A Spatial Feature Engineering Algorithm for Creating Air Pollution Health Datasets

Raja Sher Afgun Usmani [1], Thulasyammal Ramiah Pillai [1], Ibrahim Abaker Targio Hashem [1], NZ Jhanjhi [1], and Anum Saeed [1]

[1]Affiliation not available

November 8, 2023

Abstract

Air pollution is one of the significant causes of mortality and morbidity every year. In recent years, many researchers have focused their attention on the associations of air pollution and health. These studies used two types of data in their studies, i.e., air pollution data and health data. Feature engineering is used to create and optimize air quality and health features. In order to merge these datasets residential address, community/county/block/city and hospital/school address are used. Using residence address or any location becomes a spatial problem when the Air Quality Monitoring (AQM) stations are concentrated in urban areas within the regions and an overlap in the AQM stations in urban areas coverage area, which raises the question that how to associate the patients with the relevant AQM station. Also, in most of the studies the distance of patients to the AQM stations is also not taken into account. In this study, we propose a four-part spatial feature engineering algorithm to find the coordinates for health data, calculate distances with AQM stations and associate health records to the nearest AQM station. Hence, removing the limitations of current air pollution health datasets. The proposed algorithm is applied as a case study in Klang Valley, Malaysia. The results show that the proposed algorithm can generate air pollution health dataset efficiently and the algorithm also provides the radius facility to exclude the patients who are situated far away from the stations.

A Spatial Feature Engineering Algorithm for Creating Air Pollution Health Datasets

Raja Sher Afgun Usmani
School of Computer Science and Engineering
Taylor's University
Subang Jaya, Selangor, Malaysia
rajasherafgunusmani@sd.taylors.edu.my

Thulasyammal Ramiah Pillai
School of Computer Science and Engineering
Taylor's University
Subang Jaya, Selangor, Malaysia
thulasyammal.ramiahpillai@taylors.edu.my

Ibrahim Abaker Targio Hashem
Faculty of Computer Science and Information Technology
University of Malaya
Kuala Lumpur, Malaysia
ibrahimabaker@um.edu.my

NZ Jhanjhi
School of Computer Science and Engineering
Taylor's University
Subang Jaya, Selangor, Malaysia
noorzaman.jhanjhi@taylors.edu.my

Anum Saeed
Faculty of Electrical & Computer Engineering
Center for Advance Studies in Engineering
Islamabad, Pakistan
anum.se87@gmail.com

***Abstract*—Air pollution is one of the significant causes of mortality and morbidity every year. In recent years, many researchers have focused their attention on the associations of air pollution and health. These studies used two types of data in their studies, i.e., air pollution data and health data. Feature engineering is used to create and optimize air quality and health features. In order to merge these datasets residential address, community/county/block/city and hospital/school address are used. Using residence address or any location becomes a spatial problem when the Air Quality Monitoring (AQM) stations are concentrated in urban areas within the regions and an overlap in the AQM stations in urban areas coverage area, which raises the question that how to associate the patients with the relevant AQM station. Also, in most of the studies the distance of patients to the AQM stations is also not taken into account. In this study, we propose a four-part spatial feature engineering algorithm to find the coordinates for health data, calculate distances with AQM stations and associate health records to the nearest AQM station. Hence, removing the limitations of current air pollution health datasets. The proposed algorithm is applied as a case study in Klang Valley, Malaysia. The results show that the proposed algorithm can generate air pollution health dataset efficiently and the algorithm also provides the radius facility to exclude the patients who are situated far away from the stations.**

Keywords— air pollution, feature engineering, health, air quality, hospitalization, mortality, morbidity, data science, spatial data

I. INTRODUCTION

Ambient air pollution is globally regarded as a serious public health concern. According to WHO, 4.2 million deaths were attributed to ambient air pollution in 2016 [1]. Cardiovascular and respiratory diseases are the leading cause of deaths due to ambient air pollution [2]. Air pollution is also linked to other diseases like cancer and asthma [2], [3]. It is one of the major concerns in low and middle-income countries, where the air pollution levels are still on the rise, as well as high-income countries, where air pollution levels have decreased relatively in past decades [2]. Recently, the focus of researchers has been on air pollution and its health impacts. The researchers are looking into the techniques and policies that can minimize these health impacts.

The studies in this domain use different types of datasets, but in general, two datasets are necessary for conducting the health impact of air pollution study. The first dataset needed is air pollution dataset, and secondly, the health dataset is needed. Usually, the air pollution data is captured through the air quality monitoring (AQM) stations distributed across regions. The health data for the same region/state/county is obtained to associate with the air pollution data [4]. Residence address is one of the main parameters used to associate the person to the region [5].

Feature engineering is used by researchers to extract meaningful information from these datasets using the domain knowledge [6]. Feature engineering is also used to create and optimize air pollution and health features. Feature engineering provides the researchers with the flexibility of choice means, less complex algorithms will provide good accuracy with good features [7]. Using residence address or any location becomes a spatial problem when the AQM stations are concentrated in urban areas within the regions [8]. As the region/state/county is used to merge these datasets, there is an obvious overlap in the AQM stations in urban areas, which raises the question that how to associate the patients with the relevant AQM station.

