

Impact of Climate and Environmental Factors on Crop Yields in the World

Problem Description

The global agriculture landscape plays an important role in the 21st century. Today there are 8 billion of us in the world, with the human population expected to grow at a rapid pace until the end of the century. Agriculture, as the bedrock of food security and a pivotal contributor to global economies, faces an unprecedented era of uncertainty.

The challenge of ensuring food security and eradicating hunger is foundational to the attainment of global sustainable development goals. However, the growing sensitivity of food production to the impacts of climate and environmental factors present a challenge. The intricate relationships between rising temperatures, shifting precipitation patterns, pesticides, greenhouse gas emissions and other environmental factors have brought forth a pressing question – How will the world's agricultural systems adapt and respond to these multifaceted challenges?

Answering this question through predicting crop yields for the Top 10 produce in the world just might help address the challenges ahead.

Problem Significance

This project embarks on a crucial exploration, aiming to unravel the intricate web of factors shaping crop yields worldwide. Its significance extends beyond agricultural concerns, touching upon food security, socioeconomic stability, and environmental sustainability. As global populations continue to rise and climate change accelerates, understanding the consequences of these factors become more important.

The outcome of this research hopes to inform agricultural practices, guide policy formulation, prepare us against the uncertainties of the future, ultimately contributing to the welfare of nations and the preservation of our planet.

Research Questions

1. How do variations in temperature and precipitation patterns correlate with changes in crop yields across different regions and time periods?
2. What is the relationship between pesticide usage and crop yield fluctuations?
3. How do greenhouse gas emissions impact climate variables such as temperature and precipitation, and how do these changes subsequently affect crop yields?
4. Are there specific climate and environmental factors, beyond temperature and precipitation, that significantly influence crop yields?
5. How can machine learning models effectively predict crop yields based on historical climate and environmental data, and which machine learning algorithms perform best for this task?

Planned Approach

1. **Data Collection and Cleaning:** Gather historical climate data (e.g., temperature and precipitation), greenhouse gas emissions, pesticide usage, as well as crop yield data for various regions around the world. Ensure that the data is cleaned and fit for use.

2. **Exploratory Data Analysis (EDA):** Study and explore dataset to apprehend their predominant traits, discover patterns, locate outliers, and identify relationships between variables. Perform descriptive statistical analysis on time series trend, autocorrelation, seasonality, and stationarity.
3. **Feature Engineering:** Calculate relevant features from the climate data, such as growing season length, heat stress days, and cumulative rainfall.
4. **Data Splitting:** Split your dataset into training, validation, and test sets. Time-based splitting might be appropriate to maintain the temporal order of the data.
5. **Model Selection and Development:** Choose appropriate machine learning methods to predict crop yields based on climate and environmental variables.
6. **Evaluation and Model Tuning:** Finding the optimal values of hyperparameters to maximise model performance.

Potential Machine Learning Methods

1. **Time Series Forecasting:**
 - **LSTM (Long Short-Term Memory):** LSTM networks are well-suited for time series forecasting tasks. Train model to predict future crop yields based on historical climate data.
2. **Regression:**
 - **Random Forest Regression:** Random forests can capture nonlinear relationships between climate variables and crop yields.
3. **Ensemble Methods:**
 - **Gradient Boosting:** Ensemble methods like XGBoost or LightGBM can effectively handle complex relationships and interactions in the data, providing accurate yield predictions.
4. **Geospatial Analysis:**
 - **Spatial Autoregressive Models:** If geospatial information is present, these models can capture spatial dependencies in crop yields across regions.
5. **Neural Networks:**
 - **Feedforward Neural Networks:** Design a feedforward neural network with hidden layers to capture complex relationships between climate variables and crop yields.

Data Sources

1. **Food and Agriculture Organisation (FAO):** <http://www.fao.org/home/en/>
 - Comprehensive repository of agricultural and climate-related data. This includes historical crop yield data, information on crop types, climate variables, and geographical information.
2. **World Bank - Country Climate and Development Report (CCDR):** [https://databank.worldbank.org/source/country-climate-and-development-report-\(ccdr\)](https://databank.worldbank.org/source/country-climate-and-development-report-(ccdr))
 - Wealth of data related to climate, development, and agriculture. Greenhouse gas emissions come from this source and would require data manipulation.
3. **World Bank Climate Knowledge Portal (Alternative):** <https://climateknowledgeportal.worldbank.org/download-data>
 - A range of climate-related datasets, including historical climate data and climate projections such as temperature and precipitation data.