

# RapidMiner Training



SIMPLIFY  
COLLABORATE

PREDICT TO ACT

## Why should you attend this training

The RapidMiner Basics Pt. 1 & Pt. 2 classes are required for one to become a certified RapidMiner Analyst.

This training is a four-days course, split into two parts of two days each, exploring the possibilities of performing data mining and predictive analytics with RapidMiner Studio and RapidMiner Server. Where Part 1 of the course takes a clean, simplified business example to build a strong foundation, Part 2 explores a similar business case with some of the messiness of the real world added in.

This class resembles a mentor-mentee relationship with the entire group performing as members of a data science team. After successfully completing this course, participants will have a decent understanding of how RapidMiner software works and is used. The participants will be able to prepare data and create predictive models in standard data environments typically found within most analyst positions, as well as in many more uncommon environments, and will be ready to extend their knowledge with advanced topics such as text mining, image processing, and big data.

Practical exercises during class prepare the participants to transfer the knowledge gained and apply it to their own data mining problems, solving them quickly and easily. Since the class labs are hands-on, performed on the students' own laptops, the students will be taking their actual classwork home with them to jumpstart their application to the real world.

## After the training, students will have the ability to:

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

The RapidMiner Basics courses are also the perfect preparation for the Analyst level certification exam.

## About Go Training

Go Training applies effective pedagogical methodologies that demonstrate case studies and hands-on practical skills, in addition to explaining clearly how things work in principle. Every course that we conduct is delivered by a subject matter expert who holds the academic qualification and working experience in that specialization. On the days when they are not teaching, our trainers work on consultancy projects and technical deliveries. Their work has received numerous recognition and awards in the industry. Our team of trainers has been invited as keynote speakers at numerous international conferences, and as principal consultants for various industries.

Date: 22 – 25 September 2015  
(Tuesday - Friday)  
Time: 0900 - 1700  
Venue: MSC Malaysia K-Workers Development  
Centre (KDC)  
Block 3730, Persiaran APEC, Cyber 8,  
63000 Cyberjaya, Selangor.

**HRDF + MSC Malaysia MyProCert Program  
Claimable**

## Course Outline

### RapidMiner Basics Pt. 1

- Overview
  - Business Scenario, Analytics
  - Data Mining in the Enterprise, CRISP-DM
- Basic usage
  - User Interface, Creating and Handling RapidMiner Repositories
  - Starting a New RapidMiner Project, Operators and Processes
  - Loading Data, Storing Data, Processes, and Results
- 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, Naive 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

### RapidMiner Basics Pt. 2

- Overview
  - Business Case Changes, Intro Course Recap, Loading New Data
- EDA
  - Multiple Sources, Understanding New Attributes
  - Schema Relationships
- Data preparation
  - Joins, Aggregation, Multi-level Aggregation
  - Pivot, Set Theory, Calculated Values, Regular Expressions
  - Changing Value Types, Balancing Data, Outlier Detection
  - Feature Selection, Dimensionality Reduction
- Predictive Models (Sample Varies)
  - SVM, Random Forest, K-Means Clustering, Neural Networks
  - Logistic Regression, Meta Learning
- Model evaluation
  - Advanced Performance Criteria, ROC Plots
  - Comparison between Models, Lift Chart
  - Significance Tests, Logging Results
  - Validation of Preprocessing and Preprocessing Models
- Deployment
  - Sharing data, models, and processes
  - Exporting processes as web service
  - Basics of Report Creation
  - Managing Processes and Services

## About the Instructor

Jaden graduated from Wichita State University (USA) in 2010, with a Bachelor's Degree in Aerospace Engineering. His primary design work in college was the design-build-fly of a radio-controlled aircraft. Other notable efforts include the investigation of the aerodynamic behaviors of an automobile and the Global Design Challenge, a collaborative airframe design competition sponsored by Spirit Aerosystems. The close-to-industry project management and collaboration settings in these projects played a part in preparing Jaden well for a jump-start in his career.



Jaden began his career as an Application Engineer in TechSource Systems, the SEA sole distributor of MATLAB. As a consultant, he focused on technical computing areas such as data analytics, artificial intelligence, image processing, application deployment, etc. During this time, Jaden has supported many engineering and research professionals and assisted them in creating more efficient workflows and obtaining better results.

In 2013, Jaden transitioned to the MATLAB technical sales role, taking charge of commercial accounts in Malaysia and Indonesia. He works with decision-making executives and has established a thorough understanding on business requirements that necessitates adoption of new technologies. In 2014, Jaden has co-founded Quandatics, which focuses on developing and providing advanced analytics services and solutions. Jaden is now heading the business development and technical marketing teams in Quandatics and his clientele-base includes MNCs, GLCs, government agencies, and higher-education institutions.

Go Training  
wholly owned by iRadar Sdn Bhd  
HRDF Approved Training Provider (Category A)

No. 36, Jalan IMJ 1, Taman Industri Malim  
Jaya, 75250 Melaka, Malaysia.

t +606 336 6016

f +606 252 3059

w [www.gotraining.com.my](http://www.gotraining.com.my)

[f] [fb.com/gotraining.com.my](https://fb.com/gotraining.com.my)

[in] [linkedin.com/company/gotraining](https://linkedin.com/company/gotraining)

To register, please contact:

m +6010 663 1852

e [yiwei@gotraining.com.my](mailto:yiwei@gotraining.com.my)