

DS-670

Assignment 3 – Expected contribution and State of Art

Pollution is when natural resources like Air, Water, Land and other parts of environments starts to become unsuitable or unsafe to use. Air pollution is one of the most dangerous pollutions humans are facing right now. Air pollution may result in various harmful diseases and mortality rate will also fall due to this. There are many factors which results in air pollution, humans are also one of the factors for causing air pollution and important factor. Mostly air pollution is caused due to motor vehicles and many recent electronic devices. There are many ways to calculate the air pollution, sensors can be used to understand the percentage of different parameters which cause the air pollution.

Air pollution will have different parameters to consider. Air pollution can be measured using Air quality Index (AQI) metric. This is measured using nearly 449 observation points in total. If AQI increases it also increase chances of health related issues. Sensors can be used for measuring parameters. For the sake of dataset they used one sensor for each traffic sensors available in the area. These sensors will give information and values like Carbon Monoxide levels, Nitrogen dioxide level, Sulphur Dioxide level, Particulate Matter and ozone index level as specified accordingly in Air Pollution Index from wiki. Sensor measures the values by initially assigning it to a value from 25 to 100 according what it is measuring and how dense it is for example Carbon Monoxide. Next for every 5 minutes previous value will be added by a random number from 1 to 10 if it's value is below 20. Same way if the value is above 210, a random number from 1 to 10 is subtracted from the previous value. Else a random value from -5 to +5 is added to last value. It is followed this way because the values would not fall in low and high suddenly and keep the values more realistic and confine them in bounds.

The dataset is from Citypulse website, where it provides with the air pollution dataset of Brasov in Romania. There are 449 different locations where data is collected. Dataset consists of 449 different csv files, each file for each location. Each csv file have parameters like location (latitude and longitude's), timestamp (it is recorded for every 5 minutes), ozone level, particulate matter level, carbon monoxide level, sulphur dioxide and nitrogen dioxide. Dataset is from the dates August 1 2014 to September 30 2014, 2 months of data which is recorded every 5 minutes. Using this dataset we can find some pattern in the air pollution and help in bringing that down. There are some ideas or implementations which we follow through in this paper for better understanding and conclude with some insight.

From the dataset we can find the locations and their respective gas emissions. From these factors we can find some patterns like correlation between time and pollution. We can also perform time-series calculations and can predict the pollution on that particular location at this particular interval of time. From this analysis we can mark the highly polluted areas and may avoid frequent visits in these related areas. These can be calculated by the dataset we have, there are 449 locations and each location have one separate file. Each file has around 18000 records with location and emission values. So each location can be taken at a time and sum all the values. These can be categorized into different time series like calculate for monthly, daily and hourly. From this we can

find a pattern and may predict for the next day or hour. These can be very helpful in understanding the pollution distribution and help people in avoiding those areas. We can use Brasov city data and use the location data and find the values of the pollution and show case the highly polluted areas in the city by heatmap. Using these heatmap we can find the highly polluted areas so that public can avoid those in case of travelling. This can also help in taking measures for reducing the pollution in those areas, may be by planting more trees or making vehicles to take alternate route. Some ideas of showing the polluted areas are by making specific areas green where it is not polluted and red where it is highly polluted.

Using different regression models and predicting the next day's or next hour's pollution in that particular area. We can implement all these with different tools available like tableau, R, SQL. As this is not a huge data we can simply load it into R and run regression models. If it is a huge data set and even SQL cannot handle the performance, we can go for big data technologies like Hadoop, MapReduce and Spark. If size of dataset is reasonable enough to run on normal databases, it is better to use it as it is faster and economical for the performance.

Some areas are hugely populated, in these areas pollution is high automatically. When there is huge population, they need more transportation systems. Transportation systems may include different modes of transportations. Even emissions from flights are causing damage to ozone layer. This in turn will reflect on living organisms on earth like cancer. Pollution will have huge impact on health of living organisms. Humans are the main factors for these air, water and other pollutions. Air pollution is mainly caused due to emissions from the motor vehicles or any related equipment. Due to this emission of poisonous gases into the air, they undergo some chemical reactions with the atmosphere and result in bad environment. This environmental change in turn reflects on human health and affects badly on future generations.

Pollution is relatively proportional to the population of the area. For example less densely populated areas have less transportation so this results in less gas emissions and which results in less pollution. So correlation between pollution and population is also considered in study. This makes huge difference in understanding and finding the patterns. If we did not consider the population and calculated the pollution, we cannot rely on that results and it is not a good model to consider for further calculations or predicting the pollution. One limitation for this approach is that we need to have access to the population dataset and we need to have location data too for that population. We have the pollution dataset and locations based on sensors. But we do not have sensors for calculating the population. Even if we had, it would not be same locations as the sensor data. The population and pollution locations will operate on different areas. So this approach cannot be accurate enough, though we can find the city population and can calculate. But this will be for whole and cannot be relied on it.

