

KNIME



- **Modern data analytics platform that enables sophisticated statistical and data mining analysis on small, medium or large data**
- **Visual workbench chains data access, data transformation, initial investigation, powerful predictive analytics and visualization**
- **Provides the ability to develop reports or automate the deployment of new insights back into production systems**
- **KNIME Analytics Platform is open source and available under GPL license**
- **It can be extended with Commercial Software to include professional support, productivity and collaboration functionality**

KNIME Website Links

Web Links

<http://www.knime.org>

This is the first place to look for information about KNIME products. The open source Desktop version can be downloaded here.

<http://www.knime.org/learning-hub>

This is the landing page to learn more specific KNIME functionalities. Learning material - as web sites, videos, webinars, courses, and more - is organized by topic, like text mining or chemistry, or basic KNIME functionalities, etc...

<http://tech.knime.org/forum>

In the www.knime.org site you can find a number of resources. What I find particularly useful is the KNIME Forum. Here you can ask questions about how to use KNIME or about how to extend KNIME with new nodes. Someone from the KNIME community answers always and quickly.

<http://tech.knime.org/knime-labs>

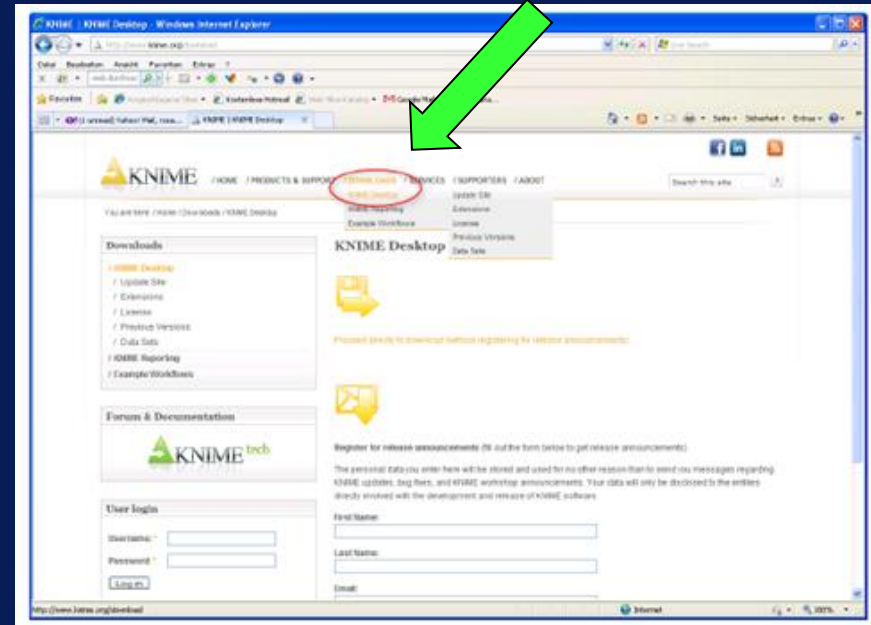
This site contains nodes still under development; i.e. the beta version of new nodes. You can already download them and use them, but they are not of product/release quality yet.

<http://knime.org/supporters-0>

This is the site where all contributing supporters (partners, providers, and sponsors) are listed.

Download KNIME

- Go to www.knime.org
- Select the "Download" tab
- Select "KNIME Desktop" in the submenu
- Select one of the two options (with registration or without registration, as you prefer)
- Choose the version that suits your environment (Windows/Linux, 32 bit/64 bit)
- Accept the terms and conditions
- Start downloading
- You will end up with a zipped or a self-extracting archive file (*.zip or *.exe)
- Unpack your file in the destination folder
- This creates a new subfolder in the selected destination folder named "knime_ 2.X.x" where x represents the minor release



KNIME Download and Installation


Installation

License

Starting with Version 2.1, KNIME is released under the GNU General Public License, Version 3 (including certain additional permissions according to Sec. 7 of the GPL Ver. 3). It is also available - through the dual licensing model - under customized licenses. If you wish to receive KNIME under a different license than the GPL, please contact us at contact@knime.org to discuss licensing arrangements.

