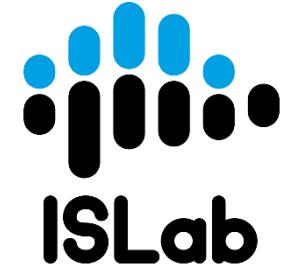


University of Minho
School of Engineering



Machine Learning

Intelligent Learning and Decisioning

ADI@LEI-3º ano - 2º Semestre
Filipe Gonçalves, Inês Alves, César Analide

Part II – February 2022

Contents

2

Terminology of AI

Intro to Knime

Workflow

Hands On

- Terminology of AI
- Intro to Knime
- Building a Workflow
- Hands On

Technologies of the next decade

3

TERMINOLOGY OF AI

Intro to Knime

Workflow

Hands On

 #1 Artificial Intelligence AI /Machine Learning / Deep Learning	 #2 Internet of Things IOT , IIOT, Sensors & Wearables	 #3 Mobile/Social Internet Advancements - Search/Social/ Messaging/Livestreams	 #4 Blockchain Distributed Ledger Systems, Apps, Infrastructure, Technologies Cryptocurrencies & DApps	 #5 Big Data + Predictive Analytics
 #6 Automation Information, Task, Process, Machine, Decision & Action	 #7 Robots Cons.,/Comm./Indus., Robots, Drones & Autonomous Vehicles	 #8 Immersive Media - #VR/ #AR/ #MR/ 360° Video?Gaming	 #9 Mobile Technologies Infrastructure, networks, standards, services & devices	 #10 Cloud Computing, SaaS, IaaS, PaaS & MESH Apps
 #11 3D Printing Additive Manufacturing & Rapid Prototyping	 #12 CX Customer Journey, Experience Commerce & Personalization	 #13 EnergyTech Efficiency, Energy Storage & Decentralized Grid	 #14 Cybersecurity Security, Intelligence Detection, Remediation & Adaptation	 #15 Voice Assistants Interfaces, Chatbots & Natural Language Processing
 #16 Nanotechnology Computing, Medicine, Machines + Smart Dust	 #17 Collaborative Tech. Crowd, Sharing, Workplace & Open Source Platforms & Tools	 #18 Health Tech. Advanced Genomics, Bionics & Health Care Tech.	 #19 Human-Computer Interaction Facial/Gesture Recognition, Biometrics, Gaze Tracking	 #20 Geo-spatial Tech. GIS, GPS, Mapping & Remote Sensing, Scanning, Navigation
 #21 Advanced Materials Composites, Alloys, Polymers, Biomimicry, Nanomanufacturing	 #22 New Touch Interfaces Touch Screens, Haptics, 3D Touch, Paper, Feedback & Exoskeletons	 #23 Wireless Power	 #24 Clean Tech. Bio-/Enviro-Materials + Solutions, Sustainability, Treatment & Efficiency	 #25 Quantum Computing + Exascale Computing
 #26 Smart Cities + Infrastructure & Transport	 #27 Edge/Computing + Fog Computing	 #28 Faster, Better Internet Broadband incl. Fiber, 5G, Li-Fi , LPN and LoRa	 #29 Proximity Tech Beacons, .RFID, Wi-Fi, Near-Field Communications & Geofencing	 #30 New Screens TVs, Digital Signage, OOH, MicroLEDS & Projections
<h2>THE 30 TECHNOLOGIES OF THE NEXT DECADE</h2>				
		Created by: Sean Moffitt @seanmoffitt , Managing Director, @Wikibrands		
				

Motivation

4

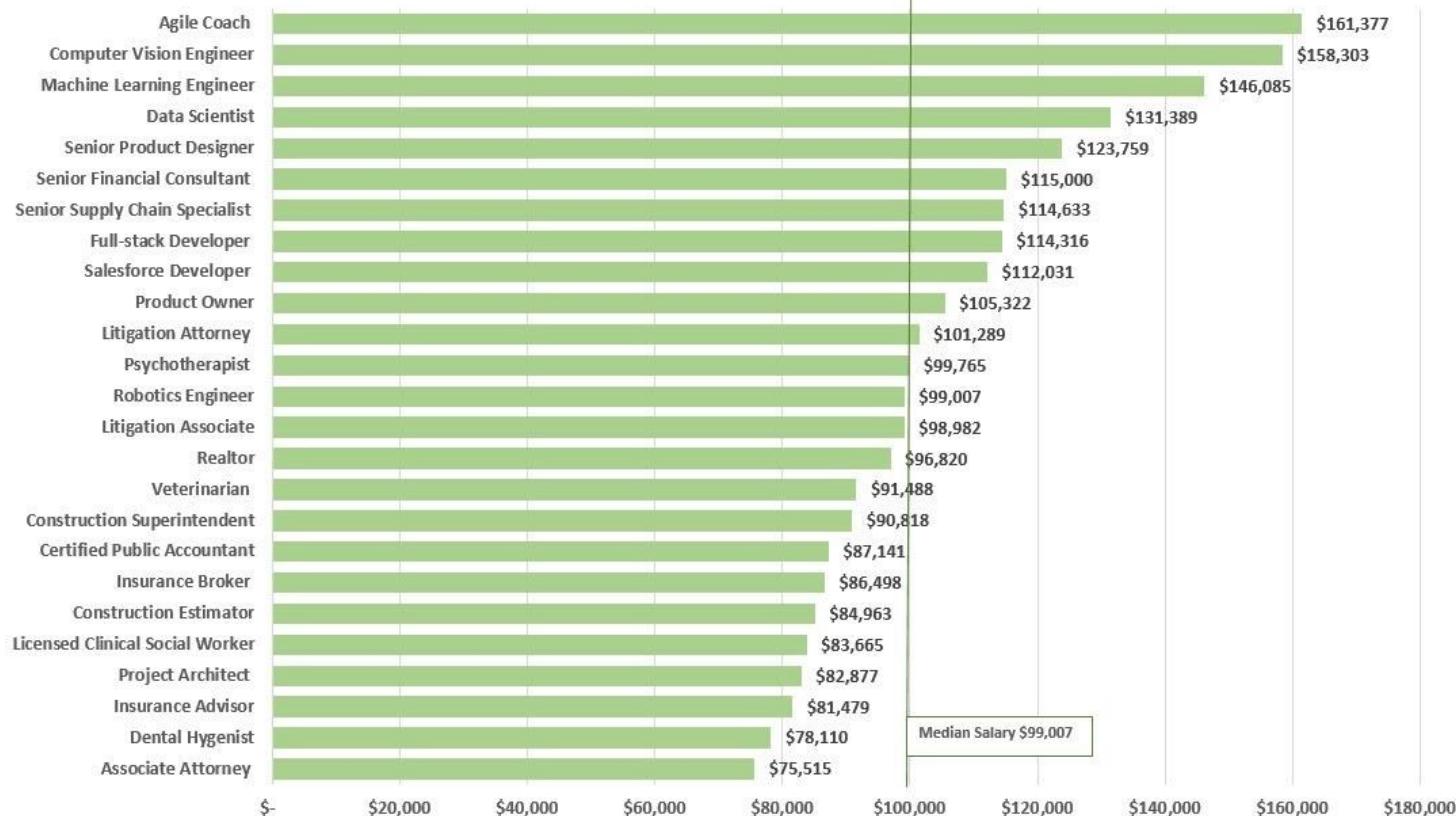
TERMINOLOGY OF AI

Intro to Knime

Workflow

Hands On

Indeed's Best Jobs In the U.S.
Average Base Salary, 2019



(<https://www.forbes.com/sites/louis columbus/2019/03/17/machine-learning-engineer-is-the-best-job-in-the-u-s-according-to-indeed/#2d134a177bb0>)

Motivation

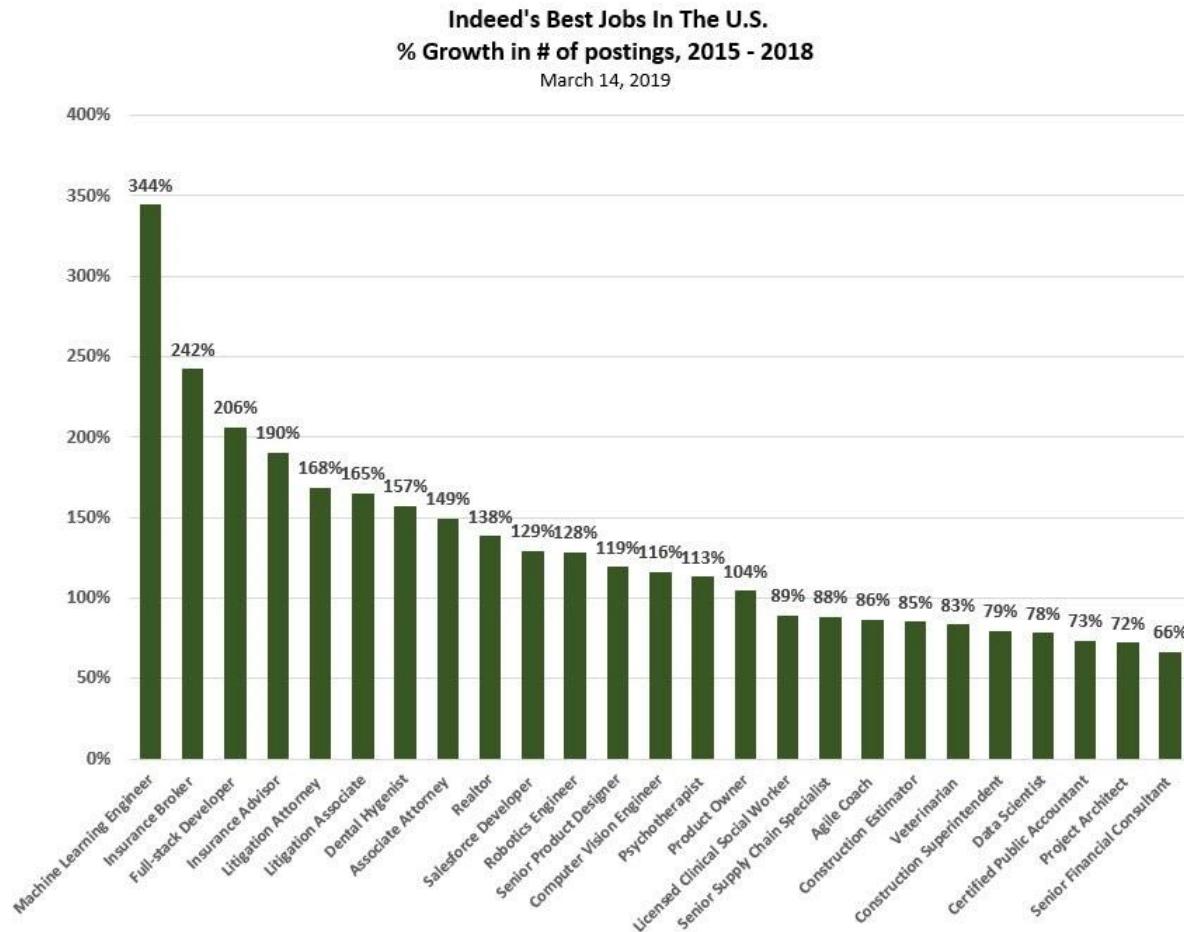
5

TERMINOLOGY OF AI

Intro to Knime

Workflow

Hands On



(<https://www.forbes.com/sites/louis columbus/2019/03/17/machine-learning-engineer-is-the-best-job-in-the-u-s-according-to-indeed/#2d134a177bb0>)

Terminology of AI

6

TERMINOLOGY OF AI

Intro to Knime

Workflow

Hands On

- Artificial Intelligence
- Machine Learning
- Deep Learning
- Data Science
- ...



Terminology of AI

7

TERMINOLOGY OF AI

Intro to Knime

Workflow

Hands On

Machine Learning vs **Data Science**

There is **no universal adherence!!!**

Terminology of AI

8

TERMINOLOGY OF AI

Intro to Knime

Workflow

Hands On

Machine Learning vs

- A -> B system
- PT-PT=**Aprendizagem Automática** (?)
- “*Field of study that gives computer the ability to learn without being explicitly programmed.*”

Arthur Samuel

Usually results in a **software artefact**

Data Science

- Analyse sets of data (datasets)
- PT-PT=**Ciência dos Dados** (?)
- Science of extracting knowledge and insights directly from data

