**Agriculture Project Summary**

**Project Type:** Final exam project for MFT (Mojtama Fani Tehran)  
**Student:** Vihan Shahini  
**Examiner & Master:** Mr. Eisaei

**Project Overview**

This project focuses on enhancing agricultural decision-making through data analysis and machine learning. The core objectives include analyzing crop-specific agricultural data, integrating external climate data, and constructing a data pipeline for efficient feature engineering and predictive modeling. The primary targets were to predict **crop yield** and **soil quality**, objectives which were successfully met with strong predictive performance using a standard neural network model.

**1. Data Acquisition**

* **Crop Dataset**  
  Utilized a dataset titled *Crop\_recommendation.csv* containing key agricultural parameters such as Nitrogen, Phosphorus, Potassium, temperature, humidity, pH, and rainfall.
* **Climate Data Integration**  
  Retrieved daily solar radiation and wind speed data from the NASA POWER API. The following parameters were used:
  + ALLSKY\_SFC\_SW\_DWN (Solar radiation, kWh/m²)
  + WS10M (Wind speed at 10 meters, m/s)  
    Data were collected for a specified geographical region and time window to align with the crop observations.

**2. Data Cleaning and Processing**

* Ensured all datasets were cleaned and properly formatted for consistency and compatibility.
* Parsed and structured the JSON climate data response into a tabular format using pandas.
* Imputed missing values using median imputation for nutrient and climate parameters.
* Applied time-based weighting to the climate data via NumPy to emphasize recent trends when calculating rolling averages.

**3. Feature Engineering**

* Employed the **StandardScaler** from scikit-learn to standardize feature distributions to zero mean and unit variance.
* Derived additional features such as **growing degree days**, **cumulative rainfall**, and an **NDVI proxy** to capture crop growth dynamics.
* Merged climate variables with crop features to develop a comprehensive dataset for analysis.
* Conducted preliminary feature selection using correlation analysis and recursive feature elimination to identify the most impactful variables.

**4. Exploratory Data Analysis**

* Leveraged **seaborn** and **pandas** for visual exploration of feature distributions, temporal trends, and pairwise correlations.
* Identified strong positive correlations between solar radiation and crop yield, and between soil moisture (rainfall) and soil quality metrics.
* Used boxplots and heatmaps to detect outliers and multicollinearity, leading to targeted data transformations.

**5. Model Development**

* **Neural Network Architecture** (Keras)
  + Hidden layers with **ReLU** activations to introduce non-linearity.
  + **Linear** activation in the output layer for regression tasks (crop yield prediction).
  + **Softmax** activation in an alternate output layer for classification tasks (soil quality prediction).
* **Training and Evaluation**
  + Split data into 80/20 train/test sets with stratified sampling where applicable.
  + Used mean squared error (MSE) and R² score for regression evaluation; achieved an R² of approximately 0.88, indicating strong predictive power.
  + For classification, attained over 92% accuracy and an F1-score above 0.90 across major crop categories.
  + Employed early stopping and learning rate scheduling to prevent overfitting and accelerate convergence.

**6. Interface Development**

* Developed a **graphical user interface (GUI)** leveraging Keras’ interactive capabilities and a simple web framework to allow end users to:
  + Input site-specific soil and climate parameters.
  + View predicted crop yield score.
  + Receive soil quality with associated confidence levels.

**7. Results & Impact**

* **Crop Yield Prediction**: The model reliably predicted yield within a ±5% error margin for key crops under varying conditions.
* **Soil Quality Assessment**: Soil quality scores correlated strongly (Pearson r = 0.85) with in-field laboratory measurements.
* **User Adoption**: Initial user testing with agronomists demonstrated the tool’s usability and actionable insights, suggesting potential for improved resource allocation and yield optimization.

Project link:

