Crop Recommendation Model Report

Team: Bitrithm
Members: HDKM Surawera, PIPN Pemaratna
Task No: 01

Data Exploration and Preprocessing

We first use pandas and ydata_profile libraries to get a brief idea about the dataset. The best thing is that it didn't have any NULL values, so it made the data preprocessing part a lot easier.

Then we look for redundant columns, because we have to reduce the number of features to obtain a higher precision model. So we use pandas to remove Total_Nutrients, Temperature_Humidity, and Log_Rainfall.

Then we saw the Label and Label_Encoded column, so here if we take the Label_Encoded column as the target feature, we don't have to do anything in the categorical variable encoding part. So we also remove the Label column.

We then tried to scale numerical features using StandardScaler, but when testing with new data we had a problem that was not predicted correctly. Then we removed the scaler part.

I put all those preprocessing steps inside a function so we can use it when predicting new data.

Model Training

I divided the data set into training (80%) and test (20%) sets.
When we train models we use some models like Support Vector Machine, Decision Tree Classifier, and Random Forest Classifier. Of those models, the best model was the Random Forest Classifier (with 100 estimators), so we chose it for the task.

Model Evaluation

For model evaluation, we used Accuracy and decided which is the best model with Accuracy and got an accuracy of around 0.99 for the test data set with Random Forest Classifier.

Prediction of New Data

I included new sample data in the Python code so you can test the model with that.

Instructions for Running the Code

- Ensure that Python and necessary libraries (e.g., pandas, scikit-learn) are installed.
- In the Python code it includes functions for train_model, preprocess_data, predict_new_data.
- In the main function, you can comment out the train_model function and uncomment the predict_new_data function, then it will predict new data.