Usually results in **slides and reports**

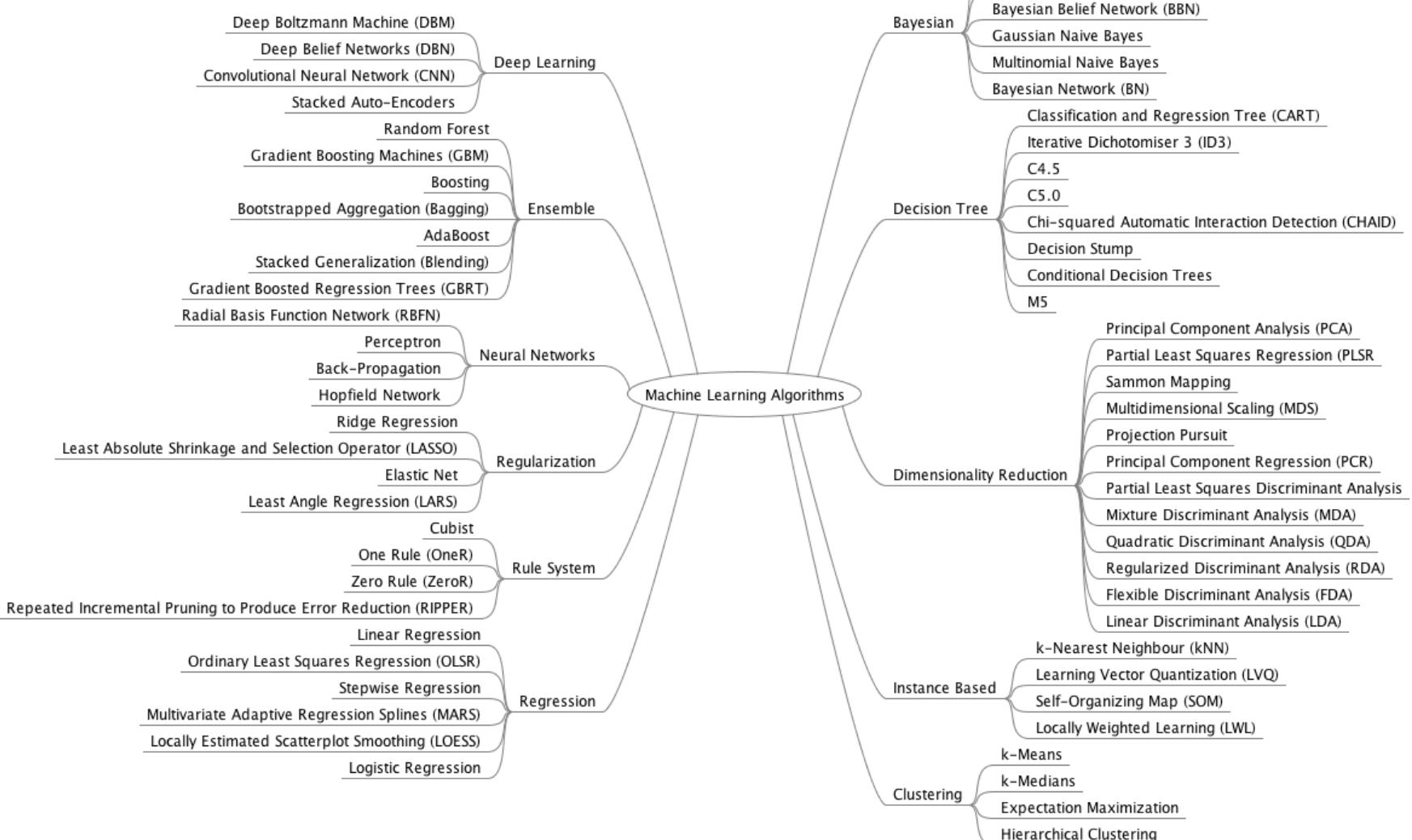
There is **no universal adherence!!!**

Terminology of AI

9

TERMINOLOGY OF AI

Intro to Knime





Open for Innovation ®
KNIME

0: Building a Simple Classifier

Simple Model Training for Classification
This workflow demonstrates how a simple classifier is built and applied to new data. It also illustrates the use of KNIME's linking capabilities, which allow interactive views to be connected within the same workflow.

```
graph LR; subgraph Workflow [Building a Simple Classifier]; subgraph ReadData [Read airline data]; ER[Excel Reader (XLS)] --> CM[Color Manager]; end; CM --> P[Partitioning]; P --> DL[Decision Tree Learner]; DL --> PMML[PMML Writer]; P --> DP[Decision Tree Predictor]; DP --> S[Scorer]; subgraph Predict [Apply decision tree model]; DP --> BC[Bar Chart (JavaScript)]; end; subgraph Output [PMML Writer, Scorer, Confusion matrix accuracy measurements]; PMML; S; end; subgraph Visualizations [Interactive Views]; SP[Scatter Plot (JavaScript)]; end;
```

Workflow Components:

- Excel Reader (XLS)**: Read airline data
- Color Manager**
- Partitioning**
- Decision Tree Learner**: Train model to predict departure delays
- PMML Writer**
- Decision Tree Predictor**: Apply decision tree model
- Scorer**: Confusion matrix accuracy measurements
- Bar Chart (JavaScript)**
- Scatter Plot (JavaScript)**

Node Repository Categories:

- IO
- Manipulation
- Views
- Analytics
- Mining
- Statistics
- Distance Calculation
- PMML
- Database
- Other Data Types

Workflow Labels:

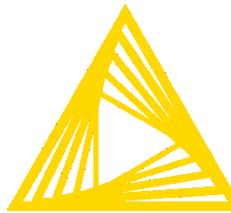
- Read airline data
- Statistics
- Train model to predict departure delays
- Apply decision tree model

Scatter Plot (Sepal Length vs Sepal Width):

Sepal Length	Sepal Width	Species
4.3	3.0	Iris-setosa
4.4	3.4	Iris-setosa
4.5	3.0	Iris-setosa
4.5	2.3	Iris-setosa
4.6	3.6	Iris-setosa
4.7	3.1	Iris-setosa
4.8	3.4	Iris-setosa
4.9	2.5	Iris-setosa
5.0	3.4	Iris-setosa
5.1	3.7	Iris-setosa
5.2	3.5	Iris-setosa
5.3	3.0	Iris-setosa
5.4	3.9	Iris-setosa
5.5	3.8	Iris-setosa
5.6	3.0	Iris-setosa
5.7	3.2	Iris-setosa
5.8	3.0	Iris-setosa
5.9	3.2	Iris-setosa
6.0	3.0	Iris-setosa
6.1	3.0	Iris-setosa
6.2	3.4	Iris-setosa
6.3	3.3	Iris-setosa
6.4	3.0	Iris-setosa
6.5	3.0	Iris-setosa
6.6	3.4	Iris-setosa
6.7	3.0	Iris-setosa
6.8	3.0	Iris-setosa
6.9	3.1	Iris-setosa
7.0	3.0	Iris-setosa
7.1	3.0	Iris-setosa
7.2	3.0	Iris-setosa
7.3	3.0	Iris-setosa
7.4	3.0	Iris-setosa
7.5	3.0	Iris-setosa
7.6	3.0	Iris-setosa
7.7	3.0	Iris-setosa
7.8	3.0	Iris-setosa
7.9	3.0	Iris-setosa
8.0	3.0	Iris-setosa
4.3	1.3	Iris-versicolor
4.4	1.4	Iris-versicolor
4.5	1.3	Iris-versicolor
4.5	1.5	Iris-versicolor
4.6	1.4	Iris-versicolor
4.7	1.4	Iris-versicolor
4.8	1.3	Iris-versicolor
4.9	1.5	Iris-versicolor
5.0	1.3	Iris-versicolor
5.1	1.3	Iris-versicolor
5.2	1.4	Iris-versicolor
5.3	1.3	Iris-versicolor
5.4	1.5	Iris-versicolor
5.5	1.4	Iris-versicolor
5.6	1.3	Iris-versicolor
5.7	1.5	Iris-versicolor
5.8	1.3	Iris-versicolor
5.9	1.4	Iris-versicolor
6.0	1.3	Iris-versicolor
6.1	1.3	Iris-versicolor
6.2	1.3	Iris-versicolor
6.3	1.3	Iris-versicolor
6.4	1.3	Iris-versicolor
6.5	1.3	Iris-versicolor
6.6	1.3	Iris-versicolor
6.7	1.3	Iris-versicolor
6.8	1.3	Iris-versicolor
6.9	1.3	Iris-versicolor
7.0	1.3	Iris-versicolor
7.1	1.3	Iris-versicolor
7.2	1.3	Iris-versicolor
7.3	1.3	Iris-versicolor
7.4	1.3	Iris-versicolor
7.5	1.3	Iris-versicolor
7.6	1.3	Iris-versicolor
7.7	1.3	Iris-versicolor
7.8	1.3	Iris-versicolor
7.9	1.3	Iris-versicolor
8.0	1.3	Iris-versicolor
4.3	5.1	Iris-virginica
4.4	4.9	Iris-virginica
4.5	4.9	Iris-virginica
4.5	4.9	Iris-virginica
4.6	4.9	Iris-virginica
4.7	4.9	Iris-virginica
4.8	4.9	Iris-virginica
4.9	4.9	Iris-virginica
5.0	4.9	Iris-virginica
5.1	4.9	Iris-virginica
5.2	4.9	Iris-virginica
5.3	4.9	Iris-virginica
5.4	4.9	Iris-virginica
5.5	4.9	Iris-virginica
5.6	4.9	Iris-virginica
5.7	4.9	Iris-virginica
5.8	4.9	Iris-virginica
5.9	4.9	Iris-virginica
6.0	4.9	Iris-virginica
6.1	4.9	Iris-virginica
6.2	4.9	Iris-virginica
6.3	4.9	Iris-virginica
6.4	4.9	Iris-virginica
6.5	4.9	Iris-virginica
6.6	4.9	Iris-virginica
6.7	4.9	Iris-virginica
6.8	4.9	Iris-virginica
6.9	4.9	Iris-virginica
7.0	4.9	Iris-virginica
7.1	4.9	Iris-virginica
7.2	4.9	Iris-virginica
7.3	4.9	Iris-virginica
7.4	4.9	Iris-virginica
7.5	4.9	Iris-virginica
7.6	4.9	Iris-virginica
7.7	4.9	Iris-virginica
7.8	4.9	Iris-virginica
7.9	4.9	Iris-virginica
8.0	4.9	Iris-virginica

Bar Chart:

Species	sepal length	sepal width	petal length	petal width
Iris-setosa	~5.0	~3.0	~1.0	~0.2
Iris-versicolor	~7.0	~3.0	~4.5	~1.5
Iris-virginica	~7.9	~3.0	~5.0	~1.5



