STAT451 Project Proposal

[Proposed Title] Crop recommendation system

[Group 22] Karen Nam, Salahuddin Syed, Seonyeong Heo, Tumi Samuel-Ipaye, Yein Choi

Introduction

Precision agriculture helps farmers to make informed decisions regarding crop selection, cultivation techniques, and farm management. This approach leverages data on soil, climate, and environmental conditions to recommend the most suitable crops for a given area, enhancing productivity and sustainability. The crop recommendation dataset in this study aims to support such decisions, focusing on key soil and climate parameters. This dataset, constructed from sources on rainfall, climate, and soil fertility data specific to India, provides a foundation for building predictive models that guide farmers in selecting optimal crops based on local soil and environmental conditions.

Methods

Dataset

- Dependent variable (y): Crop label, 22 classes
    - Type: Categorical (object)
    - Description: Indicates the most suitable crop for the given soil and environmental conditions.
    - Purpose: This variable serves as the target for predictive models, aiming to classify samples into appropriate crop categories.
- Independent variables (X): Environmental and soil conditions
    1. N – Nitrogen content
        - Type: int64
        - Description: Ratio of nitrogen content in the soil, a crucial nutrient influencing plant growth and productivity.
    2. P – Phosphorous content
        - Type: int64
        - Description: Ratio of phosphorous in the soil, supporting root development and flowering.
    3. K – Potassium content
        - Type: int64
        - Description: Ratio of potassium in the soil, aiding in resistance to diseases and improving water efficiency in plants.
    4. Temperature
        - Type: float64
        - Description: Temperature in degrees Celsius, influencing plant growth rates, photosynthesis, and respiration.
    5. Humidity
        - Type: float64
        - Description: Relative humidity percentage, affecting transpiration rates and the overall water needs of plants.
    6. pH
        - Type: float64

        - Description: Soil pH level, indicating acidity or alkalinity, which can influence nutrient availability and microbial activity.
    7. Rainfall
        - Type: float64
        - Description: Average rainfall in millimetres, which helps determine water availability for crop growth and the suitability of specific crops.

Pre-processing

- Standardization
- Data splitting
- K-fold cross validation

Models

- Decision Tree
- Random Forest
- kNN
- Logistic Regression
- SVM
- Ensemble

Evaluation metrics

- Precision
- Recall
- Accuracy

Results

- Comparison among models
    - Evaluate and rank the performance of all models using the evaluation metrics.