Assignment 10: Draft Manuscript

Prithvi Yarlagadda

Results:

We tried different prediction models to come up with predicting air pollution in Brasov city. We can use this models to predict pollutions of different cities. These predictions are also based on the parameters like correlation between area with pollution and population with 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.

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.

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 its 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.

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. 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 in 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 is in turn reflect on human health and effect badly on future generations.

In this project we are going to use python in Zeppelin. In Apache Zeppelin we can write python scripts by using %python tag before writing python scripts. But we need to use spark integrated python too as for now we have only limited dataset and in future if we are going to build large scale project based on this, we might need to leverage spark parallel processing and high volume processing. For using python in spark we need to use pyspark and this will change the interpreter into python on spark. Apache Zeppelin provides all these inbuilt so we do not have to reconfigure any setting to get this functionalities. More over Apache Zeppelin is open source which gives huge advantage to programmers and application builders and contribute in the development of tools which are helpful.

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.