

Business Analytics and Data-Driven Decision Making

Text Analytics - 03 & 04: KNIME Introduction and Hands- on Session

Raghava Mukkamala

Associate Professor & Director,
Centre for Business Data Science

Copenhagen Business School, Denmark

Email: rrm.digi@cbs.dk, Centre: <https://cbsbda.github.io/>

Slides based on KNIME Analytics Platform

Source: <https://www.knime.com/sites/default/files/2021-07/slides-l1-ds.pdf>



Outline

- KNIME Analytics Platform for Data Scientists: Basics
- Hands-on Sessions:
 - Sentiment Analysis using KNIME on Diabetes Discussion Forums
 - Topic Modeling Using Online IMDB Movie Reviews

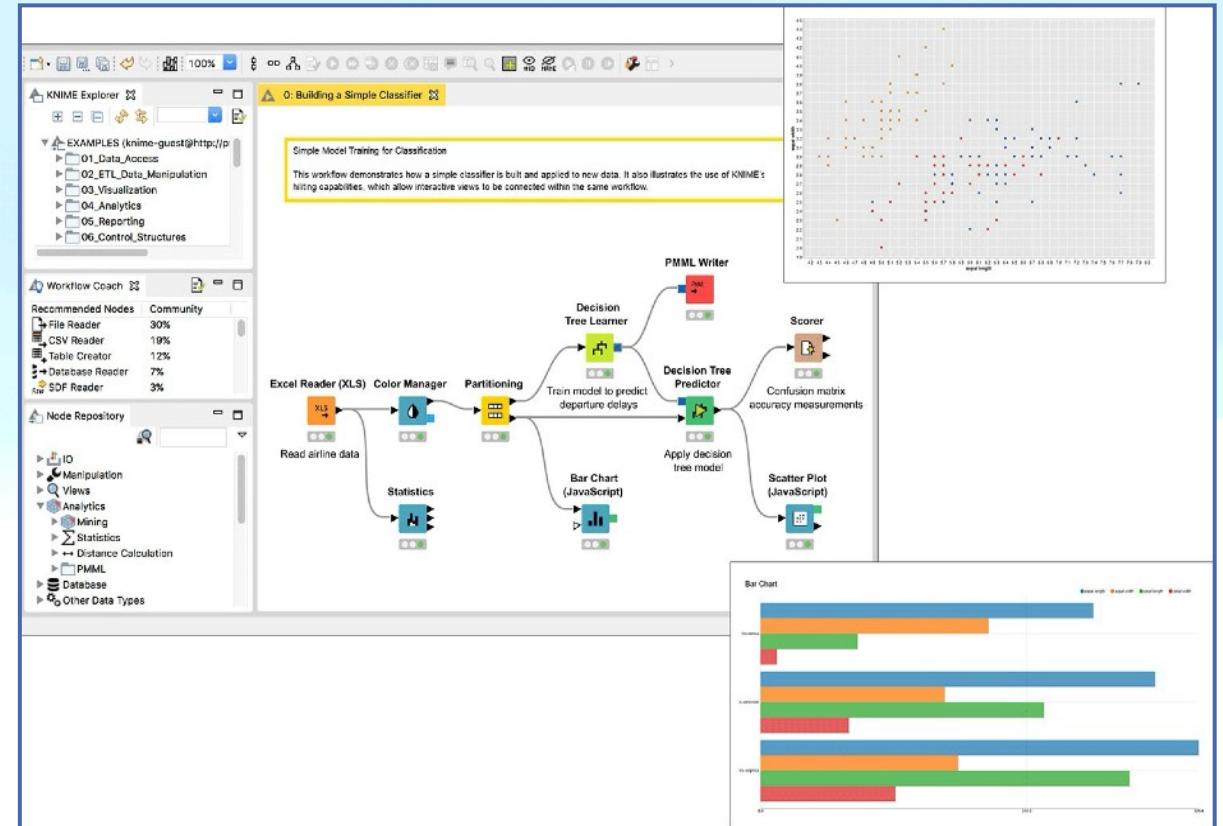
OVERVIEW: KNIME ANALYTICS PLATFORM



The slides are taken and adapted from:
KNIME AG: [L1-DS] KNIME Analytics Platform for Data Scientists: Basics
<https://www.knime.com/sites/default/files/2021-07/slides-l1-ds.pdf>

What is KNIME Analytics Platform?

- A tool for data analysis, manipulation, visualization, and reporting
- Based on the graphical programming paradigm
- Provides a diverse array of extensions:
 - Machine learning,
 - Text Mining,
 - Network Mining



KNIME Classic look (before 5.0 version)

KNIME Analytics Platform 5.0 - Modern look

KNIME Analytics Platform 5

Get started with KNIME Analytics Platform 5

Open for Innovation
KNIME

Examples

- Combine Clean and ...
- Countf and Sumf
- Non-standard format...

Find more resources on the KNIME Community Hub

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 Analytics Platform - /Users/raghava/knime-workspace-50

Home KNIME_project-Sample-01

Execute all Cancel all Reset all 100%

Repository

Search Nodes

ID

- Excel Reader
- Excel Writer
- Microsoft Authenticator
- Google Authenticator
- Google Sheets Reader
- Google Sheets Writer
- CSV Reader
- CSV Writer
- Show all

Manipulation

- Row Filter
- Column Filter
- Concatenate
- Value Lookup
- Row Aggregator
- Table Splitter
- Table Cropper
- Cell Extractor
- Show all

Views

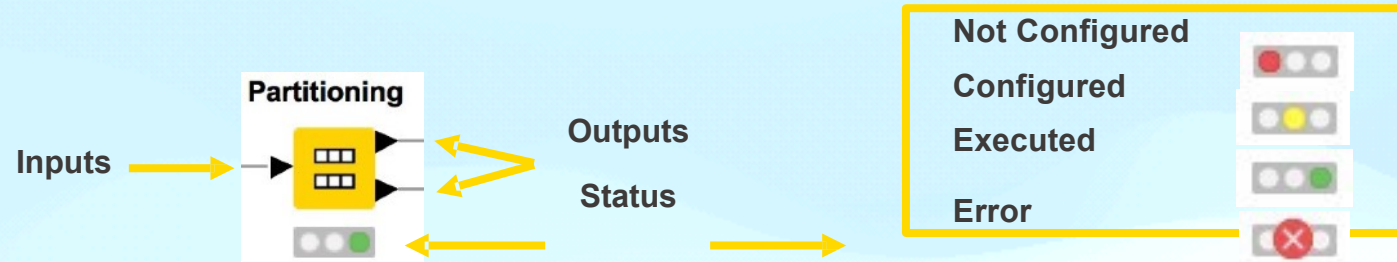
- Bar Chart
- Line Plot
- Pie Chart
- Stacked Area Chart
- Scatter Plot
- Statistics
- Heatmap
- Histogram
- Show all

Start building your workflow by dropping your data or nodes here.

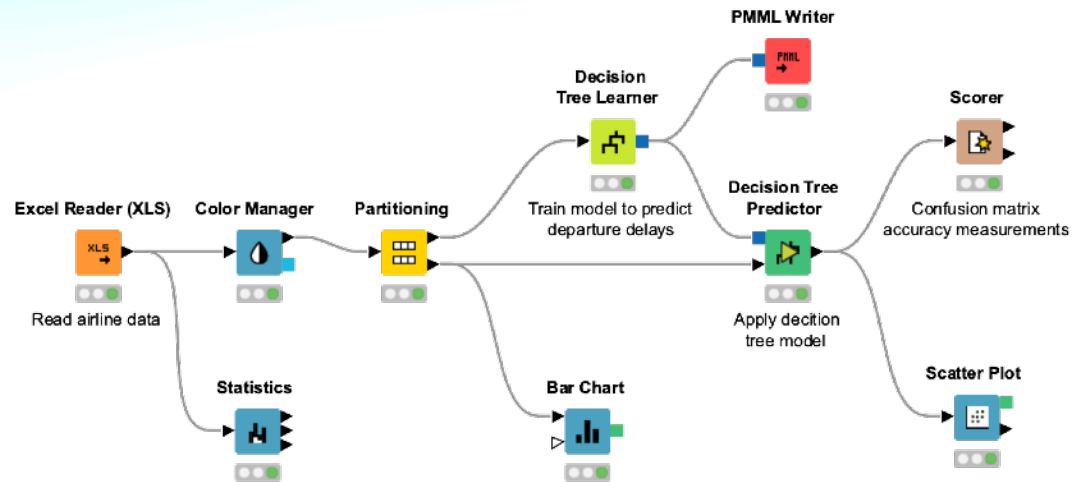
To show the node output, please select a configured or executed node.

