Climate Change Impact on Crop Resilience and Economic Outcomes Across the World

Team 7

Akshitha Komatireddy

Varshith Vuyyuru

Vishal Reddy Kota

Sai Pranav Beesetti

Kezia Shiny Pothumudi

Chandana Gangaraju

Mani Sai Bollam

George Mason University

AIT 582-002 Application of Metadata in Complex Big Data Problems (Fall 2024)

Professor: Lam Phung

October 7,2024

Project Milestone-1

Problem Statement:

The "The Real Harvest - Visualization & Statistics" dataset shows that understanding the complex relationship between soil health, climate conditions, and crop output across regions is difficult. The collection contains soil health indices, extreme weather events, and crop production statistics to help researchers study how soil quality affects agricultural output. Despite agricultural improvements, many places still have low crop yields due to poor soil health from excessive chemical fertilizer use, unsustainable agricultural practices, and unfavorable weather. This highlights the need for targeted treatments and sustainable farming to increase soil health and food security.

Droughts, floods, and heatwaves affect crop yield, which is a major concern. The collection helps identify trends and correlations between climatic events and crop production. Understanding these linkages helps policymakers and farmers determine crop susceptibilities to climatic variability and alter their strategies. Given the continual climate change that threatens global food production, agricultural systems need adaptation to be resilient.

This dataset helps improve discussions about sustainable farming and soil health management. Stakeholders can encourage crop rotation, cover cropping, and organic farming by emphasizing the relationship between soil health, climate, and agricultural productivity. Using data-driven insights to address these issues boosts agricultural output, environmental sustainability, and climate change resilience, benefiting farmers and consumers.

Motivation:

"The Real Harvest - Visualization & Statistics" dataset's extensive and complex insights regarding soil health, extreme weather events, and crop yield drive me to investigate it. This dataset allows researchers to apply analytical abilities to real-world agricultural problems, making the research both academically and socially significant. We want to contribute to food security and environmental sustainability discussions by studying agricultural productivity. This dataset will improve our understanding of agricultural systems and allow us to propose concrete ideas to help farmers adjust to climate change uncertainty. This research matches our data science academic and career goals.

Literature Review:

Agricultural research has provided substantial evidence of a link between healthy soil and fruit yields. Because it provides essential nutrients and allows for biological activity, healthy soil is fundamental for plant growth. Soil degradation reduces crop yields, according to studies, highlighting the need of sustainable farming practices including crop rotation and organic farming. Improving soil health is crucial for reducing the effects of climate change on agriculture and increasing food security.

Worldwide, agricultural systems are feeling the effects of more frequent and severe extreme weather events caused by climate change. Droughts, floods, and extremely hot or cold weather can significantly reduce agricultural yields, according to studies. If you want to come up with adaptive strategies, you need to understand their consequences. In order to safeguard food production against unpredictable weather patterns, recent studies advocate for the integration of climate resilience into agricultural planning.

The ability to analyze data is crucial in modern agriculture since it allows for the examination of complex relationships between soil quality, weather, and harvest success. Discoveries made possible by sophisticated analytical techniques can shed light on vast datasets, providing policymakers and farmers with valuable insights. In order to advance sustainable farming methods and ensure food availability in the face of environmental challenges, data-driven methodologies are essential.

Proposed Approach:

Climate change is seriously compromising environmental sustainability, economic stability, and food security by influencing global agricultural systems more and more. Rising temperatures, erratic precipitation patterns, and more severe weather events all help to create an unparalleled difficulty for agricultural output. By directly affecting crop yields, soil quality, and farming operation profitability, these climate-related elements could help to cause food shortages and poverty in susceptible regions.

Though world attempts to slow down climate change are commendable, it is crucial to quantify and understand how exactly these changing weather patterns affect agriculture. Furthermore, other adaptation techniques—such as crop rotation, water management, and technological innovations—offer varied degrees of resilience; yet, the efficiency of these methods depends on the locality and crop type.

This study aims to investigate in great detail the intricate connection between climatic variables (such as temperature, precipitation, and extreme weather events) and agricultural results (such as crop yield, soil health, and economic repercussions). It will also evaluate how agricultural adaptation plans would reduce these effects and identify the best approaches to sustain output in the face of increasing climate stress.

Research Questions.

1. How are Extreme weather events influencing the crop yield per(MT/HA) in various countries and regions ?

   Extreme weather events have a significant impact on the economy, thereby increasing the agricultural sector's susceptibility. The hypothesis posits that regions that are subjected to severe weather conditions may experience substantial economic losses as a result of the disruption of their infrastructure productivity and production in agriculture.

2. In what manner does irrigation, pesticide use, and the application of fertilizers affect the soil health and productivity of the crop across the regions?

