



**Universidade do Minho**  
Departamento de Informática

# **Ferramentas de Aprendizagem por Máquina**

*(Machine Learning Tools)*

**LEI @ 2025/2026, 2º sem**

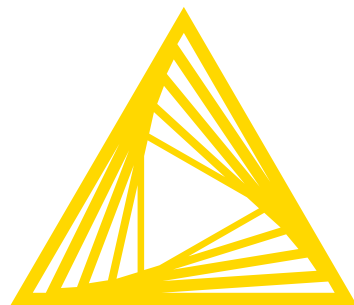


- Introdução à plataforma KNIME
- Construção de fluxos de análise de dados (KNIME *workflows*)
- Experimentação (*hands on*)





**Universidade do Minho**  
Departamento de Informática

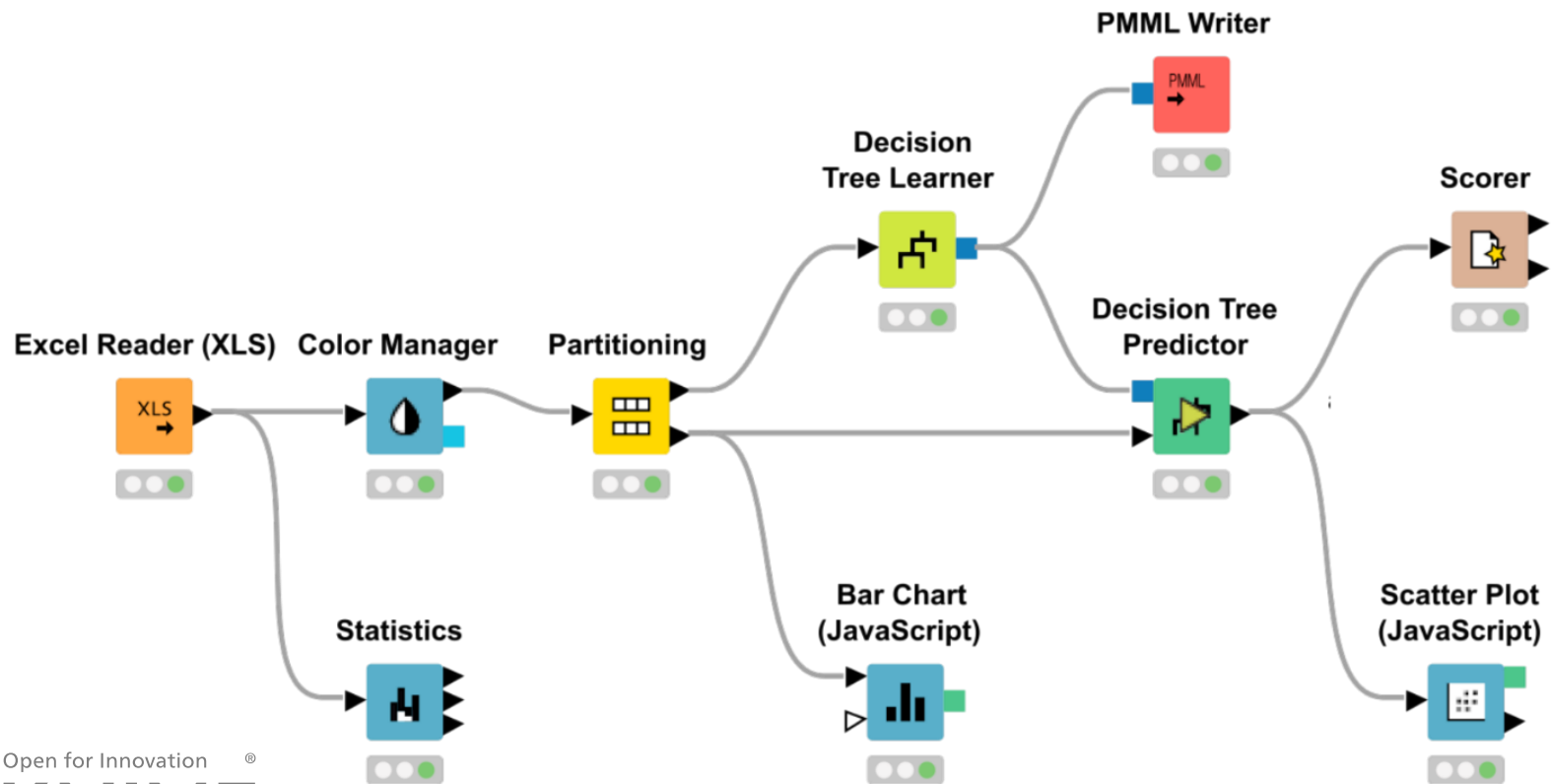


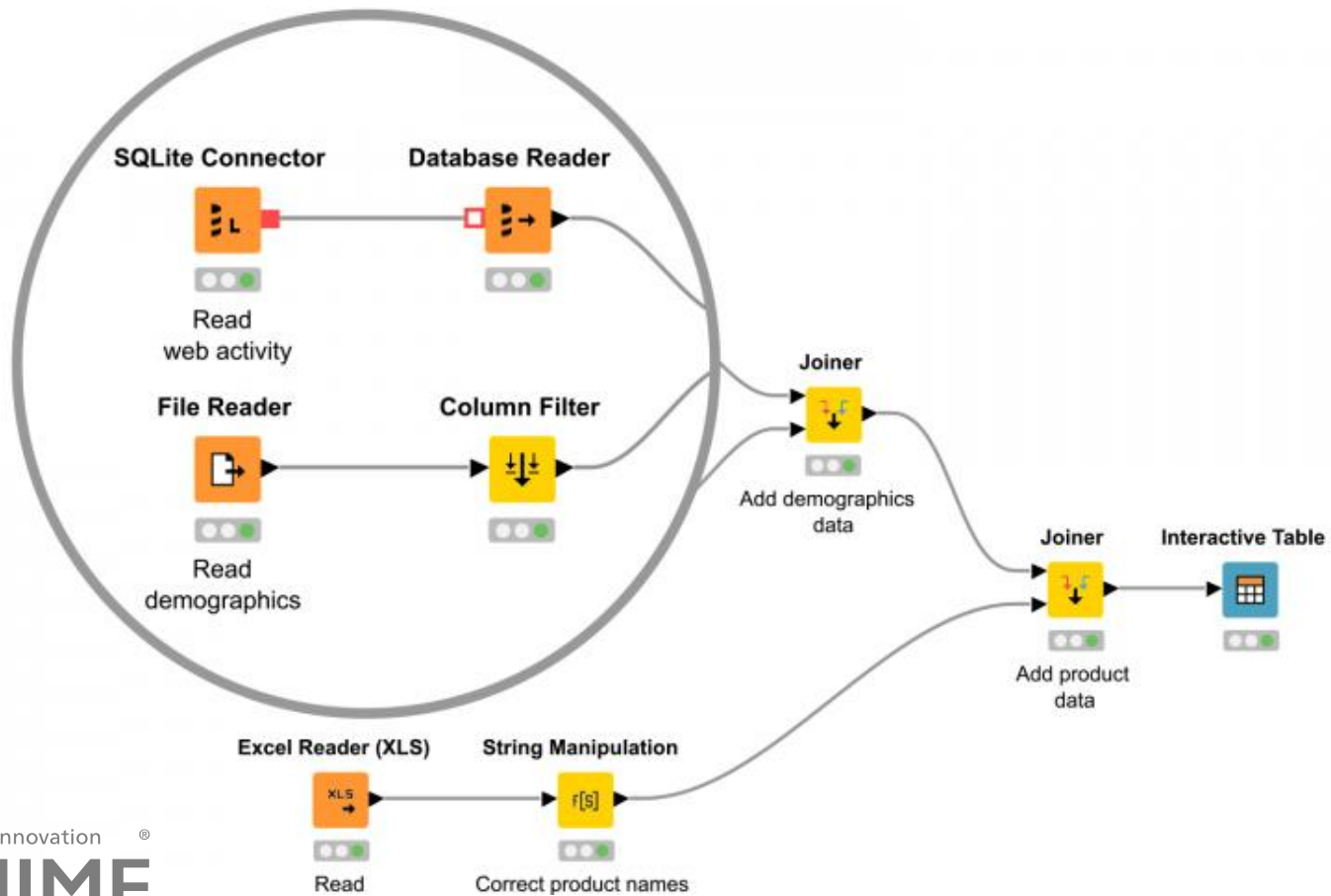
Open for Innovation<sup>®</sup>

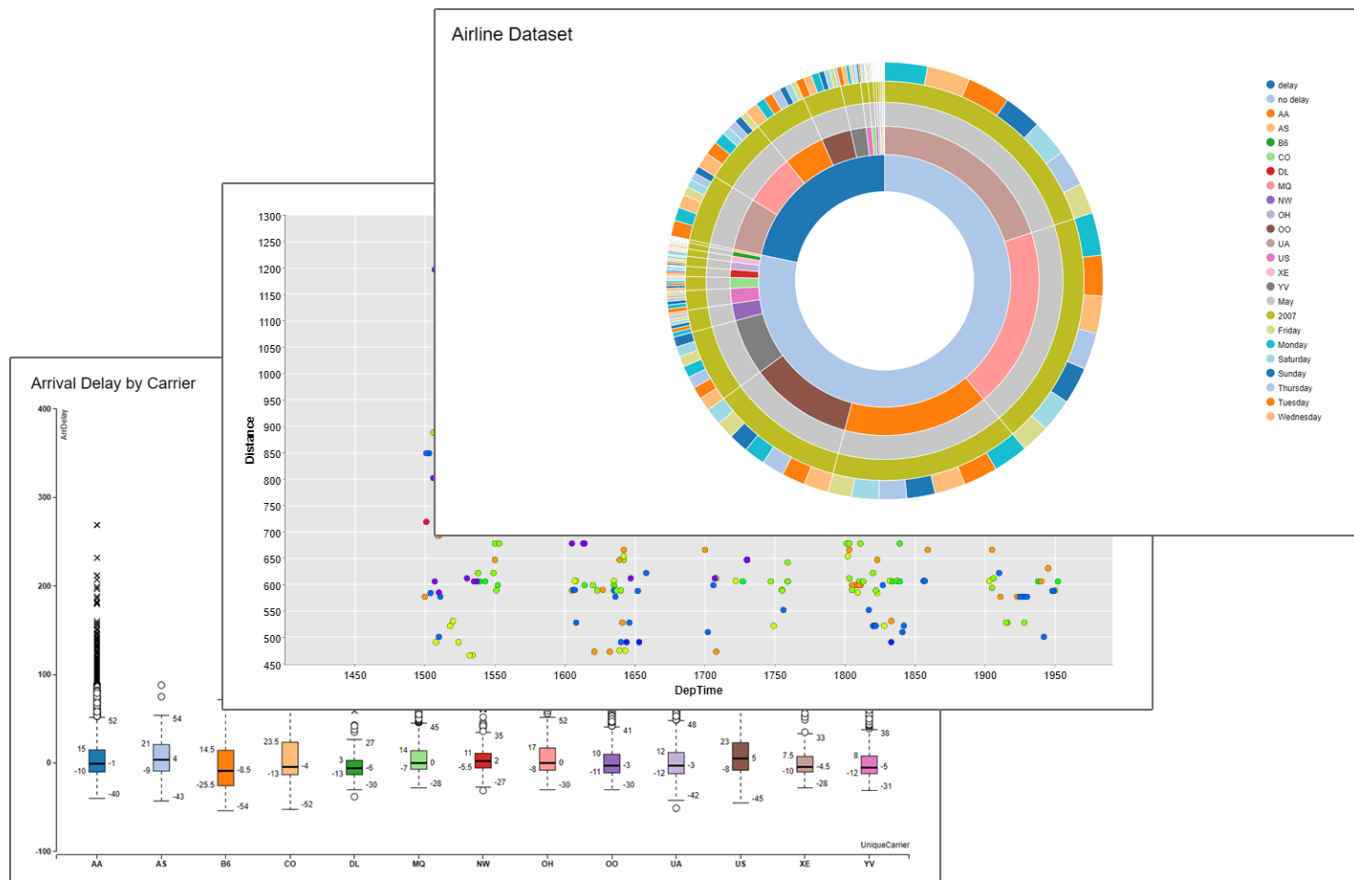
**KNIME**

**KNIME | Open for Innovation**











Home

KNIME Analytics Platform 5

Help

Preferences

Menu

## Get started with KNIME Analytics Platform 5



### Local space



The local space is the folder on your computer to store and access KNIME workflows and data produced by your workflows.



Create workflow  
in your local space.

### KNIME Community Hub ([hub.knime.com](https://hub.knime.com))

Sign in

Connect to the KNIME Community Hub to find workflows, nodes and components, and collaborate in spaces.





Home

Building a Simple Classifier ×

Help