Versions

The table describes the KNIME versions and on which platform they are available or not.



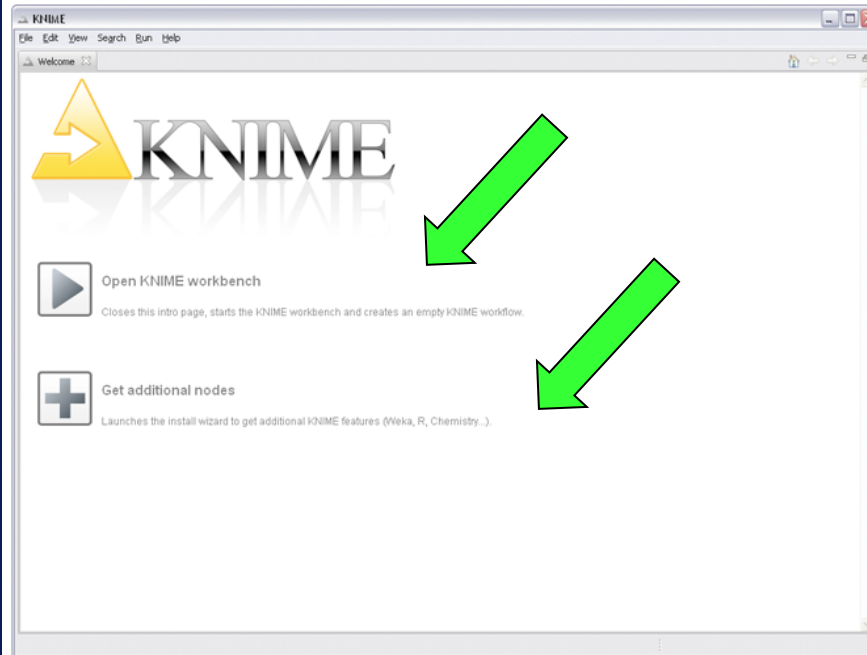
	Linux	Windows	Mac OS X
KNIME (32bit)	yes	yes	no
KNIME (64bit)	yes	yes	yes
KNIME Developer Version (32bit)	yes	yes	no
KNIME Developer Version (64bit)	yes	yes	yes

Installation

Download one of the above versions, unzip it to any directory. For Windows click the *knime.exe* file, and for Linux click *knime* in order to start KNIME.

Welcome Screen / Additional features

When KNIME is started for the first time a welcome screen appears.



From here you can

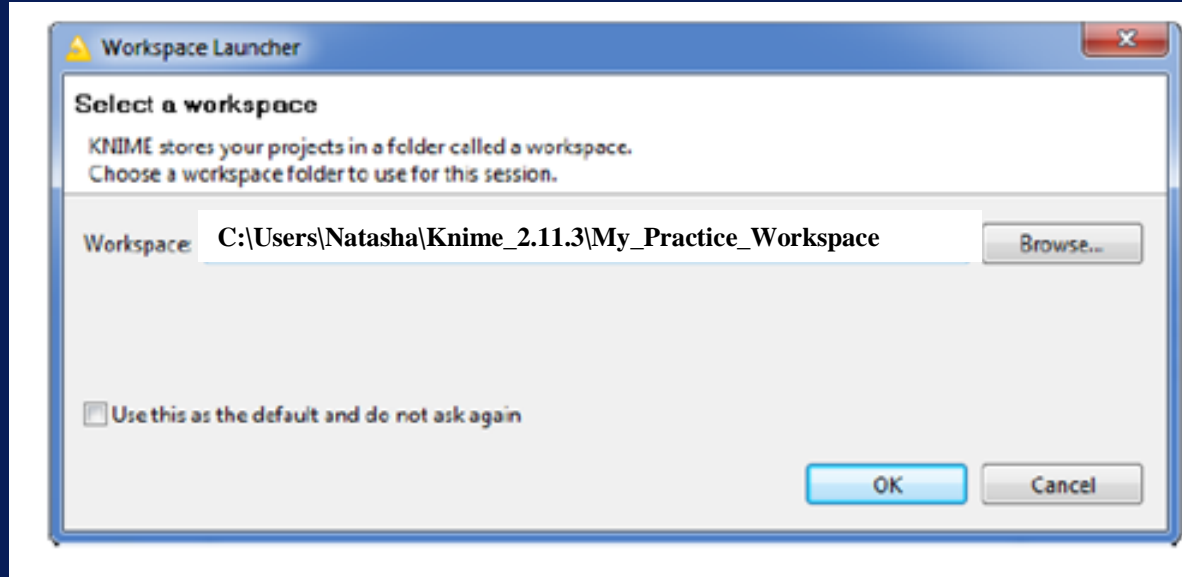
1. Open KNIME workbench: opens the KNIME workbench to immediately start exploring KNIME, build own workflows and explore your data.
2. Get additional nodes: in addition to the ready-to-start basic KNIME installation there are additional plug-ins for KNIME e.g. an R and Weka integration, or the integration of the Chemistry Development Kit with additional nodes for the processing of chemical structures, compounds, etc. You can download these features later from within KNIME (File, Update KNIME...) as well