Visual KNIME Workflows

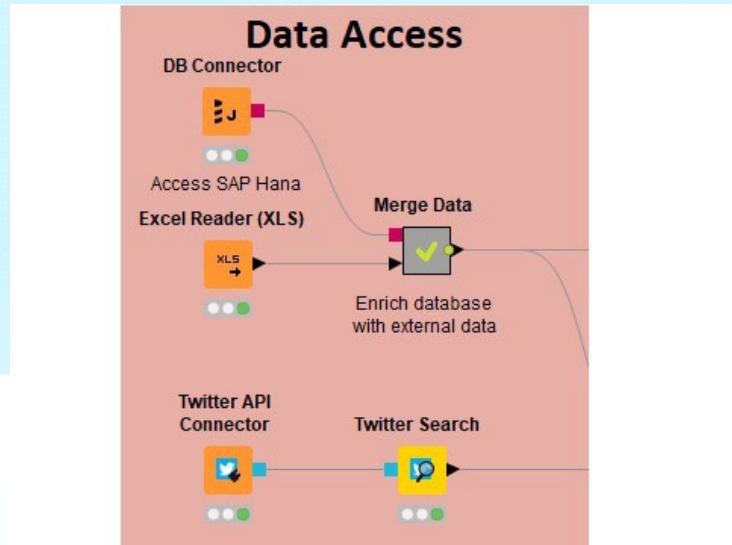
- **NODES** perform individual tasks on data
- Nodes has ports for data in and out



Nodes are combined to create **WORKFLOWS**

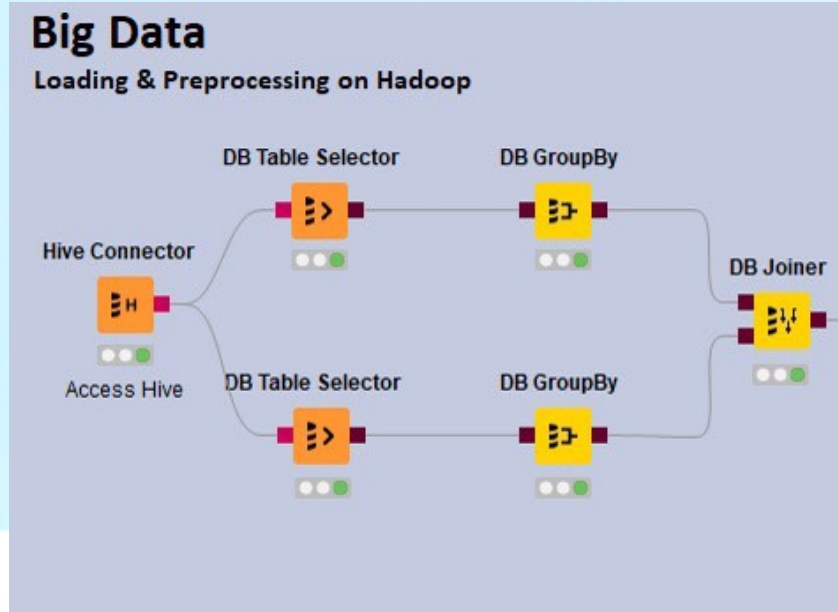


Data Access



- Databases
 - MySQL, PostgreSQL, Oracle
 - Theobald
 - any JDBC (DB2, MS SQL Server)
 - Amazon DynamoDB
- Files
 - CSV, txt, Excel, Word, PDF
 - SAS, SPSS
 - XML, JSON, PMML
 - Images, texts, networks
- Other
 - Twitter, Google
 - Amazon S3, Azure Blob Store
 - Sharepoint, Salesforce
 - Kafka
 - REST, Web services

