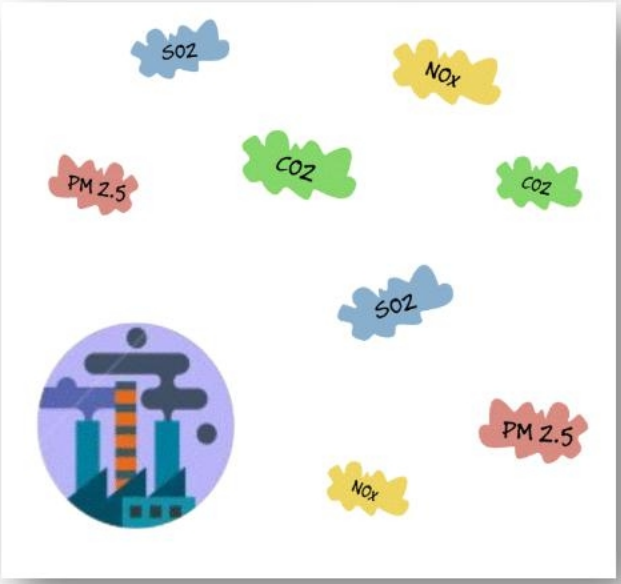
Air Quality Index Prediction

By -

Ekansh Kapoor (1828240)

INTRODUCTION

Air Quality Index (AQI)



AQI helps in understanding the level at which air is polluted and the associated health effects that might concern.

EPA calculates the AQI for five major air pollutants : ground-level ozone, particulate matter, carbon monoxide, sulfur dioxide, and nitrogen dioxide.

For each of these pollutants, EPA has established national air quality standards to protect public health.

The EPA has developed the pollutant standard index (PSI) for introducing consistency in providing information regarding the air quality throughout the US. The system is based on a scale of 0-500.

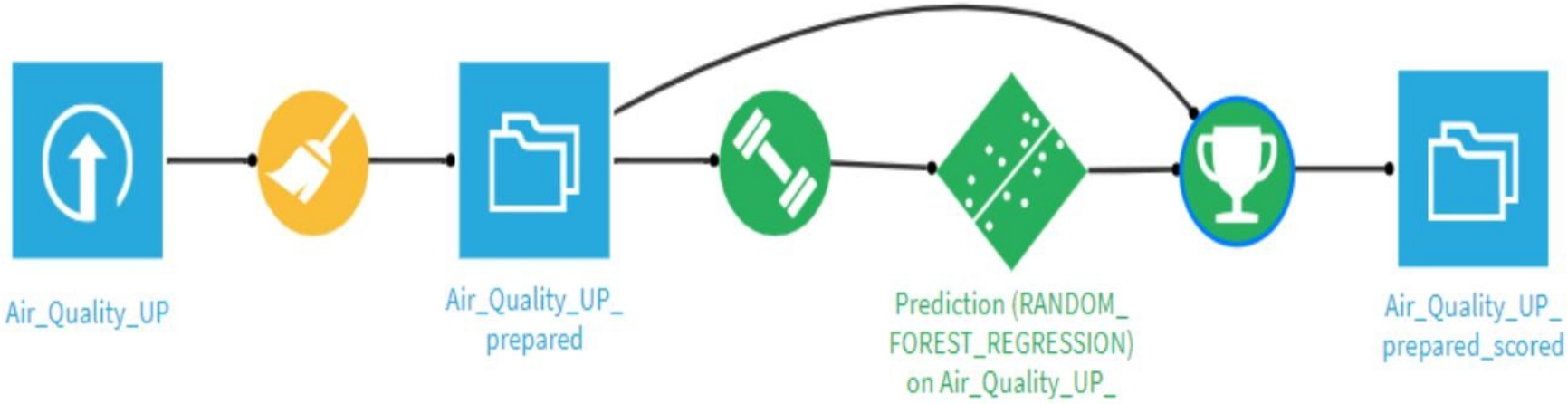
OBJECTIVE

The Air quality index aims to help the public easily understand the air quality and protect people's health.

The project aims to achieve the following:

- Inform public regarding overall status of air quality through a summation parameter that is easy to understand
- Inform citizens about associated health impacts of air pollution exposure; and
- Rank cities/towns for prioritizing actions based on measure of AQI.

MODEL


Air_Quality_UP
Air_Quality_UP_
prepared
Prediction (RANDOM_
FOREST_REGRESSION)
on Air_Quality_UP_
Air_Quality_UP_
prepared_scored

Data Info & Features Used In model

1. Data Consisted of 8 columns (SO2 , NO2 , RSPM , SPM , Date ,Newdate ,Location , State)

2. Some of the Data rows had NaN in it(Data Missing) so it was appropriate to drop those columns

3. As it is a Historical Data Regression Model were used.

4. Target variable AQI was not present in the data set it had to be created later on

FEATURES

SO2, NO2 ,RSPM ,SPM , DATE were the columns with higest correlation with our Target Variable.

Model Information

AIR QUALITY INDEX

Random Forest Regression model

Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.

AQI HEALTH MESSAGE

Air Quality Index Levels of Health Concern	Numerical Value	Meaning
Good	0 to 50	Air quality is considered satisfactory, and air pollution poses little or no risk.
Moderate	51 to 100	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
Unhealthy	151 to 200	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects.
Very Unhealthy	201 to 300	Health warnings of emergency conditions. The entire population is more likely to be affected.
Hazardous	301 to 500	Health alert: everyone may experience more serious health effects.

Tools Used:

- Dataiku 5.1 = End to End Data Science Studio
- Python, HTML, CSS, PHP
- Designmodo.com
- Jupyter Notebook
- Goggle Colaboratory
- Putty app, gen
- Anaconda Navigator
- Django
- Amazon web services
- Bracket App

Costs

List new projections of costs

- Include original estimates
  - Understand source of differences in these numbers – be ready for questions

If there are cost overruns

- Summarize why
- List corrective or preventative action you've taken
- Set realistic expectations for future expenditures

Technology

List technical problems that have been solved

List outstanding technical issues that need to be solved

- Summarize their impact on the project

List any dubious technological dependencies for project

- Indicate source of doubt
- Summarize action being taken or back up plan