

Team: DC20035

Member: Zhenyang Wang, Yanzhi Shen

Project: level-5, New Pollution

Abstract

Pollutants are a mix of harmful particles and gases in the atmosphere that results in air pollution. Higher and unhealthy concentrations of air pollutants lead to diseases, unsustainable household and industrial productions, and uninhabitable ecosystems. This research aims to seek possible solutions to improve the air quality by analyzing factors that affect the pollutants value. This research uses a time series dataset of daily observations of four pollutants in the DC, Virginia and Maryland areas from 2000 to 2010. Additionally, the research introduces outer datasets from industry, highway, climate and weather to comprehensively study the effects of various factors on air pollutants.

This research covers linear regression, decision tree and gradient boosting regression to fit the dataset. Besides, multivariate time-series methods, the VAR and the LSTM model, are trained and tested. We also introduce the Poisson regression, a good model to fit count variables, to detect the distribution of certain hours that have the max value of the pollutant molecule. Our results show that human sources, such as household, industry and transport, and natural sources, such as the hour time of the day and the weather all have their unique effects on air pollutants. Results also reveal the specific factor that has the most impact on the pollutant concentrations in each area, and thus provide clues to give customized solution recommendations in a more efficient way. This research gives predictions to pollutants produced in the DC, Virginia and Maryland areas with different time ranges. Also, from a sustainable perspective, the research also emphasizes the need to take into account both human and natural factors when dealing with air pollution, create a solution matrix, and thus improve the sustainability of productions and ecosystems.