Project Report:

Mouna Loa CO2 Prediction

1. Project Objective

The objective of this project is to analyze the atmospheric Carbon Dioxide (CO2) levels collected at the Mouna Loa Observatory and develop a predictive model to forecast future CO2 concentrations. A Streamlit web application has been built to visualize the data and the model's forecasts.

2. Methodology

2.1 Data Preprocessing

Source: `archive.csv` containing monthly CO2 measurements.

Handling Missing Values: Missing values in the `Carbon Dioxide (ppm)` column (represented as -99.99) were replaced with `NaN` and then filled using backward fill (`bfill`) followed by forward fill (`ffill`).

Date Conversion: A 'Date' column was created from 'Year' and 'Month' columns, and set as the index for time series analysis.

2.2 Exploratory Data Analysis (EDA)

Visualization: The time series plot shows a clear upward trend and seasonality in CO2 levels.

Stationarity: The Augmented Dickey-Fuller (ADF) test was performed. The p-value indicates non-stationarity, which is expected for data with a strong trend.

Decomposition: Seasonal decomposition revealed the underlying trend, seasonal, and residual components.

2.3 Model Selection: SARIMAX

A **Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX)** model was chosen due to the data's seasonality and trend.

Parameters:

Order (p, d, q): `(1, 1, 1)`

Seasonal Order (P, D, Q, s): `(1, 0, 1, 12)` (Monthly seasonality)

These parameters allow the model to capture:

AR(1): Dependence on the previous term.

I(1): Differencing to handle trend.

MA(1): Dependence on the previous error term.

Seasonal Component: Captures the yearly cycle of CO_2 fluctuation.

2.4 Model Evaluation

The model was evaluated using the last 20% of the data as a test set.

Metrics:

MAE (Mean Absolute Error): Measures the average magnitude of errors.

RMSE (Root Mean Squared Error): Measures the square root of the average squared differences between prediction and actual observation.

3. Streamlit Application

A user-friendly web interface was developed using Streamlit (`app.py`).

Features:

1. **File Upload**: Users can upload the `archive.csv` dataset.

2. **Visualization**: Interactive plots of raw data and seasonal decomposition.

3. **Forecasting**:

   - displays the SARIMAX model forecasts agains the test set.

   - Allows users to specify the number of future months to forecast.

4. **Metrics**: Displays MAE and RMSE to assess model performance.

4. Conclusion

The SARIMAX model successfully captures the trend and seasonality of the Mouna Loa CO_2 data. The Streamlit application provides an accessible tool for monitoring and forecasting these critical environmental metrics.