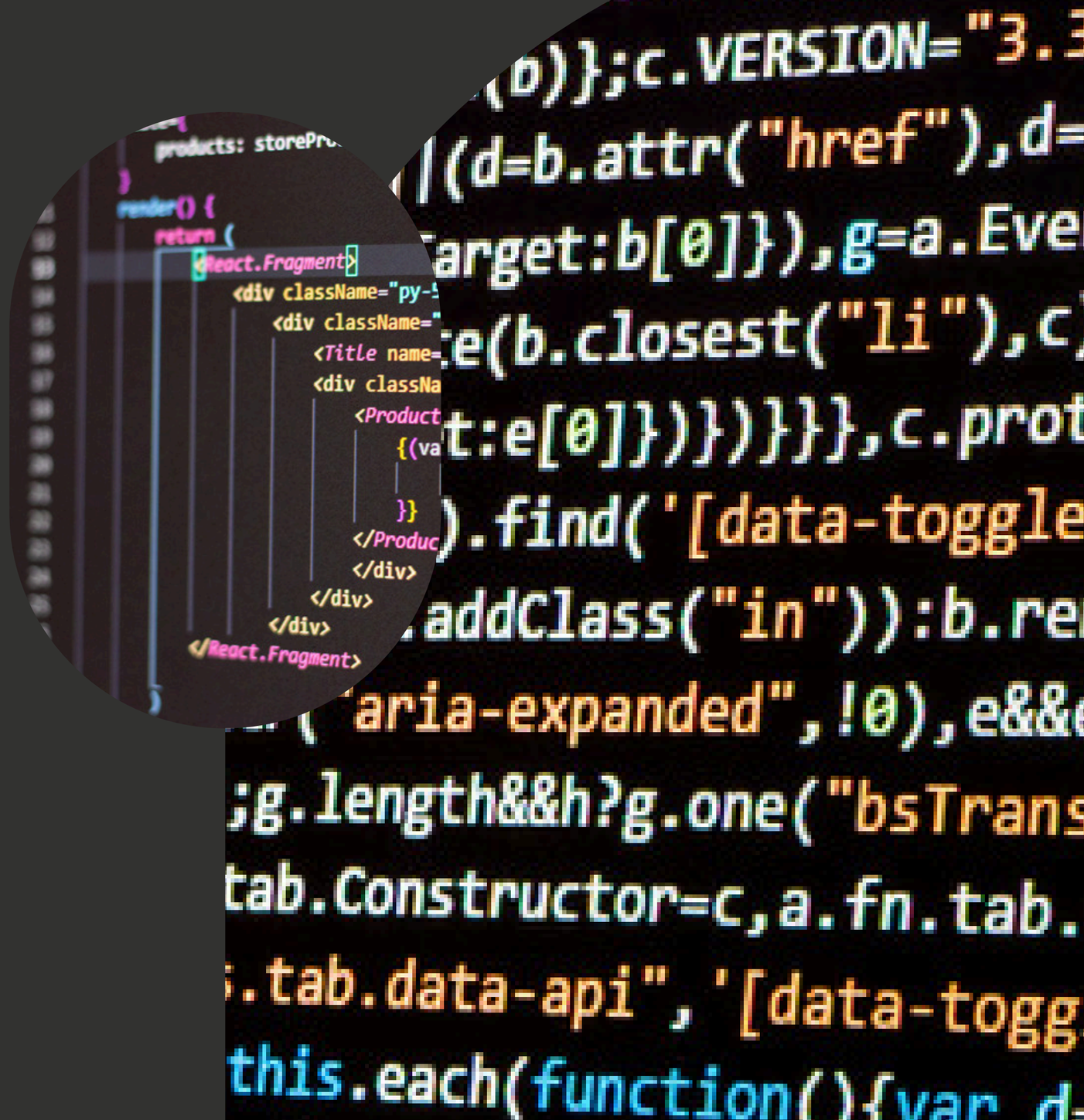
Presentation Project UE3.1

Group 6:

1. Sofiane BESSILA
2. Grégoire THOMAS
3. Sopharith NOU
4. Panha CHHAY

Date: 11 April 2024

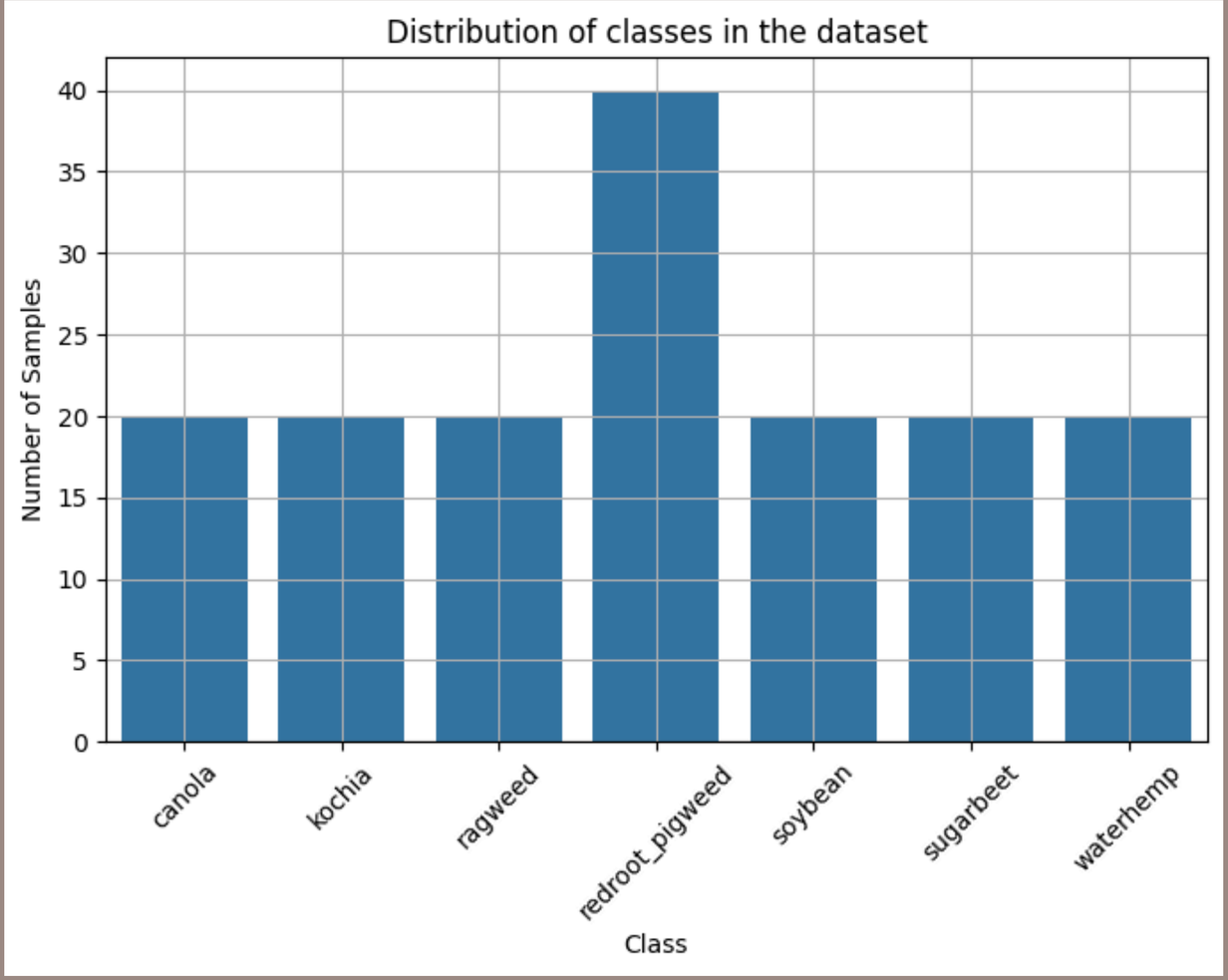


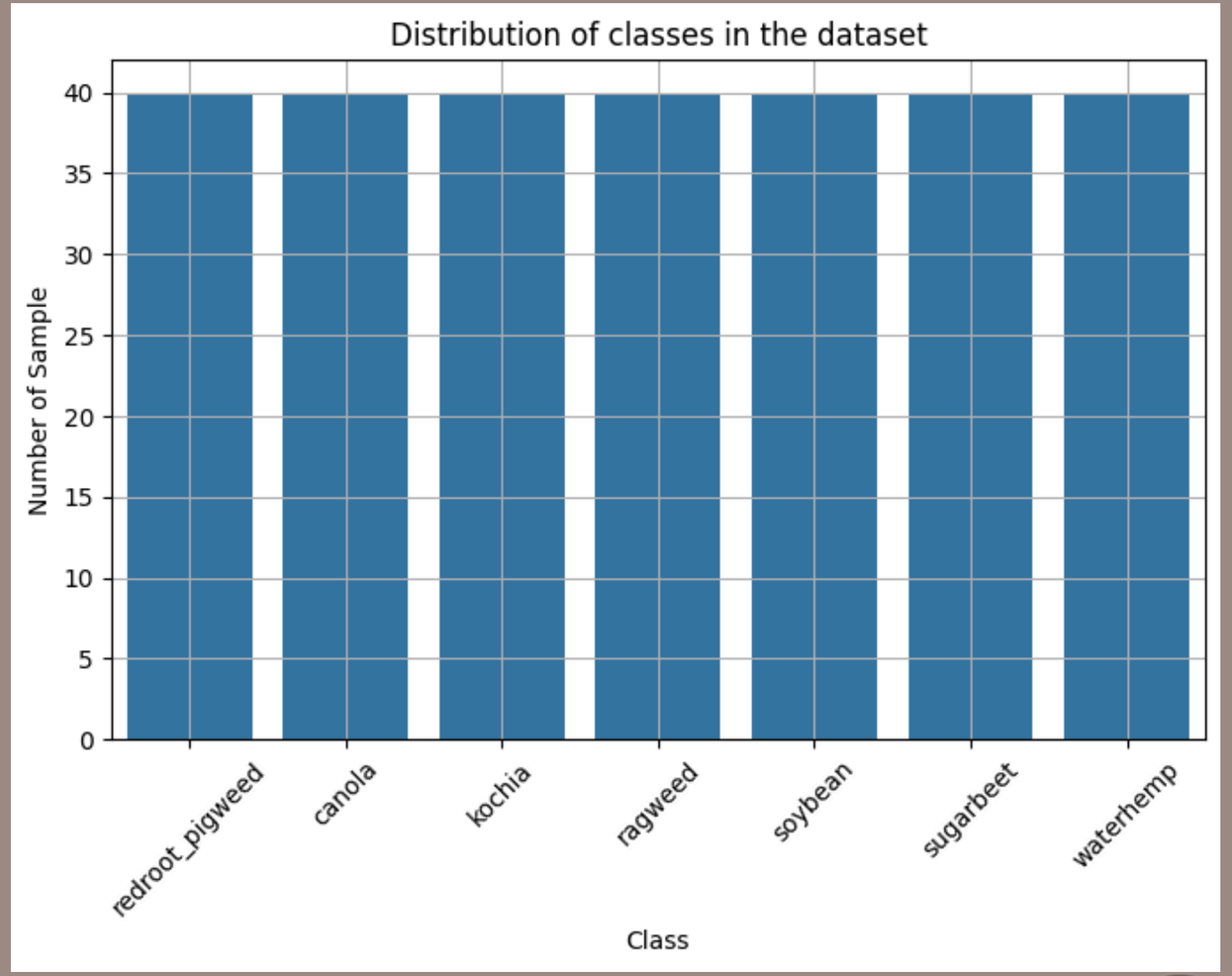
Human Agency

Seamless integration with existing tools

I. Random Forest Approach

