

Lab3: H2O AutoML in KNIME

This lab demonstrates how to train and evaluate a classification model using **H2O AutoML** inside KNIME.

Learning goals

Prepare a dataset for AutoML.

Train H2O AutoML models with a reproducible train/test split.

Evaluate predictions using KNIME's Scorer and interpret the confusion matrix and key metrics.

Dataset Description

The **Mall Customer** dataset contains **200 customers**. Each row is one subject. The goal is to predict the **Spending Class (High or Low Spender)** of the customer.

Target Variable

High Spender: "yes" if spending score ≥ 60 ; else "no".

Features:

CustomerID

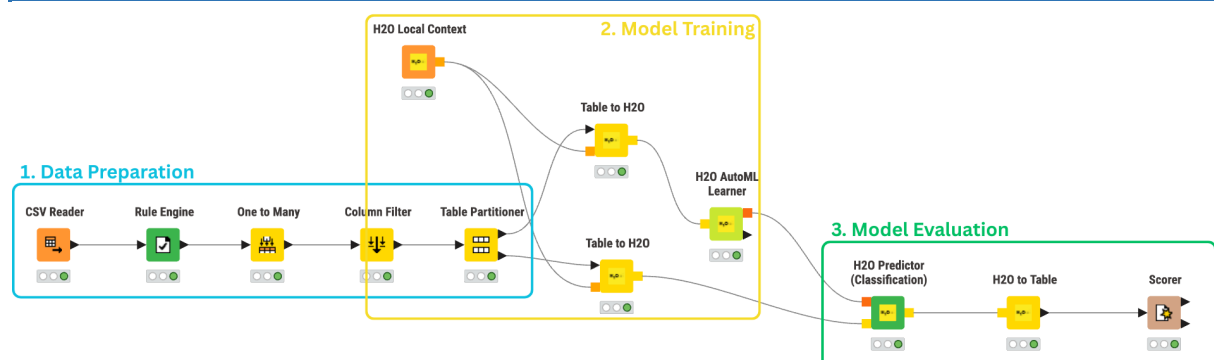
Gender: Male/Female

Age: (years)

Annual Income: (k\$)

Spending Score: Derived Metric (0-100)

KNIME Instructions



1) Data Preparation (blue block)