Open for Innovation ®

KNIME

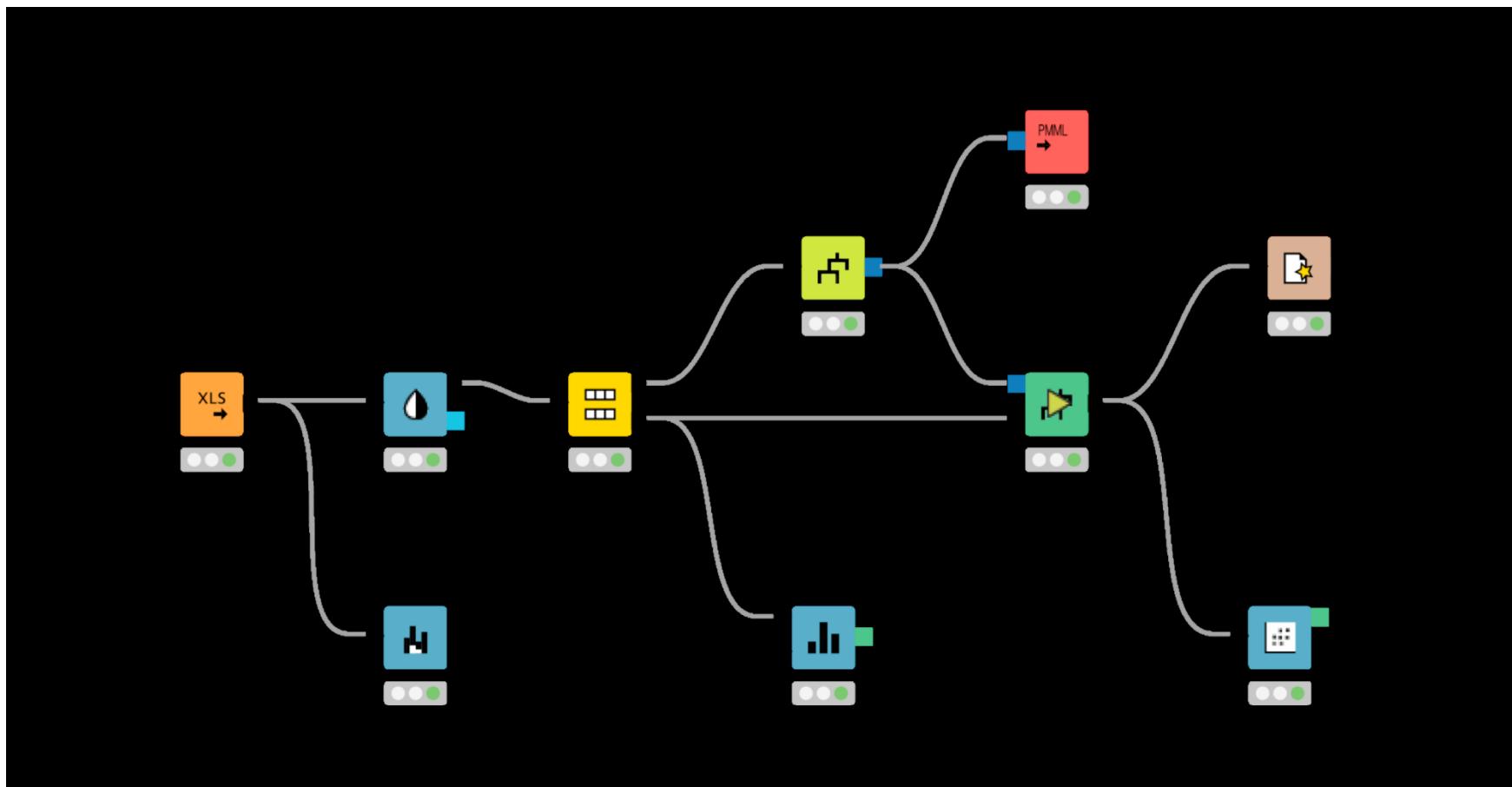
11

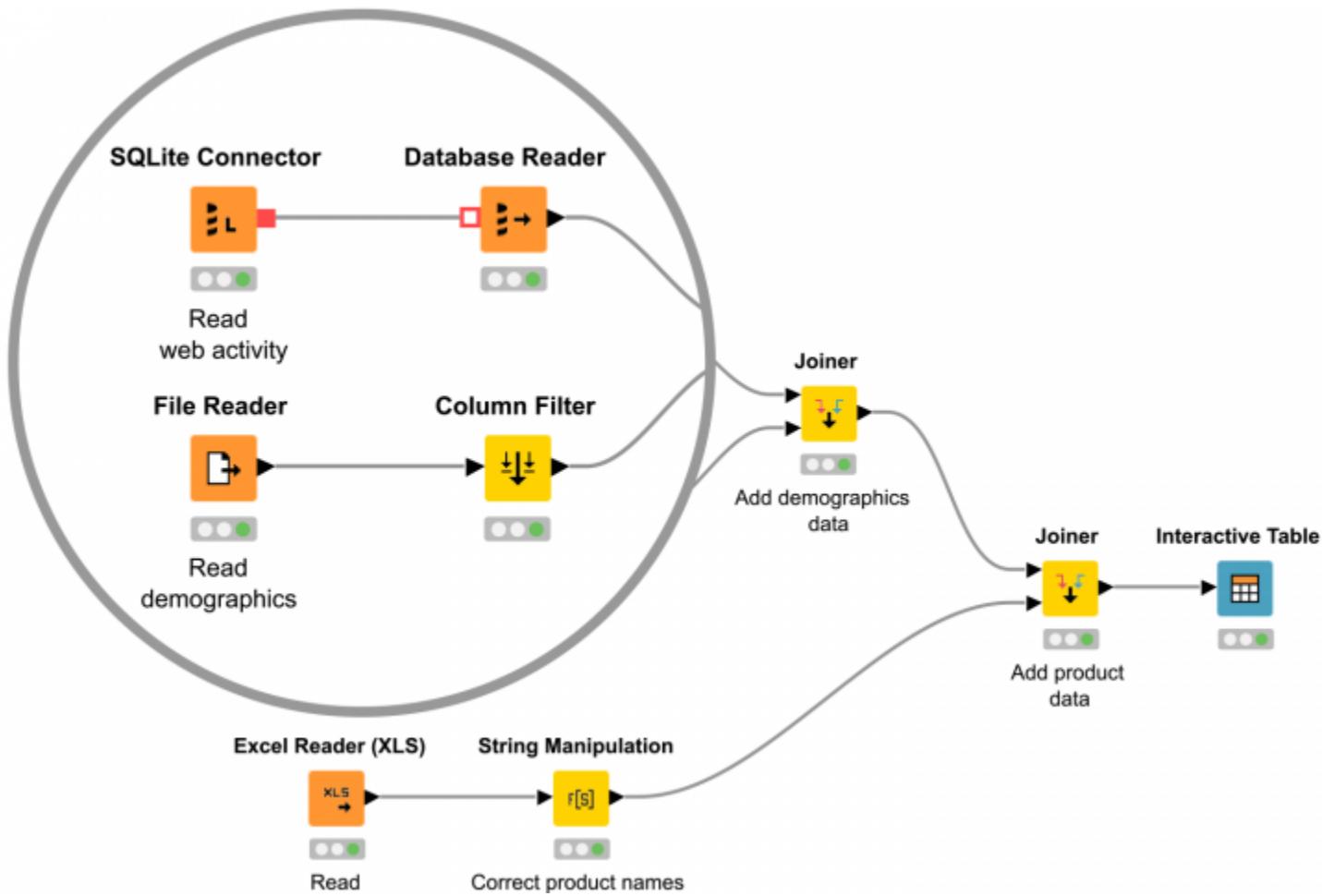
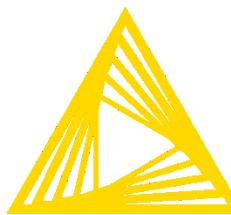
Terminology of AI

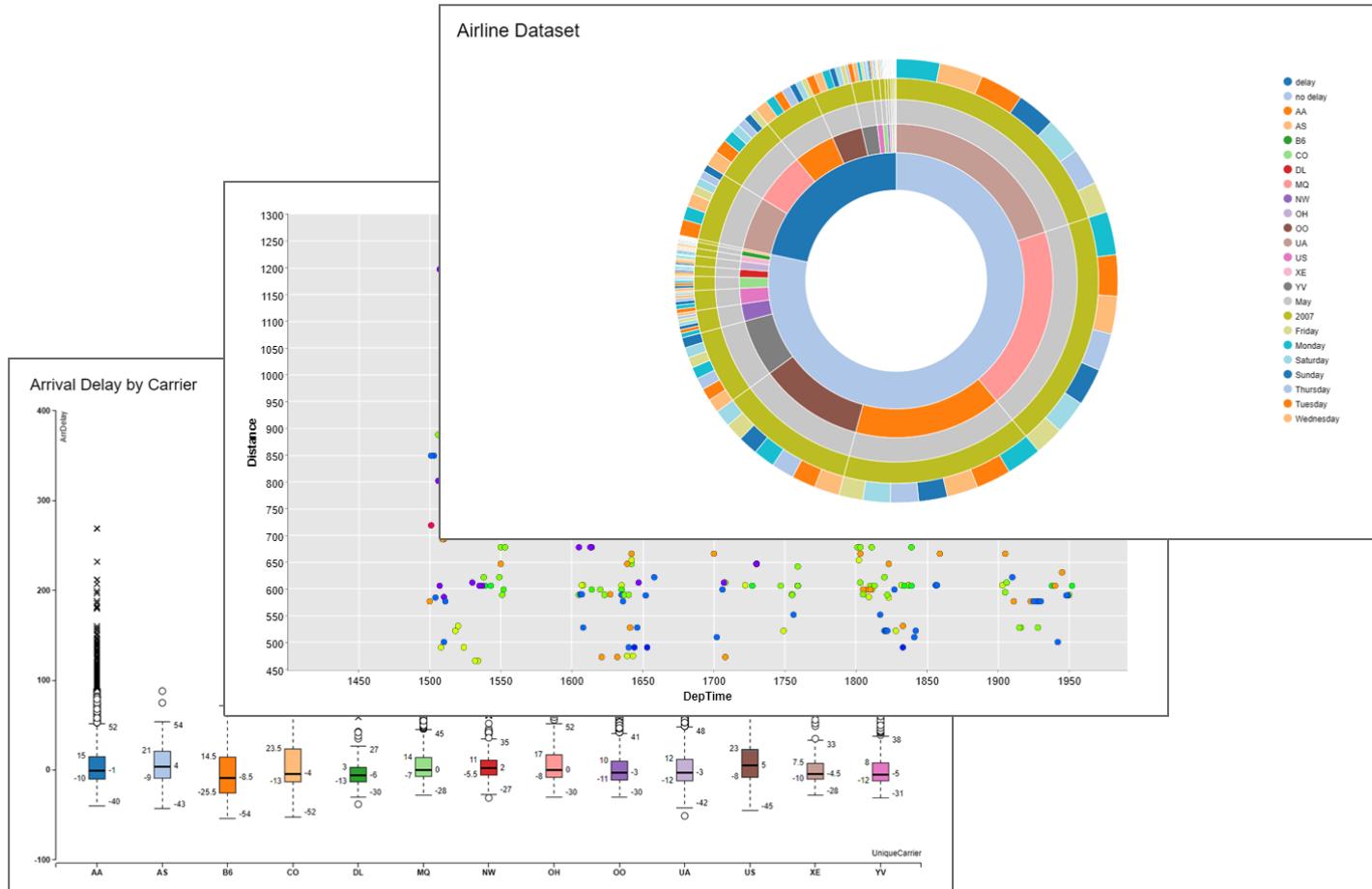
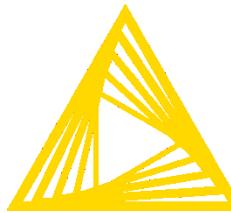
INTRO TO KNIME

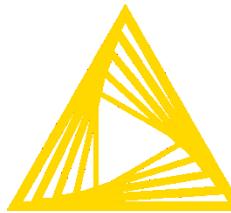
Workflow

Hands On



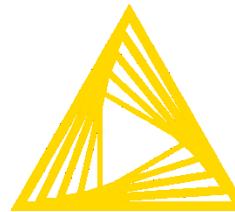






- **KNIME Analytics Platform** is one of the most popular **open source platforms** used to automate the data science process
- Released in 2006, it is **free** and **open-source**, continuously integrating new developments
- **Additional features** and functionality can be added via **KNIME extensions**
- A **Gartner's leader** for Data Science and Machine Learning Platforms for the last six years





KNIME Analytics Platform

File Edit View Help

KNIME Explorer

- > My-KNIME-Hub (hub.knime.com)
- > EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
- > Example Workflows

Workflow Coach

Recommended Nodes	Community
File Reader	24%
CSV Reader	18%
Excel Reader (XLS)	17%
Table Creator	12%

Node Repository

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

Welcome to KNIME Analytics Platform

Welcome

Looks like you're using KNIME for the first time...

Get started with this example

Open workflow

Looking for more examples? Visit the KNIME Hub

KNIME Hub

Sign up for introductory emails

These messages will get you up and running as quickly as possible.

Sign up

Console Outline

An outline is not available.

(Quick) Hands On

16

Terminology of AI

INTRO TO KNIME

Workflow

Hands On



(Quick) Hands On

17

Terminology of AI

INTRO TO KNIME

Workflow

Hands On

KNIME Analytics Platform

File Edit View Node Help

100% Quick Access

KNIME Explorer : KNIME_Aula_Exercicio

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

Workflow Coach

Recommended Nodes

- File Reader
- CSV Reader
- Excel Reader (XLS)
- Table Creator
- Database Reader (legacy)
- Table Reader
- List Files
- Database Connection T...

Node Repository

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

Console Outline

Search workflows, nodes, and more...

