

RapidMiner & DataScience: Advanced

Course Overview

RapidMiner & DataScience: 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 & DataScience: Foundations* and build a more sophisticated analytical model while reinforcing their familiarity with the graphical interface and all the major product features and functionality.

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. Participants will be able to prepare data, create and validate predictive models, deploy models, and will be ready to extend their knowledge to advanced topics such as *RapidMiner Server: Deployment and Web Apps*, *Big Data Analytics with RapidMiner Radoop*, and *Text & Web Mining with RapidMiner*.

Practical exercises during the course prepare students to take the knowledge gained and apply to their own respective data mining problems, solving them quickly and easily. Since the class labs are hands-on and performed on the participants' personal laptops, students will take actual classwork home with them, which will provide a jumpstart to the real world.

Target Audience

Advanced Analysts and Data Scientists

Prerequisites

Basic knowledge of computer programs and mathematics

RapidMiner & DataScience: Foundations

Course Objectives

After the training, students will have the ability to:

- Perform all common data preparations
- Build sophisticated analytical predictive models
- Evaluate model quality with respect to different criteria
- Deploy analytical predictive models

- Utilize more complex functionality of RapidMiner Studio
- Apply more sophisticated analytical approaches

Course Outline

- **Overview**

- ◇ Business Case
- ◇ Intro Course Review
- ◇ Loading New Data

- **EDA: Exploratory Data Analysis**

- ◇ Multiple Sources
- ◇ Joins & Set Theory
- ◇ Understanding New Attributes

- **Data Preparation**

- ◇ Advanced Data ETL (Extract, Transform, and Load)
- ◇ Aggregation & Multi-Level Aggregation
- ◇ Pivot & De-Pivot
- ◇ Calculated Values
- ◇ Regular Expressions
- ◇ Changing Value Types
- ◇ Feature Generation and Feature Engineering
- ◇ Loops
- ◇ Macros

- **Predictive Modeling Algorithms**

- ◇ Support Vector Machines
- ◇ Neural Networks
- ◇ Logistic Regression

- **Model Construction and Evaluation**

- ◇ Advanced Performance Criteria
- ◇ ROC Plots
- ◇ Comparison between Models
- ◇ Sampling
- ◇ Weighting
- ◇ Feature Selection: Forward Selection
- ◇ Feature Selection: Backward Elimination
- ◇ Validation of Preprocessing and Preprocessing Models
- ◇ Optimization & Logging Results

- **Additional Workshops**

- ◇ Principal Components Analysis
- ◇ Logistic Regression
- ◇ Performance (Cost) Model Optimization