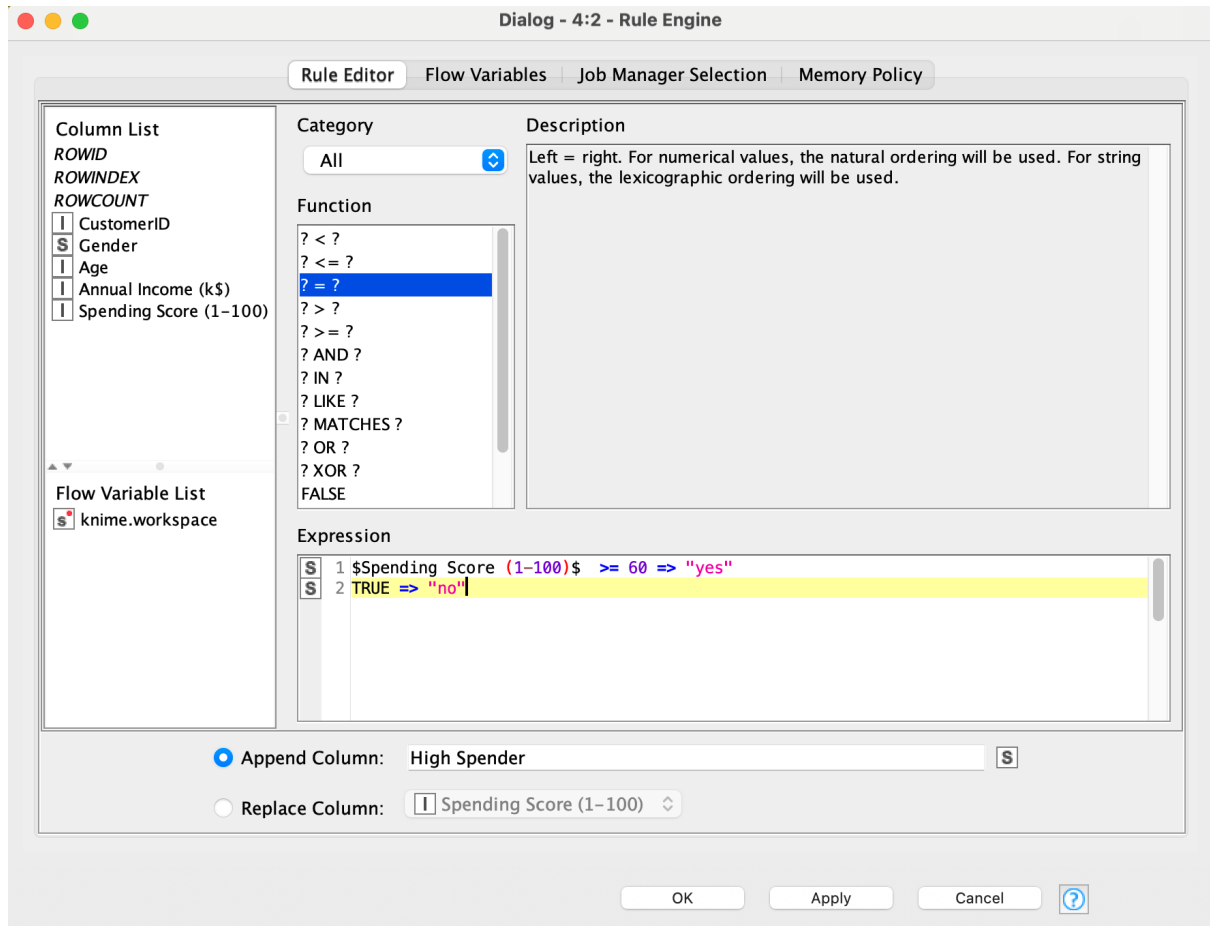
1.1 CSV Reader

Load your lab CSV file.

- Configure the file path.

1.2 Rule Engine

Create/clean the target label (class column).



1.3 One to Many

Encode categorical predictors into dummy/indicator columns.

One to Many ×

Columns to transform

Manual

Wildcard

Regex

Q

Search

Aa

Available columns

T

High Spender

Columns to transform

T

Gender

Any unknown column

>

>>

<

<<

☒ Remove included columns from output

1.4 Column Filter

Select the final modeling schema.

Column Filter

Column filter

Manual Wildcard Regex Type

Q Search

Aa

Excludes

Includes

123 CustomerID

123 Spending Score (1-1...

123 Female

>

>>

<

<<

123 Age

123 Annual Income (k\$)

T High Spender

123 Male

Any unknown column

1.5 Table Partitioner

Split into training and test sets.

Table Partitioner ✕

First partition type 📊 ?

Relative (%)

Absolute

Relative size

80 ⬆
⬇

Sampling strategy

Random

Stratified

Linear

First rows

Group column

📄 High Spender

⬇

☒ Fixed random seed

2026 ⬆
⬇

If input table is empty

Fail

Output empty table(s)

2) Model Training (yellow block)

2.1 H2O Local Context

Initialize the local H2O backend.

2.2 Table to H2O (Train)

Convert the KNIME training table into an H2OFrame.

Connect: Table Partitioner (Train) → Table to H2O.

2.3 Table to H2O (Test)

Convert the KNIME test table into an H2OFrame.

Connect: Table Partitioner (Test) → Table to H2O.

2.4 H2O AutoML Learner

Train a leaderboard of models automatically.

The screenshot shows the 'Dialog - 4:7 - H2O AutoML Learner' window. It has four tabs: 'General Settings', 'Algorithm Settings', 'Advanced Settings', and 'Flow Variables'. The 'General Settings' tab is active.

Target Column: A dropdown menu shows 'High Spender' with a blue icon to its right.

Selection Method: Three radio buttons are present: 'Manual Selection' (selected), 'Wildcard/Regex Selection', and 'Type Selection'.

Exclude Section (outlined in red):

- Header: 'Exclude'
- Filter: A text box with a funnel icon and the word 'Filter'. Below it, the text 'No columns in this list' is displayed.
- Enforce exclusion: A radio button that is selected.

Include Section (outlined in green):

