University of Minho
School of Engineering

ISLab

Machine Learning and Decision-Making

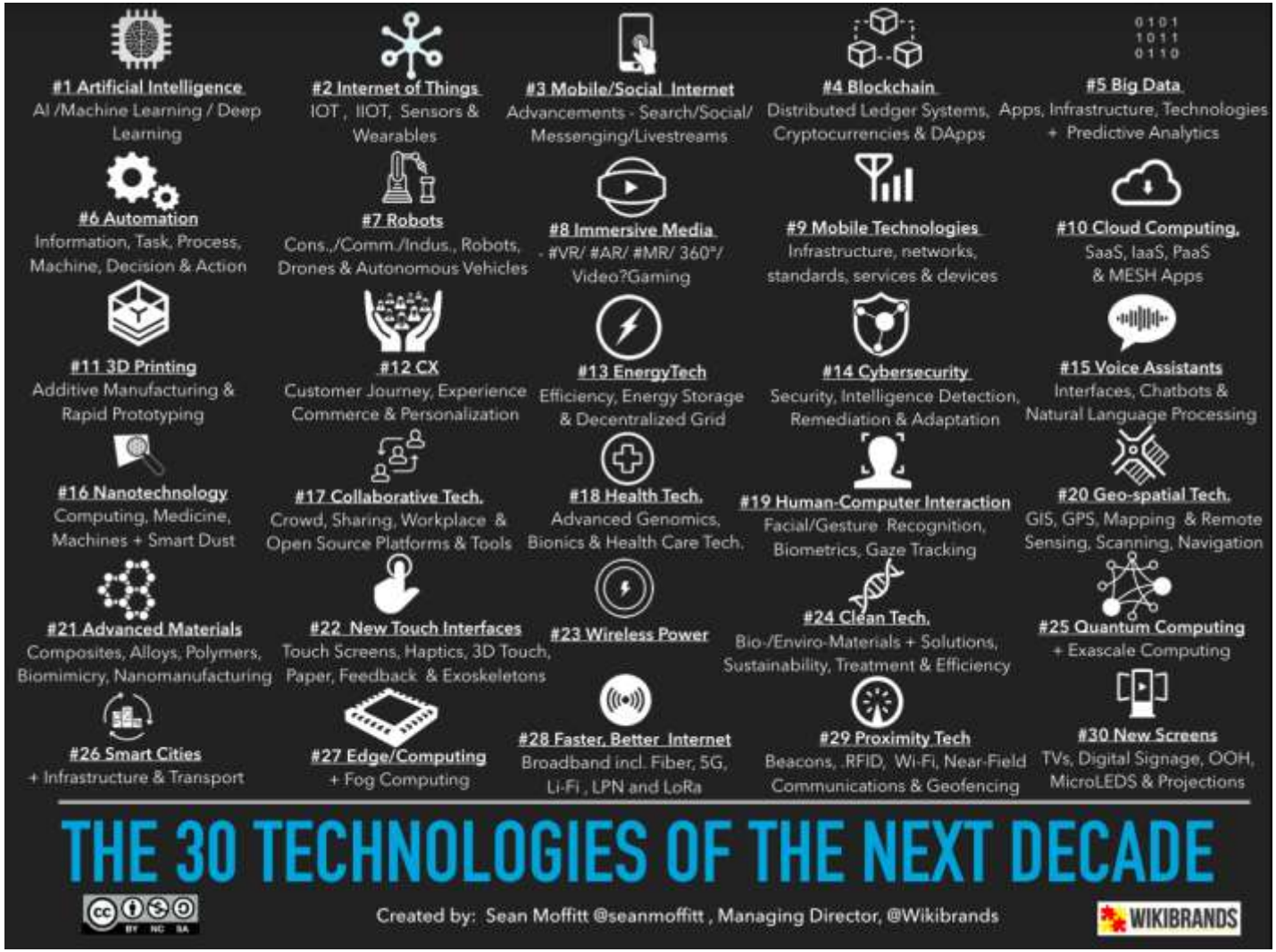
ADI @ LEI/3º, MiEI/4º - 2º Semestre

Filipe Gonçalves, Inês Alves, Cesar Analide

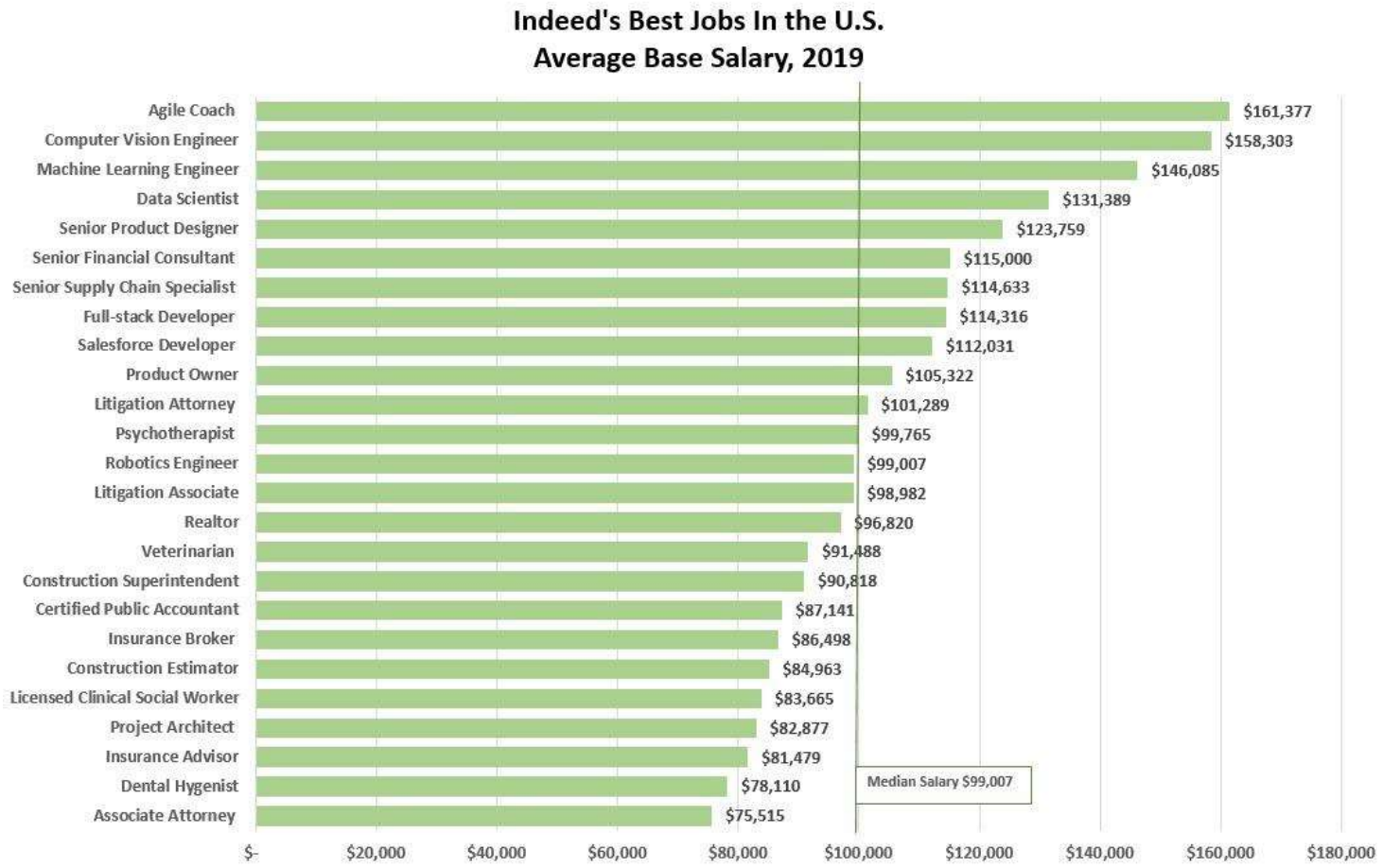
Part II – February 2022

Contents

Technologies of the next decade

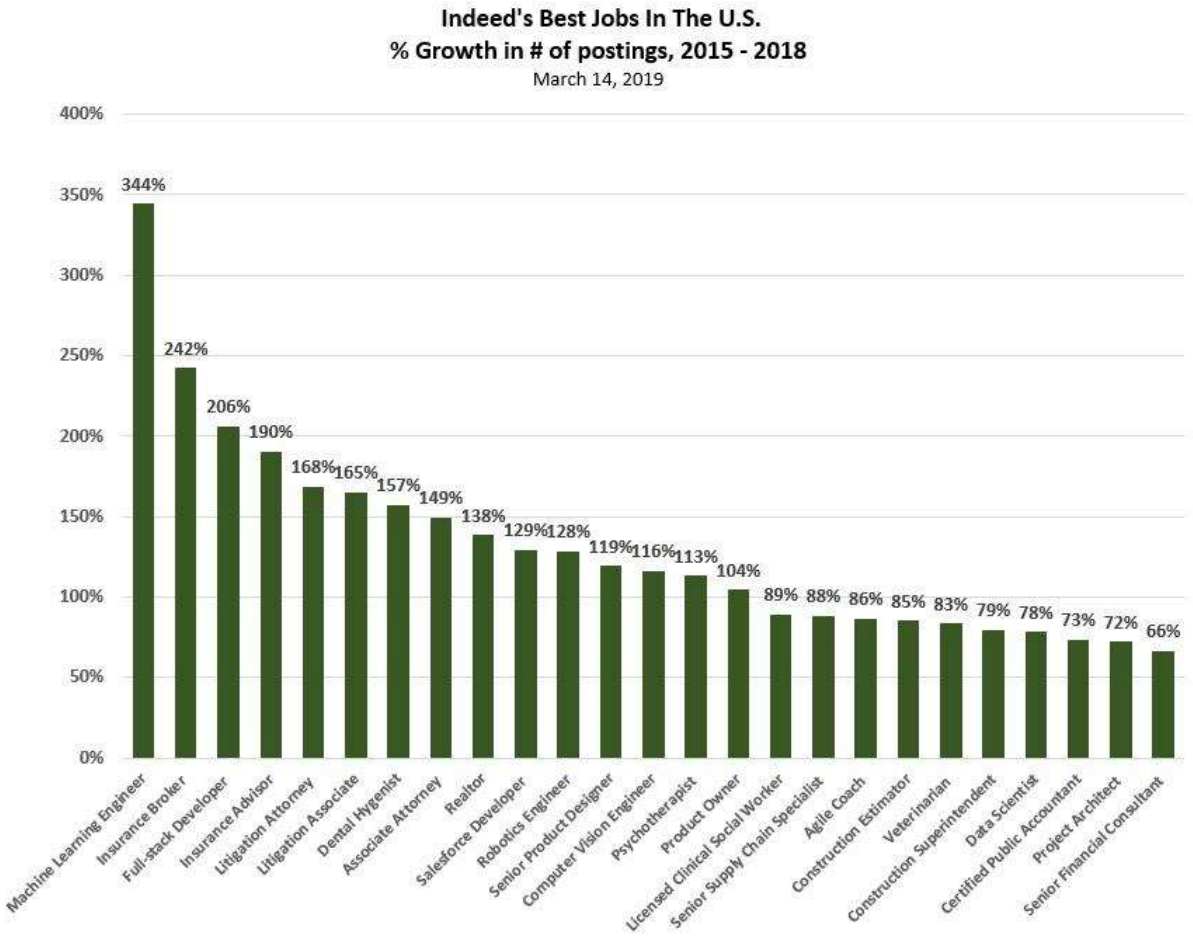


Motivation



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

Motivation



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

Terminology of AI

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

Terminology of AI

Machine Learning vs **Data Science**

There is no universal adherence!!!

