

Data Science with KNIME for Experienced Users

Model selection, evaluation and reusable analytical workflows

FA-0727 · 2 days

Data & Analytics

Introduction

Extend existing KNIME workflow skills with selected machine-learning examples and systematic evaluation. Review preparation, model goals, settings and comparison criteria using supplied datasets and compatible nodes.

The two-day workshop combines a small implementation exercise with broader algorithm demonstrations. It is not a beginner KNIME course or a promise to optimise every listed model in depth.

Learning outcomes

- Build on existing workflows to acquire and prepare data for modelling.
- Select an appropriate regression, classification or clustering goal.
- Compare selected learner settings and features using separate evaluation data.
- Interpret model measures and distinguish tuning from independent validation.
- Inspect supported model-exchange options and retain reproducible workflow evidence.

Prerequisites

- Basic education in mathematics (understanding of linear and non-linear mathematics at high-school level)
- Very basic understanding of API and web requests
- Experience and understanding of filing system

- Experience in Using KNIME for analytics

A compatible KNIME Analytics Platform installation with the extensions and supplied workflows needed for the examples.

Remote participants need reliable connectivity and conferencing tools; a second screen is optional and certification attendance rules are not implied.

Training outline

1. Day 1 — Preparation, regression and evaluation

- Review node/workflow execution, automation and small flat-file or approved API data inputs.
- Explore a supplied iterative/recursive workflow and historical data example where useful; preserve time order when it affects evaluation.
- Machine-learning goals, training/evaluation separation and the effect of preparation decisions.
- Implement a selected linear-regression or regression-tree workflow; compare polynomial-regression behaviour through a worked example.
- Inspect parameters and features using appropriate validation; do not use held-out evaluation data to optimise the workflow.

2. Day 2 — Classification, clustering and model exchange

- Compare naïve Bayes, decision trees, k-nearest neighbours, SVM and logistic regression conceptually; practise selected examples rather than five full deep-dive labs.
- Review distance metrics, evaluation criteria and algorithm suitability for the sample data.
- Clustering concepts: k-means and its cluster-count choice, hierarchical approaches and DBSCAN; compare assumptions and selected results.
- Use a bounded parameter-search demonstration; choosing k depends on evidence and purpose, not a guaranteed best grid result.
- Inspect PMML/model interchange where supported by the selected nodes; validate input schema and compatibility on reload or scoring.
- Assess the workflow, document limitations and identify further practical study.