# ML Project Proposal:

**Crop Yield Analysis and Prediction**

## Submitted By:

## Thota Rahul [AM.EN.U4CSE22257] Kowshik Padala [AM.EN.U4CSE22245] Teja Sai Sathwik [AM.EN.U4CSE22271]

**Chethan Kalyan C [AM.EN.U4CSE22216]**

## 2. Abstract

## The program focuses on using machine learning to improve agriculture by developing models that predict crop yields, recommend the most suitable crops based on soil and weather conditions, recommend suitable fertilizers, and predict sales prices. The goal is to create intelligent models that learn from historical agricultural and weather data to make decisions. By analyzing soil patterns, weather, crop data, and economic models,  these models will provide insights that will help farmers improve crop selection, increase production, and increase business decisions. The success of these models will depend on their ability to accurately predict outcomes such as crop yields and price changes, and provide recommendations for fertilizer and crop selection. This approach aims to provide technological solutions to the agricultural community that improve decision-making, ultimately leading to permaculture and better resource management.

## 3. Introduction

**Motivation:**Crop forecasting and optimization are essential to ensure agricultural food security and efficient use of resources. Forecast models can help farmers make decisions on crop selection, fertilizer use, and market pricing.

**Benefits of Solution:**The solution will facilitate better agricultural planning, reduce costs, and increase crop production. Farmers and stakeholders can use this information to maximize profits and make more informed business decisions.  
 **Solution Use:**The models developed in the project will be used to provide instant advice and forecasts to farmers and agricultural organizations. This will enable strategic planning and resource allocation in agriculture.

## 4. ML Algorithms

**We will be using the following machine-learning models for each sub-project:   
1. Crop Recommendation:**

* Logistic Regression
* Random Forest
* Support Vector Machine (SVM)
* K-Nearest Neighbors (KNN)

**2. Crop Yield Prediction:**

* Linear Regression
* Decision Tree
* Random Forest
* Gradient Boosting

**3. Fertilizer Recommendation:**

* Naive Bayes
* Random Forest
* Decision Tree
* Support Vector Machine (SVM)

**4. Price of Agricultural Commodities:**

* Linear Regression
* Ridge Regression
* Lasso Regression
* XGBoost

**Evaluation Metrics:**

**For classification models:**

* Accuracy
* Precision
* Recall
* F1-score

**For regression models:**

* Mean squared error (MSE)
* Mean absolute error (MAE)
* R-squared

## 5. Dataset Finalization

**1. Crop Recommendation Dataset:  
Source:** [Click Here](https://data.mendeley.com/datasets/8v757rr4st/1)  
Features include soil properties and climatic conditions to recommend suitable crops.  
  
**2. Crop Yield Prediction Dataset:**  
Source: [Click Here](https://www.kaggle.com/datasets/bassammakorvi/crop-production-based-on-different-states-of-india)  
Contains historical crop yield data based on regions, seasons, and other factors.  
  
**3. Fertilizers Recommendation Dataset:**  
Source: [Click Here](https://www.kaggle.com/code/pazindushane/fertilizers-recommendation/input)  
Data on soil and crop types to predict the appropriate fertilizers.  
  
**4. Price of Agricultural Commodities Dataset:**  
Source: [Click here](https://www.kaggle.com/datasets/anshtanwar/current-daily-price-of-various-commodities-india/data)  
Includes daily market prices of various agricultural commodities in India.

## 6. Assumptions

We consider the dataset to be complete, accurate, and representative of current agriculture. Any missing data will be imputed using appropriate data imputation techniques. We also assumed that the climate and soil conditions remained the same throughout the study period.

* Price Of Crop is an assumption.
* The unit in Crop Price Datasets is taken as Quintile.
* The units in Crop Yield Prediction are taken as Acres, Ton.