

# Machine learning engineer, Bootcamp (5 mois) (Titre RNCP)

by DataScientest

*Cours pratique - 105j - 735h00 - Réf. 3BZ*

Devenez expert en modèles prédictifs afin d'anticiper l'évolution des données et les tendances futures. Le Machine Learning Engineer développe des systèmes d'intelligence artificielle (IA) pour rechercher, développer et générer des algorithmes capables d'apprendre et de prédire. 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 « Machine Learning Engineer » des Mines Paris- PSL Executive ainsi que la certification de niveau 7 « Expert en ingénierie de l'intelligence artificielle » RNCP38587. Contactez-nous dès maintenant pour connaître les prochaines dates !

## Objectifs pédagogiques

À l'issue de la formation, le participant sera en mesure de :

- ✓ Mesurer l'apport de l'intelligence artificielle dans la stratégie du système d'information de l'entreprise.
- ✓ Élaborer et mettre en production des modèles et algorithmes d'analyse, de gestion et de traitement de la donnée.
- ✓ Concevoir et piloter une infrastructure d'acquisition, de stockage, de traitement et de restitution de données.
- ✓ Piloter un projet d'intelligence artificielle.

## Public concerné

Toutes les personnes ayant une appétence pour la data souhaitant se reconverter ou faire évoluer ses compétences.

## 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.

## PARTICIPANTS

Toutes les personnes ayant une appétence pour la data souhaitant se reconverter ou faire évoluer ses compétences.

## 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 animent la formation sont des spécialistes des matières abordées. 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 qu'ils enseignent. 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

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

## 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. En complément la réalisation d'une période en entreprise de 130 jours minimum est obligatoire pour valider le titre. La validation des compétences développées au cours de la formation Machine Learning Engineer vous permettra d'obtenir : • Un certificat de l'École des Mines Paris - PSL Executive Education • Le titre RNCP de niveau 7 "Expert en ingénierie de l'intelligence artificielle", enregistré au RNCP en date du 26-01-2022 sous le n°RNCP38587

## Méthodes et moyens pédagogiques

### 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.

### Tutorat

Un formateur expert accompagne l'apprenant tout au long de sa formation. Il échange régulièrement avec lui sur son projet fil rouge et l'accompagne lors de points de mentorat (individuel). Plusieurs formateurs animent également les différentes masterclass (classes collectives) et répondent aux questions des apprenants à tout moment depuis un forum dédié. En complément, de nombreuses séances de questions-réponses peuvent être organisées pour aider les apprenants.

### Pédagogie et pratique

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 35 et 40 heures par semaine.

## MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les formations pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque formation ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le participant a bien assisté à la totalité de la session.

## MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

## ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Pour toute question ou besoin relatif à l'accessibilité, vous pouvez joindre notre équipe PSH par e-mail à l'adresse [psh-accueil@orsys.fr](mailto:psh-accueil@orsys.fr).

## Programme de la formation

### 1 Prochaines dates de sessions

- Octobre 2025 : Début au 07/10/25
- Novembre 2025 : Début au 04/11/25
- Décembre 2025 : Début au 02/12/25

### 2 Acculturation Data

- Acculturation Data : métiers de la Data, gérer une équipe Data, définitions des termes les plus utilisés en Data.
- Différentes sources et types de données : stockage, gouvernance, bases de données relationnelles.
- Pandas : librairie, jeux de données, DataFrames, manipulation de données.
- Fondamentaux de Python : variables et types, fonctions, boucles.

### 3 Data Product Manager

- Gestion de projet : enjeux, modèles de gestion de projet, conduite du changement.
- Méthode Agile : outils et méthodes, application à la Data, impact dans l'entreprise.
- Projet de Data Product Management : product strategy, product discovery, product delivery.

### 4 Chef de projet

- RGPD et éthique : contexte Data, anonymisation et pseudonymisation.
- Suite Data Product Management : product strategy, product discovery, product delivery.

### 5 Fondamentaux Python

- Programmation : variables, listes et Tuples, boucles et types, fonctions, classes et modules.
- NumPy : NumPy Array, opérations matricielles et manipulation d'un NumPy Array, indicateur statistique.
- Data Quality (en option) : principes, variables numériques et textuelles, nettoyage, gestion des valeurs manquantes.

