PREDICTING THE IMPACT OF CLIMATE CHANGE ON GLOBAL CROP YIELDS USING MACHINE LEARNING

HAOQIAN ZHENG

INTRODUCTION

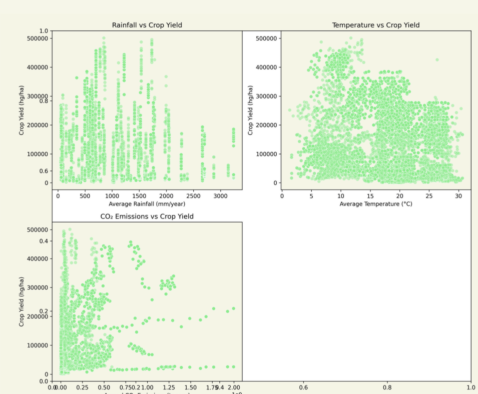
The impact of climate change on agriculture has emerged as a critical issue with profound implications for global food security.

This project focuses on predicting the impact of climate change on global crop yields focusing on potatoes and maize using machine learning techniques to uncover complex relationships between climate indicators and agricultural productivity.

DATA OVERVIEW

This project uses data from following sources: rainfall and temperature from the World Bank, crop yields from the FAO, and CO_2 emissions from Our World in Data. These datasets help explore how changes in climate, such as rainfall, temperature, and emissions, affect global crop production.

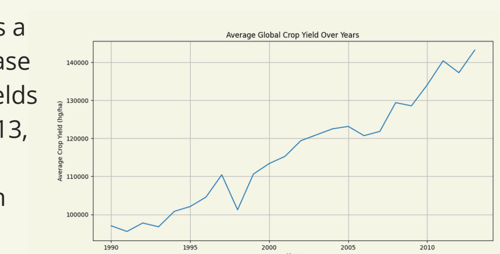
DATA VISUALIZATIONS

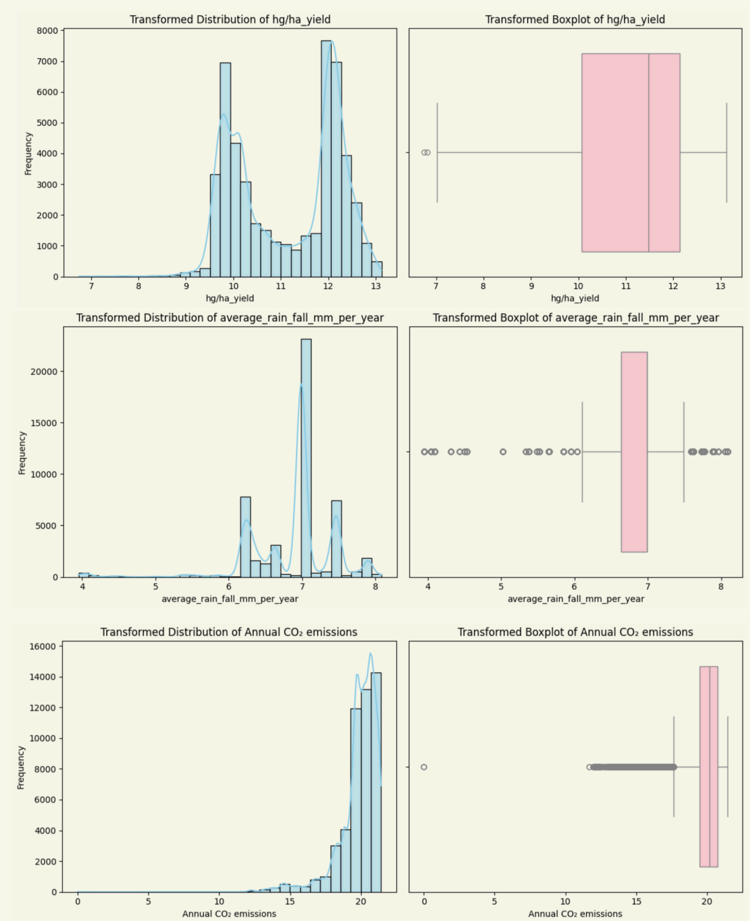


The plots illustrate relationships between rainfall, temperature, CO_2 emissions, and crop yield, showing varying trends and potential correlations.

Moderate rainfall and temperatures optimize yields, whereas extreme values lower productivity. CO2 impacts crop yields indirectly through complex climate interactions.

The graph shows a consistent increase in global crop yields from 1990 to 2013, indicating improvements in agricultural productivity.





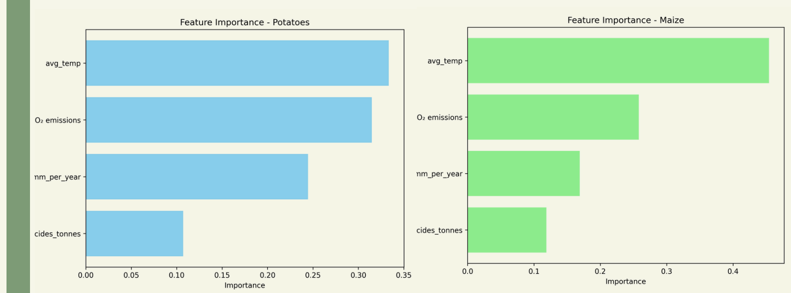
The plots show log-transformed distributions and boxplots for crop yield, rainfall, and CO_2 emissions. Log transformation reduces skewness, making data more symmetric and suitable for analysis.

