Data Scientist, temps partiel (9 mois)
by DataScientest

Formation en ligne - 15h00
Réf : 4DB - Prix 2025 : 7 190€ HT

Devenez expert en développement de modèles d'analyse de données avec ce parcours proposé par notre partenaire DataScientest. Un Data Scientist développe des modèles d'analyse complexes visant à extraire des informations à partir de bases de données. Cette formation certifiante se déroule à distance dans un format hybride mêlant temps d'échanges synchrones avec un formateur expert, exercices pratiques et modules E-learning. Basée sur la pédagogie Learning By Doing, vous réaliserez un projet fil rouge en équipe afin de mettre en pratique vos connaissances. Lors de votre inscription, vous serez rattaché à l'une des promotions Datascientest.

A l'issue de cette formation, vous obtiendrez un certificat « Data Scientist » des Mines Paris- PSL Executive ainsi que le bloc de compétence 2 de la certification RNCP « Expert en ingénierie de l'intelligence artificielle».

Contactez-nous dès maintenant pour connaître les prochaines dates !

PARTICIPANTS
Toutes les personnes ayant une appétence pour la programmation et la manipulation des données.

PRÉREQUIS
Un diplôme ou un titre de niveau bac+3 ainsi qu'un bon niveau en mathématiques.
Pour les candidats ne présentant pas le niveau de qualification requis, une dérogation est possible sur dossier.

COMPÉTENCES DU FORMATEUR
Les experts qui ont conçu la formation et qui accompagnent les apprenants dans le cadre d'un tutorat sont des spécialistes des sujets traités. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION
La progression de l'apprenant est évaluée tout au long de sa formation au moyen de QCM, d'exercices pratiques, de tests ou d'échanges pédagogiques.
Sa satisfaction est aussi évaluée à l'issue de sa formation grâce à un questionnaire.

MOYENS PÉDAGOGIQUES ET TECHNIQUES
Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : documentation et support de cours, exercices pratiques d'application et corrigés des exercices, études de cas ou présentation de cas réels.
ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Une attestation de fin de formation est fournie si l'apprenant a bien suivi la totalité de la formation.

MODALITÉS ET DÉLAIS D'ACCÈS
L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES
Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Nettoyer les données.

Analyser les données à l'aide d'outils et méthodes de visualisation et de statistiques.

Concevoir des modèles prédictifs d'intelligence artificielle.

PÉDAGOGIE ET PRATIQUES

Lors de l'inscription, l'apprenant est affecté à une promotion (dates à définir lors de l'inscription) et reçoit son calendrier de formation.

Le parcours de formation est découpé en « Sprint » de plusieurs semaines sur une thématique dédiée. Chaque semaine l'apprenant est convié à un temps d'échange avec le formateur qui se présente sous la forme de masterclass (classe collective) ou de points de mentorat (individuel).

Pendant 80% du temps, l'apprenant travaille en autonomie sur la plateforme d'enseignement. Tous les modules intègrent des exercices pratiques permettant de mettre en œuvre les concepts développés en cours.

L'apprenant doit également travailler en binôme ou trinôme sur un projet fil rouge tout au long de la formation. Cela lui permettra de développer et faire reconnaître ses compétences.
En complément, des événements et ateliers thématiques sont régulièrement proposés pour permettre aux apprenants de découvrir les dernières innovations en matière de Data Science.
Afin de suivre efficacement la formation, nous estimons le temps travail nécessaire entre 8 et 10 heures par semaine.

ACTIVITÉS DIGITALES

Cours et exercices en ligne, masterclass collective, séances de questions/réponses, classes de soutien, accompagnement par mail, projet fil rouge, coaching carrière individualisé, social learning.

CERTIFICATION

Pour clôturer la formation, l'équipe pédagogique évaluera le projet fil rouge de l'apprenant à l'aide d'un rapport écrit et d'une soutenance à distance.
La validation des compétences développées au cours de la formation Data Scientist vous permettra d'obtenir :

- Un certificat de l'École des Mine Paris - PSL Executive Education

- Le bloc de compétence 2 de la certification RNCP de niveau 7 "Expert en ingénierie de l'intelligence artificielle" enregistrée au RNCP sous le n°RNCP38587.

