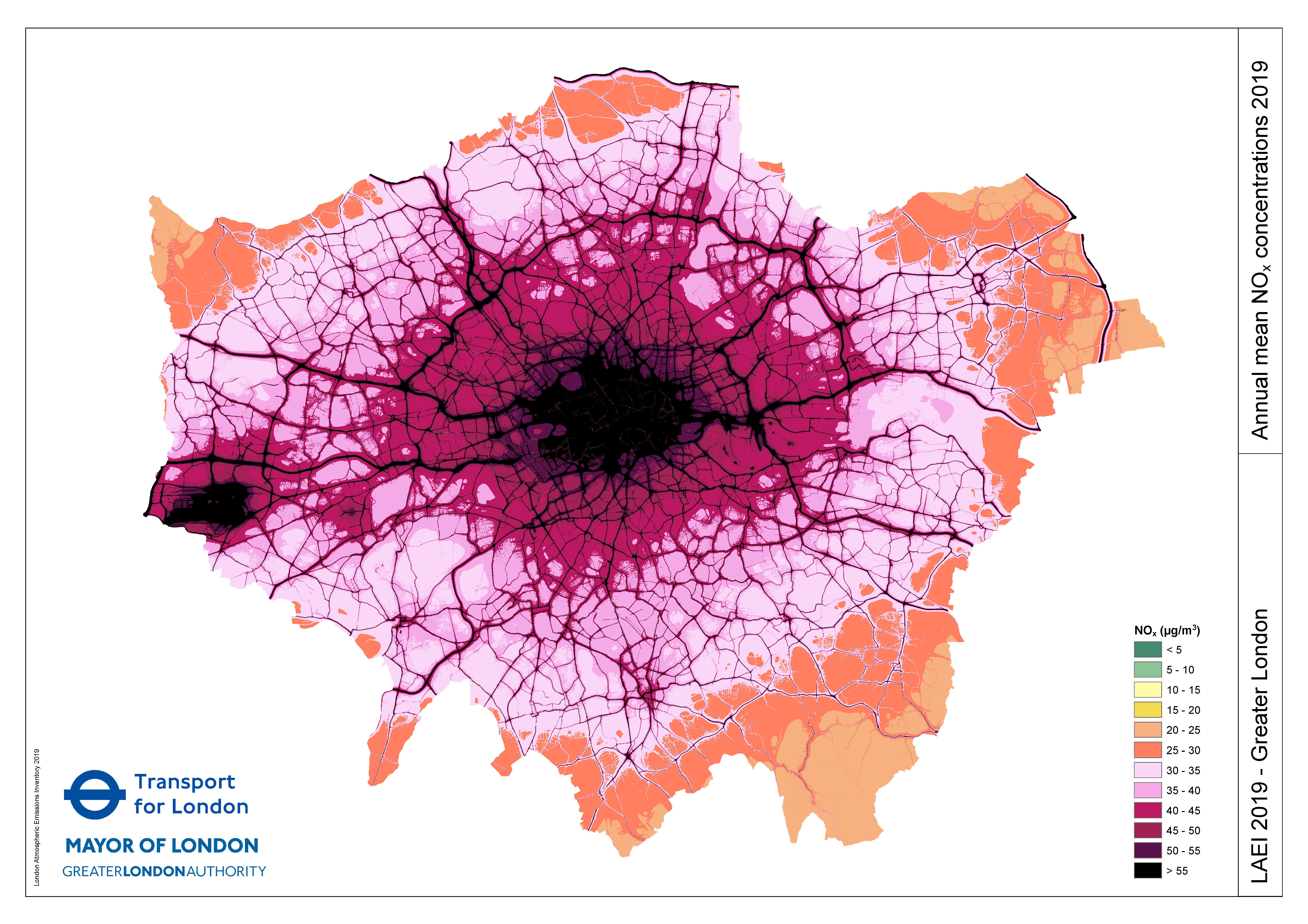


Figure 1: NOx concentrations 2019

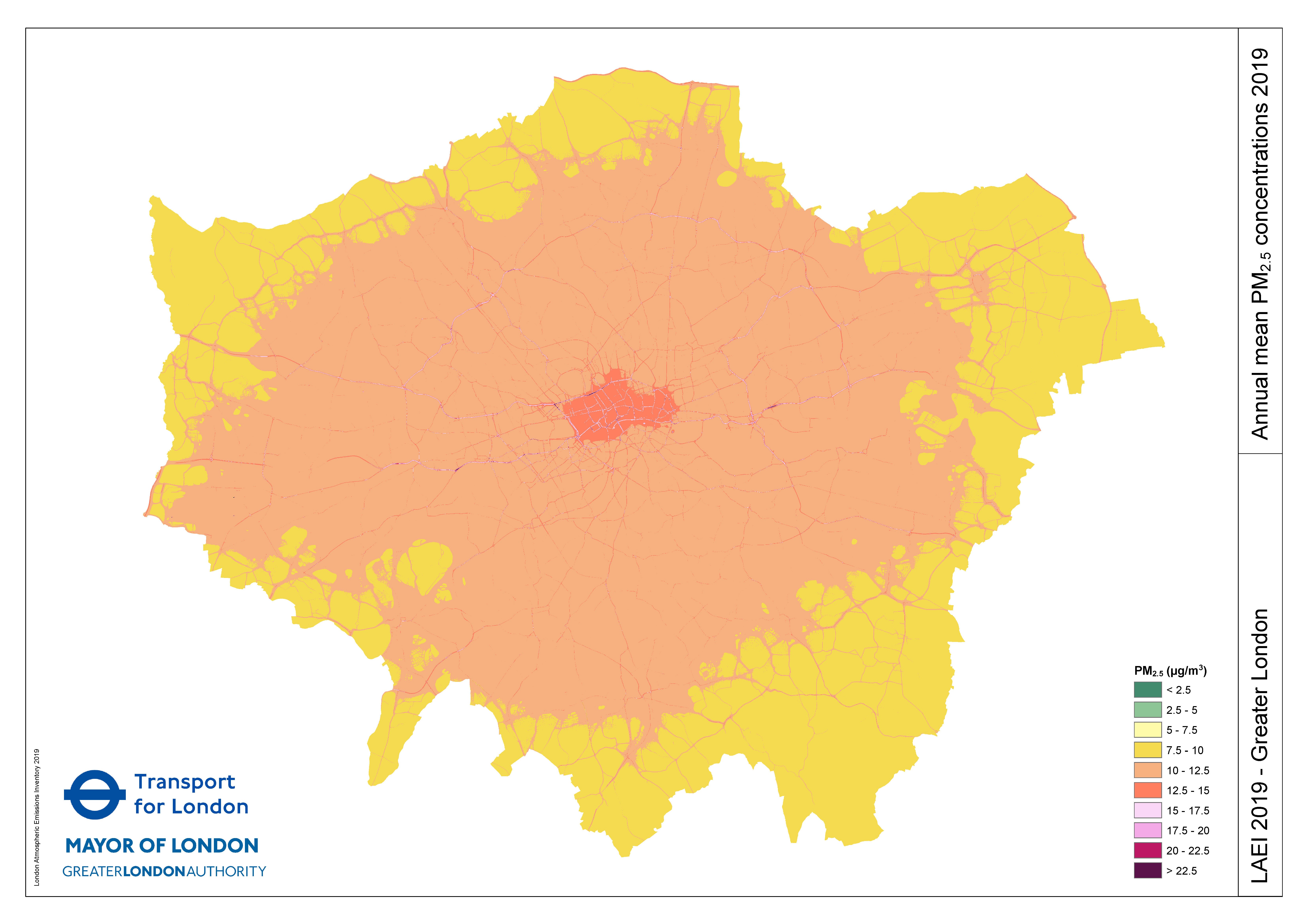


Figure 2: pm25 Concentrations 2019

# Data Inferences/ Cleaning

In our project we created a Jupyter Notebook to clean, prepare and analyze data, utilizing Pandas to handle and process the dataset, Matplotlib for visualization and Sklearn for pre-processing, algorithms and metrics.

## Loading the Dataset

Data is loaded into DataFrame objects using pandas.read\_excel() or pandas.readcsv() depending on file format. Pandas.info() provided the dataset’s characteristics, such as data types, non-null values, and column names. This step facilitated understanding the overall structure of the dataset and identification of key characteristics, missing values, and incorrect data types.

## Handling Missing Data

There are many options to handle missing data, we consider two suitable for this dataset outlined (Han, 2023). The aim is to ensure datasets contain no non-null values without introducing bias (Bell, 2020), and is implementing using pandas isnull(), dropna() and fillna() functions.

* + **Option 1: Drop Rows with Missing Data:**

This option is effective with minimal missing data and when its exclusion does not affect the overall analysis.

* + **Option 2: Impute Missing Data**

Imputation is useful when we want to preserve the dataset’s size without introducing bias.

## Detecting and Removing Duplicates

We checked for duplicate rows to eliminate redundant data entries that could distort our analysis using pandas.duplicated() and drop\_duplicates() functions.

## Adjusting Data Types

We examined the data types of each column by looking at the dtypes attribute, to ensure consistency and correctness and converted features to more appropriate data types using astype(). This was essential as many models only accept numeric data and to ensure accurate computations.

## Renaming Columns

We renamed features where necessary to improve readability and clarity, using the pandas rename() function, making it easier for team members and stakeholders to interpret the data correctly.

