

Programmation Python (préparation des pré-requis au développement d'IA)

5 j (35 heures)

Ref : IAD005

Public

Développeurs, administrateurs et architectes

Pré-requis

Connaissances de base en programmation (variables, boucles, fonctions...)
Connaissance en machine learning
Connaissance de statistiques descriptives et de probabilités
Expérience professionnelle en analyse de données appréciée mais non indispensable

Moyens pédagogiques

Formation réalisée en présentiel ou à distance selon la formule retenue
Nombreux exercices pratiques et mises en situation, échanges basés sur la pratique professionnelle des participants et du formateur, formation progressive en mode participatif. Vidéoprojecteur, support de cours fourni à chaque stagiaire

Modalités de suivi et d'évaluation

Feuille de présence émarginée par demi-journée par les stagiaires et le formateur
Exercices de mise en pratique ou quiz de connaissances tout au long de la formation permettant de mesurer la progression des stagiaires
Questionnaire d'évaluation de la satisfaction en fin de stage
Auto-évaluation des acquis de la formation par les stagiaires
Attestation de fin de formation

Cette formation intensive de 5 jours vous permettra de maîtriser les fondamentaux du langage Python et son utilisation pour la data science. Vous apprendrez à manipuler les structures de données, développer des scripts et exploiter les principales bibliothèques Python dédiées à l'analyse de données comme Numpy, Pandas, Matplotlib et Scikit-learn. À travers de nombreux exercices pratiques, vous découvrirez comment collecter, nettoyer, transformer et visualiser des données. Vous serez également initié aux bases du machine learning en mettant en œuvre des algorithmes supervisés et non supervisés.

À l'issue de la formation, vous disposerez de solides compétences en Python pour réaliser des projets complets de data science.

Objectifs

Maîtriser la syntaxe et les concepts clés de Python
Savoir manipuler les différents types de données et structures de contrôle
Être capable de développer des scripts et programmes Python modulaires
Connaître les bonnes pratiques de programmation et de gestion de projet

Programmation Python (préparation des pré-requis au développement d'IA)

- Découvrir l'écosystème data science de Python et ses principales librairies
- Savoir explorer et visualiser des données avec Pandas et Matplotlib
- Comprendre les bases du machine learning et les différents types d'algorithmes
- Mettre en œuvre des modèles supervisés et non supervisés avec Scikit-learn
- Être en mesure d'appliquer Python à des projets concrets de data science

Programme détaillé

FONDAMENTAUX DE PYTHON

INTRODUCTION A PYTHON

- Historique et philosophie du langage
- Installation et prise en main de l'environnement de développement
- Syntaxe de base : indentation, commentaires, variables, types de données...

STRUCTURES DE DONNEES PYTHON

- Listes, tuples, dictionnaires, ensembles
- Slicing, compréhensions de listes
- Manipulation et méthodes des différentes structures

STRUCTURES DE CONTROLE

- Instructions conditionnelles (if/elif/else)
- Boucles for et while
- Fonctions et passage d'arguments

BONNES PRATIQUES DE PROGRAMMATION

- Conventions de nommage, PEP 8
- Gestion des erreurs et exceptions
- Modules et packages

INTRODUCTION A LA PROGRAMMATION ORIENTEE OBJET (POO)

- Concepts de base de la POO
- Définition de classes et d'objets en Python
- Héritage et polymorphisme

PYTHON POUR LA DATA SCIENCE

PRESENTATION DE L'ECOSYSTEME DATA SCIENCE PYTHON

Vue d'ensemble des bibliothèques essentielles : Numpy, Pandas, Matplotlib...

Jupyter Notebook pour les projets data science

NUMPY POUR LE CALCUL SCIENTIFIQUE

Objets ndarray et vectorisation des calculs

Slicing, broadcasting, masques booléens

Algèbre linéaire et statistiques avec Numpy

INTRODUCTION A PANDAS

Structures de données Pandas : Series et DataFrame

Lecture/écriture de fichiers CSV, Excel, SQL...

Extraction et filtrage de données

MANIPULATIONS DE DONNEES AVEC PANDAS

Opérations sur les indexes et les colonnes

Transformation de variables (apply, map, cut...)

Gestion des données manquantes

Concaténation et fusion de DataFrame

INTRODUCTION A LA DATA VISUALISATION AVEC MATPLOTLIB

Création de graphiques simples : courbes, histogrammes, camemberts...

Personnalisation des graphiques

Figures, subplots et multiples graphiques

DATA VISUALISATION ET STATISTIQUES

DATA VISUALISATION AVANCEE AVEC SEABORN

Graphiques statistiques : dispersion, densité, pairplot...

Utilisation de données catégorielles

Personnalisation avancée des graphiques

STATISTIQUES AVEC SCIPY

Statistiques descriptives

Tests statistiques paramétriques et non-paramétriques

PROJET DIRIGE : ANALYSE EXPLORATOIRE D'UN JEU DE DONNEES

Import, nettoyage et transformation des données

Analyses statistiques univariées et multivariées

Visualisations avancées et interprétation des résultats

INTRODUCTION AU MACHINE LEARNING

CONCEPTS FONDAMENTAUX DU MACHINE LEARNING

Types d'apprentissage : supervisé, non supervisé, par renforcement
Compromis biais-variance, overfitting et régularisation
Évaluation et validation croisée

ALGORITHMES DE ML SUPERVISES AVEC SCIKIT-LEARN

Pré-processing des données : encodage, scaling...
Entraînement et évaluation d'un modèle
Algorithmes de classification (KNN, régression logistique, SVM...)
Algorithmes de régression (linéaire, polynomiale, régularisée...)

ALGORITHMES DE ML NON SUPERVISES AVEC SCIKIT-LEARN

Réduction de dimension (ACP, t-SNE...)
Clustering (K-means, DBScan...)

INTERPRETATION ET EXPLICABILITE DES MODELES

Importance des variables
Outils d'explicabilité (LIME, SHAP...)

PROJET DIRIGE : PREDICTION SUPERVISEE SUR UN JEU DE DONNEES REEL

Analyse du problème et des données
Feature engineering et sélection de variables
Comparaison et optimisation de différents algorithmes
Interprétation des résultats

DATATHON : COMPETITION KAGGLE

PRESENTATION DU CHALLENGE ET DES DONNEES

Analyse exploratoire et feature engineering
Essais de différentes approches de modélisation
Soumission et évaluation des résultats

PRESENTATION DES PROJETS DEVANT LA CLASSE

Pitch des démarches et des résultats obtenus
Retours du formateur et des autres participants

CONCLUSION ET PERSPECTIVES

Récapitulatif des points clés de la formation

Conseils pour approfondir et monter en compétences

Panorama des métiers et des opportunités en data science