To the best knowledge of the authors, there are very few studies that predict the health impact of air pollution and most studies in this domain focus on impact and associations of ambient air pollution and health. The most significant hurdle in conducting a study in this domain is the availability of datasets and lack of any appropriate spatial feature engineering method that can combine these datasets appropriately and provide the combined dataset in a usable format. Keeping these conditions in mind, we propose a spatial feature engineering algorithm to combine air pollution and health datasets. The proposed algorithm will generate a new dataset in a simple, usable format. The main contributions of this research work are as below:

- Algorithm to find coordinates of health records automatically.
- Algorithm to calculate the distance between AQM stations and health records.
- Algorithm to associate the health record with the closest station based on the calculated distance. Hence, eliminating the limitation of concentrated AQM stations

in urban areas, with the additional radius facility to exclude spatially distant records.

The rest of the paper is organized as follows: In Section II, First, we talk about the background and motivation to study air pollution and its impacts. At the end of the section, we talked about the motivation to develop our feature engineering algorithm. Section III presents the details of the case study, the datasets we are using and the feature engineering algorithm in detail. Section IV includes the results and discussions. The conclusions are presented in Section V, and lastly, future work is presented in Section VI.

II. BACKGROUND AND MOTIVATION

Air pollution is one of the major causes of death every year. According to yearly mortality statistics by the World Health Organization (WHO), tobacco use causes 7 million deaths; AIDS causes 1.2 million deaths; tuberculosis causes 1.1 million deaths, and malaria causes 0.7 million deaths [1]. In the same year, there were 6.4 million deaths attributed to air pollution worldwide, with 4.2 million deaths due to ambient air pollution and 2.8 million attributed to indoor air pollution. If not controlled aggressively, the projection of deaths from ambient air pollution in 2060 is 6-9 million. Additionally, ambient air pollution is categorized as an important risk factor for neurodegenerative diseases in adults and neurodevelopmental disorders in children [1].

In recent years, many researchers have focused their attention on the associations of air pollution and health. Various medical specialty reports have consistently highlighted a relation between particulate air pollution and not the only aggravation of cardiovascular and respiratory illnesses but also a drastically rising number of deaths among older people. These associations are reportedly unlikely to be elaborated by any confounder; they mainly represent cause and effect. However, the nature of the urban particulate cloud is where the explanation lies. It may contain up to 100000 nanometer-sized particles per mL, in what may be a gravimetric concentration of only 100-200 µg/m3 of pollutant [9]. It also suggests that such ultra-fine particles can provoke alveolar inflammation, with the release of mediators capable of causing aggravation of lung disease and increasing blood clots formation. This explains the increasing number of cardiovascular deaths associated with urban pollution.

Since sulphur dioxide has generally decreased in concentration globally [10] and in Asia [11], attention has shifted to ozone, nitrogen dioxide, and particulates. From a global perspective, millions of people living in rural areas in developing countries consume biomass fuels in concentrations that are, by magnitude, higher than recently observed in developed countries. Over 2 million children die due to acute respiratory infections that are caused by these exposures [12].

80% of premature deaths are due to stroke and Coronary heart disease (CHD), also attributed to air pollution, making it the most common cause of premature death from air pollution in Europe [13]. The mechanisms by which cardiovascular disease is caused by air pollution are observed to be identical to those causing respiratory disease.

Cardiovascular diseases account for 60—80% of air pollution-related deaths [14]. Hence, the contribution of air pollution to heart disease is presented in the 2016 report of air quality in Europe [15]. The report covered 41 European countries. The report indicates that air pollution caused 444,000 premature deaths from stroke and coronary heart disease. According to the 2013 figures, air pollution caused 416,000 premature deaths in the European Union [15].

Table I presents a comprehensive comparison of health-related studies around the world. We looked into the major diseases and conditions that are linked with air pollution and the air pollutants that are associated with these diseases. These studies used two types of data in their studies, i.e., air pollution data and health data. In order to merge these datasets residential address [3], [4], [16]–[26], post/zip code [27]–[30], community/county/block/city [31]–[43], hospital/school address [26], [44]–[46] was used. As discussed in Section I, using these parameters for the association of air pollution and health datasets is inefficient as there is a concentration of AQM stations in urban regions. Researchers rely on average in order to use community/county/block/city parameter with multiple AQM stations. In most of the studies, the distance of patients to the AQM stations is also not taken into account. To cater to this problem, we propose a spatial feature engineering algorithm, which automatically finds the appropriate AQM station and associate the patient with it.

TABLE I. AIR POLLUTION (AP) AND THE ADVERSE HEALTH EFFECTS

Disease / Condition	***NOx***	***NO***	***NO2***	***CO***	***PM***	***PM10***	***PM2.5***	***O3***	***SO2***
Cancer			[3], [19], [27]				[3], [27], [32]		
Acute exposure					[47]		[48]		
Asthma	[22]		[23], [35]	[22]	[49]	[36]			
Asthma ER Visits								[45], [46]	
Asthma hospitalizations			[37]	[37]	[25], [37]				[37]
Birth outcomes		[28]	[28]			[28]	[28]		
Chronic exposure							[48]		
Cardiovascular		[20]	[20], [26], [44]	[38]	[39]	[17]	[16], [40], [50]	[44], [51]	[44], [52]
Mortality	[53]	[53]	[21], [34], [53], [54]	[38]			[29]	[54]	
Oxidative stress			[26]						
Blood pressure							[55]		

Breast cancer			[56]						
Prostate cancer			[56]						
Respiratory			[30], [34]		[41], [42], [57]		[58]		
Respiratory hospital admissions			[42], [44]					[31], [44]	[33], [44]
Cardiopulmonary					[47]				
Morbidity							[29]		
Preeclampsia	[24]						[24]		
Preterm birth	[24]						[24]		
Pregnant women							[59]		
Premature death	[60]								
Rhinitis		[43]	[18], [61]	[23]		[43]		[43]	

III. METHODOLOGY

In this section, we explain the datasets used in this study and the spatial feature engineering algorithm in detail. Although the case study is based in Klang Valley, Malaysia, the algorithm is not case-study specific. The proposed algorithm can be used to associate air quality and health dataset in any location around the world. The MapQuest API used to find coordinates of address is also freely available with data for all countries.