1. Download Knime
2. Install it!
3. Try it!

Nodes and Workflows

18

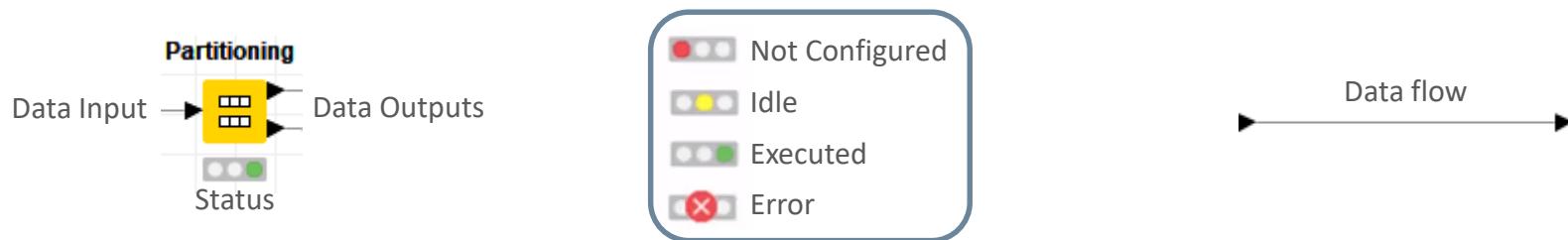
Terminology of AI

INTRO TO KNIME

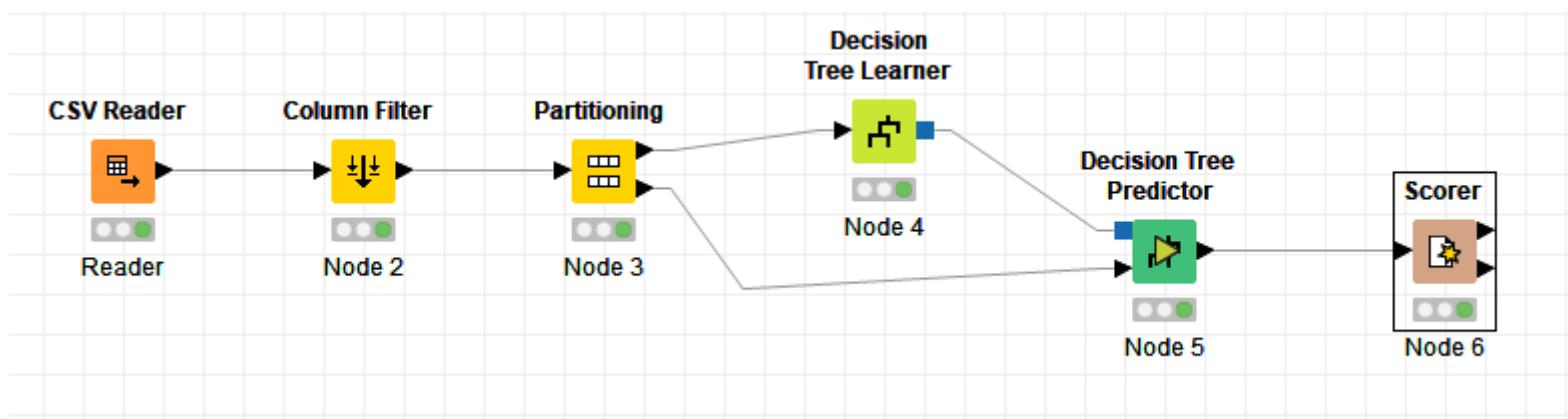
Workflow

Hands On

Nodes



Workflow



KNIME Extensions

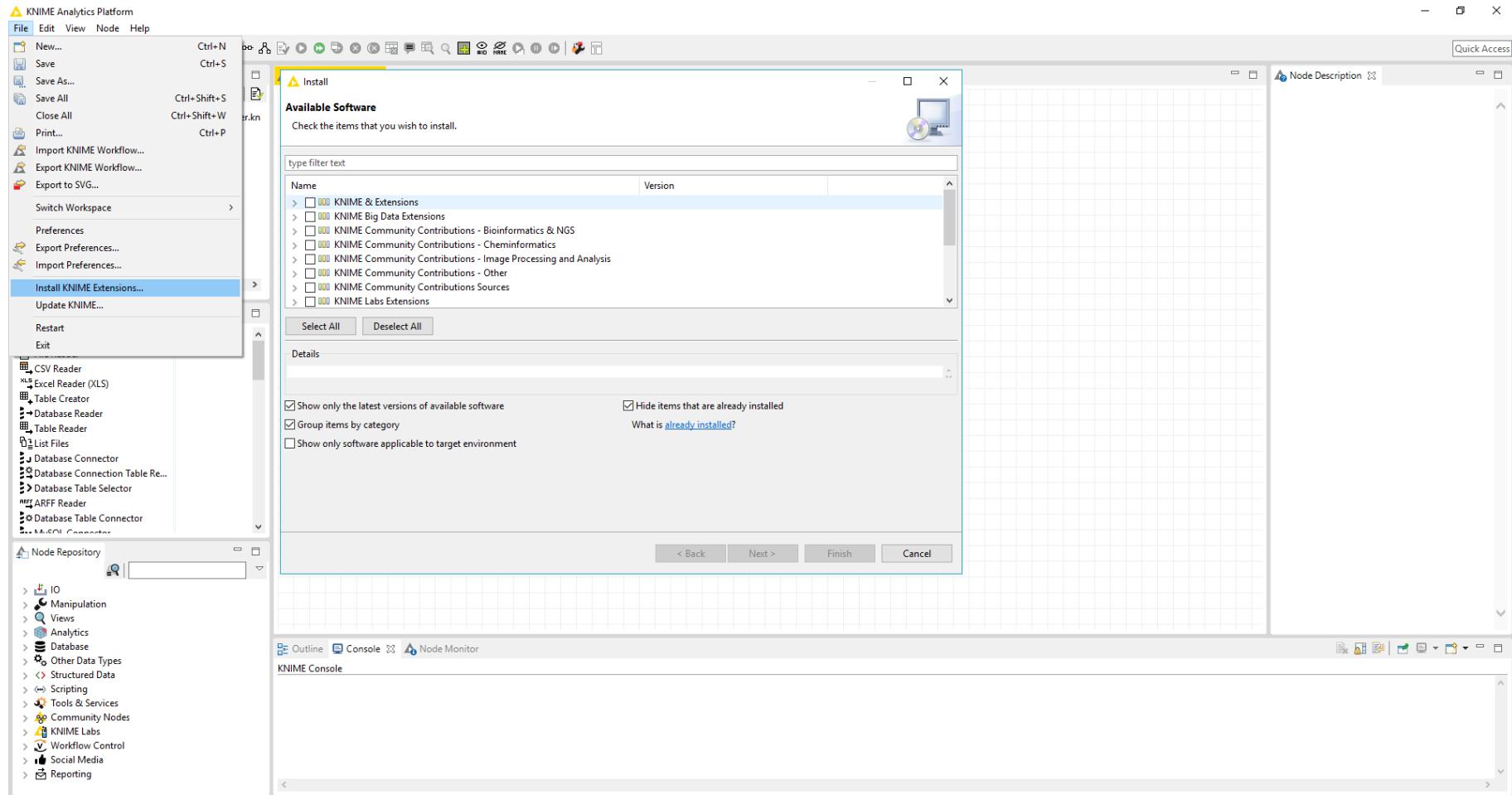
19

Terminology of AI

INTRO TO KNIME

Workflow

Hands On



Main Views

20

Terminology of AI

INTRO TO KNIME

Workflow

Hands On

The screenshot displays the KNIME Analytics Platform interface with several labeled components:

- Knime Explorer**: Located in the top-left corner of the main workspace.
- Workflow Coach**: A sidebar on the left containing a list of recommended nodes and their usage statistics.
- Node Repository**: A sidebar on the left listing categories such as IO, Manipulation, Views, Analytics, Database, and others.
- KNIME_Aula_1**: The active project name in the top-left corner of the workspace.
- Workflow Building Area**: The central area where a workflow is being built. It shows a sequence of nodes: CSV Reader → Column Filter → Partitioning → Decision Tree Learner (Node 4) → Decision Tree Predictor (Node 5) → Scorer (Node 6). The 'Partitioning' node has a self-loop arrow.
- Console Outline Others**: A box at the bottom center of the workspace.
- Node Description**: A detailed description of the 'Decision Tree Learner' node, including its function and parameters. It also includes a 'Dialog Options' section and a 'Node Description' box.
- KNIME Console**: A terminal-like window at the bottom of the workspace.

Node Description

21

Terminology of AI

INTRO TO KNIME

Workflow

Hands On

The screenshot shows the KNIME Node Description interface for the "Scorer" node. At the top left, there is a small icon of the node followed by its name "Scorer". Below this, there is a diagram showing the node's inputs and outputs: one input port and two output ports. The first output port is labeled "Confusion Matrix". The main area is titled "Scorer" and contains the following content:

Description

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.

Data Table Structure View

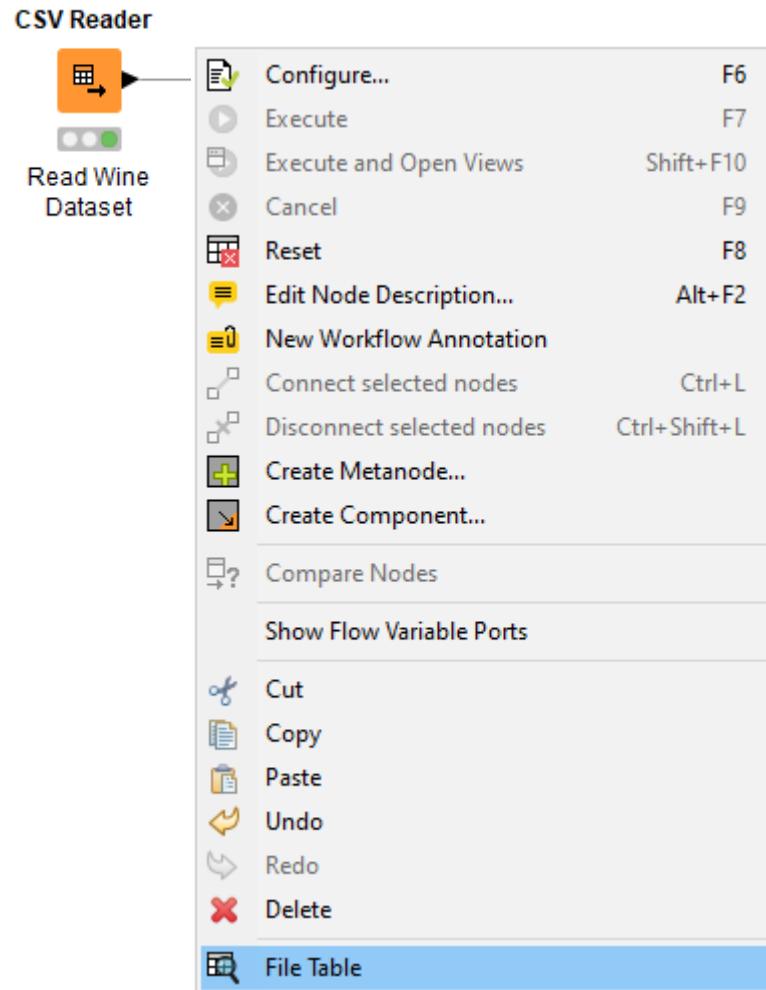
22

Terminology of AI

INTRO TO KNIME

Workflow

Hands On



Data Table Structure View

23

Terminology of AI

INTRO TO KNIME

Workflow

Hands On

Column Headers

Data Type (Double)

Data Type (String)

Row ID

Data Cells

The screenshot shows a data table from the 'winequality-red.csv' file. The columns are labeled: Row ID, fixed acids, volatile acids, citric acid, residual sugar, chlorides, Sort Descending, Sort Ascending, No Sorting, alcohol, and quality. A context menu is open over the 'Sort Descending' button, listing options: Standard Double, Percentage, Full Precision, Gray Scale, Bars, Standard Complex Number, and Default. The 'quality' column is highlighted with a yellow border. Arrows point from the labels to specific parts of the interface: 'Column Headers' to the first row, 'Data Type (Double)' to the 'alcohol' column, 'Data Type (String)' to the 'quality' column, 'Row ID' to the 'Row ID' column, and 'Data Cells' to the data values in the table.

Row ID	fixed acids	volatile acids	citric acid	residual sugar	chlorides	Sort Descending	Sort Ascending	No Sorting	alcohol	quality
Row0	7.4	0.7	0	1.9	0.076				=5	=5
Row1	7.8	0.88	0	2.6	0.098				=5	=5
Row2	7.8	0.76	0.04	2.3	0.092				=5	=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	3.36	0.57
Row9	7.5	0.5	0.36	6.1	0.071	17	102	0.998	3.35	0.8
Row10	6.7	0.58	0.08	1.8	0.097	15	65	0.996	3.28	0.54
Row11	7.5	0.5	0.36	6.1	0.071	17	102	0.998	3.35	0.8
Row12	5.6	0.615	0	1.6	0.089	16	59	0.994	3.58	0.52
Row13	7.8	0.61	0.29	1.6	0.114	9	29	0.997	3.26	1.56
Row14	8.9	0.62	0.18	3.8	0.176	52	145	0.999	3.16	0.88
Row15	8.9	0.62	0.19	3.9	0.17	51	148	0.999	3.17	0.93
Row16	8.5	0.28	0.56	1.8	0.092	35	103	0.997	3.3	0.75
Row17	8.1	0.56	0.28	1.7	0.368	16	56	0.997	3.11	1.28
Row18	7.4	0.59	0.08	4.4	0.086	6	29	0.997	3.38	0.5
Row19	7.9	0.32	0.51	1.8	0.341	17	56	0.997	3.04	1.08
Row20	8.9	0.22	0.48	1.8	0.077	29	60	0.997	3.39	0.53
Row21	7.6	0.39	0.31	2.3	0.082	23	71	0.998	3.52	0.65
Row22	7.9	0.43	0.21	1.6	0.106	10	37	0.997	3.17	0.91
Row23	8.5	0.49	0.11	2.3	0.084	9	67	0.997	3.17	0.53
Row24	6.9	0.4	0.14	2.4	0.085	21	40	0.997	3.43	0.63
Row25	6.3	0.39	0.16	1.4	0.08	11	23	0.996	3.34	0.56
Row26	7.6	0.4	0.24	1.8	0.08	4	11	0.996	3.29	0.5

Log View

24

Terminology of AI

INTRO TO KNIME

Workflow

Hands On

The screenshot shows the KNIME Analytics Platform interface with the 'Log View' window open. The log window displays a list of DEBUG-level log entries from a workflow named 'KNIME_Aula_1'. The log entries show various stages of the workflow execution, including node edits and model training.

```
2018-09-26 11:50:26,149 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:3(1) -> 0:4( 1)]
2018-09-26 11:50:26,149 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 11:50:26,150 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:4(1) -> 0:5( 1)]
2018-09-26 11:50:26,150 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 11:50:26,150 : DEBUG : main : WorkflowRootEditPart : : : part: NodeContainerEditPart( Decision Tree Predictor 0:5 (EXECUTED) )
2018-09-26 11:50:26,151 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:4(1) -> 0:5( 1)]
2018-09-26 11:50:26,151 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 11:50:26,151 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:3(2) -> 0:5( 2)]
2018-09-26 11:50:26,151 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 11:50:26,152 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:5(1) -> 0:6( 1)]
2018-09-26 11:50:26,152 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 11:50:26,152 : DEBUG : main : WorkflowRootEditPart : : : part: NodeContainerEditPart( Scorer 0:6 (EXECUTED) )
2018-09-26 11:50:26,153 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:5(1) -> 0:6( 1)]
2018-09-26 11:50:26,153 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 12:02:38,357 : DEBUG : main : NodeContainer : : Setting dirty flag on KNIME_Aula_1 0
2018-09-26 12:02:38,359 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:1(1) -> 0:2( 1)]
2018-09-26 12:02:38,359 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 12:02:38,360 : DEBUG : KNIME-Workflow-Notifier : WorkflowEditor : : Workflow event triggered: WorkflowEvent [type=WORKFLOW_DIRTY;node=@old=null;new=null;t
2018-09-26 12:02:38,364 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:2(1) -> 0:3( 1)]
2018-09-26 12:02:38,364 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 12:02:38,364 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:3(2) -> 0:5( 2)]
2018-09-26 12:02:38,364 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: bendpoints: 369, 197,
2018-09-26 12:02:38,364 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:3(1) -> 0:4( 1)]
2018-09-26 12:02:38,364 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 12:02:38,364 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:4(1) -> 0:5( 1)]
2018-09-26 12:02:38,364 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 12:02:38,364 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:5(1) -> 0:6( 1)]
2018-09-26 12:02:40,973 : DEBUG : main : NodeContainerEditPart : : : Score@0:0 (EXCLUDED)
2018-09-26 12:02:40,973 : DEBUG : main : NodeContainerEditPart : : : Decision Tree Predictor 0:5 (EXECUTED)
2018-09-26 12:02:42,964 : DEBUG : main : NodeContainerEditPart : : : Decision Tree Predictor 0:5 (EXECUTED)
2018-09-26 12:14:04,445 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:1(1) -> 0:2( 1)]
2018-09-26 12:14:04,445 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 12:14:04,446 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:2(1) -> 0:3( 1)]
2018-09-26 12:14:04,446 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 12:14:04,446 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:3(2) -> 0:5( 2)]
2018-09-26 12:14:04,446 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: bendpoints: 369, 197,
2018-09-26 12:14:04,446 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:3(1) -> 0:4( 1)]
2018-09-26 12:14:04,446 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 12:14:04,446 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:4(1) -> 0:5( 1)]
2018-09-26 12:14:04,446 : DEBUG : main : ConnectionContainerEditPart : : : modelling info: null
2018-09-26 12:14:04,446 : DEBUG : main : ConnectionContainerEditPart : : : refreshing visuals for: STD[0:5(1) -> 0:6( 1)]
2018-09-26 12:14:12,653 : DEBUG : main : NodeContainerEditPart : : : Partitioning 0:3 (EXECUTED)
```

The interface includes a navigation bar with File, Edit, View, Help, and various tool icons. The left sidebar contains the KNIME Explorer, Node Repository, Outline, Workflow Coach, Other..., Reset Perspective..., Quick Node Insertion..., and Open KNIME log (which is currently selected). The bottom navigation bar includes Outline, Console, Node Monitor, and tabs for Node, State, Flow Variables, and Port 0.

