Climate Data Analysis and Prediction Web Application

Matt Martinez

The purpose of this project was to explore United Nations Sustainable Development Goal #13: “Take urgent action to combat climate change and its impacts.” This project provides access to an interactive tool for visualizing historical climate trends and exploring model-based future projections of key climate indicators. The motivation stems from the growing need to understand and communicate the impacts of climate change, enabling users to explore temperature and CO2 data dynamically.

The application allows users to select a geographical location, specify a historical date range, and define a future time horizon (1-50 years) for predictions. The application then fetches, processes, and visualizes historical US National Average Temperature, Global CO2 concentrations, and selected location-specific mean temperatures. It also generates and displays future predictions for national temperature and CO2.

Datasets

1. US National Average Temperature: Yearly average temperature data for the United States from NOAA accessed through The New York Post’s github. It provides monthly temperature data from 1895-2019 for the Lower 48 states. They calculated annual mean temperature trends using linear regression.
2. Global CO2 Concentration: Yearly atmospheric CO2 concentration data in parts per million (ppm) from 1959-2024 scraped from a NOAA dataset on Statista
3. Location-Specific Weather: Daily mean temperature data fetched from the Open-Meteo Historical Weather API based on user-selected latitude, longitude, and date range.

Daily weather data from the Open-Meteo API is processed and aggregated (daily, monthly, or yearly) for visualization based on the selected date range. Linear regression is applied to historical US national average temperature vs. year and global CO2 vs. year to identify long-term trends. Future Global CO2 is predicted using polynomial regression trained on the full historical CO2 dataset. Future US national temperature is predicted using a SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors) model. This

model is trained on the full historical national temperature data, incorporating the predicted future CO2 values as an external influencing factor.

The analysis of historical data consistently reveals statistically significant upward trends in both US National Average Temperature and Global CO2 concentrations. A strong positive correlation (typically $r > 0.8$, depending on the exact data range used) is observed between national temperature and CO2, underscoring their interconnectedness. The application visualizes historical climate data and generates model-based predictions. The analysis confirms established long-term warming trends and rising CO2 levels, with a strong linkage between them. The SARIMAX and Polynomial Regression models project these numbers to continue increasing, highlighting the ongoing nature of climate change. This tool allows users to explore historical patterns and understand potential future trajectories.

Key challenges include the fragility of the Statista web scraper (if used for CO2 data updates), potential changes in the Open-Meteo API, and the need for periodic review/retuning of predictive models as new data becomes available. Future enhancements could include integrating more climate datasets and implementing alternative or more advanced prediction models.