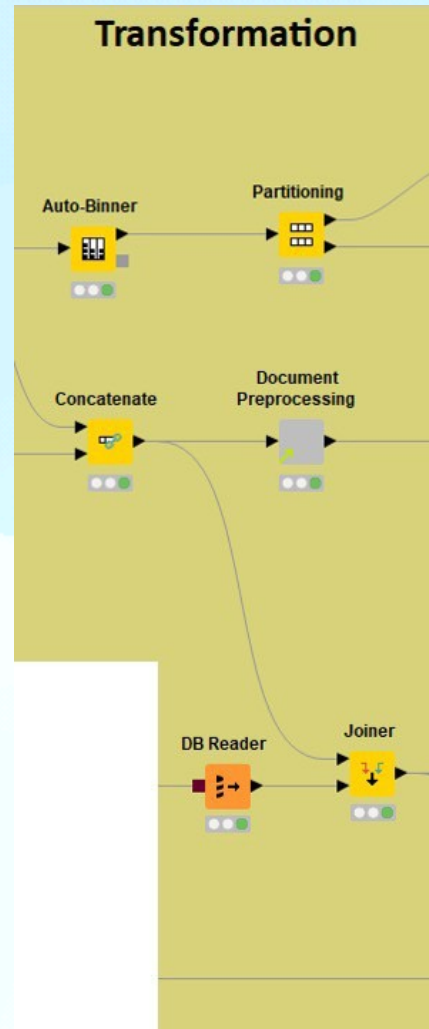
Big Data



- Spark & Databricks
- HDFS support
- Hive
- Impala
- In-database processing

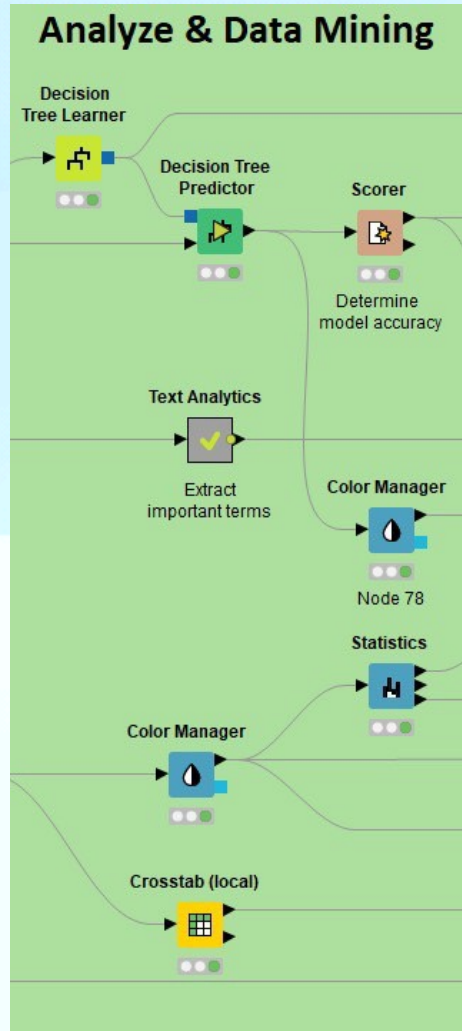


Data Transformation



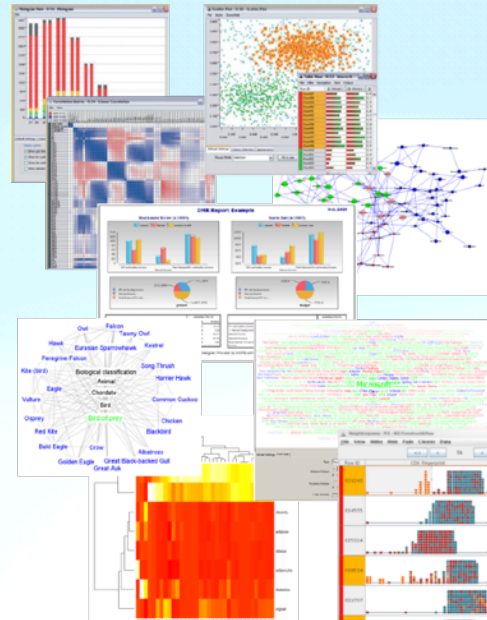
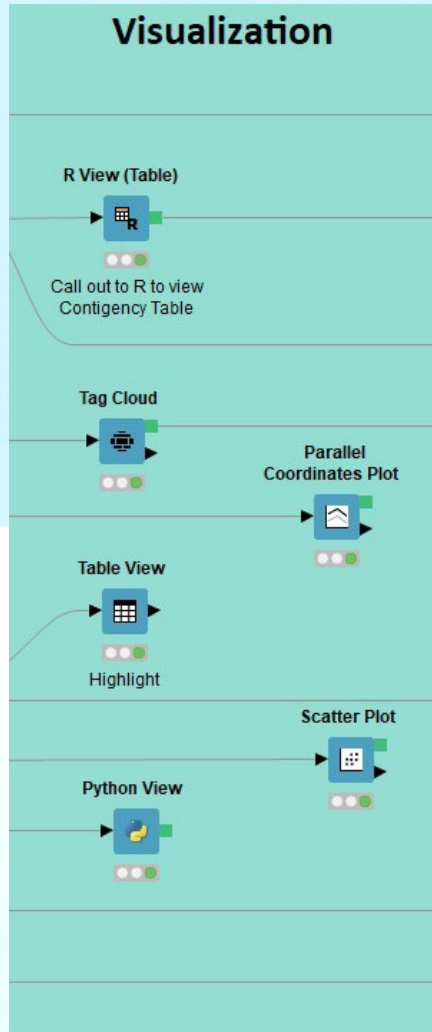
- Preprocessing
 - Row, column, matrix based
- Data blending
 - Join, concatenate, append
- Aggregation
 - Grouping, pivoting, binning
- Feature Creation and Selection

Analysis & Data Mining



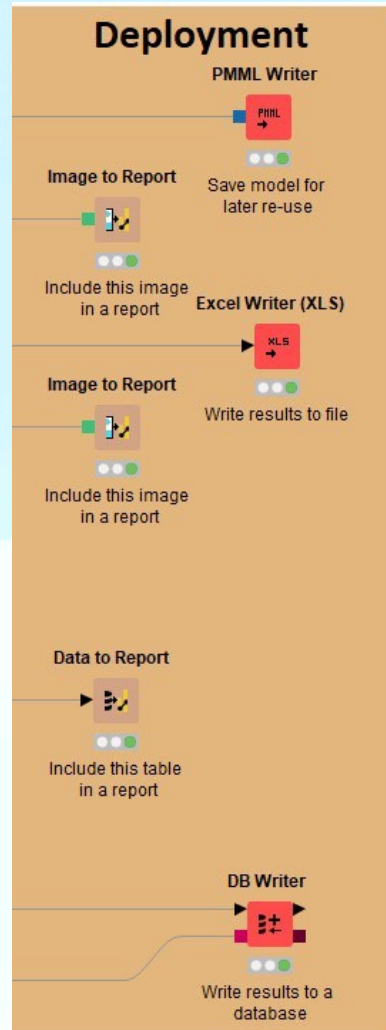
- Regression
 - Linear, logistic
- Classification
 - Decision tree, ensembles, SVM, MLP, Naïve Bayes
- Clustering
 - k-means, DBSCAN, hierarchical
- Validation
 - Cross-validation, scoring, ROC
- Deep Learning
 - Keras, DL4J
- External
 - R, Python, Weka, H2O, Keras

Visualization



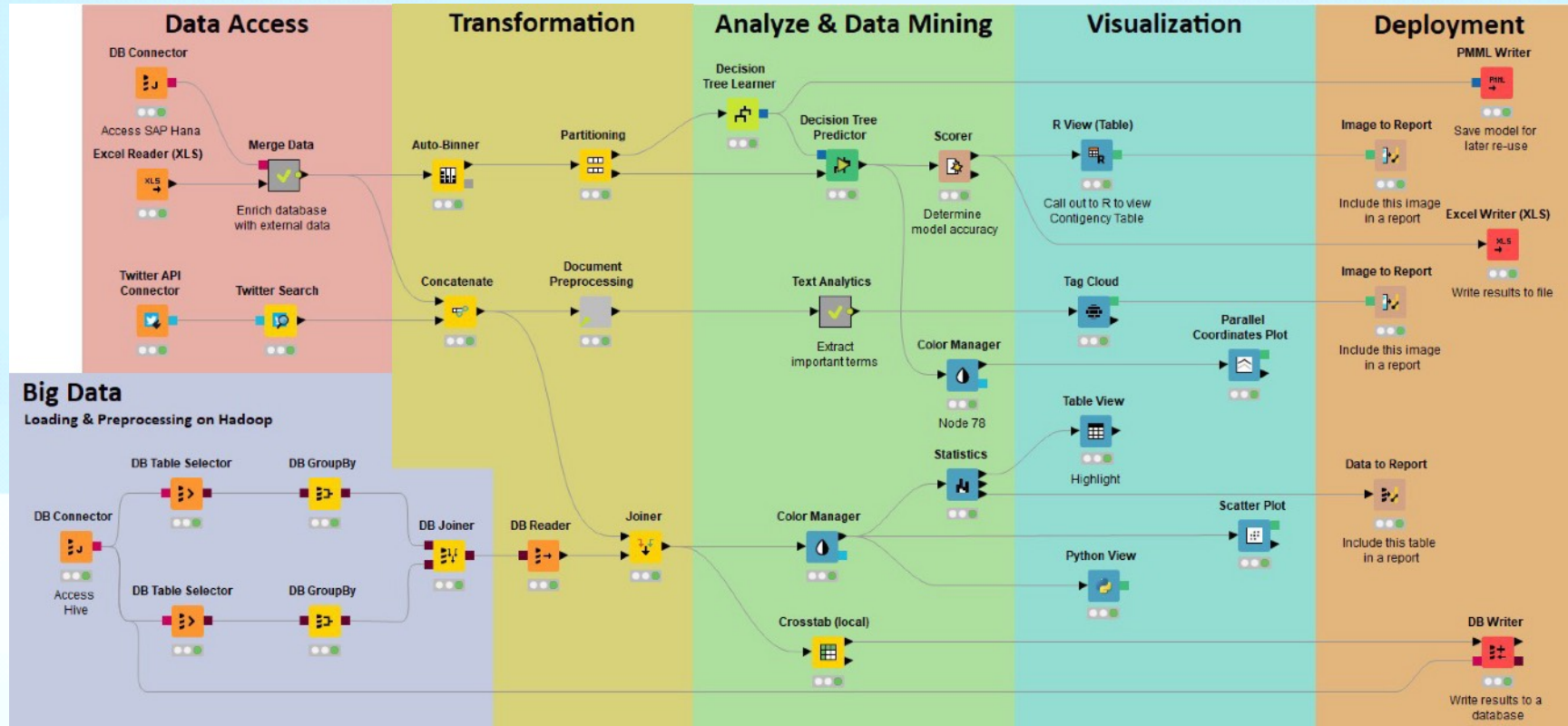
- Interactive Visualizations
- JavaScript-based nodes
 - Scatter Plot, Box Plot, Line Plot
 - Networks, ROC Curve, Decision Tree
 - Plotly Integration
 - Adding more with each release!
- Misc
 - Tag cloud, open street map, molecules
- Script-based visualizations
 - R, Python