A. Data

This study is conducted using the Klang valley, Malaysia, as a case study. Klang Valley is an urban region in Malaysia that is centered in Kuala Lumpur and includes its adjoining cities and towns in the state of Selangor. We used two datasets in this case study, i.e., air pollution dataset from the air quality monitoring stations, provided by Department of Environment (DOE), Malaysia and hospitalizations dataset, provided by Ministry of Health (MOH), Malaysia. The details of the datasets are provided below.

1) Air pollution dataset

The air quality monitoring in Malaysia was carried out through a private company known as Alam Sekitar Malaysia Sdn Bhd (ASMA) until 2017. ASMA was appointed by the DOE and the Malaysian Meteorological Department (METMalaysia), and it was responsible for collecting, processing, analyzing, and distributing air pollutant measurements. There are 66 air quality monitoring stations across Malaysia, 14 Manual sampling (High Volume Sampler) stations were operated by METMalaysia, and ASMA operated 52 continuous air-quality monitoring (CAQM) stations [62]. The DOE has increased the number of CAQM station to 68 [63]. In this study, we are using the data from eight CAQM stations in Klang Valley. The location-wise details of these CAQM stations are provided in Figure 1.

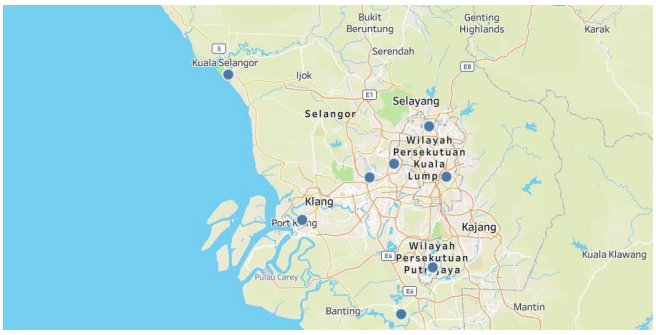

FIGURE 1: AIR QUALITY MONITORING STATIONS IN KLANG VALLEY, MALAYSIA

The air pollution dataset contained daily readings for the major air pollutants ,i.e., PM10, CO, O3, NOx, NO2, NO and SO2,. The air pollution dataset contained data for 11 years, i.e., 2006-2016. This dataset is generated by using the novel feature engineering algorithm for air pollution datasets [64]. The statistics of the dataset are provided in the Table II.

TABLE II. DESCRIPTIVE STATISTICS - AIR POLLUTION DATASET (2011-2016)

Pollutant	Mean	Median	Mode	Min	Max
PM_{10}	51.15864	45.70833	40.375	10.70588	580.875
CO	0.861812	0.790435	0.728696	0.04	5.967619
O_3	0.019124	0.018304	0.019826	0	0.0802
NO_x	0.035931	0.032304	0.031	0.000235	0.154625
NO_2	0.019705	0.018826	0.017	0.000058	0.065619
NO	0.016225	0.013	0.008	0	0.112913
SO_2	0.0031	0.002739	0.002	0	0.034154

2) Hospitalizations dataset

The hospitalizations dataset is provided by MOH, Malaysia. The dataset spans the same duration as air pollution dataset, i.e., 2006-2016. The hospitalizations dataset contains the daily cardio-respiratory hospitalization records of patients in Klang valley. Table III lists down the columns in the dataset.

TABLE III. PARAMETERS IN HOSPITALIZATIONS DATASET

#	Parameter	#	Parameter
1	Gender	6	Post code
2	Age	7	State
3	Ethnicity	8	Admit Date
4	Citizenship	9	Discharge Date
5	Address	10	Diagnosis

B. The algorithm

The algorithm proposed in this study is a four-part algorithm. The first part of the algorithm is responsible for finding coordinates of a patient's using its address. For this purpose, we utilize the MapQuest free online web mapping service [65]. Algorithm 1 accepts all the hospitalization records and creates a MapQuest web request to find the coordinates for the address.

Algorithm 1: Find Coordinates

Data: Hospitalization Records
Result: Coordinates for hospitalization records
Step 1: Function *FindCoordinates*
Step 2: allRecords = Load all hospitalization records
Step 3: **foreach** record in allRecords **do**
Step 4: **if** record does not have proper address **then**
Step 5: continue
Step 6: **end**
Step 7: webrequest = Create MapQuest webrequest
Step 8: completeAddress = record.Address + record.postcode
Step 9: webrequest.Address = completeAddress
Step 10: jsonResult = webrequest.getResponse()
Step 11: record.Latitude = jsonResult.Latitude
Step 12: record.Longitude = jsonResult.Longitude
Step 13: Save record
Step 14: **end**
Step 15: end

The second part of the algorithm finds the distance in meters between two sets of coordinates. Algorithm 2 is derived from the System.Device.Location namespace from the .Net Framework [66]. The Power function returns a specified power of the number, the Cos function returns the cosine value of the given angle, the Sin function returns the sine value of the given angle, the Atan2 function returns the angle whose tangent is the quotient of two given numbers, and the Sqrt function returns the square root value of a given number.

