Unsupervised Machine Learning with RapidMiner

Course Overview

Unsupervised Machine Learning with RapidMiner is a one-day expert seminar regarding unsupervised machine learning data science techniques. It introduces basic concepts in clustering and the most common clustering methods, such as k-means clustering, DBSCAN, and hierarchical & agglomerative clustering. It also introduces key ideas in anomaly detection and provides instruction in both distance-based and density-based methods for identifying outliers. Finally, it introduces recommender systems and association models and their use.

After successfully completing this course, participants will have a solid understanding of how RapidMiner Studio supports the most common unsupervised machine learning tasks. Participants will be able to successfully execute techniques for clustering different types of data according to a variety of different methods and will understand the most important differences between these methods and how to choose the most suitable one based on the project objectives. They will also be prepared to utilize several different approaches to anomaly and outlier detection. And they will have a basic understanding of recommender systems and how to create them in RapidMiner. Practical exercises during the course prepare students to take the knowledge gained and apply it to their own unsupervised learning 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 requires either the successful completion of the basic-level training courses (RapidMiner & Data Science: Foundations and RapidMiner & Data Science: Advanced) or successful completion of the RapidMiner Analyst Certification exam (or functional equivalence in terms of knowledge of RapidMiner and basic data science).

Course Outline

- Introduction to Unsupervised Machine Learning Problems
- Introduction to Anomaly Detection & HBOS
- Distance-Based and Density-Based Outlier Detection
- k-Means Clustering
- Clustering Performance and Labels
- DBSCAN
- Hierarchical & Agglomerative Clustering
- Recommender Systems and Association Rules Models