**Proposed solution:**

As per our analysis and findings from EDA this is the solution we are proposing.

**Objective**

* The **main objective** of this project is to predict the **impact of climate change on global crop yields** using machine learning models.
* By analyzing historical climate data and crop yields, we aim to develop a model that forecasts future crop production under different climate scenarios.

**Key Steps**

1. **Data Collection**
   * We gather climate data such as **temperature, precipitation, CO₂ levels**, and crop yield information.
   * These datasets come from publicly available sources, ensuring reliability and real-world applicability.
2. **Exploratory Data Analysis (EDA)**
   * The goal of EDA is to understand relationships between climate variables and crop yields.
   * For example, we analyze **how an increase in temperature affects wheat production** or how **variability in rainfall influences maize yields**.
3. **Model Development**
   * We use different machine learning models, including:
     + **Random Forest** (a tree-based ensemble learning method)
     + **Gradient Boosting** (a more refined boosting technique)
     + **Neural Networks** (deep learning models for complex pattern recognition)
   * These models will help **predict yield variations** under different climate conditions.
4. **Adaptation Strategies**
   * Based on model predictions, farmers and policymakers can **adjust farming practices**, such as:
     + Changing **irrigation** methods to cope with droughts.
     + **Altering planting times** to optimize yield based on projected temperature shifts.
     + Promoting **crop diversification** to ensure food security in extreme weather conditions.
5. **Reporting & Visualization**
   * We plan to develop **interactive dashboards and reports** that allow users to explore climate and crop yield relationships.
   * Stakeholders, such as government agencies, can use this data for better decision-making.

**Outcome**

* This project will provide **real-time insights** into how climate change affects agriculture.
* The insights can guide **policy changes, farming adjustments, and future food security planning**.

**Model & Algorithm:**

**Data Pipeline**

The data pipeline ensures smooth processing from input to prediction. It includes:

1. **Input Data**
   * Climate parameters (temperature, precipitation, CO₂)
   * Crop yield data (historical production trends)
   * Agricultural practices (farming techniques, soil conditions)
2. **Preprocessing**
   * **Cleaning the data** (removing missing values, handling outliers)
   * **Normalizing values** to standardize numerical features
   * **Splitting the data** into training and testing sets
3. **Feature Engineering**
   * Creating new variables such as **growing degree days**, which measure the accumulation of heat required for crop growth.

**Model Training**

The following **machine learning models** are used to predict crop yields:

1. **Random Forest**
   * A tree-based algorithm that combines multiple decision trees.
   * It is useful because it handles **non-linear relationships** between climate and crop yields well.
2. **Gradient Boosting**
   * An advanced boosting technique that improves prediction accuracy by reducing bias and variance.
3. **Neural Networks**
   * A deep learning model that captures **complex dependencies** between climate factors and agricultural outputs.

**Evaluation Metrics**

To measure model performance, we use:

1. **RMSE (Root Mean Squared Error)**: Measures the average error in predictions. Lower values indicate better accuracy.
2. **MAE (Mean Absolute Error)**: Represents the mean difference between actual and predicted values.
3. **R² Score (Coefficient of Determination)**: Tells how well our model explains the variation in crop yield data.

**Interpretability with SHAP (SHapley Additive Explanations)**

* Machine learning models, especially deep learning, are often seen as **black boxes** because they lack interpretability.
* We use **SHAP values** to explain how much each feature (like temperature or CO₂ levels) contributes to the final prediction.
* This makes our model more **transparent and reliable** for stakeholders.

**Model Output**

1. **Predictions of Crop Yields** under different climate scenarios, such as:
   * **RCP 4.5 (Moderate Climate Change)**
   * **RCP 8.5 (Severe Climate Change)**
2. **Identification of High-Risk Regions and Crops**:
   * Helps prioritize areas where urgent **adaptation strategies** are needed.
   * Supports farmers in making **informed planting decisions**.

**Conclusion:**

* Our project **bridges the gap** between climate science and agricultural decision-making using machine learning.
* By predicting how **temperature, precipitation, and CO₂ levels** affect crops, we can provide **real-time solutions** for mitigating food shortages.
* The **combination of machine learning models, SHAP analysis, and adaptation strategies** ensures that our approach is both **accurate and actionable**.
* The results of this study can assist in **policy formulation, sustainable agriculture, and long-term food security planning**.