KNIME Workspace

The Workspace is the directory where all current workflows and preferences are saved

It can be located anywhere (you specify) on the hard drive

By Default it is in the KNIME folder – but you can change that!



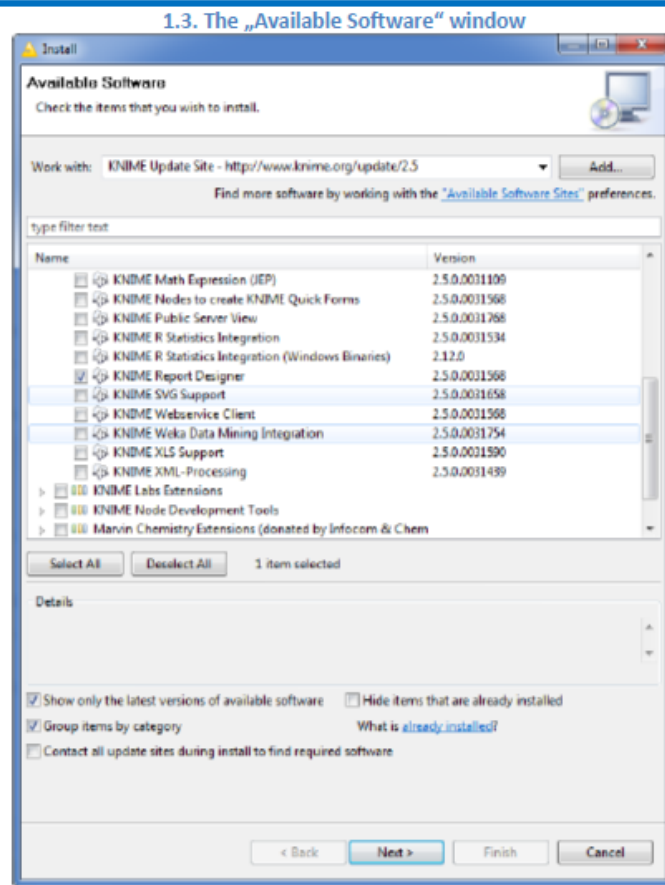
To install a new KNIME extension:

- From the Top Menu, select **"Help" -> "Install New Software"**
- In the "Available Software" window, in the "Work with" textbox, select the URL with the KNIME update site (usually named KNIME Update Site - <http://www.knime.org/update/2.x>)
- Select the extensions that you need. In this example, we chose to install the "KNIME Report Designer" in the category "KNIME & Extensions".
- Click the **"Next"** button on the bottom right and follow the wizard instructions

Once the KNIME extension has been installed and KNIME has been restarted, you should see a new node/category in the "Node Repository" in the KNIME workbench (see a few sections ahead).