The right panel displays the 'Partitioning' dialog, which provides options for handling input table partitions. It includes sections for Absolute (specifying absolute row counts), Relative (specifying a percentage of rows), Take from top (putting top rows into first output), Linear sampling (downsampling sorted columns), Draw randomly (random sampling), and Stratified sampling (optional stratification).

Views' Customization

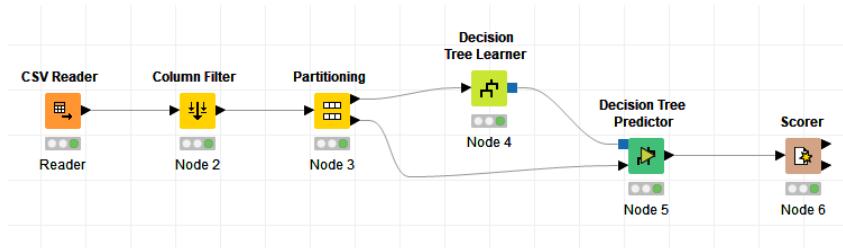
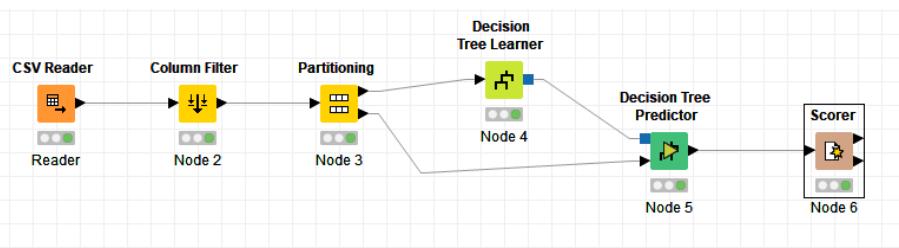
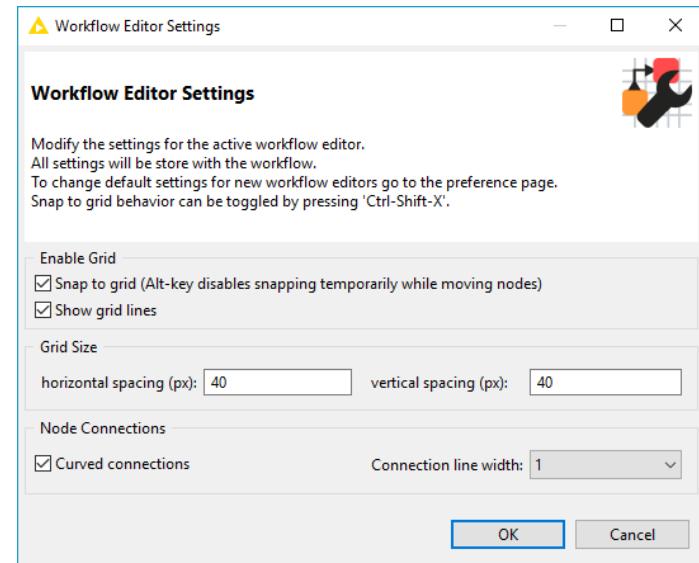
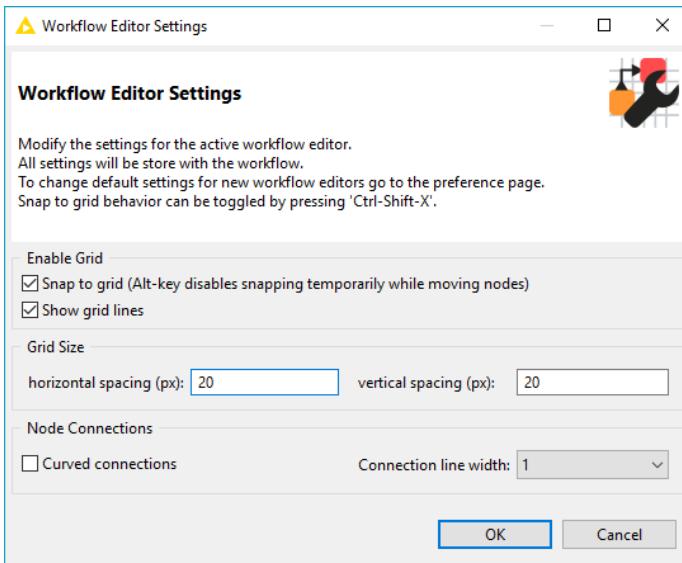
25

Terminology of AI

INTRO TO KNIME

Workflow

Hands On



Building a Simple Workflow

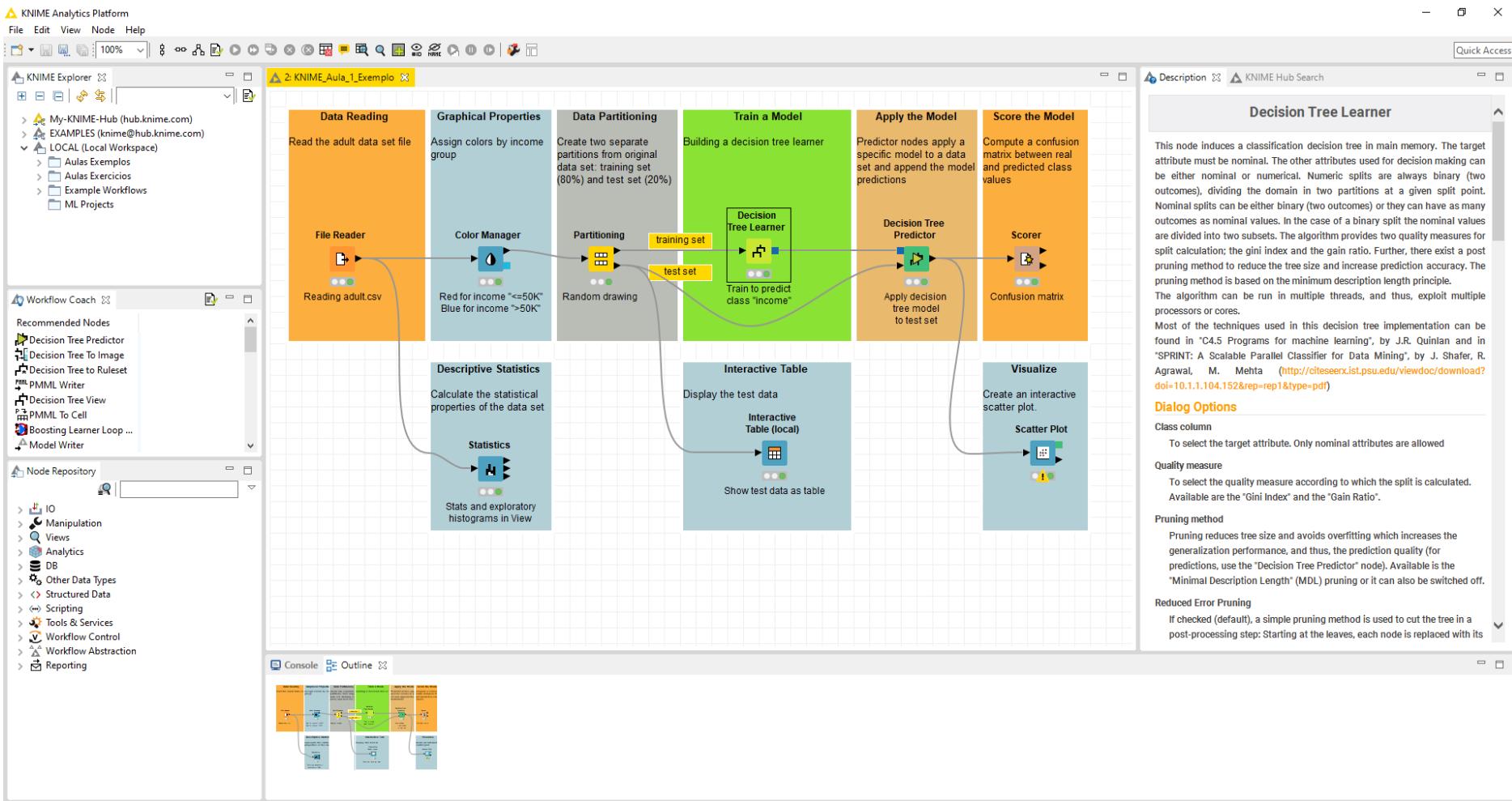
26

Terminology of AI

Intro to Knime

WORKFLOW

Hands On



Node Context Options

Data Loader

27

Terminology of AI

Intro to Knime

WORKFLOW

Hands On

KNIME Analytics Platform

File Edit View Node Help

100% 8 100% 100% 100%

Quick Access

KNIME Explorer 2: KNIME_Aula_1_Exemplo Description KNIME Hub Search

My-KNIME-Hub (hub.knime.com) EXAMPLES (knime@hub.knime.com)

LOCAL (Local Workspace)

- Aulas Exemplos
- Aulas Ejercicios
- 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

Configure... 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

F6 F7 Shift+F10 F9 F8 Alt+F2 Ctrl+L Ctrl+Shift+L

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 Configuration dialog

Color Manager

Partitioning Configuration dialog

Decision Tree Learner Configuration dialog

Decision Tree Predictor Configuration dialog

Scorer Configuration dialog

Confusion matrix

Interactive Table Configuration dialog

Scatter Plot Configuration dialog

File Reader File Reader node configuration

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 checkmarks 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.

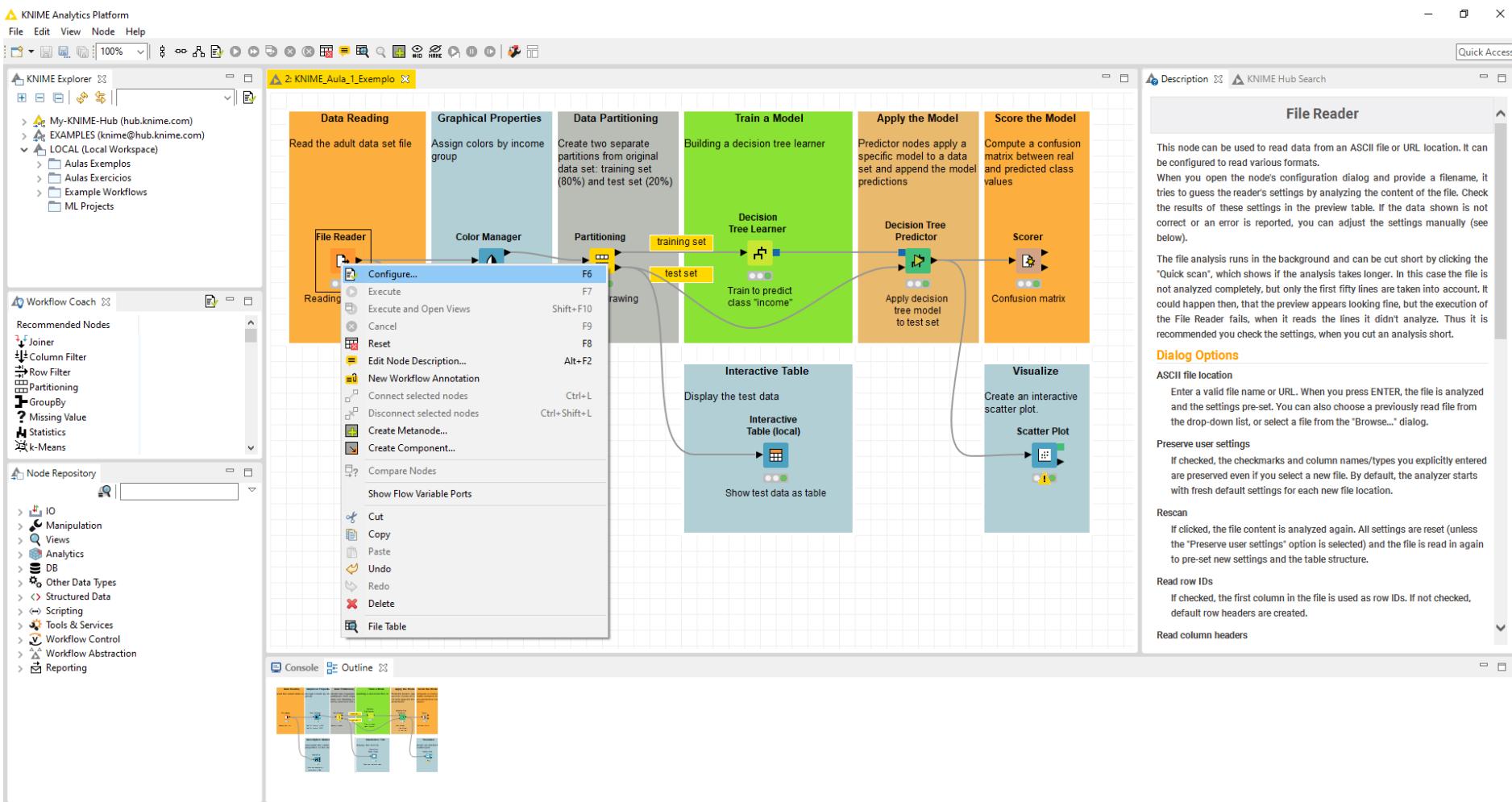
Rescan If clicked, the file content is analyzed again. All settings are reset (unless the "Preserve user settings" option is selected) and the file is read in again to pre-set new settings and the table structure.

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

Console

Outline



Node Context Options

Data Loader

28

Terminology of AI

Intro to Knime

WORKFLOW

Hands On

The screenshot shows the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' panel lists workspaces and projects. The 'Workflow Coach' panel provides recommendations for nodes. The 'Node Repository' panel lists various node categories. The main workspace contains a workflow titled '2-KNIME_Aula_1_Exemplo'. This workflow starts with a 'File Reader' node connected to a 'Color Manager' node. The 'Color Manager' node then connects to two parallel 'Descriptive Statistics' nodes. One path from the 'Color Manager' leads to a 'Statistics' node, which then connects to a 'Scorer' node. The other path leads to a 'Confusion matrix' node. A 'Dialog - 2:1 - File Reader (Reading adult.csv)' dialog box is open, showing configuration options for reading the 'adult.csv' file. The right side of the screen displays the 'File Reader' node's documentation, including its purpose, configuration options like 'File location' and 'Preserve user settings', and 'Dialog Options'.

