



PROJECT - 1

AIR QUALITY PREDICTION

AIR QUALITY INDEX



An air quality index (AQI) is used by government agencies to communicate to the public how polluted the air currently is or how polluted it is forecast to become. AQI information is obtained by averaging readings from an air quality sensor, which can increase due to vehicle traffic, forest fires, or anything that can increase air pollution. Pollutants tested include:

PM2.5

Particulate Matter
(2.5 micrometres)

PM10

Particulate Matter
(10 micrometres)

O₃

Trioxygen

SO₂

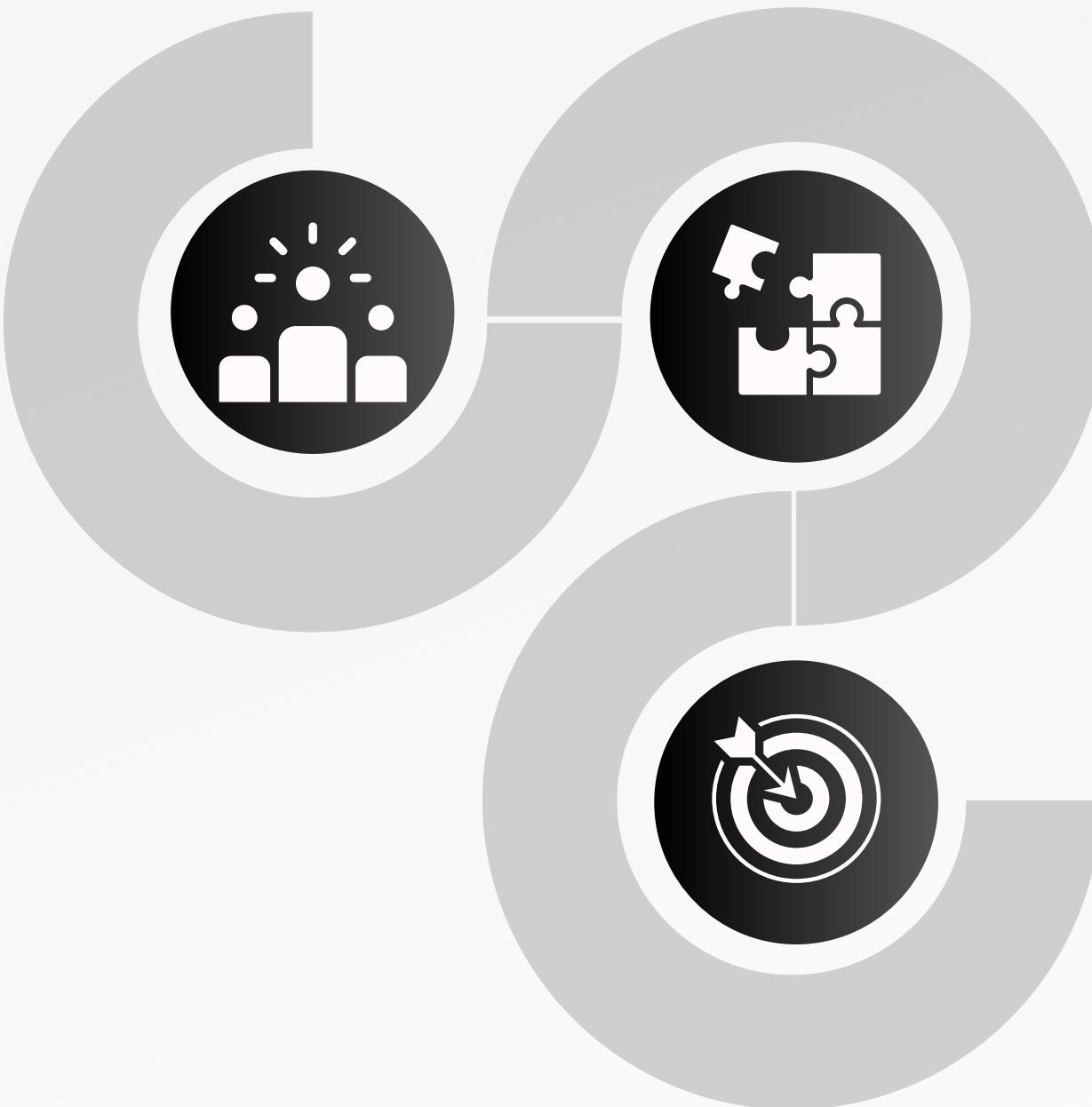
Sulfur Dioxide

CO

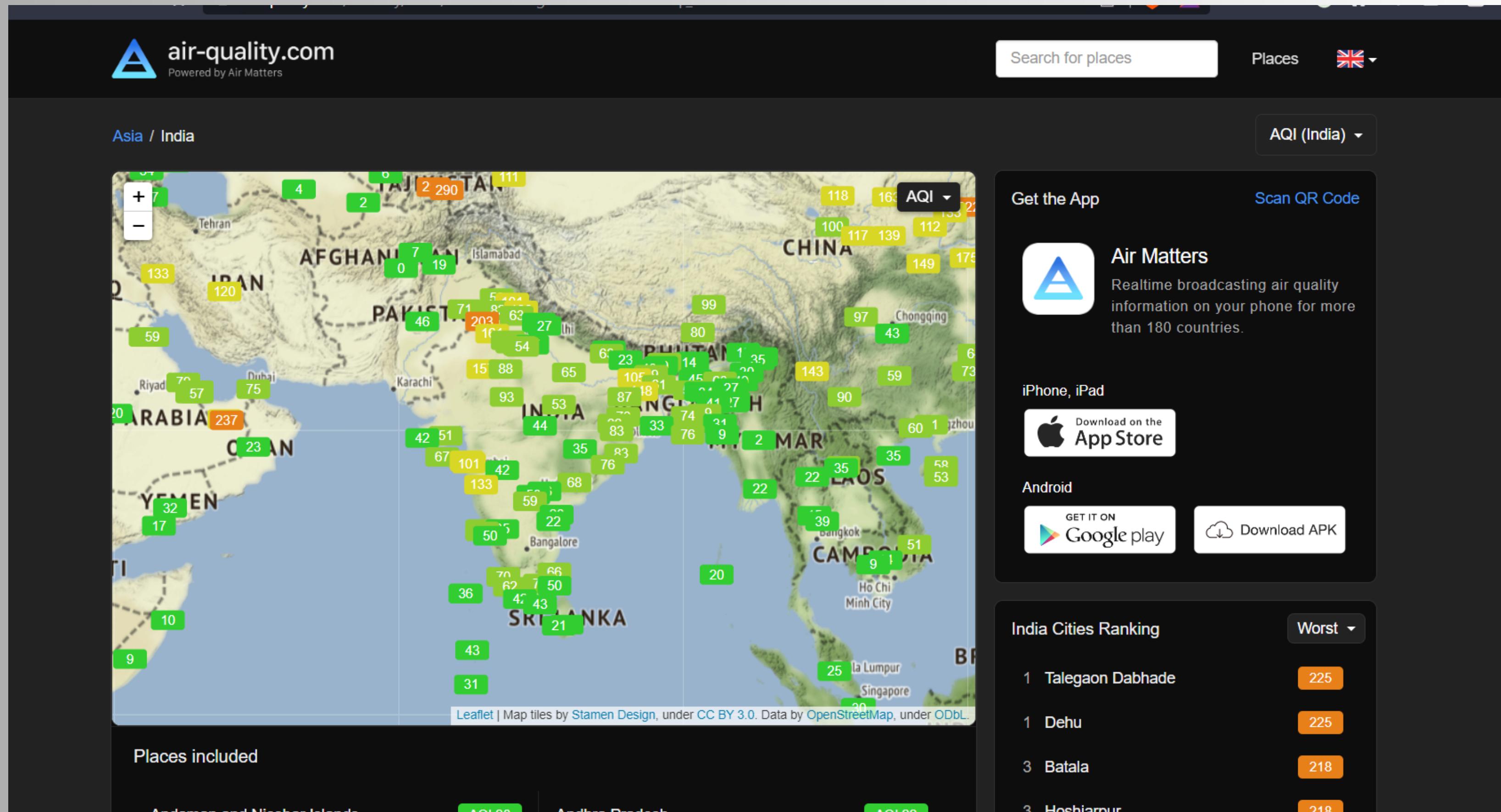
Carbon Monoxide

ARCHITECTURE DIAGRAM

- 01** Web Scrapping
- 02** Exploratory Data Analysis (EDA)
- 03** Feature Engineering
- 04** Machine Learning Model Training
- 05** Model Evaluation and Deployment



WEB SCRAPING SITE



TECH STACK USED

BeautifulSoup

Beautiful Soup is a Python library used for web scraping. It allows you to parse HTML and XML documents, navigate the parse tree, and extract data from the documents in a convenient way. Web scraping is the process of extracting information or data from websites.

KNN Imputer

In order to fill-in for the missing values in our dataset we used the KNN Imputer that estimates the missing values based on the KNN algorithm.

PowerTransformer

We used the PowerTransformer class from `sklearn.preprocessing` in order to scale the numerical features in our dataset.