Preferences

Menu

Save

Undo

Redo

Execute all

Cancel all

Reset all

93%

Nodes

Search all nodes

IO

Excel Reader

Excel Writer

Microsoft Authenticator

CSV Reader

CSV Writer

Table Creator

SharePoint Online Connector

File Reader

Show all

Manipulation

Row Filter

Column Filter

Concatenate

Value Lookup

Row Aggregator

Table Splitter

String Cleaner

Table Cropper

Show all

Views

1. Download Knime

2. Install it!

3. Try it!

[knime.com/downloads](https://www.knime.com/downloads)

To show the node output, please select a configured or executed node.



Home

Building a Simple Classifier ×

Help

Preferences

Menu

Execute all

Cancel all

Reset all

93%

Nodes

Search all nodes

IO

Excel Reader

Excel Writer

Microsoft Authenticator

CSV Reader

CSV Writer

Table Creator

SharePoint Online Connector

File Reader

Show all

Manipulation

Row Filter

Column Filter

Concatenate

Value Lookup

Row Aggregator

Table Splitter

String Cleaner

Table Cropper

Show all

Views

# QUICK HANDS ON

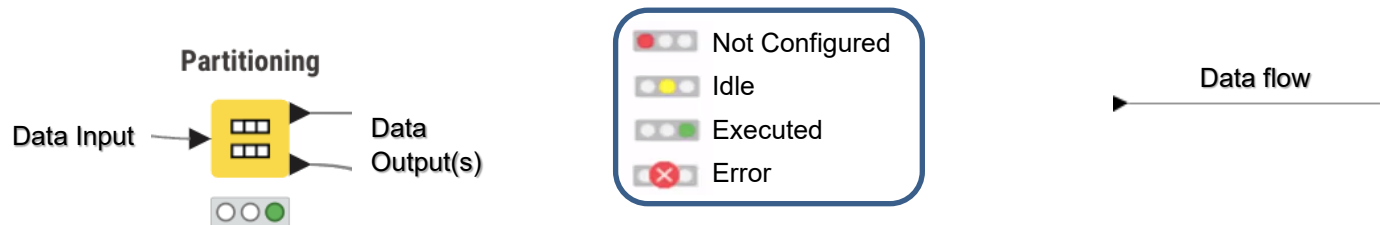
To show the node output, please select a configured or executed node.