# Machine Learning Model Goals and Techniques

The primary goal of the machine learning model is to predict atmospheric emissions and pollution levels in London. Providing insights from this data can help the Department of Environment mitigate pollution. The model will achieve this by analysing historical data on key pollutants such as nitrogen oxides (NOx) and particulate matter (PM2.5) (LAEI 2019 Summary, 2021). This will allow for more informed decisions regarding urban planning, traffic regulation, and public health advisories, particularly in roads prone to high emissions.

Four key machine learning approaches are used: **linear regression**, **logistic regression**, **random forest regression**, and **principal component analysis (PCA)**.

## Machine learning techniques used and their applications

|  |  |  |
| --- | --- | --- |
| **ML Technique** | **Description** | **Application** |
| Linear Regression | A basic regression method used to predict the value of a continuous target variable based on one or more predictor variables | To establish relationships between pollutant concentrations and environmental variables  (Aram *et al*., 2024) |
| Logistic Regression | A classification technique that estimates the probability of a binary outcome based on input features | Helps classify pollution levels by estimating the probability of exceeding safe thresholds  (Aram *et al*., 2024) |
| Random Forest Regression | An ensemble method that builds multiple decision trees and averages their predictions to provide more accurate and stable results | Extract the most relevant information on each pollutant  (Suman, 2024) |
| Principal Component Analysis (PCA) | A dimensionality reduction technique that transforms a dataset into a smaller set of uncorrelated variables called principal components | Determines variance in the four gases  (Jiang, 2023) |

Table 1: Machine Learning Approaches

# Python Code Review

## Function use overview

|  |  |
| --- | --- |
| **Function** | **Use** |
| Group by | Remove duplicate entries for a feature, normally using sum |
| filter | Extract target columns from DataFrame |
| map | Add values from one DataFrame to another based on column values |
| apply | In conjunction with lambdas to apply equation to a whole column |
| drop | Remove unnecessary or temporary columns |
| dropna | Cleanup data |

Table 2: Principal DataFrame Functions used

|  |  |
| --- | --- |
| **Function** | **Use, informed by**  Boschetti, A. and Massaron, L. (2018) andSciKit Learn. (2024). |
| fit | Train the model |
| predict | Test and use the model |
| mean\_absolute\_error | Mean of absolute differences between real and predicted points |
| mean\_squared\_error | Square root of sum of squared differences, more sensitive to large differences. |
| r2\_score | Coefficient of determination, determines the linear fit of predictors and target. |
| score | Mean accuracy of test set and labels |
| accuracy\_score | Percentage of correctly classified labels |
| precision\_score | Average of relevant results for Linear Regression or correct labels for each classification in Logistic Regression |
| recall\_score | Averaged comparison of relevant results to relevant labels in Linear Regression or correctly classified labels divided by total count of labels in a set in Logistic regression |
| f1\_score | Harmonic average of precision and recall. |
| roc\_auc\_score | For Logistic Regression, shows performance changes over classifier thresholds |

Table 3: Model Function used

## Data Sources

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | Part | Filename | Sheets | Reason |
| 1 | One | laei-2019-major-roads-vkm  -flows-speeds.xlsx | laei-2019-major-roads | Speed, Annual Average Daily Traffic (AADT) and Vehicle Kilometre (VKM) by Topographic Identifier (TOID) |
| 2 | One | LAEI2019-nox-pm-co2-major  -roads-link-emissions.xlsx | NOx\_Road\_Link\_Emissions  PM25\_Road\_Link\_Emissions | Emission data by TOID |
| 3 | Two | LAEI-2019-Emissions-Summary  -including-Forecast.xlsx | Emissions by Grid ID | Emission data by Grid Coordinate |
| 4 | Two | laei\_LAEI2019v3\_  CorNOx15\_NOx.csv | N/A | Concentrations of NOx by Grid Coordinate |
| 5 | Two | laei\_LAEI2019v3\_  CorNOx15\_PM25.csv | N/A | Concentrations of PM25 by Grid Coordinate |

Table 4: Data Sources

## Python Code

Set data source disk location variables and imports.

