

Formation Python, Coder avec Pandas, NumPy & Scikit-learn

Présentation

Ce cours intensif de 5 jours vous initie à l'utilisation de Python dans le contexte de la Data Science, en vous appuyant sur les bibliothèques incontournables :

- NumPy pour les calculs numériques,
- Pandas pour la manipulation de données tabulaires,
- Matplotlib/Seaborn pour la visualisation,
- Scikit-learn pour les modèles de machine learning.

À travers des cas concrets et des exercices pratiques, vous apprendrez à importer, transformer, visualiser et modéliser des jeux de données.

Durée : 35,00 heures (5 jours)

Tarif INTRA : Nous consulter

Objectifs de la formation

À l'issue de la formation, les participants seront capables de :

- Utiliser NumPy pour les calculs numériques sur tableaux multidimensionnels
- Manipuler efficacement des données tabulaires avec Pandas
- Créer des visualisations claires avec Matplotlib et Seaborn
- Préparer les données pour la modélisation (feature engineering, normalisation)
- Appliquer des algorithmes de machine learning avec Scikit-learn
- Évaluer les performances d'un modèle prédictif

Prérequis

- Bonnes bases en Python (variables, fonctions, boucles, listes)
- Connaissances de base en statistiques descriptives (moyenne, écart-type, corrélation...)
- Notions de manipulation de fichiers CSV ou Excel

Public

- Analystes de données, data analysts



- Développeurs ou ingénieurs souhaitant se spécialiser en Data Science
- Responsables métier ou chefs de projet souhaitant monter en compétences sur la donnée
- Toute personne intéressée par l'analyse et la modélisation des données avec Python

Programme de la formation

Jour 1 : Fondamentaux et NumPy

- Introduction à Python pour la Data Science
- Jupyter Notebook et environnement Anaconda
- Introduction à NumPy
- Tableaux ndarray, indexation, slicing
- Fonctions mathématiques vectorisées
- Opérations sur matrices, statistiques de base
- Génération de données aléatoires (simulateurs, distributions)
- Manipulation d'axes, reshape, broadcast
- Introduction aux visualisations avec Matplotlib

TP

- Simulation de données + opérations statistiques sur matrices

Jour 2 : Manipulation de données avec Pandas

- Séries et DataFrames : création, import/export (CSV, Excel, JSON)
- Accès aux données (loc, iloc, at, iat)
- Nettoyage : gestion des valeurs manquantes, duplication, typage
- Filtres conditionnels, tri, regroupement (groupby)
- Agrégations et statistiques descriptives
- Merge, join, concaténation de DataFrames

TP

- Nettoyage et analyse d'un jeu de données (ex : Airbnb, Titanic, données RH)

Jour 3 : Visualisation de données

- Visualisation avec Matplotlib : courbes, histogrammes, barres
- Introduction à Seaborn : styles, palettes, intégration avec Pandas
- Visualisation statistique : boxplots, heatmaps, pairplots
- Analyse exploratoire de données (EDA)
- Détection d'anomalies visuelles, corrélations, distribution
- Graphiques croisés dynamiques

TP

- Exploration graphique d'un dataset complet (ex : santé, marketing, transport)

Jour 4 : Introduction au Machine Learning avec Scikit-learn

- Introduction au Machine Learning : supervision vs non-supervision
- Séparation train/test, train_test_split, pipelines
- Modèles supervisés : régression linéaire, k-NN, arbre de décision
- Évaluation de modèle : précision, rappel, F1-score, courbe ROC
- Méthodes de validation croisée (cross_val_score)
- Feature scaling, normalisation

TP

- Prédiction d'un score (ex : prix d'un logement, revenu, classification binaire)

Jour 5 : Atelier projet & approfondissements

- Feature engineering : création de variables, encodage
- Algorithmes avancés (Random Forest, SVM – introduction)
- GridSearch et tuning des hyperparamètres

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.

Riches de leur expérience sur le sujet, ils sauront accompagner vos collaborateurs dans leur montée en compétence.

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.
- Équilibre 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.
- Évaluation 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.

Accessibilité

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