Algorithm 2: Calculate distance in meters

Data: Two sets of coordinates
Result: Distance in meters between the provided coordinates
Step 1: Function *CalculateDistanceInMeters (lat1, long1, lat2, long2)*
Step 2: PI=3.141592653589793
Step 3: d1 = lat1*(PI/180.0)
Step 4: num1 = long1*(PI/180.0)
Step 5: d2 = lat2*(PI/180.0)
Step 6: num2 = long2*(PI/180.0)
Step 7: d3 = Power(Sin((d2 - d1) / 2.0), 2.0) + Cos(d1) * Cos(d2) * Power(Sin(num2 / 2.0), 2.0)
Step 8: return 6376500.0 * (2.0 * Atan2(Sqrt(d3), Sqrt(1.0 - d3)));
Step 9: end

The third part of the algorithm is responsible for finding the distance of patients from the air quality monitoring stations. Algorithm 3 reads all the hospitalization records and stations. The most important feature of this part of the algorithm is the radius feature. A researcher can specify the maximum distance between the AQM station and the patient, making sure that irrelevant patients are not associated with the dataset. It iterates through all the hospitalization records and use the algorithm 2 to find the distance in meters between the record and the stations. In the end, it picks the smallest distance and associates the record with the station with the smallest distance.

Algorithm 3: Find Distance

Data: Hospitalization records, station records
Result: Distance of hospitalization records from stations
Step 1: Function *FindDistance*
Step 2: allRecords = Load all hospitalization records
Step 3: allStations = Load all stations
Step 4: radius = Set maximum distance from station
Step 5: **foreach** record in allRecords **do**
Step 6: allStationDistances = reset allStationDistances
Step 7: **foreach** station in allStations **do**
Step 8: distance = CalculateDistanceInMeters (record.Latitude,record.Longitude, station.Latitude, station.Longitude)
Step 9: Add distance,station to allStationDistances
Step 10: **end**
Step 11: smDistSt = Get station with smallest distance from allStationDistances
Step 12: **if** smDistSt.distance <= radius **then**
Step 13: record.station = smallestDistanceStation
Step 14: Save record
Step 15: **end**
Step 16: **end**
Step 17: end

The last part of the algorithm generates new dataset by combining the two datasets. Algorithm 4 reads all stations, air pollution datasets and updated hospitalizations dataset. It iterates over the stations and gets all the dates in the datasets for that station. In the next step, it iterates through these dates and combines the air pollution reading of that date with the number of hospitalizations on the same date for the station. In the end, it saves the new record.

Algorithm 4: Associate Datasets

Data: Hospitalization records, Air pollution records
Result: Air pollution health dataset
Step 1: Function *AssociateDatasets*
Step 2: allHospRecords = Load all hospitalization record
Step 3: allApRecords = Load all air pollution records
Step 4: allStations = Load all stations
Step 5: **foreach** station in allStations **do**
Step 6: allDates = Get all dates from datasets for the station
Step 7: **foreach** date in allDates **do**
Step 8: apRecord = Get apRecord on that datefor the station from allApRecords
Step 9: hospCountRecord = Get hospitalizations count on that date for the station from allHospRecords
Step 10: aphealthRecord = create single row object with date,apRecord,hospCountRecord
Step 11: Save apheathRecord
Step 12: **end**
Step 13: **end**
Step 14: end

IV. RESULTS & DISCUSSIONS

The results indicate that the proposed algorithm generates air pollution health dataset efficiently. In the third part of the algorithm, the radius parameter in the algorithm plays an integral part as inclusion criteria. It has a significant impact on the number of hospitalizations included in the generated dataset.

Table IV demonstrates the included hospitalization records with a varying radius of stations. It is clearly visible that the radius parameter of the proposed algorithm has a substantial impact on the new dataset, with excluded records dropping from 341,234 to 150,171 with a change of radius from 5,000 meters to 10,000. The trend of reduction in excluded records continues when the radius value is increased.

TABLE IV. INCLUDED HOSPITALIZATIONS RECORDS WITH VARYING RADIUS

Radius	Station	Records
5,000 meters	Raja Zarina	43,484
	Petaling Jaya	14,907
	Shah Alam	3,824
	Sains Kuala Selangor	1,804
	Putrajaya	1,685
	Cheras, Kuala Lumpur	32,498
	Batu Muda, Kuala Lumpur	78,853
	Bukit Changgang	364
	Excluded Records	**341,234**
10,000 meters	Raja Zarina	94,932
	Petaling Jaya	21,349
	Shah Alam	27,663
	Sains Kuala Selangor	3,690
	Putrajaya	15,026
	Cheras, Kuala Lumpur	86,641
	Batu Muda, Kuala Lumpur	117,884
	Bukit Changgang	1,297
	Excluded Records	**150,171**
20,000 meters	Raja Zarina	110,786
	Petaling Jaya	22,793
	Shah Alam	45,259
	Sains Kuala Selangor	7,214
	Putrajaya	46,240
	Cheras, Kuala Lumpur	98,669
	Batu Muda, Kuala Lumpur	131,095
	Bukit Changgang	7,379
	Excluded Records	**49,218**
	Total Records	**518,653**

Table V shows the sample data from the generated dataset. The generated dataset is in a simple row-format. The sample dataset contains the air pollutants readings for the specific date, along with the count of hospitalizations on the same date as well. All eight monitoring stations from case study are represented in the sample dataset.