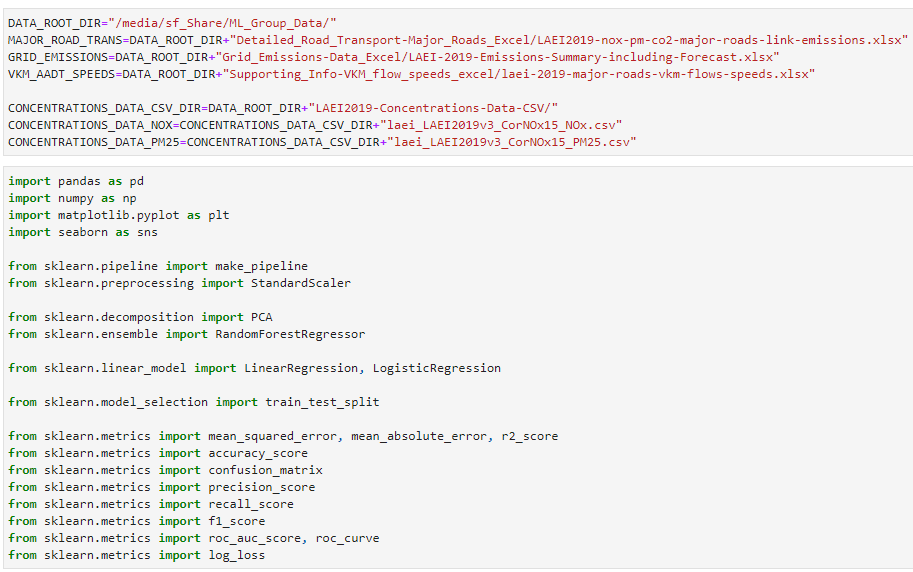


Figure 3: Data source definition and library import

The code is split into three parts.

### Part One: Correlate vehicle metrics with emissions by TOID.

1. Load part one sources, ref Table 4: Data Sources



Figure 4: Load Part One data

1. Filter 2019 data.

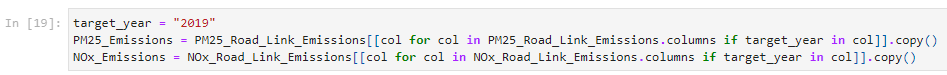


Figure 5: List comprehension to copy relevant columns

1. Combine Vehicle Speeds, VKM and AADT with NOx and CO2 emissions by TOID.
   1. Emissions are summed to reduce data dimensions

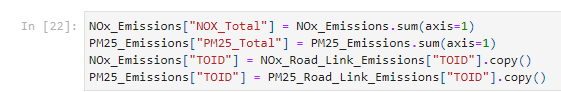


Figure 6: Sum emissions from vehicle classes

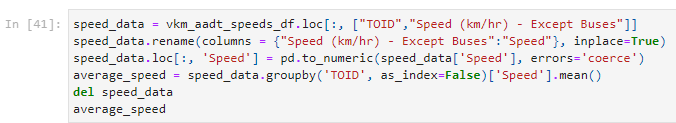


Figure 7: Retrieve speed data, grouping by mean

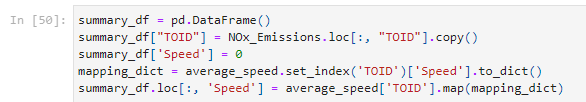


Figure 8: Use map to create a summary DataFrame correlated by TOID

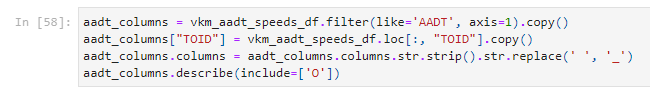


Figure 9: Load AADT by TOID. (VKM uses same method)

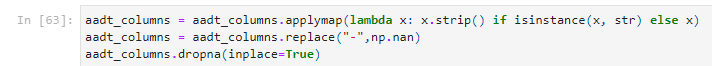


Figure 10: Remove whitespace and drop non-numeric values



Figure 11: Use merge when adding to summary\_df

1. Use RFR for each pollutant, to narrow vehicle statistics feature list.

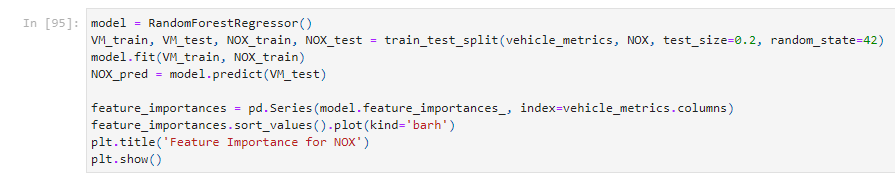


Figure 12: Training and use of RFR

1. Use PCA to verify chosen contributing features.

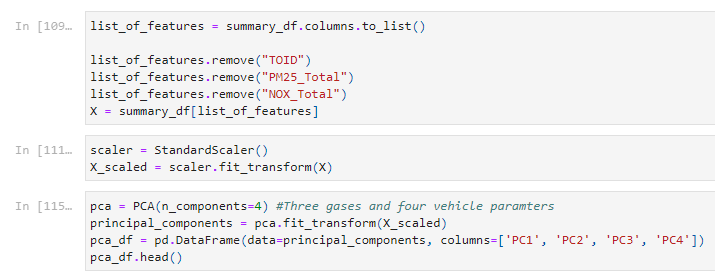


Figure 13: Preparation for and use of PCA

### Part Two: Correlate pollutant emissions with concentrations by grid coordinate.

1. Load relevant sources, ref Table 4: Data Sources**.**



Figure 14: Load Part Two data

1. Filter on 2019 data and extract useful columns from “Emissions by Grid ID” and combine Easting and Northing to make grouping against coordinate easier.

