

Construire, entraîner et valider des modèles de Machine Learning performants

Date et durée
Code formation : IA034FR Durée : 2 jours Nombre d'heures : 14 heures
Description
<p>La valeur d'un projet d'intelligence artificielle repose sur la capacité à produire des modèles prédictifs fiables et généralisables. Cette formation de 2 jours vous permet de dépasser la théorie pour acquérir les compétences opérationnelles nécessaires à la construction de modèles de Machine Learning robustes. Vous apprendrez à naviguer à travers toutes les étapes critiques, du nettoyage des données brutes jusqu'à l'optimisation fine des hyperparamètres.</p> <p>Le parcours met l'accent sur la méthodologie expérimentale : comment choisir le bon algorithme (régression, classification, clustering), comment préparer efficacement ses données et surtout, comment valider rigoureusement ses résultats pour éviter les pièges du surapprentissage. Vous manipulerez les bibliothèques de référence comme Scikit-learn pour mettre en œuvre ces concepts.</p> <p>À travers une série d'ateliers pratiques sur des jeux de données réels, vous développerez votre intuition pour interpréter les métriques de performance et prendre les bonnes décisions de modélisation. Vous repartirez avec une boîte à outils complète pour concevoir des solutions d'IA performantes et auditables.</p>
Objectifs
<p>À l'issue de cette formation, vous atteindrez les objectifs de compétences suivants :</p> <ul style="list-style-type: none">• décrire les étapes du cycle de vie d'un modèle de machine learning, de la conception au déploiement ;• sélectionner et paramétriser les algorithmes les plus adaptés à une problématique métier donnée ;• appliquer les techniques de validation et de mesure de performance pour qualifier un modèle ;• détecter et corriger les risques de surapprentissage (overfitting) et les biais potentiels ;• optimiser les hyperparamètres pour maximiser la robustesse et la capacité de généralisation des modèles.
Points forts
<ul style="list-style-type: none">• Approche complète : vous maîtriserez l'intégralité de la chaîne de valeur, de la donnée brute jusqu'au modèle optimisé et validé.• Focus qualité : vous apprendrez à déjouer les pièges classiques comme le surapprentissage pour garantir des modèles fiables en production.• Pratique intensive : vous ancrerez vos acquis grâce à 6 ateliers techniques couvrant la préparation, l'entraînement et l'optimisation.• Boîte à outils : vous repartirez avec la maîtrise des bibliothèques standards (Scikit-learn, Pandas) et des méthodes d'évaluation professionnelles.
Modalités d'évaluation

Travaux Pratiques

Pré-requis

Suivre cette formation nécessite les prérequis suivants :

- **Expérience professionnelle :** une première expérience en manipulation de données.
- **Connaissances de base :**
 - une maîtrise des bases du langage Python et de ses bibliothèques de manipulation de données ;
 - des notions fondamentales en statistiques et en machine learning.

Public

Cette formation s'adresse aux profils techniques souhaitant se spécialiser. Le public inclut notamment :

- les **data analysts et data scientists débutants** cherchant à structurer leur approche ;
- les **développeurs** souhaitant acquérir des compétences spécifiques en Machine Learning ;
- les **chefs de projet IA et product owners techniques** désirant comprendre la mécanique interne des modèles ;
- **Toute personne impliquée** dans la conception de modèles prédictifs.

Programme

Module 1 : s'approprier la démarche de modélisation prédictive

- La définition des objectifs et la typologie des problèmes (classification, régression, clustering).
- Le cycle de vie complet d'un modèle de Machine Learning.

Travaux pratiques

- Identifier le type de problème ML à traiter à partir de différents cas d'usage métiers.

Module 2 : préparer et structurer les données

- Le nettoyage des données et le traitement des valeurs manquantes.
- L'encodage des variables catégorielles pour les rendre exploitables par les algorithmes.
- Les techniques de normalisation et de standardisation des features.

Travaux pratiques

- Préparer et nettoyer un jeu de données brut en utilisant Pandas et Scikit-learn.

Module 3 : entraîner les modèles

- Le panorama des algorithmes classiques : régression linéaire, arbres de décision, SVM, KNN.
- Les stratégies de paramétrage et la séparation des données (train/test/validation).

Travaux pratiques

- Entraîner plusieurs modèles concurrents sur un même jeu de données réel.

Module 4 : évaluer les performances

- L'analyse des métriques de classification (accuracy, precision, recall, F1-score).

- L'analyse des métriques de régression (RMSE, MAE, R²).
- L'utilisation de la matrice de confusion et des courbes ROC/AUC.

Travaux pratiques

- Réaliser une évaluation comparative de plusieurs modèles et interpréter les résultats.

Module 5 : valider et fiabiliser les modèles

- La mise en œuvre de la validation croisée (Cross-validation, K-fold).
- La détection et le traitement du surapprentissage (overfitting) et du sous-apprentissage.
- L'application de techniques de régularisation pour améliorer la généralisation.

Travaux pratiques

- Mettre en œuvre une procédure de validation croisée et analyser les écarts de performance.

Module 6 : optimiser et sélectionner la solution finale

- L'automatisation de la recherche d'hyperparamètres (Grid Search, Random Search).
- La sélection des features pertinentes et l'analyse de l'importance des variables.
- Les outils d'interprétabilité des modèles (SHAP, LIME).

Travaux pratiques

- Optimiser un modèle complexe avec GridSearchCV et interpréter les variables clés.

Les noms de marques et logos éventuellement cités dans cette fiche (ex. Python, Scikit-learn, Pandas) sont la propriété de leurs détenteurs respectifs. Leur mention à des fins pédagogiques ne constitue ni un engagement ni un partenariat.