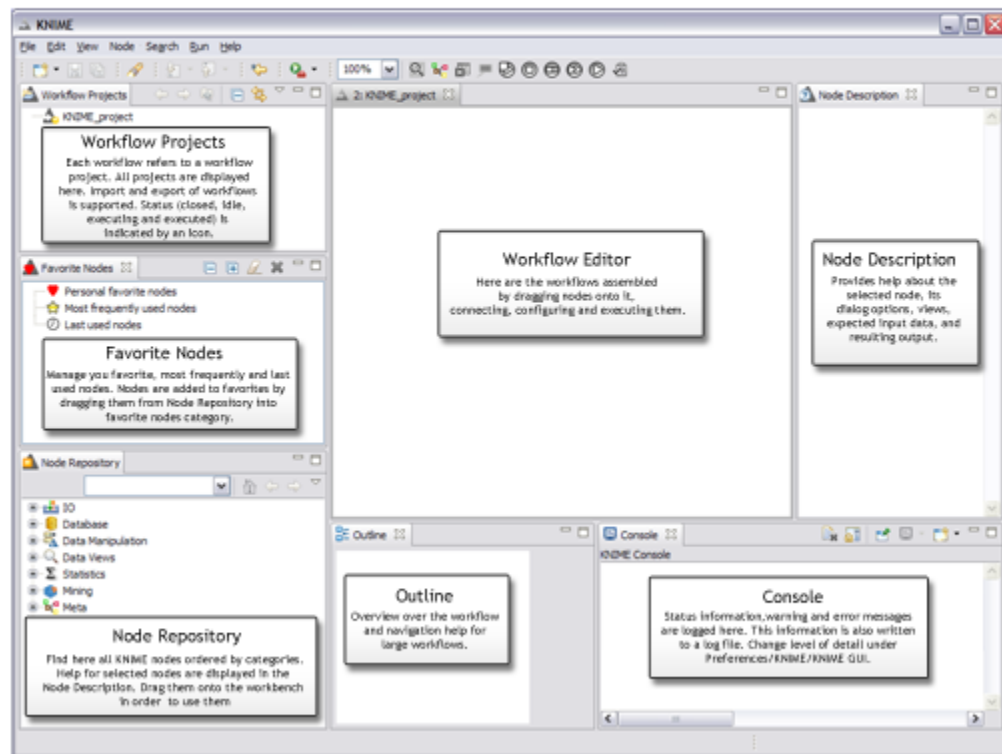
For example, after installing the KNIME Report Designer, you should see a category "Reporting" in the "Node Repository" panel on the left.

Another way to reach the "Available Software" window and then install a new KNIME extension is the following: from the Top Menu, select **"File" -> "Install KNIME Extensions"**.