Node Context Options

Model Learner

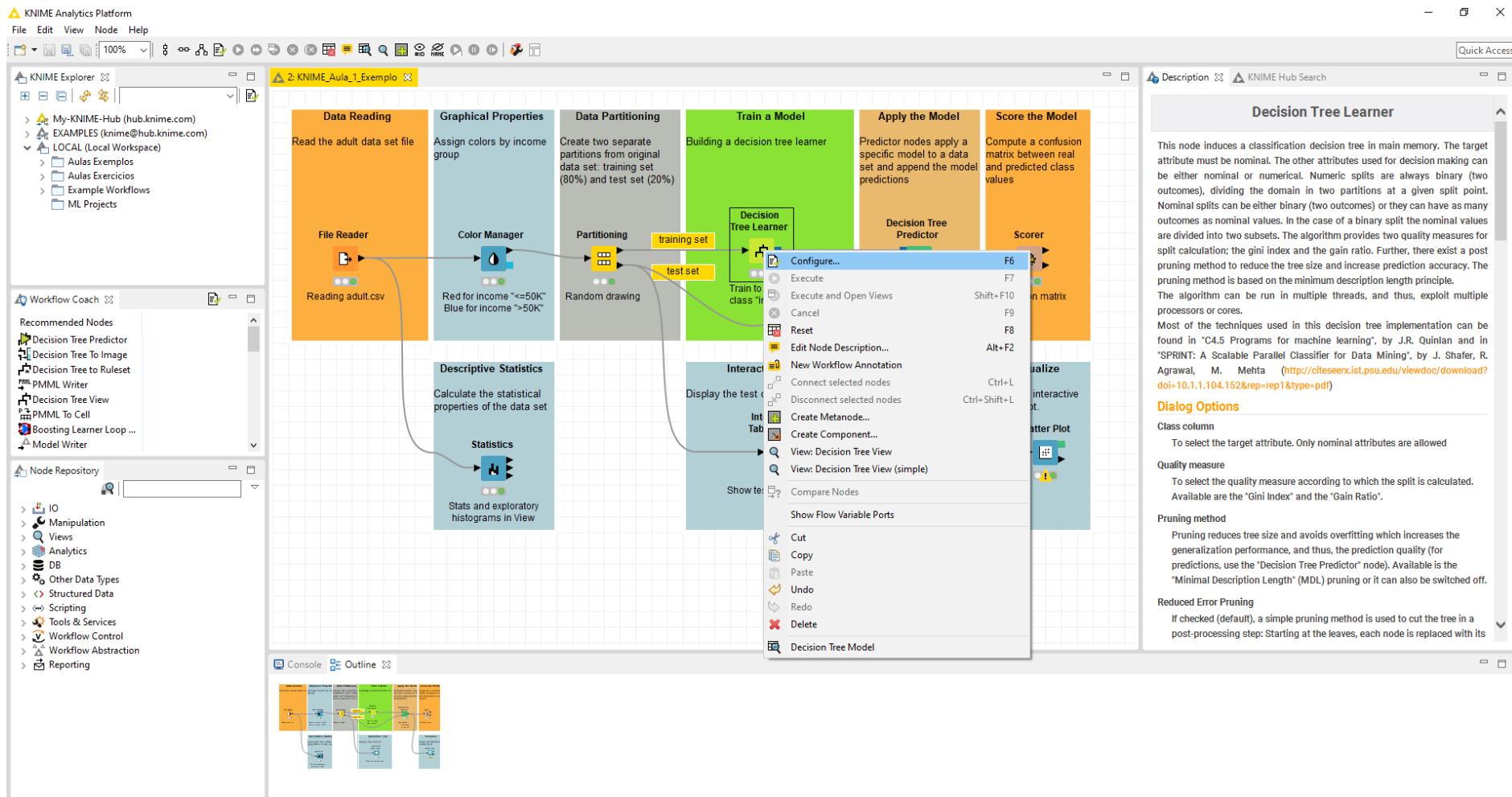
29

Terminology of AI

Intro to Knime

WORKFLOW

Hands On



Node Context Options

Model Learner

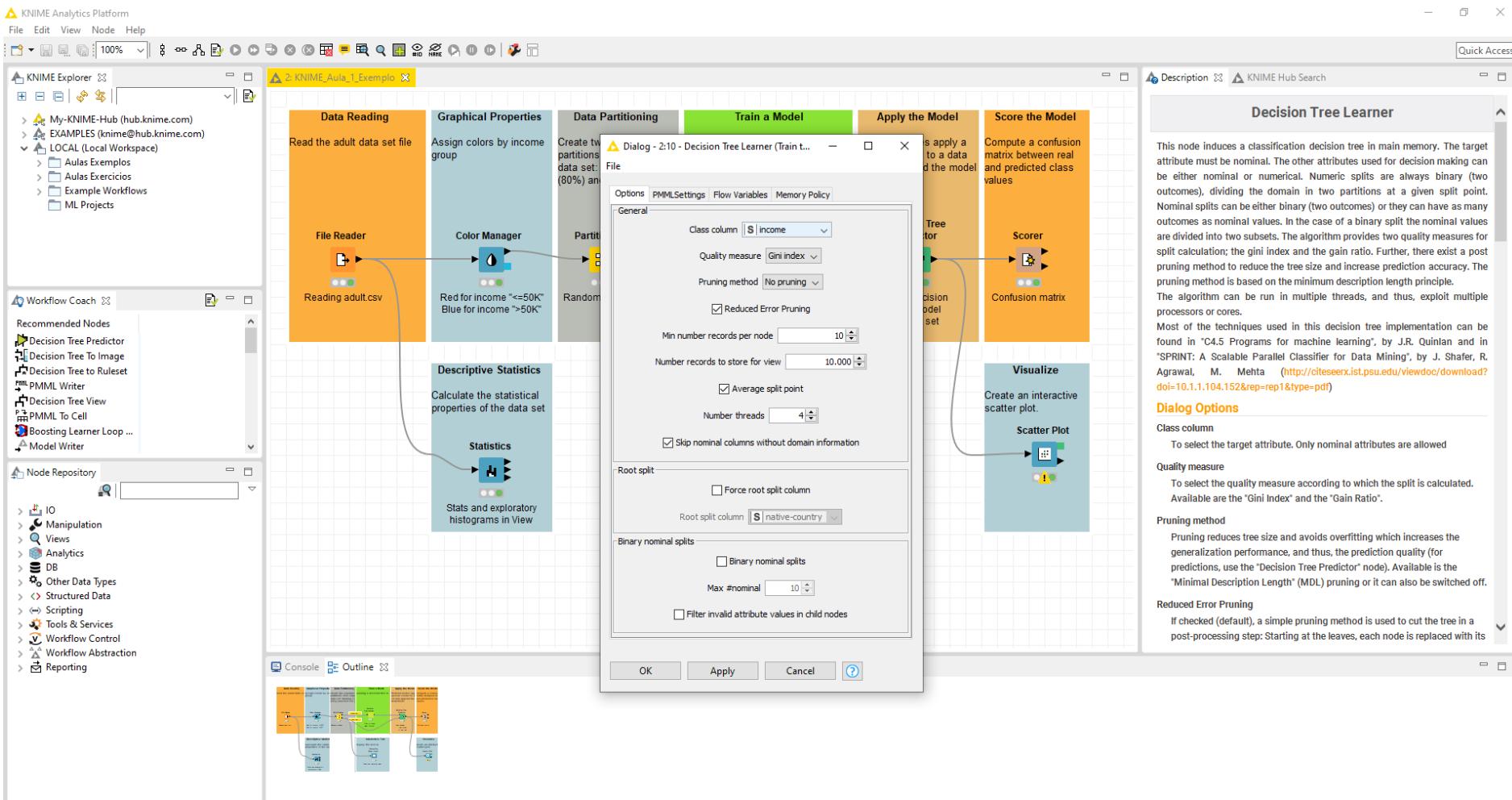
30

Terminology of AI

Intro to Knime

WORKFLOW

Hands On



Node Context Options

Model Learner

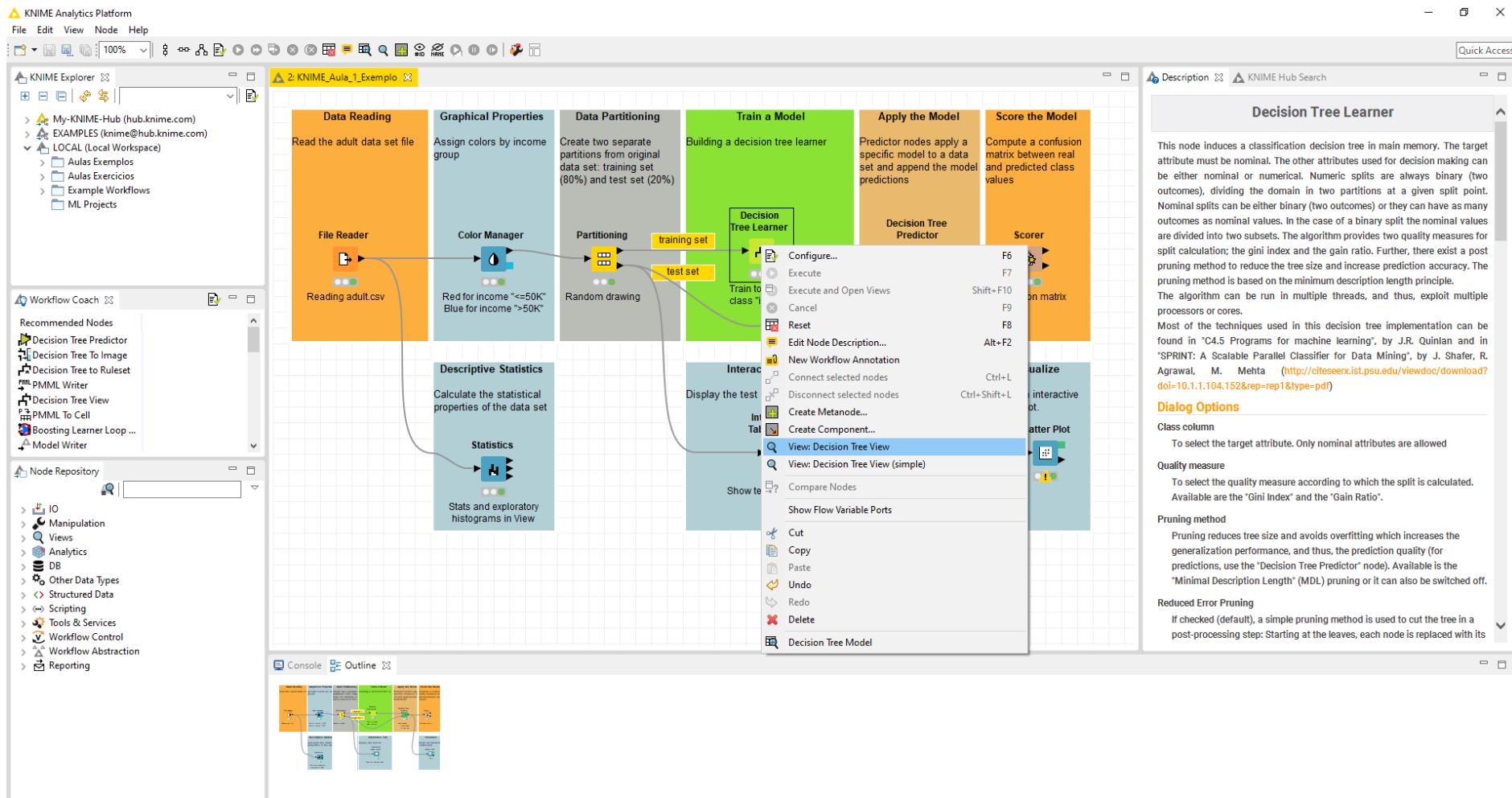
31

Terminology of AI

Intro to Knime

WORKFLOW

Hands On



Node Context Options

Model Learner

32

Terminology of AI

Intro to Knime

WORKFLOW

Hands On

KNIME Analytics Platform

File Edit View Node Help

100% Quick Access

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
 - Aulas Ejemplos
 - Aulas Ejercicios
 - 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

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

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

File Hilit Tree

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

Descriptive Statistics: Calculate the statistic properties of the data set

Statistics: Stats and exploratory histograms in View

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 principle of the algorithm is to find the best split for each node. Thus, exploit multiple parallelism where possible.

Implementation can be J.R. Quinlan and R. Rivest, "Learning Decision Trees", by J. Shafer, R. Rivest, and P. Mehta, MIT Press, Cambridge, MA, 1993. viewdoc/download?doi=10.1.1.103.103&rep=rep1&type=pdf

are allowed

split is calculated.

