

XGBoost training

2 days (14 hours)

Presentation

XGBoost (Extreme Gradient Boosting) is a performance-optimized machine learning library used for classification and regression tasks. It is based on gradient boosting, integrating advanced regularization techniques, missing value management and parallel optimization.

Our XGBoost training course will enable you to master the construction of robust and interpretable predictive models. You'll discover how to train, optimize, explain and deploy high-performance models, using real-world case studies.

You'll learn how to prepare data, optimize hyperparameters, visualize decisions and put your models into production with auditability and reproducibility.

At the end of the course, you'll be able to build XGBoost models from A to Z, from exploratory analysis to integration into business systems.

Like all our training courses, this one is based on the latest stable version of [XGBoost v3.0.3](#) and takes a resolutely practical and operational approach.

Objectives

- Understand how gradient boosting and XGBoost work
- Build classification and regression models
- Optimize performance with advanced tuning techniques
- Interpret and document model results
- Deploy an XGBoost model and ensure its traceability

Target audience

- Data Scientists
- Data Analysts
- Machine Learning Engineers
- Statisticians

Prerequisites

- Basic knowledge of Python, Pandas and scikit-learn
- Fundamental knowledge of supervised machine learning
- Notions in statistics and tabular data analysis

Our XGBoost training course

Understanding the basics of boosting and XGBoost

- Introduction to supervised machine learning and ensemble models
- Principle of boosting and combining weak models
- Comparison between bagging, random forest and boosting
- Architecture and specific features of XGBoost (regularization, split finding, etc.)
- Configuring the Python environment for XGBoost
- Workshop: Implementing a simple boosting model with scikit-learn and XGBoost

Data preparation and pipeline optimization

- Tabular data management and variable encoding
- Handling missing values and outliers
- Use of DMatrix for optimized training
- Train/test separation and cross-validation
- Best preprocessing practices for boosting
- Workshop: Building a data preparation pipeline with Pandas and XGBoost

Training a classification model with XGBoost

- Definition of main hyperparameters
- Evaluation metrics: accuracy, logloss, AUC
- Variable importance and split analysis
- Early stopping, L1/L2 regularization and overlearning management
- Tree visualization and local interpretability
- Workshop: Modeling a customer churn or credit scoring problem with XGBoost

Performance optimization and advanced tuning

- Using GridSearchCV, RandomizedSearchCV and Optuna
- Analysis of validation curves and choice of best parameters
- Class balance management in unbalanced datasets
- Ensemble strategies with XGBoost (bagging + boosting)
- Best practices for reproducibility and robustness
- Workshop: Tuning an XGBoost model for a public dataset

Regression and business use cases for XGBoost

- Adapt XGBoost to regression problems
- Case studies: price forecasting, sales estimation, simple time series
- Comparison with other models (linear, RF, LGBM)
- Choice of objectives and management of specific metrics (RMSE, MAE)
- Interpreting business results and reporting
- Workshop: Forecasting real estate prices or sales volumes with XGBoost

Deployment, audit and go-live

- Saving and loading models with joblib or pickle
- Deployment via REST API (Flask or FastAPI)
- Explicability: SHAP, LIME, global vs. local importance
- Production performance monitoring
- Quality control, documentation, model auditability
- Workshop: Putting an XGBoost model into production with integrated interpretability

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced IT technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the forthcoming course, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical training: 60% hands-on, 40% theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples and

and group work sessions.

Validation

At the end of the session, a multiple-choice questionnaire is used to check that skills have been correctly acquired.

Certification

A certificate will be issued to each trainee who has completed the entire training course.