



# Machine learning avec R

---

## Référence et durée

R\_ML  
2 jours

## Objectifs

La formation répond à plusieurs objectifs

- Avoir une vision transverse des principaux algorithmes supervisés et non supervisés
- Savoir mettre en œuvre les algorithmes dans R sur un problème de classification ou de régression (apprentissage supervisé d'une cible binaire ou continue)

## Public et prérequis

Cette formation s'adresse à un public connaissant R d'analystes, data miners, chargés d'études statistiques, directeurs d'études. Des connaissances statistiques et informatiques sont préférables.

## Programme détaillé

- Introduction
  - Algorithmes, vocabulaire
  - Panorama des packages R pour l'apprentissage (caret, e1071, gbm ...)
  - Démarche d'apprentissage : feature extraction (décomposition de matrices, n-grams, hashing, tf-idf)
  - Démarche d'apprentissage : échantillonnage, validation, cross-validation, grid search, feature selection, métriques de validation (mape, AUC, ...), biais-variance, erreur d'apprentissage, erreur réelle
  - Démarche d'apprentissage : les algorithmes
    - Les méthodes supervisées : knn, SVM, Réseaux de neurones, réseau bayésien naif, régression pénalisée, boosting, random forest
    - Les méthodes non supervisées : Clustering, Règles d'associations et de séquences, Traitement du texte (text mining)
- Mise en pratique d'algorithmes non supervisés
  - Analyse de données textuelles : feature extraction, analyse latente, clustering
- Mise en pratique d'algorithmes supervisés
  - Exercice de mise en œuvre d'un modèle (apprentissage supervisé d'une cible binaire ou continue) à partir d'un grand nombre de variables
  - Mise en œuvre d'au moins 3 techniques parmi régression pénalisée, randomforest, boosting, SVM, réseaux de neurones
  - Métriques d'Evaluation et comparaison de modèles