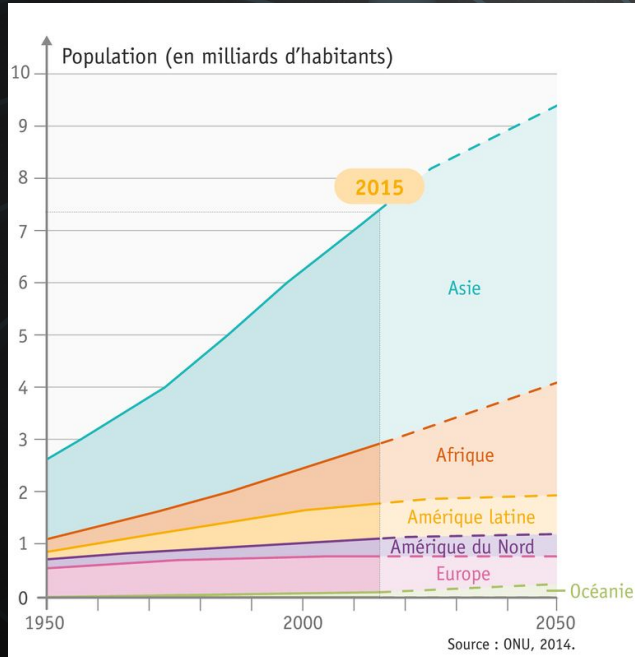


LE MACHINE LEARNING AU SERVICE DE L'AGRICULTURE DE PRÉCISION

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)

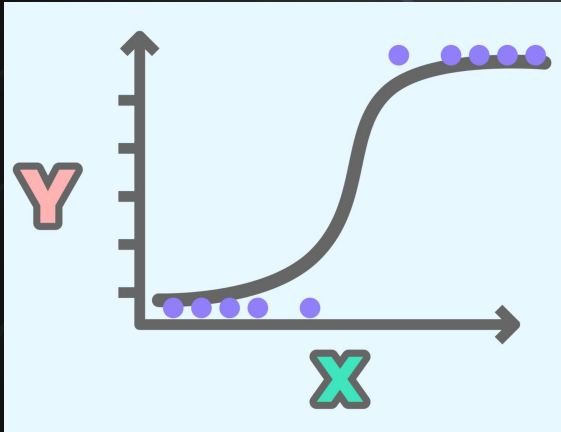
7 variables de description des sols :

Azote - Phosphore - Potassium - Température - Humidité - pH - Pluie

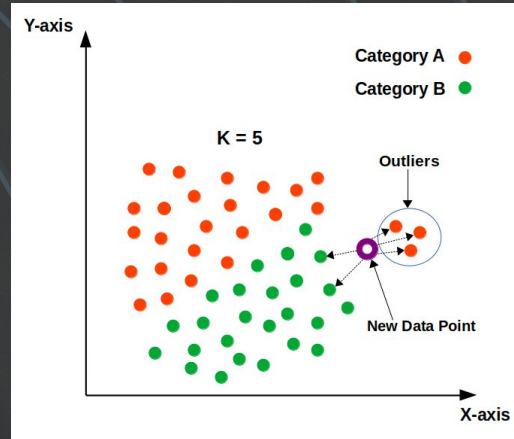
Source du jeu de données : Katarya et al., 2020