Deployment



- Database
- Files
 - Excel, CSV, txt
 - XML
 - PMML
 - to: local, KNIME Server, Amazon S3, Azure Blob Store
- BIRT Reporting

Over 2000 Native and Embedded Nodes Included:



Data Access

MySQL, Oracle, ...
 SAS, SPSS, ...
 Excel, Flat, ...
 Hive, Impala, ...
 XML, JSON, PMML
 Text, Doc, Image, ...
 Web Crawlers
 Industry Specific
 Community / 3rd

Transformation

Row
 Column
 Matrix
 Text, Image
 Time Series
 Java Python
 Community / 3rd

Analysis & Mining

Statistics
 Data Mining
 Machine Learning
 Web Analytics Text
 Mining Network
 Analysis Social
 Media Analysis
 R, Weka, Python
 Community / 3rd

Visualization

R
 JFreeChart
 JavaScript Plotly
 Community / 3rd

Deployment

via BIRT
 PMML
 XML, JSON
 Databases Excel,
 Flat, etc. Text,
 Doc, Image
 Industry Specific
 Community / 3rd

Install KNIME Analytics Platform

- Select the KNIME version for your computer:
 - Mac
 - Windows – 32 or 64 bit
 - Linux
- Download the archive and extract the file, or download installer package and run it

<https://www.knime.com/downloads>

Windows

KNIME Analytics Platform for Windows (installer) <i>The installer adds an icon to the desktop and suggests suitable memory settings</i>	Download (459 MB)
KNIME Analytics Platform for Windows (self-extracting archive) <i>The self-extracting archive only creates a folder holding the KNIME installation</i>	Download (463 MB)
KNIME Analytics Platform for Windows (zip archive)	Download (547 MB)

Linux

KNIME Analytics Platform for Linux	Download (583 MB)
------------------------------------	-----------------------------------

Mac

KNIME Analytics Platform for macOS (10.13 and above)	Download (438 MB)
--	-----------------------------------

Find out what's new in the latest KNIME 4.4 release [here](#).

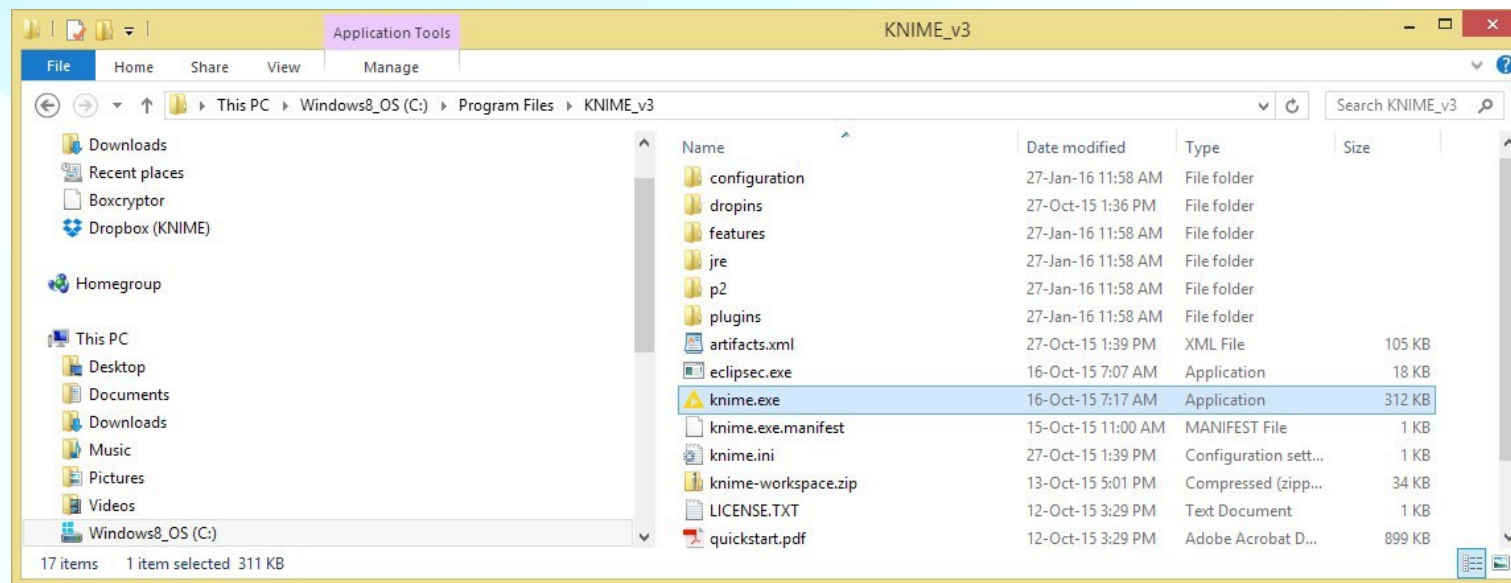
If you are interested in a previous version of KNIME Analytics Platform, please click [here](#).

Start KNIME Analytics Platform

- Use the shortcut created by the installer

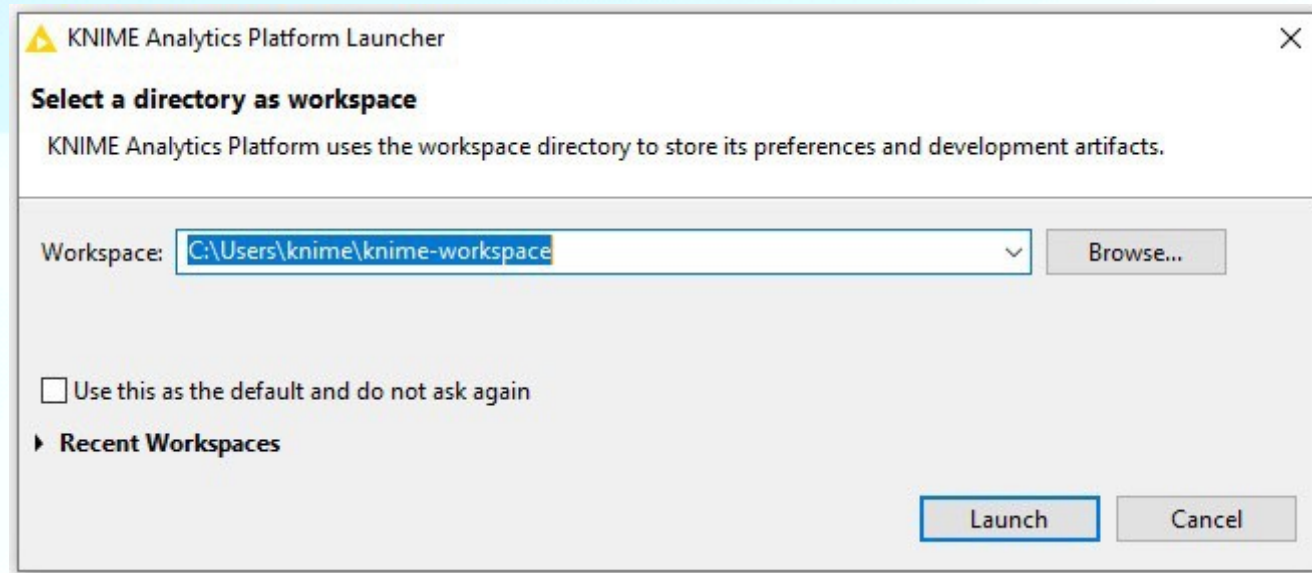


- Or go to the installation directory and launch KNIME via the knime.exe



The KNIME Workspace

- The workspace is the **folder/directory** in which workflows (and potentially data files) are stored for the current KNIME session.
- Workspaces are portable (just like KNIME)



The KNIME Analytics Platform Workbench (classic UI)

The screenshot displays the KNIME Analytics Platform Workbench interface. The central area is the **Workflow Editor**, showing a workflow titled "My first Workflow" with four nodes: **File Reader** (read adult.csv), **Row Filter** (keep only records born in the US), **Column Filter** (remove gender), and **Table Writer** (write table). The **Row Filter** node is highlighted with a yellow box.