TABLE V. SAMPLE AIR POLLUTION HOSPITALIZATION DATASET

Station	Date	CO	O3	PM10	NOx	NO2	NO	SO2	Hospitalizations
Raja Zarina	1/1/2006	0.7345	0.0117	45.4545	0.0346	0.0220	0.0128	0.0027	13
Petaling Jaya	1/1/2006	1.6413	0.0250	50.0833	0.0541	0.0300	0.0239	0.0040	3
Shah Alam	1/1/2006	0.8070	0.0278	37.0417	0.0251	0.0170	0.0083	0.0027	10
Sains Kuala Selangor	1/1/2006	N/A	N/A	40.6667	N/A	N/A	N/A	N/A	1
Putrajaya	1/1/2006	0.5887	0.0291	30.3913	0.0194	0.0139	0.0055	0.0010	1
Cheras, Kuala Lumpur	1/1/2006	1.2413	0.0240	33.9583	0.0416	0.0190	0.0228	0.0010	4
Batu Muda, Kuala Lumpur	1/3/2009	1.1314	0.0033	28.7500	0.0360	0.0160	0.0200	0.0021	15
Bukit Changgang	4/1/2010	0.3583	0.0147	41.0000	0.0096	0.0073	0.0023	0.0010	4

The generated dataset can be used for various associations and predictive models. Also, we have generated the dataset for the hospitalization count for the case study, but the researchers can alter this to include any other characteristic of health records as well.

V. CONCLUSION

Air pollution is one of the biggest environmental challenges in the 21st century, and it has a substantial health effects. These researchers are using the air pollution and health datasets to study the health impacts of air pollution. The parameters used to merge these datasets are residential address and other location parameters. With air quality monitoring stations concentrated in an urban region, it becomes a spatial problem that how to associate the patients with the monitoring stations around them. Also, most of the studies do not take into account the distance of station from the patients coordinates. In this study, we propose a spatial feature engineering algorithm to cater to these limitations. The proposed algorithm also includes a radius facility to exclude patients that live far away from AQM stations.

VI. FUTURE WORK

The future directions include addition of more parameters in both datasets. Additionally, using only the residential address as the spatial parameter can be considered as the limitation of this algorithm but it can be improved upon by adding more spatial parameters in future work.

ACKNOWLEDGMENTS

This research is funded by Taylor's University under the research grant application ID (TUFR/2017/004/04) entitled as "Modeling and Visualization of Air-Pollution and its Impacts on Health". We are also thankful to Department of Environment, Malaysia and Ministry of Health, Malaysia for providing the air quality monitoring station dataset and hospitalization dataset respectively.

REFERENCES

[1] WHO. (2017) The cost of a polluted environment, 1.7 million child deaths a year, says who. Accessed: 2019-03-28. [Online]. Available: https://www.who.int/en/news-room/detail/06-03-2017-the-cost-of-a-polluted-environment-1-7-million-child-deaths-a-year-says-who

[2] J. Burns, H. Boogaard, S. Polus, L. M. Pfadenhauer, A. C. Rohwer, A. M. van Erp, R. Turley, and E. Rehfuess, "Interventions to reduce ambient particulate matter air pollution and their effect on health," Cochrane Database of Systematic Reviews, no. 5, 2019.

[3] M. C. Turner, E. Gracia-Lavedan, M. Cirac, G. Castano-Vinyals, N. Malats, A. Tardon, R. Garcia-Closas, C. Serra, A. Carrato, R. R. Jones˜ et al., "Ambient air pollution and incident bladder cancer risk: Updated analysis of the spanish bladder cancer study," International journal of cancer, vol. 145, no. 4, pp. 894–900, 2019.

[4] L. A. T. Cox Jr and D. A. Popken, "Has reducing fine particulate matter and ozone caused reduced mortality rates in the united states?" Annals of epidemiology, vol. 25, no. 3, pp. 162–173, 2015.

[5] Y.-L. Huang and S. Batterman, "Residence location as a measure of environmental exposure: a review of air pollution epidemiology studies," Journal of Exposure Science and Environmental Epidemiology, vol. 10, no. 1, p. 66, 2000.

[6] Wikipedia. (2020) Feature engineering. Accessed: 2020-2-6. [Online]. Available: https://en.wikipedia.org/wiki/Feature engineering

[7] J. Thanaki, Python natural language processing. Packt Publishing Ltd, 2017.

[8] C. Vitolo, M. Scutari, M. Ghalaieny, A. Tucker, and A. Russell, "Modeling air pollution, climate, and health data using bayesian networks: A case study of the english regions," Earth and Space Science, vol. 5, no. 4, pp. 76–88, 2018.

[9] Seaton, D. Godden, W. MacNee, and K. Donaldson, "Particulate air pollution and acute health effects," The lancet, vol. 345, no. 8943, pp. 176–178, 1995.

[10] U. S. E. P. A. (EPA). (2020) Sulfur dioxide trends. Accessed: 2020-2-16. [Online]. Available: https://www.epa.gov/air-trends/sulfur-dioxide-trends

[11] C. A. Asia. (2020) Sulfur dioxide (so2) status and trends in asia. Accessed: 2020-2-16. [Online]. Available: https://cleanairasia.org/wpcontent/uploads/portal/files/documents/5 SO2 Status and _Trends in Asia Factsheet _21 September 2010.pdf

[12] P. Moraga, G. . C. of Death Collaborators et al., "Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the global burden of disease study 2016," The Lancet, vol. 390, no. 10100, pp. 1151–1210, 2017.