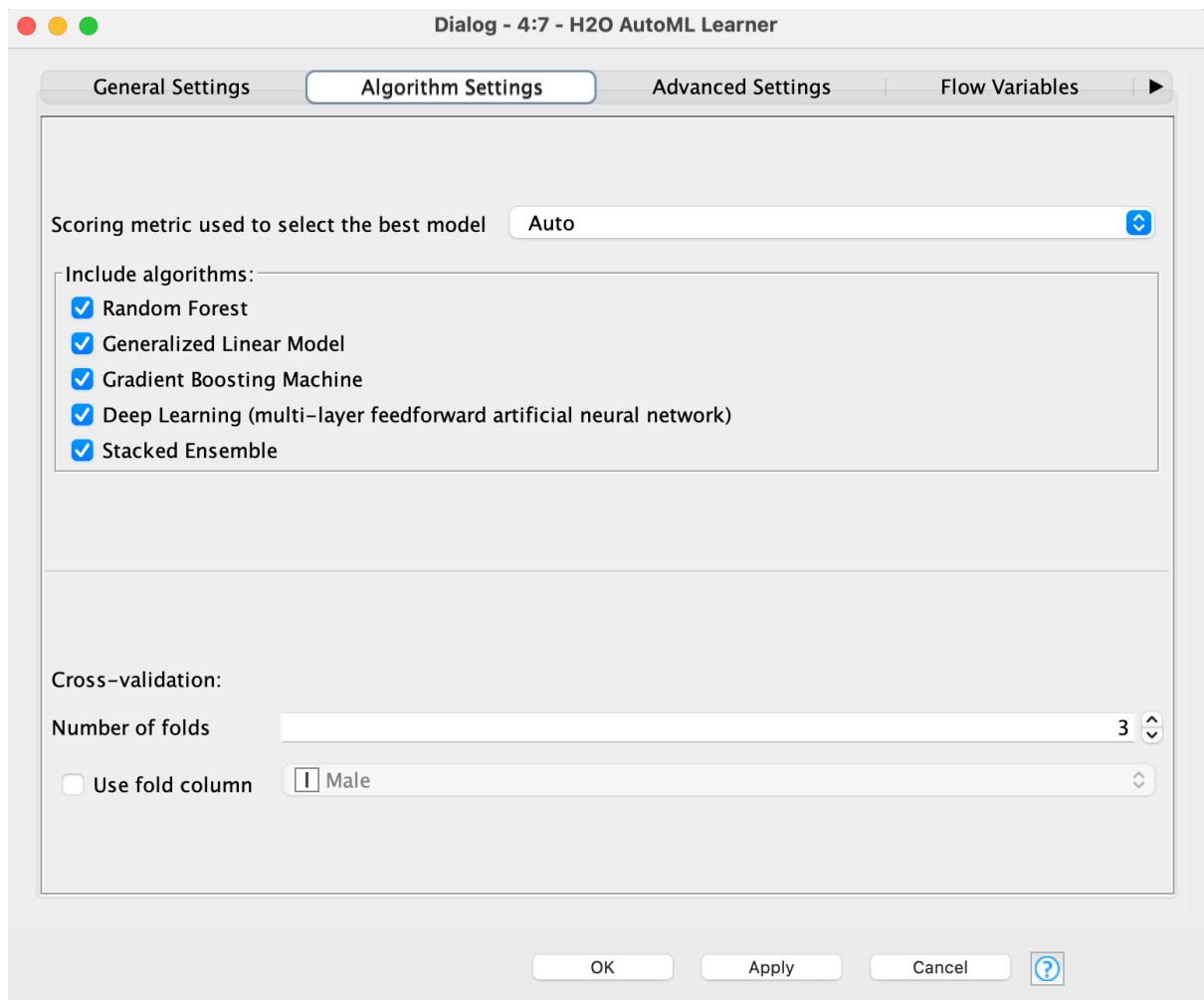
- Header: 'Include'
- Filter: A text box with a funnel icon and the word 'Filter'. Below it, a list of columns is shown: 'Age', 'Annual Income (k\$)', and 'Male'. Each column has a small square icon to its left.
- Enforce inclusion: A radio button that is not selected.

Stopping criteria:

- 'Max. runtime in seconds': A checkbox that is checked, followed by a text box containing '300' and a small up/down arrow icon.
- 'Max. number of models': A checkbox that is not checked, followed by a text box containing '10' and a small up/down arrow icon.

Use static random seed: A checkbox that is checked, followed by a text box containing '2026' and a 'New' button.

Buttons: At the bottom, there are 'OK', 'Apply', 'Cancel', and a help icon (a question mark inside a circle).



3) Model Evaluation (green block)

3.1 H2O Predictor (Classification)

Score the test H2OFrame using the trained AutoML model.

Connect:

- Model input: H2O AutoML Learner output
- Data input: Table to H2O (Test) output

Output: H2OFrame with predicted class and class probabilities.

3.2 H2O to Table

Convert predictions back to a KNIME table so you can use standard KNIME scoring and visualization nodes.

3.3 Scorer

Evaluate classification performance.

Scorer ×

First column

T High Spender ∨

Second column

T Prediction (High Spender) ∨

Sorting strategy

Insertion order ∨

☐ Reverse order

☐ Use name prefix

Missing values

Ignore

Fail