Number of the data balance

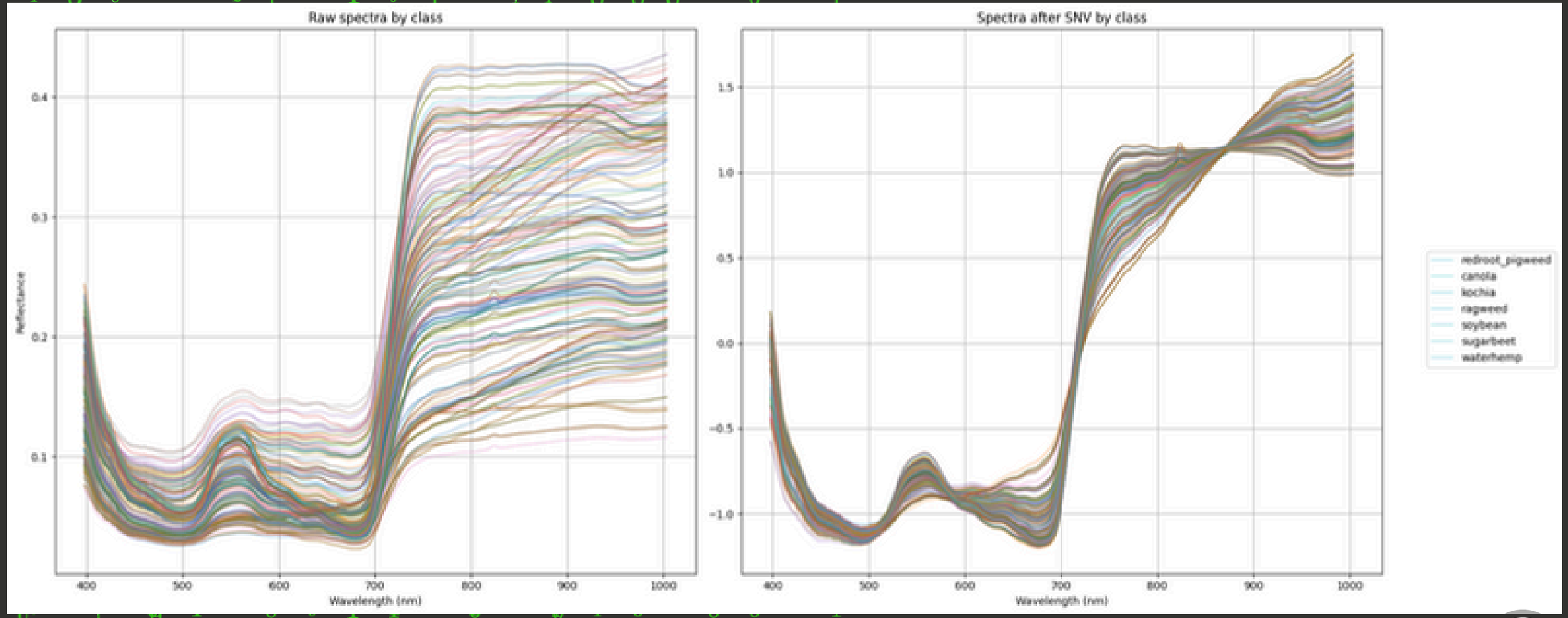




Visualization the raw data

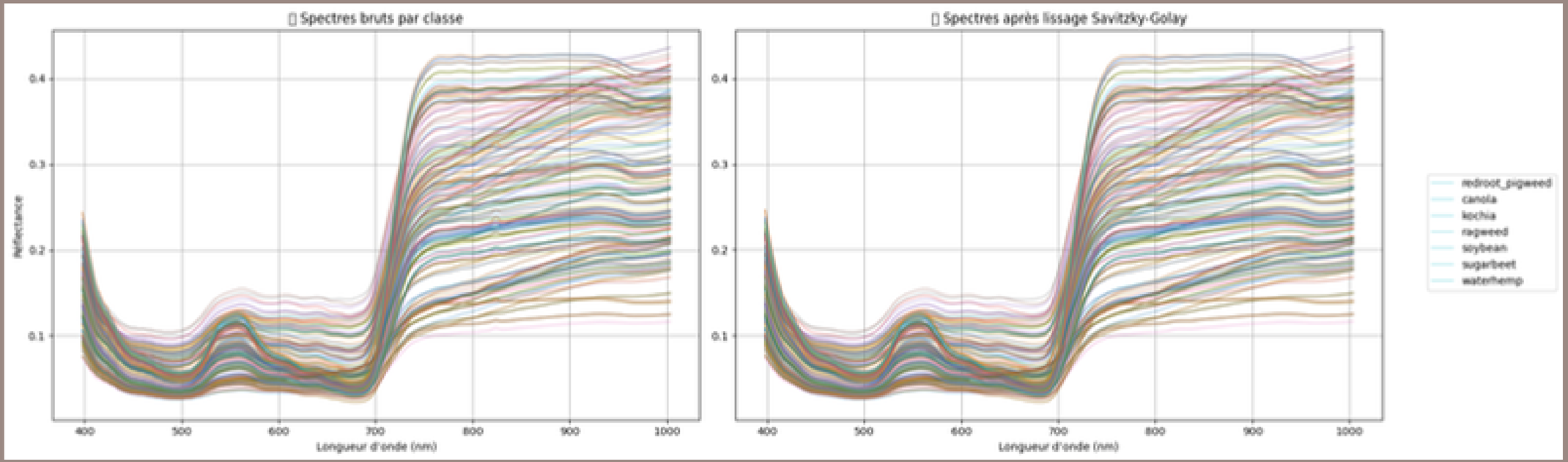
Visualization the data after run the balance

The pretreatment between Raw data VS SNV of the Spectral.

