

Formation Big Data, Python pour l'analyse de données

Présentation

Cette formation vous enseignera les fondamentaux du traitement des données en Python, de la collecte de données brutes au nettoyage et à la mise en forme de celles-ci afin de les exploiter pour créer de modèles prédictifs, établir des statistiques et des indicateurs, et bien plus encore.

Durée : 21,00 heures (3 jours)

Tarif INTRA : Nous consulter

Objectifs de la formation

- Comprendre les principes de la modélisation statistique,
- Savoir utiliser les principaux outils d'analyse et de traitement des données de Python,
- Extraire des données d'un fichier,
- Savoir appliquer les bonnes pratiques pour nettoyer et préparer les données avant analyse,
- Créer des modèles d'apprentissage simples,
- Choisir entre régression et classification en fonction du type de données,
- Évaluer les performances prédictives d'un algorithme,
- Créer des sélections et des évaluations dans de grands volumes de données pour identifier les tendances.

Prérequis

- Connaissances de base en Python et statistiques.

Public

- Développeurs en Python,
- Responsables Infocentre,
- Développeurs de logiciels,
- Programmeurs,
- Data analysts,
- Data scientists.



Programme de la formation

Introduction à la structure des données en Python

- Python pour les manipulations de données
- Listes, ensembles, strings, tuples et dictionnaires

Opérations avancées sur la structure de données intégrée

- Structures de données avancées
- Opérations de base sur les fichiers en Python

Introduction à NumPy, Pandas et Matplotlib

- Les tableaux NumPy
- Les DataFrames Pandas
- Statistiques et visualisation avec NumPy et Pandas
- Utiliser NumPy et Pandas pour calculer des statistiques descriptives de base sur le DataFrame

Data Wrangling avec Python

- Définition et processus
- Sous-ensembles, filtrage et regroupement
- Détection des valeurs extrêmes et traitement des valeurs manquantes
- Concaténer, fusionner et joindre
- Les méthodes utiles de Pandas
- Compréhension avancée des listes et fonction zip
- Formatage des données

Scraping de données avec Python

- Définition du scraping
- Les différents niveaux de difficulté sur plusieurs supports
- Lecture de données provenant de différentes sources textuelles, et non textuelles.
- Outils de scraping
- Introduction à BeautifulSoup

- Introduction à CSS Select

Scraping avancé et collecte de données

- Les bases du Web scraping, bibliothèques BeautifulSoup
- Python comme solution ETL
- Formats de données structurées : CSV, flux XML et JSON
- Lire et écrire des fichiers
- Exploiter des données de fichiers provenant de différentes sources
- Fonctions pour l'accès et le chargement de données en blocs de lignes

Implémenter un scraper

- Scraper simple avec les requêtes GET et les pages séquencées
- Identification de la stratégie de navigation sur le site pour trouver les données
- Construction de l'algorithme du scraper
- Scraper complexe : envoi de données à un site Web, obtention de résultats plus complexes
- Requêtes POST et GET
- Parcourir un site pour trouver les données
- Identifier la stratégie à adopter
- Coder le scraper

Application dans la vie réelle et conclusion du cours

- Appliquer vos connaissances à des tâches de data wrangling et data scraping de la vie réelle

Organisation

Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

Moyens pédagogiques et techniques

- Apports des connaissances communes.
- Mises en situation sur le thème de la formation et des cas concrets.
- Méthodologie d'apprentissage attractive, interactive et participative.
- Equilibre théorie / pratique : 60 % / 40 %.
- Supports de cours fournis au format papier et/ou numérique.
- Ressources documentaires en ligne et références mises à disposition par le formateur.
- Pour les formations en présentiel dans les locaux mis à disposition, les apprenants sont accueillis dans une salle de cours équipée d'un réseau Wi-Fi, d'un tableau blanc ou paperboard. Un ordinateur avec les logiciels appropriés est mis à disposition (le cas échéant).

Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation

En amont de la formation

- Recueil des besoins des apprenants afin de disposer des informations essentielles au bon déroulé de la formation (profil, niveau, attentes particulières...).
- Auto-positionnement des apprenants afin de mesurer le niveau de départ.

Tout au long de la formation

- Évaluation continue des acquis avec des questions orales, des exercices, des QCM, des cas pratiques ou mises en situation...

A la fin de la formation

- Auto-positionnement des apprenants afin de mesurer l'acquisition des compétences.
- Evaluation par le formateur des compétences acquises par les apprenants.
- Questionnaire de satisfaction à chaud afin de recueillir la satisfaction des apprenants à l'issue de la formation.
- Questionnaire de satisfaction à froid afin d'évaluer les apports ancrés de la formation et leurs mises en application au quotidien.

NB : dans le cadre d'une Action collective, chaque stagiaire bénéficiaire sera contacté par un prestataire choisi par l'Opco Atlas afin d'évaluer « à chaud » la qualité de la formation suivie.

Accessibilité

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.