Course Overview
RapidMiner Basics Part 1 is a two day course focusing on data mining and predictive analytics with RapidMiner Studio. Over the course of two days students will explore a clean, simplified business use case and build a strong analytical model while becoming familiar with the graphical interface and all of the products features and functionality.

The course is structured in a mentoring fashion where the entire group performs 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, and will be ready to extend their knowledge to advanced topics such as RapidMiner Basics Part 2 and RapidMiner Server: Web Apps and Deployment.

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
 Analysts, Developers, and Administrators

Prerequisites
Basic knowledge of computer programs and mathematics

Course Objectives
After the training, students will have the ability to:

- Perform all common data preparations
- Build strong analytical predictive models
- Evaluate model quality with respect to different criteria
- Share analytical models and collaborate with team members
Course Outline

• Overview
  ◇ Business scenario
  ◇ Analytics
  ◇ Data mining in the Enterprise
  ◇ CRISP-DM

• Getting Started with RapidMiner Studio
  ◇ User Interface
  ◇ Creating and Managing RapidMiner Repositories
  ◇ Starting a new RapidMiner Project
  ◇ Operators and Processes
  ◇ Loading Data
  ◇ Storing data, processes, and Result Sets

• EDA: Exploratory Data Analysis
  ◇ Data Types
  ◇ Data Hierarchy
  ◇ Quick Summary Statistics
  ◇ Visualizing Data
  ◇ Charting

• Data Preparation
  ◇ Normalization and Standardization
  ◇ Basic Transformations of Value Types
  ◇ Handling Missing Values
  ◇ Sampling
  ◇ Filtering Examples and Attributes
  ◇ Handling Attribute Roles

• Building Better Processes
  ◇ Organizing
  ◇ Renaming
  ◇ Relative Path
  ◇ Flow Control
  ◇ Subprocesses
  ◇ Building Blocks
  ◇ Breakpoints

• Predictive Models
  ◇ Correlations
  ◇ k-Nearest Neighbor
  ◇ Naïve Bayes
  ◇ Linear Regression
  ◇ Rules
  ◇ Decision Trees
  ◇ Importance of Attributes
• Model Evaluation
  ◊ Applying Models
  ◊ Overfitting
  ◊ Splitting Data
  ◊ Evaluation Methods
  ◊ Performance Criteria

• Sharing and Collaboration
  ◊ Exporting Images
  ◊ RapidMiner Server