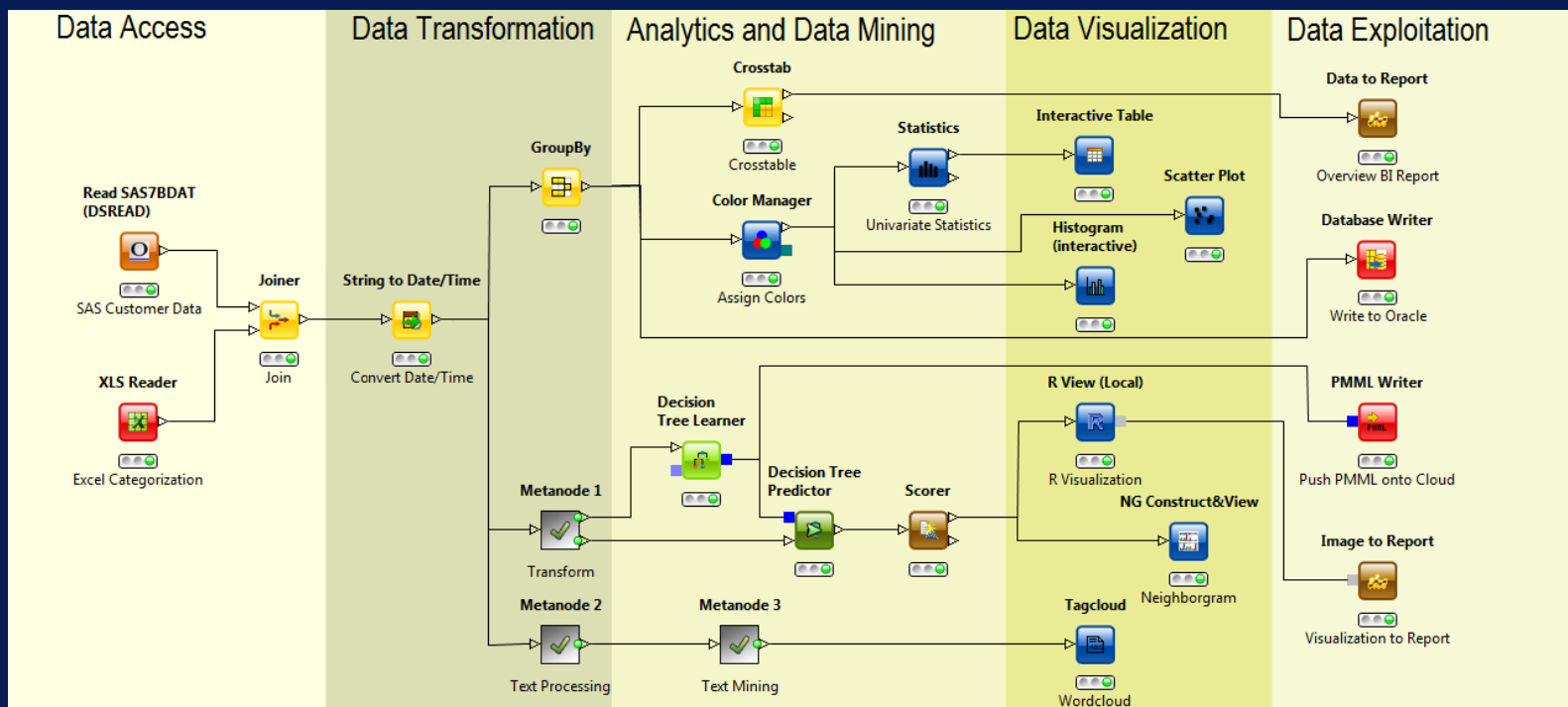
Workbench overview

The KNIME Workbench is organized as follows:



The image shows the KNIME software interface with several components labeled:

- List of Workflow Projects:** Located in the top-left pane, showing a tree view of projects including Chapter2, Exercises, my_first_workflow, Chapter3, Chapter4, Chapter5, and KNIME_project.
- Node Repository:** Located in the bottom-left pane, showing a list of nodes categorized by IO, Database, Data Manipulation, Data Views, Statistics, Mining, Chemistry, Distance Matrix, Meta, Flow Control, Misc, KNIME Labs, Time Series, Quick Form, R, Reporting, Weka, and XML.
- Workflow Editor:** The central workspace showing a workflow diagram. The workflow starts with a **File Reader** node (adult.data set), followed by a **Column Filter** node (rm "final weight"), then a **Row Filter** node (born outside US), and finally a **CSV Writer** node (write new file). A text box below the workflow states: "This workflow is my first KNIME workflow. It reads data, removes uninteresting columns and rows from the data table and writes the result to a CSV file."
- Node Description:** A panel on the right showing the description for the **File Reader** node. It states: "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..."
- KNIME Server:** A panel on the right showing the status of the KNIME Server. It displays "Workflow Server: publicserver.knime" and "Status: not connect" with a "Connect" button.
- Outline:** A panel at the bottom left showing a small overview of the workflow diagram.
- Console:** A panel at the bottom right showing the KNIME Console output. It displays a welcome message for KNIME v2.4.1.0030379, copyright information, and a warning message: "WARN CSV Writer Selected output file exists and will be overwritten!"