Terminology of AI

Machine Learning vs Data Science

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

Arthur Samuel

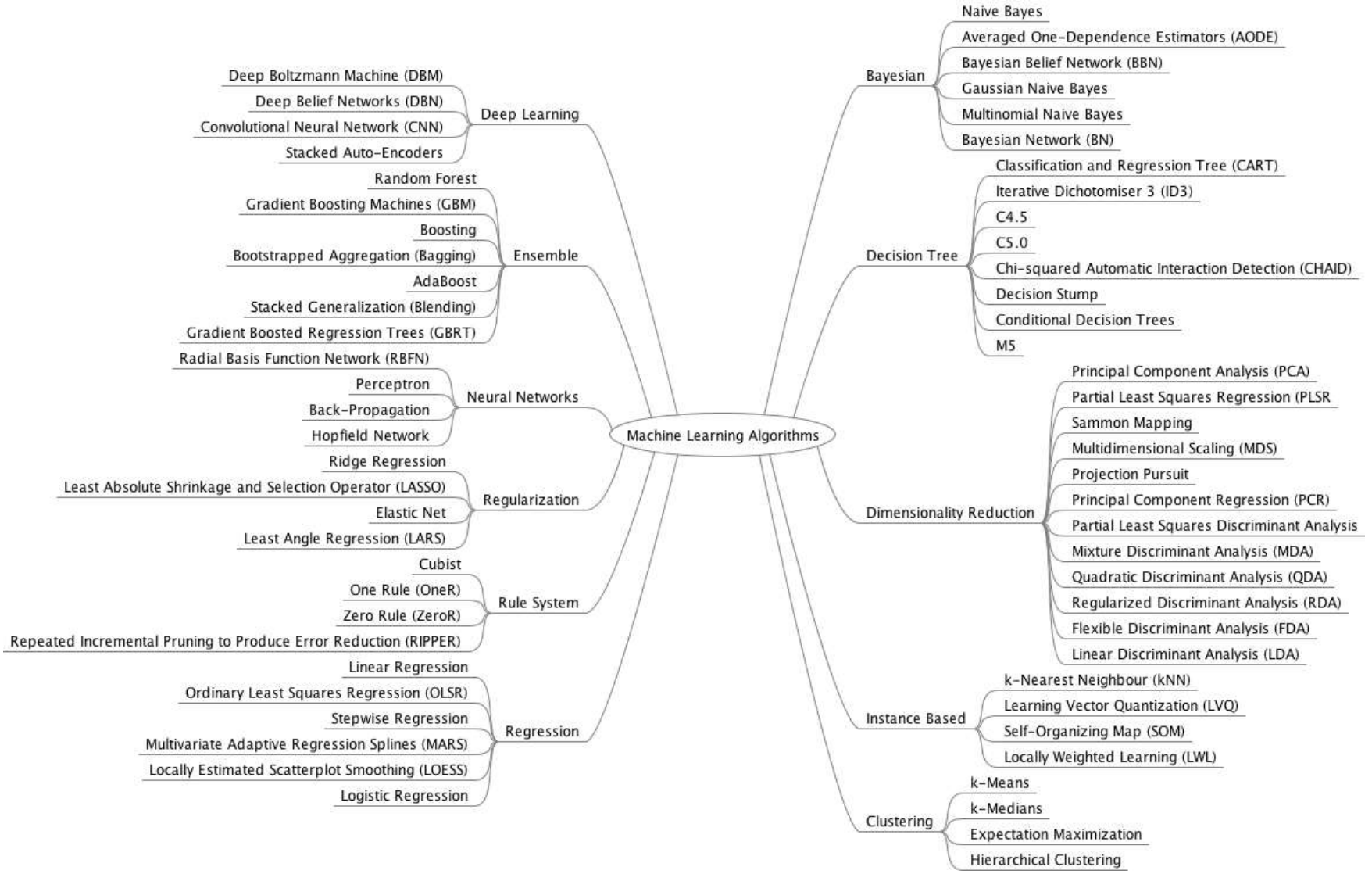
Usually results in a software artefact

- 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

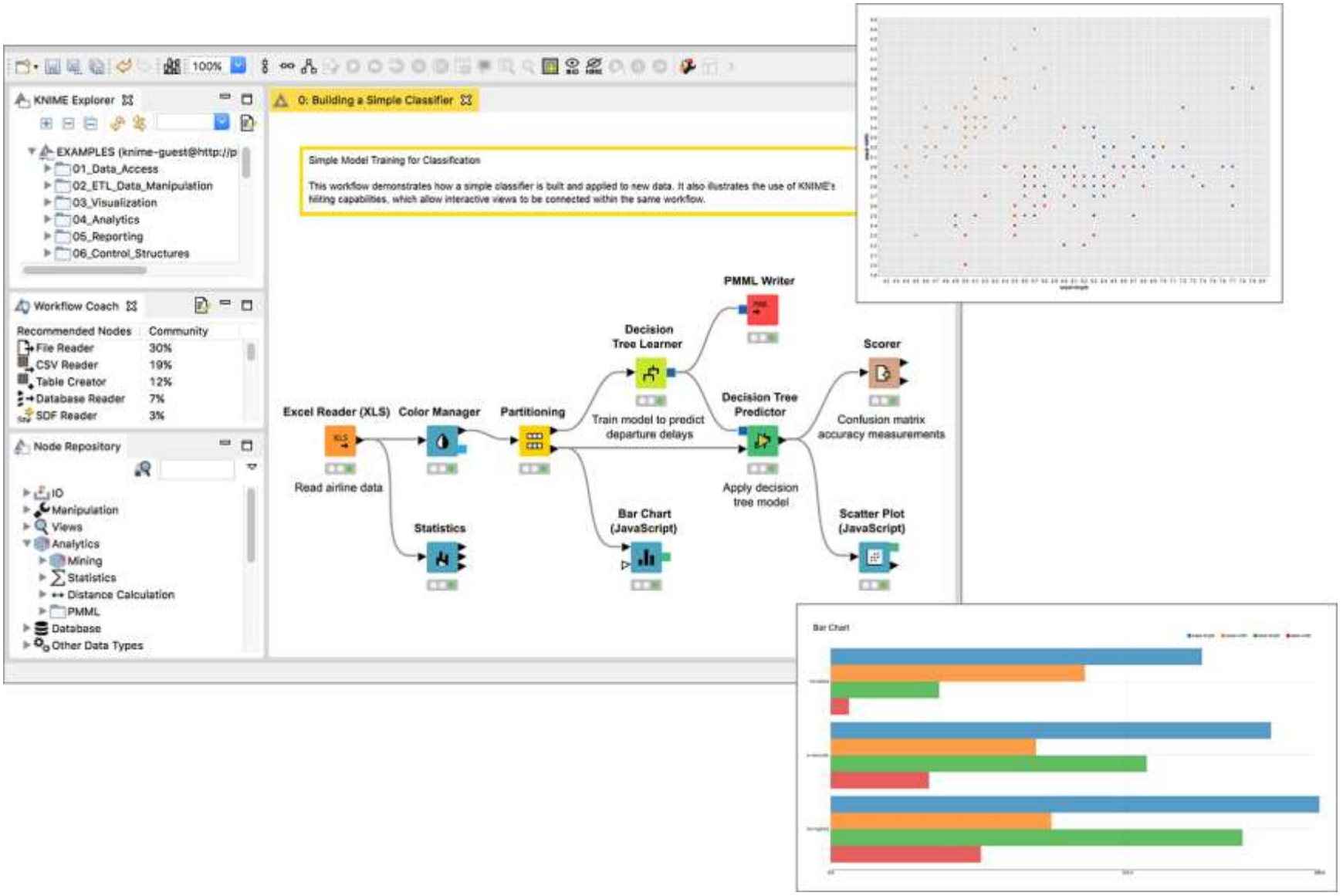
There is no universal adherence!!!

