

RapidMiner Mining Core Training

3 days (21 hours)

Presentation

Master visual data mining with RapidMiner thanks to this comprehensive, operational training course. Designed for data professionals, it will enable you to get to grips with the full range of RapidMiner Mining Core functionalities.

You'll start by exploring the RapidMiner Studio interface and the key steps in the CRISP-DM cycle. The focus is on importing, cleansing and transforming data to build reliable, actionable datasets.

You will then learn how to develop classification, regression or clustering models, evaluate them in detail, and automate your workflows with Auto Model and parameter optimization. Case studies reinforce each skill.

Finally, the training covers exporting results, integration into a business environment and best practices in visual analysis, to enable you to deploy high-performance, sustainable and replicable data projects.

As with all our training courses, this one will be presented with the latest [RapidMiner](#) updates.

Objectives

- Understand the RapidMiner ecosystem and the CRISP-DM lifecycle
- Configure the RapidMiner Studio environment and import multi-source data
- Prepare, transform and enrich datasets for analysis
- Design, train and evaluate predictive and unsupervised models
- Automate analytical workflows with Auto Model and parameter optimization
- Integrate RapidMiner into a reliable, industrialized analytical production process

Target audience

- Data analysts
- Data scientists

Prerequisites

- Mastery of SQL

RapidMiner Mining Core training program

Introduction to RapidMiner and data mining

- Positioning in the data science ecosystem
- Definition of data mining
- Lifecycle of a data mining project
- Overview of panels: repository, design, parameters
- Starting a new process
- Understanding the organization of a workflow

Importing and preparing data

- Loading files
- Database connections and Web Services
- RapidMiner repository: dataset management
- Data cleansing and transformation
- Filling in or deleting missing values
- Normalization, discretization, encoding
- Table filtering, sampling, sorting, merging and joining
- Replace Missing Values, Normalize, Filter Examples
- Generate Attributes to create derived variables

Exploratory analysis and visualization

- Mean, median, standard deviation, distribution
- Visualizing outliers
- Histograms, boxplots, scatter plots
- Heatmaps and correlation matrices
- Interactive graphics in RapidMiner

Variable selection and reduction

- Integrated methods: Weight by Information Gain, Weight by Gini Index
- Elimination of redundant or unnecessary variables
- PCA (Principal Component Analysis)

- t-SNE for high-dimensional visualization

Supervised predictive modeling

- Algorithms: Decision Tree, Naive Bayes, SVM, k-NN, Random Forest
- Use of Train, Apply Model, Cross Validation operators
- Performance analysis: accuracy, recall, precision, F1-score
- Linear regression, Ridge/Lasso regression
- Regression models in RapidMiner

Clustering and unsupervised modeling

- k-Means, k-Medoids, DBSCAN
- Distance and similarity analysis
- Silhouette Index, Davies-Bouldin Index
- Cluster visualization

Model evaluation

- Cross-validation, split training/test
- Confusion matrix
- ROC, AUC, Lift and Gains curves
- Overlearning analysis

Automation and optimization

- Auto Model in RapidMiner
- Using Auto Model for fast cases
- Customizing algorithm selection and evaluation
- Hyperparameter optimization
- Grid Search, Evolutionary Parameter Optimization
- Parameter impact measurement

Integration and export

- CSV, Excel, database export
- Report and dashboard generation
- Integration with other systems
- RapidMiner Server
- Calls via API or Web Services

Companies concerned

This training 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 with regard to the training to come, 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, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Certification

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