### 6 Data Visualisation

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

### 7 Outils de programmation

- Programmation Linux & Bash : systèmes Linux, terminal, scripts Bash.
- Git & Github : introduction à Git, dépôt, concepts Git, plateforme Github, fonctionnalités majeures, Github actions.
- Tests Unitaires : Pytest, test d'intégration, avantages, test unitaires.
- AWS Cloud Practitioner : outils Cloud avec AWS, préparation à la certification AWS Cloud Practitioner.

### 8 Machine Learning I & II

- Modèles et algorithmes de classification : Scikit-learn, algorithmes classiques, techniques de bagging et boosting.
- Méthodes de clustering : Modèles de classification non-supervisée, métriques d'évaluation pour le clustering.
- Systèmes de recommandations : fonctionnement et enjeux, méthodes content-based et collaborative.
- Méthodes de régression : linéaire simple et multiple, régularisée.

## 9 Machine Learning III

- Séries temporelles avec Statsmodels : modèles ARIMA et SARIMA, signal temporel.
- Détection d'anomalies : problématiques de détection d'anomalies, modèles supervisés, Robust Covariance, DBScan.
- Méthode de réduction de dimensions : Feature selection process, composantes principales, approche Manifold Learning.
- Reinforcement Learning : méthode de Monte- Carlo, temporal Difference Learning, SARSA et Q-Learning.

## 10 Machine Learning IV

- Text mining : expressions régulières, données textuelles, WordCloud, analyse de sentiments.
- Webscraping : langage web (HTML, CSS), BeautifulSoup, scraping sur Google.
- Éthique & Interprétabilité : utilisation de données sensibles, enjeux éthiques de l'IA, SHAP.
- MLflow : l'architecture MLFlow, MLFlow Tracking, MLFlow Projects, MLFlow Models, MLFlow Registry.

## 11 Deep Learning

- Réseaux denses : réseaux de neurones, framework Keras, réseau dense.
- Réseaux de convolution : analyse d'image, architecture LeNet, transfer Learning.
- Théorie des graphes : optimisation avec TensorFlow, callbacks et modèles personnalisés, TensorBoard.

## 12 Data Engineering

- SQL : requêter une base de données, les différents types de jointure, requêtes imbriquées.
- PySpark : calcul distribué avec PySpark, API RDD et DataFrame de Spark, processing de données distribuées, MLLib.
- API : introduction, méthodes HTTP et fonctions.

## 13 Fondamentaux du MLOps

- Docker : conteneurisation et son utilité, fonctionnement de Docker, communication, persistance, dockerfile et hub.
- AWS Solutions Architect : principales composantes du cloud , préparation à la certification AWS Solution Architect.
- Tests Unitaires : tests unitaires avec Pytest, tests d'intégration, avantages, intégration en développement.

## 14 Versioning & Isolation

- DVC & Dagshub : Data Version Control, utilisation de Dagshub, intégration en développement.
- AirFlow : concept, l'orchestration et utilité, graphe orienté acycliques, opérateurs, gestion des tâches, monitoring.
- Jenkins : automatisation, pipeline en CI/CD.

## 15 Deployment & Model Serving

- Bento ML : concepts de création, gestion et déploiement, Docker et BentoML.
- Prometheus & Grafana : utilité, Prometheus Query Language, Dashboard avec Grafana, intégration en production.
- Sécurisation des APIs : sécurité des APIs et des clés API, authentification HTTP, JSON Web Token et HTTPS.

## 16 Scaling

- Data Drift : concept et impact, techniques et outils pour la détection du drift.
- Kubernetes : déployer et gérer des conteneurs, orchestrer plusieurs services.

## 17 Orchestration platform

- ZenML : principes et rôles, suivi et gestion des expériences, intégration avec d'autres outils.
- Weight & Biases : suivi d'expérimentation, de reporting et collaboration, utilisation avec des frameworks.

## Solutions de financement

Plusieurs solutions existent pour financer votre formation et dépendent de votre situation professionnelle.

Découvrez-les sur notre page [Comment financer sa formation](#) ou [contactez votre conseiller formation](#).