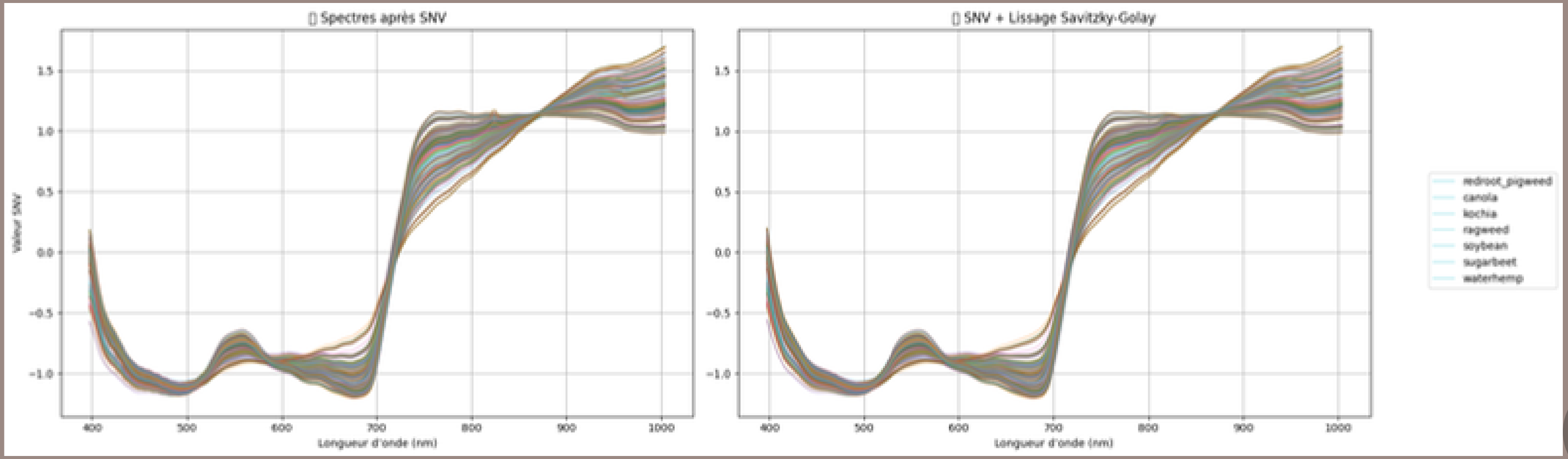


✅ Accuracy : 0.98

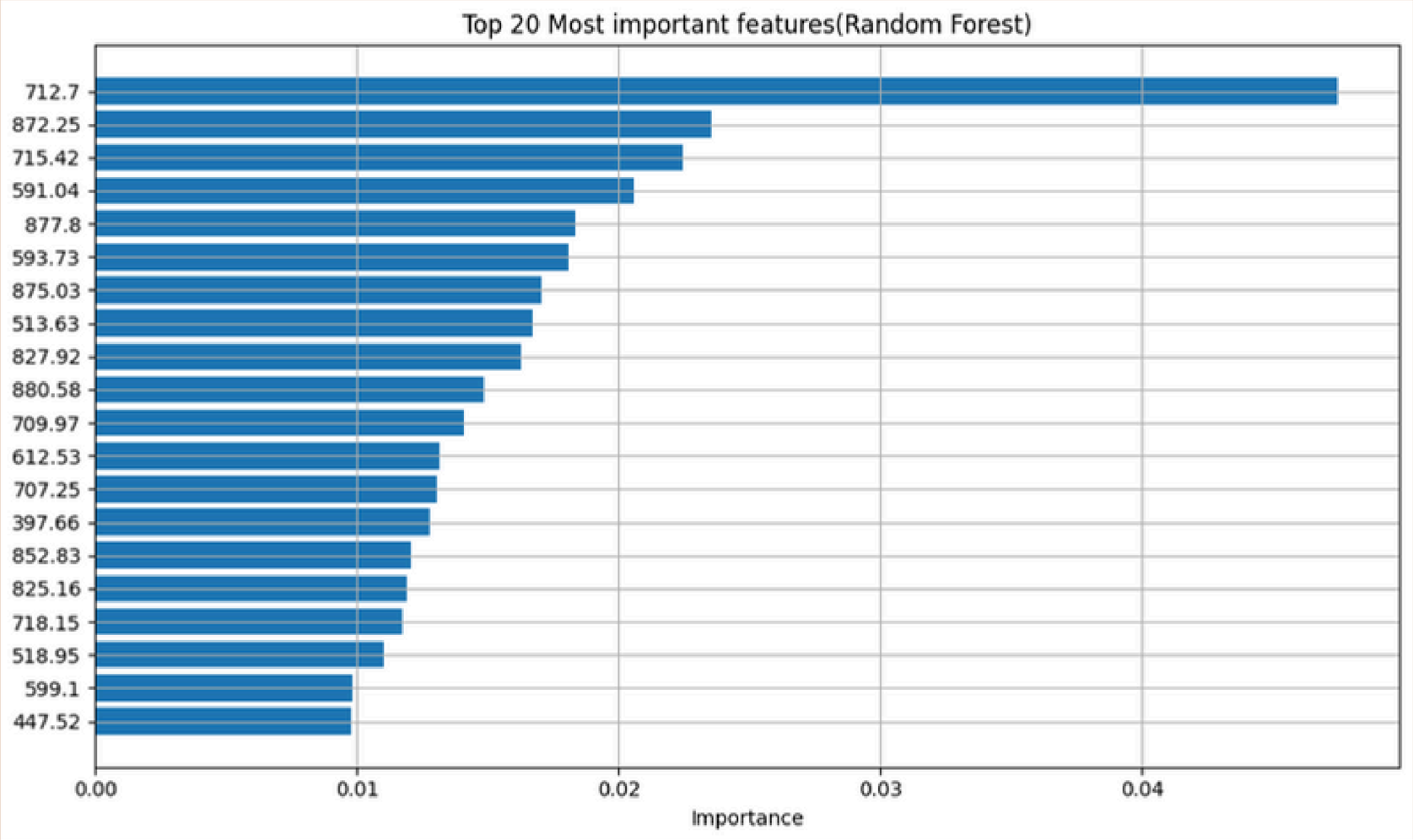
✅ Accuracy : 0.98



Method Savitzky-Golay and SNV+ Savitzky-Golay have the same **“Accuracy : 0.98”**



Feature adjustment and dimension reduction for extracting the importance data



📊 Initial dimension: 224

🔻 Dimension after reduction: 18

📈 31.50%

Noted: Important Feature has to classify the fitting data feature.