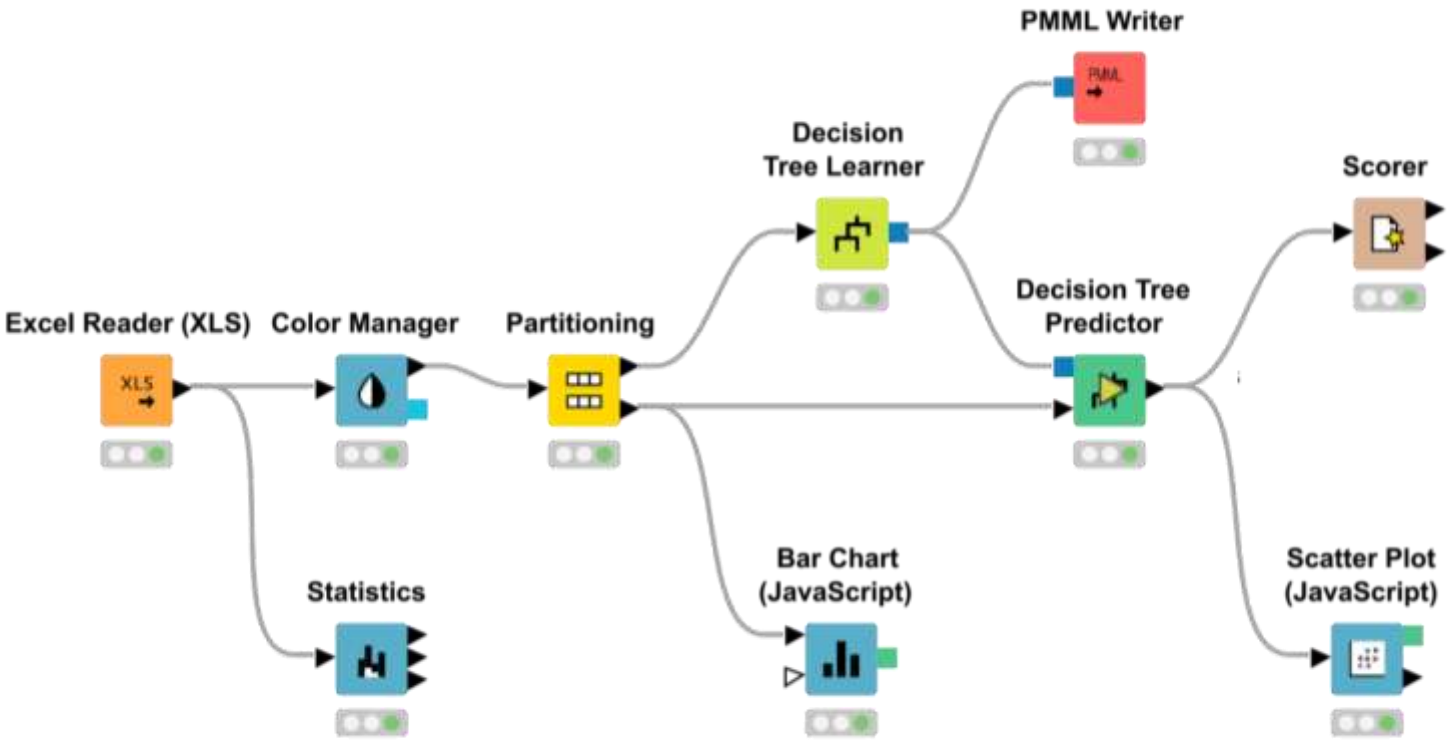
Terminology of AI

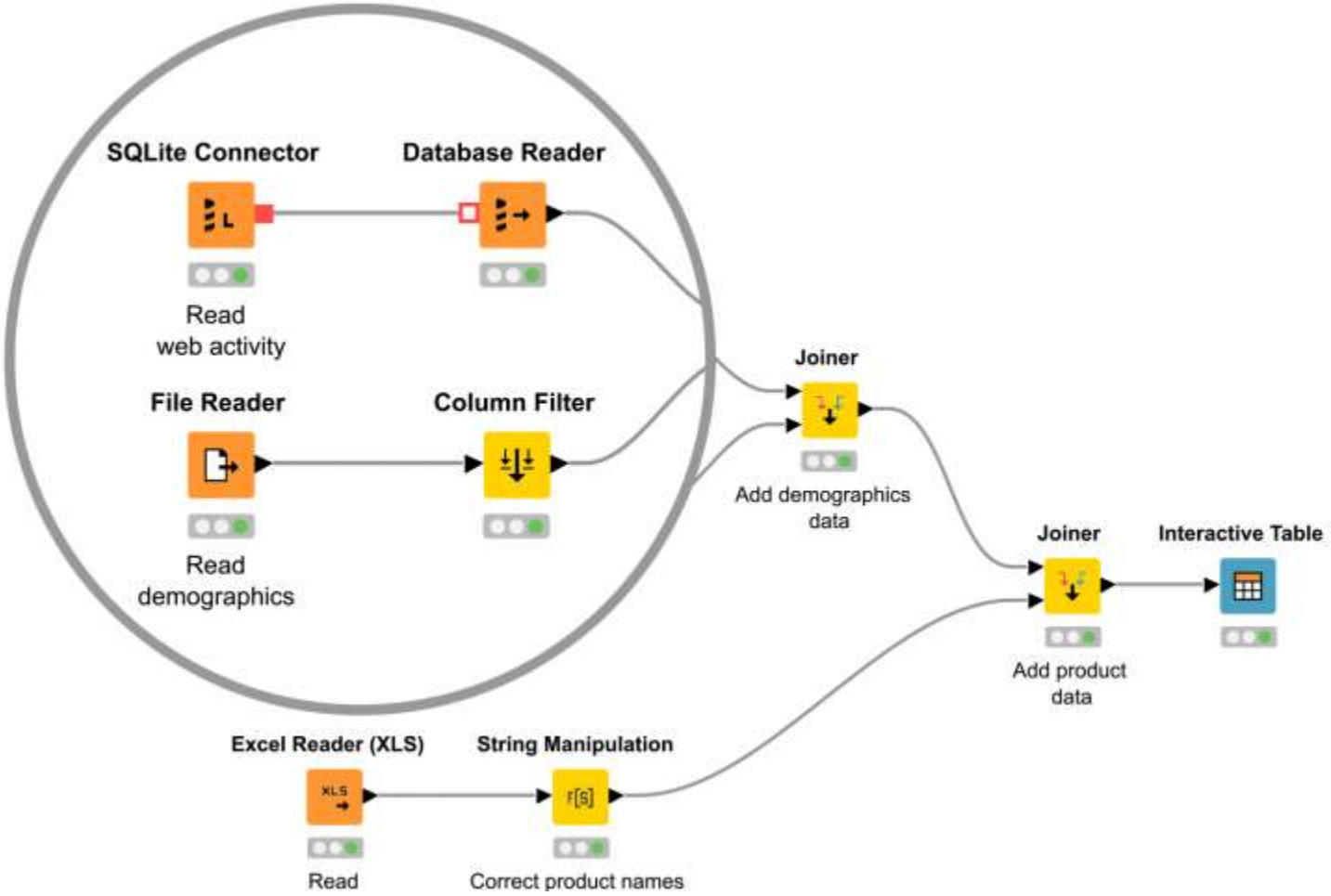


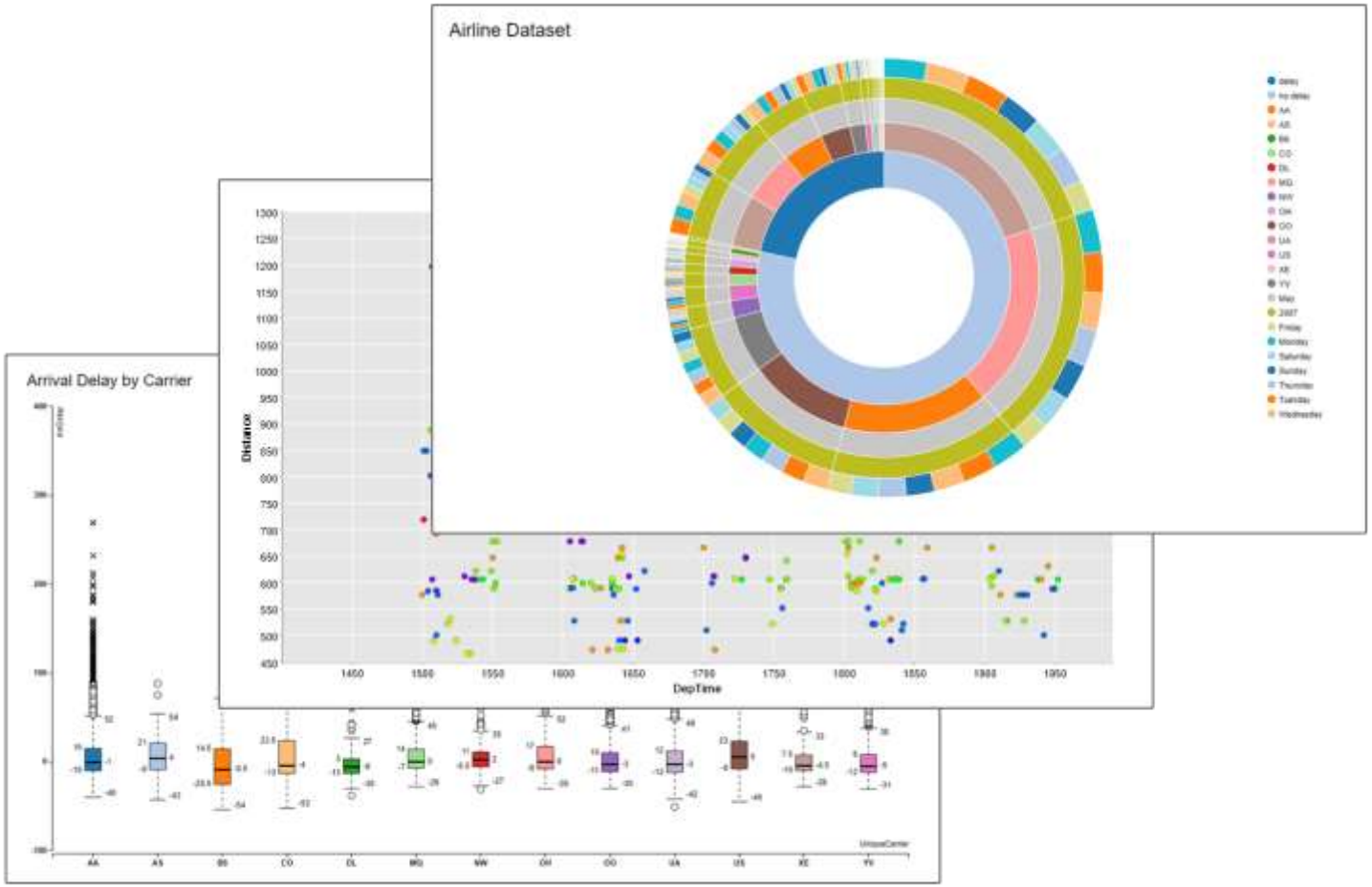
Open for Innovation ®

KNIME



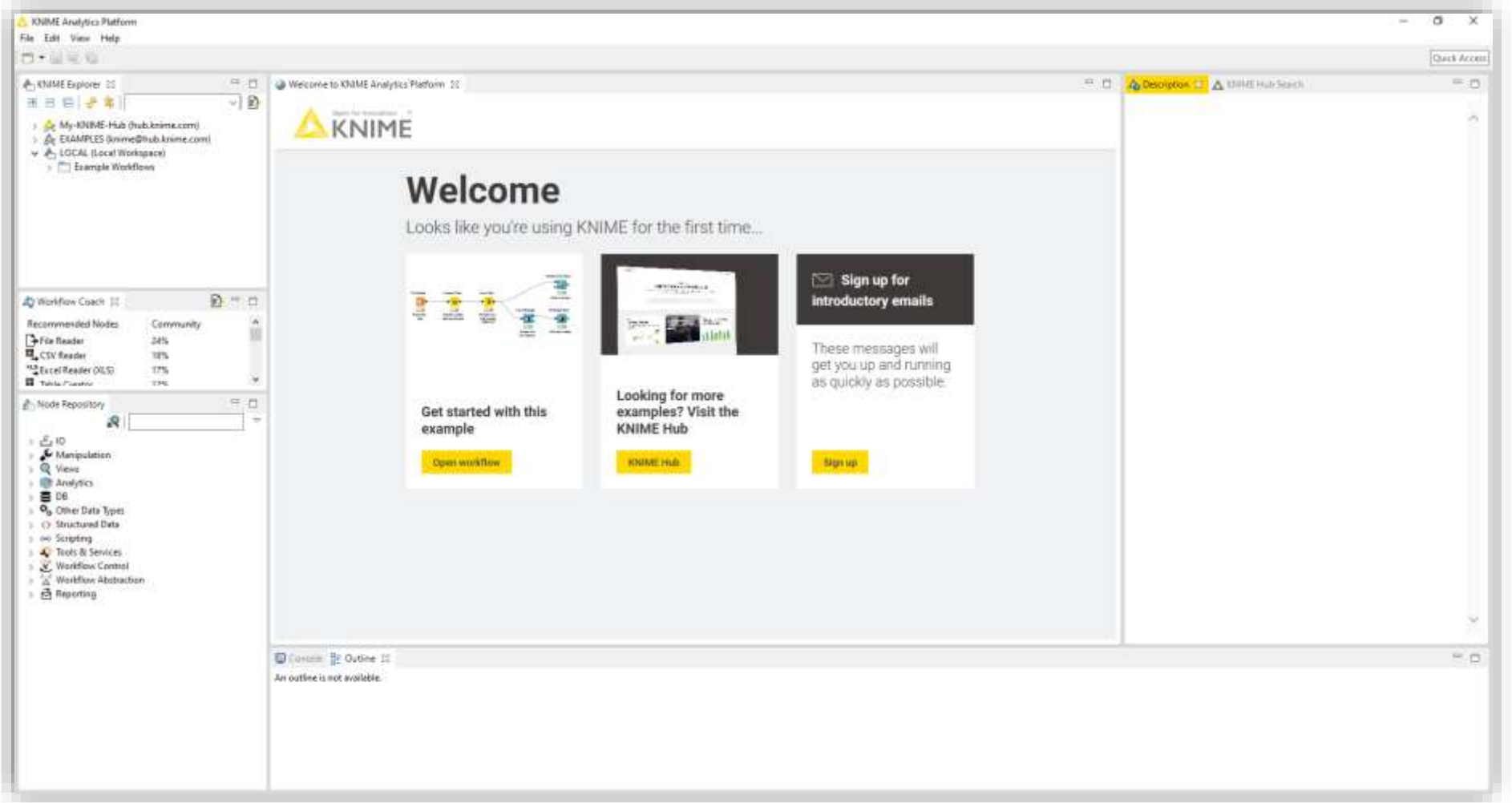

PMML Writer
Decision
Tree Learner
Scorer
Excel Reader (XLS)
Color Manager
Partitioning
Decision Tree
Predictor
Statistics
Bar Chart
(JavaScript)
Scatter Plot
(JavaScript)


SQLite Connector
Database Reader
Read
web activity
File Reader
Column Filter
Read
demographics
Joiner
Add demographics
data
Joiner
Interactive Table
Add product
data
Excel Reader (XLS)
String Manipulation
Read
Correct product names

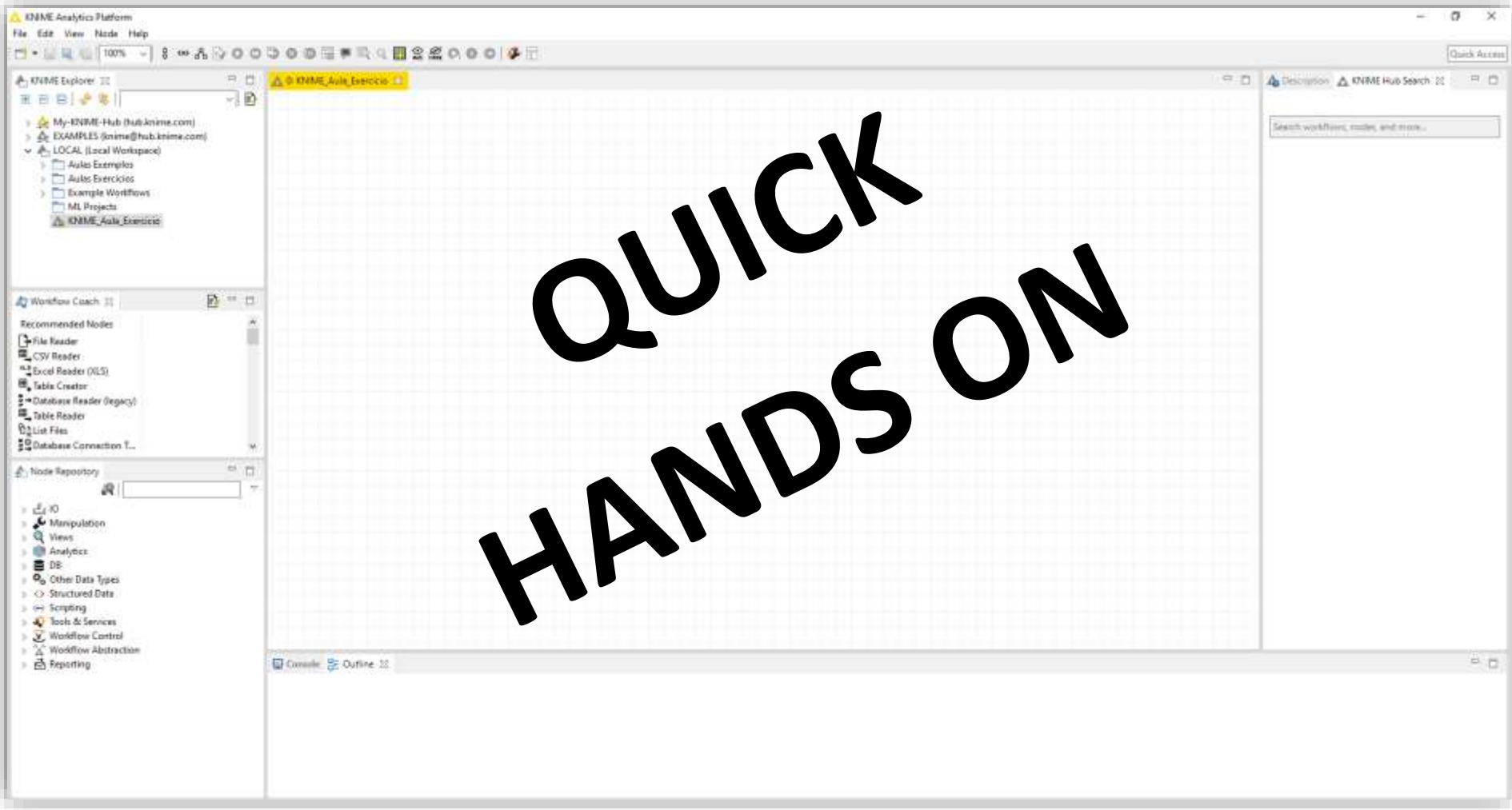

Airline Dataset
Arrival Delay by Carrier
Distance
DepTime

- 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

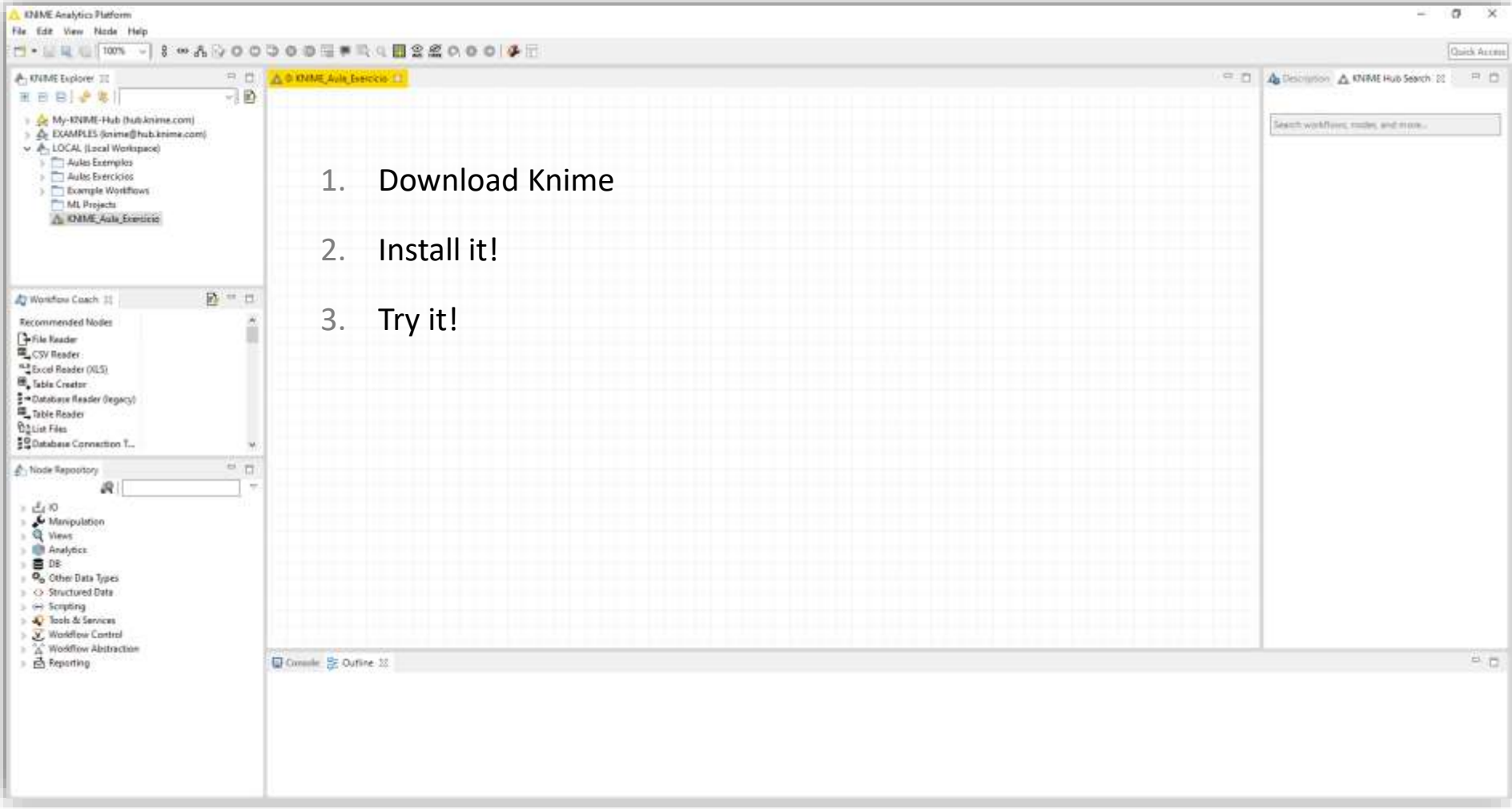
Source: Gartner (January 2019)


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%
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
KNIME
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
Description
KNIME Hub Search
Quick Access
Console
Outline
An outline is not available.

(Quick) Hands On



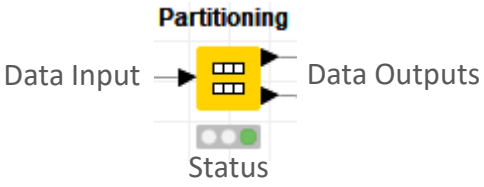
(Quick) Hands On

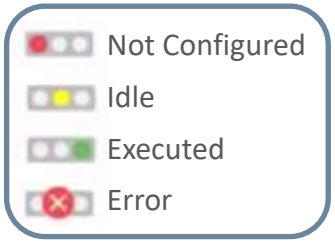


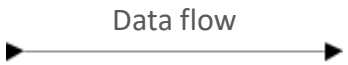
1. Download Knime
2. Install it!
3. Try it!

