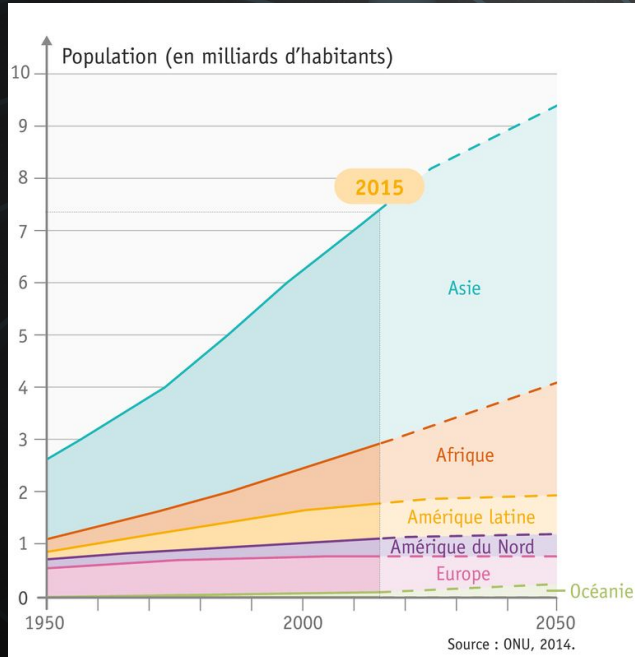


LE MACHINE LEARNING AU SERVICE DE L'AGRICULTURE DE PRÉCISION

Présenté par Elise Lonchamp, Emma Da Costa Silva, Maud Lesage

I. Introduction

A. Agriculture et défis mondiaux



Population mondiale : 9,1 milliards en 2050

Défi : nourrir toute la population

I. Introduction

B. Qu'est ce que l'agriculture de précision ?



**Gestion des parcelles
agricoles en s'appuyant sur
des données**

**Prendre des décisions
optimisées pour chaque
parcelle**

I. Introduction

C. Le rôle du Machine Learning dans l'agriculture de précision ?



Analyser les données des capteurs

Aide pour les prises de décisions

I. Introduction

D. Présentation du jeu de données



Inde : 70% de la population pratique l'agriculture

Objectif : prédire les cultures optimales en fonction des caractéristiques physico-chimiques des sols

Description : 2200 individus et 8 variables

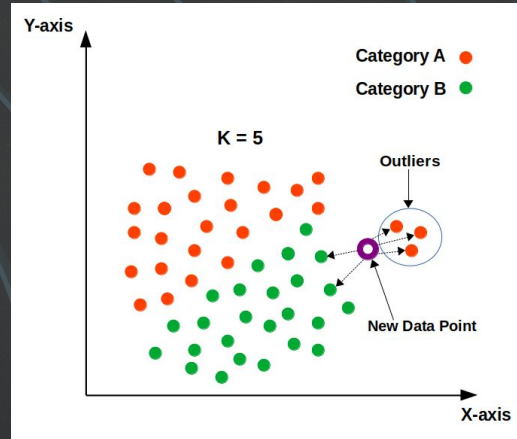
1 variable des noms des cultures (22 modalités)

7 variables de description des sols :
Azote - Phosphore - Potassium - Température -
Humidité - pH - Pluie

