

Bienvenue !

Ce cours explore différentes méthodes de **Machine Learning** pour la **classification supervisée**, c'est-à-dire l'affectation d'individus à des **catégories discrètes** (*targets*) en fonction de **variables explicatives** (*features*), qu'elles soient continues ou catégorielles.

L'objectif est de **comparer les performances** de ces méthodes en évaluant :

- Leur **précision globale** sur l'ensemble des classes
- Leur **capacité à bien prédire les classes les plus rares**, souvent les plus difficiles à classifier.

Le dataset utilisé : Covertypes

Nous utilisons le **dataset Covertypes**, issu de l'[UCI Machine Learning Repository](#). Ce jeu de données comprend :

- **581 012 observations**
- **54 variables explicatives**
- **7 classes différentes** représentant des types de couvertures forestières

Les classes sont fortement **déséquilibrées numériquement**, ce qui constitue un défi supplémentaire pour les modèles de classification.

Objectifs du cours

Nous **expérimenterons plusieurs approches**, notamment :

- **Méthodes paramétriques** : régression logistique, LDA, QDA...
- **Méthodes non-paramétriques** : KNN, arbres de décision, forêts aléatoires...
- **Modèles à base de réseaux de neurones**
- **Stratégies One-Versus-All (OVA) et One-Versus-One (OVO)** pour les modèles binaires

Ce cours vise ainsi à comparer **l'efficacité de ces différentes méthodes** et à identifier celles qui offrent les meilleurs compromis en termes de **précision, robustesse et adaptation aux déséquilibres de classe**.