Paper Title:

Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities

Paper Link:

https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8746201

1 Summary

1.1 Motivation

This Paper is driven by the significant environmental and health concerns related to air pollution in smart cities. Every year millions of people are at health risk due to air pollution. Also due to emission of this pollution from various sources it greatly contributes in greenhouse effect and also it is crucial to the global climate change process. The urgent need to address air pollution is highlighted by the fact that it is a risk factor for many types of health issues such as asthma, lung cancer, heart issues etc types of respiratory diseases. Because of rapid urbanization and the increasing demand of transportation and energy, effective air quality is must for smart cities development.

1.2 Contribution

The main contribution of this paper is to conduct a comparative analysis of four advanced regression techniques at predicting air quality in smart cities. They evaluated these techniques based on data size, processing time and prediction accuracy. They utilized Apache Spark for conducting experiments and performing pollution estimation using multiple datasets to select the most efficient model for air quality prediction which the presented based on metrics like Mean Absolute Error(MAE) and Root Mean Square Error(RMSE)

1.3 Methodology

This paper used four regression techniques which are Decision Tree Regression (DTR),Random Forest Regression(RFR), Gradient Boosting Regression and ANN Multilayer Perceptron Regression(MLPR). They compared these techniques performance and processing time on Apache Spark using five datasets of different cities.

1.4 Conclusion

This paper analyzed and compared four existing schemes. From them they concluded that Random Forest Regression was best as it has the lowest error rate and processing time, and Gradient Boosting Regression was worst as it has highest error rate and processing rate on most of the datasets.

2 Limitations

2.1 First Limitation

In this paper dataset they used is limited in terms of size, locations and characteristics. Using more locations and variations data in their dataset would provide a more comprehensive analysis.

2.2 Second Limitation

In this they have used only four regression techniques. They could have used more advanced and robust techniques to increase accuracy and reliability of the prediction of their model.

3 Synthesis

This research paper compared four machine learning algorithms for predicting air quality in smart cities. They evaluated four regression techniques based on error rate and processing time, from which they found that Random Forest Regression(RFR) is best among them for real-time air quality prediction in smart cities.