Bristol Air Quality Analysis

Ritunjai Sharma
Department of Engineering Mathematics
University Of Bristol
Bristol, United Kingdom
dv21835@bristol.ac.uk

Lakshmikanth Loya
Department of Engineering Mathematics
University Of Bristol
Bristol, United Kingdom
ff21018@brsitol.ac.uk

Sourav Singh
Department of Engineering Mathematics
University Of Bristol
Bristol, United Kingdom
vl21348@bristol.ac.uk

Gurdeep Singh Bhambra
Department of Engineering Mathematics
University Of Bristol
Bristol, United Kingdom
ml21460@bristol.ac.uk

***Abstract*—Nitrogen Dioxide (NO_2) and Particulate Matter such as PM_{10} are one of the leading causes of air pollution worldwide. Literature suggests that these are in turn affected by factors such as traffic volumes and weather conditions. Therefore, COVID-19 provides a unique opportunity to help understand the effect of traffic volumes and weather on the NO_2 and PM_{10} concentration levels in Bristol. In order to understand their effect on air quality, several datasets from sources such as Open Data Bristol and Department for Transport pertaining to weather, traffic, air quality etc. were utilized. Tableau visualisations were used to analyze the average annual NO_2 and PM_{10} concentration levels in Bristol in the recent years. Also, to understand the current state as well as the future of electric vehicles in Bristol, FBProphet based time series modeling was used to forecast the expected number of electric and SORN vehicles for the coming 6 years. Furthermore, to quantify the effect of weather and traffic volumes on the NO_2 concentration levels in Bristol, a case study of Bristol Temple way was carried out whereby, machine learning based model such as Random Forest Regressor were used for NO_2 concentration prediction. The final model generalizes well to unseen data and is able to effectively predict the NO_2 concentration levels at a given hour based on weather sensor readings and traffic volumes for that hour.**

I. Introduction

The two major components that impact the air quality are NO_2 and PM_{10}(Particulate Matter). Therefore, it is imperative that we understand their harmful effects before diving deeper into this project. NO_2 is usually formed in the atmosphere when fossils such as coal, oil, gas and diesel are burned at high temperatures. If the NO_2 rises to more than the suggested levels, it harms the vegetation and reduces visibility. It also affects the respiratory system of humans and increases the person's vulnerability to lung diseases like asthma, COPD etc. PM refers to a mixture of solid and liquid particles of different shapes, sizes, and compositions. Man-made PM is mainly attributed to the combustion of fossil fuels in vehicles, industries, and homes. Coarse particles (PM_{10} particles less than 10 microns (μm) in diameter) pose the most significant risk because they can be drawn deeper into the lung. Exposure to high concentrations of PM_{10} can also result in severe health impacts leading to premature deaths. Acid rain and climate change are both exacerbated by pollution caused by PM_{10}. Particulate pollution can alter weather patterns, induce drought, cause global warming, and induce the ocean to acidify.

Literature suggests that traffic can be one of the leading factors affecting the NO_2 concentration levels at a given location. In March 2020, the United Kingdom enacted a lockdown to prevent the spread of COVID-19 among its people. The mobility of people was reduced drastically with the halt in public and private transportations along with a significant reduction in economic activity. These restrictions possibly had a positive influence on air quality, which gives us an excellent opportunity to research and understand more about how traffic volumes and weather can affect air quality.

Furthermore, since electric vehicles emit little to no NOx, they have recently started becoming a part of the effort to help stabilize NO_2 concertation levels in various parts of the world. Therefore, naturally, we wanted perform an analysis of the number of electric vehicles in Bristol to understand if they can be effectively utilized to help reduce NO_2 concentration levels in the coming future. Also, Statutory Off Road Notification (SORN) vehicles are also another indirect means of reducing NO_2 concentration as vehicles that have been registered as off-road and cannot be driven on public roads. Therefore, understanding their current and future state in Bristol is also an important and useful step.

II. Literature Review

Before going ahead with the analysis, we researched about various techniques and models that have been utilized for helping understand air quality and pollution in different parts of the world. Previous works include research on the impact of the Covid-19 lockdown on air pollution. For this, Random Forest Algorithm was applied to capture historical relationships between the pollutants and compare the predictions to actual pollution values after the COVID-19 lockdowns were imposed. The model is trained on data available pre covid, i.e., 2014-19, for four pollutants, namely NO_2, PM_{10}, O_3 (ozone) and O_x (total oxidant) [1]. Also, another interesting paper focussed on the traffic caused near the school zones in Bristol using datasets based on air quality, weather, holidays, and traffic. This indicates that the problem can be geospatial. Machine learning algorithms like Elastic net and Gradient Boosting to predict the pollutant concentration at that zone were implemented [2].

Another paper focussed on an interpretable model for nitrogen oxide (NO_2) levels in Bristol indicating the structural changes and describing their dynamics. Data related to NO_x levels from 5 specific points in the city of Bristol from Jan

2010 to March 2021 were studied by taking daily averages. Using linear regression, a control for meteorological and seasonal effects such as temperature, wind speed and direction, and the day of the week was achieved, followed by analysing the residuals [3]. Studies related to parametric approaches have also be done that take the air quality metrics from several monitoring stations at an hourly rate for two years prior to the covid-lockdown. The parametric study aims to see if the concentrations of NO_2, $PM_{2.5}$, and O_3 in March 2020 are statistically different from those in March 2018 and March 2019 by comparing quantitative variables in two paired samples [4].

