Project Idea: This project aims to develop predictive models for critical environmental and socio-economic indicators related to climate change for various countries. The indicators, spanning from 1960 to 2022, include urban population growth, CO2 emissions, freshwater withdrawals, agricultural land usage, and other relevant metrics. The primary goal is to forecast these indicators for future years, allowing us to analyze the long-term impacts of climate change on these variables. Additionally, the project seeks to explore the correlation between these predicted indicators and the Global Hunger Index (GHI) in each country, offering insights into how climate change may influence hunger and food security, aligning with SDG 13 (Climate Action) and SDG 2 (Zero Hunger).

Relevance to Sustainable Development Goals (SDGs): This project is directly relevant to the United Nations Sustainable Development Goals, particularly SDG 13 (Climate Action) and SDG 2 (Zero Hunger). By predicting future trends in environmental and socio-economic indicators, the project contributes to understanding the potential future impacts of climate change on global hunger. For instance, rising CO2 emissions and decreasing agricultural land could exacerbate food shortages, leading to higher GHI scores. By identifying these trends early, the project can inform policymakers and stakeholders, enabling them to take proactive measures to mitigate these effects and promote sustainable development.

Literature Examples:

1. **Machine Learning in Agriculture: Crop Yield Prediction by Hames Sherif**
    - *Summary:* This project focuses on improving food security in regions with desert and semi-arid climates in Africa by identifying environmental and economic factors affecting crop yields. The goal is to create a predictive model that estimates the yield of staple crops in specific countries within these regions. The variability in crop yield makes management challenging, and this research aims to identify key influencing factors to help farmers and policymakers make informed decisions on crop management, resource allocation, and agricultural strategies.
    - *Relevance :*This literature example is highly relevant to your project as it addresses similar challenges in predicting agricultural outcomes influenced by environmental and socio-economic factors. It provides a foundation for understanding how predictive models can be used to estimate crop yields and manage food security in regions with varying climate conditions. This aligns with your project's goals of forecasting indicators related to climate change and assessing their impact on global hunger, thus contributing to SDG 13 and SDG

2. **LSTM-Based Time Series Forecasting for Climate and Environmental Data (International Journal of Climate Change Strategies, 2021):**
   - *Summary:* This paper focused on using Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), to predict long-term climate and environmental trends. LSTM models were particularly effective in capturing the temporal dependencies and patterns in the data, making them ideal for time-series forecasting tasks like those in our project.
   - *Relevance:* The use of LSTM models in this paper aligns with our approach of employing deep learning for time-series predictions. The paper provides valuable insights into the implementation and optimization of LSTM models, which will be crucial for accurately forecasting our selected indicators.

Describe Your Data: The dataset for this project is sourced from the [Humanitarian Data Exchange]() and covers a range of environmental and socio-economic indicators from 1960 to 2022. Key indicators include urban population growth, CO_2 emissions (from various fuel sources), freshwater withdrawals, agricultural land usage, and more. The data is in CSV format and consists of time-series data for multiple countries.

- **Data Size:** The dataset is extensive, with several indicators recorded annually for multiple countries over a span of more than six decades.
- **Preprocessing Steps:**
  - *Handling Missing Values:* Techniques such as interpolation or filling with mean/mode values will be employed.
  - *Normalization:* Since the indicators have varying scales, normalization will be applied to bring all features to a comparable range.
  - *Feature Extraction:* We will create derived features that capture trends, seasonality, or other relevant patterns in the data to improve model performance.

Approach (Machine Learning or Deep Learning): Given the temporal nature and complexity of the data, we will adopt a hybrid approach, utilizing both deep learning and machine learning models.

- **Deep Learning:**
  - *LSTM with Keras:* LSTM networks are well-suited for time-series forecasting due to their ability to remember long-term dependencies. We'll

use LSTM models to predict indicators like CO2 emissions, urban population growth, and freshwater withdrawals, as these involve complex temporal relationships.

- **Machine Learning:**
  - *Regression Models:* We will employ a variety of regression models, including Random Forest, AdaBoost, ExtraTrees, Gradient Boosting, and SVR, to predict indicators where deep learning may not be necessary or to provide ensemble predictions for better accuracy.
  - *Justification:* The combination of machine learning and deep learning allows us to handle the different complexities of our data. LSTM will capture long-term dependencies in the time-series data, while ensemble learning methods in machine learning will provide robust predictions by reducing the variance and bias.