



BIG DATA ANALYTICS AVEC PYTHON

Public	Responsables Infocentre (Datamining, Marketing, Qualité...), utilisateurs et gestionnaires métiers de bases de données.
Durée	4 jours - 28 heures
Pré-requis	Connaissances de base en Python. Connaissances de base en statistiques
Objectifs	Comprendre le principe de la modélisation statistique Choisir entre la régression et la classification en fonction du type de données Évaluer les performances prédictives d'un algorithme Créer des sélections et des classements dans de grands volumes de données pour dégager des tendances
Méthodes pédagogiques	Pour bien préparer la formation, le stagiaire remplit une évaluation de positionnement et fixe ses objectifs à travers un questionnaire. La formation est délivrée en présentiel ou distanciel (e-learning, classe virtuelle, présentiel et à distance). Le formateur alterne entre méthodes démonstratives, interrogatives et actives (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification. Cette formation est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par Audit Conseil Formation.
Moyens techniques	1 poste de travail complet par personne De nombreux exercices d'application Mise en place d'ateliers pratiques Remise d'un support de cours Passage de certification(s) dans le cadre du CPF Remise d'une attestation de stage
Modalité d'évaluation des acquis	Evaluation des besoins et objectifs en pré et post formation Evaluation technique des connaissances en pré et post formation Evaluation générale du stage
Délai d'accès	L'inscription à cette formation est possible jusqu'à 5 jours ouvrés avant le début de la session
Accessibilité handicapés	Au centre d'affaires ELITE partenaire d'ACF à 20 m. Guide d'accessibilité à l'accueil.

INTRODUCTION À LA MODÉLISATION

- Introduction au langage Python.
- Introduction au logiciel Jupiter Notebook.
- Les étapes de construction d'un modèle.
- Les algorithmes supervisés et non supervisés.
- Le choix entre la régression et la classification.

PROCÉDURES D'ÉVALUATION DE MODÈLES

- Les techniques de ré-échantillonage en jeu d'apprentissage, de validation et de test.
- Test de représentativité des données d'apprentissage.
- Mesures de performance des modèles prédictifs.
- Matrice de confusion, de coût et la courbe ROC et AUC.

LES ALGORITHMES SUPERVISÉS

- Le principe de régression linéaire univariée.
- La régression multivariée.
- La régression polynomiale.
- La régression régularisée.
- Le Naïve Bayes.
- La régression logistique.

LES ALGORITHMES NON SUPERVISÉS

- Le clustering hiérarchique.
- Le clustering non hiérarchique.
- Les approches mixtes.

ANALYSE EN COMPOSANTES

- Analyse en composantes principales.
- Analyse factorielle des correspondances.
- Analyse des correspondances multiples.
- Analyse factorielle pour données mixtes.
- Classification hiérarchique sur composantes principales.

ANALYSE DE DONNÉES TEXTUELLES

- Collecte et prétraitement des données textuelles.
- Extraction d'entités primaires, d'entités nommées et résolution référentielle.
- Étiquetage grammatical, analyse syntaxique, analyse sémantique.
- Lemmatisation.
- Représentation vectorielle des textes.
- Pondération TF-IDF.
- Word2Vec.

NOUS CONTACTER

Siège social

16, ALLÉE FRANÇOIS VILLON
38130 ÉCHIROLLES

Téléphone

04 76 23 20 50 - 06 81 73 19 35

Suivez-nous sur les réseaux sociaux, rejoignez la communauté !

 ACF Audit Conseil Formation



@ACF_Formation

Dernière mise à jour : 11/09/2025

PROFIL Formateur : Les formateurs sont recrutés selon plusieurs critères :
Expérience, pédagogie, dynamisme et prévoyance.