BIG DATA - PYTHON POUR L'ANALYSE DE DONNÉES

Public	Développeurs en Python, responsables infocentre, développeurs de logiciels, programmeurs, data analysts, data scientists.
Durée	3 jours - 21 heures
Pré-requis	Maîtrise de la programmation Python
Objectifs	Comprendre le principe de la modélisation statistique Savoir utiliser les principaux outils de traitement et d'analyse de données pour Python Savoir appliquer les pratiques optimales en matière de nettoyage et de préparation des données avant l'analyse Choisir entre la régression et la classification en fonction du type de données Apprendre à mettre en place un modèle d'apprentissage simple Être capable d'extraire des données d'un fichier
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
Planning	Du 28/07/2025 au 30/07/2025 Du 20/10/2025 au 22/10/2025 Du 17/11/2025 au 19/11/2025
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.

PRÉSENTATION DE L'ÉCOSYSTÈME PYTHON SCIENTIFIQUE

- Panorama de l'écosystème scientifique de Python : les librairies incontournables.
- Savoir ou trouver de nouvelles librairies et juger de leur pérennité.
- Les principaux outils et logiciels open source pour la data science.

TRAVAILLER LES DONNÉES AVEC PYTHON

- Le socle scientifique Python : la SciPy Stack.
- Les bonnes pratiques pour bien démarrer votre projet de data science avec Python.
- Les formats de fichiers scientifiques et les librairies pour les manipuler.
- Pandas : l'analyse de données tabulaires (fichiers csv, excel...), statistiques, pivots, filtres, recherche...
- Numpy : calcul numérique et algèbre linéaire (les vecteurs, matrices, images).
- L'extraction des données,la préparation , le nettoyage.

INTRODUCTION À LA MODÉLISATION

- 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-échantillonnage 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 Naive Bayes.
- La régression logistique.

LES ALGORITHMES NON SUPERVISÉS

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

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 : 29/09/2023