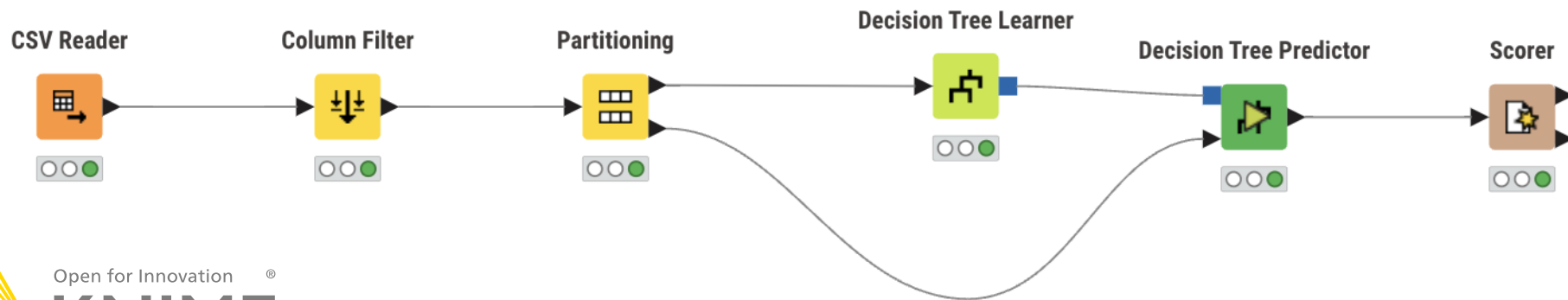
# Nodos e Fluxos

## *Nodes and Workflows*

### ■ Nodos *Nodes*



### ■ Fluxos *Workflows*





# Extensões KNIME

## KNIME *Extensions*

The screenshot displays the KNIME Modern UI interface. The top bar includes a 'Home' button and a breadcrumb 'Building a Simple Classifier'. Below this is a toolbar with icons for saving, undo, redo, and buttons for 'Execute all', 'Cancel all', and 'Reset all'. The left sidebar contains a 'Nodes' panel with a search bar and a grid of node categories: 'Excel Reader', 'Excel Writer', 'Microsoft Authenticator', 'CSV Reader', 'CSV Writer', 'Table Creator', 'SharePoint Online Connector', 'File Reader', and a 'Show all' button. A 'Manipulation' section is also visible, containing 'Row Filter', 'Column Filter', 'Concatenate', 'Value Lookup', 'Row Aggregator', and 'Table Splitter'. The main workspace shows a workflow with the following nodes: 'CSV Reader' (orange), 'Column Filter' (yellow), 'Partitioning' (yellow), 'Decision Tree Learner' (green), and 'Decision Tree Predictor' (green). The 'Decision Tree Predictor' node is highlighted with a red box. A 'Help' menu is open in the top right corner, listing options: 'Check for updates', 'Show KNIME log in File Explorer', 'Install extensions' (with a description: '... to access additional functionality, including complex data type processing and advanced algorithms.'), 'Switch workspace' (with a description: '... to access KNIME workflows and data from a different folder on your computer.'), and 'Switch to classic user interface' (with a description: '... to use the classic user interface. Switch back again, with the button "Open KNIME Modern UI" in the top right corner.').



# Extensões KNIME

## KNIME *Extensions*

The screenshot displays the KNIME software interface with the 'Available Software' dialog box open. The dialog box is titled 'Install' and contains the following elements:

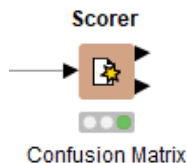
- Available Software**: A section with the instruction 'Check the items that you wish to install.' and a search bar labeled 'type filter text'.
- Table of Available Software**: A table with two columns, 'Name' and 'Version'. The table lists the following items:
  - > KNIME & Extensions
  - > KNIME Big Data Extensions
  - > KNIME Community Extensions - Bioinformatics & NGS
  - > KNIME Community Extensions - Cheminformatics
  - > KNIME Community Extensions - Image Processing and Analysis
  - > KNIME Community Extensions - Other
  - > KNIME Hub & Server Extensions
  - > KNIME Labs Extensions
  - > KNIME Node Development Tools
  - > KNIME Partner Extensions
- Details**: A section with a text area for details.
- Filters**: A section with checkboxes for filtering the software:
  - ☒ Show only the latest versions of available software
  - ☒ Group items by category
  - ☐ Show only software applicable to target environment
  - ☐ Hide items that are already installed
- Buttons**: 'Select All', 'Deselect All', '< Back', 'Next >', 'Cancel', and 'Finish'.

The background interface shows the 'Nodes' panel on the left with a search bar and a list of nodes. The 'Manipulation' section at the bottom lists nodes like 'Row Filter', 'Column Filter', 'Concatenate', 'Value Lookup', 'Row Aggregator', 'Table Splitter', 'String Cleaner', and 'Table Cropper'.



# Descrição dos Nodos

## *Node Description*



Description x KNIME Hub Search

Scorer

Compares two columns by their attribute value pairs and shows the confusion matrix, i.e. how many rows of which attribute and their classification match. Additionally, it is possible to highlight cells of this matrix to determine the underlying rows. The dialog allows you to select two columns for comparison; the values from the first selected column are represented in the confusion matrix's rows and the values from the second column by the confusion matrix's columns. The output of the node is the confusion matrix with the number of matches in each cell. Additionally, the second out-port reports a number of **accuracy statistics** such as True-Positives, False-Positives, True-Negatives, False-Negatives, Recall, Precision, Sensitivity, Specificity, F-measure, as well as the overall accuracy and **Cohen's kappa**.

Dialog Options

First column

The first column represents the real classes of the data.

Second column

The second column represents the predicted classes of the data.

Sorting strategy

Whether to sort the labels according to their appearance, or use the lexical/numeric ordering.

Reverse order

Reverse the order of the elements.

Use name prefix

The scores (i.e. accuracy, error rate, number of correct and wrong classification) are exported as flow variables with a hard coded name. This option allows you to define a prefix for these variable identifiers so that name conflicts are resolved.

### Missing Values

Choose how to treat missing values in either the reference or prediction column. Default is to ignore them (treat them as if the row did not exist). Alternatively, you can expect the table to not contain missing values in these two columns. If they do, the node will fail during execution.

### Ports

#### Input Ports

- 0 Table containing at least two columns to compare.

#### Output Ports

- 0 The confusion matrix.
- 1 The accuracy statistics table.