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

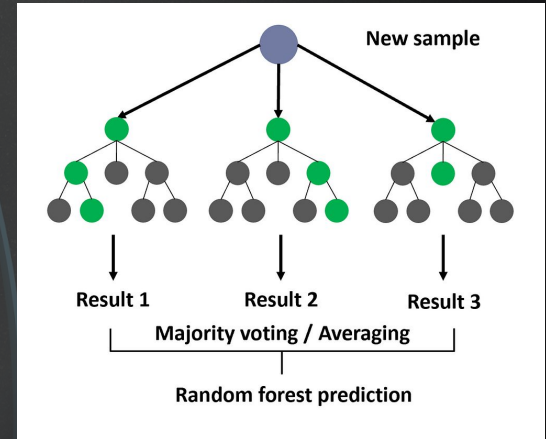
A. Présentation des méthodes



Régression logistique



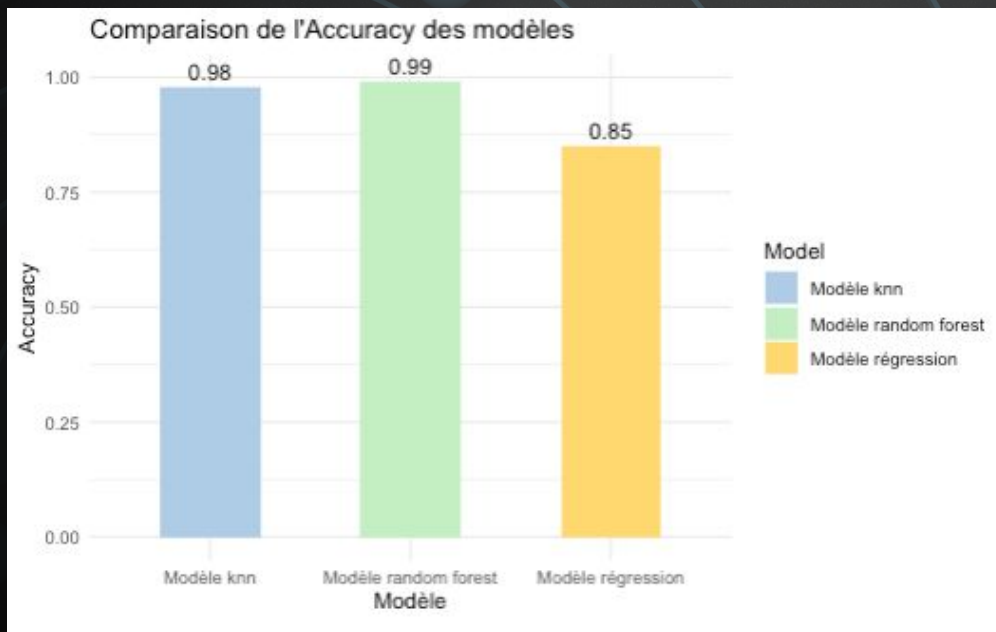
Méthode KNN (plus proches voisins)