Nodes and Workflows

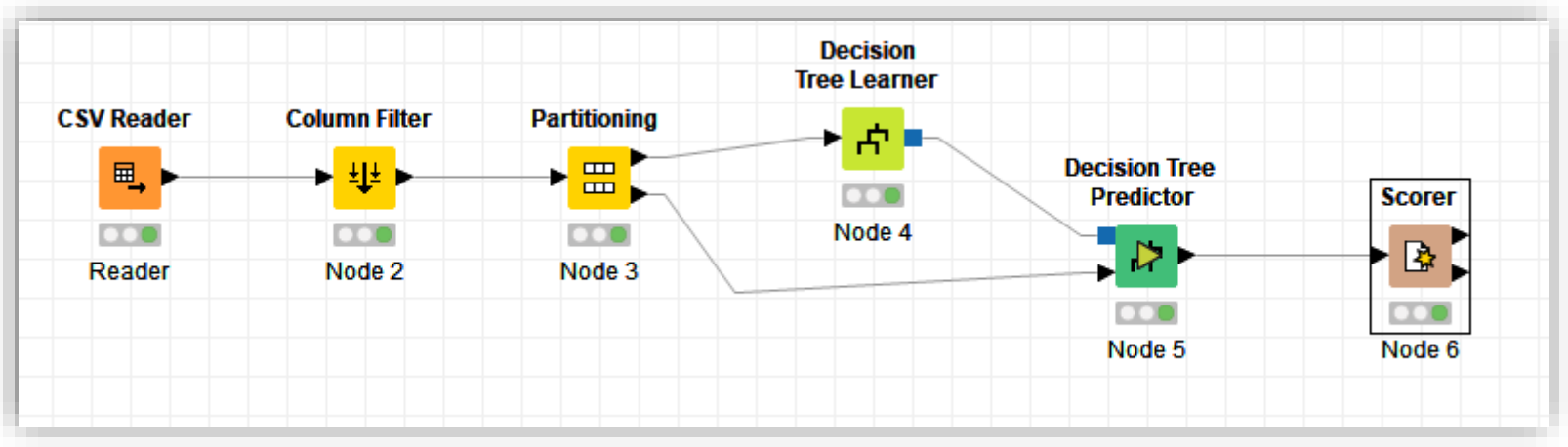
Nodes



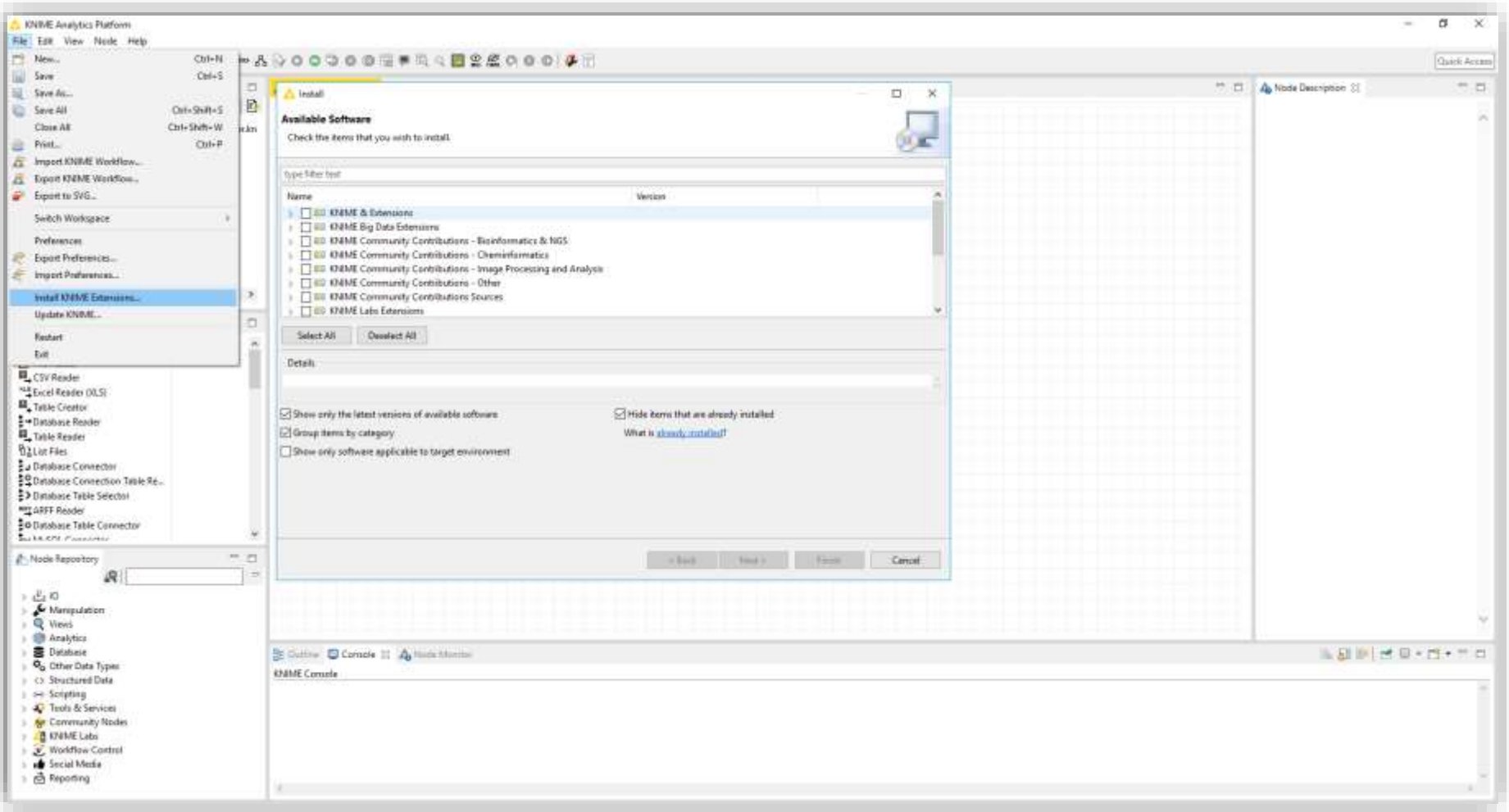




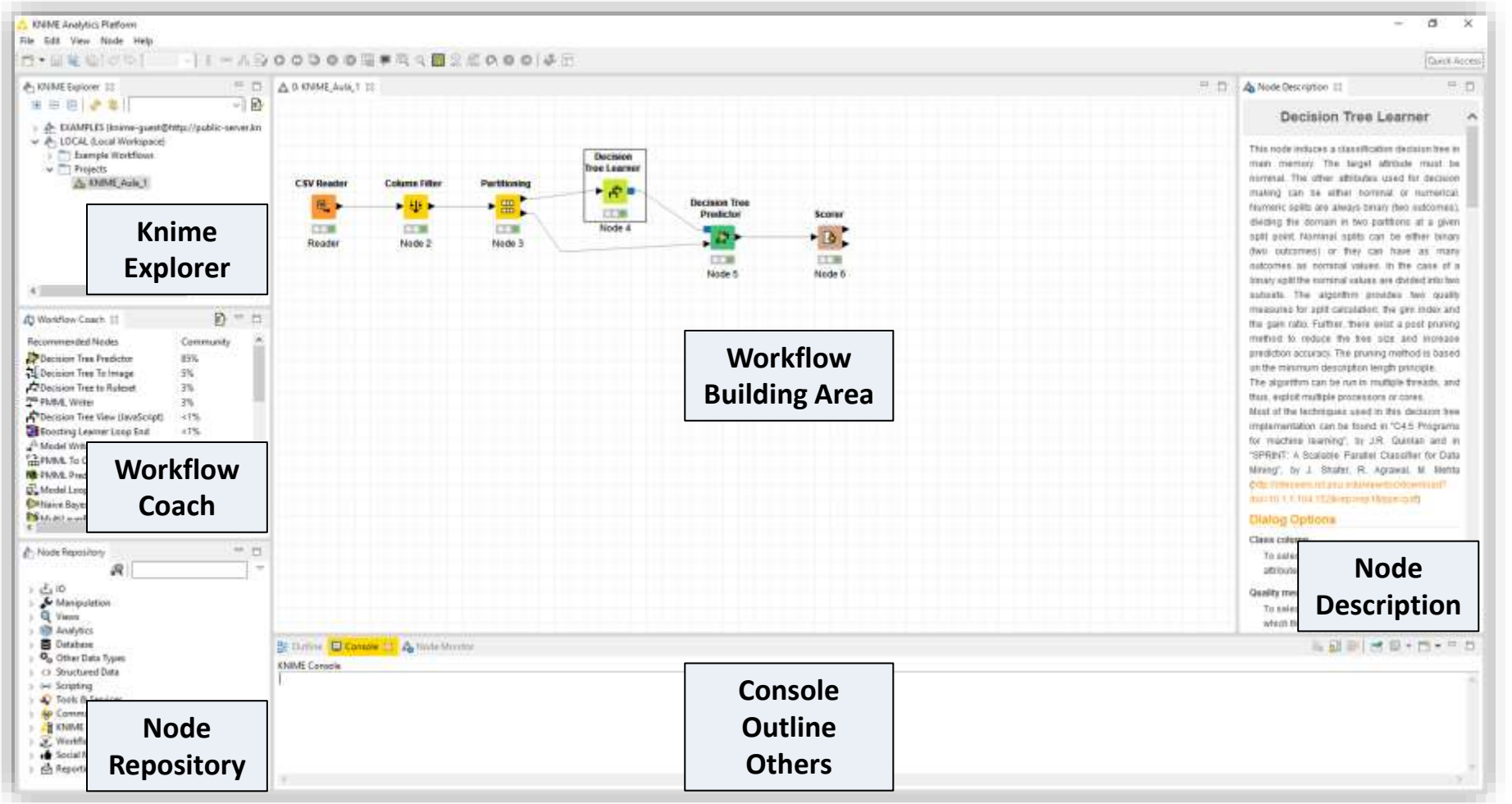
Workflow



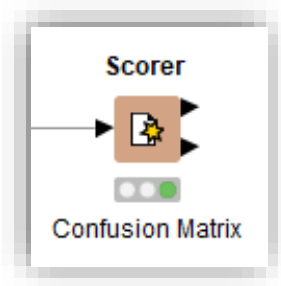
KNIME Extensions



Main Views


Knime Explorer
Workflow Coach
Node Repository
Workflow Building Area
Node Description
Console Outline Others
CSV Reader
Column Filter
Partitioning
Decision Tree Learner
Decision Tree Predictor
Scorer
Reader
Node 2
Node 3
Node 4
Node 5
Node 6
Decision Tree Learner

Node Description



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 hilight 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.

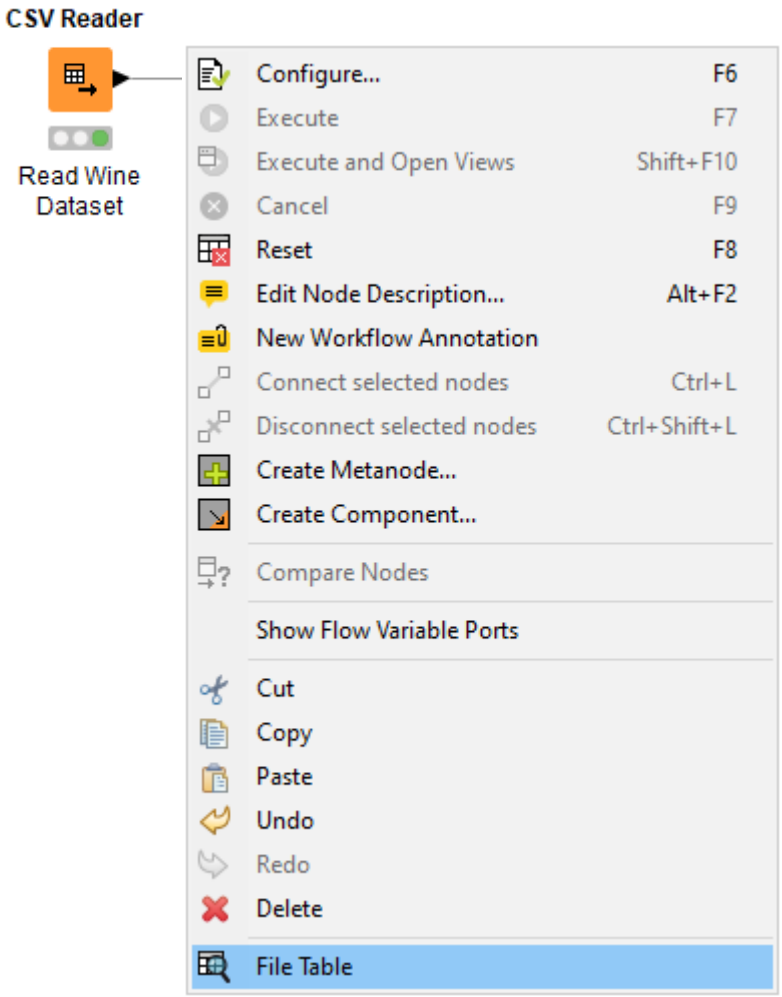
Output Ports	
0	The confusion matrix.
1	The accuracy statistics table.

Views

Confusion Matrix

Displays the confusion matrix in a table view. It is possible to hilight 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



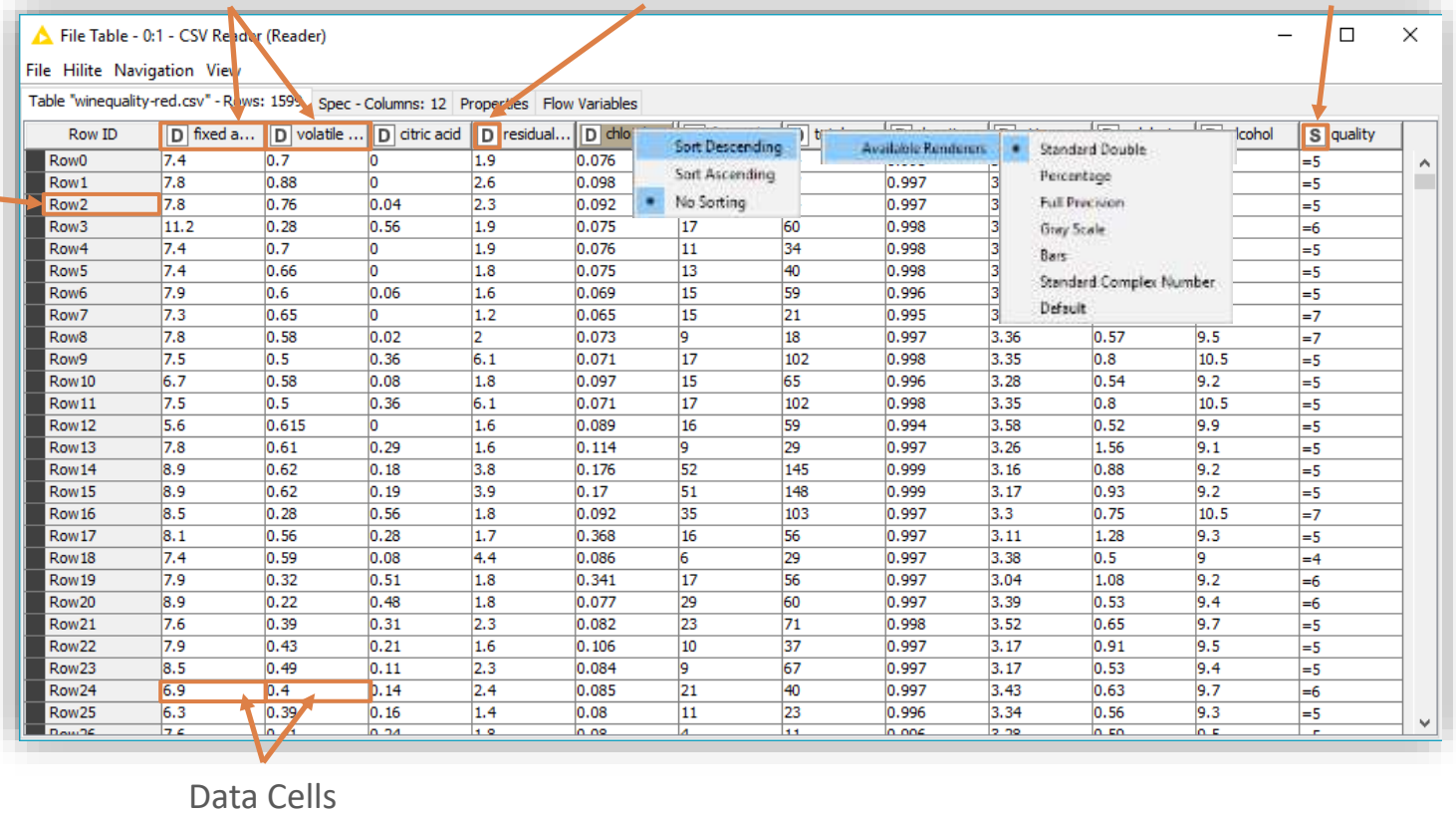
Data Table Structure View

Column Headers

Data Type (Double)