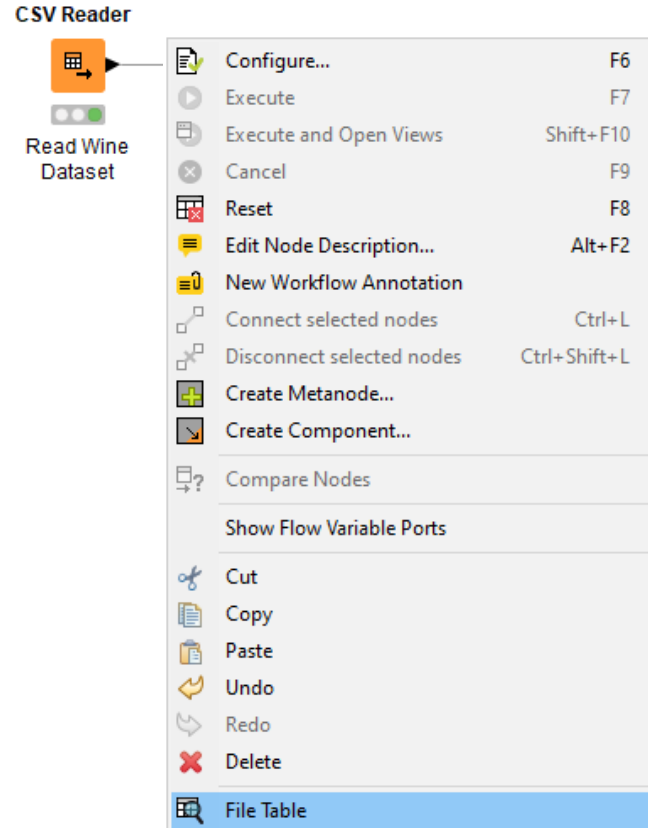
### Views

#### Confusion Matrix

Displays the confusion matrix in a table view. It is possible to highlight cells of the matrix which propagates highlighting to the corresponding rows. Therefore, it is possible for example to identify wrong predictions.



## Visualização do Ficheiro de Dados *Data Table Structure View*





# Visualização do Ficheiro de Dados

## *Data Table Structure View*

Column Headers

Data Type (Double)

Data Type (String)

Row ID

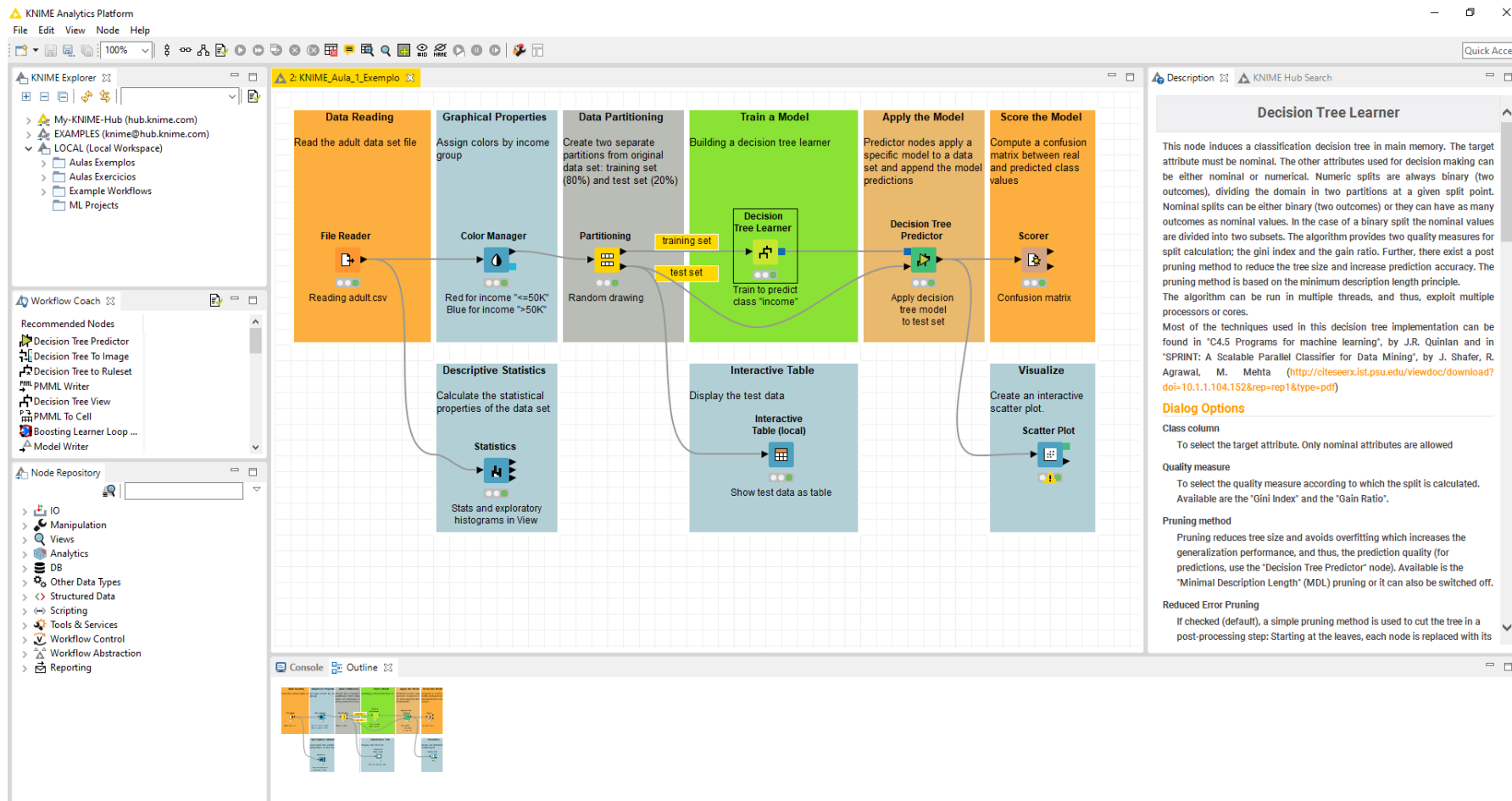
Data Cells

Row ID	D fixed a...	D volatile ...	D citric acid	D residual...	D chlo	t	t	t	t	t	t	alcohol	S quality
Row0	7.4	0.7	0	1.9	0.076								=5
Row1	7.8	0.88	0	2.6	0.098								=5
Row2	7.8	0.76	0.04	2.3	0.092								=5
Row3	11.2	0.28	0.56	1.9	0.075	17	60		0.998	3			=6
Row4	7.4	0.7	0	1.9	0.076	11	34		0.998	3			=5
Row5	7.4	0.66	0	1.8	0.075	13	40		0.998	3			=5
Row6	7.9	0.6	0.06	1.6	0.069	15	59		0.996	3			=5
Row7	7.3	0.65	0	1.2	0.065	15	21		0.995	3			=7
Row8	7.8	0.58	0.02	2	0.073	9	18		0.997	3	3.36	0.57	9.5
Row9	7.5	0.5	0.36	6.1	0.071	17	102		0.998	3	3.35	0.8	10.5
Row10	6.7	0.58	0.08	1.8	0.097	15	65		0.996	3	3.28	0.54	9.2
Row11	7.5	0.5	0.36	6.1	0.071	17	102		0.998	3	3.35	0.8	10.5
Row12	5.6	0.615	0	1.6	0.089	16	59		0.994	3	3.58	0.52	9.9
Row13	7.8	0.61	0.29	1.6	0.114	9	29		0.997	3	3.26	1.56	9.1
Row14	8.9	0.62	0.18	3.8	0.176	52	145		0.999	3	3.16	0.88	9.2
Row15	8.9	0.62	0.19	3.9	0.17	51	148		0.999	3	3.17	0.93	9.2
Row16	8.5	0.28	0.56	1.8	0.092	35	103		0.997	3	3.3	0.75	10.5
Row17	8.1	0.56	0.28	1.7	0.368	16	56		0.997	3	3.11	1.28	9.3
Row18	7.4	0.59	0.08	4.4	0.086	6	29		0.997	3	3.38	0.5	9
Row19	7.9	0.32	0.51	1.8	0.341	17	56		0.997	3	3.04	1.08	9.2
Row20	8.9	0.22	0.48	1.8	0.077	29	60		0.997	3	3.39	0.53	9.4
Row21	7.6	0.39	0.31	2.3	0.082	23	71		0.998	3	3.52	0.65	9.7
Row22	7.9	0.43	0.21	1.6	0.106	10	37		0.997	3	3.17	0.91	9.5
Row23	8.5	0.49	0.11	2.3	0.084	9	67		0.997	3	3.17	0.53	9.4
Row24	6.9	0.4	0.14	2.4	0.085	21	40		0.997	3	3.43	0.63	9.7
Row25	6.3	0.39	0.16	1.4	0.08	11	23		0.996	3	3.34	0.56	9.3
Row26	7.6	0.39	0.24	1.8	0.08	14	11		0.996	3	3.28	0.60	9.5