New Model with Random Forest of the Important Feature

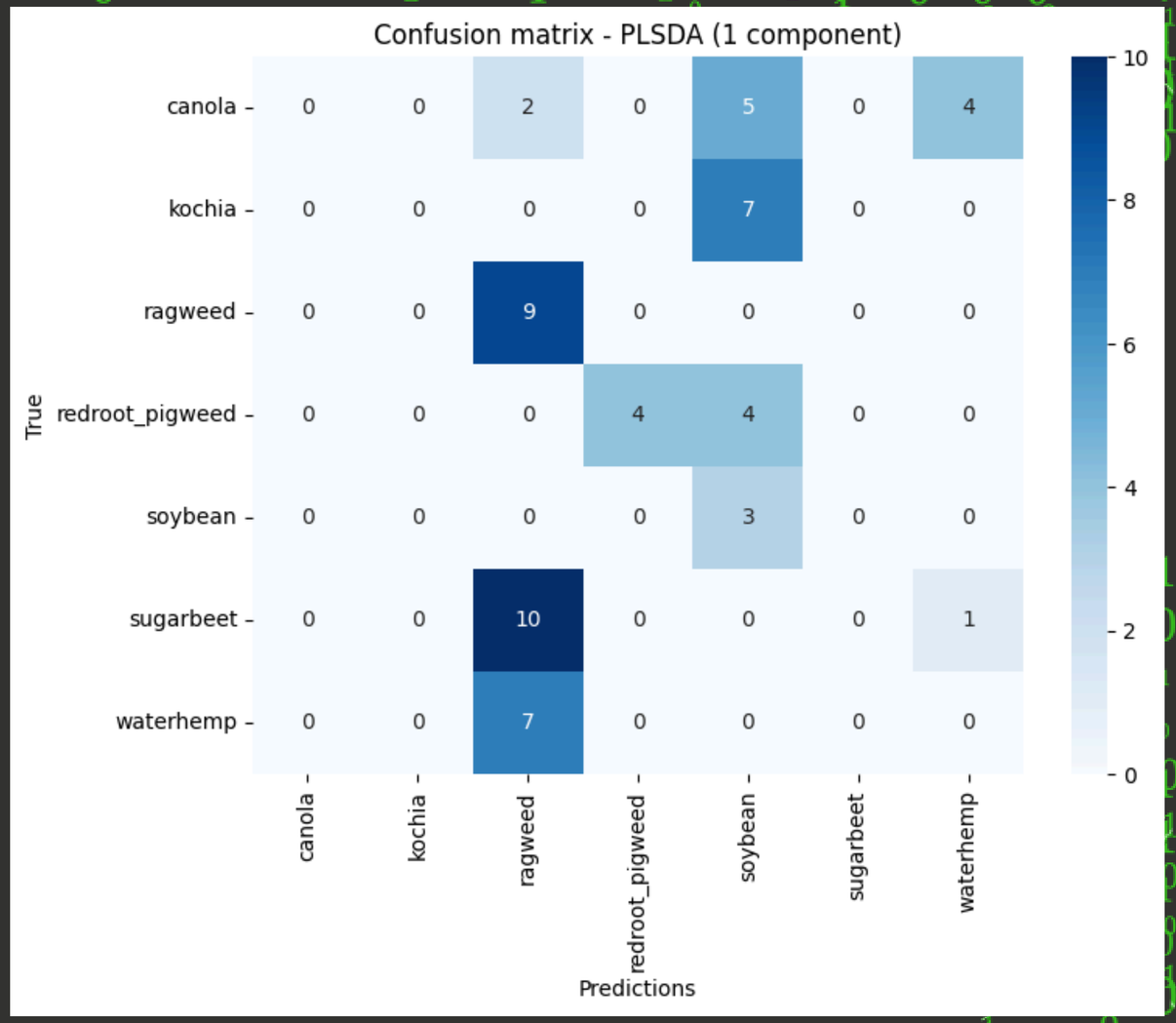
```
✅ Accuracy (modèle réduit) : 0.9821428571428571

📄 Rapport de classification (modèle réduit) :
                precision    recall  f1-score   support

        canola       1.00      1.00      1.00         8
        kochia       0.89      1.00      0.94         8
       ragweed       1.00      0.88      0.93         8
redroot_pigweed      1.00      1.00      1.00         8
       soybean       1.00      1.00      1.00         8
     sugarbeet       1.00      1.00      1.00         8
     waterhemp       1.00      1.00      1.00         8

      accuracy                           0.98        56
     macro avg       0.98      0.98      0.98        56
  weighted avg       0.98      0.98      0.98        56

🧮 Matrice de confusion :
[[8 0 0 0 0 0 0]
 [0 8 0 0 0 0 0]
 [0 1 7 0 0 0 0]
 [0 0 0 8 0 0 0]
 [0 0 0 0 8 0 0]
 [0 0 0 0 0 8 0]
 [0 0 0 0 0 0 8]]
```