Méthode Random Forest

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

B. Modèles complets

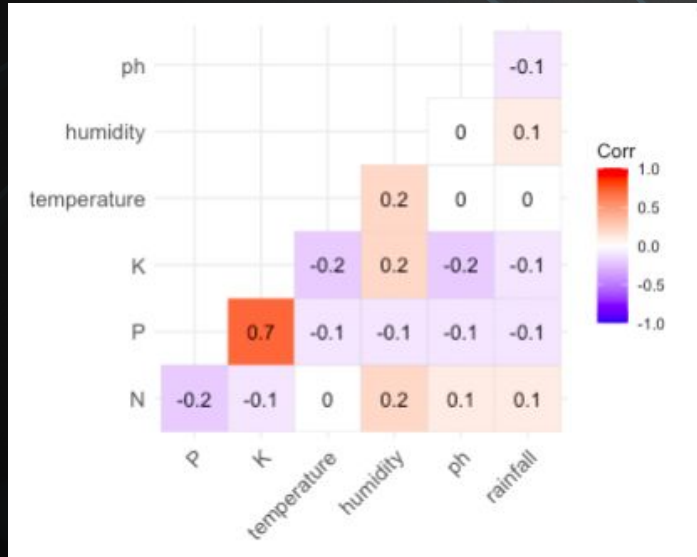


Modèle complet : Culture \sim P + K + N
+ pluie + humidité + température +
pH

Modèle KNN : $k = 1$

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

C. Modèles simplifiés : corrélation entre les variables



Corrélation importante entre P et K : 0,7

Colinéarité = information redondante →
pouvoir prédictif biaisé

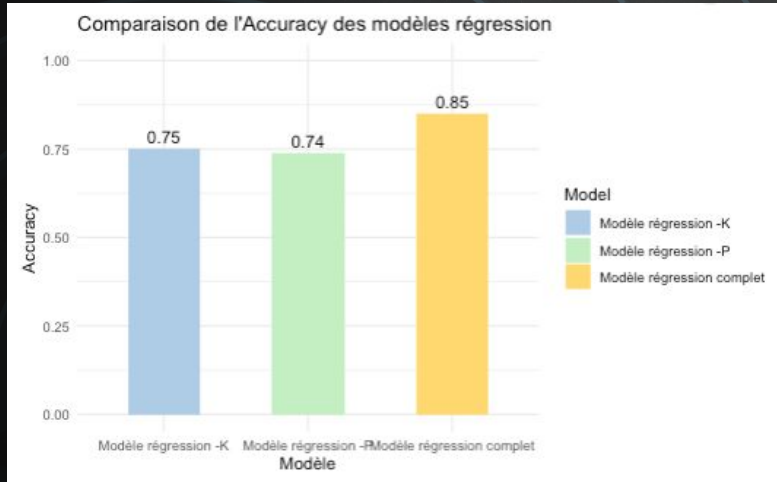
Modèle sensible à la colinéarité : régression

