

RapidMiner & Data Science: Advanced

Course Overview

RapidMiner & Data Science: Advanced is a two-day course focusing on data mining and predictive analytics with RapidMiner Studio. Over the course of two days, students will expand their knowledge gained in *RapidMiner & Data Science: Foundations* and build more sophisticated analytical models while utilizing more complex functions and features of RapidMiner Studio.

The course is structured in a mentoring fashion where the entire group performs tasks alongside the instructor as members of a data science team. After successfully completing this course, participants will have a solid understanding of how RapidMiner Studio functions and will be able to use it on a wide variety of real-world problems. They will understand automation techniques inside RapidMiner and they will also be introduced to more advanced concepts in modern data science such as feature selection, sampling, and weighting, and several new modern machine learning algorithms for predictive modeling.

Practical exercises during the course prepare students to take the knowledge gained and apply it to their own complex data challenges. The class exercises and labs are hands-on, so students will internalize the topics covered, which will provide a jumpstart to the real world application of these techniques.

Prerequisites & Target Audience

This class is aimed at Analysts and Data Scientists. It assumes a basic knowledge of computer programming principles and higher mathematics (through calculus). It also assumes successful completion of the *RapidMiner & Data Science: Foundations* course, or a similar level of expertise in RapidMiner Studio and basic knowledge of applied statistics or data science.

Course Objectives

After the training, students will have the ability to use RapidMiner to:

- Perform more advanced data preparations and transformations for data mining using loops & macros
- Implement feature generation and feature engineering to generate new predictive attributes using transformations, aggregations, and regular expressions
- Build sophisticated analytical predictive models using a wider variety of machine learning algorithms
- Evaluate model quality with respect to new performance criteria
- Deploy analytical predictive models
- Use sampling and weighting to address imbalanced samples
- Utilize feature selection and variable reduction techniques

Course Outline

- Overview
 - Business Case
 - Foundations Course Review
- EDA: Exploratory Data Analysis
 - Advanced Data ETL (Extract, Transform, and Load)
 - Loops
 - Macros
 - Joins & Set Theory
 - Appending New Attributes & Examples
- Feature Generation and Feature Engineering
 - Aggregation & Multi-Level Aggregation
 - Pivot & De-Pivot
 - Calculated Values
 - Date Transformations & Calculations
 - Regular Expressions
 - Changing & Formatting Value Types
- Model Construction and Evaluation
 - Advanced Performance Criteria
 - ROC Plots & Area Under the Curve
 - Comparisons between Models
 - Sampling
 - Weighting
 - Validation of Preprocessing and Preprocessing Models
 - Optimization & Logging Results
- Feature Selection
 - Attribute Correlation
 - Forward Selection
 - Backward Elimination
- New Predictive Modeling Algorithms
 - Neural Networks
 - Support Vector Machines