

Fiche du cours

Titre :

ML400 - Machine Learning Avancé (Ensembles, Time Series & MLOps)

Description :

Approfondir le Machine Learning appliqué au niveau production : feature engineering avancé, tuning par optimisation bayésienne, ensembles (XGBoost/LightGBM/CatBoost, stacking/blending), calibration & seuils, données déséquilibrées, séries temporelles (CV temporelle, features retardées), interprétabilité (PDP/ICE, SHAP), anomalies (IsolationForest) et MLOps : suivi d'expériences (MLflow), validation de données, serving (API), monitoring & drift.

Objectifs :

- Concevoir des pipelines robustes (prétraitements + modèle) en évitant la data leakage.*
- Maîtriser hyperparameter tuning (recherche bayésienne), cross-validation avancée (stratifiée, time-based).*
- Traiter le déséquilibre (class weights, SMOTE/undersampling), calibrer les probabilités et choisir les seuils selon le coût.*
- Exploiter des ensembles modernes (GBDT : XGBoost/LightGBM/CatBoost) et des méta-modèles (stacking/blending).*
- Gérer séries temporelles : features glissantes, backtesting, CV bloquée, fuites temporelles.*
- Interpréter & auditer : importances, PDP/ICE, SHAP, biais/équité (aperçu).*
- Passer en MLOps : MLflow (expériences/modèles), Great Expectations (qualité des données), API (FastAPI – concept), monitoring & drift (concepts), packaging & reproductibilité.

Chapitres :

1. Pipelines & anti-leakage : ColumnTransformer, ordre des étapes, CV avancée*
2. Tuning avancé : recherche bayésienne, espaces d'hyperparamètres, early stopping*
3. Déséquilibre & coût : métriques (PR/F1), resampling (SMOTE), coûts & seuils*
4. Calibration : Platt/Isotonic, fiabilité, décision au seuil*
5. Ensembles GBDT I : XGBoost/LightGBM/CatBoost – réglages clés & régularisation*
6. Ensembles GBDT II : stacking/blending, bagging vs boosting, leakage en stacking*
7. Time Series I : features (lags, rolling stats), split temporel, backtesting*
8. Time Series II : covariables, drift saisonnier, évaluation réaliste*
9. Unsupervised/Anomalies : IsolationForest, One-Class SVM, seuils & alertes*
10. Interprétabilité : PDP/ICE, SHAP local/global, limites & pièges*

11. MLOps I : MLflow (traçage runs, modèles), Great Expectations (tests de données)*

12. MLOps II : serving (API, schémas d'entrées), versionnage, monitoring & drift (concepts), packaging & doc

À la fin :

À l'issue du cours, vous saurez mettre à l'échelle vos workflows ML : tuner efficacement, stabiliser les performances (CV réaliste, séries temporelles), expliquer les décisions (SHAP), gérer déséquilibre & calibration, et préparer la production (suivi d'expériences, validation des données, plan de serving/monitoring) — un livrable portfolio-ready.

Vous automatiserez le suivi et la comparaison des modèles (MLflow) pour garantir la reproductibilité.

Vous produirez une documentation et un runbook opérationnel (métriques, seuils d'alerte, gestion du drift) facilitant la maintenance et l'amélioration continue.