

Programme de formation Data Science et Machine Learning avec Python

• Objectifs

Durant cette formation de 4 jours, les stagiaires apprennent à utiliser différentes méthodes d'exploration de données, ils comprennent le principe de la modélisation statistique, sont en mesure de choisir entre régression et classification selon leur problématique et peuvent évaluer les performances prédictives d'un modèle donné.

• Pré requis

Des connaissances de base en statistiques et sur Python

• Durée

4 jours

• Public

Administrateurs, analystes, développeurs

• Plan de formation

Introduction à la formation et rappels sur le langage Python

Les types de données dans Python
Importation-exportation de données
Techniques pour tracer des courbes et des graphiques
Introduction au logiciel Jupiter Notebook

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