

Analytics and Data Manipulation with KNIME

No-code data preparation and introductory predictive workflows

FA-0718 · 3 days

Data & Analytics

Introduction

Turn a small analytical question into a visual KNIME workflow using sample spreadsheet, CSV and selected external data. Practise defining the problem, preparing data, exploring results and reviewing introductory predictive models.

No coding experience is required. The three-day course prioritises a small end-to-end exercise and selected model demonstrations; it does not imply that no-code tools remove the need for sound data and model evaluation.

Learning outcomes

- Define a business question, relevant data and a suitable analysis workflow.
- Acquire, filter, concatenate, join, aggregate and export sample data.
- Clean values and apply selected type conversions, discretisation, normalisation and pivoting.
- Inspect visualisations and compare candidate regression or classification models.
- Explain basic train/evaluation separation and the limits of feature-importance interpretations.
- Explore a small time-series example and combine an approved external source with local data.

Prerequisites

- Familiarity with using a personal computer.
- Good command of the English language.

- Comfortable using the internet.
- Basic understanding of File systems.

Access to a compatible KNIME Analytics Platform installation and the required example workflows/extensions on a supported operating system.

Approved sample data and a supplied or authorised API endpoint for external-source exercises; no paid Hub service or programming is required.

Training outline

1. Day 1 – Problem framing and data acquisition

- Data science and analytics: define a question, intended decision and evidence requirements before choosing a model.
- Translate the question into a visual process flow and identify required data, quality checks and appropriate methods.
- Explore the KNIME workspace, node configuration and execution; read spreadsheet and CSV samples.
- Filter, join, concatenate and aggregate tables; inspect keys, row counts and types.
- Use a prepared API or other approved source and inspect its response; discuss authentication and source limits.

2. Day 2 – Data preparation and exploration

- Identify cleaning needs, missing values, inconsistent formats and duplicate records; document handling decisions.
- Apply selected conversions, JSON/XML parsing, pivoting, discretisation and normalisation.
- Integrate a small external dataset with local data and check correspondence rather than assuming every source is compatible.
- Create views and summaries, check interpretation against the data and export a prepared result.

3. Day 3 – Models, evaluation and time-series awareness

- Distinguish regression and classification and choose an example target and features.
- Explore supplied linear-regression, k-nearest-neighbour and XGBoost workflows using the necessary extensions.
- Separate fitting from evaluation, inspect selected model settings and compare results against an appropriate baseline.
- Discuss feature contributions as model-dependent evidence rather than proof of causality.

- Inspect a small time series for aggregation and seasonality; introduce a supplied ARIMA/residual example with chronological evaluation and stated assumptions.
- Review the end-to-end workflow, its data limitations and the further study needed before operational use.