   Climate change mitigation in food production involves regulating irrigation, herbicides, and fertilizers to improve crop resilience and yield. These approaches improve environmental prediction and management, leading to a more stable farming industry.

3. What is the impact of changes of mean temperature and total accumulative rainfall on crop production over various regions having different crop commodities?

Variations in average temperature and total precipitation can significantly impact agricultural output, affecting crop growth, yield, and overall quality. The variations in these modifications may depend on the specific crop being farmed.

4. What is the best solution that has been implemented by various geographical areas towards various challenges within the climatic changes?

   This theory suggests that it is impossible for one location to successfully adopt a climate change solution for another due to the unique environmental, social, and economic variables in each place.

Preliminary Results so far...

After going through the dataset, we have found quiet few null values across the columns. Not taking any chances we have algorithms to clean the data and load the null and missing values with median and mode values and performing the research with the cleaned data.

[1]:
```
import numpy as np
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
```

[2]:
```
data = pd.read_csv("C:/Users/manis/OneDrive/Desktop/Assignments/AIT - 582 project/climate_change_impact_on_agriculture_2024.csv")
data.head()
```

[2]:

	Year	Country	Region	Crop_Type	Average_Temperature_C	Total_Precipitation_mm	CO2_Emissions_MT	Crop_Yield_MT_per_HA	Extreme_Weather_Events	Irrigation_A
0	2001	India	West Bengal	Corn	1.55	447.06	15.22	1.737	8	
1	2024	China	North	Corn	3.23	2913.57	29.82	1.737	8	
2	2001	France	Ile-de-France	Wheat	21.11	1301.74	25.75	1.719	5	
3	2001	Canada	Prairies	Coffee	27.85	1154.36	13.91	3.890	5	
4	1998	India	Tamil Nadu	Sugarcane	2.19	1627.48	11.81	1.080	9	

[3]:
```
#Cleaning the Data

print("\nMissing values in each column:")
print(data.isnull().sum())
```

```
Missing values in each column:
Year                          0
Country                       0
Region                        0
Crop_Type                     0
Average_Temperature_C         0
Total_Precipitation_mm        0
CO2_Emissions_MT              0
Crop_Yield_MT_per_HA          0
Extreme_Weather_Events        0
Irrigation_Access_%           0
Pesticide_Use_KG_per_HA       0
Fertilizer_Use_KG_per_HA      0
Soil_Health_Index             0
Adaptation_Strategies         0
Economic_Impact_Million_USD   0
dtype: int64
```

```
[4]: #Filling Numerical and categorical Values with median and mode
numeric_cols = data.select_dtypes(include=['float64', 'int64']).columns
data[numeric_cols] = data[numeric_cols].fillna(data[numeric_cols].median())
categorical_cols = data.select_dtypes(include=['object']).columns
for col in categorical_cols:
    data[col] = data[col].fillna(data[col].mode()[0])
print("\nMissing values after cleaning:")
print(data.isnull().sum())
print("\nCleaned DataFrame:")
print(data.head())
```

```
Missing values after cleaning:
Year                           0
Country                        0
Region                         0
Crop_Type                      0
Average_Temperature_C          0
Total_Precipitation_mm         0
CO2_Emissions_MT               0
Crop_Yield_MT_per_HA           0
Extreme_Weather_Events         0
Irrigation_Access_%            0
Pesticide_Use_KG_per_HA        0
Fertilizer_Use_KG_per_HA       0
Soil_Health_Index              0
Adaptation_Strategies          0
Economic_Impact_Million_USD    0
dtype: int64


Cleaned DataFrame:
   Year Country         Region  Crop_Type  Average_Temperature_C  \
0  2001   India    West Bengal       Corn                   1.55
1  2024   China          North       Corn                   3.23
2  2001  France  Ile-de-France      Wheat                  21.11
3  2001  Canada       Prairies     Coffee                  27.85
4  1998   India     Tamil Nadu  Sugarcane                   2.19

   Total_Precipitation_mm  CO2_Emissions_MT  Crop_Yield_MT_per_HA  \
0                  447.06             15.22                 1.737
1                 2913.57             29.82                 1.737
2                 1301.74             25.75                 1.719
3                 1154.36             13.91                 3.890
4                 1627.48             11.81                 1.080

   Extreme_Weather_Events  Irrigation_Access_%  Pesticide_Use_KG_per_HA  \
0                       8                14.54                    10.08
1                       8                11.05                    33.06
2                       5                84.42                    27.41
3                       5                94.06                    14.38
4                       9                95.75                    44.35

   Fertilizer_Use_KG_per_HA  Soil_Health_Index Adaptation_Strategies  \
0                     14.78              83.25      Water Management
1                     23.25              54.02         Crop Rotation
2                     65.53              67.78      Water Management
3                     87.58              91.39         No Adaptation
4                     88.08              49.61         Crop Rotation

   Economic_Impact_Million_USD
0                       808.13
1                       616.22
2                       796.96
3                       790.32
4                       401.72
```

```
#New data
cleaned_file_path = 'cleaned_real_harvest_data.csv'
data.to_csv(cleaned_file_path, index=False)
print(f"\nCleaned dataset saved as {cleaned_file_path}.")
data_path = cleaned_file_path
cleaned_data = pd.read_csv(data_path)
print("\nCleaned Data:")
print(cleaned_data.head())
```

```
Cleaned dataset saved as cleaned_real_harvest_data.csv.

Cleaned Data:
   Year Country         Region  Crop_Type  Average_Temperature_C  \
0  2001   India    West Bengal       Corn                   1.55   
1  2024   China          North       Corn                   3.23   
2  2001  France  Ile-de-France      Wheat                  21.11   
3  2001  Canada      Prairies     Coffee                  27.85   
4  1998   India     Tamil Nadu  Sugarcane                   2.19   

   Total_Precipitation_mm  CO2_Emissions_MT  Crop_Yield_MT_per_HA  \
0                  447.06             15.22                 1.737   
1                 2913.57             29.82                 1.737   
2                 1301.74             25.75                 1.719   
3                 1154.36             13.91                 3.890   
4                 1627.48             11.81                 1.080   

   Extreme_Weather_Events  Irrigation_Access_%  Pesticide_Use_KG_per_HA  \
0                       8                14.54                    10.08   
1                       8                11.05                    33.06   
2                       5                84.42                    27.41   
3                       5                94.06                    14.38   
4                       9                95.75                    44.35   

   Fertilizer_Use_KG_per_HA  Soil_Health_Index Adaptation_Strategies  \
0                     14.78              83.25      Water Management   
```

