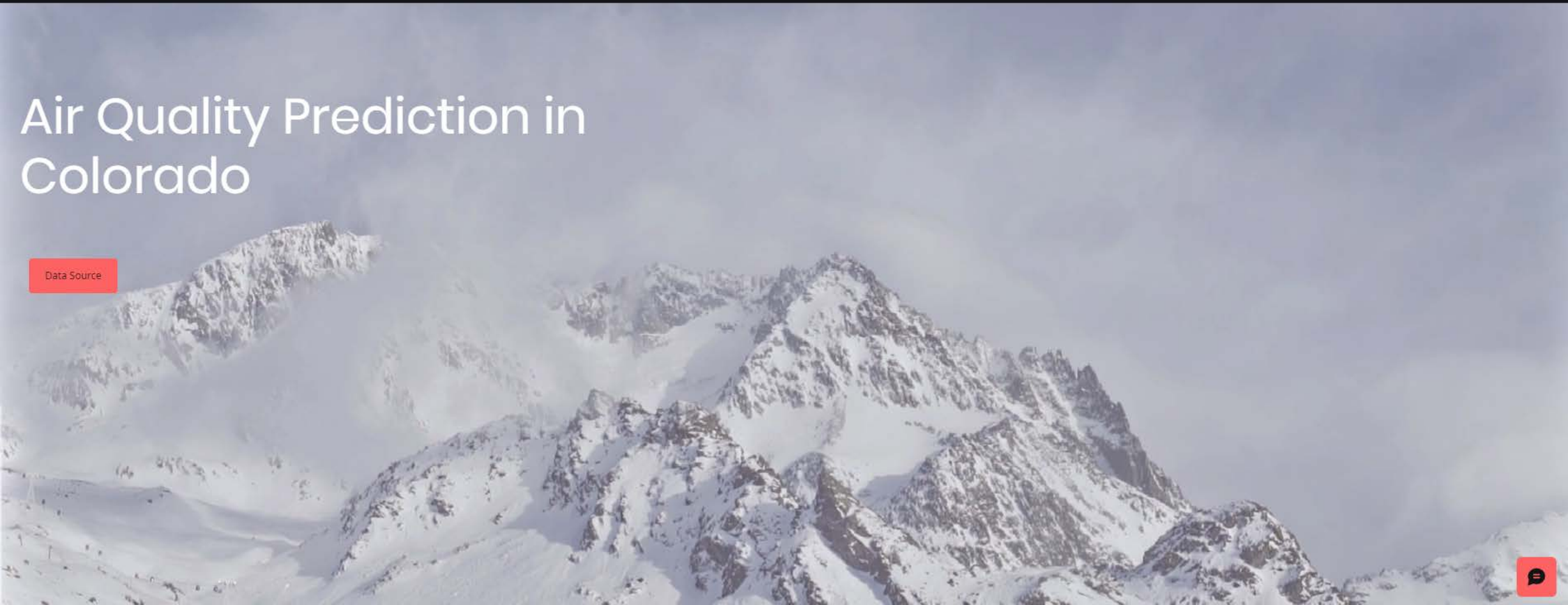


Air Quality Prediction in Colorado

Data Source

Project Goals

The objective of this project is to develop a machine learning model that predicts air quality levels (classified as Good, Poor, Moderate, and Satisfactory) at various locations in Colorado State.

Data Collection and Preprocessing

The project uses pollutant data collected from the EPA's AQS API, followed by cleaning and removing outliers. Missing values are handled, and time-based features are created to enhance the dataset's usability for analysis.

Feature Engineering

New features are generated to improve model performance, including pollutant interactions, location-based data, and temporal variables such as the time of day and season to capture relevant trends.

