

Mis à jour le 07/08/2025

S'inscrire

Formation XGBoost

2 jours (14 heures)

Présentation

XGBoost (Extreme Gradient Boosting) est une bibliothèque de machine learning optimisée pour la performance, utilisée pour les tâches de classification et de régression. Elle repose sur le gradient boosting, intégrant des techniques avancées de régularisation, gestion des valeurs manquantes et optimisation parallèle.

Notre formation XGBoost vous permettra de maîtriser la construction de modèles prédictifs robustes et interprétables. Vous découvrirez comment entraîner, optimiser, expliquer et déployer des modèles performants, en utilisant des cas concrets issus du monde réel.

Vous apprendrez à préparer les données, optimiser les hyperparamètres, visualiser les décisions et mettre en production vos modèles avec auditabilité et reproductibilité.

À l'issue de la formation, vous serez en mesure de construire des modèles XGBoost de A à Z, de l'analyse exploratoire à leur intégration dans des systèmes métiers.

Comme toutes nos formations, celle-ci s'appuie sur la dernière version stable de [XGBoost v3.0.3](#) et privilégie une approche résolument pratique et opérationnelle.

Objectifs

- Comprendre le fonctionnement du gradient boosting et d'XGBoost
- Construire des modèles de classification et de régression
- Optimiser les performances avec des techniques de tuning avancé
- Interpréter et documenter les résultats modèles
- Déployer un modèle XGBoost et assurer sa traçabilité

Public visé

- Data Scientists
- Data Analysts
- Ingénieurs Machine Learning
- Statisticiens

Pré-requis

- Maîtrise des bases Python, Pandas et scikit-learn
- Connaissances fondamentales en machine learning supervisé
- Notions en statistiques et analyse de données tabulaires

Notre formation XGBoost

Comprendre les fondements du boosting et d'XGBoost

- Introduction au machine learning supervisé et aux modèles d'ensemble
- Principe du boosting et de la combinaison de modèles faibles
- Comparaison entre bagging, random forest et boosting
- Architecture et spécificités de XGBoost (régularisation, split finding, etc.)
- Configuration de l'environnement Python pour XGBoost
- Atelier : Implémenter un modèle de boosting simple avec scikit-learn et XGBoost

Préparation des données et optimisation du pipeline

- Gestion des données tabulaires et encodage des variables
- Traitement des valeurs manquantes et outliers
- Utilisation des matrices DMatrix pour un entraînement optimisé
- Séparation train/test et validation croisée
- Bonnes pratiques de preprocessing pour le boosting
- Atelier : Construction d'un pipeline de préparation des données avec Pandas et XGBoost

Entraîner un modèle de classification avec XGBoost

- Définition des hyperparamètres principaux
- Métriques d'évaluation : accuracy, logloss, AUC
- Importance des variables et analyse des splits
- Early stopping, régularisation L1/L2 et gestion du surapprentissage
- Visualisation des arbres et interprétabilité locale
- Atelier : Modéliser un problème de churn client ou crédit scoring avec XGBoost

Optimisation des performances et tuning avancé

- Utilisation de GridSearchCV, RandomizedSearchCV et Optuna
- Analyse des courbes de validation et choix des meilleurs paramètres
- Gestion de l'équilibre classe dans des datasets déséquilibrés
- Stratégies d'ensemble avec XGBoost (bagging + boosting)
- Bonnes pratiques de reproductibilité et de robustesse
- Atelier : Tuning de modèle XGBoost pour un dataset public

Régression et cas d'usage métiers de XGBoost

- Adapter XGBoost à des problèmes de régression
- Études de cas : prévision de prix, estimation de vente, séries temporelles simples
- Comparaison avec d'autres modèles (linear, RF, LGBM)
- Choix des objectifs et gestion des métriques spécifiques (RMSE, MAE)
- Interpréter les résultats métiers et reporting
- Atelier : Prévision de prix immobiliers ou volume de ventes avec XGBoost

Déploiement, audit et mise en production

- Sauvegarde et chargement de modèles avec joblib ou pickle
- Déploiement via API REST (Flask ou FastAPI)
- Explicabilité : SHAP, LIME, importance globale vs locale
- Suivi des performances en production
- Contrôles qualité, documentation, auditabilité du modèle
- Atelier : Mise en production d'un modèle XGBoost avec interprétabilité intégrée

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des

séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.