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

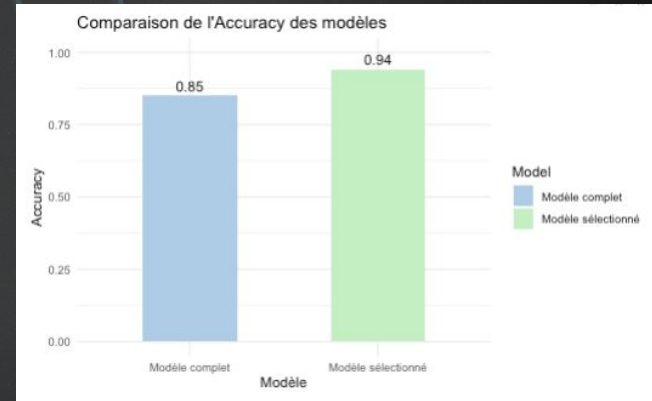
Suppression des variables P et K



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

Modèle sélectionné

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

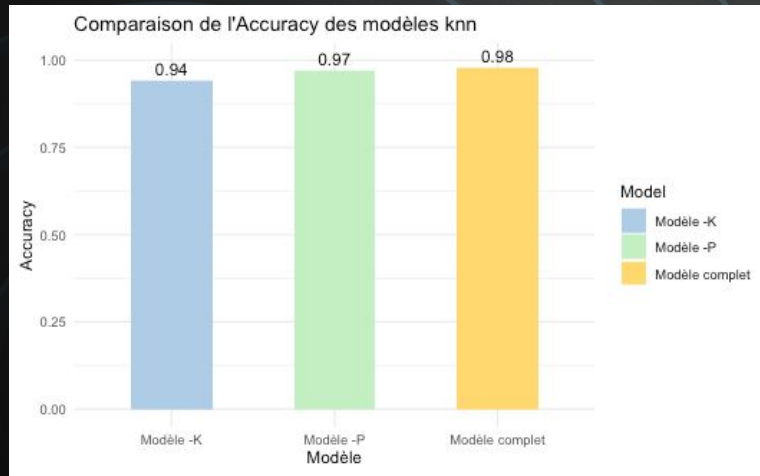


Accuracy = 0.94 > 0.85 → meilleur modèle **8**

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

C. Modèles simplifiés : KNN

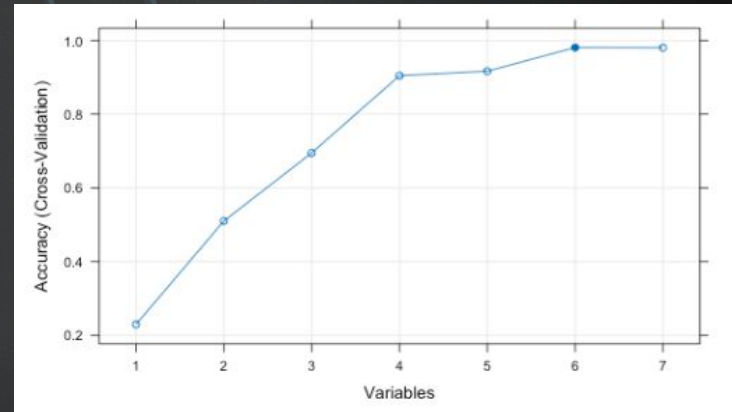
Suppression des variables P et K



→ Moins bonne performance que le modèle complet.

Modèle sélectionné

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

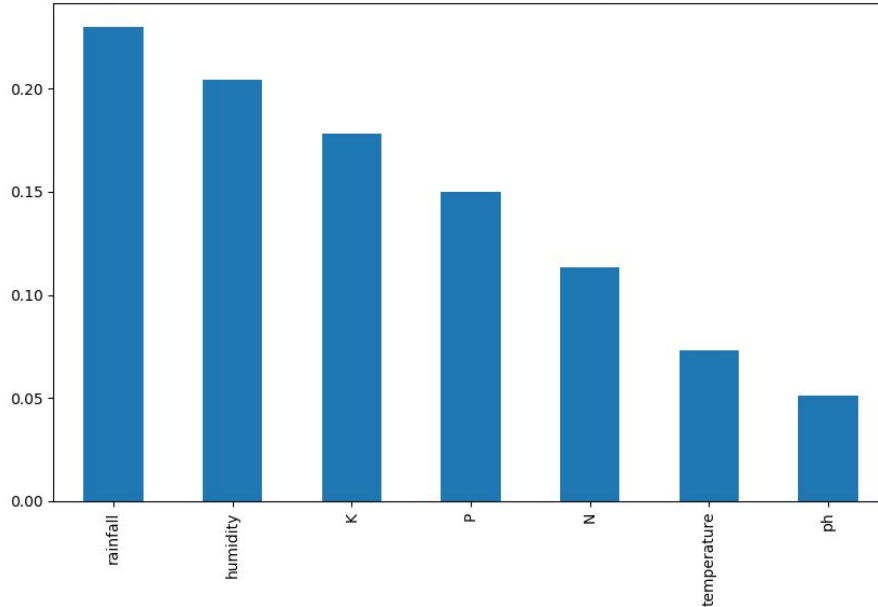


Accuracy = 0.9807 > 0.9801 → très proche **9**

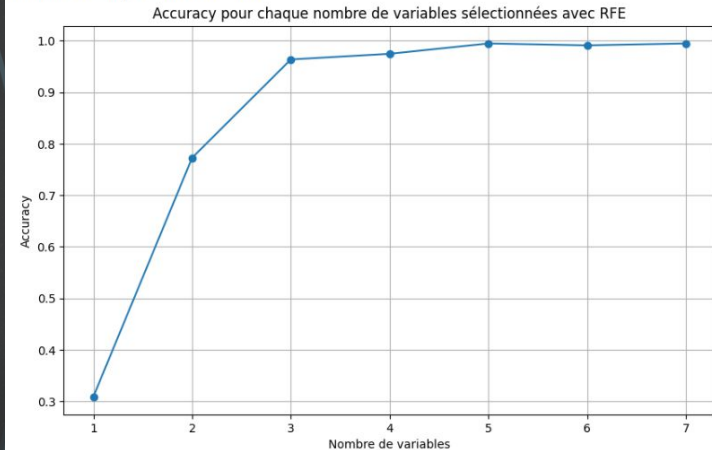
II - Prédiction de la meilleure culture pour un profil de sol donné

