Final Portfolio Project Proposal

Module: 5CS037 - Concepts and Technologies of AI
Student Name: Milan Sherpa
University ID: 2462987

1. Regression Task: Global Air Pollution Prediction

- **Research Question:** How accurately can we predict the specific concentration of PM2.5 pollutants in various global cities based on other atmospheric pollutant levels like Carbon Monoxide (CO), Ozone, and Nitrogen Dioxide (NO2)?

- **Dataset Description:** The "air pollution dataset.csv" contains over 23,000 records of air quality metrics from cities worldwide. Key independent variables include CO AQI Value, Ozone AQI Value, and NO2 AQI Value, with the **PM2.5 AQI Value** serving as the continuous target variable. A primary preprocessing challenge will be handling geographic variations and normalizing values across different countries.

- **Connection to SDG:** This project aligns with **SDG 11: Sustainable Cities and Communities**. By forecasting PM2.5 levels, the analysis supports urban air quality management and helps in creating healthier, more sustainable living environments.

2. Classification Task: E-commerce Customer Churn Prediction

- **Research Question:** Can machine learning models accurately classify whether an e-commerce customer will churn based on their engagement metrics and purchasing behavior?
- **Dataset Description:** The "ecommerce_customer_churn_dataset.csv" includes detailed customer profiles with features such as Membership Years, Login Frequency, Session Duration, and Cart Abandonment Rate. The target variable is **Churned** (Binary: 0 for retained, 1 for churned). Foreseen challenges include addressing missing values in social media engagement scores and managing potential class imbalances between churned and active users.

- **Connection to SDG:** This project aligns with **SDG 9: Industry, Innovation, and Infrastructure**. Utilizing AI to predict customer behavior fosters innovation in digital infrastructure and helps businesses build more resilient and efficient economic models.