Research paper titled "predict air pollution in Brasov city with regression models" is concentrated on what are the parameters that influence the results of air pollution in Brasov city. In this study they followed some methodologies to predict the pollution in urban air pollution. Also explained about how vehicle transportation is important and what are the minimum requirements used for these kind of transportations. By this how our civilization is advancing and improving technological advancements. On the other hand these advancements are also having negative impacts on the world we live in. These negative impacts will result in the environment changes, health concerned issues and our world, earth we live in will day by day become unsuitable for humans. These are serious issues every nation and every individual should consider. Our future generation will depend on the actions we take now. So considering seriousness of the environment, this study helps in understanding distribution and level of air pollution in small area compared to the world Brasov city.

They tried different regression models to come up with better prediction of the air pollution. In my research and I am going to consider these methods to understand how it is distributed and predicted and I will implement those on our dataset. I also consider the parameters like correlation between area and pollution, population and pollution. Because we cannot say that these part of city have high pollution based on parameters, there could be more vehicles running in these areas at some particular time interval and which results in these kind of increase in air pollution. So population should also be a main factor in deciding and predicting the air pollution.

Another research paper titled “Modelling the impact of road traffic on air pollution in urban environment case study: A new overpass in the city of Craiova” represents the work of authors on correlation between average speed of vehicles and volume of emissions pollution evaluations. They also considered the parameters and factors like infrastructure and area of travel for better understanding. For example in highways, average speed is comparatively more than the average speed of vehicles in normal roads these factors are also important in higher understanding. They also stated that by implementing an overpass in the Craiova city, it can reduce the emissions of gases like CO, C₆H₆, NO₂ and O₃. They performed their research on Craiova city dataset. The abstract mainly started with stating population is being increased day by day and due to increase in population there is a need to increase in transportation systems and technical advancements. Transportation is the main source of air pollution now-a-days. There will be effect on humans, animals, infrastructure, landscapes and plants due to air pollution. In this paper they performed the correlation between average speed and gas emissions. As speed is also a considerable factor for gas emissions and in turn results in air pollution.

In “Traffic Air Pollution and Mortality Rate Advancement Periods” they directly calculating the relation between pollution and mortality rate. Air pollution is directly related with mortality rates. The effect of air pollution in respect to different reasons for death in a population is of general wellbeing significance and has not been introduced. In this study, increase in rate time periods is related with air pollution exposures were evaluated. They performed the tests at a centre in Hamilton, Ontario, Canada, in the vicinity of 1985 and 1999 for pulmonary testing. Cox regression model was used to model mortality from all normal causes from 1992 to 2001 in connection to various organs in body. As discussed previous air pollution will not only make the world unsafe and unsuitable for future generations to live it also have huge effect on human health. Not only humans, even animals and plants. Due to damage caused by air pollution on human health, it reduces the mortality rate of the humans. In our research we are not going into mortality rate and how humans are being affected by the air pollution. We are going to perform some regression models to see how the data is distributed. This distribution of the data is related to dataset of Brasov city. Where there are locations and values for the different emissions of gases.

Other similar articles provided in the reference are following the similar approach and finding the predictions of mortality rates, how it is effected to one's health. These research also considered different factors varying from simple highways to as complicated as emission of gases with respect to speed. These all factors play major role in air pollution. In our research we consider the location, time and different gases as we have access to this data only and perform different regression models and test which one is best suited and gives more accurate results in predicting the pollution for next hour or next day. This paper will also provide with the areas which are highly polluted. This can be used in various field to make the world better place by implementing some measures reduce the pollution or showing the public what are the dangerous places to live or travel regarding air pollution.