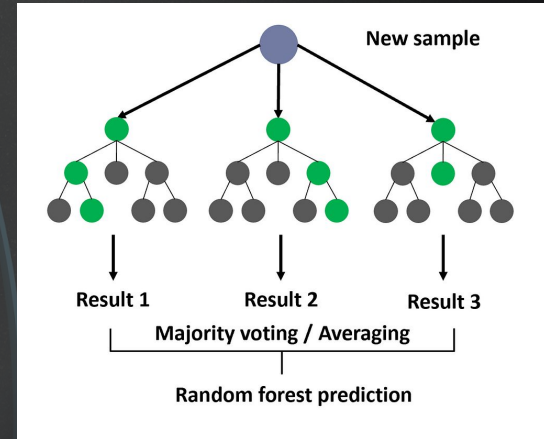
Source : Katarya et al., 2020

II - Prédiction de la meilleure culture pour un profil de sol donné

A. Présentation des méthodes



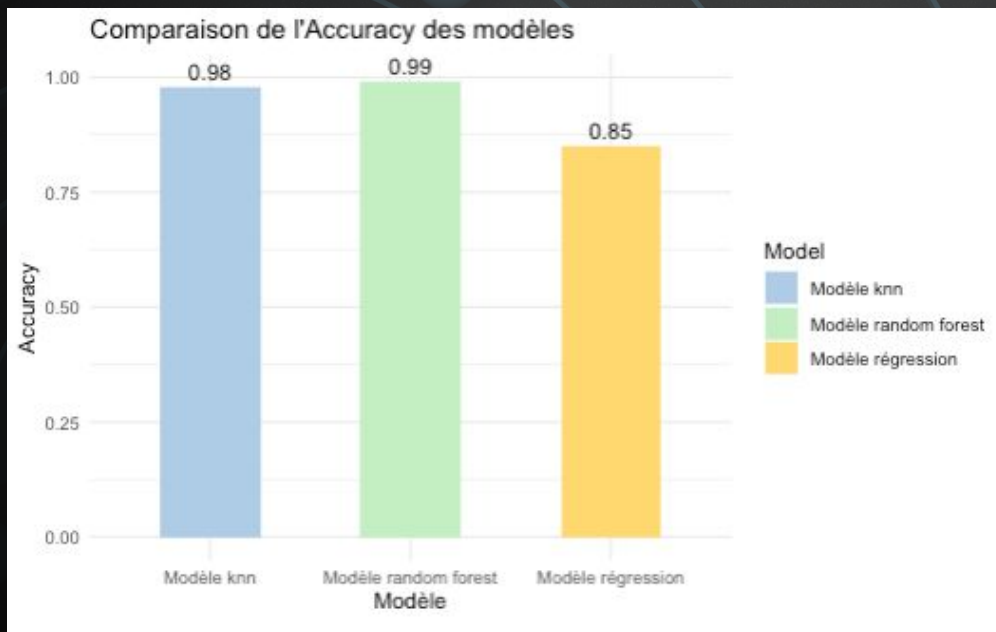
Méthode KNN (plus proches voisins)



Méthode Random Forest

II - Prédiction de la meilleure culture pour un profil de sol donné

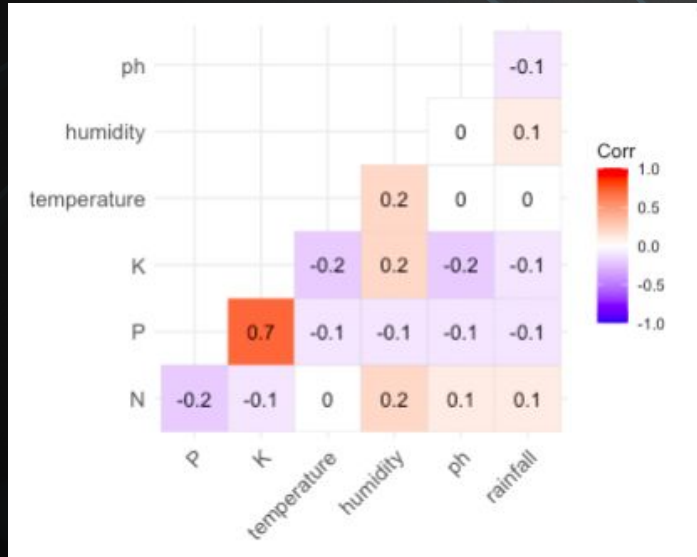
B. Modèles complets



Modèle complet : Culture ~ P + K + N
+ pluie + humidité + température +
pH

II - Prédiction de la meilleure culture pour un profil de sol donné

C. Modèles simplifié : corrélation entre les variables



Corrélation importante entre P et K : 0,7

Colinéarité = information redondante → pouvoir prédictif biaisé (Japa et al., 2019)