C. Modèles simplifiés : Random Forest

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

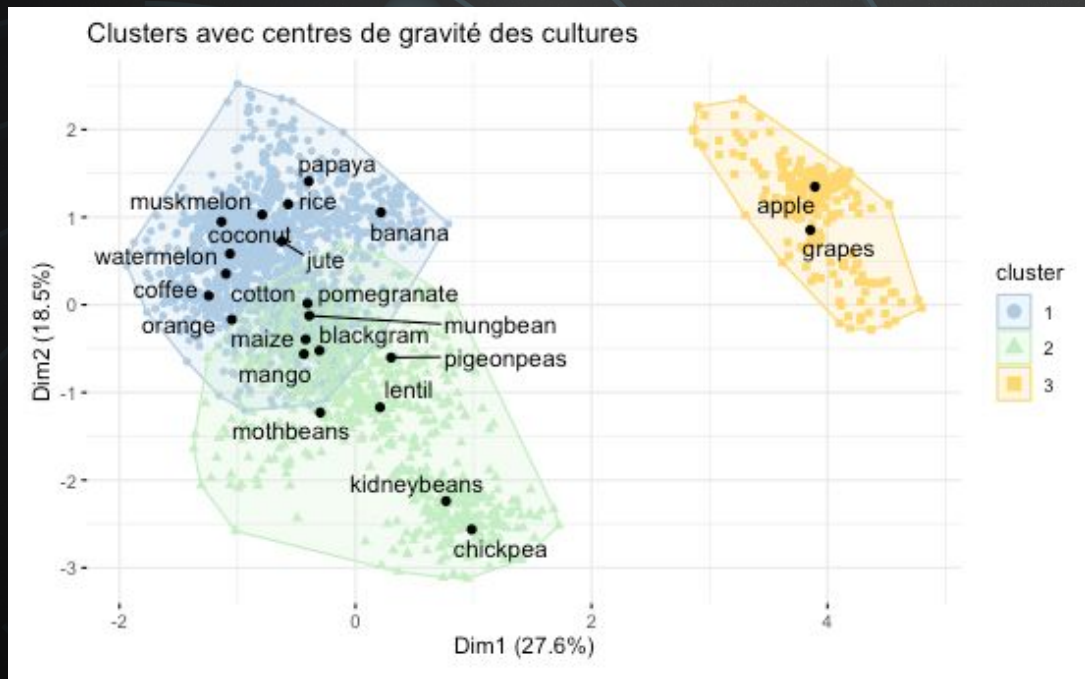


```
Variables sélectionnées avec 1 variables: ['humidity']
Accuracy avec 1 variables sélectionnées: 0.3090909090909091
Variables sélectionnées avec 2 variables: ['humidity', 'rainfall']
Accuracy avec 2 variables sélectionnées: 0.7727272727272727
Variables sélectionnées avec 3 variables: ['K', 'humidity', '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	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

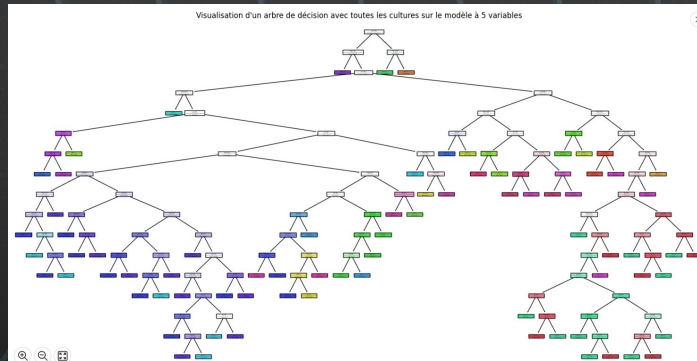
	Accuracy	Sensibilité	Spécificité
Modèle complet	0,93	Plus faible pour la classe jute : 0,79	Plus faible pour la classe riz : 0,94
Modèle sélectionné	0,94	Amélioration pour la classe maïs (1 contre 0,96)	Amélioration pour la classe coton (1 contre 0,96)

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

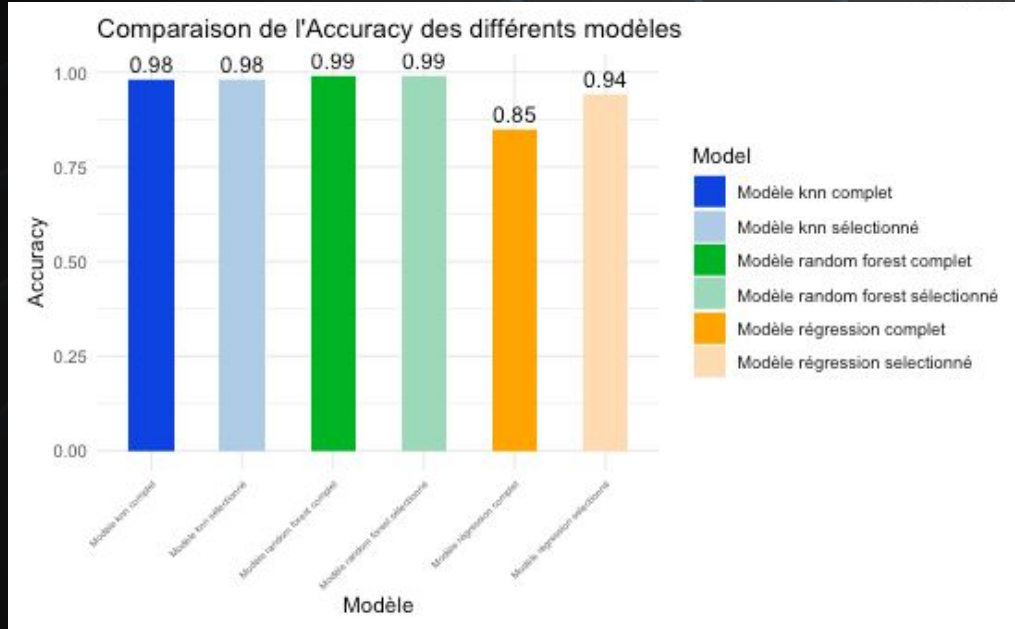
III - Test de la capacité du modèle à différencier les cultures similaires

B. Prédiction 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



IV - Conclusion



Classement des méthodes en fonction de leurs performances :

- 1) Random Forest
- 2) KNN
- 3) Régression logistique