Many Available Data Nodes

Read

- File Reader
- ARFF Reader
- CSV Reader
- Line Reader
- Table Reader
- PMML Reader
- Model Reader
- Read XLS Sheet Names
- XLS Reader

Write

- CSV Writer
- ARFF Writer
- Table Writer
- PMML Writer
- Model Writer
- XLS Sheet Appender
- XLS Writer

Other

- Modular Data Generation
- Data Generator
- Create Table Structure
- Extract System Properties
- Extract Context Properties
- Image Column Writer
- Image Port Writer
- List Files
- Read PNG Images
- Table Creator

Many Available Community Nodes

- Community Nodes
 - KNIME Image Processing
 - CDK
 - DYMATRIX Customer Intelligence
 - EMBL-EBI
 - Groovy Scripting
 - HCS Tools
 - Matlab Scripting
 - NGS
 - Palladian
 - Python Scripting
 - R Scripting
 - RDKit
 - REST Nodes
 - Talete
 - Vernalis
 - CheS-Mapper

Many Statistics and Data Mining Nodes

Statistics

- Hypothesis Testing
 - Single sample t-test
 - Independent groups t-test
 - Paired t-test
 - One-way ANOVA
- Regression
 - Linear Regression Learner
 - Polynomial Regression Learner
 - Logistic Regression Learner
 - Regression Predictor
- Linear Correlation
- Correlation Filter
- Statistics
- Crosstab
- Value Counter

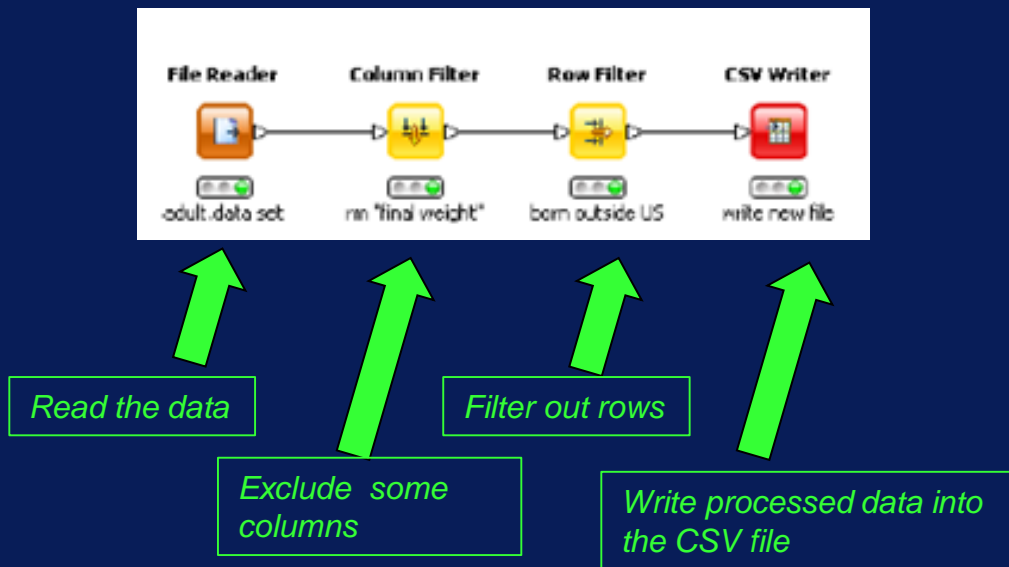
Mining

- Bayes
- Clustering
- Rule Induction
- Neural Network
- Decision Tree
 - Decision Tree Learner
 - Decision Tree Predictor
 - Decision Tree To Image
- Misc Classifiers
- Ensemble Learning
 - PMML Ensembles
 - Utility Nodes
- Bagging
- Boosting Learner
- Boosting Predictor
- Delegating

- Item Sets / Association Rules
- MDS
- PCA
- SVM
 - LIBSVM
 - SVM Learner
 - SVM Predictor
- Scoring
 - Enrichment Plotter
 - Entropy Scorer
 - Numeric Scorer
 - ROC Curve
 - Scorer
- Meta
 - Cross Validation
 - Feature Selection

Workflows in KNIME

- Sequence (pipeline) of the analysis steps
- Example
 - Read Data
 - Clean Data
 - Train a model
 - Etc.