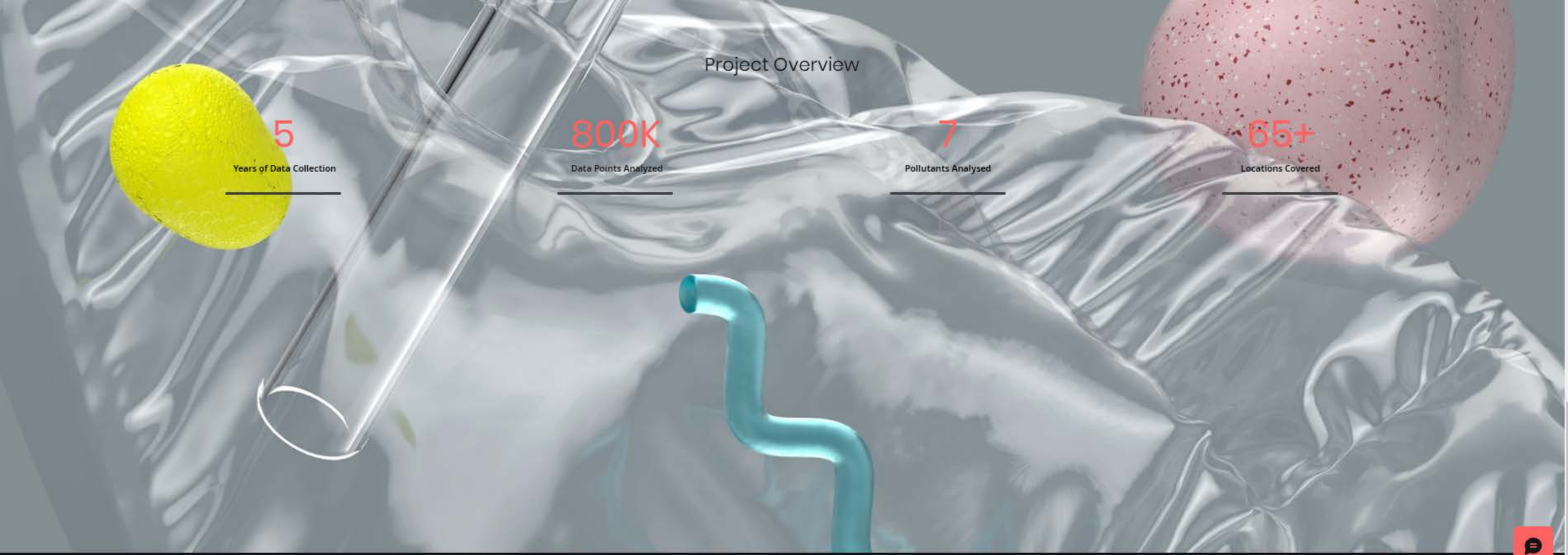
Model Training

Multiple machine learning models, such as Logistic Regression, Decision Trees, and Random Forests, are trained on the pre-processed data to predict air quality levels at various locations in Colorado.

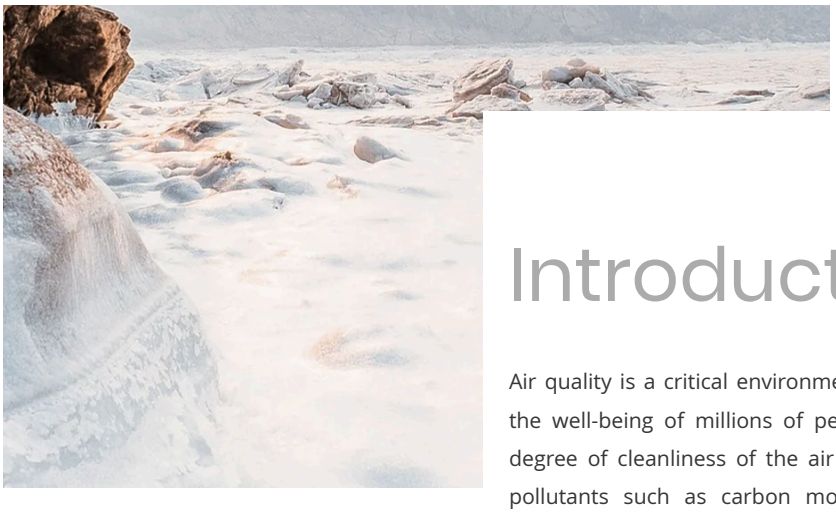
Model Optimization

Hyperparameter tuning techniques like grid search and random search are employed to refine model accuracy and stability, ensuring optimal performance for real-time air quality predictions.


Data Collection Method
The dataset used for this project will be collected via the AQS API provided by the United States Environmental Protection Agency (EPA). This API allows access to comprehensive air quality monitoring data, including pollutants and meteorological factors, collected across the U.S.
API Link


Project Overview
5
Years of Data Collection
800K
Data Points Analyzed
7
Pollutants Analysed
65+
Locations Covered

Colorado Air Quality Prediction



Introduction

Air quality is a critical environmental and public health issue, influencing the well-being of millions of people around the world. It refers to the degree of cleanliness of the air we breathe, which can be impacted by pollutants such as carbon monoxide, sulfur dioxide, and particulate matter. These pollutants, often originating from industrial emissions, transportation, and natural events like wildfires, can have severe consequences on human health and the environment. In Colorado, with its diverse landscapes ranging from bustling cities to vast rural areas, monitoring air quality is vital for ensuring a healthy living environment. This project aims to predict air quality levels at various locations across Colorado using machine learning models and data collected from the EPA's Air Quality System (AQS) API.