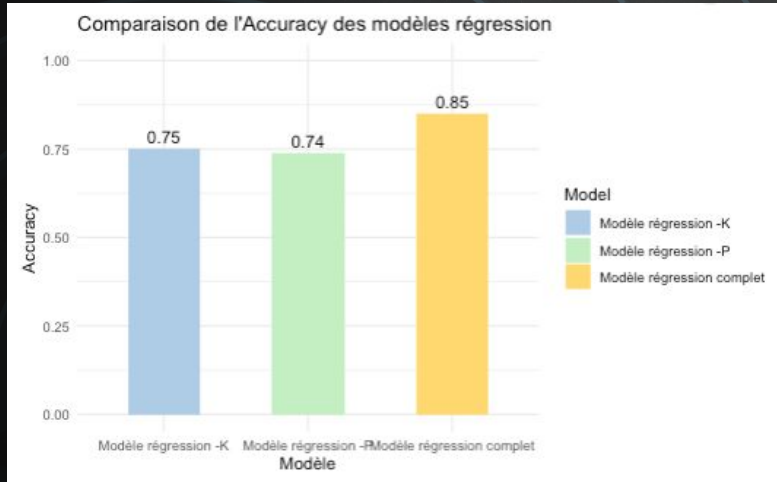
Modèle sensible à la colinéarité : régression

Modèles plus robustes : KNN et Random Forest

II - Prédiction de la meilleure culture pour un profil de sol donné

C. Modèles simplifiés : Régression logistique multinomiale

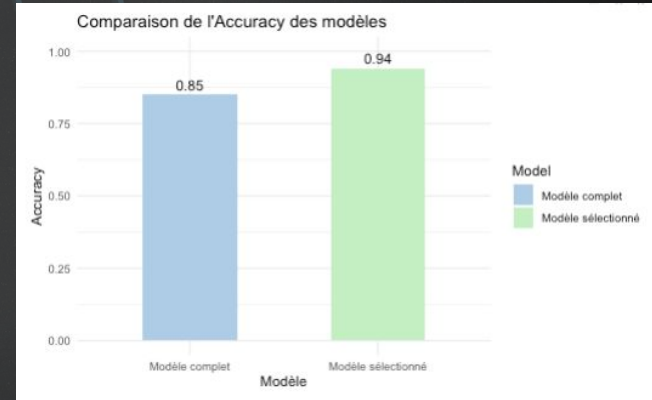
Suppression des variables P et K



→ Les variables P et K ne semblent pas être colinéaires.