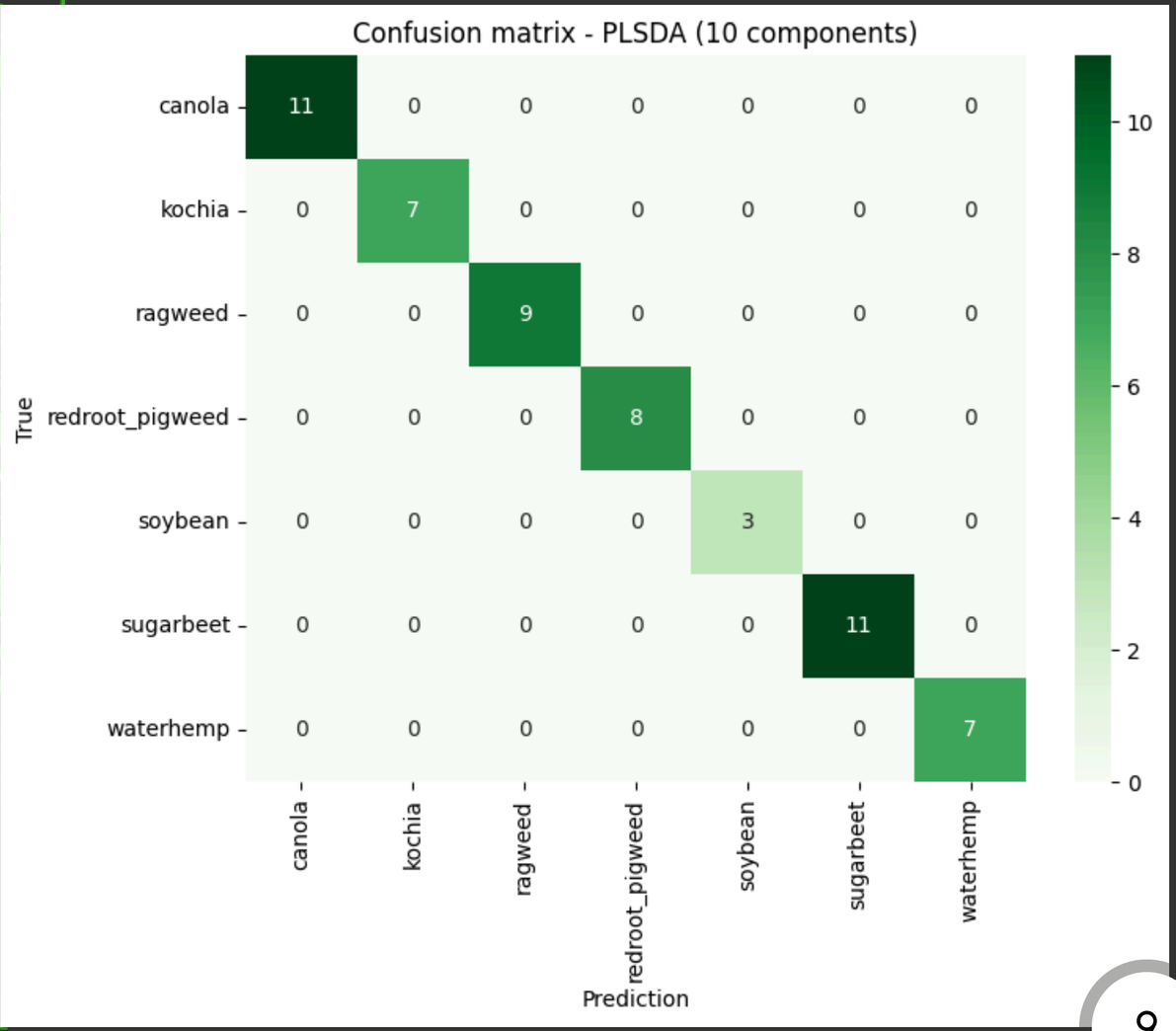
Result of Hyperparameter optimization

```
Fitting 5 folds for each of 108 candidates, totalling 540 fits
🌟 Meilleurs hyperparamètres : {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}
🎯 Meilleure accuracy (cross-validation) : 0.9911111111111112

📈 Accuracy sur test set : 0.9821428571428571

📄 Rapport de classification :
              precision    recall  f1-score   support

      canola       1.00      1.00      1.00         8
      kochia       0.89      1.00      0.94         8
     ragweed       1.00      0.88      0.93         8
redroot_pigweed       1.00      1.00      1.00         8
     soybean       1.00      1.00      1.00         8
   sugarbeet       1.00      1.00      1.00         8
   waterhemp       1.00      1.00      1.00         8

    accuracy                           0.98        56
   macro avg       0.98      0.98      0.98        56
weighted avg       0.98      0.98      0.98        56
```