Significance of the Project

The significance of air quality prediction cannot be overstated. Poor air quality is linked to a wide array of health issues, from respiratory problems and cardiovascular diseases to developmental issues in children and premature death. Colorado's urban centers, like Denver, are particularly vulnerable due to their higher population density and traffic congestion, while rural areas face threats from agricultural activities and wildfires. Predicting air quality allows for timely interventions, such as issuing health advisories or enforcing stricter pollution controls. By identifying areas at risk of poor air quality, we can mitigate the negative health impacts on the state's residents, making this a crucial tool for both individuals and policymakers.

Air quality affects everyone, but certain groups are more vulnerable than others. Children, the elderly, and individuals with pre-existing health conditions, such as asthma or heart disease, are at greater risk of experiencing adverse effects from pollution. Additionally, those who live near industrial areas or heavily trafficked roads are more exposed to harmful pollutants. The project will serve not only the general population but also environmental agencies, public health organizations, and policymakers who need to base decisions on accurate and timely air quality data. With the ability to predict air quality levels, these stakeholders can take preventive action, ensuring that vulnerable communities

Despite the availability of air quality monitoring systems like the EPA's AQS API, gaps remain in our ability to make accurate, real-time predictions across different regions. Existing systems provide historical data and current air quality measurements but often lack predictive capabilities, especially in regions with rapidly changing environmental conditions. Moreover, much of the public remains unaware of how daily air quality fluctuations can impact their health. This project seeks to fill those gaps by developing a machine learning model that not only predicts air quality levels but also makes the information accessible to the public through an interactive web application. This ensures that communities are better informed and prepared to respond to poor air quality in real-time.

Several initiatives have been undertaken globally to monitor air quality, including the development of sophisticated air pollution models. However, many models focus on large-scale predictions without accounting for local variances. In Colorado, where geographical and climatic diversity can cause air quality to change rapidly, there is a need for more localized, accurate predictions. This project's focus on predicting air quality across specific Colorado locations such as Denver, Welby, and Elk Springs will provide a more nuanced understanding of pollution levels. By combining pollutant data with machine learning techniques, this project hopes to offer more precise predictions and actionable insights for improving air quality.

Questions Related to the Dataset

- What are the primary pollutants contributing to poor air quality in Colorado?
- How do pollutant levels vary across different locations within Colorado?
- Can seasonal trends be identified in the pollutant data?
- Which pollutants have the most significant impact on air quality levels?
- How does air quality in urban areas like Denver compare to rural areas like Elk Springs?
- What time of day typically experiences the worst air quality in Colorado?
- Are there any correlations between temperature and air quality in different regions of the state?
- How does traffic congestion in Colorado cities affect air pollution levels?
- Can machine learning models predict sudden drops in air quality due to wildfire events?
- How can real-time air quality predictions be used to inform public health advisories effectively?

Colorado Air Quality Prediction

The Team

Who We Are

Sanjay Mythili

Student, MS in Data Science @ CU Boulder

I am a Data Science graduate student at the University of Colorado Boulder, passionate about leveraging data, technology, and human insight to drive innovation. With a strong focus on Data Science and Artificial Intelligence, I have explored how data can fuel impactful solutions. I have previously published 3 research papers, and my areas of interest span computer vision, Natural Language Processing, and Generative AI.

720-934-3029 sanjay.mythili@colorado.edu

Dhanush Chandrasekar

Student, MS in Data Science @ CU Boulder

A performance and goal driven Electrical and Electronics Engineering professional with core skills and data science/ programming knowledge, seeking an opportunity to work in a challenging environment that leverages AI and data to build and optimize Electrical systems using machine learning algorithms and related areas, thereby enhancing individual professional growth aligned to the vision of the organization.

720-671-5293 dhanush.chandrasekar@colorado.edu

Shiva Sriram

Student, MS in Data Science @ CU Boulder

I am an aspiring data scientist pursuing a Master's degree in Data Science at the University of Colorado Boulder, with a strong interest in data analytics, machine learning, and predictive modeling. I'm keen on developing impactful data-driven solutions, leveraging my skills in Python, R, SQL, and cloud computing. I have a passion for working with large datasets, building statistical models, and data visualization using tools like Seaborn and Tableau. With strong communication skills and critical thinking, I excel in solving complex problems efficiently while expanding my technical expertise.

720-239-2610 shiva.sriram@colorado.edu