Modèle sélectionné

Méthode stepwise : Culture \sim P + K + pluie + humidité

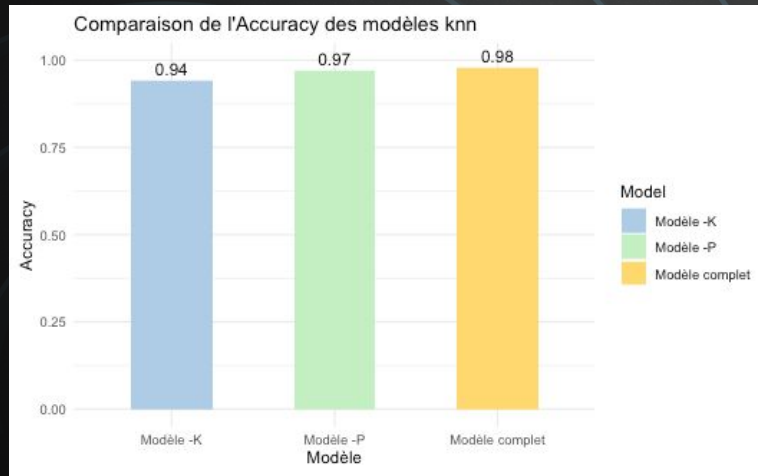


Accuracy = 0.94 > 0.85 → meilleur modèle **8**

II - Prédiction de la meilleure culture pour un profil de sol donné

C. Modèles simplifiés : KNN

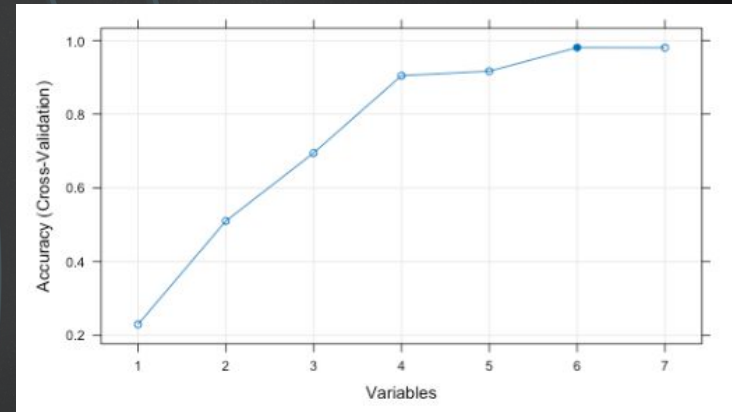
Suppression des variables P et K



→ Moins bonne performance que le modèle complet.

Modèle sélectionné

Méthode RFE (Kuhn, 2019): Culture ~ P + K + N + pluie + humidité + température

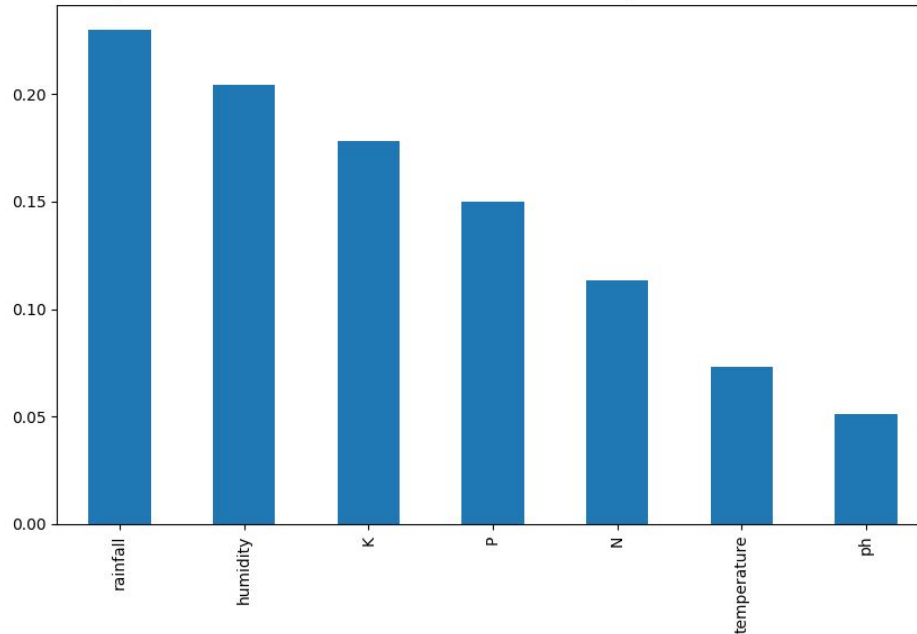


Accuracy = 0.9807 > 0.9801 → très proche **9**

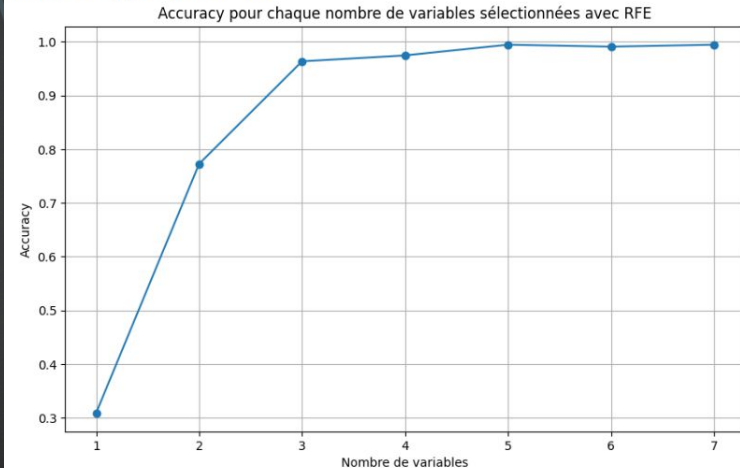
II - Prédiction de la meilleure culture pour un profil de sol donné