Data Type (String)

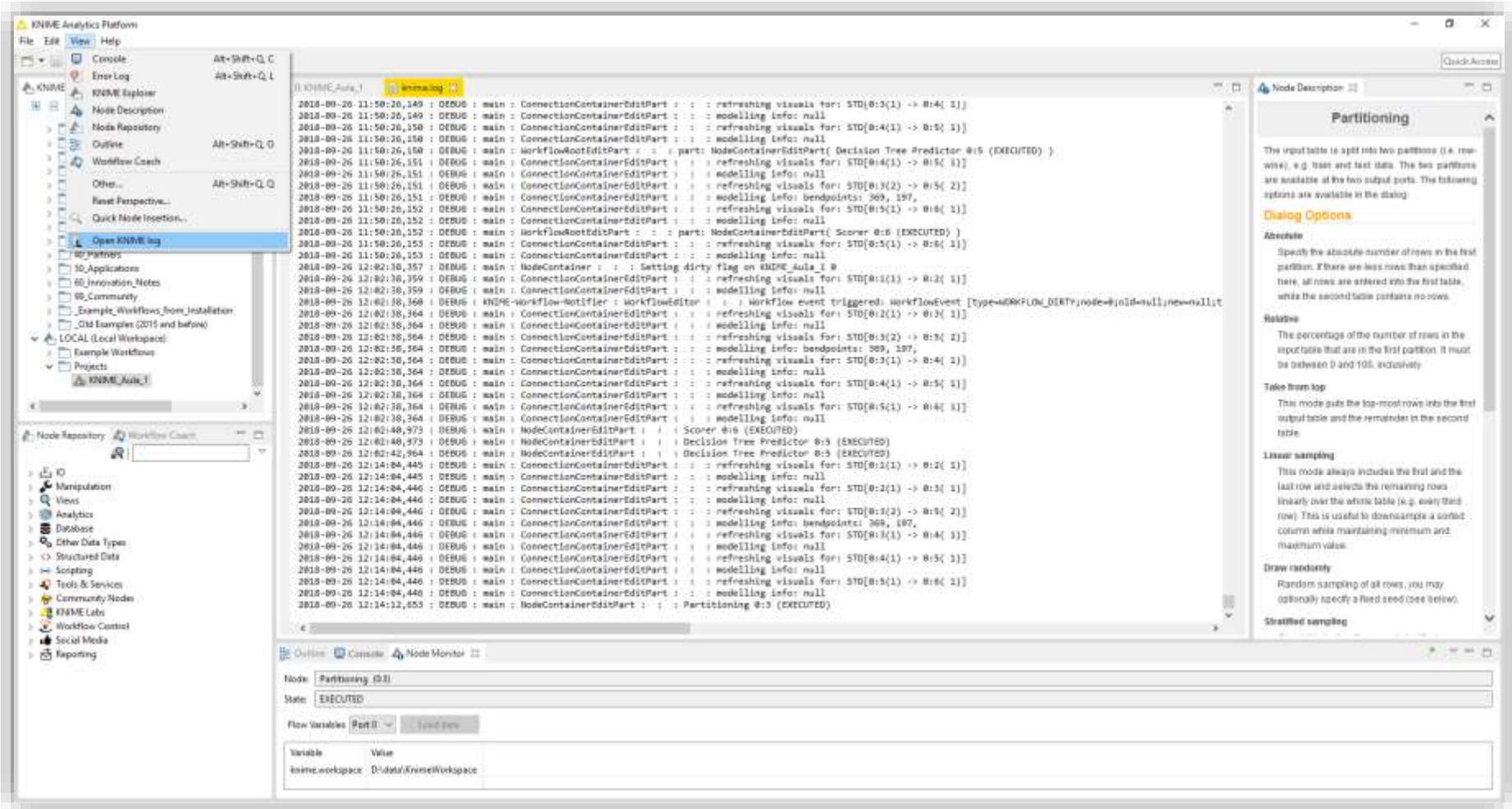
Row ID



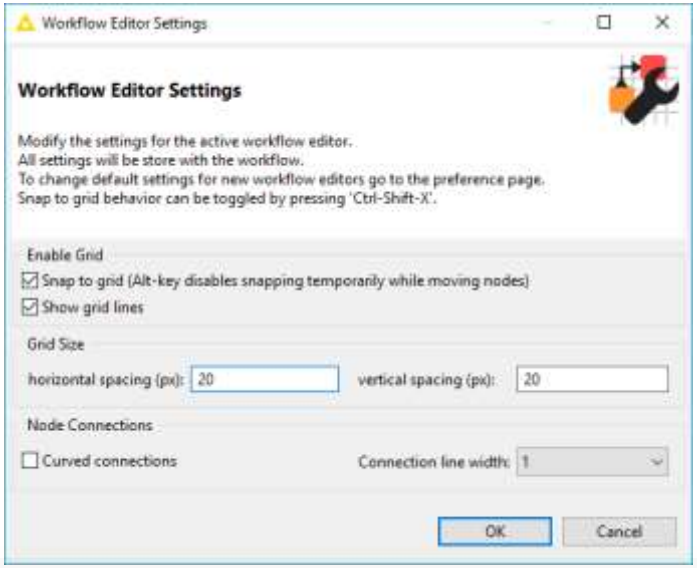
Row ID	fixed a...	volatile ...	citric acid	residual...	chlo						alcohol	quality
Row0	7.4	0.7	0	1.9	0.076							=5
Row1	7.8	0.88	0	2.6	0.098			0.997	3			=5
Row2	7.8	0.76	0.04	2.3	0.092			0.997	3			=5
Row3	11.2	0.28	0.56	1.9	0.075	17	60	0.998	3			=6
Row4	7.4	0.7	0	1.9	0.076	11	34	0.998	3			=5
Row5	7.4	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	9.5	=7
Row9	7.5	0.5	0.36	6.1	0.071	17	102	0.998	3.35	0.8	10.5	=5
Row10	6.7	0.58	0.08	1.8	0.097	15	65	0.996	3.28	0.54	9.2	=5
Row11	7.5	0.5	0.36	6.1	0.071	17	102	0.998	3.35	0.8	10.5	=5
Row12	5.6	0.615	0	1.6	0.089	16	59	0.994	3.58	0.52	9.9	=5
Row13	7.8	0.61	0.29	1.6	0.114	9	29	0.997	3.26	1.56	9.1	=5
Row14	8.9	0.62	0.18	3.8	0.176	52	145	0.999	3.16	0.88	9.2	=5
Row15	8.9	0.62	0.19	3.9	0.17	51	148	0.999	3.17	0.93	9.2	=5
Row16	8.5	0.28	0.56	1.8	0.092	35	103	0.997	3.3	0.75	10.5	=7
Row17	8.1	0.56	0.28	1.7	0.368	16	56	0.997	3.11	1.28	9.3	=5
Row18	7.4	0.59	0.08	4.4	0.086	6	29	0.997	3.38	0.5	9	=4
Row19	7.9	0.32	0.51	1.8	0.341	17	56	0.997	3.04	1.08	9.2	=6
Row20	8.9	0.22	0.48	1.8	0.077	29	60	0.997	3.39	0.53	9.4	=6
Row21	7.6	0.39	0.31	2.3	0.082	23	71	0.998	3.52	0.65	9.7	=5
Row22	7.9	0.43	0.21	1.6	0.106	10	37	0.997	3.17	0.91	9.5	=5
Row23	8.5	0.49	0.11	2.3	0.084	9	67	0.997	3.17	0.53	9.4	=5
Row24	6.9	0.4	0.14	2.4	0.085	21	40	0.997	3.43	0.63	9.7	=6
Row25	6.3	0.39	0.16	1.4	0.08	11	23	0.996	3.34	0.56	9.3	=5

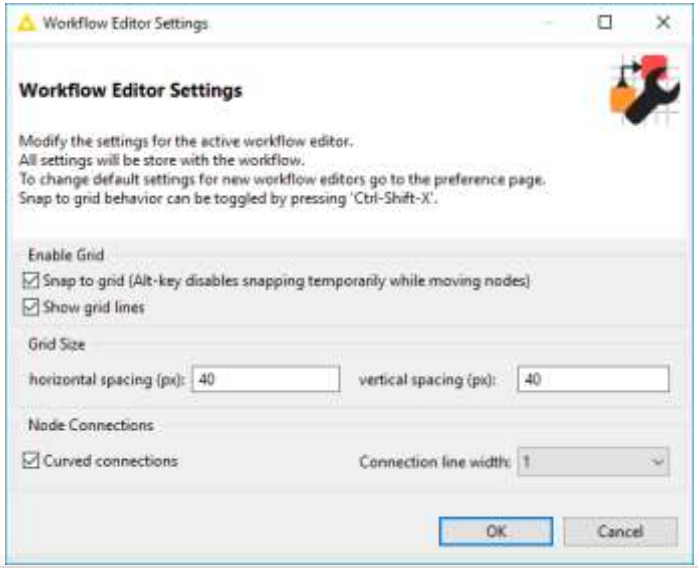
Data Cells

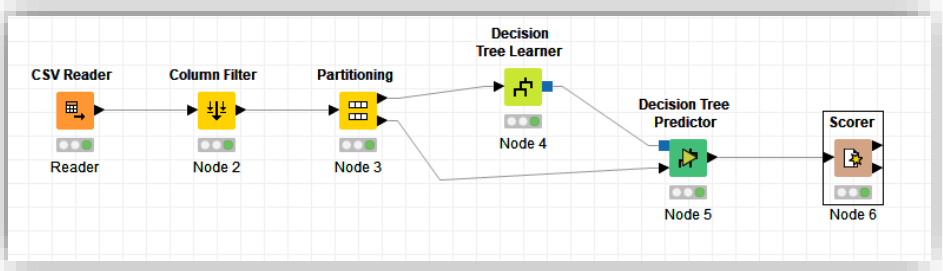
Log View

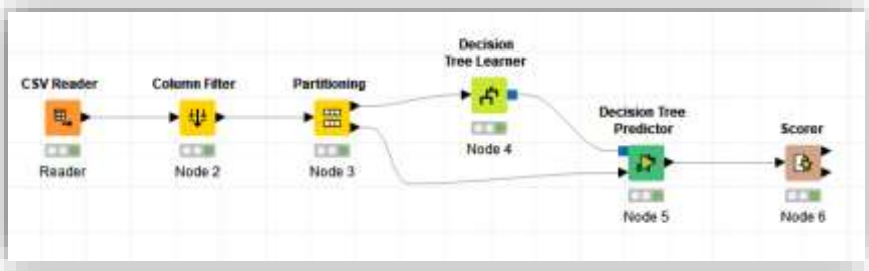

KNIME Analytics Platform
File Edit View Help
Console
Error Log
KNIME Explorer
Node Description
Node Repository
Outline
Workflow Coach
Other...
Reset Perspective...
Quick Node Insertion...
Open KNIME log
Node Repository
IO
Manipulation
Views
Analytics
Database
Other Data Types
Structured Data
Scripting
Tools & Services
Community Nodes
KNIME Labs
Workflow Control
Social Media
Reporting
Node Description
Partitioning
Dialog Options
Absolute
Relative
Take from top
Linear sampling
Draw randomly
Stratified sampling
Node Monitor
Node: Partitioning (0:3)
State: EXECUTED
Flow Variables
Variable
Value
knime.workspace
D:\data\KnimeWorkspace

