Predicting NO2 concentrations

This manuscript (permalink) was automatically generated from sfiala2/498_NO2_pred@1d1295c on December 5, 2020.

Authors

- **Tessa Clarizio**
  · tessac2
- **Jane Roe**
  · janeroe
- **Jane Roe**
  · janeroe
- **Jane Roe**
  · janeroe
- **Jane Roe**
  · janeroe
- **Jane Roe**
  · janeroe

Abstract

test commit on abstract

1. Introduction

NO2 is defined by the U.S. Environmental Protection Agency (EPA) as a criteria air pollutant, meaning it poses a risk to human and environmental health. The primary National Ambient Air Quality Standard (NAAQS) is set at a 53 ppb annual average [1]. NO2 can cause respiratory irritation and can aggravate respiratory diseases such as asthma (US EPA, n.d., B). NO2 can also react with other chemicals in the atmosphere to form both particulate matter (PM) and tropospheric ozone (US EPA, n.d., B). PM and ozone are also criteria air pollutants and are harmful to human health. NO2 also contributes to the formation of acid rain, smog, and nutrient pollution in coastal waters (US EPA, n.d., B). The primary source of NO2 emissions is fossil fuel combustion, particularly from traffic and power plants (US EPA, n.d., B).

Therefore, understanding and predicting the spatial variability of NO2 emissions is of great importance to public health. However, prediction of air quality can be complicated due to the number of factors that affect local air quality, ranging from meteorology to land use. Machine learning models are a useful tool to interpret and find relationships in complex data.

[introduce Bechle study...] Bechle et al (2015) explores the impact of.. [Grace please add here]

This report proposes a machine learning model to predict NO2 concentrations spatially. First, a literature review was undertaken to understand what machine learning models have typically performed well in predicting air quality. Next, an exploratory data analysis (EDA) was performed on the Bechle et al (2015) dataset. Finally, multiple linear regression, neural network and random forest models were built and results were compared to see which method had the lowest mean-squared error (MSE).

2. Methods

2.1 Literature Review

2.1.1 PM2.5

2.1.2 AQI/API

2.2 Exploratory Data Analysis

2.3 Model

2.3.1 Multiple Linear Regression

2.3.2 Neural Networks

2.3.3 Random Forest

3. Results

4. Discussion

References

1. **Primary National Ambient Air Quality Standards (NAAQS) for Nitrogen Dioxide**
   OAR US EPA
   US EPA (2016-07-01) https://www.epa.gov/no2-pollution/primary-national-ambient-air-quality-standards-naaqs-nitrogen-dioxide