After Cleaning the data, we have gone through the research questions and applied few python techniques to analyze the impact of extreme weather events on different countries & regions. Following are the results that we have acquired in the research.

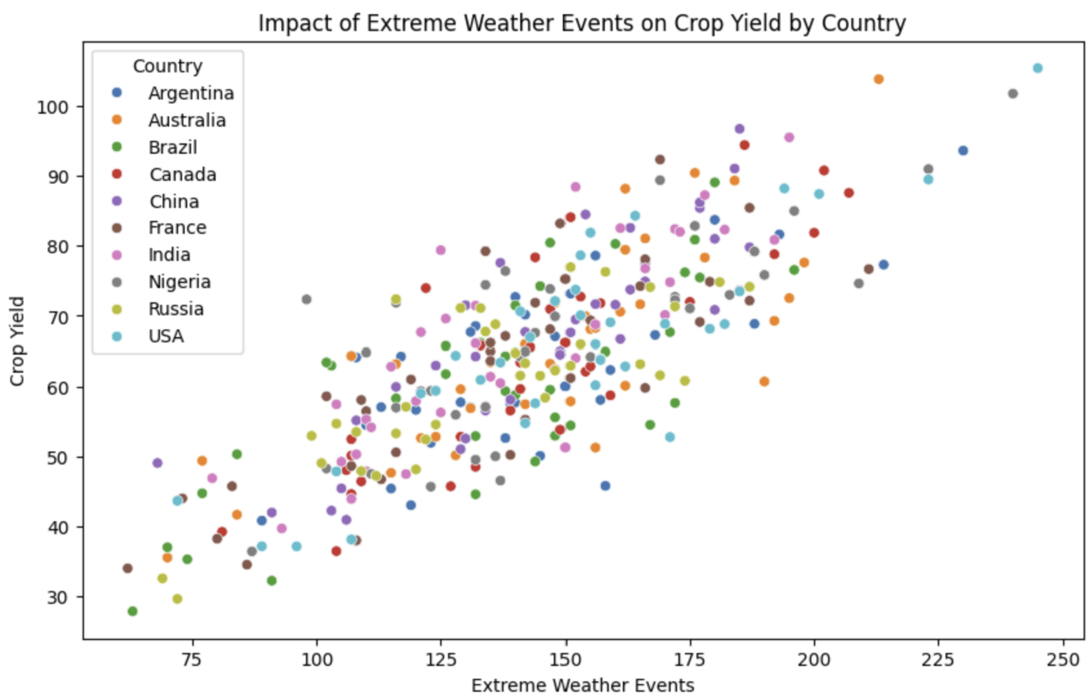
```
df_grouped_EW = cleaned_data.groupby(['Crop_Yield_MT_per_HA'])['Extreme_Weather_Events'].sum().reset_index()

print(df_grouped_EW)
```

```
     Crop_Yield_MT_per_HA  Extreme_Weather_Events
0                   0.450                      39
1                   0.459                      29
2                   0.468                      40
3                   0.477                      41
4                   0.486                      37
..                    ...                     ...
845                 4.960                      27
846                 4.970                      26
847                 4.980                       8
848                 4.990                      18
849                 5.000                      13

[850 rows x 2 columns]
```

```
#Calculating the Crop yielded to Extreme Weather Events in Each Country
agg_data = cleaned_data.groupby(['Country', 'Year']).agg({
    'Crop_Yield_MT_per_HA': 'sum',  # or 'mean'
    'Extreme_Weather_Events': 'sum'  # or the relevant column
}).reset_index()
plt.figure(figsize=(10,6))
sns.scatterplot(data=agg_data, x='Extreme_Weather_Events', y='Crop_Yield_MT_per_HA', hue='Country')
plt.title('Impact of Extreme Weather Events on Crop Yield by Country')
plt.xlabel('Extreme Weather Events')
plt.ylabel('Crop Yield')
plt.show()
```