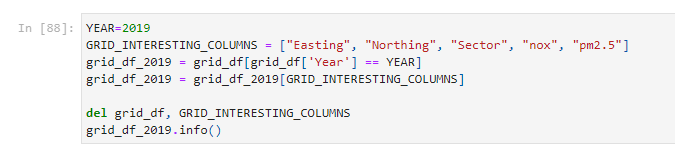


Figure 15: Use column filtering to reduce source data dimensions

1. Combine Easting and Northing to make grouping by coordinate easier



Figure 16: Use astype and concatenation to combine coordinates into single feature

1. Group NOx and pm2.5 by coordinate and calculate the percentage concentration attributable to Road Transport sector, based on emissions.

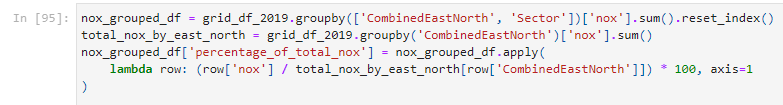


Figure 17: Use sum and lamda to calculate the percentage of total nox for a coordinate



Figure 18: Filter on non-zero values and by road transport

1. Combine the x and y coordinates for concentration data which is synonymous with Easting and Northing.

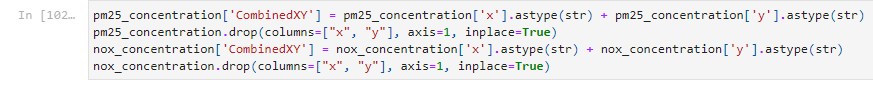


Figure 19: Combine grid coordinates in concentration data

1. Estimate percentage of concentration attributable to road traffic



Figure 20: Use apply to calculate percentage concentration

1. Create LIR models predicting pollutant ug/m3 concentration given tonnes/year emissions.

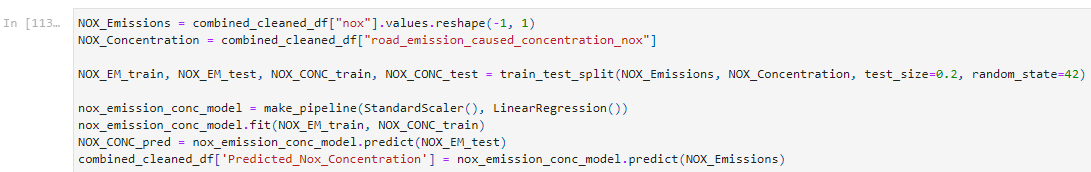


Figure 21: Create LIR models for each pollutant using a pipeline

### Part Three: Create final models predicting concentrations of pollutant from vehicle metrics.

1. Use the LIRs from Part Two to predict concentrations of pollutant based on emissions.



Figure 22: Apply LIR model to feature

1. Create two Logistic Regression Models predicting whether minimum concentrations of pollutant will be surpassed or not.

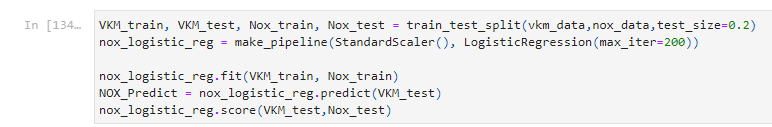


Figure 23: Create LOR for each pollutant based on key features

1. Assess model performance (This was done for all models). See Chapter 5.

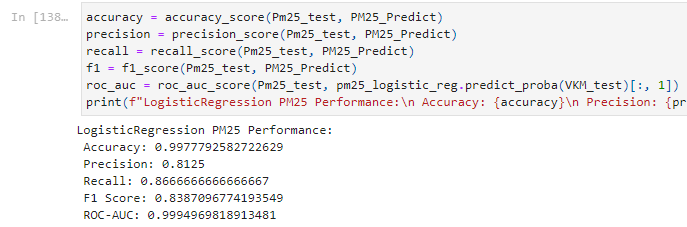


Figure 24: Assess performance

# Findings, Analysis and predictions

The algorithms developed combine data manipulation techniques with machine learning models to analyze and predict air pollution levels from vehicle emissions data.

