**Le Machine Learning dans l’agriculture de précision**

**Da Costa Silva E, Lesage M, Lonchampt E.**

**Institut Agro de Rennes, Octobre 2024**

D'ici 2050, la population mondiale atteindra 9,1 milliards d'habitants, entraînant une augmentation de 70 % des besoins alimentaires. Cette situation, combinée à la réduction des terres agricoles disponibles en raison de l'urbanisation rapide, nécessite des innovations majeures dans la gestion agricole. L’agriculture de précision, un domaine en plein essor, utilise les données et la technologie pour optimiser la production agricole, en intégrant la variabilité biophysique des sols dans le processus décisionnel.

Le travail proposé a été réalisé à partir d’un jeu de données collecté en Inde, un pays où 70 % de la population pratique l’agriculture, contribuant à environ 17 % du PIB national. Le jeu de données inclut 2200 observations réparties sur huit variables : six variables physico-chimiques des sols, une variable météorologique et une variable décrivant le type de culture adaptée. L’objectif est d’utiliser le machine learning pour prédire les cultures les plus adaptées en fonction des caractéristiques du sol et de l'environnement, optimisant ainsi l’usage des ressources agricoles.

Les méthodes de machine learning envisagées pour cette étude comprennent un modèle de régression, un modèle basé sur la similarité tel que les k-nearest neighbors (KNN), ainsi qu’un modèle d’ensemble comme les forêts aléatoires (random forests). Ces deux derniers modèles sont particulièrement efficaces pour améliorer la précision des prédictions, car ils intègrent et exploitent les interactions complexes entre les différentes variables explicatives. Dans un premier temps, les trois modèles seront optimisés à l'aide de techniques de sélection de variables, notamment la méthode stepwise et l'élimination récursive de variables (RFE). Cela permettra de déterminer les caractéristiques les plus influentes pour la prédiction des cultures. Par la suite, l’accuracy des modèles complets et des modèles sélectionnés sera évaluée afin d’identifier celui qui offre la meilleure performance pour prédire la culture appropriée en fonction des caractéristiques du sol. En complément, une nouvelle série de prédictions sera réalisée sur un sous-ensemble de cultures similaires, sélectionnées à l’aide d’une classification ascendante hiérarchique. Cette étape vise à évaluer la capacité des modèles à différencier efficacement les cultures qui ont des besoins agronomiques proches.

En termes de perspectives, l’application de ces méthodes pourrait être étendue à d’autres régions géographiques et cultures, permettant une gestion plus fine et durable des ressources agricoles. De plus, l'intégration de méthodes plus avancées comme le deep learning, avec les réseaux de neurones multi-couches (MLP), ouvre des perspectives intéressantes pour améliorer encore la précision des recommandations.

**Mots-clés** : agriculture de précision, prédiction des cultures, KNN, random forest, échantillonnage