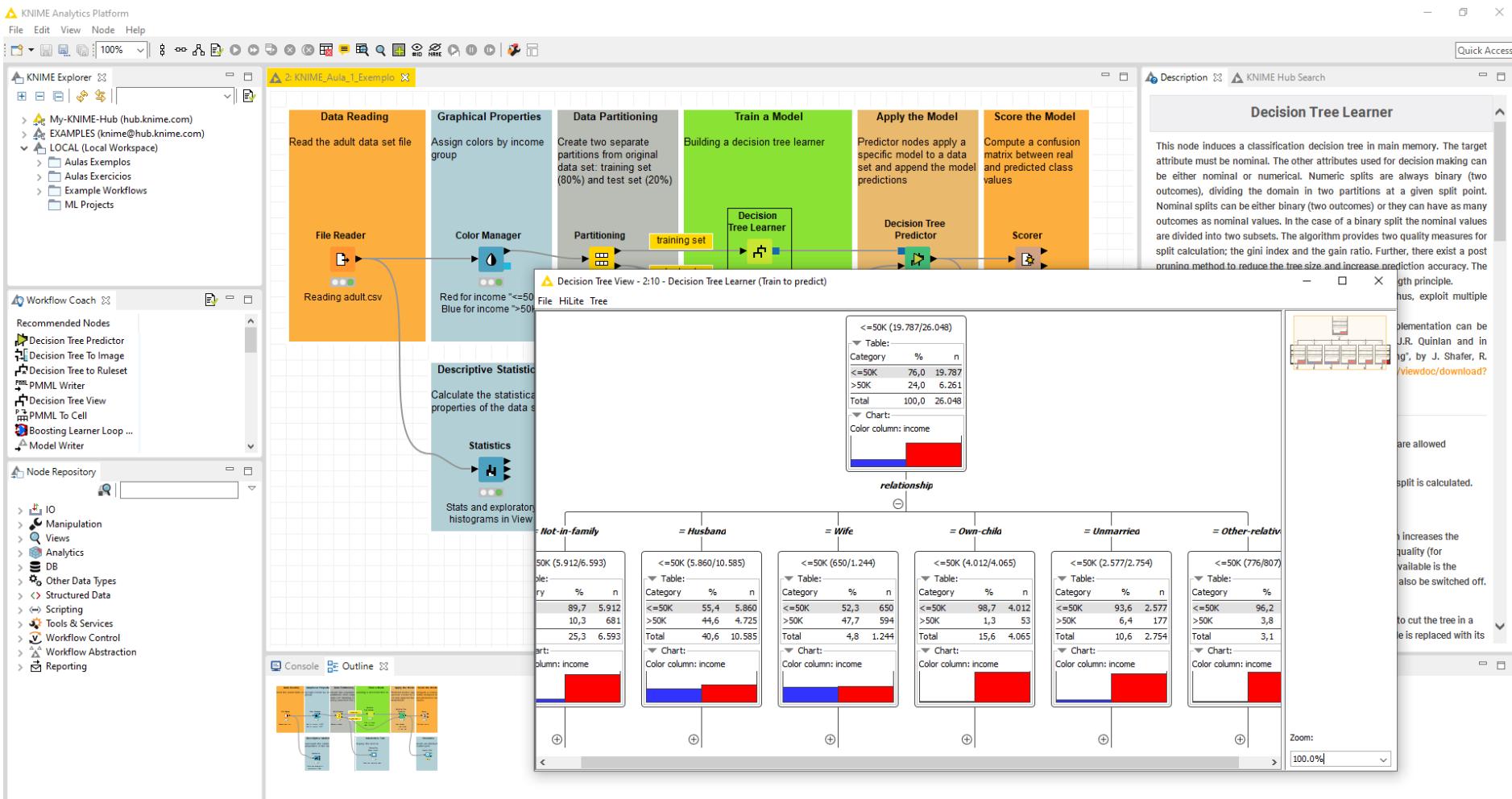
increases the quality for available is the also be switched off.

to cut the tree in a node is replaced with its

Console

Outline

Zoom: 100.0%



Node Context Options

Scorer

33

Terminology of AI

Intro to Knime

WORKFLOW

Hands On

KNIME Analytics Platform

File Edit View Node Help

100% Quick Access

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
 - Aulas Exemplos
 - Aulas Ejercicios
 - 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

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

Scorer

Compared 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 output port contains statistics such as True-Positives, False-Negatives, Recall, Precision, Sensitivity, Specificity and the overall accuracy and Cohen's kappa.

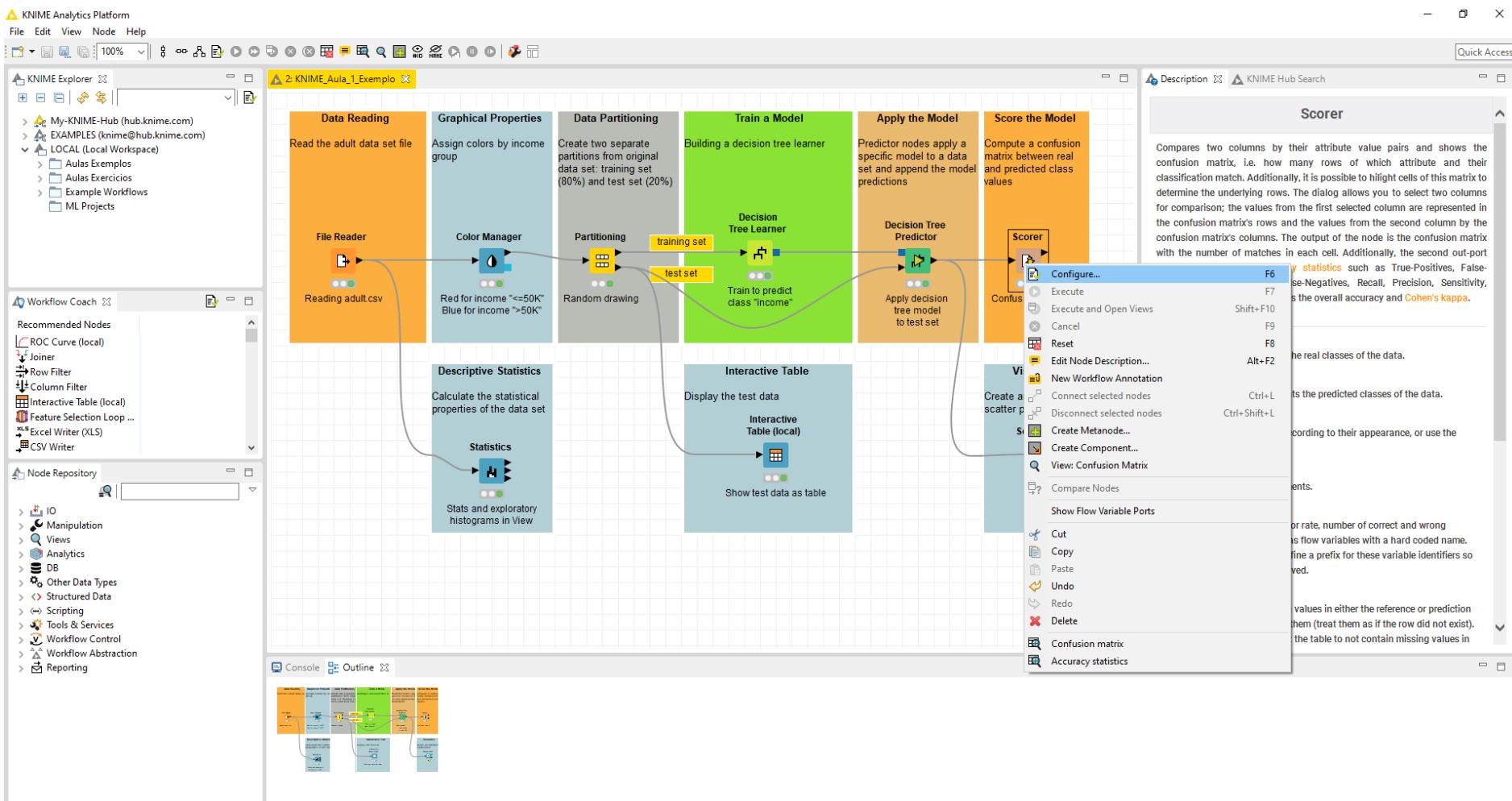
Description

Configure...

- Execute F6
- Execute and Open Views F7
- Cancel Shift+F10
- Reset F9
- Edit Node Description... F8
- New Workflow Annotation Alt+F2
- Connect selected nodes Ctrl+L
- Disconnect selected nodes Ctrl+Shift+L
- 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

34

Terminology of AI

Intro to Knime

WORKFLOW

Hands On

KNIME Analytics Platform

File Edit View Node Help

100% Quick Access

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
 - Aulas Ejemplos
 - Aulas Ejercicios
 - 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

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 adult.csv

Color Manager: Red for income " $\leq 50K$ " Blue for income " $>50K$ "

Partitioning: Random draw

Decision Tree Learner

Decision Tree Predictor

Scorer: Dialog - 2/6 - Scorer (Confusion matrix)

Confusion matrix

Descriptive Statistics: Calculate the statistical properties of the data set

Statistics: Stats and exploratory histograms in View

Scatter Plot

Scorer Dialog Options

First column: \$|income

Second column: \$|Prediction (income)

Sorting of values in tables: Insertion order, Reverse order

Provide scores as flow variables: Use name prefix

In case of missing values...: Ignore, Fail

OK, Apply, Cancel, ?

Scorer Description

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 output 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.

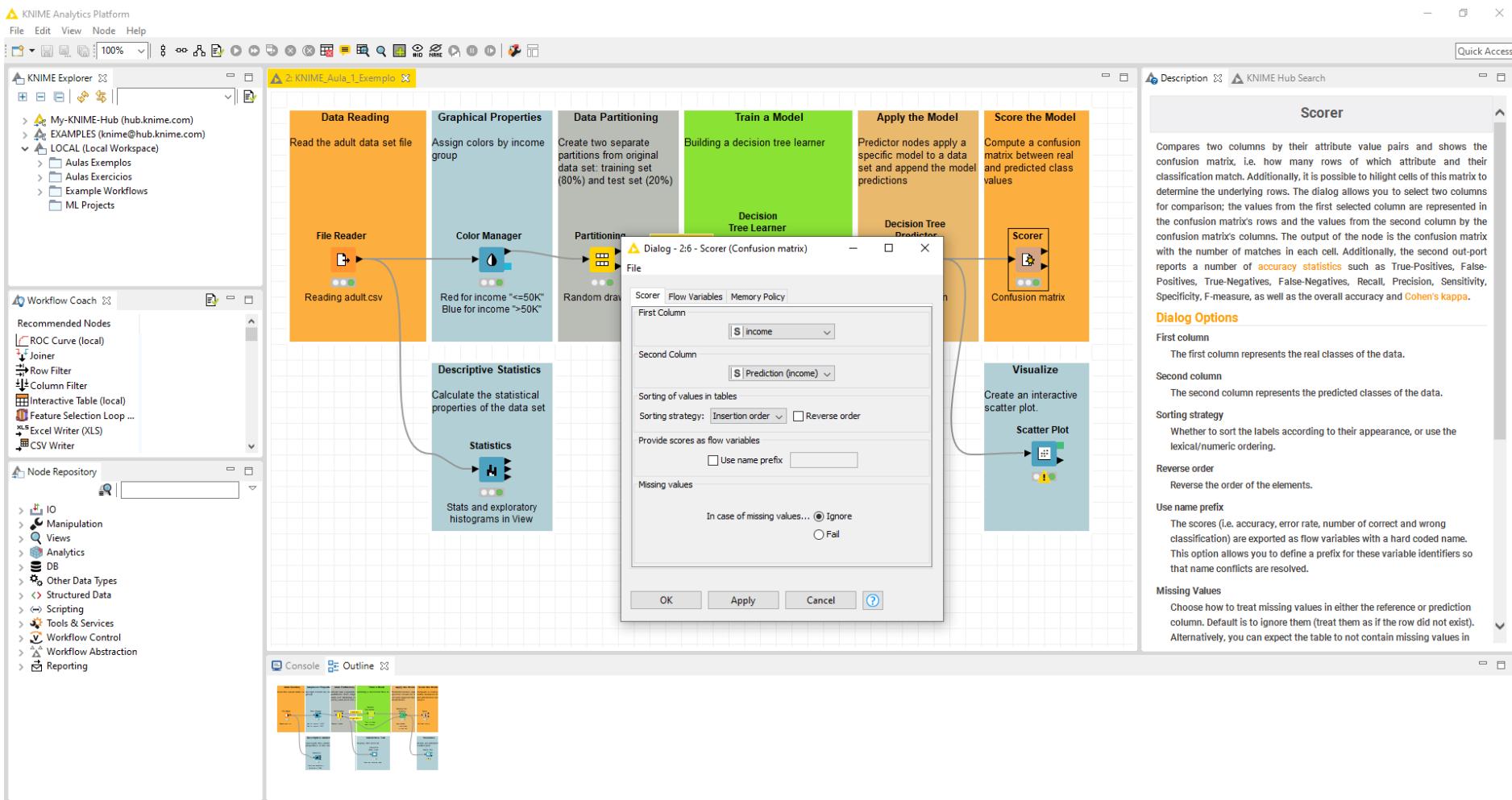
Reverse order: Reverse the order of the elements.