- Crop Yield and Rainfall: Rainfall appears concentrated around specific ranges, with fewer extreme outliers, which may influence crop yield variability.
- Crop Yield and CO_2 Emissions: Despite normalization, CO_2 emissions remain skewed with several outliers, indicating possible indirect or non-linear impacts on crop yield.
- Rainfall and CO_2 Emissions: Both variables exhibit distinct patterns, suggesting their independent contributions to environmental conditions influencing crop productivity.

Log transformation reduces skewness, making data more symmetric and suitable for analysis.

FEATURE SELECTIONS

Feature selection identified average temperature and CO_2 emissions as key drivers for both maize and potato yields.



Rainfall showed moderate importance, while pesticide usage contributed the least. These insights underscore the dominant role of climate factors, particularly temperature, in influencing crop productivity across different agricultural systems.

STATISTICAL METHOD

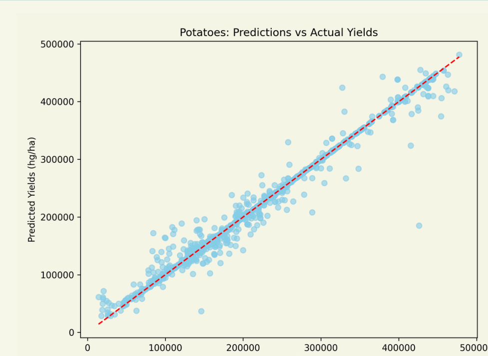
Predictive Modeling:

- Random Forest Regression: Used for predicting crop yields and determining variable importance.
- Logistic Regression: Applied to classify high vs. low yields based on the median.

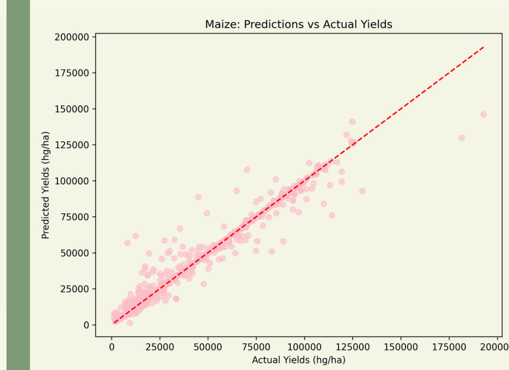
Model Evaluation:

- Metrics like MAE, RMSE, R^2, and MAPE were used for evaluating regression models.
- Accuracy, confusion matrix, and classification report for logistic regression.

RESULTS



The plot shows predicted vs. actual potato yields. Most points align near the red diagonal, indicating accurate predictions



The plot illustrates maize yield predictions versus actual values. Most points closely follow the diagonal, indicating strong prediction accuracy.

MODELS	RESULTS
RANDOM FOREST(POTATOES)	98.79%
RANDOM FOREST(MAIZE)	97.96%
LOGISTIC REGRESSION	50.93%

Random Forest models achieved high accuracy (98.79% for potatoes, 97.96% for maize), outperforming Logistic Regression (50.93%).

This highlights Random Forest's effectiveness in capturing complex relationships, making it more suitable for predicting crop yields under varying climate conditions.

CONCLUSIONS AND FUTURE WORK

Conclusions:

- Random Forest achieved the highest accuracy of 98.79% for potatoes and 97.96% for maize, outperforming Logistic Regression (50.93%).
- Temperature and CO_2 emissions emerged as the most significant predictors of crop yields.
- Random Forest effectively captures complex relationships, providing reliable predictions to inform agricultural planning and climate adaptation strategies.

Future work:

- Explore additional climate variables, advanced models (e.g., deep learning), and region-specific analyses to enhance prediction accuracy and resilience against climate impacts.

Carbon Cost:
Estimated CO_2 emissions: 1.632258349438576e-06 kg

Reference: World Bank Data Catalog. "Global Precipitation Anomaly Data." World Bank Data Catalog. Available: https://datacatalog.worldbank.org/search/dataset/0040276 , Food and Agriculture Organization (FAO). "Crops and Livestock Products Dataset." FAOSTAT. Available: https://www.fao.org/faostat/en/#data/QCL, Patel, R., "Crop Yield Prediction Dataset," Kaggle. Available: https://www.kaggle.com/datasets/patelris/crop-yield-prediction-dataset/data?select=yield.csv,Ritchie, H., and Roser, M., "CO_2 and Greenhouse Gas Emissions," Our World in Data. Available: https://ourworldindata.org/co2-emissions, Statista. "Global Precipitation Anomaly Statistics," Statista. Available: https://www.statista.com/statistics/1293084/global-precipitation-anomaly/