References:

- Cofaru, C. (2011, February). Strategies of Developing Road Transport by Controlling Automotive Emissions to Reduce Local and Global Environment Impacts. In Proceedings of 6th IASME/WSEAS International Conference on Energy & Environment (pp. 23-25).
- Cofaru, C. O. R. N. E. L. I. U., Timar, J. A. N. O. S., Florea, D. A. N. I. E. L. A., & Covaciu, D. I. N. U. (2009, September). Study Regarding the Simulation of Urban Traffic Noise in Braşov City. In WSEAS Conference of Advances in Environmental and Geological Science and Engineering, Transylvania University of Brasov (pp. 154-157).
- Tarulescu, S., Tarulescu, R., & Soica, A. (2009). Approximation of the carbon monoxide concentration resulting from the road traffic using experimental measurements. *Annals of DAAAM & Proceedings*, 1865-1867.
- Peptenatu, D., Pintilii, R. D., & Draghici, C. (2011). Environmental risk management of urban growth poles regarding national importance. *International Journal of Environmental Science & Technology*, 8(4), 737-746.
- Mocanu, R., & Steinnes, E. (2002). Air pollution monitoring in Romania. In *Global Atmospheric Change and its Impact on Regional Air Quality* (pp. 43-48). Springer Netherlands.
- COFARU, C., & TARULESCU, S. Study Regarding Air Pollution Produced by Vehicles and a Prediction Model of Chemical Pollutants. In WSEAS Proceedings of the 2nd International Conference on Environmental and Geological Science and Engineering (pp. 218-224).
- Marinescu, N. (2011). Regional Policy In Practice: The Case Of The Brasov Metropolitan Area. *Studia Universitatis Babes Bolyai-Negotia*, (2), 79-88.
- Murena, F. (2007). Air quality nearby road traffic tunnel portals: BTEX monitoring. *Journal of Environmental Sciences*, 19(5), 578-583.
- Iordache, S., & Dunea, D. (2013). Cross-spectrum analysis applied to air pollution time series from several urban areas of Romania. *Environmental Engineering and Management Journal*, 12(4), 677-684.
- Mitran, G., Ilie, S., Tabacu, I., & Nicolae, V. (2012). MODELING THE IMPACT OF ROAD TRAFFIC ON AIR POLLUTION IN URBAN ENVIRONMENT CASE STUDY: A NEW OVERPASS IN THE CITY OF CRAIOVA. *Environmental Engineering & Management Journal (EEMJ)*, 11(2).
- Hordijk, L., & Amann, M. (2007). How science and policy combined to combat air pollution problems. *Envtl. Pol'y & L.*, 37, 336.
- Tarulescu, S., Beles, H., Soica, A., Chioreanu, N., & Mitran, T. (2014). Predict Air Pollution in Brasov City with Regression Models. In *Applied Mechanics and Materials* (Vol. 659, pp. 617-622). Trans Tech Publications.
- Stelian, T. A. R. U. L. E. S. C. U., Horia, B. E. L. E. S., Adrian, S. O. I. C. A., Nicolae, C. H. I. O. R. E. A. N. U., & Tudor, M. I. T. R. A. N. (2014). Predict Air Pollution in Brasov City with Regression Models. *Applied Mechanics & Materials*, (659).
- Stohl, A., Eckhardt, S., Forster, C., James, P., & Spichtinger, N. (2002). On the pathways and timescales of intercontinental air pollution transport. *Journal of Geophysical Research: Atmospheres*, 107(D23).

Krzyzanowski, M., Kuna-Dibbert, B., & Schneider, J. (2005). *Health effects of transport-related air pollution*. WHO Regional Office Europe.

Colville, R. N., Hutchinson, E. J., Mindell, J. S., & Warren, R. F. (2001). The transport sector as a source of air pollution. *Atmospheric environment*, 35(9), 1537-1565.

Akimoto, H. (2003). Global air quality and pollution. *Science*, 302(5651), 1716-1719.

Costabile, F., & Allegrini, I. (2008). A new approach to link transport emissions and air quality: An intelligent transport system based on the control of traffic air pollution. *Environmental Modelling & Software*, 23(3), 258-267.

Zhang, L., Jacob, D. J., Boersma, K. F., Jaffe, D. A., Olson, J. R., Bowman, K. W., ... & Dibb, J. E. (2008). Transpacific transport of ozone pollution and the effect of recent Asian emission increases on air quality in North America: an integrated analysis using satellite, aircraft, ozonesonde, and surface observations. *Atmospheric Chemistry and Physics Discussions*, 8(2), 8143-8191.

Hickman, J., Hassel, D., Joumard, R., Samaras, Z., & Sorenson, S. (1999). Methodology for calculating transport emissions and energy consumption.

Dataset Collection. (n.d.). Retrieved February 04, 2017, from <http://iot.ee.surrey.ac.uk:8080/datasets.html#pollution>

(n.d.). Retrieved February 04, 2017, from <http://iot.ee.surrey.ac.uk:8080/datasets/pollution/readme.txt>

Air quality index. (n.d.). Retrieved February 04, 2017, from https://en.wikipedia.org/wiki/Air_quality_index