# Building a Simple Workflow





# Node Context Options: Data Loader

KNIME Analytics Platform

File Edit View Node Help

100%

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
  - Aulas Exemplos
  - Aulas Exercicios
  - Example Workflows
  - ML Projects

Workflow Coach

Recommended Nodes

- Joiner
- Column Filter
- Row Filter
- Partitioning
- GroupBy
- Missing Value
- Statistics
- k-Means

Node Repository

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Workflow Control
- Workflow Abstraction
- Reporting

2: KNIME\_Aula\_1\_Exemplo

**Data Reading**  
Read the adult data set file

**Graphical Properties**  
Assign colors by income group

**Data Partitioning**  
Create two separate partitions from original data set: training set (80%) and test set (20%)

**Train a Model**  
Building a decision tree learner

**Apply the Model**  
Predictor nodes apply a specific model to a data set and append the model predictions

**Score the Model**  
Compute a confusion matrix between real and predicted class values

**File Reader**  
Reading

**Color Manager**  
Color Manager

**Partitioning**  
Partitioning

**Decision Tree Learner**  
Train to predict class "income"

**Decision Tree Predictor**  
Apply decision tree model to test set

**Scorer**  
Confusion matrix

**Interactive Table**  
Display the test data

**Visualize**  
Create an interactive scatter plot

**File Reader Configuration Dialog**

- Execute
- Execute and Open Views
- Cancel
- Reset
- Edit Node Description...
- New Workflow Annotation
- Connect selected nodes
- Disconnect selected nodes
- Create Metanode...
- Create Component...
- Compare Nodes
- Show Flow Variable Ports
- Cut
- Copy
- Paste
- Undo
- Redo
- Delete
- File Table

**File Reader Description**

This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats.

When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).

The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.

**Dialog Options**

**ASCII file location**

Enter a valid file name or URL. When you press ENTER, the file is analyzed and the settings pre-set. You can also choose a previously read file from the drop-down list, or select a file from the "Browse..." dialog.

**Preserve user settings**

If checked, the checkpoints and column names/types you explicitly entered are preserved even if you select a new file. By default, the analyzer starts with fresh default settings for each new file location.

**Rescan**

If clicked, the file content is analyzed again. All settings are reset (unless the "Preserve user settings" option is selected) and the file is read in again to pre-set new settings and the table structure.

**Read row IDs**

If checked, the first column in the file is used as row IDs. If not checked, default row headers are created.

**Read column headers**



# Node Context Options: Data Loader

The screenshot displays the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' pane shows a project structure with 'LOCAL (Local Workspace)' containing 'Aulas Exemplos', 'Aulas Exercicios', 'Example Workflows', and 'ML Projects'. Below it, the 'Workflow Coach' pane lists recommended nodes like 'Joiner', 'Column Filter', 'Row Filter', 'Partitioning', 'GroupBy', 'Missing Value', 'Statistics', and 'k-Means'. The 'Node Repository' pane at the bottom shows various node categories including 'IO', 'Manipulation', 'Views', 'Analytics', 'DB', 'Other Data Types', 'Structured Data', 'Scripting', 'Tools & Services', 'Workflow Control', 'Workflow Abstraction', and 'Reporting'.

The main workspace shows a workflow titled '2- KNIME\_Aula\_1\_Exemplo'. It includes a 'Data Reading' node labeled 'File Reader' with the description 'Read the adult data set file'. This node is connected to a 'Graphical Properties' node labeled 'Color Manager' with the description 'Assign colors by income group'. The 'Color Manager' node is further connected to a 'Descriptive Statistics' node labeled 'Statistics' with the description 'Calculate the statistical properties of the data'. The 'Statistics' node is connected to a 'Visualize' node labeled 'Scatter Plot' with the description 'Create an interactive scatter plot'.

The 'File Reader' node's configuration dialog is open, showing the 'Settings' tab. The 'Enter ASCII data file location' field contains 'knime://knime.workflow/adult.csv'. The 'Basic Settings' section includes checkboxes for 'read row IDs' (unchecked), 'read column headers' (checked), 'ignore spaces and tabs' (checked), and 'Java-style comments' (unchecked). The 'Column delimiter' is set to a comma. The 'Advanced...' button is visible. The 'Preview' section shows a table of data with columns: Row ID, age, workclass, fnlwgt, education, education-num, marital-status, and more. The table contains 25 rows of data.

The 'Dialog - 2:1 - File Reader (Reading adult.csv)' window is also visible, showing the 'Settings' tab. The 'Enter ASCII data file location' field contains 'knime://knime.workflow/adult.csv'. The 'Basic Settings' section includes checkboxes for 'read row IDs' (unchecked), 'read column headers' (checked), 'ignore spaces and tabs' (checked), and 'Java-style comments' (unchecked). The 'Column delimiter' is set to a comma. The 'Advanced...' button is visible. The 'Preview' section shows a table of data with columns: Row ID, age, workclass, fnlwgt, education, education-num, marital-status, and more. The table contains 25 rows of data.

The 'Description' pane on the right shows the 'File Reader' node's description: 'This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats. When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below). The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.' Below the description, the 'Dialog Options' section lists 'ASCII file location', 'Preserve user settings', 'Rescan', 'Read row IDs', and 'Read column headers'.