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

35

Terminology of AI

Intro to Knime

WORKFLOW

Hands On

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 provides statistics such as True-Positives, False-Negatives, Recall, Precision, Sensitivity, the overall accuracy and Cohen's kappa.

real classes of the data.

the predicted classes of the data.

according to their appearance, or use the

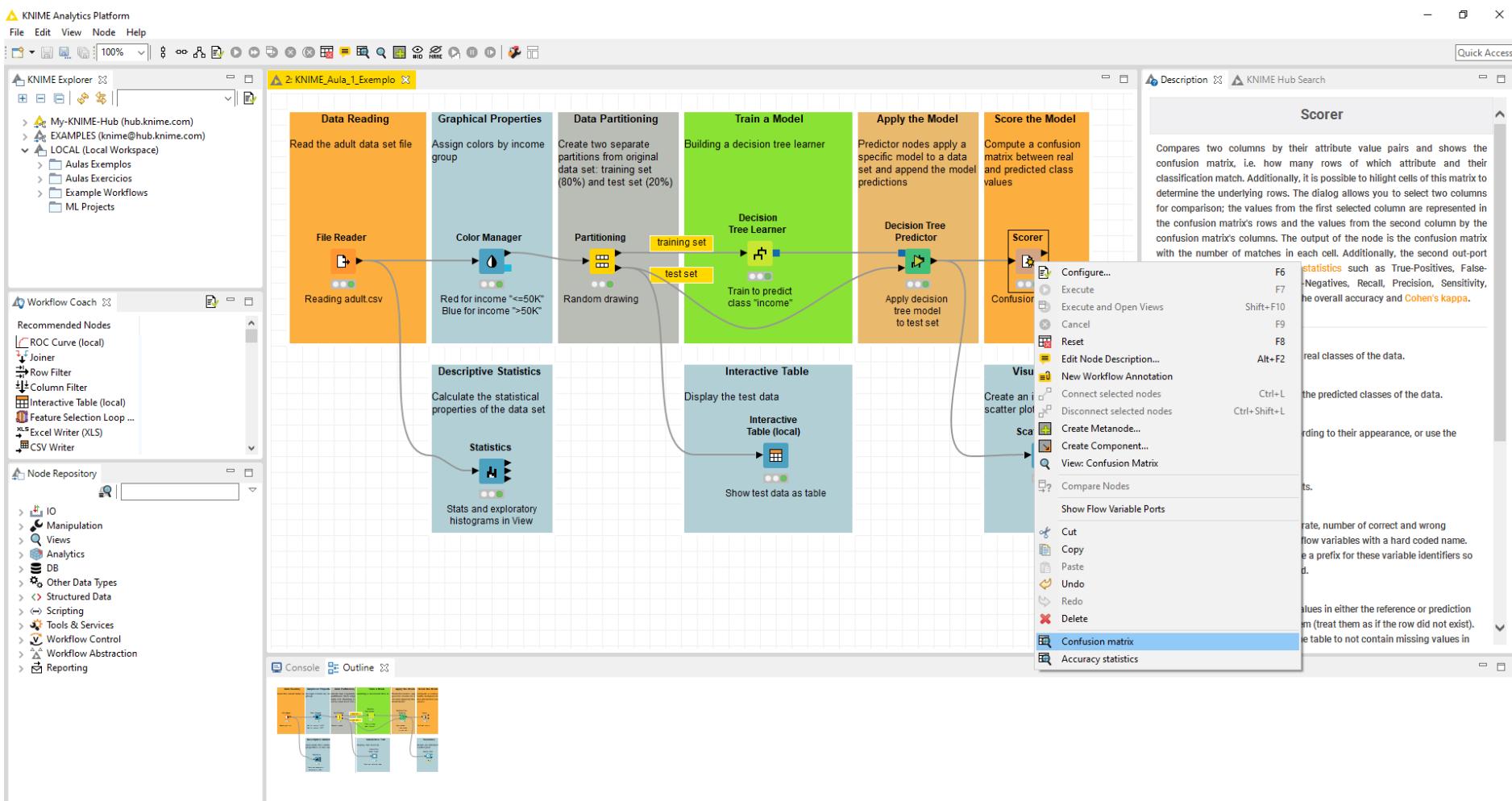
ts.

rate, number of correct and wrong flow variables with a hard coded name. Use a prefix for these variable identifiers so they do not conflict with other variables.

values in either the reference or prediction column (treat them as if the row did not exist). Set the table to not contain missing values in the rows.

Confusion matrix

Accuracy statistics



```

graph LR
    FR[File Reader] --> CM[Color Manager]
    CM --> P[Partitioning]
    P -- training set --> DTL[Decision Tree Learner]
    P -- test set --> DTP[Decision Tree Predictor]
    DTL --> DTP
    DTP --> S[Scorer]
    S --> IT[Interactive Table]
    S --> CM_S[Statistics]
    CM_S --> IT
  
```

Node Context Options

Scorer

36

Terminology of AI

Intro to Knime

WORKFLOW

Hands On

KNIME Analytics Platform

File Edit View Node Help

100% Quick Access

KNIME Explorer

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
- LOCAL (Local Workspace)
 - Aulas Ejemplos
 - Aulas Ejercicios
 - 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

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

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

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

Descriptive Statistics: Calculate the statistical properties of the data set

Statistics: Stats and exploratory histograms in View

Scorer: Confusion matrix

Visualize: Create an interactive scatter plot.

Scatter Plot

Correct classified: 5,463
Wrong classified: 1,050
Accuracy: 83,878 %
Error: 16,122 %
Cohen's kappa (κ) 0,531

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 output 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.

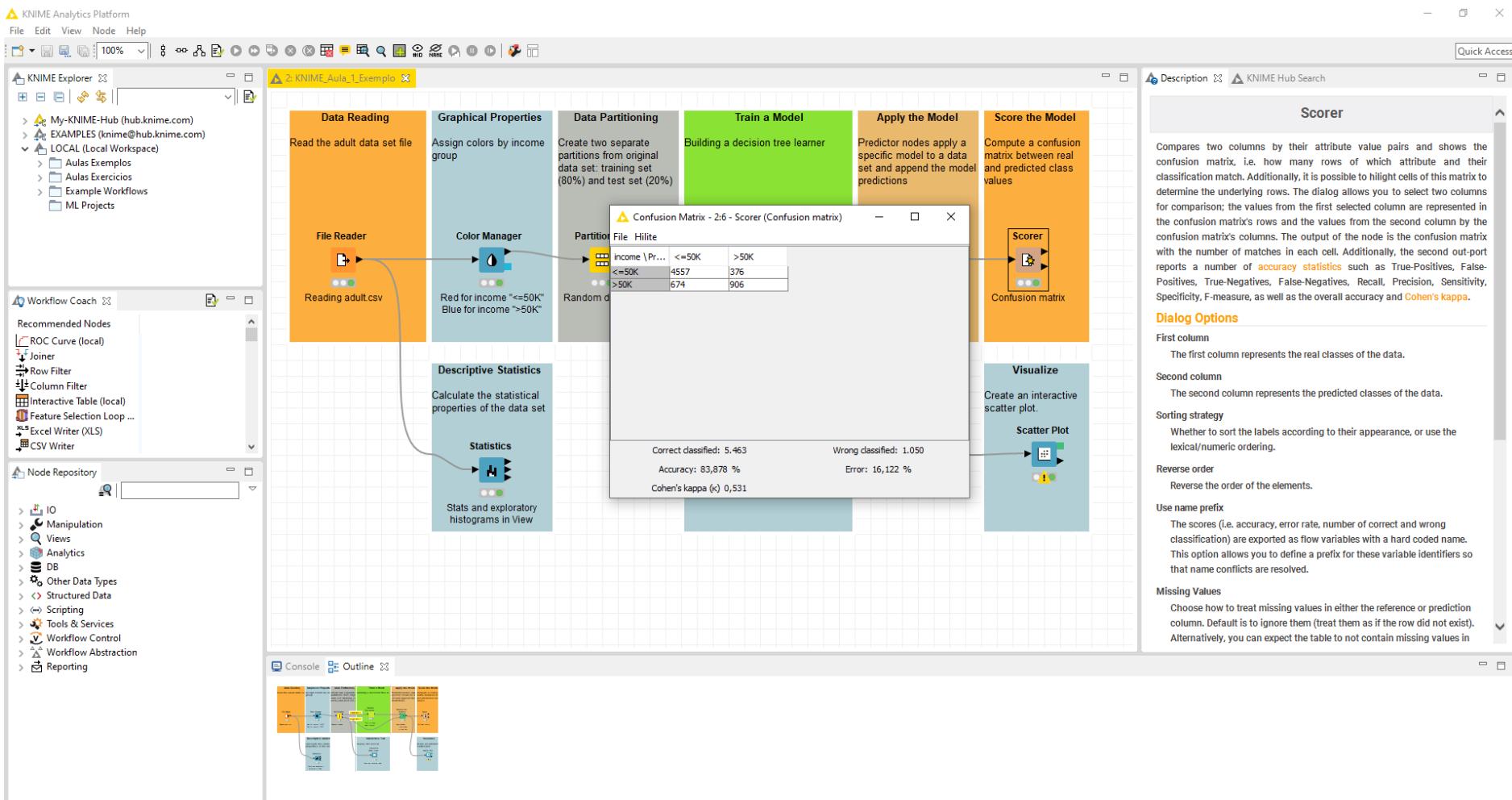
Reverse order: Reverse the order of the elements.

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



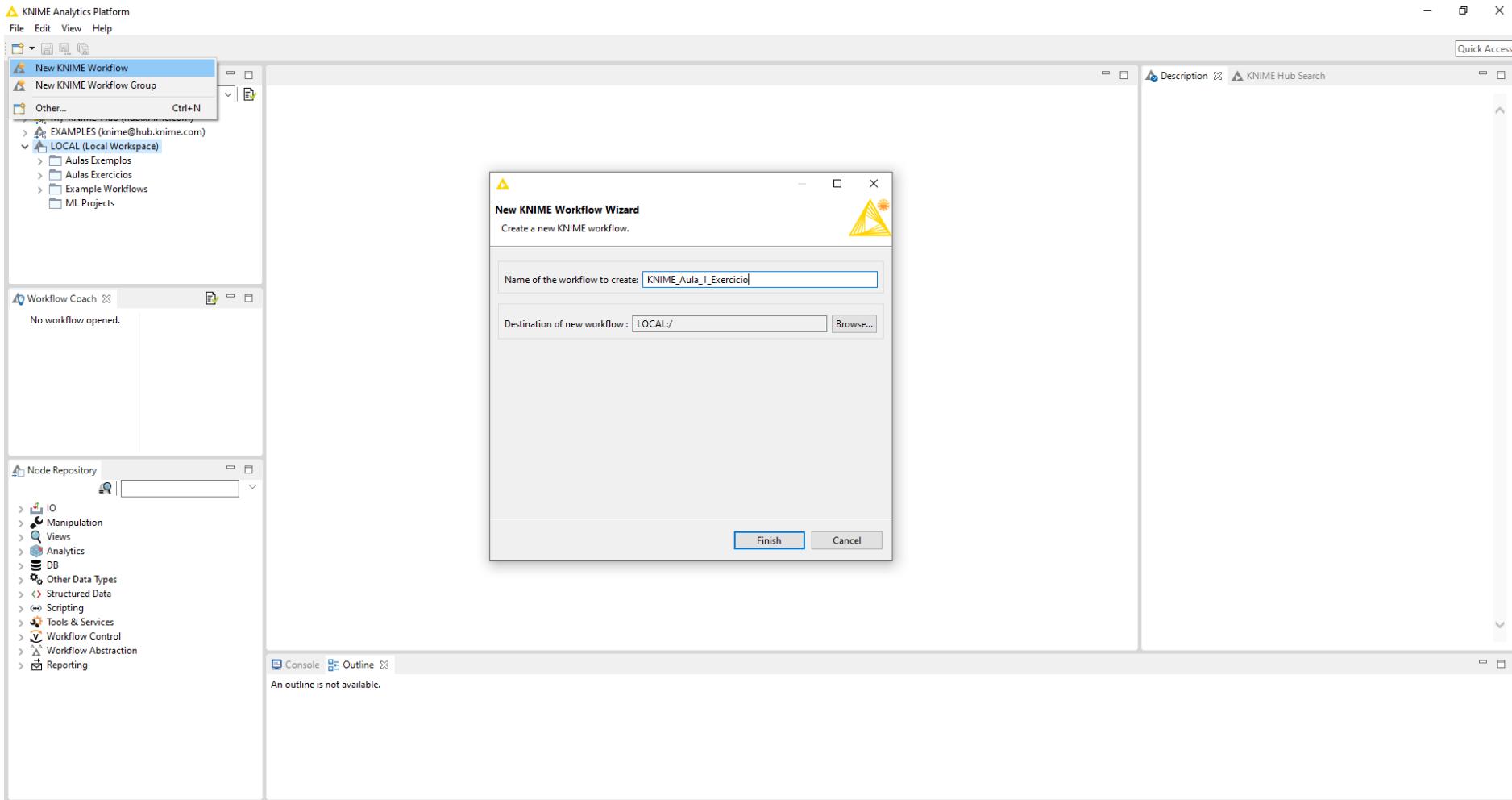
Hands On

37

Terminology of AI

Intro to Knime

Workflow

HANDS ON

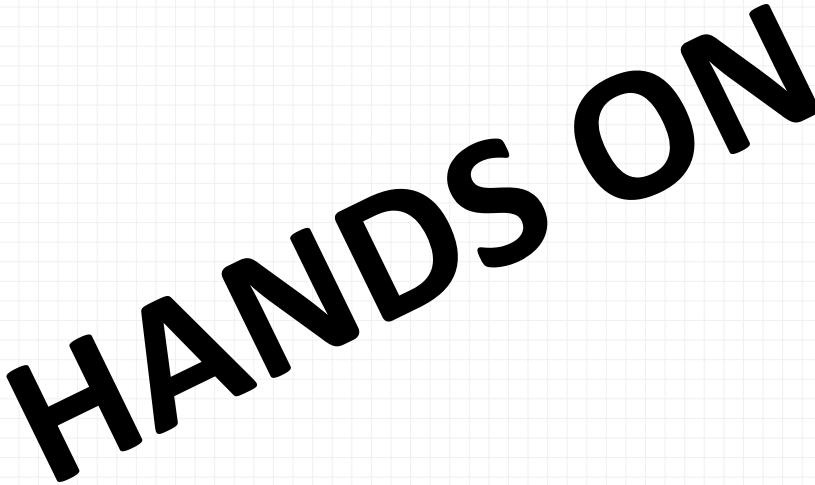
Hands On

38

Terminology of AI

Intro to Knime

Workflow

HANDS ON

The image shows the KNIME Analytics Platform interface. The central workspace is filled with a large, hand-drawn style word "HANDS ON" written diagonally from bottom-left to top-right. The interface includes a top menu bar with File, Edit, View, Node, Help, and various toolbars. On the left, there's a KNIME Explorer showing local workspaces like "LOCAL (Local Workspace)" and "Aulas Exemplos", and a Node Repository listing categories such as IO, Manipulation, Views, Analytics, DB, and Reporting. A Workflow Coach panel is also visible on the left. On the right, there's a Description panel and a KNIME Hub Search with a search bar. At the bottom, there are tabs for Console and Outline.