III. METHODOLOGY

A. Dataset Collection

Firstly, a major challenge with this project was the collection of relevant datasets as the search space was huge. Therefore, to have a structured approach to this problem, we generated a Logical Model for Data Collection. As shown in Fig. 1, we have divided our collected datasets into four levels. Our target datasets, i.e., "Air quality datasets", are positioned in level 0. Similarly, datasets that directly impact air quality such as "Traffic", "Number of registered vehicles", etc., are in level -1. Again, datasets indirectly impacting air quality have been positioned in level -2. Datasets capturing air quality effects such as "Deaths due to respiratory diseases" have been placed in level 1.

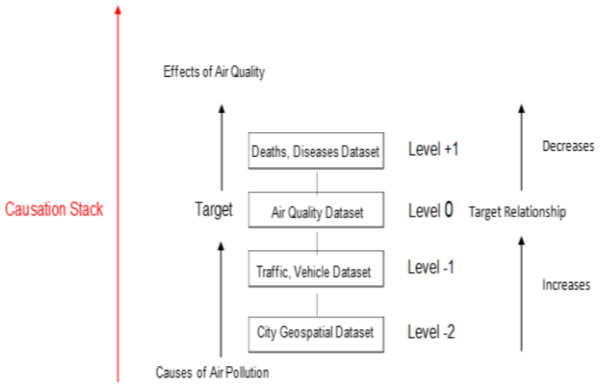

Fig. 1. Logical Model for Data Collection

B. Visual Analytics

To effectively visualise PM_{10} and NO_2 concentration levels in Bristol, Tableau was utilised. Map visual channels and circular marks were used to identify the locations of various PM_{10} and NO_2 sensors in Bristol. Next, the colour channel was used to convey information about the annual average concentration of the sensors, e.g., the red colour represents a yearly average concentration of more than 40 $\mu g/m^3$ and blue represents an annual average concentration of less than 40 $\mu g/m^3$.

C. Time Series Forecasting for the count of Electric Vehicles in Bristol

To forecast the number of electric vehicles expected in the coming years, in Bristol, FBProphet based time series forecasting was used. FBProphet is an open-source forecasting model developed by Facebook's core Data Science team. FBProphet has recently gained a lot of traction in the Data Science industry due to its intuitive implementation. FBProphet is based on an additive model where non-linear trends are fit with yearly, weekly and daily seasonality. The model primarily consists of the sum of three functions of time and an error term: growth g(t), seasonality s(t), holidays h(t), and error term e(t):

$$y(t) = g(t) + s(t) + h(t) + e(t) \quad (1)$$

The growth part g(t) helps model changes that are not periodic. The seasonality part s(t) helps periodic model changes and the holiday component h(t) helps model holidays and special days where there might be a sudden change in values.

D. NO_2 Concentration Prediction

We first created a baseline model for predicting the NO_2 concentration levels at Bristol Temple Way using Linear Regression. Linear Regression is a class of supervised machine learning models that tries to model a relationship between two variables by fitting a linear equation to the observed data. In such a model, one variable acts as the dependent variable and the other as the explanatory variable. A linear regression equation takes the following form:

$$y=\beta_0+\beta_1 x_1 \quad (2)$$

where x_1 is the explanatory variable, y is the dependent variable, β_0 is the intercept, and β_1 is the slope. The goal is to find optimal values of β_0 and β_1 that minimise a given loss function, such as Mean Squared Error, using techniques such as gradient descent.

For the final NO_2 concentration prediction, Random Forest Regressor Model was used. Random Forest Regressor is a tree-based machine learning model that fits several decision trees on multiple subsamples of the training data. The prediction is the average of the individual tree outputs. This helps improve performance as the final prediction outperforms the separate decision tree outputs and helps control overfitting. Fig 2. [5] provides a comprehensive visual

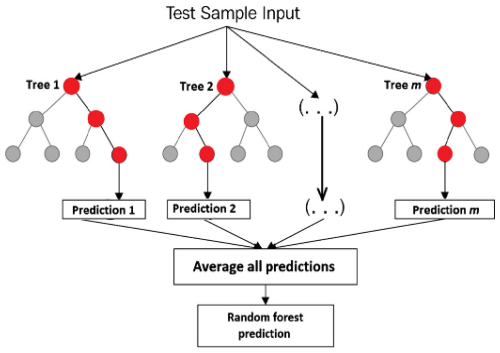

Fig. 2. Working of Random Forest Regressor Model

representation of the working of a Random Forest Regressor Model.

E. SHAP values

SHAP Values (an acronym from SHapley Additive exPlanations) break down a prediction of a given machine learning model to show the impact of each feature on the prediction. Furthermore, SHAP summary plots give us a birds-eye view of feature importance and what is driving it. The horizontal position of the features on the y-axis in the

summary plot represents their respective importance. Features higher up on the y-axis denote a higher impact on the model's predictions. Also, the blue-red colour scale indicates the individual feature values. Therefore, red represents a higher magnitude for the given feature, and blue denotes a low magnitude for the given feature. The vertical position on the x-axis indicates the impact of the feature on the prediction of the model. Values towards the right on the x-axis denote that the feature increases the magnitude of the output variable prediction by the model, and values towards the left denote that the feature decreases the magnitude of the output variable prediction