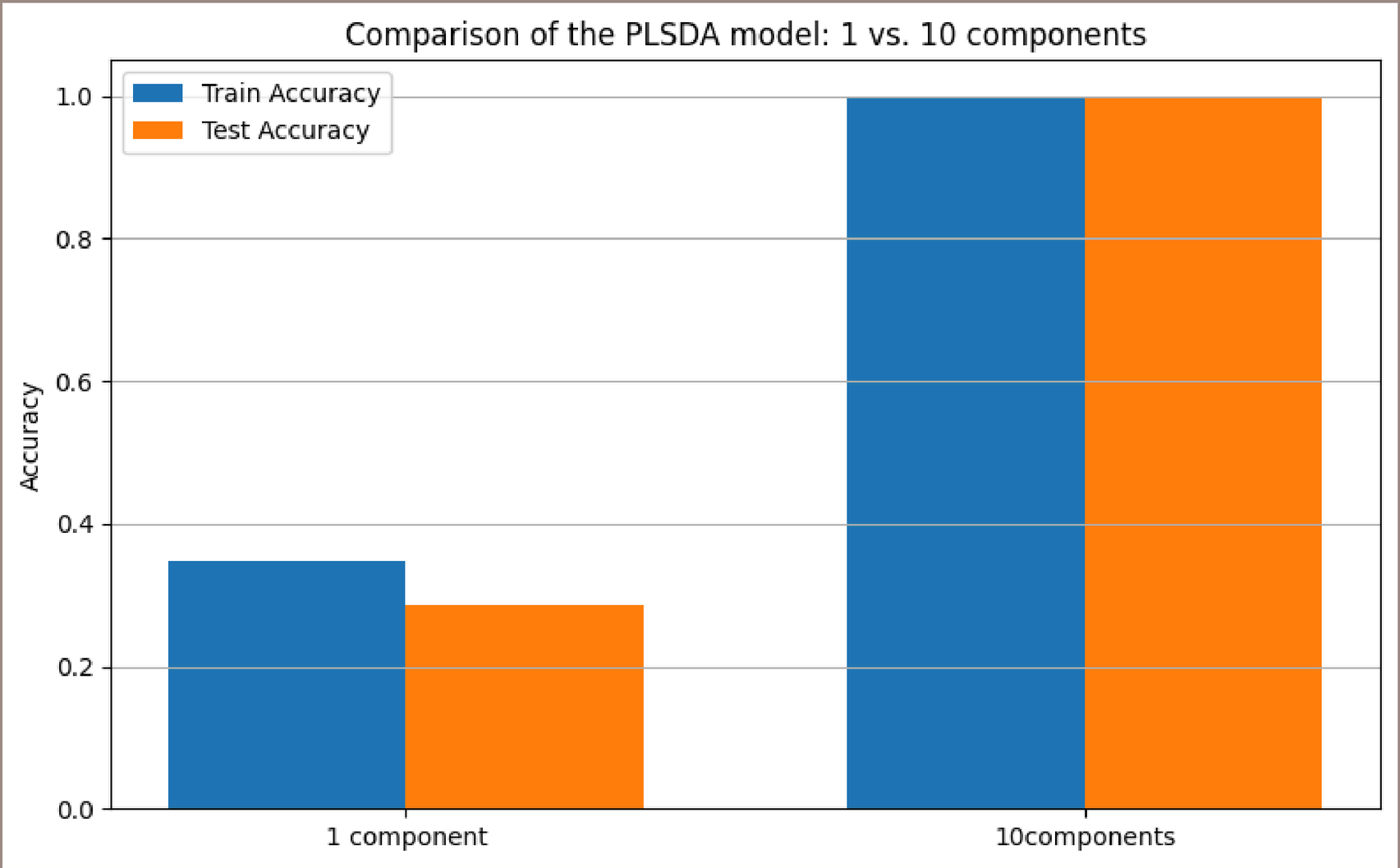
```
📊 Scores de validation croisée (5-fold) : [1.         1.         0.98214286 1.         1.        ]
✅ Moyenne : 0.9964
📉 Écart-type : 0.0071
```

II. PLS Approach





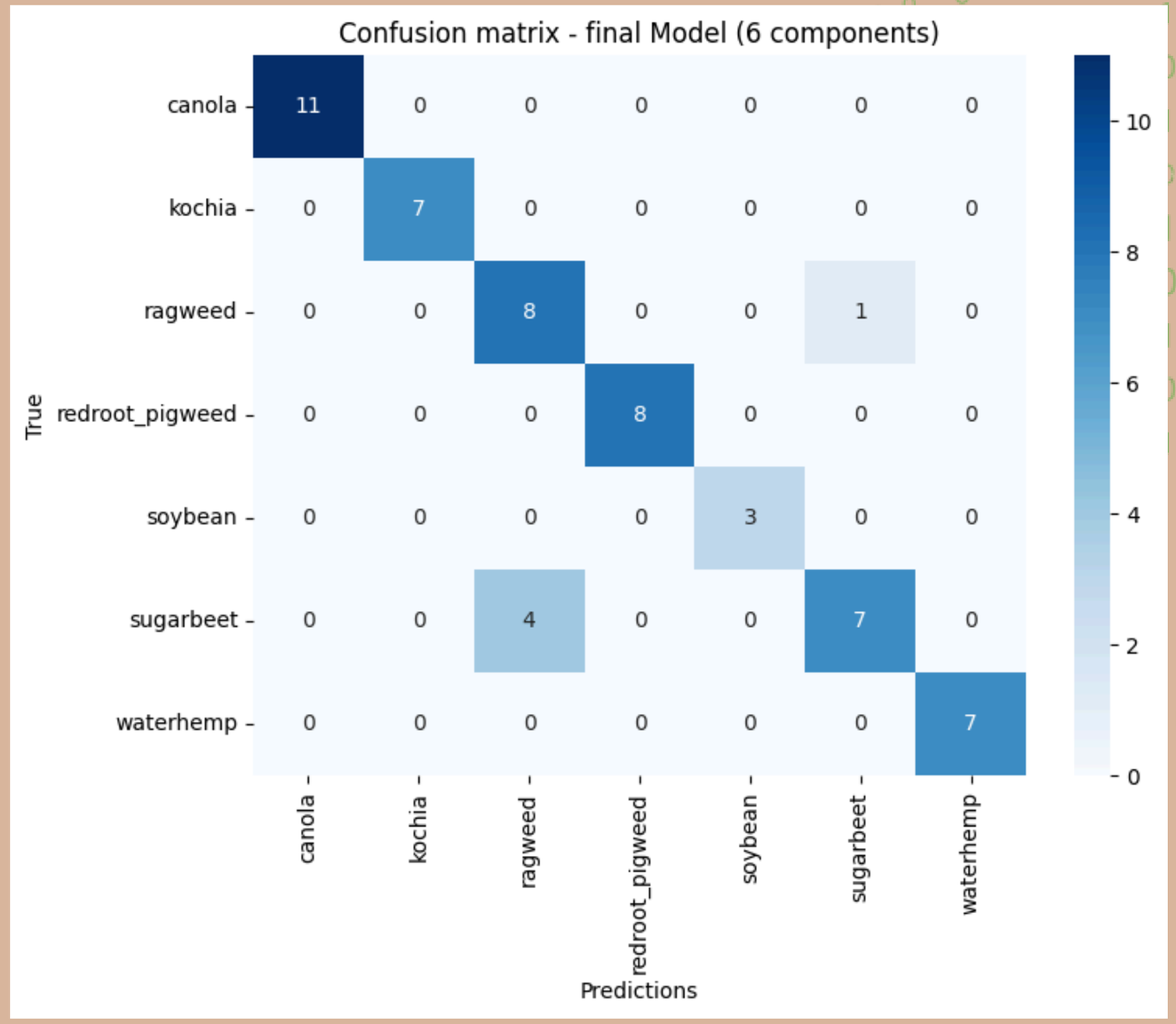
Compare Accuracy and SD



Extremely significant Difference

- Average accuracy (1 component) : 0.29 ± 0.03
- Average accuracy (10 components) : 1.00 ± 0.01

