USING MACHINE LEARNING APPROACHES FOR THE PREDICTION OF AIR QUALITY INDEX

Jacqueline Kennady
Supervisor: Jason Barron
MSc. Data Science

Institiúid Teicneolaíochta Cheatharlach
INSTITUTE of TECHNOLOGY CARLOW
At the heart of South Leinster

About the Research

- Air pollution in urban environment has risen steadily in the last several decades.
- Urbanization and industrialization are the main reason for the depletion of air quality.
- Air quality index is the scale to measure the quality of air (Baralis et al., 2016).
- Earlier classical method such as probability and statistics are used to measure the quality of air.
- Now the AQI prediction is done through machine learning algorithm, deep learning methods and neural network.
- In this research various supervised machine learning algorithms are used to predict the accurate AQI.

Methodology

Collect meteorological data and corresponding weather data from environmental monitoring center.

Select the features from the dataset by examining the correlation between the factors.

Hyper parameter tuning will be applied for better accuracy.

Train models by machine learning regression models with the prepared data and scrutinize the result.

Choose the model with most appropriate test result.

What is the goal?

- The objective of this research is to predict the air quality based on meteorological and weather data.
- Various supervised machine leaning algorithms will be evaluated in comparison to others to identify the most suitable approach for the prediction.

Data Presentation

The dataset contain 1 million records of meteorological as well as weather station data from the China meteorological administration.

PM2.5 Concentrations Across China In 2013 And 2017 (µg/m3)
2013 2017 WHO Standard China Standard
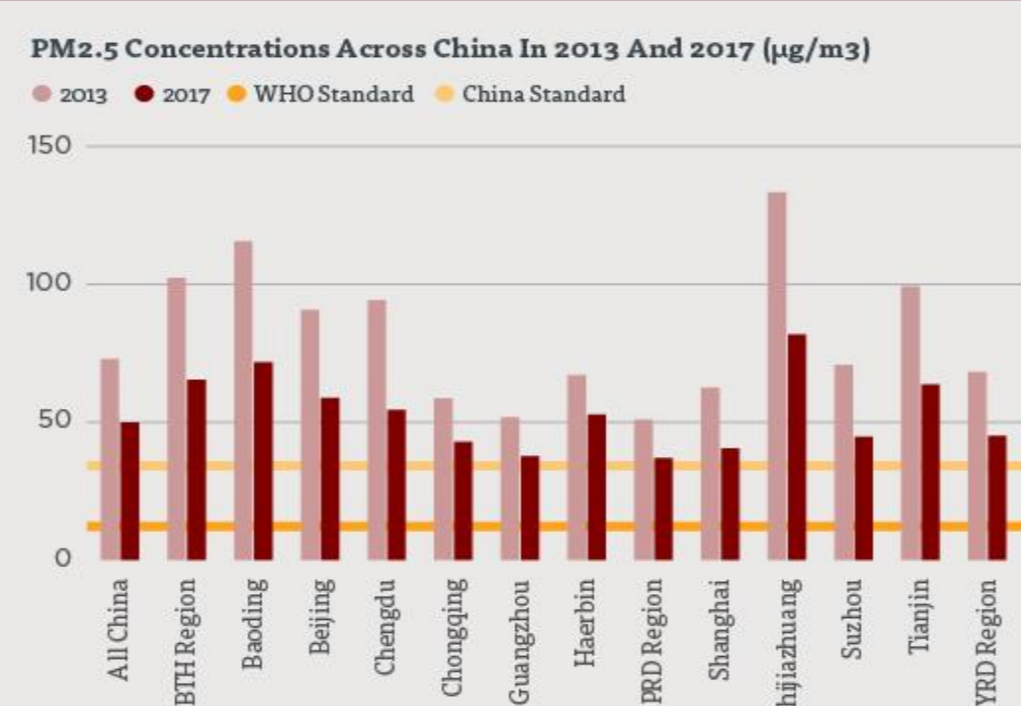
Note: Data are from the China National Environmental Monitoring Center. "All China" refers to the 204 prefectures for which the balanced sample of monitors has available data. "BTH" refers to Beijing-Tianjin-Hebei; "PRD" refers to the Pearl River Delta; "YRD" refers to the Yangtze River Delta. The cities shown in this figure are the ten most populated prefectures in our sample.

Figure 1:PM concentration across china

Technologies

jupyter python NumPy scikit learn pandas matplotlib

A Brief on Previous Work

- Ishan et.al described that ARIMA have been widely used to find out level of air pollutants as well as air quality and it achieved highest popularity due to its adaptability to represent a wide range of process and its statistical properties(Ishan et.al.,2017).
- Chao Zhang et.al specified machine learning models are very useful to find out the hidden relationship between data (Chao Zhang et.al,2020).
- When predicting the time series the machine learning approach like ANN, GP, SVM have been shown to outperform ARIMA.
- (Ruijun Yang et al.,2015) compared the result obtained using a multilinear regression model to the ones achieved by an ANN where ANN produce more robust result.

Conclusion

- The system propose different algorithm to predict the AQI with greater accuracy.
- The proposed system will help common people as well as those in the meteorological department to detect and predict pollution levels.
- This system can be extended by adding real time traffic data which will give more accurate prediction in AQI in on the time basis.

Email: c00565740@itcarlow.ie

Supervisor E-mail: Jason.Barron@itcarlow.ie

Student ID: C00265740

Submission Date:09/04/2021

Reference:

- Elena Baralis, Tania Cerquitelli, Silvia Chiusano, Paolo Garza, and Moham-mad Reza Kavoosifar. 2016. Analyzing air pollution on the urban environment.
- Ming Li, Ruiyuan Li, Zhangqing Shan, Yu Zheng, Xiuwen Yi ,Eric Chang, andTianrui Li. 2015. Forecasting Fine-Grained Air Quality Based on Big Data.
- Verma, Ishan, Rahul Ahuja, HardikMeisheri, andLipikaDey. "Air pollutant severity rediction using Bi-directional LSTM Network." In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 651-654. IEEE, 2018.
- Yang, Ruijun, Feng Yan, and Nan Zhao. "Urban air quality based on Bayesian network." In 2017 IEEE 9th Fig. 10. RH w.r.t NO Fig. 11. RH w.r.t NO2.