IV. DATA DESCRIPTION/ PREPARATION

A. Data Description

Air Quality Data: The primary datasets used for air quality analysis were 'Air Quality Data Continuous', 'Air Quality (NO2 diffusion tube) data' and 'Luftdaten Air Quality (PM) data' from Open Data Bristol. 'Air Quality Data Continuous' provides historical and live, continuous, hourly air quality data from Bristol City Council and Defra. NOx readings (containing NO2 data) are from chemiluminescent NOx analysers in this dataset. 'Air Quality (NO_2 diffusion tube) data' provides annual NO_2 concentrations using diffusion tubes and has the significant advantage of multiple sites for sensor readings. However, this dataset is discontinuous with multiple missing values and was only used for visualisation purposes. 'Luftdaten Air Quality (PM) data' consisted of data from low-cost PM (Particulate Matter) Luftdaten sensors in Bristol and was also used only for visualisation purposes.

Traffic Data: The dataset 'Traffic Count Data' from Open Data Bristol was primarily used for traffic insights. This dataset consists of hourly traffic flow data collected using SCOOT traffic counters. These traffic counters are induction loops embedded in roads that help detect when a vehicle passes that road. Additionally, the dataset 'Traffic' for Local Authority of Bristol by the Department for Transport was also initially visualised to get insights into the number of vehicles miles driven per year in Bristol from 2014 to 2020.

Weather Data: The dataset 'Meteorological Data Bristol Lulsgate' from Open Data Bristol was used for weather data. This dataset provides observed meteorological data at a half-hourly interval from Bristol Airport and is supplied by IEM and converted to SI units by Bristol City Council.

Electric Vehicle Data: The dataset 'Licensed plug-in cars, LGVs and quadricycles by local authority United Kingdom' provided jointly by Department for Transport and Driver and Vehicle Licensing Agency was used to get insights into the number of registered electric vehicles in Bristol. This dataset contains battery-electric, plug-in hybrid electric and fuel cell electric vehicles. It provides the count of these vehicles at the end of each quarter for 2014-2021. Additionally, the dataset 'Licensed vehicles by body type and local authority: United Kingdom' by the Department for Transport was also used to get insights into the annual total number of registered vehicles in Bristol for 2014-2020.

Additional Data: In addition to data about air quality, traffic, electric vehicles and weather, other datasets were also utilised to help provide context for analysis. This includes datasets such as 'Quality of Life 2020-21 (citywide trend)' from Open Data Bristol to gauge how concerned people are in Bristol regarding climate change and traffic congestion, as well as 'Vehicles with a Statutory Off-Road Notification by postcode district and body type: United Kingdom' from the Department for Transport to gain insights about the number of registered SORN vehicles in Bristol

B. Data Quality

Many of the datasets collected for the analysis required pre-processing, cleaning, aggregation, handling missing values etc. Some of the issues identified and the subsequent steps taken to address them are given below.

Meta Data: Some of the datasets on Open Data Bristol didn't have metadata for various attributes/features and required manual searching on the internet to understand their meaning. Also, several datasets, mainly from Department for Transport, had metadata embedded in the dataset itself (CSV file) and thus, required manual cleaning to use the data.

Filtering and Aggregation: The majority of the datasets from the Department for Transport were at national, local authority and postcode levels and had separate spreadsheets for each year. Thus, manual filtering and aggregation were required to filter out and aggregate the data for the City of Bristol for the different years, under one table.

Data Inconsistency: Many of the datasets had inconsistent formats, especially for datetime attributes, e.g., the datasets 'Meteorological Data Bristol Lulsgate' and 'Air Quality Data Continuous' had the datetime attributes in the UTC time zone, but the 'Traffic Count Data' had the datetime attribute in BST time zone. Furthermore, converting the BST time zone to UTC for 'Traffic Count Data' proved to be an additional challenge requiring the manipulation of the parameters of the to_datetime function from Pandas in Python.

Lack of Required Information: The information about the proximity of traffic sensors to the various pollutant sensors was not readily available. Therefore, manual visualisation of the traffic and pollutant sensors had to be carried out using a tableau map to identify the traffic sensors in close proximity to the Temple Way pollutant sensor.

Merging Datasets: Merging the traffic, weather and air quality datasets proved one of the biggest challenges. Firstly, these datasets had different granularities, with 'Traffic Count Data' and 'Air Quality Data Continuous' datasets having hourly readings, whereas the dataset 'Meteorological Data Bristol Lulsgate' had half-hourly readings. Therefore, the weather sensor readings had to be aggregated to hourly measures to combine with traffic and air quality datasets. Furthermore, the datetime attribute of the above datasets was split into Year, Month, Day, and Hour to merge them seamlessly. Additionally, pivoting was used to get the traffic data into the required format for merging with the air quality and weather datasets.

Handling Missing Values: The majority of the datasets used for analysis, especially for forecasting, didn't have a large number of missing values. Therefore, the rows with missing values were dropped from these datasets. However, some datasets such as 'Vehicles with a Statutory Off Road Notification by postcode district and body type: United Kingdom' from the Department for Transport had no entry for the 4th quarter of 2021 and thus required imputation by mean based on the values of the other three quarters of 2021.