Views' Customization

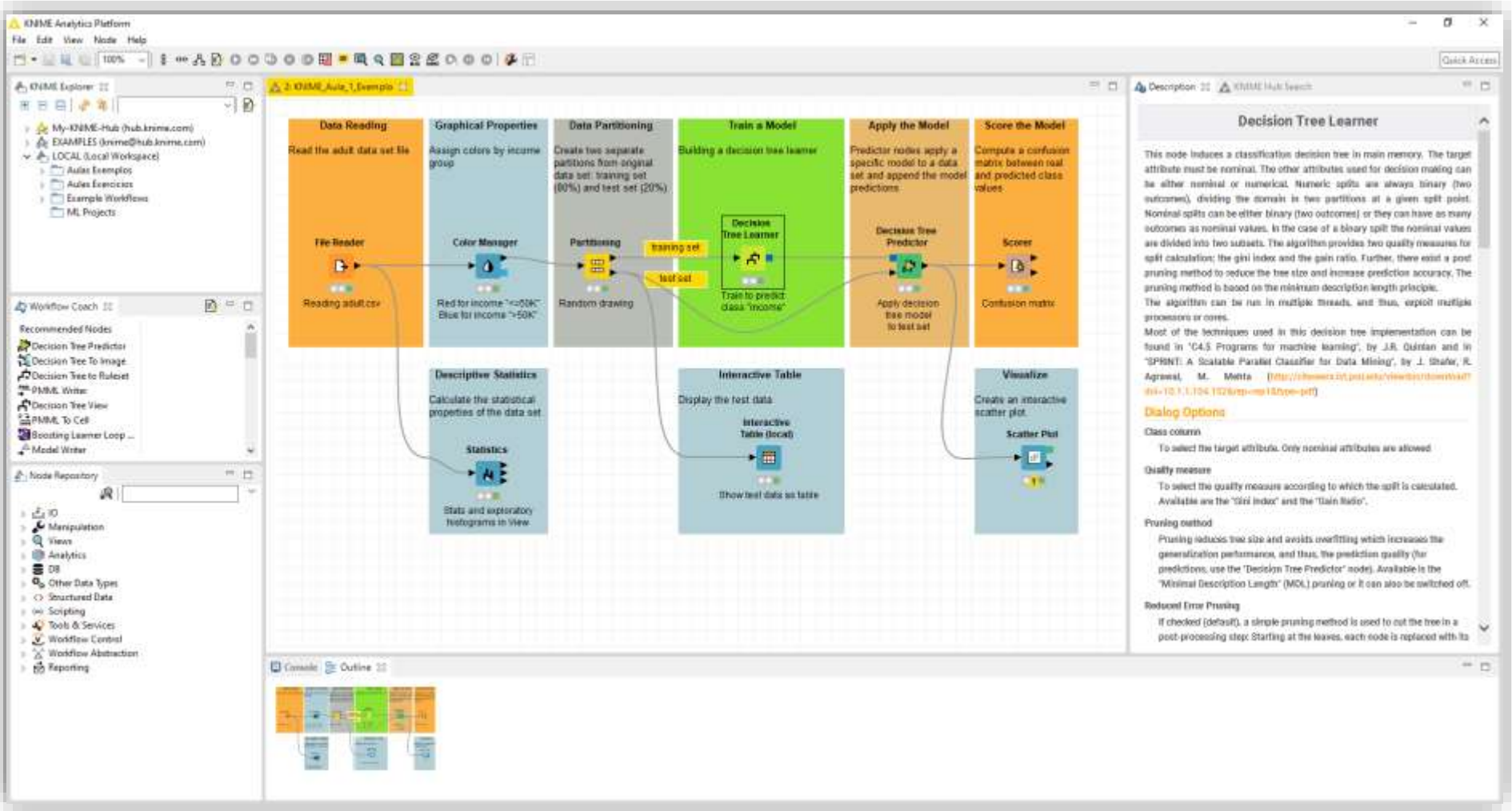




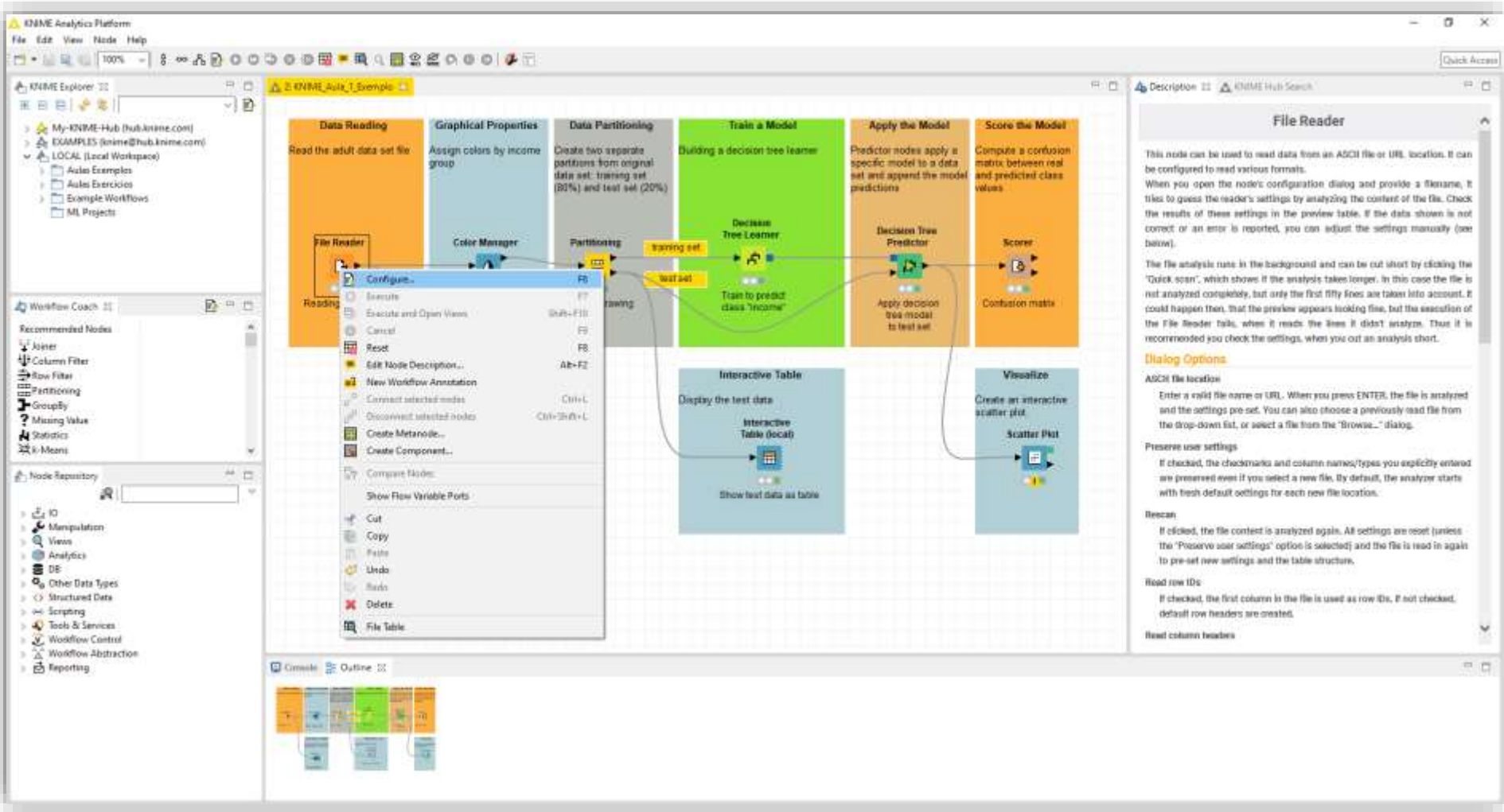




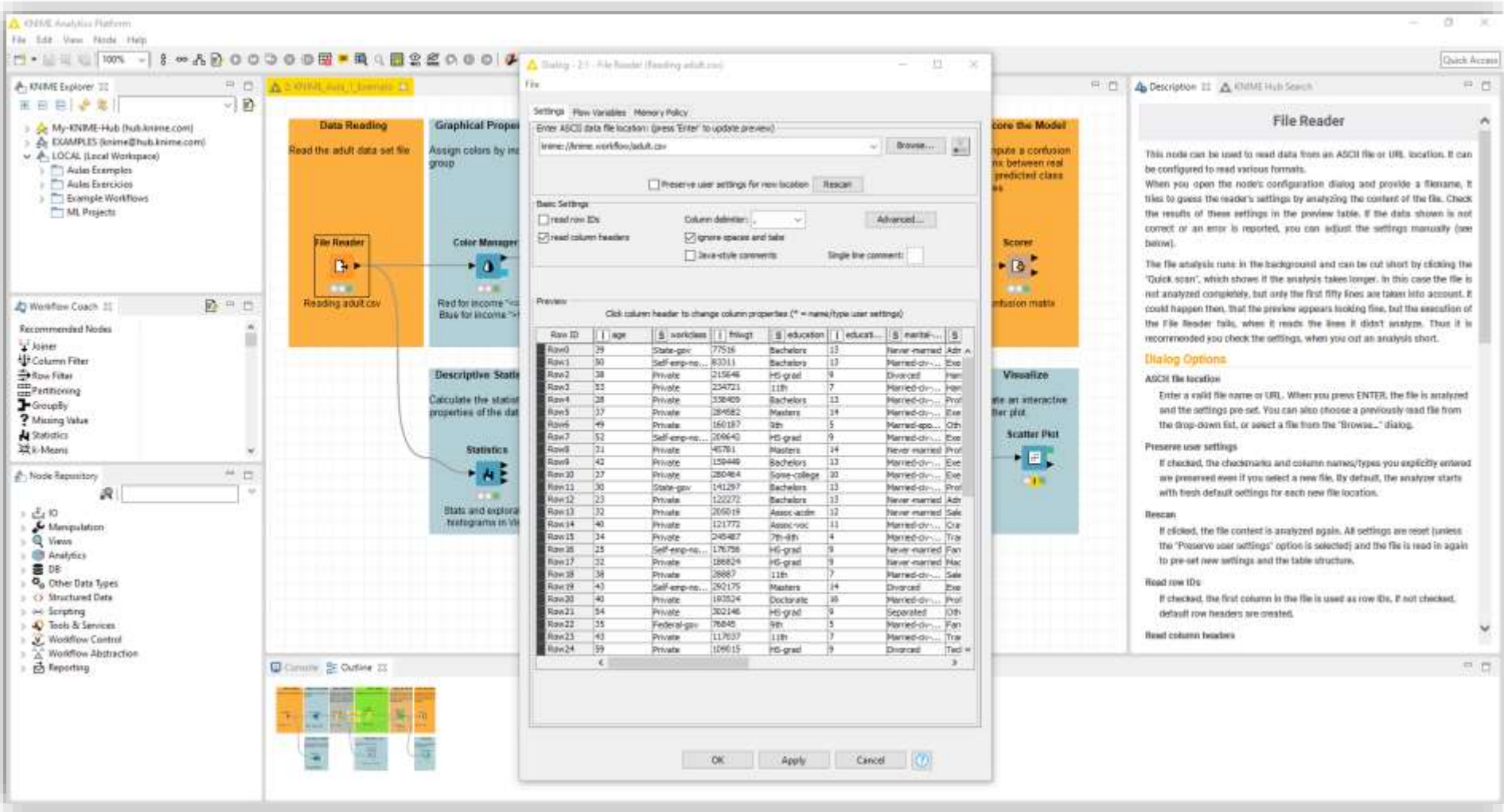
Building a Simple Workflow



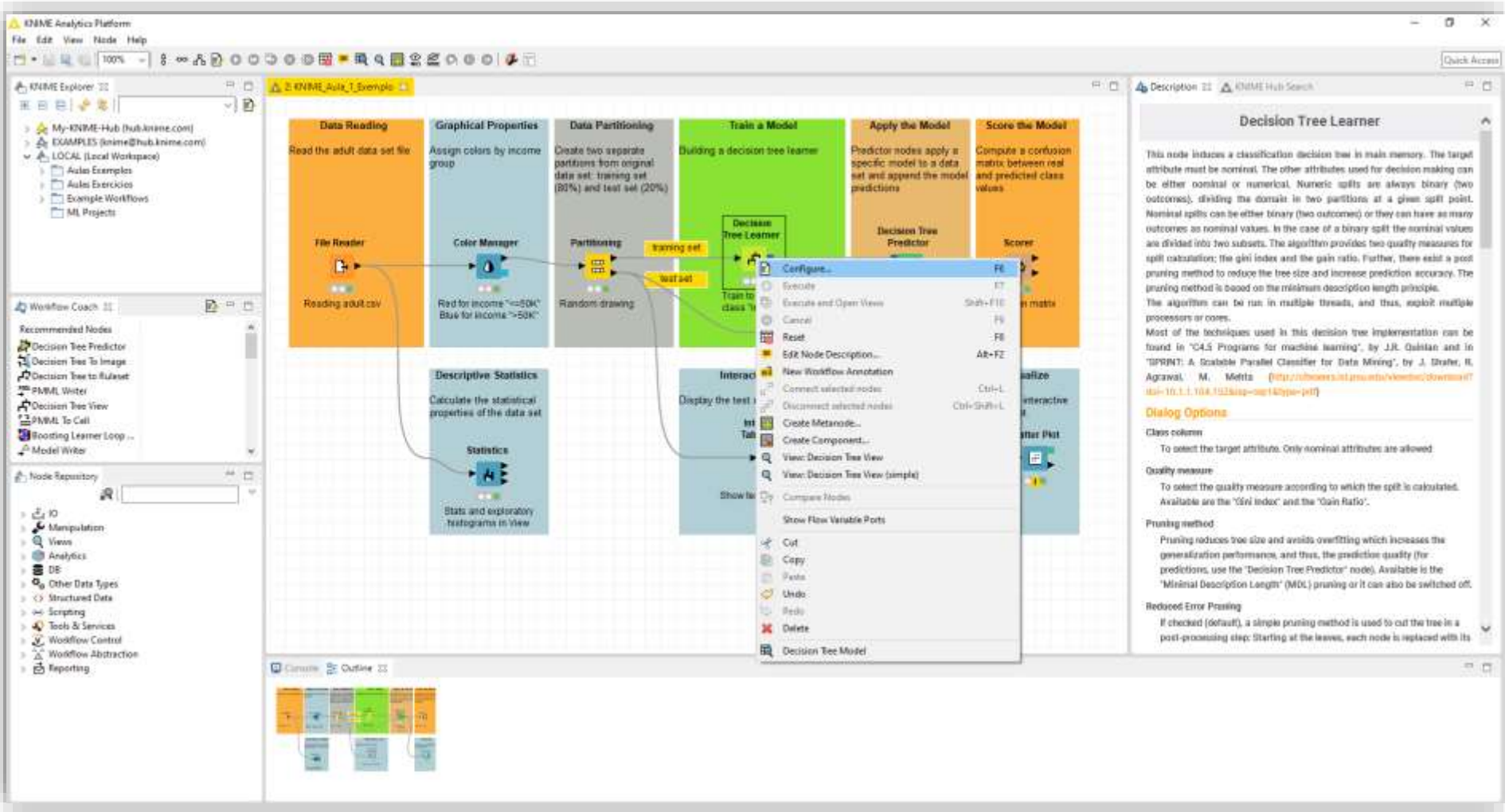
Node Context Options
Data Loader



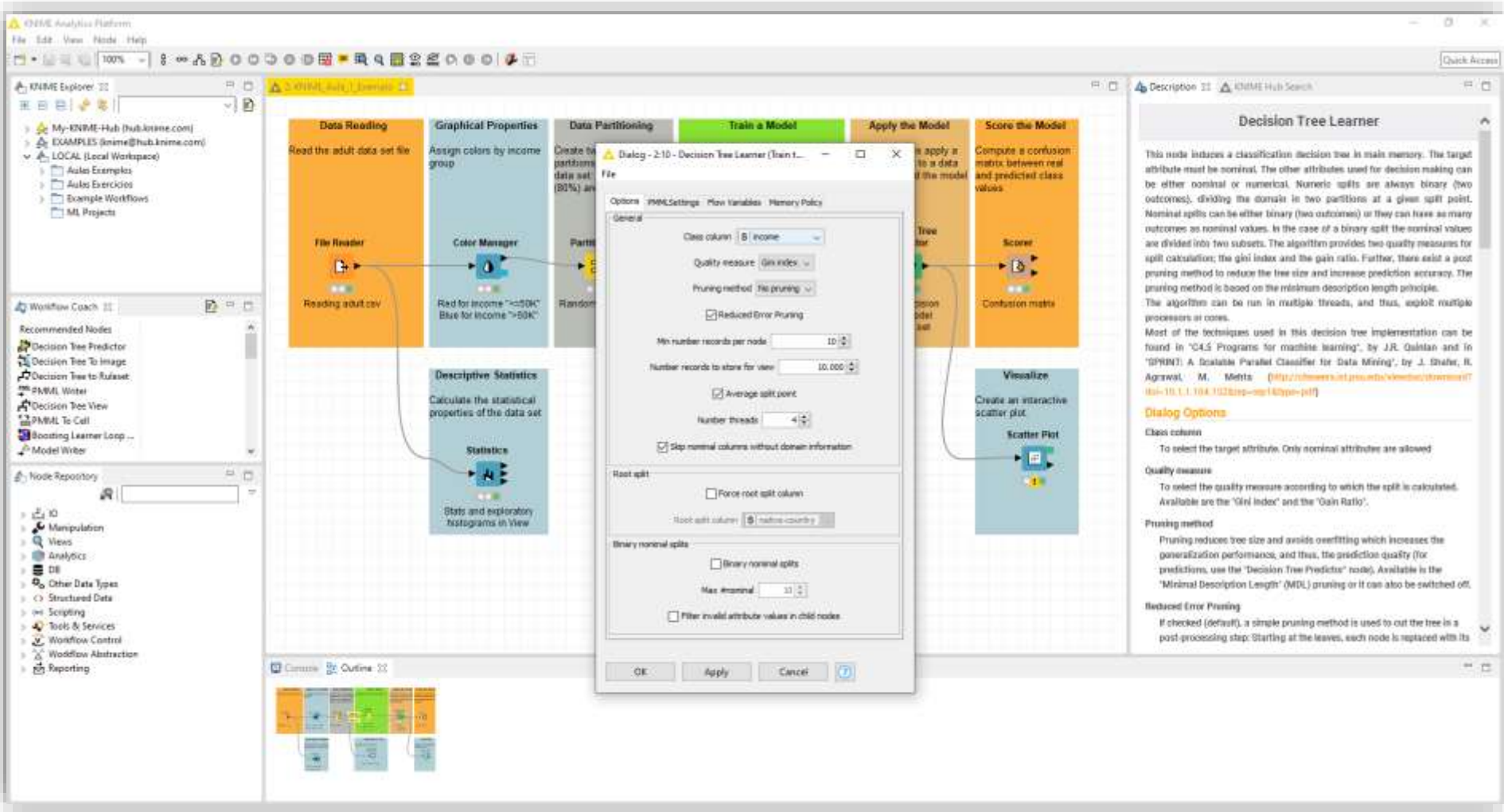
Node Context Options
Data Loader



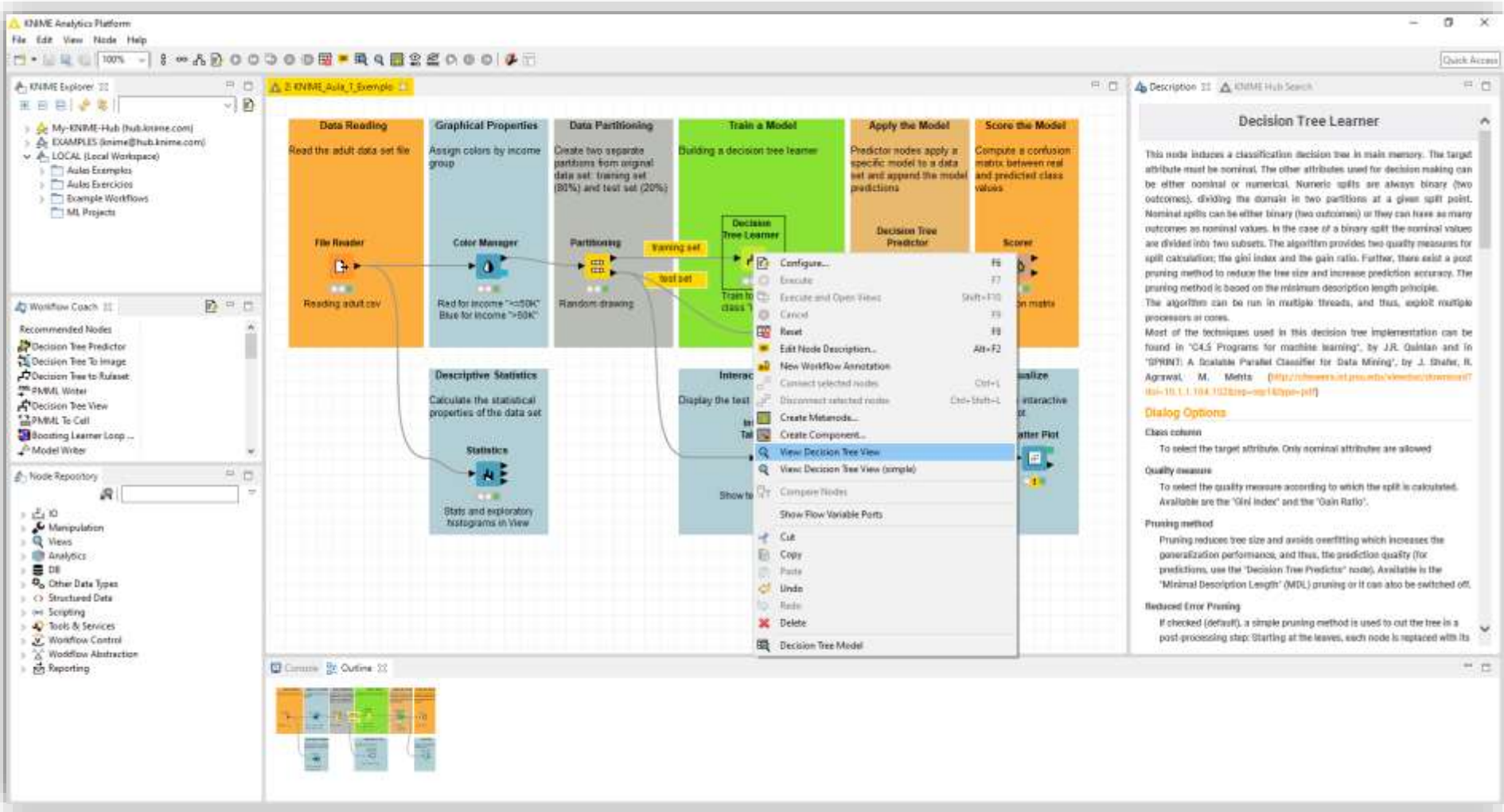
Node Context Options
Model Learner



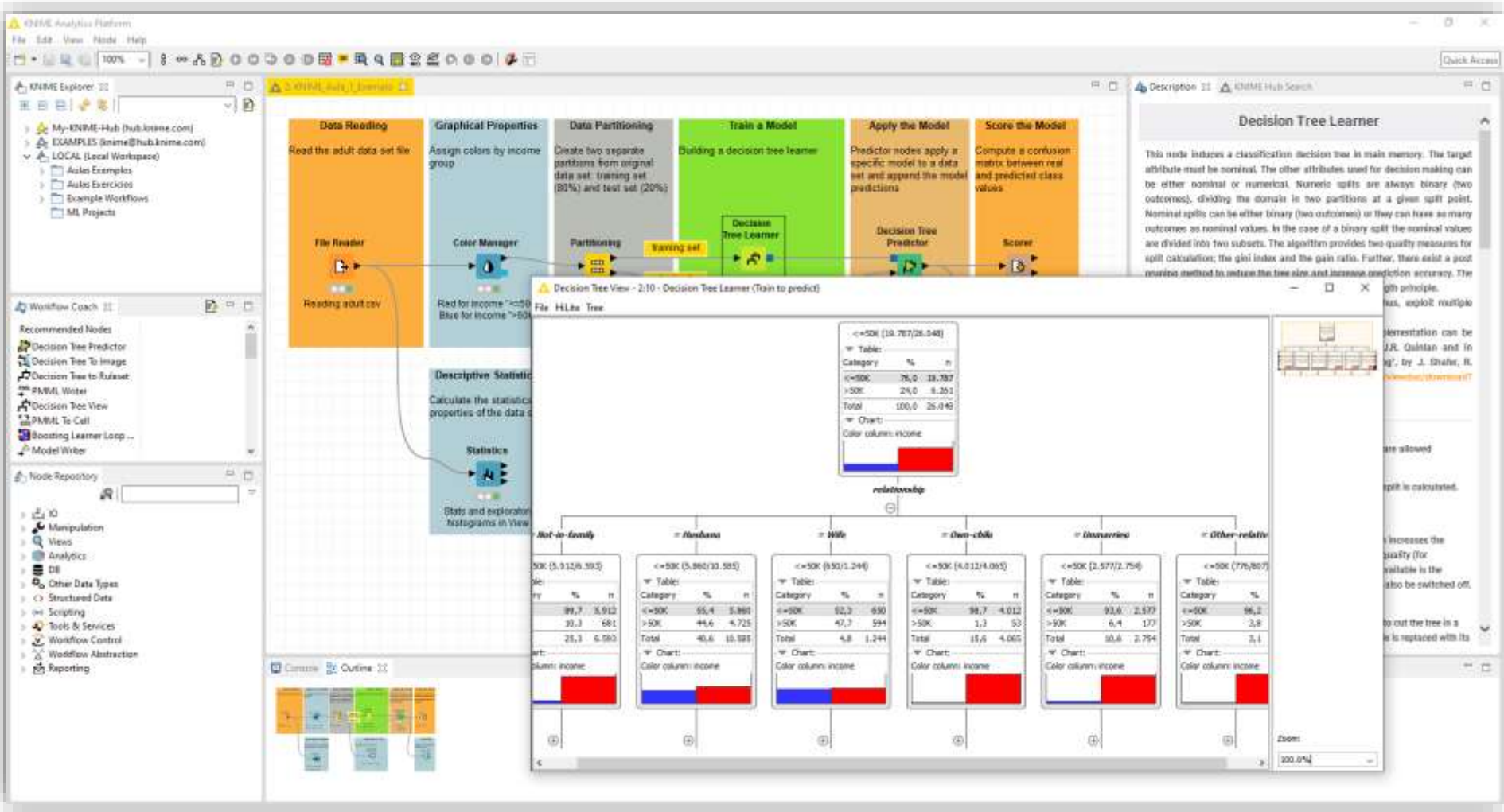
Node Context Options
Model Learner



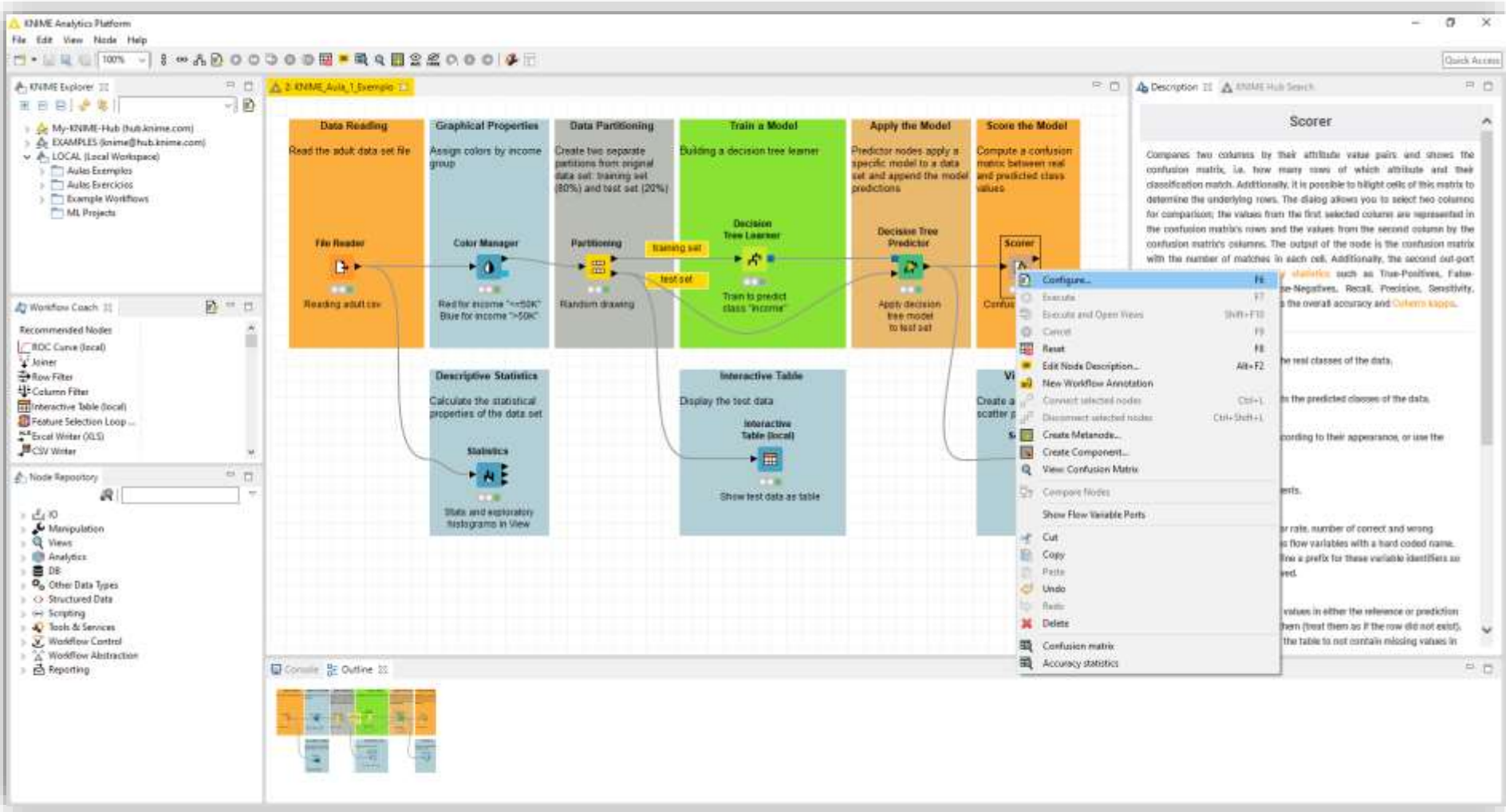
Node Context Options
Model Learner



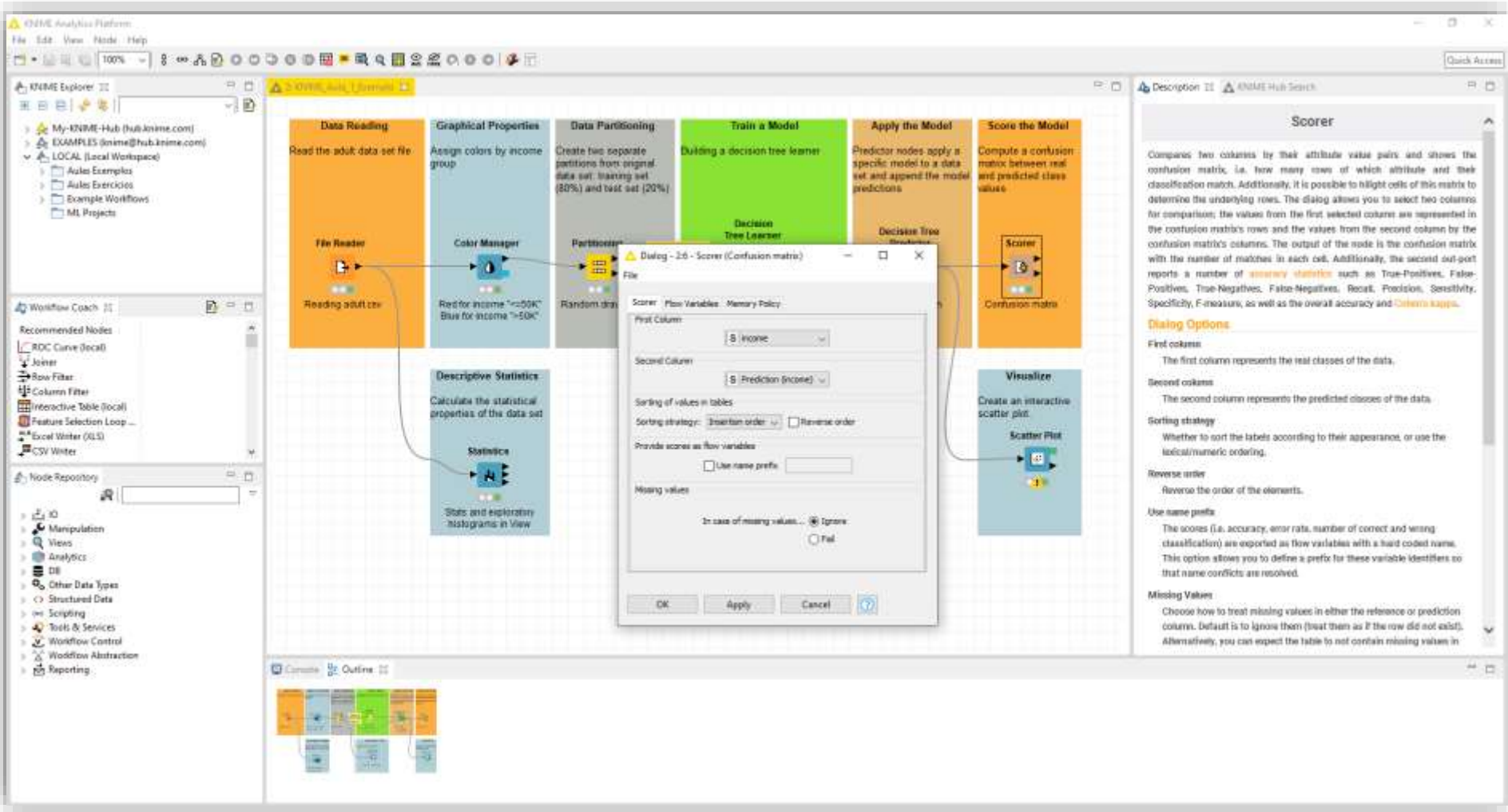
Node Context Options