## Part One

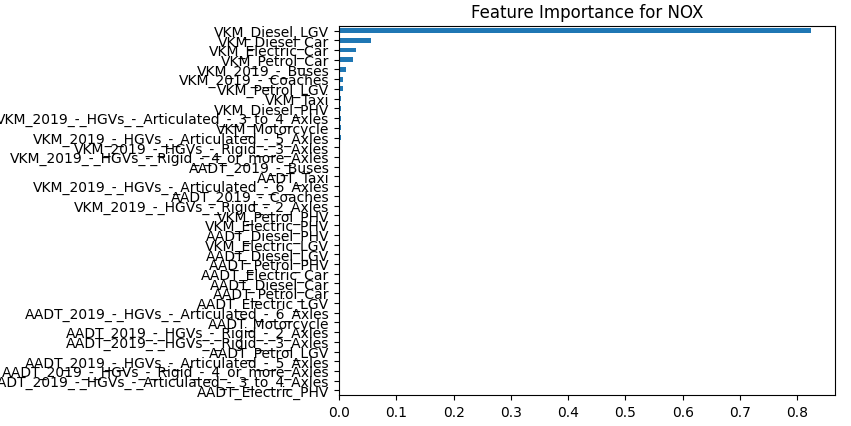


Figure 25: RFR for NOx Emissions

|  |  |  |
| --- | --- | --- |
| **Metric** | **Value** | **Meaning** |
| Mean\_absolute\_error | 0.06 | On average a prediction is incorrect by 0.06 units which is around 7% of the largest value on this scale. |
| Mean\_squared\_error | 0.03 | To get a meaningful value the square root is required, which is very small meaning large errors are rare. |
| R2\_score | 0.94 | This is a good result and indicates that most NOx emissions are explained by the model. |

Table 5: NOx LIR concentration performance metrics

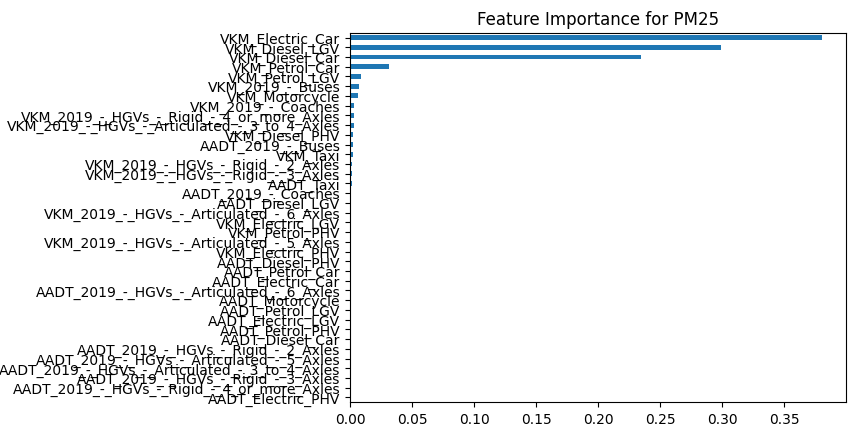


Figure 26: RFR for pm25 Emissions

|  |  |  |
| --- | --- | --- |
| **Metric** | **Value** | **Meaning** |
| Mean\_absolute\_error | 0.00067 | On average a prediction is incorrect by a very small margin |
| Mean\_squared\_error | 1.01e-05 | To get a meaningful value the square root is required, which is very small meaning large errors are rare |
| R2\_score | 0.97 | This is a good result and indicates that the majority of pm25 emissions are explained by the model. |

Table 6: NOx LIR concentration performance metrics

In the above RFR models four features stood out, VKM\_Electric\_Car, VKM\_Diesel\_Car, VKM\_Petrol\_Car and VKM\_Diesel\_LGV.



Figure 27: PCA results for four principal components

## Part Two

Exploring the relationship between emissions and pollutant concentrations based on grid coordinates using LIR models.

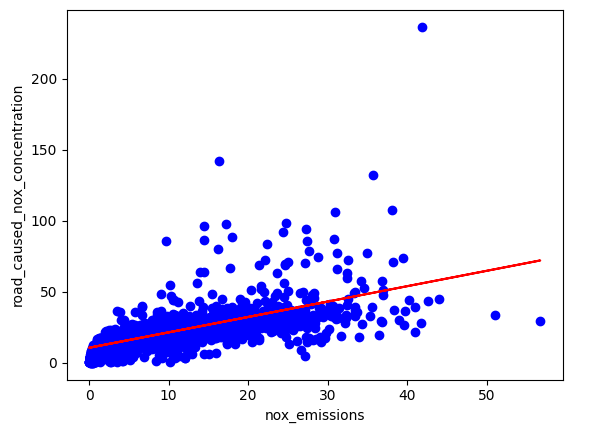


Figure 28: NOx concentration LIR model

|  |  |  |
| --- | --- | --- |
| **Metric** | **Value** | **Meaning** |
| Mean\_absolute\_error | 5.31 | On average a prediction is incorrect by 5.31 units |
| Mean\_squared\_error | 130.59 | Taking the square root of this value we get 11.42, which indicates the outliers are a significant influence |
| R2\_score | 0.4 | Indicates other factors are influencing concentrations, beyond emissions data |