# Node Context Options: Model Learner

KNIME Analytics Platform

File Edit View Node Help

100%

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
  - Aulas Exemplos
  - Aulas Exercicios
  - Example Workflows
  - ML Projects

Workflow Coach

Recommended Nodes

- Decision Tree Predictor
- Decision Tree To Image
- Decision Tree To Ruleset
- PMML Writer
- Decision Tree View
- PMML To Cell
- Boosting Learner Loop ...
- Model Writer

Node Repository

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Workflow Control
- Workflow Abstraction
- Reporting

2: KNIME\_Aula\_1\_Exemplo

**Data Reading**  
Read the adult data set file  
File Reader  
Reading adult.csv

**Graphical Properties**  
Assign colors by income group  
Color Manager  
Red for income "<=50K"  
Blue for income ">50K"

**Data Partitioning**  
Create two separate partitions from original data set: training set (80%) and test set (20%)  
Partitioning  
Random drawing

**Train a Model**  
Building a decision tree learner  
Decision Tree Learner  
Train to class w

**Apply the Model**  
Predictor nodes apply a specific model to a data set and append the model predictions  
Decision Tree Predictor

**Score the Model**  
Compute a confusion matrix between real and predicted class values  
Scorer

**Descriptive Statistics**  
Calculate the statistical properties of the data set  
Statistics  
Stats and exploratory histograms in View

**Interactions**  
Display the test results  
Show test results

Context Menu:

- Configure... F6
- Execute F7
- Execute and Open Views Shift+F10
- Cancel F9
- Reset F8
- Edit Node Description... Alt+F2
- New Workflow Annotation Ctrl+L
- Connect selected nodes
- Disconnect selected nodes Ctrl+Shift+L
- Create Metanode...
- Create Component...
- View: Decision Tree View
- View: Decision Tree View (simple)
- Compare Nodes
- Show Flow Variable Ports
- Cut
- Copy
- Paste
- Undo
- Redo
- Delete
- Decision Tree Model

Console Outline

Decision Tree Learner

This node induces a classification decision tree in main memory. The target attribute must be nominal. The other attributes used for decision making can be either nominal or numerical. Numeric splits are always binary (two outcomes), dividing the domain in two partitions at a given split point. Nominal splits can be either binary (two outcomes) or they can have as many outcomes as nominal values. In the case of a binary split the nominal values are divided into two subsets. The algorithm provides two quality measures for split calculation; the gini index and the gain ratio. Further, there exist a post pruning method to reduce the tree size and increase prediction accuracy. The pruning method is based on the minimum description length principle. The algorithm can be run in multiple threads, and thus, exploit multiple processors or cores.

Most of the techniques used in this decision tree implementation can be found in "C4.5 Programs for machine learning", by J.R. Quinlan and in "SPRINT: A Scalable Parallel Classifier for Data Mining", by J. Shafer, R. Agrawal, M. Mehta (<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.152&rep=rep1&type=pdf>)

**Dialog Options**

**Class column**  
To select the target attribute. Only nominal attributes are allowed

**Quality measure**  
To select the quality measure according to which the split is calculated. Available are the "Gini Index" and the "Gain Ratio".

**Pruning method**  
Pruning reduces tree size and avoids overfitting which increases the generalization performance, and thus, the prediction quality (for predictions, use the "Decision Tree Predictor" node). Available is the "Minimal Description Length" (MDL) pruning or it can also be switched off.

**Reduced Error Pruning**  
If checked (default), a simple pruning method is used to cut the tree in a post-processing step: Starting at the leaves, each node is replaced with its



# Node Context Options: Model Learner

The screenshot displays the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' shows a project structure with 'LOCAL (Local Workspace)' containing 'Aulas Ejemplos', 'Aulas Ejercicios', 'Example Workflows', and 'ML Projects'. Below it, the 'Workflow Coach' lists recommended nodes like 'Decision Tree Predictor', 'Decision Tree To Image', 'Decision Tree To Ruleset', 'PMML Writer', 'Decision Tree View', 'PMML To Cell', 'Boosting Learner Loop ...', and 'Model Writer'. The 'Node Repository' at the bottom left shows categories like IO, Manipulation, Views, Analytics, DB, Other Data Types, Structured Data, Scripting, Tools & Services, Workflow Control, Workflow Abstraction, and Reporting.

The main workspace shows a workflow titled '2: KNIME\_Aula\_1\_Ejemplo'. The workflow consists of several nodes: 'File Reader' (Data Reading), 'Color Manager' (Graphical Properties), 'Partitioning' (Data Partitioning), 'Train a Model' (Model Learner), 'Apply the Model' (Apply the Model), 'Score the Model' (Score the Model), and 'Visualize' (Visualize). The 'Train a Model' node is highlighted, and its context menu is open, showing options like 'General', 'PMML Settings', 'Flow Variables', and 'Memory Policy'. The 'General' tab is selected, showing settings for the 'Decision Tree Learner' node.

The 'Decision Tree Learner' dialog box is open, showing the following settings:

- Class column:
- Quality measure:
- Pruning method:
- ☒ Reduced Error Pruning
- Min number records per node:
- Number records to store for view:
- ☒ Average split point
- Number threads:
- ☒ Skip nominal columns without domain information
- Root split: ☐ Force root split column
- Root split column:
- Binary nominal splits: ☐ Binary nominal splits
- Max #nominal:
- ☐ Filter invalid attribute values in child nodes

On the right, the 'Description' pane shows the 'Decision Tree Learner' node's documentation. It explains that the node induces a classification decision tree in main memory, where the target attribute must be nominal. It also describes the quality measures (Gini index and gain ratio) and the pruning methods (Reduced Error Pruning and Minimal Description Length).



# Node Context Options: Model Learner

KNIME Analytics Platform

File Edit View Node Help

100%

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
  - Aulas Exemplos
  - Aulas Exercicios
  - Example Workflows
  - ML Projects

Workflow Coach

Recommended Nodes

- Decision Tree Predictor
- Decision Tree To Image
- Decision Tree To Ruleset
- PMML Writer
- Decision Tree View
- PMML To Cell
- Boosting Learner Loop ...
- Model Writer

Node Repository

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Workflow Control
- Workflow Abstraction
- Reporting

2: KNIME\_Aula\_1\_Exemplo

**Data Reading**  
Read the adult data set file  
File Reader  
Reading adult.csv

**Graphical Properties**  
Assign colors by income group  
Color Manager  
Red for income "<=50K"  
Blue for income ">50K"

**Data Partitioning**  
Create two separate partitions from original data set: training set (80%) and test set (20%)  
Partitioning  
Random drawing

**Train a Model**  
Building a decision tree learner  
Decision Tree Learner  
Train to class

**Apply the Model**  
Predictor nodes apply a specific model to a data set and append the model predictions  
Decision Tree Predictor

**Score the Model**  
Compute a confusion matrix between real and predicted class values  
Scorer  
Confusion matrix

**Descriptive Statistics**  
Calculate the statistical properties of the data set  
Statistics  
Stats and exploratory histograms in View

**Interactions**  
Display the test  
Show test

Context Menu:

- Configure...
- Execute
- Execute and Open Views
- Cancel
- Reset
- Edit Node Description...
- New Workflow Annotation
- Connect selected nodes
- Disconnect selected nodes
- Create Metanode...
- Create Component...
- View: Decision Tree View
- View: Decision Tree View (simple)
- Compare Nodes
- Show Flow Variable Ports
- Cut
- Copy
- Paste
- Undo
- Redo
- Delete
- Decision Tree Model

Console

Outline

Quick Access

Description

KNIME Hub Search

## Decision Tree Learner

This node induces a classification decision tree in main memory. The target attribute must be nominal. The other attributes used for decision making can be either nominal or numerical. Numeric splits are always binary (two outcomes), dividing the domain in two partitions at a given split point. Nominal splits can be either binary (two outcomes) or they can have as many outcomes as nominal values. In the case of a binary split the nominal values are divided into two subsets. The algorithm provides two quality measures for split calculation; the gini index and the gain ratio. Further, there exist a post pruning method to reduce the tree size and increase prediction accuracy. The pruning method is based on the minimum description length principle. The algorithm can be run in multiple threads, and thus, exploit multiple processors or cores. Most of the techniques used in this decision tree implementation can be found in "C4.5 Programs for machine learning", by J.R. Quinlan and in "SPRINT: A Scalable Parallel Classifier for Data Mining", by J. Shafer, R. Agrawal, M. Mehta (<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.152&rep=rep1&type=pdf>)

### Dialog Options

**Class column**  
To select the target attribute. Only nominal attributes are allowed

**Quality measure**  
To select the quality measure according to which the split is calculated. Available are the "Gini Index" and the "Gain Ratio".

