Import pandas as pd

Import numpy as np

Import matplotlib.pyplot as plt

Import seaborn as sns

From sklearn.model\_selection import train\_test\_split

From sklearn.ensemble import RandomForestRegressor

From sklearn.metrics import mean\_squared\_error, r2\_score

# Load dataset

Df = pd.read\_csv(“air\_quality\_data.csv”) # Replace with your dataset file

# Display basic info

Print(df.head())

Print(df.info())

# Handle missing values

Df.fillna(df.mean(numeric\_only=True), inplace=True)

# Feature and target selection

X = df[[‘PM2.5’, ‘PM10’, ‘NO2’, ‘CO’, ‘SO2’, ‘O3’]] # Adjust based on your dataset

Y = df[‘AQI’] # Target variable

# Split dataset

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Initialize and train model

Model = RandomForestRegressor(n\_estimators=100, random\_state=42)

Model.fit(X\_train, y\_train)

# Predict

Y\_pred = model.predict(X\_test)

# Evaluation

Mse = mean\_squared\_error(y\_test, y\_pred)

Rmse = np.sqrt(mse)

R2 = r2\_score(y\_test, y\_pred)

Print(f”RMSE: {rmse}”)

Print(f”R² Score: {r2}”)

# Visualization

Plt.figure(figsize=(10, 6))

Sns.scatterplot(x=y\_test, y=y\_pred)

Plt.xlabel(“Actual AQI”)

Plt.ylabel(“Predicted AQI”)

Plt.title(“Actual vs Predicted AQI”)

Plt.grid(True)

Plt.show()