Task: 01

Team Name: KernelNewbies

Approach Taken

(1) First, we identified the input features to be N, P, K, temperature, humidity, pH, and rainfall. So, we omitted the inherited columns such as Total_Nutrients, Temperature_Humidity, Log_Rainfall, and Label_Encoded. Numerical features are standardized using *scalar.fit_transform* Categorical data were already encoded in the given data set. *print(data.isnull().sum())* showed that there were no missing values.

(2) The ML model was built using *RandomForestClassifier*. This doesn't require extensive pre-processing and has the ability to create a robust and accurate model. The data set was split into testing-training sets with a ratio of 1:4. This was done using *train_test_split.*

(3) Accuracy of the model was tested using *accuracy_score.* Which returned 0.9977272727272727. Which is a good score. The model can suggest the appropriate corps based on the given environmental conditions with an accuracy of 99%.

(4) User enters inputs about a new environment and the joblib model is used to make predictions based on the user input. The top 3 predictions are displayed along with the accuracy.

Challenges Faced

- Whether to use another leaning algorithm other than *RandomForestClassifier.* Finally, we decided to use Random Forest because of its accuracy and ease of use (little pre-processing).

Insights Gained from Model Evaluation

- The model has an accuracy of 0.9977272727272727.
- The model predicted the labels of the given dataset with a probability of approximately 1.
- When tested for new environments the model kept predicting the correct label with a considerably good probability.

How to Improve Model's Performance

- Try to modify the parameters of *RandomForestClassifier* to achieve a better accuracy.
- We have already performed feature selection; feature engineering can combine the selected features to better suite the relationships between features.

How to Run the Code

(1) Use *python Intellihack_KernelNewbies_01_code.py* to run the code. (assuming the data set is in the file *Crop_Dataset - Crop_Dataset.csv*)
(2) It will display the model accuracy and ask for the features of the new environment.
(3) After entering the features of the new environment, it will display the predicted top 3 labels.

Sample Test Cases

Case 01 – Made up Environment

python Intellihack_KernelNewbies_01_code.py
Accuracy of the Model: 0.9977272727272727
N: 90
P: 60
K: 17
temperature: 23
humidity: 62
ph: 6.1
rainfall: 134.6
Top Three Crops that Matches Your Input:
barley: 0.59
tomatoes: 0.29
cauliflower: 0.08

Case 02 – Data Related to Strawberries

python Intellihack_KernelNewbies_01_code.py
Accuracy of the Model: 0.9977272727272727
N: 63
P: 67
K: 50
temperature: 43
humidity: 91.1
ph: 6.62
rainfall: 91.12
Top Three Crops that Matches Your Input:
strawberries: 0.97
lettuce: 0.01
potatoes: 0.01