V. Results and Discussions

A. Analysis of PM_{10} concentration levels in Bristol

We have realised that monitoring of particulate matter in the city of Bristol is quite limited. However, while the primary focus in Bristol has been on achieving compliance with NO_2 limits [6], it is important not to forget that particulate matter, especially $PM_{10,}$ also has harmful effects, above the average value of 40 µg/m^3 annually [7]. Therefore, to get an estimate of the PM_{10} concentration levels in Bristol, 'Luftdaten Air Quality (PM) data' from Open Data Bristol was utilised to plot a map visualisation in Tableau for the annual average PM_{10} concentrations at the monitoring sites that recorded PM_{10} readings continuously for the years 2019-2021. We can see from Fig 3. that the average annual PM_{10} concentration in 2019 was below 40 µg/m^3 for the majority of the monitoring sites, with only two areas exceeding this threshold.

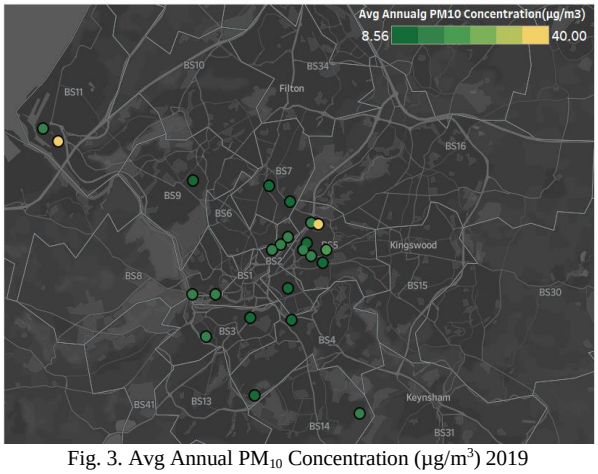

Fig. 3. Avg Annual PM_{10} Concentration (µg/m^3) 2019

Also, we can see from Fig 4, none of the monitoring sites had an average annual PM_{10} concentration of more than 40 µg/m^3 in 2020. Since PM_{10} is emitted by sources such as industrial processes [8] and motor vehicle exhaust [9], the reduction in PM_{10} concentration levels in 2020 can be attributed to lockdowns imposed due to the COVID-19 pandemic leading to reduced traffic and industrial activity. Also, the average annual concentration levels have continued to stay below 40 µg/m^3 in 2021 for all monitoring sites, although at slightly higher levels than in 2020. This is understandable given the easing of lockdown restrictions in 2021. Overall, the above analysis suggests that the PM_{10} concentration levels in Bristol are not a major cause of

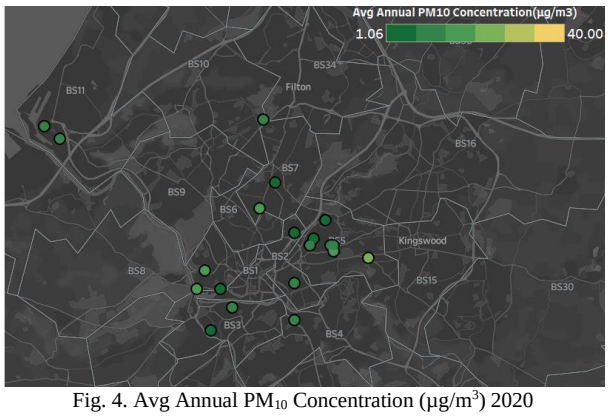

Fig. 4. Avg Annual PM_{10} Concentration (µg/m^3) 2020

concern. However, since no threshold has been identified yet by WHO below which no damage to health is observed due to PM_{10} [10], it is advisable that a close watch is kept on the PM_{10} concertation levels in Bristol and efforts are made to reduce the concentration levels even further.

B. Analysis of NO_2 concentration levels in Bristol

To analyse the concentration levels of NO_2 in Bristol, the dataset 'Air Quality (NO2 diffusion tube) data' from Open Data Bristol was use. This is because the monitoring sensors of this dataset cover a larger proportion of Bristol's area compared to the sensors of the dataset 'Air Quality Data Continuous'. Like PM_{10} concentration level analysis, the average annual NO_2 concentration levels at the monitoring sites were visualised using a Tableau Map Visualization.

We can see from Fig 5. that many monitoring sites reported an average annual NO_2 concentration of more than 40 µg/m3 in 2019 (orange and red dots).

The sensors at Colston Avenue and Parson St. had the highest average annual NO_2 concentration readings in 2019. Thus, serve as potential areas where more work needs to be done to reduce the NO_2 levels.

However, we can see from Fig 6. that NO_2 concentration levels dropped drastically in 2020, with only five monitoring sites reporting an average annual NO_2 concentration of more than 40 µg/m^3. This suggests a strong link between NO_2 concentration levels and traffic since the drop in concentration

