Machine learning avec Python (scikit-learn)

Conduire un projet de machine learning de bout en bout avec Python et scikit-learn : préparation des données, choix et réglage des algorithmes, évaluation et interprétation des modèles.

En pratique

DURÉE
21 h (3 jours)

MODALITÉS
Présentiel · Distanciel · Intra-entreprise · Inter-entreprises

LIEU
Orléans (45), dans vos locaux ou en classe virtuelle

EFFECTIF
3 à 8 participants

TARIF
Inter : 1 790 € HT par participant ; intra : sur devis

DÉLAI D'ACCÈS
30 jours ouvrés entre la demande et le début de la formation, sous réserve d'un nombre suffisant de participants

VALIDATION
Attestation de fin de formation

NIVEAU
Développeur

Objectifs pédagogiques

À l'issue de la formation, le stagiaire sera capable de :

- Formuler un problème métier en problème d'apprentissage supervisé ou non supervisé
- Préparer les données : encodage, normalisation, valeurs manquantes, pipelines
- Entraîner et comparer des algorithmes de régression et de classification
- Évaluer un modèle avec les bonnes métriques et éviter le surapprentissage
- Interpréter un modèle et préparer sa mise en production

Public visé

- Data analysts, ingénieurs et développeurs qui veulent passer au prédictif
- Data scientists débutants

Prérequis

- Pratique de Python et de pandas
- Bases de statistique

Programme

Jour 1 — Les fondamentaux

Le machine learning, à quoi ça sert ?

- Supervisé, non supervisé : cas d'usage et démarche
- Le cycle d'un projet de bout en bout

Préparer les données

- Exploration, valeurs manquantes, variables catégorielles
- Pipelines scikit-learn et séparation apprentissage-test

Premiers modèles

- Régression linéaire et logistique
- Plus proches voisins, arbres de décision

Jour 2 — Des modèles performants

Méthodes d'ensemble

- Forêts aléatoires, gradient boosting (XGBoost, LightGBM)

Évaluer et régler

- Validation croisée, recherche d'hyperparamètres
- Métriques : RMSE, précision, rappel, F1, AUC ; classes déséquilibrées

Apprentissage non supervisé

- Clustering (k-moyennes, DBSCAN), réduction de dimension (ACP)
- Détection d'anomalies

Jour 3 — Interpréter et industrialiser

Interpréter

- Importance des variables, valeurs SHAP
- Biais et équité des modèles

Vers la production

- Sauvegarder un modèle, l'exposer par une API
- Suivi des performances et réentraînement : introduction au MLOps
- Projet final

Méthodes pédagogiques

- Possibilité de travailler sur les données des participants (en intra, après accord)
- Supports de cours, codes source et corrigés remis aux participants
- Formation dispensée par un consultant data et IA qui conçoit et met en production ce qu'il enseigne

Modalités d'évaluation

- Questionnaire de positionnement en début de formation
- Exercices pratiques corrigés tout au long de la formation
- Projet final : modèle prédictif complet sur un jeu de données inédit
- Questionnaire de satisfaction à chaud

Accessibilité

Formation accessible aux personnes en situation de handicap. Contactez-nous en amont (contact@dcid-consulting.fr) afin d'étudier ensemble les adaptations nécessaires.

Financement

DCID Consulting n'est pas certifié Qualiopi à ce jour : cette formation est financée directement par l'employeur et n'est éligible ni au CPF ni à une prise en charge par un OPCO.

Contact et inscription

contact@dcid-consulting.fr · 02 38 25 33 90 ·dcid-consulting.fr/formations/machine-learning-python/