C. Modèles simplifiés : Random Forest

Importance des variables dans le modèle Random Forest pour l'ensemble des cultures

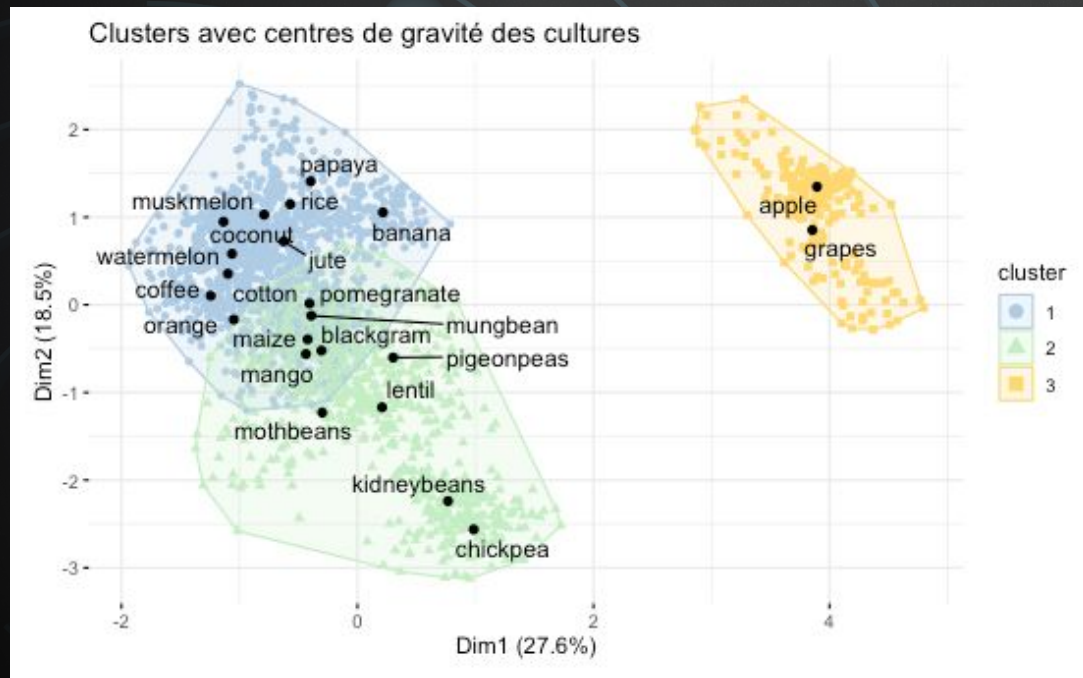


```
Variables sélectionnées avec 1 variables: ['humidity']
Accuracy avec 1 variables sélectionnées: 0.3090909090909091
Variables sélectionnées avec 2 variables: ['humidity', 'rainfall']
Accuracy avec 2 variables sélectionnées: 0.7727272727272727
Variables sélectionnées avec 3 variables: ['K', 'humidity', 'rainfall']
Accuracy avec 3 variables sélectionnées: 0.9636363636363636
Variables sélectionnées avec 4 variables: ['P', 'K', 'humidity', 'rainfall']
Accuracy avec 4 variables sélectionnées: 0.9745454545454545
Variables sélectionnées avec 5 variables: ['N', 'P', 'K', 'humidity', 'rainfall']
Accuracy avec 5 variables sélectionnées: 0.9945454545454545
Variables sélectionnées avec 6 variables: ['N', 'P', 'K', 'temperature', 'humidity', 'rainfall']
Accuracy avec 6 variables sélectionnées: 0.9909090909090909
Variables sélectionnées avec 7 variables: ['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall']
Accuracy avec 7 variables sélectionnées: 0.9945454545454545
Nombre optimal de variables: 5
Meilleure accuracy: 0.9945454545454545
```