**Pruning method**  
Pruning reduces tree size and avoids overfitting which increases the generalization performance, and thus, the prediction quality (for predictions, use the "Decision Tree Predictor" node). Available is the "Minimal Description Length" (MDL) pruning or it can also be switched off.

**Reduced Error Pruning**  
If checked (default), a simple pruning method is used to cut the tree in a post-processing step: Starting at the leaves, each node is replaced with its



# Node Context Options: Model Learner

KNIME Analytics Platform

File Edit View Node Help

100%

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
  - Aulas Exemplos
  - Aulas Exercicios
  - Example Workflows
  - ML Projects

Workflow Coach

Recommended Nodes

- Decision Tree Predictor
- Decision Tree To Image
- Decision Tree To Ruleset
- PMML Writer
- Decision Tree View
- PMML To Cell
- Boosting Learner Loop ...
- Model Writer

Node Repository

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Workflow Control
- Workflow Abstraction
- Reporting

2: KNIME\_Aula\_1\_Example

Data Reading: Read the adult data set file. File Reader (Reading adult.csv)

Graphical Properties: Assign colors by income group. Color Manager (Red for income "<=50K", Blue for income ">50K")

Data Partitioning: Create two separate partitions from original data set: training set (80%) and test set (20%). Partitioning

Train a Model: Building a decision tree learner. Decision Tree Learner

Apply the Model: Predictor nodes apply a specific model to a data set and append the model predictions. Decision Tree Predictor

Score the Model: Compute a confusion matrix between real and predicted class values. Scorer

Decision Tree View - 2:10 - Decision Tree Learner (Train to predict)

File HiLite Tree

Table:

Category	%	n
<=50K	76,0	19.787
>50K	24,0	6.261
Total	100,0	26.048

Chart:

Color column: income

relationship

Not-in-family (5.912/6.593)

= Husband (5.860/10.585)

= Wife (650/1.244)

= Own child (4.012/4.065)

= Unmarried (2.577/2.754)

= Other relative (776/807)

Zoom: 100.0%





# Node Context Options: Scorer

KNIME Analytics Platform

File Edit View Node Help

100%

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
- Aulas Exemplos
- Aulas Exercicios
- Example Workflows
- ML Projects

Workflow Coach

Recommended Nodes

- ROC Curve (local)
- Joiner
- Row Filter
- Column Filter
- Interactive Table (local)
- Feature Selection Loop ...
- Excel Writer (XLS)
- CSV Writer

Node Repository

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Workflow Control
- Workflow Abstraction
- Reporting

2: KNIME\_Aula\_1\_Exemplo

Data Reading

Read the adult data set file

File Reader

Reading adult.csv

Graphical Properties

Assign colors by income group

Color Manager

Red for income "<=50K"  
Blue for income ">50K"

Descriptive Statistics

Calculate the statistical properties of the data set

Statistics

Stats and exploratory histograms in view

Data Partitioning

Create two separate partitions from original data set: training set (80%) and test set (20%)

Partitioning

Random drawing

training set

test set

Train a Model

Building a decision tree learner

Decision Tree Learner

Train to predict class "income"

Apply the Model

Predictor nodes apply a specific model to a data set and append the model predictions

Decision Tree Predictor

Apply decision tree model to test set

Score the Model

Compute a confusion matrix between real and predicted class values

Scorer

Confusion matrix

Interactive Table

Display the test data

Interactive Table (local)

Show test data as table

Context Menu:

- Configure... F6
- Execute F7
- Execute and Open Views Shift+F10
- Cancel F9
- Reset F8
- Edit Node Description... Alt+F2
- New Workflow Annotation
- Connect selected nodes Ctrl+L
- Disconnect selected nodes Ctrl+Shift+L
- Create a scatter plot
- Create Metanode...
- Create Component...
- View: Confusion Matrix
- Compare Nodes
- Show Flow Variable Ports
- Cut
- Copy
- Paste
- Undo
- Redo
- Delete
- Confusion matrix
- Accuracy statistics

Description

KNIME Hub Search

Scorer

Compares two columns by their attribute value pairs and shows the confusion matrix, i.e. how many rows of which attribute and their classification match. Additionally, it is possible to highlight cells of this matrix to determine the underlying rows. The dialog allows you to select two columns for comparison; the values from the first selected column are represented in the confusion matrix's rows and the values from the second column by the confusion matrix's columns. The output of the node is the confusion matrix with the number of matches in each cell. Additionally, the second out-port shows statistics such as True-Positives, False-Negatives, Recall, Precision, Sensitivity, the overall accuracy and Cohen's kappa.

the real classes of the data.

ts the predicted classes of the data.

according to their appearance, or use the

ents.

or rate, number of correct and wrong

is flow variables with a hard coded name.

fine a prefix for these variable identifiers so

ved.

values in either the reference or prediction

them (treat them as if the row did not exist).

the table to not contain missing values in

Console

Outline





# Node Context Options: Scorer

KNIME Analytics Platform

File Edit View Node Help

100%

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
  - Aulas Exemplos
  - Aulas Exercicios
  - Example Workflows
  - ML Projects

Workflow Coach

Recommended Nodes

- ROC Curve (local)
- Joiner
- Row Filter
- Column Filter
- Interactive Table (local)
- Feature Selection Loop ...
- Excel Writer (XLS)
- CSV Writer

Node Repository

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Workflow Control
- Workflow Abstraction
- Reporting

2: KNIME\_Aula\_1\_Example

Data Reading: Read the adult data set file. File Reader. Reading adult.csv

Graphical Properties: Assign colors by income group. Color Manager. Red for income "<=50K" Blue for income ">50K"

Data Partitioning: Create two separate partitions from original data set: training set (80%) and test set (20%). Partitioning. Random draw

Train a Model: Building a decision tree learner. Decision Tree Learner

Apply the Model: Predictor nodes apply a specific model to a data set and append the model predictions. Decision Tree

Score the Model: Compute a confusion matrix between real and predicted class values. Scorer. Confusion matrix

Visualize: Create an interactive scatter plot. Scatter Plot

Descriptive Statistics: Calculate the statistical properties of the data set. Statistics. Stats and exploratory histograms in View

Dialog - 2:6 - Scorer (Confusion matrix)

Scorer | Flow Variables | Memory Policy

First Column: S income

Second Column: S Prediction (income)

Sorting of values in tables: Sorting strategy: Insertion order ☐ Reverse order

Provide scores as flow variables: ☐ Use name prefix

Missing values: In case of missing values... ☒ Ignore ☐ Fail

OK Apply Cancel ?

Description

## Scorer

Compares two columns by their attribute value pairs and shows the confusion matrix, i.e. how many rows of which attribute and their classification match. Additionally, it is possible to highlight cells of this matrix to determine the underlying rows. The dialog allows you to select two columns for comparison: the values from the first selected column are represented in the confusion matrix's rows and the values from the second column by the confusion matrix's columns. The output of the node is the confusion matrix with the number of matches in each cell. Additionally, the second out-port reports a number of **accuracy statistics** such as True-Positives, False-Positives, True-Negatives, False-Negatives, Recall, Precision, Sensitivity, Specificity, F-measure, as well as the overall accuracy and **Cohen's kappa**.