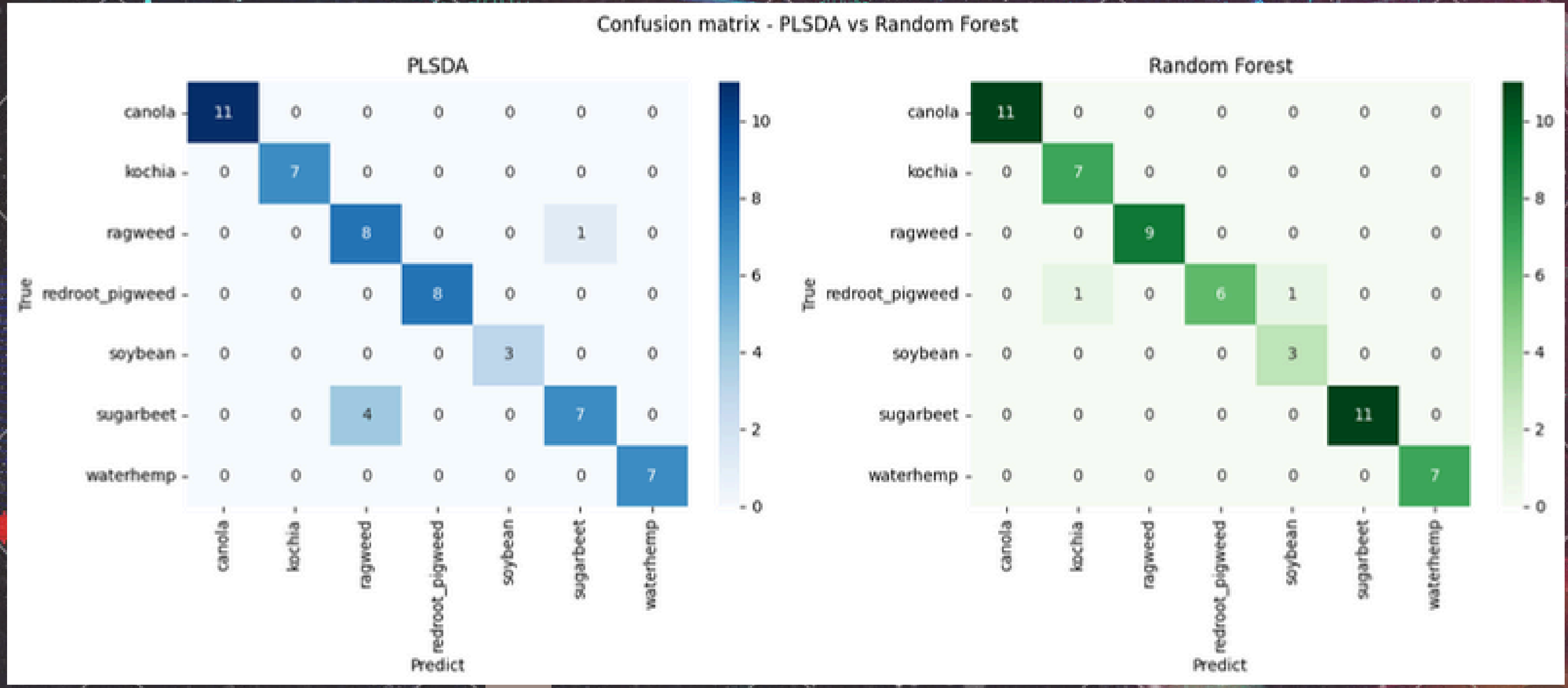
The Final Model PLS




```
📊 Cross-validation results :

🔸 1 components : Average accuracy = 0.29 | Standard deviation = 0.03
🔸 2 components : Average accuracy = 0.48 | Standard deviation = 0.06
🔸 3 components : Average accuracy = 0.67 | Standard deviation = 0.08
🔸 4 components : Average accuracy = 0.71 | Standard deviation = 0.16
🔸 5 components : Average accuracy = 0.80 | Standard deviation = 0.12
🔸 6 components : Average accuracy = 0.91 | Standard deviation = 0.05
🔸 7 components : Average accuracy = 0.95 | Standard deviation = 0.03
🔸 8 components : Average accuracy = 0.95 | Standard deviation = 0.03
🔸 9 components : Average accuracy = 1.00 | Standard deviation = 0.01
🔸 10 components : Average accuracy = 1.00 | Standard deviation = 0.01
```

The accuracy more than 90% with the low SD from 6 components

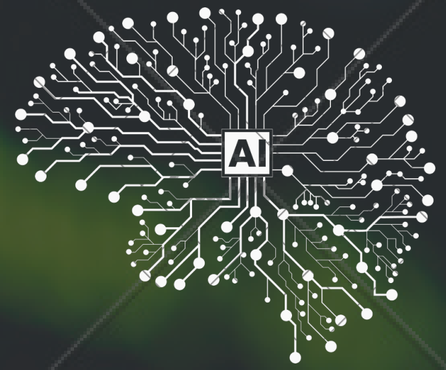
Random Forest Vs PLS



Accuracy : 0.91

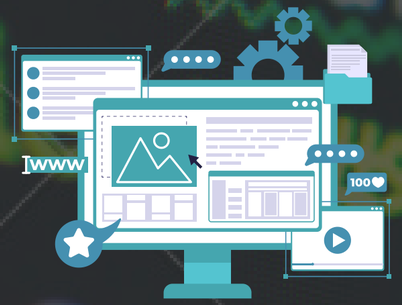
Accuracy : 0.96

III. Conclusion



- The Random Forest and PLS have good results of accuracy, but Random Forest greater than PLS.

- We should test with another dataset to evaluate these approaches.



THANK YOU