Surrounding the workflow editor are several panels:

- KNIME Explorer**: Located at the top left, showing a tree view of the workspace structure.
- Workflow Coach**: Located on the left side, displaying recommended nodes and their usage percentages.
- Node Repository**: Located at the bottom left, providing a categorized list of available nodes.
- Outline**: Located at the bottom left, showing a small overview of the workflow.
- Console & Node Monitor**: Located at the bottom right, displaying the execution status of the selected node and its output data.
- Node Description**: Located on the right side, providing detailed information about the selected node.
- KNIME Hub**: Located on the right side, showing a search interface and a welcome message.

The **Console & Node Monitor** panel shows the following output data:

ID	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours
Row0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40
Row1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13
Row2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40

The KNIME Analytics Platform Workbench (Modern UI)

Predict the income group from demographic attributes of the adult data set (census data)

Data Reading
Read the adult data set file. There is one row for each person, plus demographic info and the income group. The file is located in TheData/Basics/.

Data Partitioning
Create two separate partitions from original data set: training set (80%) and test set (20%). The test set deliberately consists of unseen data that will not be used for training.

Train a Model
This node builds a decision tree. Other Learner nodes train other models. Most Learner nodes output a PMML model (blue square output port).

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 and calculate the related accuracy measures.

Descriptive Statistics
Calculate the statistical properties of the data set attributes.

Visualize
Create interactive scatter plot.

► 1: File Table | Flow Variables

Rows: 32561 | Columns: 15

#	Row...	age Number (inte...)	workclass String	fnlwgt Number (inte...)	education String	educatio... Number (inte...)	marital-s... String	occupati... String	relations... String	race String
1	Row0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White
2	Row1	50	Self-emp-not-i...	83311	Bachelors	13	Married-civ-s...	Exec-manage...	Husband	White
3	Row2	38	Private	215646	HS-grad	9	Divorced	Handlers-clea...	Not-in-family	White
4	Row3	53	Private	234721	11th	7	Married-civ-s...	Handlers-clea...	Husband	Black
5	Row4	28	Private	338409	Bachelors	13	Married-civ-s...	Prof-specialty	Wife	Black
6	Row5	37	Private	284582	Masters	14	Married-civ-s...	Exec-manage...	Wife	White
7	Row6	49	Private	160187	9th	5	Married-spou...	Other-service	Not-in-family	Black

In KNIME 5.0 or later versions

Switching from Classical UI to Modern UI (in KNIME >5.0)

The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a workflow titled "3: Building a Simple Classifier" with the goal: "Predict the income group from demographic attributes of the adult data set (census data)". The workflow consists of five nodes:

- CSV Reader**: Read the adult data set file. There is one row for each person, plus demographic info and the income group. The file is located in TheData/Basics/.
- Partitioning**: Create two separate partitions from original data set: training set (80%) and test set (20%). The test set deliberately consists of unseen data that will not be used for training.
- Decision Tree Learner**: This node builds a decision tree. Other Learner nodes train other models. Most Learner nodes output a PMML model (blue square output port).
- Decision Tree Predictor**: Apply the Model Predictor nodes apply a specific model to a data set and append the model predictions.
- Scorer**: Compute a confusion matrix between real and predicted class values and calculate the related accuracy measures.

The interface includes a sidebar with a Node Repository, a Workflow Coach, and a Console window. The Console window shows the following output:

```
KNIME Console
*****
*** Welcome to KNIME Analytics Platform v5.1.2.v202310091347 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: /Users/raghava/knime-workspace-50/.metadata/knime/knime.log
WARN Excel Reader 5:266 The node configuration changed and the table spec will be re
WARN Excel Reader 5:266 The node configuration changed and the table spec will be re
```

Switching from Modern UI to Classic UI (in KNIME >5.0)

Click on icon i

KNIME Analytics Platform - /Users/raghava/knime-workspace-50

Home Building a Simple Classifier

Execute Cancel Reset Create metanode Create component

Repository

Search Nodes

Excel Reader Excel Writer Microsoft Authenticator

Google Authenticator Google Sheets Reader Google Sheets Writer

CSV Reader CSV Writer

Manipulation

Row Filter Column Filter Concatenate

Value Lookup Row Aggregator Table Splitter

Table Cropper Cell Extractor

Single Model Training for Classification

This workflow demonstrates how a simple classifier is built and applied to new data. Find more information on KNIME's Learning page at <http://www.knime.com/learning> (courses, tutorials, cheatsheets, books, and more)

Task

Predict the income group from demographic attributes of the adult data set (census data)

Data Reading

Read the adult data set file. There is one row for each person, plus demographic info, and the income group. The file is located in TheData/Basics/.

Data Partitioning

Create two separate partitions from original data set: training set (80%) and test set (20%). The test set deliberately consists of unseen data that will not be used for training.

Train a Model

This node builds a decision tree. Other Learner nodes train other models. Most Learner nodes output a PMML model (blue square output port).

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 and calculate the related accuracy measures.

1: File Table Flow Variables

Rows: 32561 | Columns: 15

#	Row...	age	workclass	fnlwgt	education	educatio...	marital-s...
		Number (inte...	String	Number (inte...	String	Number (inte...	String
1	Row0	39	State-gov	77516	Bachelors	13	Never-married
2	Row1	50	Self-emp-not-i...	83311	Bachelors	13	Married-civ-s...
3	Row2	38	Private	215646	HS-grad	9	Divorced
4	Row3	53	Private	234721	11th	7	Married-civ-s...
5	Row4	28	Private	338409	Bachelors	13	Married-civ-s...
6	Row5	37	Private	284582	Masters	14	Married-civ-s...
7	Row6	49	Private	160187	9th	5	Married-snou...

KNIME Analytics Platform - /Users/raghava/knime-workspace-50

Home Building a Simple Classifier

Install Extensions

Install Extensions to access additional functionality such as the ability to process complex data types, as well as to use advanced algorithms.

Install Extensions Check for updates

Switch to classic user interface

Switch to the classic KNIME Analytics Platform user interface. To switch back again, click the button "Open KNIME Modern UI" in the top right corner of the classic user interface.