### Dialog Options

**First column**  
The first column represents the real classes of the data.

**Second column**  
The second column represents the predicted classes of the data.

**Sorting strategy**  
Whether to sort the labels according to their appearance, or use the lexical/numeric ordering.

**Reverse order**  
Reverse the order of the elements.

**Use name prefix**  
The scores (i.e. accuracy, error rate, number of correct and wrong classification) are exported as flow variables with a hard coded name. This option allows you to define a prefix for these variable identifiers so that name conflicts are resolved.

**Missing Values**  
Choose how to treat missing values in either the reference or prediction column. Default is to ignore them (treat them as if the row did not exist). Alternatively, you can expect the table to not contain missing values in

Console Outline



# Node Context Options: Scorer

KNIME Analytics Platform

File Edit View Node Help

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
  - Aulas Exemplos
  - Aulas Exercicios
  - Example Workflows
  - ML Projects

Workflow Coach

Recommended Nodes

- ROC Curve (local)
- Joiner
- Row Filter
- Column Filter
- Interactive Table (local)
- Feature Selection Loop ...
- Excel Writer (XLS)
- CSV Writer

Node Repository

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Workflow Control
- Workflow Abstraction
- Reporting

2: KNIME\_Aula\_1\_Exemplo

Data Reading: Read the adult data set file. File Reader (Reading adult.csv)

Graphical Properties: Assign colors by income group. Color Manager (Red for income "<=50K", Blue for income ">50K")

Data Partitioning: Create two separate partitions from original data set: training set (80%) and test set (20%). Partitioning (Random drawing)

Train a Model: Building a decision tree learner. Decision Tree Learner (Train to predict class "income")

Apply the Model: Predictor nodes apply a specific model to a data set and append the model predictions. Decision Tree Predictor (Apply decision tree model to test set)

Score the Model: Compute a confusion matrix between real and predicted class values. Scorer (Confusion)

Descriptive Statistics: Calculate the statistical properties of the data set. Statistics (Stats and exploratory histograms in View)

Interactive Table: Display the test data. Interactive Table (local) (Show test data as table)

Visualize: Create an interactive scatter plot. Scatter Plot (Show test data as table)

Scorer

Compares two columns by their attribute value pairs and shows the confusion matrix, i.e. how many rows of which attribute and their classification match. Additionally, it is possible to highlight cells of this matrix to determine the underlying rows. The dialog allows you to select two columns for comparison: the values from the first selected column are represented in the confusion matrix's rows and the values from the second column by the confusion matrix's columns. The output of the node is the confusion matrix with the number of matches in each cell. Additionally, the second out-port shows statistics such as True-Positives, False-Negatives, Recall, Precision, Sensitivity, the overall accuracy and Cohen's kappa.

Statistics

real classes of the data.

the predicted classes of the data.

rding to their appearance, or use the

ts.

rate, number of correct and wrong

flow variables with a hard coded name.

e a prefix for these variable identifiers so

d.

values in either the reference or prediction

m (treat them as if the row did not exist).

ie table to not contain missing values in

Configure...

Execute

Execute and Open Views

Cancel

Reset

Edit Node Description...

New Workflow Annotation

Connect selected nodes

Disconnect selected nodes

Create Metanode...

Create Component...

View: Confusion Matrix

Compare Nodes

Show Flow Variable Ports

Cut

Copy

Paste

Undo

Redo

Delete

Confusion matrix

Accuracy statistics

Console

Outline



# Node Context Options: Scorer

KNIME Analytics Platform

File Edit View Node Help

100%

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
  - Aulas Exemplos
  - Aulas Exercicios
  - Example Workflows
  - ML Projects

Workflow Coach

Recommended Nodes

- ROC Curve (local)
- Joiner
- Row Filter
- Column Filter
- Interactive Table (local)
- Feature Selection Loop ...
- Excel Writer (XLS)
- CSV Writer

Node Repository

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Workflow Control
- Workflow Abstraction
- Reporting

2: KNIME\_Aula\_1.Exemplo

Data Reading: Read the adult data set file. File Reader. Reading adult.csv

Graphical Properties: Assign colors by income group. Color Manager. Red for income "<=50K" Blue for income ">50K"

Data Partitioning: Create two separate partitions from original data set: training set (80%) and test set (20%). Partition. File. Hilite. income \ Pr... <=50K >50K <=50K 4557 376 >50K 674 906

Train a Model: Building a decision tree learner

Apply the Model: Predictor nodes apply a specific model to a data set and append the model predictions

Score the Model: Compute a confusion matrix between real and predicted class values. Scorer. Confusion matrix

Descriptive Statistics: Calculate the statistical properties of the data set. Statistics. Stats and exploratory histograms in View

Visualize: Create an interactive scatter plot. Scatter Plot

Confusion Matrix - 2:6 - Scorer (Confusion matrix)

	<=50K	>50K
income \ Pr...	4557	376
<=50K	4557	376
>50K	674	906

Correct classified: 5.463  
Accuracy: 83,878 %  
Cohen's kappa (κ) 0,531

Wrong classified: 1.050  
Error: 16,122 %

Scorer

Compares two columns by their attribute value pairs and shows the confusion matrix, i.e. how many rows of which attribute and their classification match. Additionally, it is possible to highlight cells of this matrix to determine the underlying rows. The dialog allows you to select two columns for comparison: the values from the first selected column are represented in the confusion matrix's rows and the values from the second column by the confusion matrix's columns. The output of the node is the confusion matrix with the number of matches in each cell. Additionally, the second out-port reports a number of **accuracy statistics** such as True-Positives, False-Positives, True-Negatives, False-Negatives, Recall, Precision, Sensitivity, Specificity, F-measure, as well as the overall accuracy and **Cohen's kappa**.

Dialog Options

First column  
The first column represents the real classes of the data.

Second column  
The second column represents the predicted classes of the data.

Sorting strategy  
Whether to sort the labels according to their appearance, or use the lexical/numeric ordering.

Reverse order  
Reverse the order of the elements.

Use name prefix  
The scores (i.e. accuracy, error rate, number of correct and wrong classification) are exported as flow variables with a hard coded name. This option allows you to define a prefix for these variable identifiers so that name conflicts are resolved.

Missing Values  
Choose how to treat missing values in either the reference or prediction column. Default is to ignore them (treat them as if the row did not exist). Alternatively, you can expect the table to not contain missing values in

Console Outline



Home

Building a Simple Classifier x

Help

Preferences

Menu

Execute all

Cancel all

Reset all

93%

Nodes

Search all nodes

IO

Excel Reader

Excel Writer

Microsoft Authenticator

CSV Reader

CSV Writer

Table Creator

SharePoint Online Connector

File Reader

Show all

Manipulation

Row Filter

Column Filter

Concatenate

Value Lookup

Row Aggregator

Table Splitter

String Cleaner

Table Cropper

Show all

Views

# QUICK HANDS ON

To show the node output, please select a configured or executed node.



**Universidade do Minho**  
Departamento de Informática

# **Ferramentas de Aprendizagem por Máquina**

*(Machine Learning Tools)*

**LEI @ 2025/2026, 2º sem**