III - Test de la capacité du modèle à différencier les cultures similaires

A. Choix de 5 cultures similaires



Choix du sous-ensemble de cultures similaires :

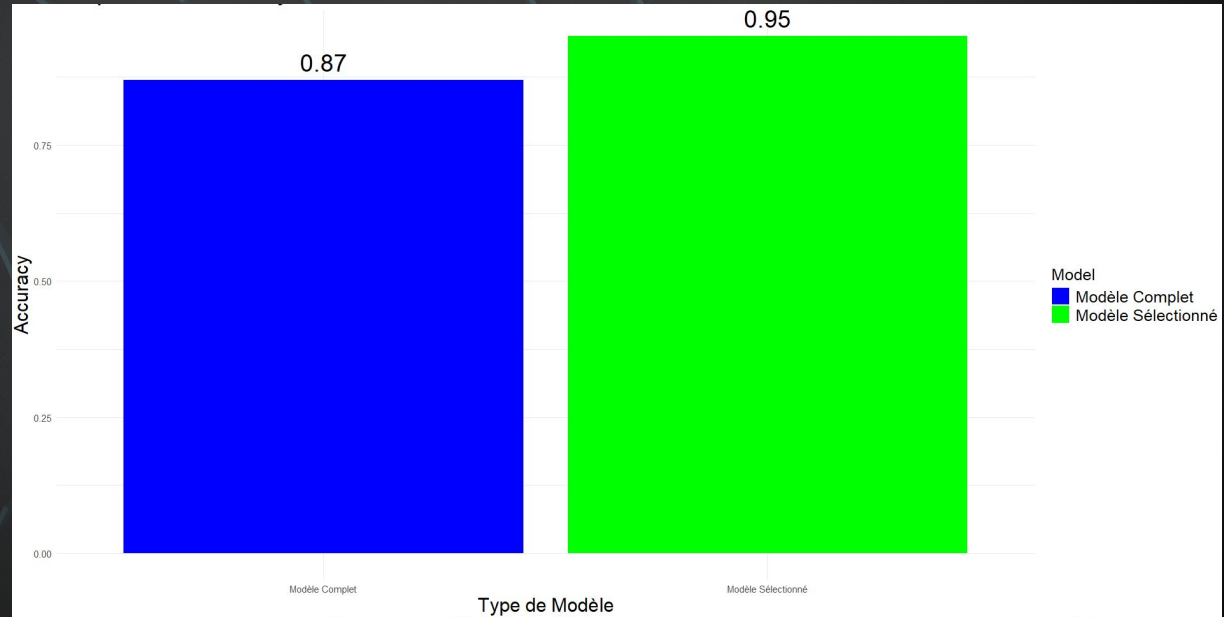
Riz - Maïs - Jute -
Coton - Café

III - Test de la capacité du modèle à différencier les cultures similaires

B. Prédiction avec un sous-ensemble de cultures similaires : Régression logistique multinomiale

**Accuracy =
0.95 > 0.87**

**→ Modèle
sélectionné
meilleur**



III - Test de la capacité du modèle à différencier les cultures similaires

B. Prédiction avec un sous-ensemble de cultures similaires : Régression logistique multinomiale

	Accuracy	Sensibilité	Spécificité
Modèle complet	0,87	Plus faible pour la classe jute : 0,57	Plus faible pour la classe maïs : 0,93
Modèle sélectionné	0,95	Amélioration : plus faible pour la classe jute : 0,90	Amélioration : plus faible pour la classe riz : 0,977