BaggingRegressor

After running our dataset on various models, we found out that BaggingRegressor is performing the best.

PROBLEM FACED

- 01** Identifying a reliable source of air quality data for urban areas that permits the users to extract data via web-scraping.
- 02** Cleaning the extracted data and handling the missing values in it, as well as scaling our data in order to get accurate predictions from our model.
- 03** Finding out which model performs the best on our dataset to predict target feature accurately.

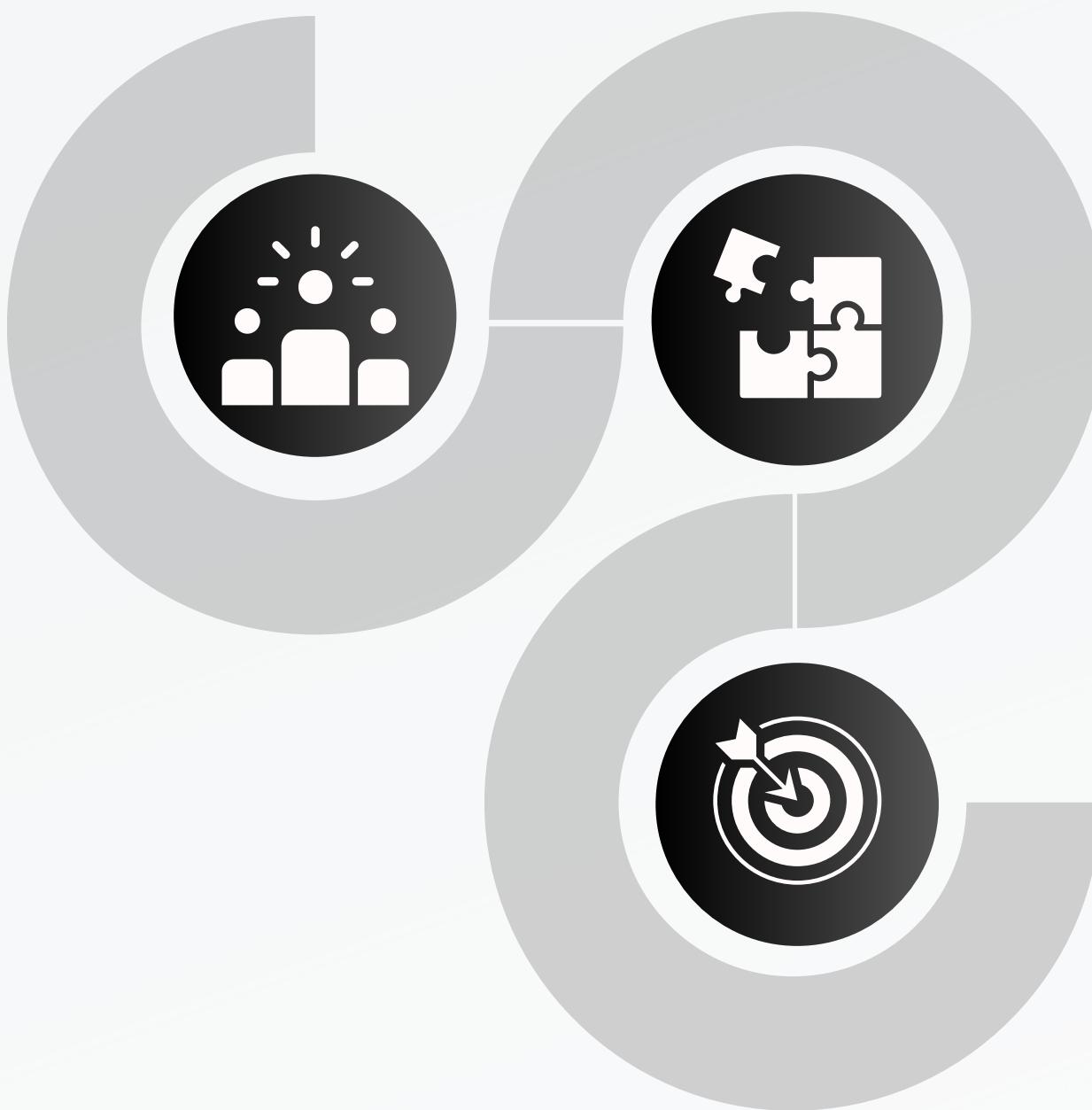


SOLUTIONS

- 01** Researched and found out a website that meets the project requirements.

- 02** Handled the missing values using Iterative Imputing techniques. Also worked with various feature scaling techniques to find out what works the best.

- 03** Trained and tested our dataset on different models and based on the MSE values got the model that works the best.



THANK YOU