**Assignment – 11**

**DS-607-Capstone**

Neha Kumari

**CONCLUSION**

K-means clustering was applied to get the healthiest area in the smart city of Denmark. K value changed from 3 to 10 until convergence was achieved. We found the cleanest area at longitude 10.18032 and latitude 56.17061 where ozone level was minimum, 81.92253 on Air Quality Index.

**DISCUSSION**

The data set is a generated from sensor network that provide real time measurements of the air pollution levels. The data is sent over wide area network, which is an Internet of things connectivity.

The ultimate goal would be to use this data into a decision support systems that aims to improve the quality of life in an urban city.

The smart city project can be integrated further to understand the impact of road traffic and on future urban planning and development.

It should fulfill the following needs:

* Monitor real time air pollution.
* Pollution exposure for residents, pedestrians and drivers on an average day.
* Generate reports to take decisions on urban planning and support.

The benefits we can expect are:

* Optimize road traffic in areas with high pollution and improve quality of life.
* Take informed decisions on urban planning.
* Increase public awareness and pollution reduction initiatives.

According to an analysis by WHO, almost 90% of urban population today breathes in polluted air such that the pollutant level is much higher than the recommended thresholds. It poses great health issues among children and adults. Also according to UN by 2030 nearly 60% of world population will live in urban areas. The reports also suggests that cities only occupy 2% of the earth’s surface but are responsible for nearly 70% of the carbon dioxide emissions in the global environment.

The other area of work in this field can be accurate forecast of air pollution levels. The sensors collect real time data and can be used for real time series forecasting. Deep Learning frameworks like Tensor Flow, azure, theano, Torch, Keras, Caffe, with computer vision are powerful tools used for real time analysis. Deep Neural Networks, artificial intelligence can precisely predict the pollution levels in several day advances. The neural network having more layers, and each layer are linked which helps in accurate predictions. ELM is a two-layer neural network that works well for predictions. First layer is random and second layer is trained. It is used for classification, regression, feature selection etc.

The training process involves hundreds of iterations wherein the algorithm learns and reduces the error, meaning the difference between the forecasted values and actual values. The weather, road traffic along with air pollution data can be needed for accurate forecasts as it also affects the pollution levels.

Further, there are many research going on, one of them is to collect highly localized data, meaning instead of collecting data from sensors, sensors being attached to the wrists of people who travel across the common routes like subway, trains, cars, pedestrian walking. This study can be immensely helpful to understand the exposure to air pollution of common people at different time of the day. The other is the government has approved building Smog Chamber in Beijing, China. The chamber would emit different pollutant mixtures and create conditions similar to smog, to study them. Like the chemical reactions that take place under sunshine and without sunshine to control heavy smog.

In our project findings, we predict the cleanest area for the day as per the data available from the sensors. In this study ozone was the deciding factor, other pollutants can also be used to see and investigate their effect. The data set was pretty clean and structured, well suited for K-means. The results may vary for each run, as the initialization of K-means is different for every run. But as per our observation in most runs the best results were obtained with K=9 or K=10.

Now the big question is how can we use these results in improving quality of life in a Smart City. We need to build the responses, preventive measures and decision systems as per the predictions of these models. For example if our model predicts high pollution level in certain region, the place could raise the congestion charge, block heavy vehicles like trucks or encourage people to use public transports, electric cars, car sharing, creating green areas initiatives, mobile apps that can direct to the greenest route for that day just like Google maps for fastest route. On a larger level where the air pollution has already reached to alarming levels like in Beijing some preventive measures are taken like the regulation to control air pollution is done. Every city will have a yearly quota to limit the pollutant emissions; the purpose is to gradually control the pollutant levels and in particular smog in these areas. Other measures like Building low emissions zones (LEZ) is also becoming popular in many parts across the world. It is a step to control pollution in urban cities and usually high pollutant emission vehicles are not allowed in these areas.

The effect of industrial gases in industrial areas can be dealt separately.

In fact many experiments and research are going on in collaboration with companies like Intel, Siemens and government organizations with the universities on this subject. There are real time Air Quality Index apps using deep learning frameworks like Tensor Flow, theano, being developed locally. In many cities like Chicago, Pittsburg, Oregon, cities in Germany the air monitoring systems are working that provide the air quality in real time for different locations in the city. Cameras along with sensors let people know the source of pollution like traffic, industrial sources, Carbon dioxide emissions etc. Further the effect of green house gases on the pollution level can be studied. It’s been on research level and the next step would be to build these data driven systems that can provide some useful insights on which decision systems can be made. The Smart City is a connected city so these findings can be useful for integrating some other smart city projects working on different or similar data sets like weather or road traffic.