Table 7: NOx LIR concentration performance metrics

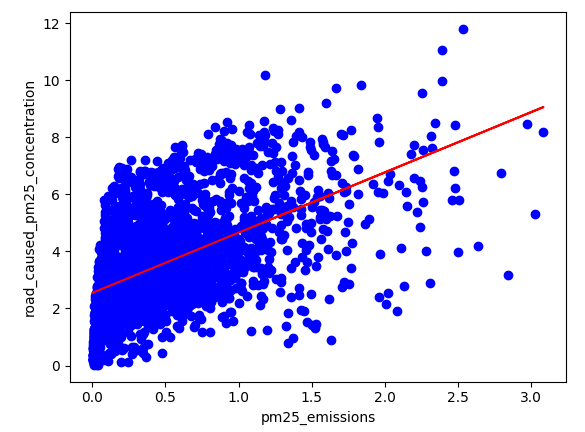


Figure 29: pm25 concentration LIR model

|  |  |  |
| --- | --- | --- |
| **Metric** | **Value** | **Meaning** |
| Mean\_absolute\_error | 1.28 | On average a prediction is incorrect by 1.28 units |
| Mean\_squared\_error | 2.6 | Taking the square root of this value we get 2.6, which indicates the spread visible in the graph is a issue |
| R2\_score | 0.28 | Indicates other factors are influencing concentrations, beyond emissions data |

Table 8: pm25 LIR concentration performance metrics

## Part Three

Synthesizes findings from earlier sections and develops LOR models forecasting whether pollutant concentrations will exceed specified thresholds based on key features, which were established as VKM values.

|  |  |  |
| --- | --- | --- |
| **Metric** | **Value** | **Meaning** |
| Accuracy | 0.93 | 93% labels correctly classified |
| Precision | 0.97 | 97% of positively labelled predictions were positive |
| Recall | 0.95 | 95% of actual positive cases were identified. |
| F1 Score | 0.96 | Balance of Precision and Recall |
| ROC-AUC | 0.98 | 98% accurate in discriminating between positive and negative labels. |

Table 9: NOx LIR concentration performance metrics

|  |  |  |
| --- | --- | --- |
| **Metric** | **Value** | **Meaning** |
| Accuracy | 0.99 | 99% labels correctly classified |
| Precision | 0.81 | 81% of positively labelled predictions were positive |
| Recall | 0.87 | 87% of actual positive cases were identified. |
| F1 Score | 0.83 | Balance of Precision and Recall |
| ROC-AUC | 0.99 | 99% accurate in discriminating between positive and negative labels. |

Table 10: pm25 LIR concentration performance metrics

## Predictions

|  |  |
| --- | --- |
| **Aspect** | **Details** |
| Model Performance | Expected to identify general trends and classify high-risk areas for pollutant concentrations. |
| Linear Regression Limitations | Lower R² values indicate limitations in precise predictions based on emissions. |
| Practical Applications | Valuable for urban planning and traffic management; helps pinpoint areas affected by vehicle emissions. |

Table 11: Predictions

# Future Improvements

|  |  |  |
| --- | --- | --- |
| **Item** | **References** | **Notes** |
| Data Cleaning | Chapter 3. Data Inferences/ Cleaning | Most values were dropped in this case, imputation could yield a better model. |
| Removal of outliers | Chapter 6. Findings, Analysis and predictions | Outliers distort the LIRs used to map emissions to concentrations of pollutant. Methods to address this must understand why they exist and by what mechanism they were created (Han, 2023). |
| Alternative prediction techniques | Chapter 6. Findings, Analysis and predictions | The distribution of the pm25 does trend but is not linear and could benefit from a more granular approach, using what Russel et al (2022) describe as non-parametric models such as Nearest-neighbour models to establish relationships. |
| Consider additional factors in the model. | Chapter 6. Findings, Analysis and predictions | Both LIR models suggest a linear relationship doesn’t not exist between the emissions from road transport and the concentration they create. Further machine learning steps, as mentioned in the above item, could be applied to improve understanding. |

Table 12: Future Improvements

# Conclusion

In conclusion, this project demonstrated the application of machine learning techniques to predicting atmospheric concentrations of two key pollutants. Utilisation of LOR, RFR and PCA models works well to highlight contributors to emissions and predict whether a threshold would be breached. The LIR models used to map emissions in tonnes/year to ug/m3 need to be improved, there is a large spread in the data, and the metrics suggest outliers and other influences into this relationship. So, although the project provides accurate classifications and predictions, future improvements may be required to increase accuracy of emission concentrations. These results provide a solid foundation for government bodies to make informed decisions on environmental policies and enable proactive measures to improve public health and environmental issues.