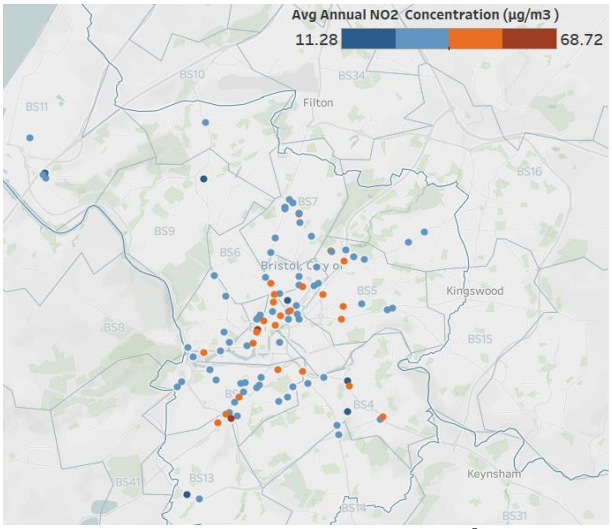

Fig. 5. Avg Annual NO_2 Concentration (µg/m^3) 2019

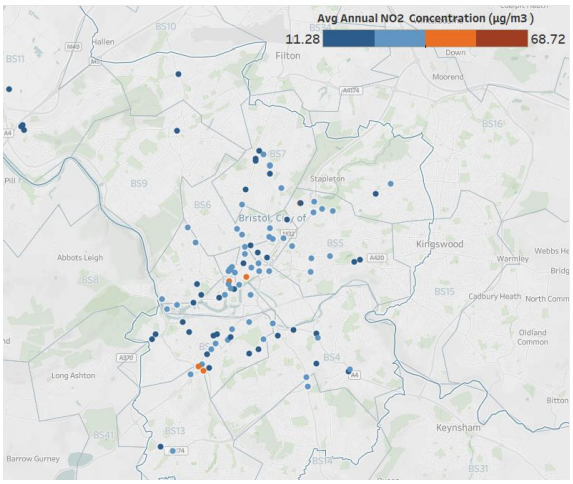

Fig. 6. Avg Annual NO_2 Concentration (µg/m^3) 2020

levels could be possibly attributed to the reduction in the amount of traffic and hence, combustion vehicles due to lockdown restrictions in 2020.

C. *Time Series Forecasting for the number of Electric and SORN Vehicles in Bristol*

As explained earlier, electric vehicles are highly beneficial for the environment, primarily due to their 0 NOx emissions. Thus, the use of electric vehicles is one possible way of curbing the concentration levels of NO_2 in Bristol. Also, SORN vehicles are another indirect way of reducing NO_2 concentration levels. SORN vehicles are primarily vehicles that have been registered as off-road and cannot be driven on public roads. This is especially useful for helping save tax and insurance cost when the owners do not aim to use their cars in the near future. However, this also helps reduce NO_2 concentration in the form of lesser vehicles on the road. To understand the current state of electric and SORN vehicles in Bristol, the dataset 'Licensed plug-in cars, LGVs and quadricycles by the local authority the United Kingdom' and 'Vehicles with a Statutory Off-Road Notification by postcode district and body type: United Kingdom' provided jointly by the Department for Transport and Driver and Vehicle Licensing Agency was used. Additionally, the dataset 'Licensed vehicles by body type and local authority: United Kingdom' by the Department for Transport was also used to get insights into Bristol's total number of registered vehicles each year. We can see from Fig 7. that the number of electric vehicles in Bristol has been on a constant rise since 2018 (yellow line), which is certainly an encouraging sign.

However, the current number of electric vehicles still forms a tiny proportion (2500 electric vehicles compared to 225,000 total registered vehicles), i.e., approximately 1% of Bristol's total number of registered vehicles. To gain insights into how the electric vehicle revolution will possibly pan out in Bristol in the future, we used FBProphet based Time Series Modeling to forecast the expected number of electric vehicles in Bristol in the coming six years. As we can see from the results in Fig 7., the trend (green line), although increasing, is not significant to allow for a substantial proportion of licensed vehicles in Bristol to be electric vehicles in the next six years. Therefore, The Bristol City Council needs to take positive

Fig. 7. Forecast of the number of electric vehicles in Bristol 2022-2028.

steps such as incentivising electric vehicles to help increase the rate of their adoption in the coming years.

Also, FBProphet based time series modeling was utilized to forecast the number of SORN vehicles in Bristol till 2026. The red line represents the current number of SORN vehicles in Bristol. The blue represents the forecasted number of SORN vehicles in Bristol. As we can see from Fig 8., the number of SORN vehicles are expected to rise in the coming years by a considerable amount. This is certainly a positive news since more and more people in Bristol are opting to register their vehicles as off-road when not in immediate use. Therefore, SORN vehicles in Bristol can serve as an alternative way of helping reduce NO_2 concentration levels in the coming years.

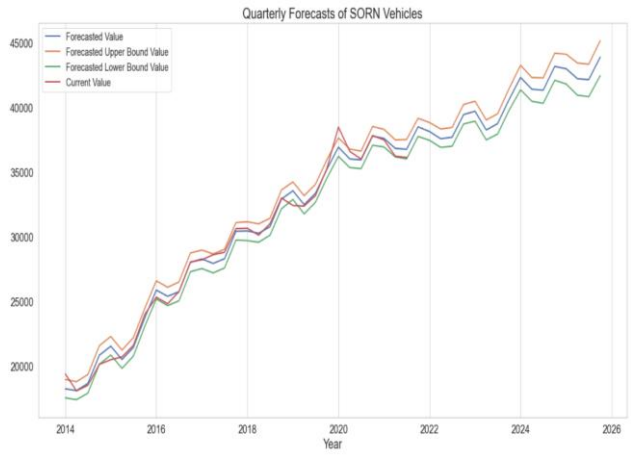

Fig. 8. Forecast of the number of SORN vehicles in Bristol 2022-2026

D. *NO_2 Concentration Prediction, Case Study- Bristol Temple Way*

We have created a machine learning model to quantify the effect of traffic volumes and weather on NO_2 concentration levels in Bristol and to check if there is a link between these features. The model can predict the NO_2 concentration levels at a given hour based on that hour's traffic and weather sensor readings. We have mainly focused on the NO_2 sensor readings from the Bristol Temple Way monitoring site as it is in close proximity to the City Centre and houses one of the most traffic-congested roads in Bristol [11]. Thus, this monitoring site provides a good opportunity to measure the

