

Machine learning engineer, Bootcamp (5 months) (Titre RNCP)

by DataScientest

Practical course - 31d - 217h00 - Ref. 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 !

Teaching objectives

At the end of the training, the participant will be able to:

- ✓ Measure the contribution of artificial intelligence to corporate information system strategy.
- ✓ Develop and implement models and algorithms for analyzing, managing and processing data.
- ✓ Design and manage a data acquisition, storage, processing and retrieval infrastructure.
- ✓ Managing an artificial intelligence project.

Intended audience

Anyone with an appetite for data who wants to retrain or upgrade their skills.

Prerequisites

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

Anyone with an appetite for data who wants to retrain or upgrade their skills.

PREREQUISITES

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.

TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more.

Participants also complete a placement test before and after the course to measure the skills they've developed.

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

Practical details

Digital activities

Online courses and exercises, group masterclasses, question/answer sessions, support classes, e-mail coaching, red thread projects, individualized career coaching, social learning.

Mentoring

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.

Pedagogy and practice

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.

Course schedule

1 Upcoming session dates

- October 2025: Start date 07/10/25
- November 2025: Start date 04/11/25
- December 2025: Start date 02/12/25

TEACHING AIDS AND TECHNICAL RESOURCES

- The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
- At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
- A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

2 Acculturation Data

- Acculturation Data: Data professions, managing a Data team, definitions of the most commonly used Data terms.
- Different data sources and types: storage, governance, relational databases.
- Pandas: library, datasets, DataFrames, data manipulation.
- Python basics: variables and types, functions, loops.

3 Data Product Manager

- Project management: issues, project management models, change management.
- Agile method: tools and methods, application to Data, impact on the company.
- Data Product Management project: product strategy, product discovery, product delivery.

4 Project management

- RGPD and ethics: Data context, anonymization and pseudonymization.
- Suite Data Product Management: product strategy, product discovery, product delivery.

5 Python fundamentals

- Programming: variables, lists and tuples, loops and types, functions, classes and modules.
- NumPy: NumPy Array, matrix operations and NumPy Array manipulation, statistical indicator.
- Data Quality (optional): principles, numerical and textual variables, cleaning, management of missing values.

6 Data Visualization

- Matplotlib: graph types, graph customization.
- Plotly: interactive graphics that can be integrated into Web pages, Widgets.
- Seaborn: distribution analysis, statistical analysis, multivariate analysis.

7 Programming tools

- Linux & Bash programming: Linux systems, terminal, Bash scripts.
- Git & Github: introduction to Git, repository, Git concepts, Github platform, major features, Github actions.
- Unit testing: Pytest, integration testing, benefits, unit testing.
- AWS Cloud Practitioner: Cloud tools with AWS, preparation for AWS Cloud Practitioner certification.

8 Machine Learning I & II

- Classification models and algorithms: Scikit-learn, classical algorithms, bagging and boosting techniques.
- Clustering methods: unsupervised classification models, evaluation metrics for clustering.
- Recommender systems: operation and challenges, content-based and collaborative methods.
- Regression methods: simple and multiple linear, regularized.

9 Machine Learning III

- Time series with Statsmodels: ARIMA and SARIMA models, time signal.
- Anomaly detection: anomaly detection issues, supervised models, Robust Covariance, DBScan.
- Dimension reduction methods: Feature selection process, principal components, Manifold Learning approach.
- Reinforcement Learning: Monte Carlo method, temporal Difference Learning, SARSA and Q-Learning.

10 Machine Learning IV

- Text mining: regular expressions, text data, WordCloud, sentiment analysis.
- Webscraping: web language (HTML, CSS), BeautifulSoup, Google scraping.
- Ethics & Interpretability: use of sensitive data, AI ethical issues, SHAP.
- MLflow: MLflow architecture, MLflow Tracking, MLflow Projects, MLflow Models, MLflow Registry.

11 Deep Learning

- Dense networks: neural networks, Keras framework, dense network.
- Convolution networks: image analysis, LeNet architecture, transfer learning.
- Graph theory: optimization with TensorFlow, callbacks and custom models, TensorBoard.

12 Data Engineering

- SQL: database queries, different types of joins, nested queries.
- PySpark: distributed computing with PySpark, Spark RDD and DataFrame APIs, distributed data processing, MLLib.
- API: introduction, HTTP methods and functions.

13 MLOps fundamentals

- Docker: containerization and its uses, how Docker works, communication, persistence, dockerfile and hub.
- AWS Solutions Architect: main cloud components, preparation for AWS Solution Architect certification.
- Unit testing: unit testing with Pytest, integration testing, benefits, integration in development.

14 Versioning & Isolation

- DVC & Dagshub: Data Version Control, using Dagshub, integration in development.
- AirFlow: concept, orchestration and utility, acyclic directed graph, operators, task management, monitoring.
- Jenkins: automation, CI/CD pipeline.

15 Deployment & Model Serving

- Bento ML: concepts of creation, management and deployment, Docker and BentoML.
- Prometheus & Grafana: benefits, Prometheus Query Language, Dashboard with Grafana, production integration.
- API security: API and API key security, HTTP authentication, JSON Web Token and HTTPS.

16 Scaling

- Data Drift: concept and impact, techniques and tools for drift detection.
- Kubernetes: deploy and manage containers, orchestrate multiple services.

17 Orchestration platform

- ZenML: principles and roles, tracking and managing experiments, integration with other tools.
- Weight & Biases: monitoring of experimentation, reporting and collaboration, use with frameworks.