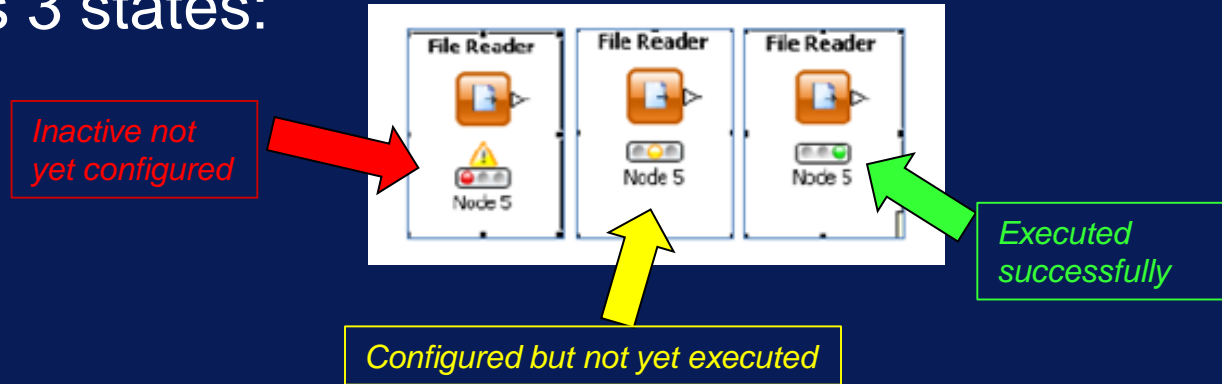


Building a Workflow

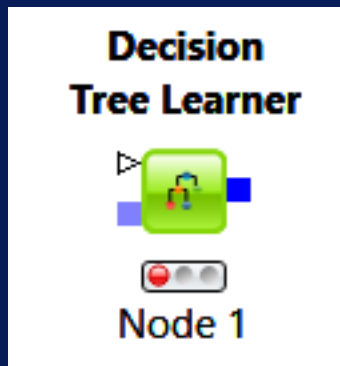
- A workflow is built by dragging nodes from the Node Repository to the Workflow Editor and connecting them
- Nodes are the basic processing units of a workflow
- Each node has a number of input- and/or output ports
- Data (or a model) is transferred via a connection from an out-port to the in-port of another node

Nodes in a Workflow

- Node is a single processing unit in the workflow
- The node takes data from the input, processes it and delivers it in the output port
- The processing can be anything from data manipulation, plotting to training Machine learning algorithms
- Every node has 3 states:



Decision Tree Node



Dialog - 0:1128 - Decision Tree Learner

File

Options PMMLSettings Flow Variables Memory Policy

General

Class column S Activity ▼

Quality measure Gini index ▼

Pruning method No pruning ▼

☒ Reduced Error Pruning

Min number records per node 2

Number records to store for view 10,000

☒ Average split point

Number threads 4

☒ Skip nominal columns without domain information

Binary nominal splits

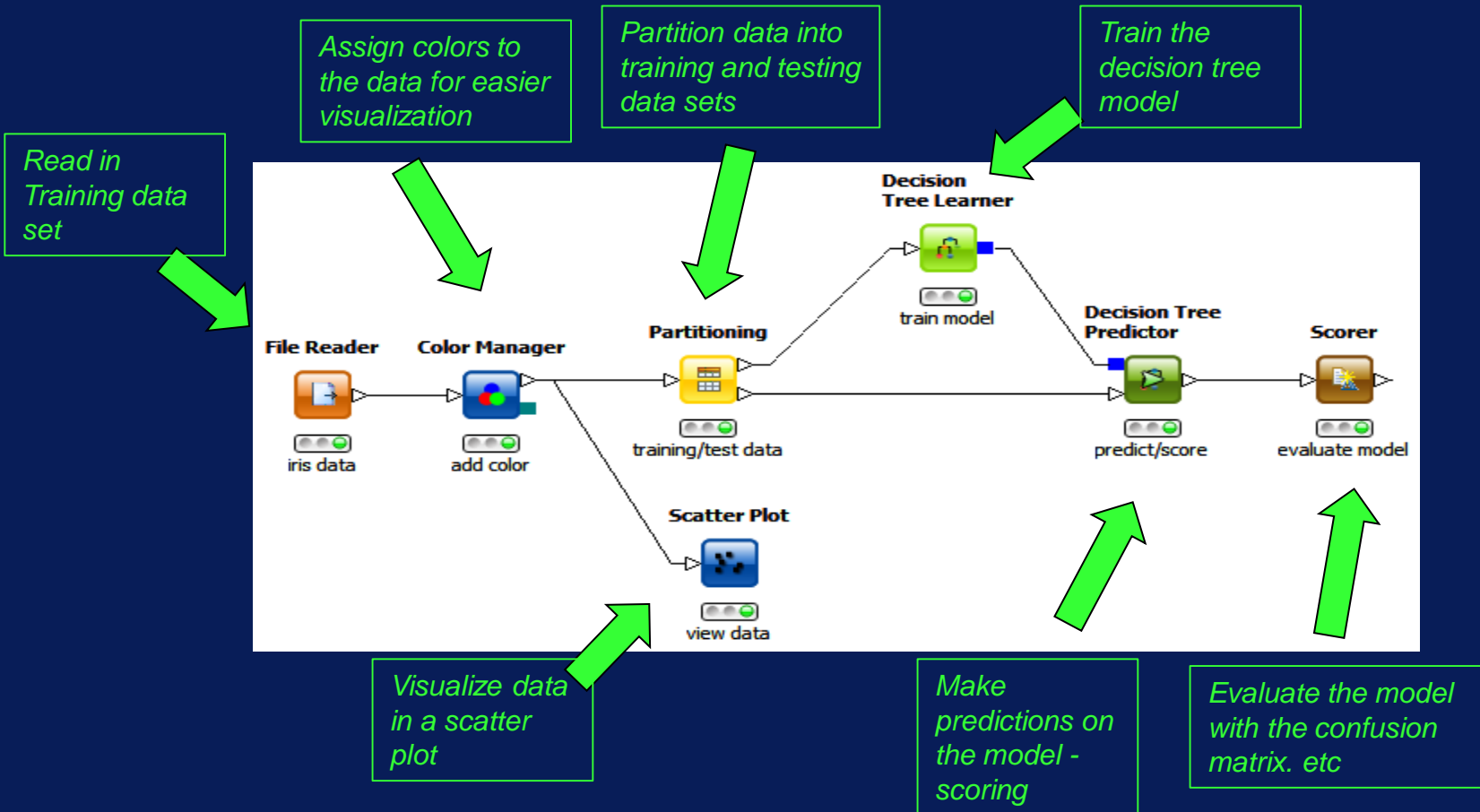
☐ Binary nominal splits

Max #nominal 10

☐ Filter invalid attribute values in child nodes

OK Apply Cancel ?

Decision Tree Workflow Example



Additional Information

<https://www.knime.org/learning-hub?src=knimeapp>

Learning Hub

If you would like to learn more about KNIME, please check out our learning materials. We have made a large range of materials available on a variety of different media channels: books, whitepapers, the KNIME TV channel on Youtube, and of course the KNIME courses and webinars.

Usage

Applications

Development

Basic KNIME

KNIME Web site:

- Installation Guide
- KNIME Interactive Views
- List of KNIME Nodes and Features
- KNIME Extensions on Labs

Books:

- KNIME Press: *KNIME Beginner's Luck*
- PRACT Publishing: *KNIME Essentials*

KNIME TV on Youtube:

- Playlist "KNIME Starter Kit"
- Playlist "Basic KNIME"
- New Database Connectors (KNIME 2.10) (including DB Connectors to Hadoop Hive)
- New Database Nodes and Big Data Extension Nodes in KNIME 2.11
- Improvements in Excel nodes as part of the "What's New" talk at UGM 2014

Courses and Webinars:

- Basic User Training
- KNIME Online Self-Training

Advanced KNIME

Reporting

KNIME Server