III - Test de la capacité du modèle à différencier les cultures similaires

B. Prédiction avec un sous-ensemble de cultures similaires : KNN

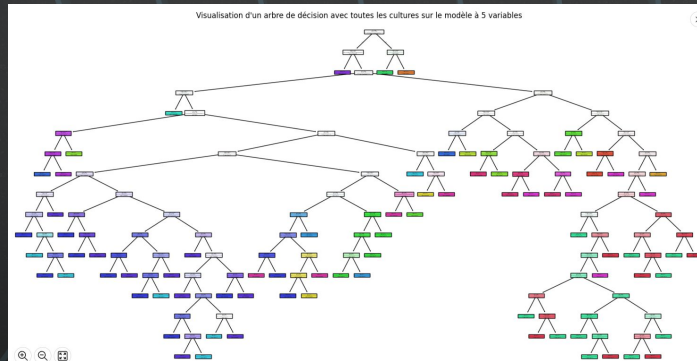
	Accuracy	Sensibilité	Spécificité
Modèle complet	0,93	Plus faible pour la classe jute : 0,79	Plus faible pour la classe riz : 0,94
Modèle sélectionné	0,94	Amélioration pour la classe maïs (1 contre 0,96)	Amélioration pour la classe coton (1 contre 0,96)

- **Confusion entre les classes jute et riz pour les deux modèles**
- **Modèle simplifié permet de prédire sans erreur la classe maïs**

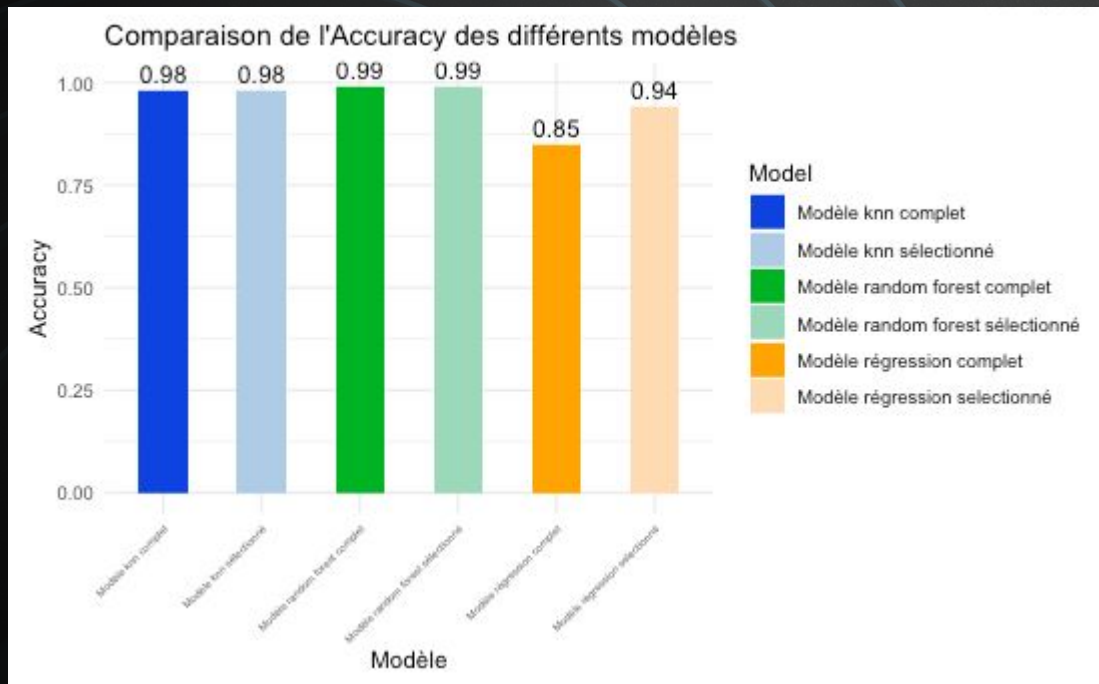
III - Test de la capacité du modèle à différencier les cultures similaires

B. Prédictions avec un sous-ensemble de cultures similaires : Random Forest

	Accuracy	Sensibilité	Spécificité
Modèle complet et Modèle sélectionné	0,99	Plus faible pour la classe riz : 0,96	Plus faible pour la classe jute : 0,99



CONCLUSION



+ **PERSEPECTIVE
S : DEEP
LEARNING ET
RÉSEAUX DE
NEURONES**

BIBLIOGRAPHIE

https://www.cellulestat.cra.wallonie.be/wp-content/uploads/2016/12/Formation_Stats_3_4_GLM_en_Pratique_Part1.pdf