[13] EU. (2016) Air quality in europe — 2016 report, european environment agency 2016. Accessed: 2019-03-28. [Online]. Available: https://www.eea.europa.eu/publications/air-quality-in-europe-2016

[14] T. Bourdrel, M.-A. Bind, Y. Bejot, O. Morel, and J.-F. Argacha, "Cardiovascular effects of air pollution," Archives of cardiovascular diseases, vol. 110, no. 11, pp. 634–642, 2017.

[15] E. H. N. paper. (2017) European heart network paper-july 2017. Accessed: 2019-03-28. [Online]. Available: http://www.ehnheart.org/projects/1087:ehnpaper-on-air-pollution-and-cardiovascular-diseases.html

[16] D. Q. Rich, W. Zhang, S. Lin, S. Squizzato, S. W. Thurston, E. van Wijngaarden, D. Croft, M. Masiol, and P. K. Hopke, "Triggering of cardiovascular hospital admissions by source specific fine particle concentrations in urban centers of new york state," Environment international, vol. 126, pp. 387–394, 2019.

[17] D. Meier-Girard, E. Delgado-Eckert, E. Schaffner, C. Schindler, N. Kunzli, M. Adam, V. Pichot, F. Kronenberg, M. Imboden, U. Frey¨ et al., "Association of long-term exposure to traffic-related pm10 with heart rate variability and heart rate dynamics in healthy subjects," Environment international, vol. 125, pp. 107–116, 2019.

[18] E. Burte, B. Leynaert, R. Bono, B. Brunekreef, J. Bousquet, A.-E. Carsin, K. De Hoogh, B. Forsberg, F. Gormand, J. Heinrich et al., "Association between air pollution and rhinitis incidence in two european cohorts," Environment international, vol. 115, pp. 257–266, 2018.

[19] M. C. Turner, D. Krewski, W. R. Diver, C. A. Pope III, R. T. Burnett, M. Jerrett, J. D. Marshall, and S. M. Gapstur, "Ambient air pollution and cancer mortality in the cancer prevention study ii," Environmental health perspectives, vol. 125, no. 8, p. 087013, 2017.

[20] S. E. Alexeeff, A. Roy, J. Shan, X. Liu, K. Messier, J. S. Apte, C. Portier, S. Sidney, and S. K. Van Den Eeden, "High-resolution mapping of traffic related air pollution with google street view cars and incidence of cardiovascular events within neighborhoods in oakland, ca," Environmental Health, vol. 17, no. 1, p. 38, 2018.

[21] G. Cesaroni, D. Porta, C. Badaloni, M. Stafoggia, M. Eeftens, K. Meliefste, and F. Forastiere, "Nitrogen dioxide levels estimated from land use regression models several years apart and association with mortality in a large cohort study," Environmental Health, vol. 11, no. 1, p. 48, 2012.

[22] N. Middleton, P. Yiallouros, N. Nicolaou, S. Kleanthous, S. Pipis, M. Zeniou, P. Demokritou, and P. Koutrakis, "Residential exposure to motor vehicle emissions and the risk of wheezing among 7-8 year-old schoolchildren: a city-wide cross-sectional study in nicosia, cyprus," Environmental Health, vol. 9, no. 1, p. 28, 2010.

[23] Q. Deng, C. Lu, Y. Li, J. Sundell, and D. Norback, "Exposure to outdoor air pollution during trimesters of pregnancy and childhood asthma, allergic¨ rhinitis, and eczema," Environmental research, vol. 150, pp. 119–127, 2016.

[24] J. Wu, C. Ren, R. J. Delfino, J. Chung, M. Wilhelm, and B. Ritz, "Association between local traffic-generated air pollution and preeclampsia and preterm delivery in the south coast air basin of california," Environmental health perspectives, vol. 117, no. 11, pp. 1773–1779, 2009.

[25] M. Lin, Y. Chen, R. T. Burnett, P. J. Villeneuve, and D. Krewski, "The influence of ambient coarse particulate matter on asthma hospitalization in children: case-crossover and time-series analyses." Environmental health perspectives, vol. 110, no. 6, pp. 575–581, 2002.

[26] E. Dons, M. Van Poppel, L. I. Panis, S. De Prins, P. Berghmans, G. Koppen, and C. Matheeussen, "Land use regression models as a tool

for short, medium and long term exposure to traffic related air pollution," Science of the total Environment, vol. 476, pp. 378–386, 2014.

[27] E. Lavigne, M.-A. B´ elair, M. T. Do, D. M. Stieb, P. Hystad, A. Van Donkelaar, R. V. Martin, D. L. Crouse, E. Crighton, H. Chen´et al., "Maternal exposure to ambient air pollution and risk of early childhood cancers: a population-based study in ontario, canada," Environment international, vol. 100, pp. 139–147, 2017.

[28] M. Brauer, C. Lencar, L. Tamburic, M. Koehoorn, P. Demers, and C. Karr, "A cohort study of traffic-related air pollution impacts on birth outcomes," Environmental health perspectives, vol. 116, no. 5, pp. 680–686, 2008.

[29] Kheirbek, J. Haney, S. Douglas, K. Ito, and T. Matte, "The contribution of motor vehicle emissions to ambient fine particulate matter public health impacts in new york city: a health burden assessment," Environmental Health, vol. 15, no. 1, p. 89, 2016.

[30] G. Ribeiro, G. S. Downward, C. U. de Freitas, F. C. Neto, M. R. A. Cardoso, M. d. R. D. de Oliveira, P. Hystad, R. Vermeulen, A. C. Nardocci et al., "Incidence and mortality for respiratory cancer and traffic-related air pollution in sao paulo, brazil,"˜ Environmental research, vol. 170, pp. 243–251, 2019.