Switch to KNIME classic user interface

Open for Innovation

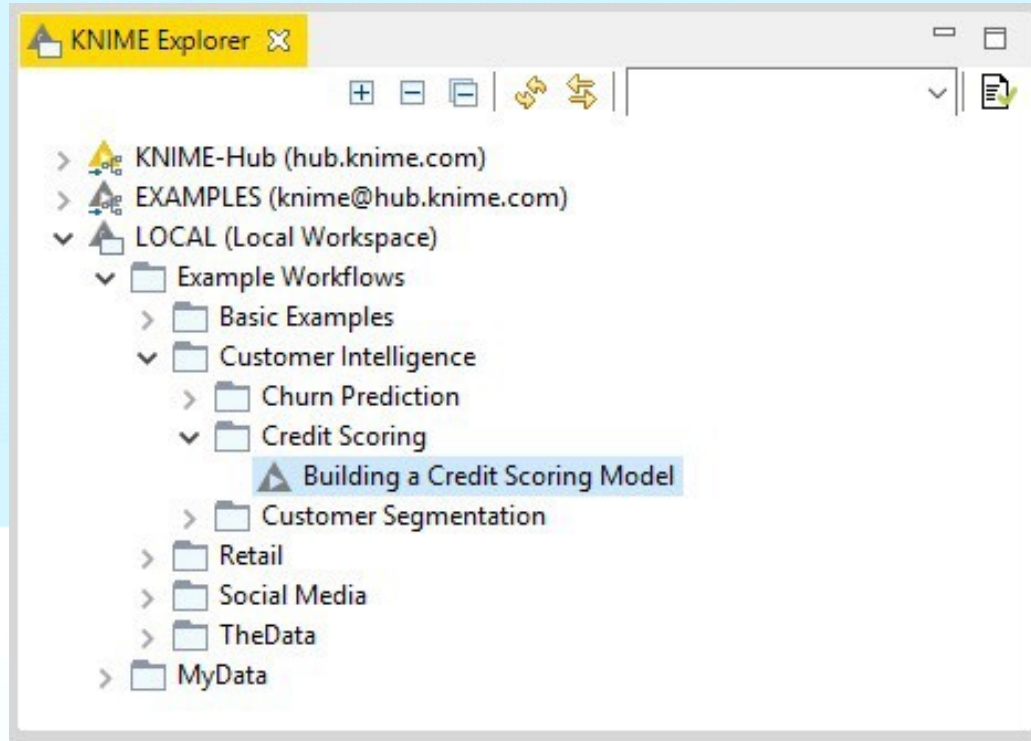
KNIME

Copyright by KNIME AG, Zurich, Switzerland
contact@knime.com
<https://www.knime.com>

This software is a bundle of multiple modules, each released under its own license. Please check the individual licenses by clicking "About KNIME" and the "Credits" button.

About KNIME Credits

KNIME Explorer





- In LOCAL you can access your own workflow projects.

- Other mountpoints allow you to connect to

- EXAMPLE Server
- KNIME Hub
- KNIME Server

- The Explorer toolbar on the top has a search box and buttons to

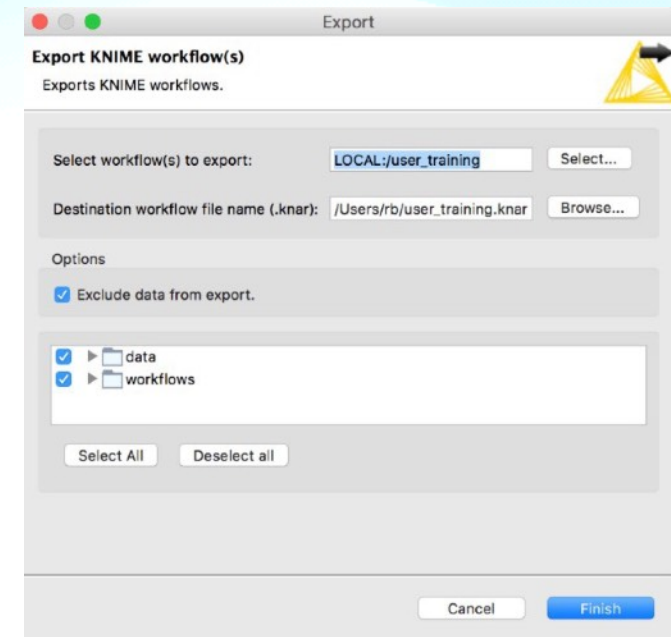
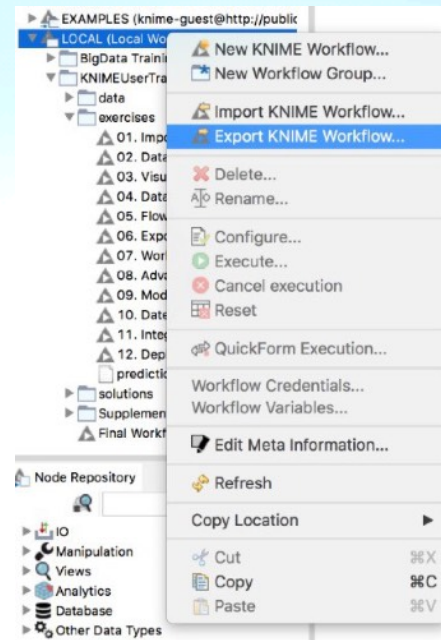
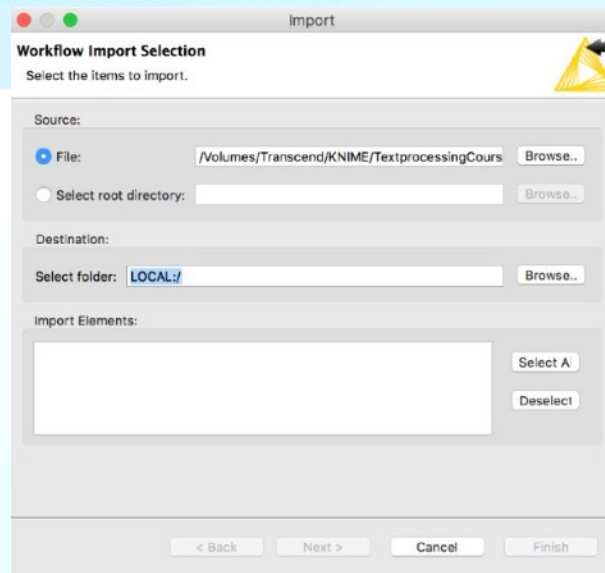
-  select the workflow displayed in the active editor
-  refresh the view

- The KNIME Explorer can contain 4 types of content:

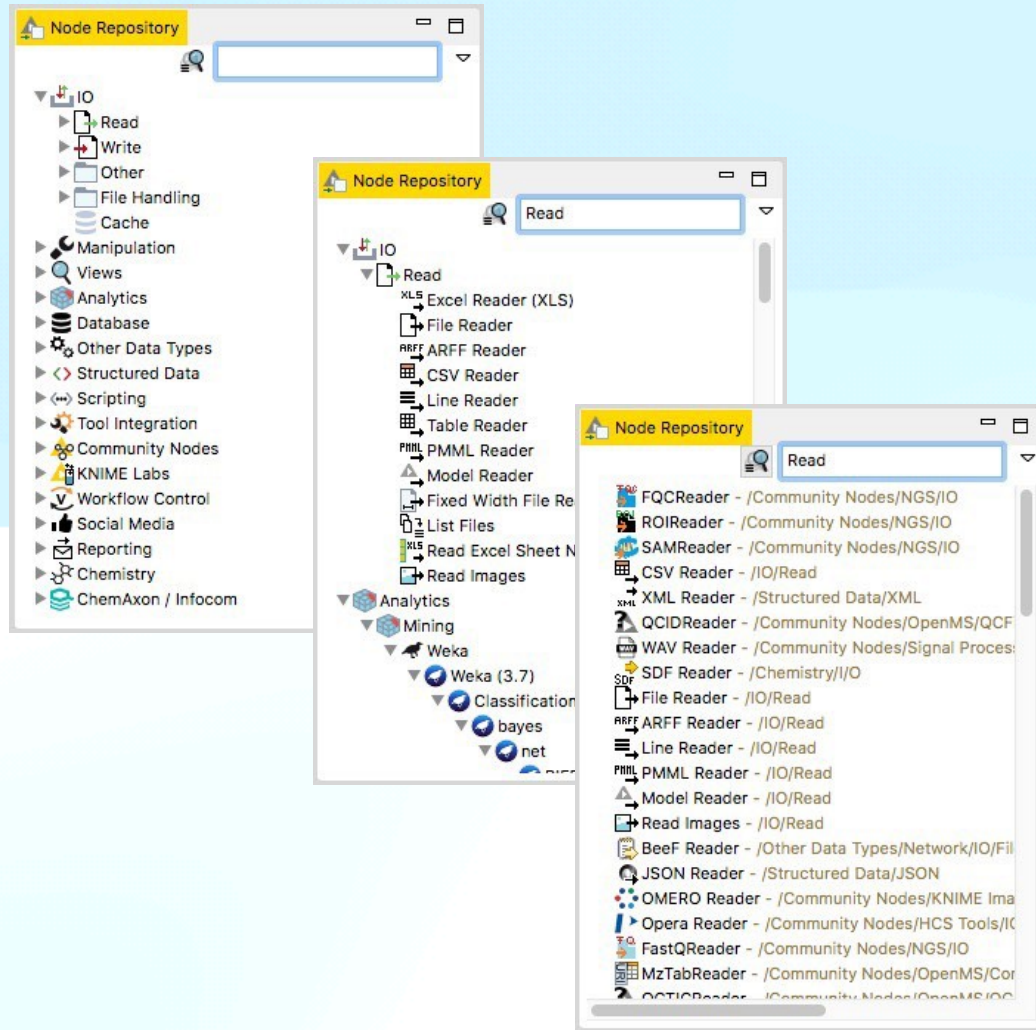
- Workflows
- Workflow groups
- Data files
- Shared Components