REFERENCES

Aram, S.A. *et al.,* (2024) *Machine learning-based prediction of air quality index and air quality grade: a comparative analysis*. Int. J. Environ. Sci. Technol*.* 21, pp.1345–1360. https://doi.org/10.1007/s13762-023-05016-2

BELL, J. (2020) *Chapter 2: Planning for Machine Learning*, *MACHINE LEARNING : hands-on for developers and technical professionals.* [Place of publication not identified]: JOHN WILEY,.

Boschetti, A. and Massaron, L. (2018) *Python data science essentials : a practitioner’s guide covering essential data science principles, tools, and techniques / Alberto Boschetti, Luca Massaron.* Third edition. Birmingham, UK: Packt Publishing.

Han, J. (2023) *Data Mining Concepts and Techniques / Jiawei Han [and two others].* Cambridge, MA, United States: Elsevier.

Jiang, X. (2023) *Research on Calibration of Air Quality Data Based on Principal Component Analysis and Multiple Linear Regression Model*. In: Zhang, J., Ruan, R., Bashir, M.J.K. (eds) Environmental Pollution Governance and Ecological Remediation Technology. ICEPG 2022. Environmental Science and Engineering. Springer, Cham. https://doi.org/10.1007/978-3-031-25284-6\_45.

LAEI 2019 Summary (2021). *Summary of London Atmospheric Emissions Inventory 2019 Update* available at <https://data.london.gov.uk/download/london-atmospheric-emissions-inventory--laei--2019/06aab8a6-79a6-40ae-8038-8303ac82a3aa/LAEI%202019%20Summary%20Note%20FINAL.pdf> [Accessed 30/9/2024]

LAEI 2019 Concentrations (2019) *Concentrations of London Atmospheric Emissions Inventory 2019* available at <https://data.london.gov.uk/download/london-atmospheric-emissions-inventory--laei--2019/7ca70bfc-2c16-419d-9dc8-97cc61f757d5/LAEI2019-Air-Quality-Maps-Images-2019.zip>

Russell, S.J. (Stuart J., Norvig, P. and Chang, M.-W. (2022) *Artificial intelligence a modern approach / Stuart J. Russell and Peter Norvig ; contributing writers: Ming-Wei Chang [and 8 others].* Fourth edition. Harlow, United Kingdom: Pearson.

SciKit Learn. (2024). *Model selection and evaluation*. Scikit-Learn.Org. Available at <https://scikit-learn.org/stable/modules/model_evaluation.html#the-scoring-parameter-defining-model-evaluation-rules> [Accessed 30/9/2024]

Suman, N., Rao, A., Ranjitha, (2024) *Forecasting air quality using random forest regression with hyperparameter optimization and LSTM network (RNN)*. AIP Conf. Proc. 2742 (1): 020046. https://doi.org/10.1063/5.0185778.

APPENDICES

Gas Names, sources and properties

|  |  |  |  |
| --- | --- | --- | --- |
| **Acronym** | **Full Name** | **Sources** | **Properties** |
| bap | Benzo[a]pyrene | Soot/smoke | Carcinogen, incomplete combustion of organic materials |
| cd | Cadmium | Industrial: metal refining, fossil fuel, waste incineration | Toxic heavy metal |
| c4h6 | Butadiene | Motor vehicles/Industrial | VOC, respiratory effects, ground-level ozone |
| c6h6 | Benzene | Vehicle exhaust, industrial, cigarette | Carcinogen |
| co | Carbon Monoxide | Incomplete combustion | Colourless/odourless |
| co2 | Carbon Dioxide | Fossil Fuels, industrial, | Major climate change |
| hc | Hydrocarbons | Vehicle emissions/Industrial | Ground level ozone |
| hcl | Hydrogen Chloride | Industrial (coal,chemicals) | Respiratory effects, env harm |
| hg | Mercury | Industrial, waste, industrial | Nervous systems, accumulates |
| n2o | Nitrous Oxide | Agriculture, combustion, industrial | Greenhouse |
| nh3 | Ammonia | Agriculture | Contributes to pm2.5, harms ecosystems |
| nmvoc | Non-Methan Volatile Organic Compounds | Industrial/Vehicle/Solvents | Ground level ozone, smog |
| nox | Nitrogen Oxides (Nitrogen dioxide/nitic acid) | Vehicle engines/power plants | Air pollution, smog and acid rain |
| pb | Lead | Industrial, waste incineration | Severe health issues |
| pcb | Polychlorinated Biphenyls | Indsutrial | Toxic to wildlife and hunans |
| pm10 | Particulate Matter less than 10 micrometers | Industrial, vehicle emissions, wildfires | Respiratory systems |
| pm2.5 | Particulate Matter less than2.5 micrometers | Industrial, vehicle emissions, wildfires | Blood, respiratory |
| so2 | Sulfur Dioxide | Fossil fuels | Major pollutant, respiratory issues |