[31] L. M. Luong, D. Phung, T. N. Dang, P. D. Sly, L. Morawska, and P. K. Thai, "Seasonal association between ambient ozone and hospital admission for respiratory diseases in hanoi, vietnam," PloS one, vol. 13, no. 9, 2018.

[32] Y. Guo, H. Zeng, R. Zheng, S. Li, A. G. Barnett, S. Zhang, X. Zou, R. Huxley, W. Chen, and G. Williams, "The association between lung cancer incidence and ambient air pollution in china: A spatiotemporal analysis," Environmental research, vol. 144, pp. 60–65, 2016.

[33] G. Goudarzi, S. Geravandi, E. Idani, S. A. Hosseini, M. M. Baneshi, A. R. Yari, M. Vosoughi, S. Dobaradaran, S. Shirali, M. B. Marzooni et al., "An evaluation of hospital admission respiratory disease attributed to sulfur dioxide ambient concentration in ahvaz from 2011 through 2013," Environmental science and pollution research, vol. 23, no. 21, pp. 22001–22007, 2016.

[34] R. Beelen, G. Hoek, P. A. van Den Brandt, R. A. Goldbohm, P. Fischer, L. J. Schouten, M. Jerrett, E. Hughes, B. Armstrong, and B. Brunekreef, "Long-term effects of traffic-related air pollution on mortality in a dutch cohort (nlcs-air study)," Environmental health perspectives, vol. 116, no. 2, pp. 196–202, 2007.

[35] L. Perez, F. Lurmann, J. Wilson, M. Pastor, S. J. Brandt, N. Kunzli, and R. McConnell, "Near-roadway pollution and childhood asthma: implications for¨ developing "win–win" compact urban development and clean vehicle strategies," Environmental health perspectives, vol. 120, no. 11, pp. 1619–1626, 2012.

[36] R. Alotaibi, M. Bechle, J. D. Marshall, T. Ramani, J. Zietsman, M. J. Nieuwenhuijsen, and H. Khreis, "Traffic related air pollution and the burden of childhood asthma in the contiguous united states in 2000 and 2010," Environment international, vol. 127, pp. 858–867, 2019.

[37] J.-T. Lee, H. Kim, H. Song, Y.-C. Hong, Y.-S. Cho, S.-Y. Shin, Y.-J. Hyun, and Y.-S. Kim, "Air pollution and asthma among children in seoul, korea," Epidemiology, vol. 13, no. 4, pp. 481–484, 2002.

[38] C. Liu, P. Yin, R. Chen, X. Meng, L. Wang, Y. Niu, Z. Lin, Y. Liu, J. Liu, J. Qi et al., "Ambient carbon monoxide and cardiovascular mortality: a nationwide time-series analysis in 272 cities in china," The Lancet Planetary Health, vol. 2, no. 1, pp. e12–e18, 2018.

[39] J. Feng and W. Yang, "Effects of particulate air pollution on cardiovascular health: a population health risk assessment," PloS one, vol. 7, no. 3, p. e33385, 2012.

[40] R. Ghosh, F. Lurmann, L. Perez, B. Penfold, S. Brandt, J. Wilson, M. Milet, N. Kunzli, and R. McConnell, "Near-roadway air pollution and coronary¨ heart disease: burden of disease and potential impact of a greenhouse gas reduction strategy in southern california," Environmental health perspectives, vol. 124, no. 2, pp. 193–200, 2015.

[41] W. J. Requia, P. Koutrakis, H. L. Roig, M. D. Adams, and C. M. Santos, "Association between vehicular emissions and cardiorespiratory disease risk in brazil and its variation by spatial clustering of socio-economic factors," Environmental Research, vol. 150, pp. 452–460, 2016.

[42] G. Barnett, G. M. Williams, J. Schwartz, A. H. Neller, T. L. Best, A. L. Petroeschevsky, and R. W. Simpson, "Air pollution and child respiratory health: a case-crossover study in australia and new zealand," American journal of respiratory and critical care medicine, vol. 171, no. 11, pp. 1272–1278, 2005.

[43] B.-F. Hwang, J. J. Jaakkola, Y.-L. Lee, Y.-C. Lin, and Y.-l. L. Guo, "Relation between air pollution and allergic rhinitis in taiwanese schoolchildren," Respiratory research, vol. 7, no. 1, p. 23, 2006.

[44] M. A. B. A. Tajudin, M. F. Khan, W. R. W. Mahiyuddin, R. Hod, M. T. Latif, A. H. Hamid, S. A. Rahman, and M. Sahani, "Risk of concentrations of major air pollutants on the prevalence of cardiovascular and respiratory diseases in urbanized area of kuala lumpur, malaysia," Ecotoxicology and environmental safety, vol. 171, pp. 290–300, 2019.

[45] R. Buchdahl, C. D. Willems, M. Vander, and A. Babiker, "Associations between ambient ozone, hydrocarbons, and childhood wheezy episodes: a prospective observational study in south east london," Occupational and environmental medicine, vol. 57, no. 2, pp. 86–93, 2000.

[46] J. Thompson, M. D. Shields, and C. C. Patterson, "Acute asthma exacerbations and air pollutants in children living in belfast, northern ireland," Archives of Environmental Health: An International Journal, vol. 56, no. 3, pp. 234–241, 2001.