Creating New Workflows, Importing and Exporting

- Right-click inside the KNIME Explorer to create a new workflow or a workflow group, or to import a workflow
- Right-click the workflow or workflow group to export

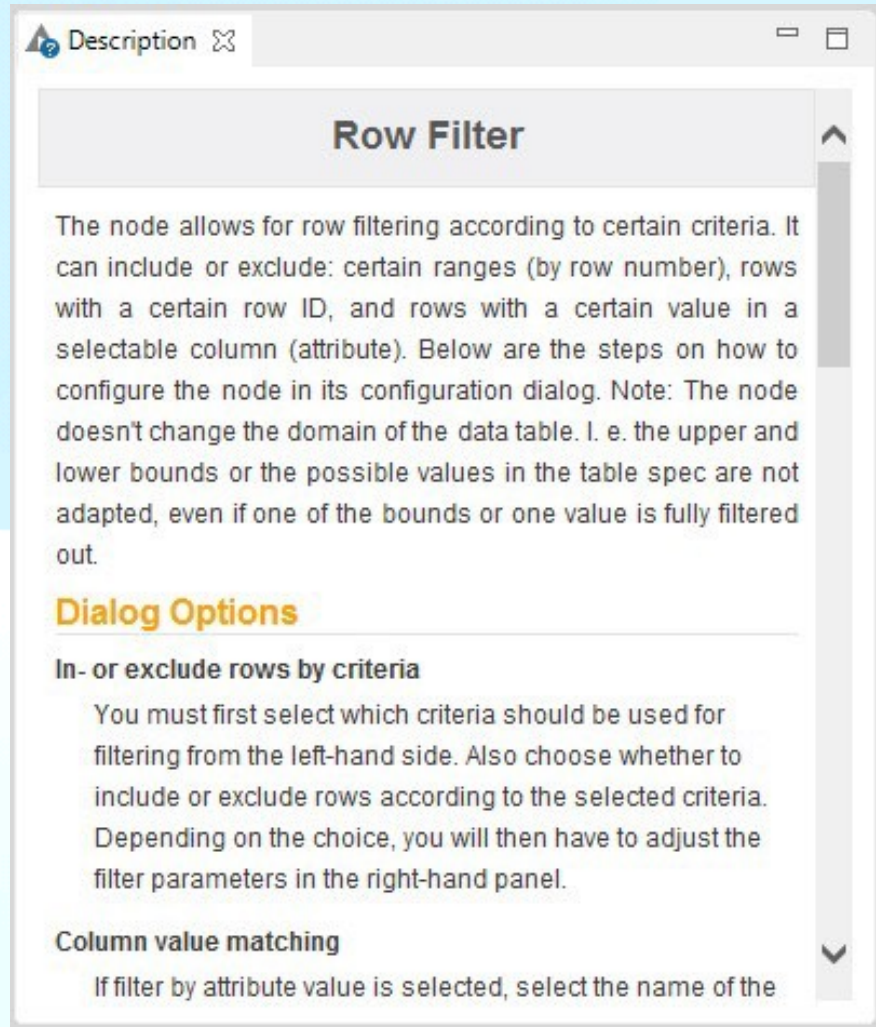


Node Repository



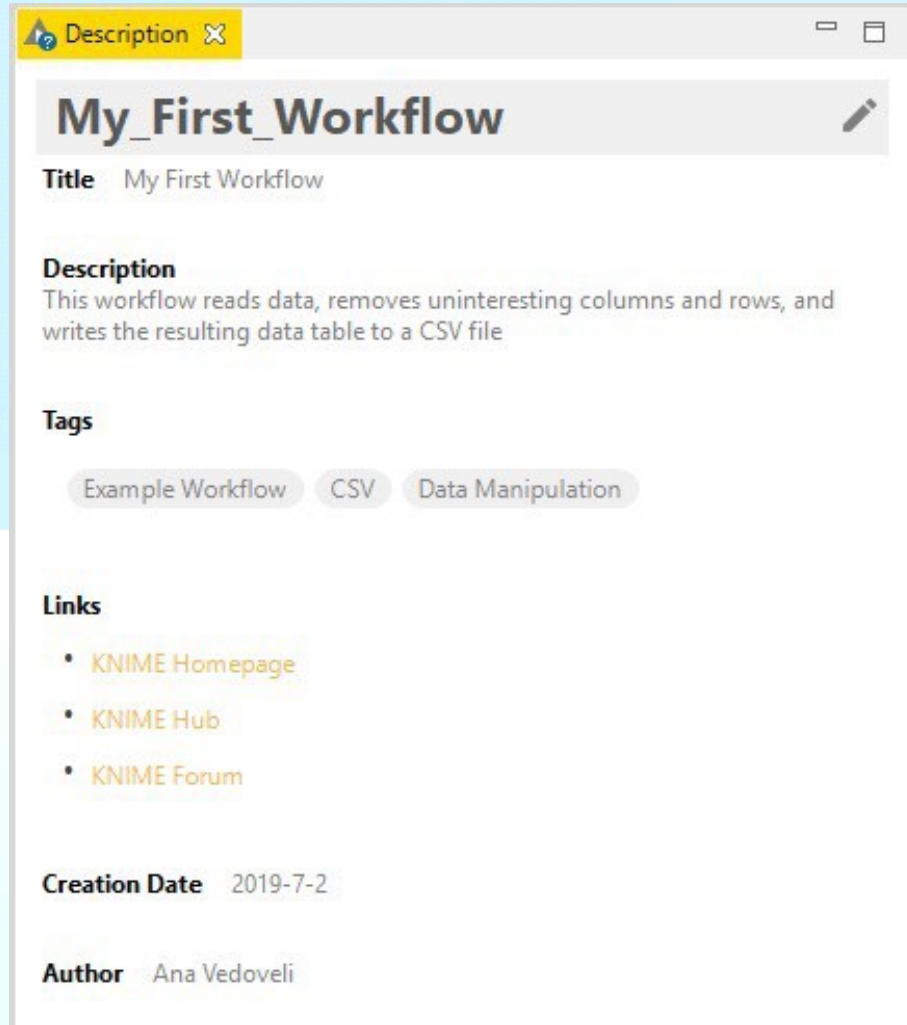
- The Node Repository lists all KNIME nodes
- The search box has 2 modes
 -  Standard Search – exact match of node name
 -  Fuzzy Search – finds the most similar node name

Description



- The Description window gives information about:
 - Node Functionality
 - Input & Output
 - Node Settings
 - Ports
 - References to literature