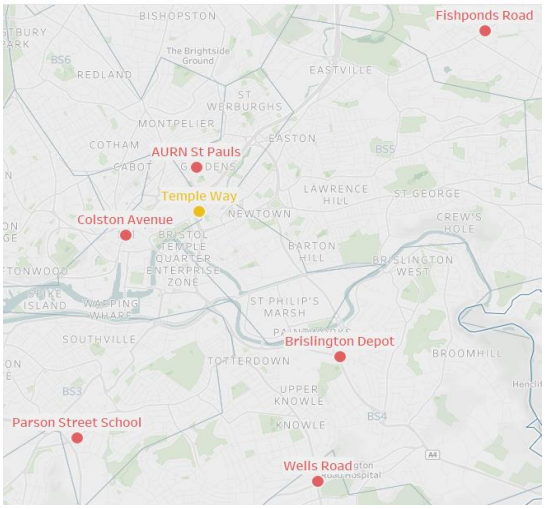

Fig. 9. Pollutant Sensor Locations (Temple Way highlighted)

effects of traffic volumes on the NO_2 concentration levels in Bristol.

For the NO_2 concentration data, the dataset 'Air Quality Data Continuous' from Open Data Bristol was used. We filtered the data to only retain the readings from the sensor at the Temple Way monitoring site (Sensor ID 500), highlighted in Fig 9.

Next, we have used the dataset 'Traffic Count Data' from Open Data Bristol and manually identified the sensors in close proximity to the Temple Way pollutant sensor. The above identified traffic sensors and the Temple Way pollutant sensor are visually represented in Fig 10. The yellow mark denotes the Temple Way pollutant sensor, whereas the red marks are the identified traffic sensors in close proximity.

We have used the dataset 'Meteorological Data Bristol Lulsgate' from Open Data Bristol and aggregated the readings to hourly measures as explained in the data description section of this report for the weather data. The final merged dataset consists of the weather, traffic and NO_2 sensor readings for the years 2020 and 2021, and the readings are hourly in frequency.

Next, we have also performed an exploratory data analysis of the above-merged datasets. As we can see from Fig 11., the EDA firstly shows that the annual average NO_2

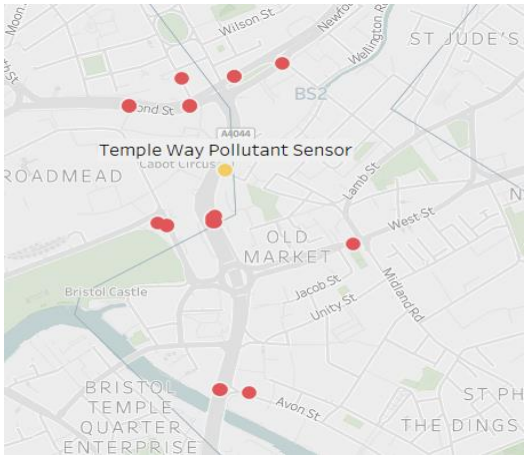

Fig. 10. Identified Traffic Sensors in close proximity to Temple Way Pollutant Sensor

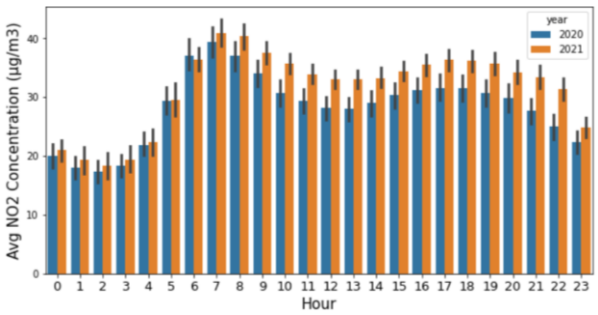

Fig. 11. Average Annual NO_2 Concentration 2020-2021 (Hourly)

concentration at Temple Way was majoritarily lesser in 2020 than in 2021 for each hour of the day. This again suggests that the reduction in traffic volumes in 2020 due to lockdown restrictions possibly helped reduce NO_2 concentrations in Bristol.

To check for a link between weather, traffic and NO_2 sensor readings, we created a correlation matrix of the various features in the merged dataset. We also created a plot for understanding the correlation of the individual features with NO_2 sensor readings. The NO_2 sensor readings seemed to have a weak to moderate correlation with the traffic sensor readings (84, 86…221 denote the traffic sensor IDs). Compared to the rest of the features, the readings from the traffic sensor with IDs 82 and 221 had the highest correlation with the NO_2 readings from the Temple Way sensor (feature no2). This is logical since the traffic sensor with IDs 82 and 221 denote the sensors at Bond Street and Avon Street and are directly situated on the road that is close to the Temple Way sensor, i.e., road A4044. This suggests that traffic volumes indeed impact the NO_2 concentration levels at Bristol Temple Way. However, since the correlation plots only explain the linear relationship between the features, we have further explored the merged dataset.