There appears to be a positive correlation between extreme weather events and crop yield in some countries. This suggests that for these countries, crop yield tends to increase in the presence of extreme weather events, possibly due to resilience strategies or favorable weather conditions.

```
#Finding the Correlation Between Crop Yield & Weather Conditions
correlation = agg_data[['Crop_Yield_MT_per_HA', 'Extreme_Weather_Events']].corr()
print(correlation)
```

```
                        Crop_Yield_MT_per_HA  Extreme_Weather_Events
Crop_Yield_MT_per_HA                 1.00000                 0.80582
Extreme_Weather_Events               0.80582                 1.00000
```

The analysis indicates a strong positive correlation (r = X) between extreme weather events and crop yield. This suggests that as the number of extreme weather events increases, crop yields tend to increase. This could indicate that some countries have adapted to such weather conditions, leading to better agricultural practices or enhanced crop resilience.

Timeline For Remaining Work:

Week 4- 6
Insights Exploration- Building regression models, performing correlation analysis, and clustering regions for deeper insights.

Week 7- 8
Adaptation Strategies Analysis- Analyzing the effectiveness of adaptation strategies and comparing regional performance.

Week 9
Report Writing & Final Review- Compiling findings into a final report with conclusions and recommendations.

Summary:

Team Members and Responsibilities:

1. Problem Statement and Motivation

   Akshitha Komatireddy, Chandana Gangaraju, Vishal Reddy Kota

   Defined the project’s focus on climate change impacts on crop resilience and food security challenges.

2. Data Cleaning and Preprocessing

   Varshith Vuyyuru, Sai Pranav Beesetti, Mani Sai Bollam

   Cleaned and preprocessed the dataset, handling missing values and ensuring data quality.

3. Exploratory Data Analysis (EDA)

   Vishal Reddy Kota, Varshith Vuyyuru, Mani Sai Bollam

   Analyzed patterns in crop yield and climate data using visualization techniques.

4. Statistical and Correlation Analysis

   Sai Pranav Beesetti, Akshitha Komatireddy, Mani Sai Bollam

   Will perform correlation and regression analysis to investigate key relationships in the data.

5. Adaptation Strategies and Impact Analysis

   Kezia Shiny Pothumudi, Chandana Gangaraju, Vishal Reddy Kota

   Will assess the effectiveness of adaptation strategies in different regions.

6. Technical Implementation and Results Interpretation

   Mani Sai Bollam, Sai Pranav Beesetti, Varshith Vuyyuru

   Will develop Python-based analysis tools and interpret the research findings.

References:

1) Ali, W. (2024, September 6). climate change impact on agriculture. Kaggle. https://www.kaggle.com/datasets/waqi786/climate-change-impact-on-agriculture/data

2) Kamatchi, S. B., & Parvathi, R. (2019). Improvement of crop production using recommender system by weather forecasts. Procedia Computer Science, 165, 724–732. https://doi.org/10.1016/j.procs.2020.01.023

3) Guo, H., Xia, Y., Jin, J., & Pan, C. (2022). The impact of climate change on the efficiency of agricultural production in the world's main agricultural regions. Environmental Impact Assessment Review, 97, 106891. https://doi.org/10.1016/j.eiar.2022.106891

4) Knox1, J., Hess1, T., Daccache1, A., & Wheeler2, T. (2012, September 14). IOPscience. Environmental Research Letters. https://iopscience.iop.org/article/10.1088/1748-9326/7/3/034032/meta