[47] S.-H. Cho, H. Tong, J. K. McGee, R. W. Baldauf, Q. T. Krantz, and M. I. Gilmour, "Comparative toxicity of size-fractionated airborne particulate matter collected at different distances from an urban highway," Environmental health perspectives, vol. 117, no. 11, pp. 1682–1689, 2009.

[48] Y. J. Lee, Y. W. Lim, J. Y. Yang, C. S. Kim, Y. C. Shin, and D. C. Shin, "Evaluating the pm damage cost due to urban air pollution and vehicle emissions in seoul, korea," Journal of Environmental Management, vol. 92, no. 3, pp. 603 – 609, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301479710003208

[49] C. Borrego, A. Costa, R. Tavares, M. Lopes, J. Valente, J. Amorim, A. Miranda, I. Ribeiro, and E. Sa, "Effects of road traffic scenarios on human´ exposure to air pollution," WIT Transactions on Ecology and the Environment, vol. 123, pp. 89–100, 2009.

[50] B.-J. Lee, B. Kim, and K. Lee, "Air pollution exposure and cardiovascular disease," Toxicological research, vol. 30, no. 2, p. 71, 2014.

[51] R. B. Devlin, K. E. Duncan, M. Jardim, M. T. Schmitt, A. G. Rappold, and D. Diaz-Sanchez, "Controlled exposure of healthy young volunteers to ozone causes cardiovascular effects," Circulation, vol. 126, no. 1, pp. 104–111, 2012.

[52] K. Newell, C. Kartsonaki, K. B. H. Lam, and O. Kurmi, "Cardiorespiratory health effects of gaseous ambient air pollution exposure in low and middle income countries: a systematic review and meta-analysis," Environmental Health, vol. 17, no. 1, p. 41, 2018.

[53] C. Johansson, B. Lovenheim, P. Schantz, L. Wahlgren, P. Almstr¨ om, A. Markstedt, M. Str¨ omgren, B. Forsberg, and J. N. Sommar, "Impacts on air¨ pollution and health by changing commuting from car to bicycle," Science of the total environment, vol. 584, pp. 55–63, 2017.

[54] C. Guerreiro, J. Horalek, F. de Leeuw, and F. Couvidat, "Benzo (a) pyrene in europe: Ambient air concentrations, population exposure and health effects,"´ Environmental pollution, vol. 214, pp. 657–667, 2016.

[55] D. P. Lamoureux, E. A. Diaz, Y. Chung, B. A. Coull, V. Papapostolou, J. Lawrence, R. Sato, and J. J. Godleski, "Effects of fresh and aged vehicular particulate emissions on blood pressure in normal adult male rats," Air Quality, Atmosphere & Health, vol. 6, no. 2, pp. 407–418, 2013.

[56] M. Shekarrizfard, M.-F. Valois, M. S. Goldberg, D. Crouse, N. Ross, M.-E. Parent, S. Yasmin, and M. Hatzopoulou, "Investigating the role of transportation models in epidemiologic studies of traffic related air pollution and health effects," Environmental research, vol. 140, pp. 282–291, 2015.

[57] J.-Z. Wu, D.-D. Ge, L.-F. Zhou, L.-Y. Hou, Y. Zhou, and Q.-Y. Li, "Effects of particulate matter on allergic respiratory diseases," Chronic diseases and translational medicine, vol. 4, no. 2, pp. 95–102, 2018.

[58] D. L. Buckeridge, R. Glazier, B. J. Harvey, M. Escobar, C. Amrhein, and J. Frank, "Effect of motor vehicle emissions on respiratory health in an urban area." Environmental health perspectives, vol. 110, no. 3, pp. 293–300, 2002.

[59] M. H. Askariyeh, S. Vallamsundar, J. Zietsman, and T. Ramani, "Assessment of traffic-related air pollution: Case study of pregnant women in south texas," International journal of environmental research and public health, vol. 16, no. 13, p. 2433, 2019.

[60] L. Hou, K. Zhang, M. A. Luthin, and A. A. Baccarelli, "Public health impact and economic costs of volkswagen's lack of compliance with

the united states' emission standards," International journal of environmental research and public health, vol. 13, no. 9, p. 891, 2016.

[61] D. Norback, J. H. Hashim, Z. Hashim, and F. Ali, "Volatile organic compounds (voc), formaldehyde and nitrogen dioxide (no2) in schools in johor¨ bahru, malaysia: Associations with rhinitis, ocular, throat and dermal symptoms, headache and fatigue," Science of The Total Environment, vol. 592, pp. 153–160, 2017.

[62] DOE. (2013) Sources of air pollution in malaysia - jabatan alam sekitar. Accessed: 2019-12-26. [Online]. Available: http://www.doe.gov.my/portalv1/wpcontent/uploads/2013/06/General-Information-of-Air-Pollutant-Index.pdf

[63] DOE. (2020) Air pollutant index of malaysia. Accessed: 2020-2-4. [Online]. Available: http://apims.doe.gov.my/public v2/home.html

[64] R. S. A. Usmani, W. N. F. B. W. Azmi, A. M. Abdullahi, I. A. T. Hashem, and T. R. Pillai, "A novel feature engineering algorithm for air quality datasets," Indonesian Journal of Electrical Engineering and Computer Science, vol. 19, no. 3, Sep. 2020.

[65] MapQuest. (2020) Official mapquest - maps, driving directions, live traffic. Accessed: 2020-03-28. [Online]. Available: https://www.mapquest.com/

[66] Microsoft. (2020) System.device.location namespace, microsoft docs. Accessed: 2020-03-28. [Online]. Available: https://docs.microsoft.com/enus/dotnet/api/system.device.location