Model Learner



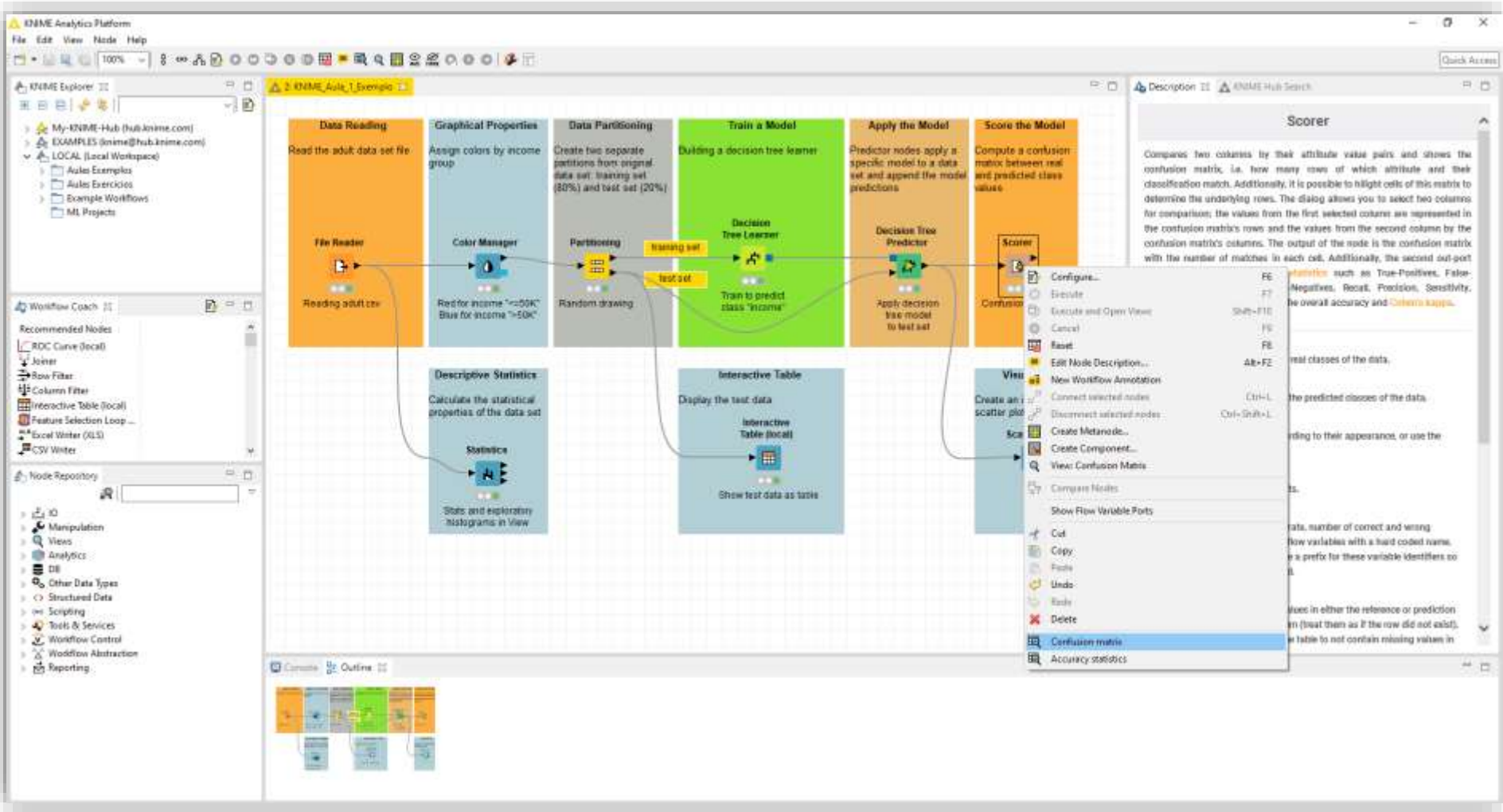
Node Context Options
Scorer



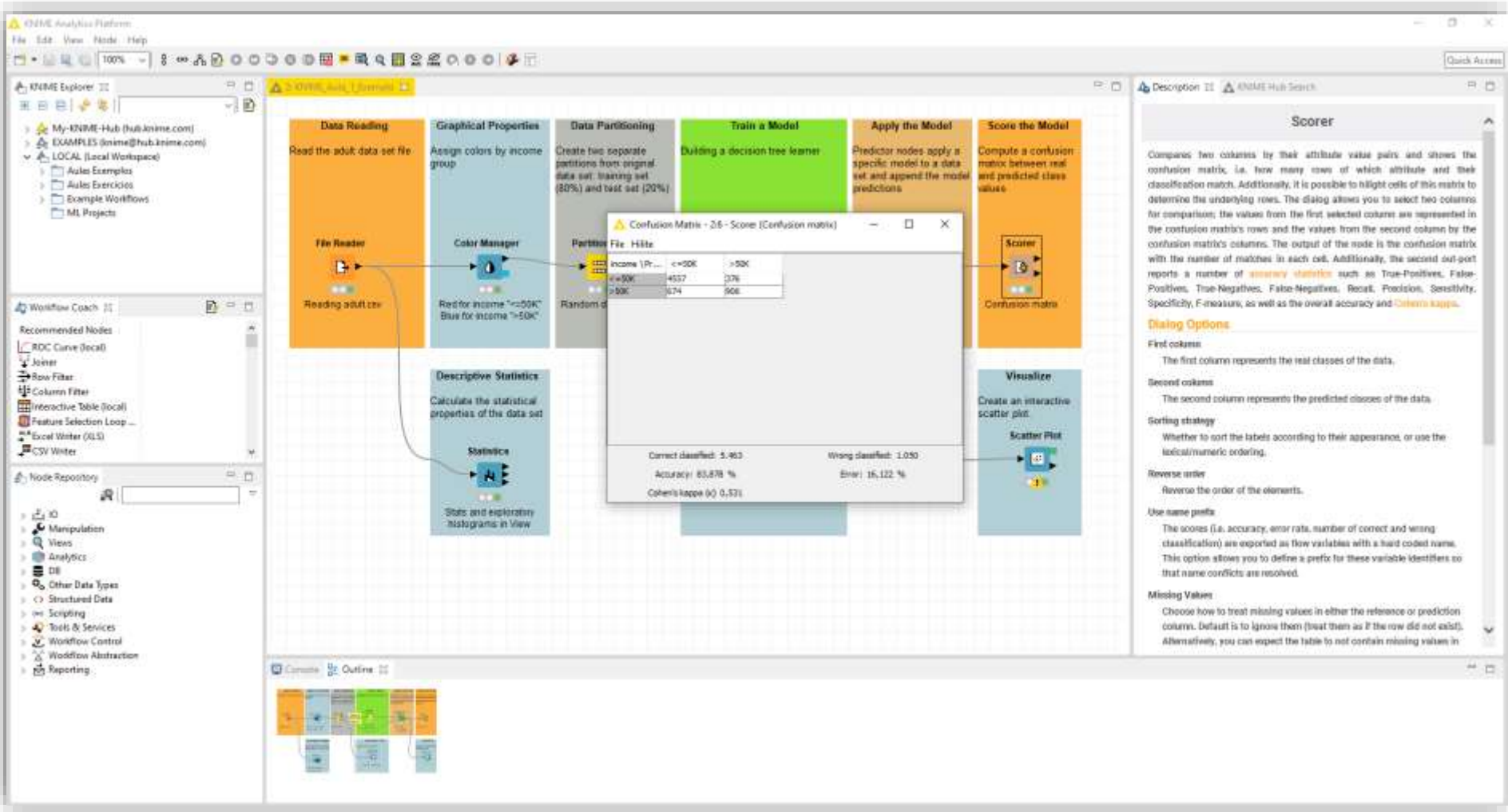
Node Context Options
Scorer



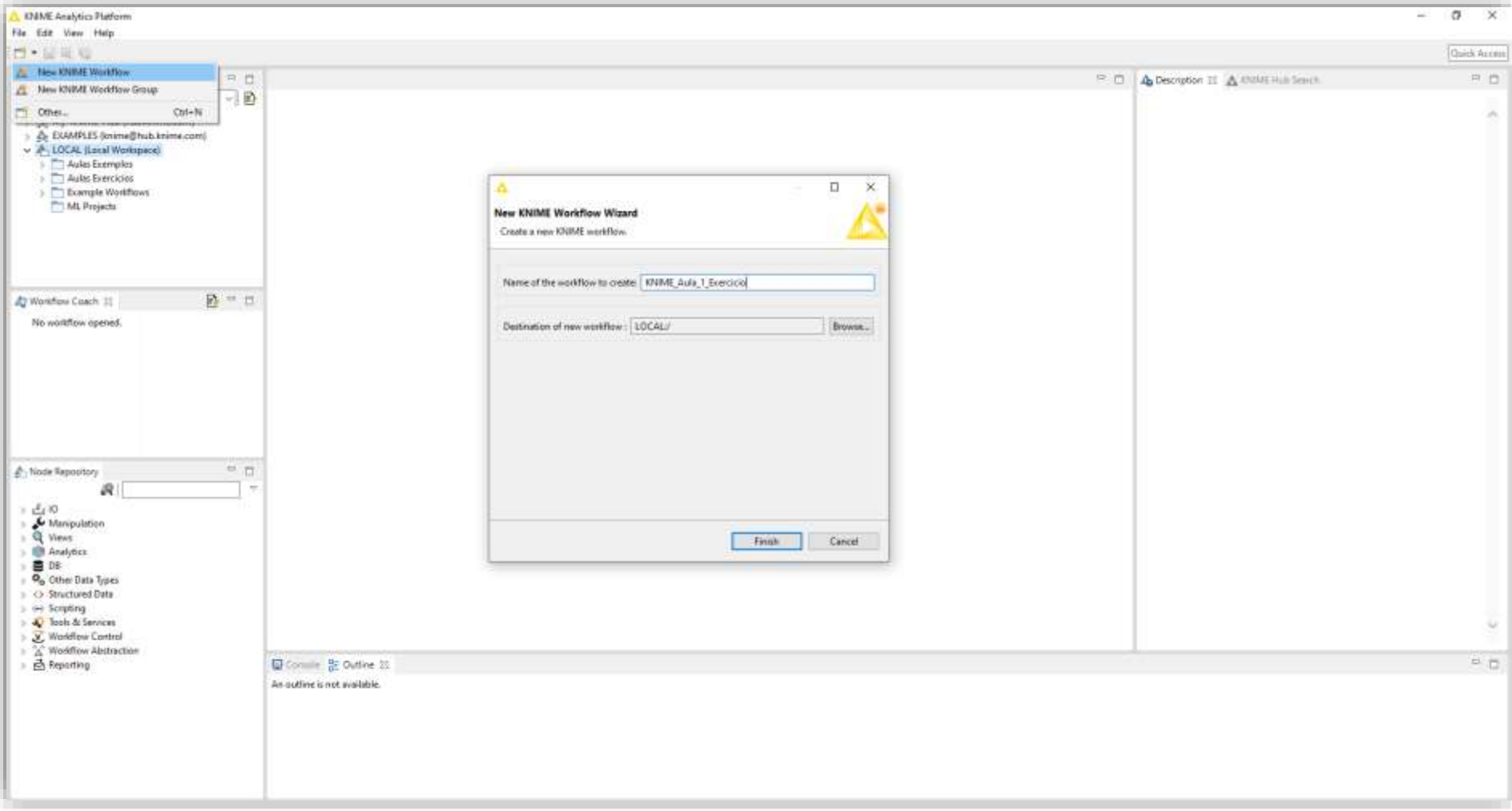
Node Context Options
Scorer



Node Context Options
Scorer



Hands On


KNIME Analytics Platform
File Edit View Help
New KNIME Workflow
New KNIME Workflow Group
Other... Ctrl+N
EXAMPLES (knime@hub.knime.com)
LOCAL (Local Workspace)
Aulas Exemplos
Aulas Exercícios
Example Workflows
ML Projects
Workflow Coach
No workflow opened.
Node Repository
IO
Manipulation
Views
Analytics
DB
Other Data Types
Structured Data
Scripting
Tools & Services
Workflow Control
Workflow Abstraction
Reporting
New KNIME Workflow Wizard
Create a new KNIME workflow.
Name of the workflow to create: KNIME_Aula_1_Exercicio
Destination of new workflow: LOCAL/
Browse...
Finish
Cancel
Description
Console
Outline
An outline is not available.

Hands On


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 Exercícios
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
Node Repository
IO
Manipulation
Views
Analytics
DB
Other Data Types
Structured Data
Scripting
Tools & Services
Workflow Control
Workflow Abstraction
Reporting
KNIME_Aula_Exercicio
Description
KNIME Hub Search
Console
Outline
HANDS ON