

POE Développeurs Python DataScience

57 j (399 heures)

Ref : POE-PYD

Public

Bac +2 à +3 avec expérience significative en Informatique
Bac +5 à +8 Scientifique ou Informatique

Pré-requis

Réussite de nos tests de recrutement
Bon relationnel, ouvert, curieux, communicant
Niveau correct en Anglais

Moyens pédagogiques

Formation réalisée en présentiel ou à distance selon la formule retenue, en continu sur 3 mois - dans certains cas, une période de stage de 5 jours pourra être prévue
Un poste par stagiaire, vidéoprojecteur ou écran interactif tactile, support de cours fourni à chaque stagiaire
Exposés, discussions techniques, démonstrations, exercices, mise en application sur un TP/projet fil rouge

Modalités de suivi et d'évaluation

Feuille de présence émargée 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

Le monde accumule une quantité extraordinaire de données dont une grande partie est publique. Afin de les analyser, de leur donner du sens et d'en dégager des informations exploitables, elles font appel aux scientifiques des données. Le data Scientist, expert de la données, transforme la donnée brute (d'entreprise ou disponible en opendata) en données ayant du sens pour son destinataire, qui souhaite se développer et être compétitif. Il interprète et analyse la donnée de sources multiples en amenant des solutions imaginatives aux problèmes. Il pourra utiliser des outils de programmation performants pour lire et nettoyer la donnée et utiliser les outils mathématiques avancés pour une analyse performante et révéler la véritable valeur de ces données.

Objectifs

Analyser, filtrer de la donnée brute
Résoudre les problèmes de son entreprise grâce à l'analyse de données

- Présenter ses résultats sous forme de graphiques
- Utiliser les statistiques et le Machine Learning pour une analyse performante des données
- Déterminer la meilleure manière de répondre aux besoins métier
- Définir les algorithmes d'analyses les plus pertinents
- Développer des modèles descriptifs et prédictifs
- Exploiter de larges volumes de données non structurées

Programme détaillé

TEAMBOOSTER - 8H

- Cohésion du groupe, travail sur le savoir-être
- points faibles, points forts et axes de progression
- présentation du projet

BASES DE PYTHON - 42H

- Environnement interactif Jupyter (Notebook, Lab)
- environnement de développement pour Python
- Linters, types de base et opérateurs, listes, tuples et dictionnaires, boucles et conditionnelles
- fonctions, modules et packages

CALCULS ET STATISTIQUES - 35H

- Frameworks Numpy et Scipy
- les tableaux et matrices
- opérations mathématiques sur les tableaux
- statistiques, algèbre linéaire, régression Linéaire, interpolation
- trigonométrie, matrices Masquées, algèbre complexe transformée de Fourier

TRAITEMENT DE LA DONNÉE - 35H

- Framework Pandas
- les Dataframes, différentes sources de données
- nettoyer, convertir les données
- croiser, concaténer des données
- analyser les données (groupage, partitionnement, statistiques)
- exporter les données

GRAPHIQUES - 35H

- Framework Matplotlib
- graphiques en 2D et en 3D
- gestion des couleurs, titres, légendes et graduation
- produire des multi-graphiques,

sauvegarde et intégrer un graphique, librairie Seaborn pour graphiques avancés

PROJET INTERMÉDIAIRE - 35H

TP et mise en application des modules abordées ci-dessus.

GESTION DE PROJET / AGILITÉ - 14H

Présentation Cycle en V, AMOE, AMOA,
exemple de procédure, agilité
Méthodologie Scrum
gestion du temps, définition et conduite d'un projet
responsabilité du chef de projet
réutilisation d'un projet dans un autre

BASE DE DONNÉES SQL POSTGRESQL - 35H

Présentation, outils de gestion d'une base de données
modélisation de base de données
requêtes en SQL
procédures stockées, vues, indexation, postGIS

BASE DE DONNÉES NO SQL - 21H

Base de données MongoDB : création et interrogation, base de données
hadoop : création et map/reduce

PYTHON SIG - 21H

Système de projection
dataframes GeoPandas
modules python de Cartographie
dialogue avec logiciels SIG : Qgis/ArcGis

PYTHON ML - 42H

Machine Learning, Deep Learning frameworks
tensorFlow et Keras, framework Spark, framework Scikit-learn
réseau de neurones, réseaux de neurones convolutifs
(CNN)
réseaux de neurones récurrents (RNN)
régression Linéaire, K-voisins, classification, apprentissage, prédiction, arbre de décision, forêts
aléatoires, qualité d'un modèle, théorème
No Free Lunch, détection d'anomalie
Isolation Forest

DEVOPS - 7H

Git
Pip
Environnements virtuels

OUTILS BI - 28H

Data Lake, data Warehouse, Data Mart
visualisation des données avec Power BI
utilisation d'un ETL avec Talend

POSTURE DU CONSULTANT - 7H

Rôle attendu des collaborateurs, relai équipe technique / fonctionnelle
gestion des situations difficiles
développer le compte Client par votre rôle de conseil
comprendre les enjeux et intégrer les étapes de la relation Client
développer l'état d'esprit
créer une proximité et renforcer la confiance, reconnaître et agir face aux opportunités commerciales,
cultiver
une démarche de disponibilité sans déborder ses propres limites, gagner en agilité et créativité

PROJET FINAL + SOUTENANCE - 35H

Mise en application des modules de formation abordées tout au long du parcours
Soutenance projet