To quantify the effect of weather and traffic volumes on NO_2 concentration levels, we have created a baseline model using Linear Regression and implemented a Random Forest Regressor Model. These models aim to predict the NO_2 concentration at a given hour based on the weather and traffic sensor readings for that hour. We have split the data into train and test sets, keeping 80% of the data for training and 20% of the data for testing the performance of the models. Furthermore, we have used cross_val_score from the sklearn library to evaluate the performance of the models comprehensively. The metrics used are mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE) and R^2.

The results for the Linear Regression model, i.e., the baseline mode, are tabulated in Table 1. As we can see from the results in Table 1., the model can roughly explain 50% of the variation (R^2 score) in the output variable (NO_2 concentration) and performs reasonably well for a baseline model.

TABLE 1. LINEAR REGRESSION MODEL RESULTS (BASELINE MODEL)

	MAE	MSE	RMSE	R^2
CV Mean Score	10.90	204.33	14.29	0.49
Test Set Score	10.79	201.4	14.19	0.50

However, the model hints toward non-linear relationships in the data, that cannot be modelled using a linear model such as Linear Regression. Therefore, we have created a Random Forest Regressor model next as it is a non-linear regression method and can perform well on heterogeneous data [12]. As we can see from the results of Table 2., the Random Forest Regressor model outperforms the Linear Regression model in terms of cross-validation scores and test set scores. The model can explain roughly 75% of the variation (R^2 Score) in the output variable (NO_2 Concentration), which shows that the model performs well given the use of raw, real-world datasets and loss of data due to dropping of rows with missing values. Also, the model generalises well to unseen data since the test set scores are comparable to the cross-validation scores.

TABLE 2. RANDOM FOREST REGRESSOR MODEL RESULTS

	MAE	MSE	RMSE	R^2
CV Mean Score	7.23	103.80	10.19	0.74
Test Set Score	6.98	99.29	9.96	0.75

To further interpret the results of the Random Forest Regressor model, SHAP values were used to understand how each feature in the merged dataset impacts the prediction of NO_2 concentration. We can see from Fig 12. that wind speed and direction had the most significant impact on the model's predictions.

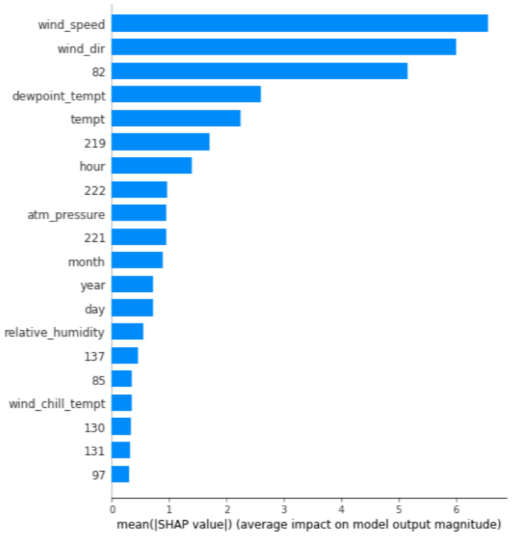

Fig. 12. Feature importance scores for Random Forest

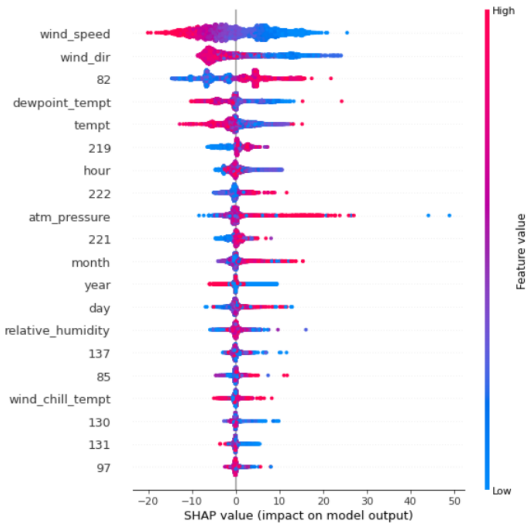

Fig. 13 Impact of various features on Random Forest Regressor model predictions based on SHAP values

This is understandable since past literature has shown a link between NO_2 concentration levels, wind speed [13], and wind direction [14]. However, the next feature that had the most significant impact on the model's predictions was the sensor readings from the traffic sensor with ID 82, as seen in the Linear Regression model's case.

In order to further understand the impact of each feature on the NO_2 concentration levels, a summary plot of the SHAP values was generated, as given in Fig 13. We can see that high wind speeds lead to lower NO_2 concentration levels. Also, higher traffic volumes lead to higher NO_2 concentration predictions, up to 20 $\mu g/m^3$ in magnitude.

The above results show that there is indeed a strong link between the traffic volumes and NO_2 concentration levels in Bristol Temple Way, with higher traffic volumes leading to higher NO_2 concentration levels.

VI. Further Work and Improvement

A. *Expanding on the NO_2 Prediction Model*

Although the developed model provides reasonable predictions for the NO_2 concentration levels at Bristol Temple Way, we would like to cater to the other pollutant monitoring sites in Bristol as well, e.g., the sensors at Fishponds Road and Colston Avenue. This would also require identifying the traffic sensors close to these respective pollutant sensors.

B. *Productionising the Machine Learning Model*

We want to productionise the model to make it more accessible and usable for the non-technical demographic. Also, the current model relies heavily on manually feeding the traffic and weather sensor readings to make NO_2 predictions. We would like to automate this process and create a Web App based on the above model. The model will be able to ingest real-time traffic and weather sensor readings and predict the NO_2 concentration at the location chosen by the user for the current hour. The model could then also predict the NO_2 concentration levels for the next X hours based on expected traffic volumes and weather forecasts for

the chosen location. Furthermore, the application could also be used by policymakers to understand how reducing the traffic volumes to particular levels affect the NO_2 concentration levels in Bristol.

