

Programme de formation Data Science et Machine Learning avec R

● Objectifs

Statistiques descriptives, prédictives ou exploratoires, la Data Science repose sur la maîtrise de diverses techniques pour l'exploration des données. Cette formation de 4 jours vous permettra d'appréhender leur utilisation et de comprendre le principe de modélisation statistique. Les stagiaires seront en mesure de faire un choix entre régression et classification, ainsi que d'évaluer la performance d'un modèle.

● Pré requis

Connaissances fondamentales en statistiques et sur le langage R

● Durée

4 jours

● Public

Administrateurs, Analystes

● Plan de formation

Introduction à la formation et rappels sur le langage R

Les types de données dans R
Importation-exportation de données
Techniques pour tracer des courbes et des graphiques

Analyse en composantes

Analyse en Composantes Principales
Analyse Factorielle des Correspondances
Analyse des Correspondances Multiple
Exercices

Modélisation

Les algorithmes supervisés et non supervisés
Le choix entre la régression et la classification
Les étapes de construction d'un modèle

Algorithmes non supervisés

Le clustering hiérarchique
Le clustering non hiérarchique
Les approches mixtes
Exercices

Algorithmes supervisés

Le principe de régression linéaire univariée
La régression multivariée
La régression polynomiale
La régression logistique
Le Naive Bayes

L'arbre de décision
Les K plus proches voisins
Exercices

Procédures d'évaluation de modèles

Les techniques de ré-échantillonnage en jeu d'apprentissage, de validation et de test
Mesures de performance des modèles prédictifs
Matrice de confusion, de coût et la courbe ROC et AUC
Exercices

Analyse de données textuelles

Quelques packages utiles
Cas de la régression linéaire multiple
Cas de l'analyse en composantes principales
ACP
Cas de la classification CAH