Project Report

Title: Predicting Health Impact Classes from Air Quality Data

---

Problem Statement

Air pollution is a major public health concern worldwide, leading to severe health complications and reduced quality of life. This project aims to develop a machine learning model that predicts health impact classes based on air quality metrics. The specific health impact classes include Very High, High, Moderate, Low, and Very Low, which represent the severity of health risks associated with varying levels of air pollution.

Aligned Sustainable Development Goals (SDGs)

This project aligns with **SDG 3: Good Health and Well-being**, which emphasizes ensuring healthy lives and promoting well-being for all at all ages. By predicting health impacts from air quality data, the project supports public health interventions and policies aimed at reducing air pollution and its adverse health effects.

Proposed Solution

To tackle this problem, we employed various machine learning algorithms, including:

1. **Random Forest Classifier**: An ensemble learning method that uses multiple decision trees to improve classification accuracy.
2. **Support Vector Machine (SVM)**: A supervised learning model that classifies data by finding the hyperplane that best separates different classes.
3. **XGBoost**: An efficient implementation of the gradient boosting framework designed for speed and performance.

The solution involves the following steps:

- Data preprocessing, including handling missing values and encoding categorical variables.
- Splitting the dataset into training and testing sets.
- Training and evaluating the models using the training set.
- Generating performance metrics to assess the effectiveness of each model.

Impact

The successful implementation of this project provides critical insights into the relationship between air quality and health outcomes. By effectively predicting health impact classes, this model can assist public health officials and policymakers in:

- Making informed decisions to mitigate air pollution.
- Designing targeted health interventions to protect vulnerable populations.
- Raising awareness about air quality issues among communities.

Internship Acknowledgment

This project was completed as part of the **IBM SkillsBuild internship**, where I gained practical experience in applying machine learning techniques to real-world problems. The internship provided valuable insights into data analysis and model development, enhancing my skills in leveraging AI for impactful solutions.