VII. CONCLUSION

In this work, we have first explored the concentration levels of PM_{10} and NO_2 in Bristol in recent years. Our analysis suggests that the concentration levels of PM_{10} in Bristol are not a cause of major concern at the moment. However, we recommend keeping a close eye on the concentration levels of PM_{10} and making efforts to reduce the concentration further.

Next, NO_2 concentration levels were analysed and we found that the average annual NO_2 concentration levels were less in 2020 than in 2021 for most of the monitoring sites. This suggests a strong link between NO_2 concentration levels and traffic, as the drop in concentration levels could be possibly attributed to the reduction in traffic volumes due to lockdown restrictions in 2020.

We have also forecasted the number of electric vehicles in Bristol for the coming six years and concluded that the current trend is not significant enough for electric vehicles to replace a large enough proportion of total registered vehicles in Bristol to make a noticeable difference. We have also foracsted the number of SORN vheciles expected in the future, in Bristol. The trend shows that the number of SORN vehciles are exptected to rise by a signifiacnt amoun in the next 4 years. This is an encouraging sign since more and more people in Bristol are opting to register their vehicles as off-road when not in immideate use.

Finally, to understand the interplay of weather and traffic volumes with NO_2 concentration levels, we developed machine learning models to predict the NO_2 concentrations at a given hour based on the traffic and weather sensor readings for that hour. The pollutant, traffic and weather data for Bristol Temple Way were utilised for this task. The Random Forest Regressor based model showed good generalisation and performance for the above data and made reasonable predictions for the NO_2 concentration levels. Furthermore, SHAP values were also used to interpret the Random Forest Regressor model results. Therefore, we have been able to effectively quantify the effect of traffic volumes and weather on the NO_2 concentration levels in Bristol. Our model analysis suggests a strong link between the traffic volumes and NO_2 concentration levels in Bristol, with higher traffic volumes leading to higher NO_2 concentration levels. Our model can thus, be used by policymakers to understand how redcuing traffic volumes to particular target levels affect the NO_2 concentration levels in Bristol. The model can also be used to effectively predict the NO_2 concentration levels in Bristol at a given hour, based on traffic and weather sensor readings for that hour.

REFERENCES

[1] Lovrić, M., Pavlović, K., Vuković, M., Grange, S.K., Haberl, M. and Kern, R. (2020). Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. Environmental pollution, p.115900. doi:10.1016/j.envpol.2020.115900.

[2] https://www.turing.ac.uk/sites/default/files/2020-04/data_study_group_network_final_report_-_bristol_city_council.pdf

[3] compass.blogs.bristol.ac.uk. (n.d.). Student Perspectives: Change in the air: Tackling Bristol's nitrogen oxide problem – Compass Blog. [online] Available at: https://compass.blogs.bristol.ac.uk/2021/04/13/change-in-the-air/ [Accessed 4 May 2022].

[4] Zambrano-Monserrate, M.A. and Ruano, M.A. (2020). Has air quality improved in Ecuador during the COVID-19 pandemic? A parametric analysis. Air Quality, Atmosphere & Health, 13(8), pp.929–938. doi:10.1007/s11869-020-00866-y.

[5] Afzal, A., Aabid, A., Khan, A., Khan, S.A., Rajak, U., Verma, T.N. and Kumar, R., 2020. Response surface analysis, clustering, and random forest regression of pressure in suddenly expanded high-speed aerodynamic flows. Aerospace Science and Technology, 107, p.106318.

[6] Air quality annual status report 2021, Bristol

[7] Anderson, J.O., Thundiyil, J.G. and Stolbach, A., 2012. Clearing the air: a review of the effects of particulate matter air pollution on human health. Journal of medical toxicology, 8(2), pp.166-175.

[8] Puliafito, S.E., Castro, F. and Allende, D., 2011. Air-quality impact of PM10 emission in urban centres. International Journal of Environment and Pollution, 46(3-4), pp.127-143.

[9] Qu, H., Lu, X., Liu, L. and Ye, Y., 2019. Effects of traffic and urban parks on PM10 and PM2. 5 mass concentrations. Energy Sources, Part A: Recovery, Utilisation, and Environmental Effects, pp.1-13.

[10] https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

[11] https://www.bristolpost.co.uk/news/bristol-news/revealed-worst-bristol-roads-congestion-3796491

[12] Lovrić, M., Pavlović, K., Vuković, M., Grange, S.K., Haberl, M. and Kern, R., 2021. Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. Environmental pollution, 274, p.115900.

[13] Grundström, M., Hak, C., Chen, D., Hallquist, M. and Pleijel, H., 2015. Variation and co-variation of PM10, particle number concentration, NOx and NO2 in the urban air–Relationships with wind speed, vertical temperature gradient and weather type. Atmospheric Environment, 120, pp.317-327.

[14] Donnelly, A., Misstear, B. and Broderick, B., 2011. Application of nonparametric regression methods to study the relationship between NO2 concentrations and local wind direction and speed at background sites. Science of the Total Environment, 409(6), pp.1134-1144.

Link to Project GitHub Repo:
https://github.com/InvalidDuck/G32-Butterfly-Data