LE PROGRAMME

dernière mise à jour : 04/2025

1) Prochaines dates de sessions

- Mai 2025 : Début au 06/05/25
- Juin 2025 : Début au 03/06/25
- Juillet 2025 : Début au 01/07/25
- Août 2025 : Début au 05/08/25
- Septembre 2025 : Début au 09/09/25
- Octobre 2025 : Début au 07/10/25
- Novembre 2025 : Début au 04/11/25
- Décembre 2025 : Début au 02/12/25

2) Fondamentaux de Python

- Programmation : types de variables, listes, tuples, boucles, fonctions, classes et modules.
- NumPy : NumPy Array, opérations matricielles, indicateur statistique et opérations.
- Pandas : introduction à la bibliothèque, chargement d'un jeu de données, Data Cleaning, Data Processing.

3) Data Visualisation

- Matplotlib : types de graphes, personnalisation de graphiques.
- Plotly : graphiques interactifs et intégrables web, Widgets.
- Seaborn : analyse de distribution, analyses statistiques, analyse multivariée.

4) Outils de programmation

- Git : gestion de version Git, dépôt Git, branches, tag et merge.
- GitHub : plateforme Github, Fork, Pull Request, Issues, partage des modifications, Github Actions.
- Tests unitaires : Pytest, tests d'intégration et fonctions, avantages des tests.
- Système Linux et Script Bash : présentation des systèmes Linux, prise en main du terminal, scripts Bash.

5) Machine Learning

- Modèles et algorithmes de classification : Scikit-learn, logistic regression, KNN, SVM, bagging et boosting.
- Méthodes de clustering : modèles de classification non supervisée, métrique d'évaluation pour le clustering.
- Méthodes de régression : régression linéaire simple et multiple, linéaire régularisée.

6) Machine Learning Avancé

- Séries temporelles avec Statsmodels : modèles ARIMA et SARIMA.
- Détection d’anomalies : problématiques de détection d'anomalies, modèles supervisés, Robust Covariance, DBScan.
- Méthode de réduction de dimensions : feature selection process, analyses de composantes, approche Manifold Learning.

7) Machine Learning Appliqué

- Text mining : expressions régulières, données textuelles, WordCloud, analyse de sentiments.
- Webscraping : langage web, BeautifulSoup, scraping sur Google.
- Éthique & Interprétabilité : données sensibles, enjeux éthiques, analyse de biais, interprétabilité des modèles.

8) Modèles complexes

- Systèmes de recommandations : introduction au système, méthodes content-based, mathématiques de recommandation.
- Reinforcement Learning : méthode de Monte-Carlo, temporal Difference Learning, SARSA et Q-Learning.
- Théorie des graphes : algorithmes Krustal et Dijkstra, détection de communautés, algorithme PageRank, NetworkX.

9) Fondamentaux Deep Learning

- Réseaux denses : réseaux de neurones, framework Keras, optimisation d'un réseau dense.
- Réseaux de convolution : analyse d'image par convolution, LeNet, transfer learning.
- TensorFlow : optimisation avec TensorFlow, callbacks et modèles personnalisés, TensorBoard.

10) Data Engineering

- SQL : requête base de données avec SQL, types de jointures, requêtes imbriquées.
- PySpark : calcul distribué avec PySpark, API RDD et DataFrame de Spark, pipeline de processing, MLLib.
- API : Introduction aux API, méthodes HTTP et fonctions.

11) MLOps

- MLFow : cycle de vie d'un projet Machine Learning, Model training, MLflow.
- Docker : présentation de la conteneurisation, fonctionnement de Docker, images et conteneurs, Dockerfile et Docker Hub.

NOS POINTS FORTS

- Séquences de courte durée
- Activités digitales variées
- Accès illimité pendant 1 an ou pendant la durée du parcours
- Tutorat personnalisé inclus ou en option
- Accès multi-device (smartphone, tablette ou ordinateur)