Sélectionner les variables améliore fortement la performance du modèle pour la régression logistique

Perspectives :

Deep learning et réseaux de neurones

Ces méthodes permettent de prédire efficacement quelles cultures les indiens doivent planter en fonction des différentes caractéristiques physico-chimiques du sol → **augmentation du rendement** → **nourrir l'humanité**

Bibliographie

J. Ariedis, D. Japa, B. Daniel, and Y. Shi, "Towards Optimizing Data Analysis for Multi-dimensional Data Sets," in *Proceedings of the 2019 International Conference on Data Science and Knowledge Engineering for Sensing Decision Support (DSKD)*, Springer, 2019, pp. 614-625. doi: 10.1007/978-3-030-12388-8_43.

R. Katarya, A. Raturi, A. Mehndiratta, and A. Thapper, "Impact of Machine Learning Techniques in Precision Agriculture," *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, Jaipur, India, 2020, pp. 1-6. doi: 10.1109/ICETCE48199.2020.9091741.

M. Kuhn, "Recursive Feature Elimination," *The caret Package*, 2017. Available: <https://topepo.github.io/caret/recursive-feature-elimination.html>.

A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," *IEEE Access*, vol. 9, pp. 4843-4873, 2021. doi: 10.1109/ACCESS.2020.3048415.