+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Mis à jour le 23/01/2024

S'inscrire

Formation Python : Analyse de Données

3 jours (21 heures)

Présentation

Python est l'un des langages le plus flexibles et polyvalents. Utilisés dans énormément de domaines, Python est le langage de l'année 2022 selon l'index TIOBE. La simplicité de sa syntaxe simplifie son apprentissage, il s'agit du langage de programmation le plus utilisé en Big Data.

Avec la montée en popularité du Big Data ces dernières années, Python est devenu un langage incontournable. Son écosystème d'API lui permet de traiter des données de différents types (SQL et NoSQL) et de piloter des outils de traitement avancé.

Notre formation Python : analyse de données vous enseignera la résolution de problèmes grâce à l'analyse de données. Vous apprendrez les principes de la modélisation statistique et des outils de traitement. Vous verrez également comment extraire les données d'un fichier et mettre en place un modèle d'apprentissage simple.

À l'issue de cette formation, vous serez capable de manipuler des données statistiques grâce au langage Python. Nous vous présenterons la dernière version du langage Python 3.12.

Objectifs

- Appréhender le principe de la modélisation statistique
- Maîtriser les outils de traitement et d'analyse de données pour Python
- Savoir extraire des données d'un fichier
- Appliquer les meilleures pratiques en matière de nettoyage et de préparation des données
- Savoir mettre en place un modèle d'apprentissage simple
- Choisir la classification et la régression en fonction du type de données
- Optimiser les performances prédictives d'un algorithme
- Créer des classements et des sélections dans de grands volumes de données

Public visé

- Développeurs en Python
- Responsables Infocentre
- Développeurs de logiciels
- Programmeurs
- Data Analysts
- Data Scientists

Pré-requis

- Connaissances de base sur Python
- Connaissances de base en statistiques

Pré-requis logiciel

Les dernières versions d'Anaconda et Python installés.

Programme de notre formation Big Data : Python pour l'Analyse de Données

Présentation des librairies

- Pandas
- NumPy et SciPy
- Matplotlib
- Scikit-Learn
- Utiliser des fonctions de librairies
- Gérer les modules et librairies

Préparation des données

- L'importance d'avoir des données intègres et préparées
- Lire et écrire des fichiers CSV
- Importer des ensembles de données
- Nettoyer et préparer les données
- Formatage des données
- Construire des pipelines de données

Modéliser ses données avec Pandas

- Manipuler ses données

- Créer des colonnes
- Transformer ses colonnes
- Interrogation, triage et filtrage
- Regroupement et agrégation pour la synthèse des données
- Fusionner et joindre des données provenant de plusieurs sources
- La concaténation
- Sélectionner des éléments
- Remplacer des valeurs

Visualisation des données avec Pandas

- Effectuer des comparaisons
- Créer des graphiques :
  - Histogramme
  - Pie chart
  - Graphique de densité Kernel
  - Graphique hexagonal
  - Scatter chart
  - Line chart
  - Boxplot
- TP : Trouver des insights à partir de ses visualisations

Calcul numérique avec NumPY

- Passer des listes Python aux tableaux NumPy
- Travailler avec des tableaux multidimensionnels
- Opérations sur les arrays, découpage et diffusion
- Les fonctions d'algèbre linéaire et matricielle
- Les principales fonctions mathématiques

Analyse statistique avec SciPy

- Les bonnes pratiques pour une analyse de données réussie
- Les 3 classes statistiques de SciPy
- Utiliser les fonctions de statistiques descriptives
- Calculer le z-score et le t-score
- TP : Analyse de plusieurs dataframes différents

Développement des modèles de machine learning

- Les points forts et les limites du machine learning
- Choisir ses modèles à l'aide de la visualisation
- Les fonctions d'optimisation et le preprocessing avec SciPy
- Les différents modèles d'apprentissage avec SciKit-Learn

Régression avec Scikit-learn

- Régression linéaire simple
- Régression linéaire multiple
- La régression logistique
- La régression polynomiale
- R-carré et EQM pour évaluer son modèle
- Prédiction et prise de décision

Classification avec Scikit-learn

- SVM
- Nearest neighbors
- Random forest
- K-Means

Algorithmes Data

- Découvrir des algorithmes destinés à l'analyse de données
- Mettre en œuvre des algorithmes d'analyse de données
- Évaluer les performances prédictives

Sociétés concernées

Cette formation s’adresse à la fois aux particuliers ainsi qu’aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d’auto-évaluation nous permettant d’apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d’anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l’acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.

Organisme de formation enregistré sous le numéro 11 75 54743 75. Cet enregistrement ne vaut pas agrément de l’État.
© Ambient IT 2015-2024. Tous droits réservés. Paris, France - Suisse - Belgique - Luxembourg