Workflow Description



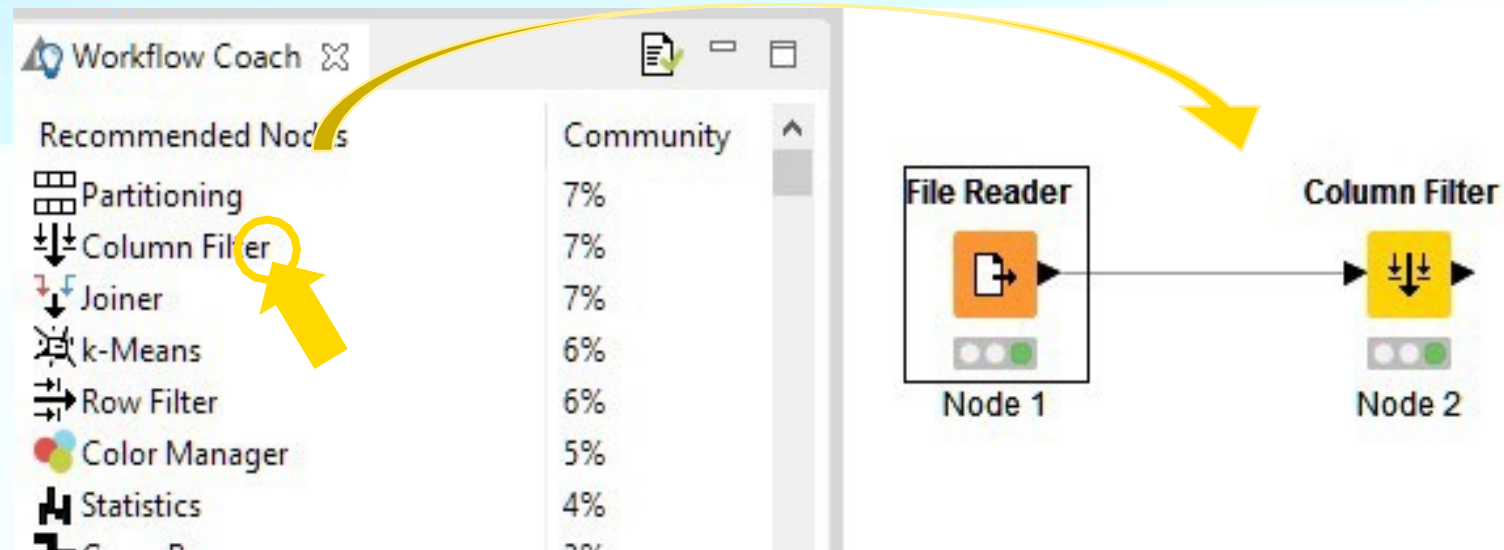
The screenshot shows a window titled "Description" with a close button. The main content area displays the following information:

- Title:** My First Workflow
- Description:** This workflow reads data, removes uninteresting columns and rows, and writes the resulting data table to a CSV file
- Tags:** Example Workflow, CSV, Data Manipulation
- Links:**
 - [KNIME Homepage](#)
 - [KNIME Hub](#)
 - [KNIME Forum](#)
- Creation Date:** 2019-7-2
- Author:** Ana Vedoveli

- When selecting the workflow, the Description window gives information about the workflow's:
 - Title
 - Description
 - Associated Tags and Links
 - Creation Date
 - Author

Workflow Coach

- Node recommendation engine
 - Gives hints about which node use next in the workflow
 - Based on KNIME communities' usage statistics
 - Based on own KNIME workflows



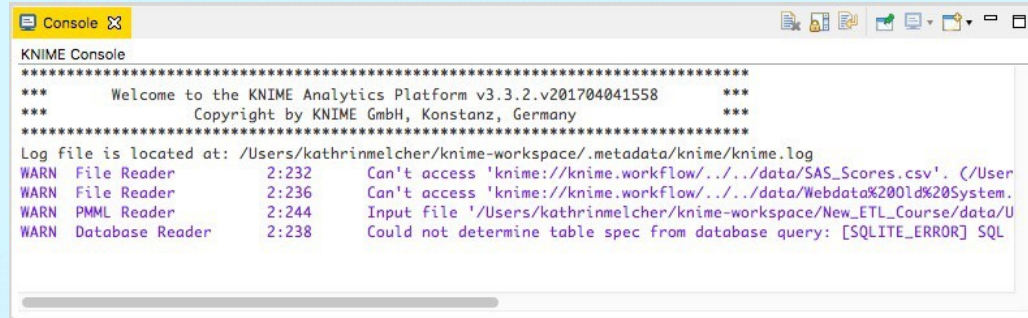
Node Monitor

- By default the Node Monitor shows you the output table of the node selected in the workflow editor
- Click on the three dots on the upper right to show the flow variables, configuration, etc.

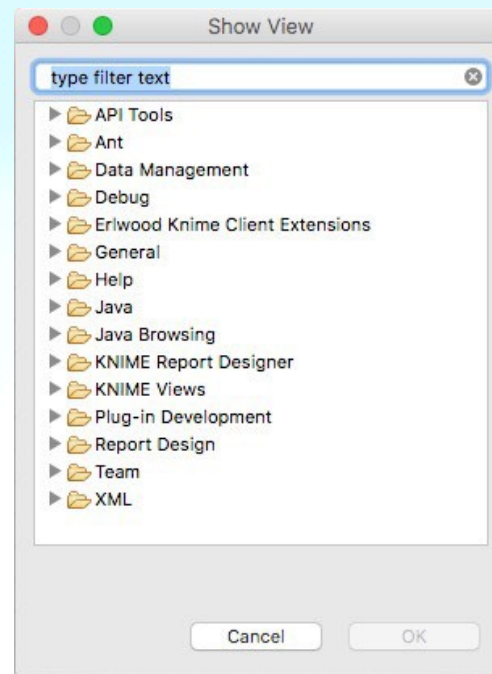
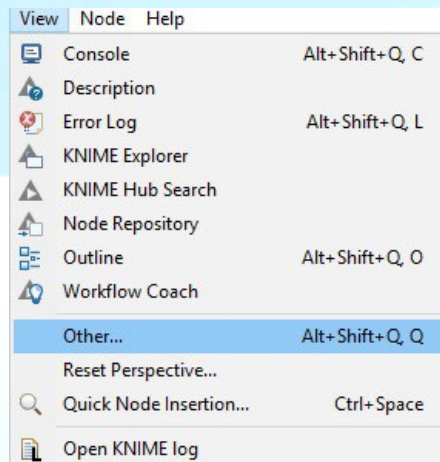
The screenshot shows the KNIME Node Monitor window for the node "Get Customers from Database (0:1207)". The state is "EXECUTED". The "Port Output" is set to "Port 0". A context menu is open, listing options: "Show Output Table" (checked), "Show Variables", "Show Configuration", "Show Entire Configuration", "Show Node Timing Information", and "Show Graph Annotations". Below the menu is a table of customer data.

ID	MaritalStatus	Gender	EstimatedYearlyIncome	NumberOfContracts	Age	Available401K	CustomerV				Products
CustomerID: 722204	S	F	80000	4	42	1	1		4	5	Private Investr
CustomerID: 489847	M	M	60000	2	46	1	1		4	3	Private Investr
CustomerID: 8444723	M	M	40000	1	32	1	2		3	0	P+B Investmer
CustomerID: 1487427	M	M	30000	2	63	1	1		2	2	P+B Investmer
CustomerID: 4693433	M	M	20000	2	63	1	1		3	4	Gold Investme
CustomerID: 7724940	M	M	30000	2	33	1	2		3	0	P+B Investmer
CustomerID: 9784443	M	M	60000	2	34	1	2		3	0	P+B Investmer
CustomerID: 3177757	M	M	70000	2	57	1	1		5	2	Fund Manager

Console and Other Views



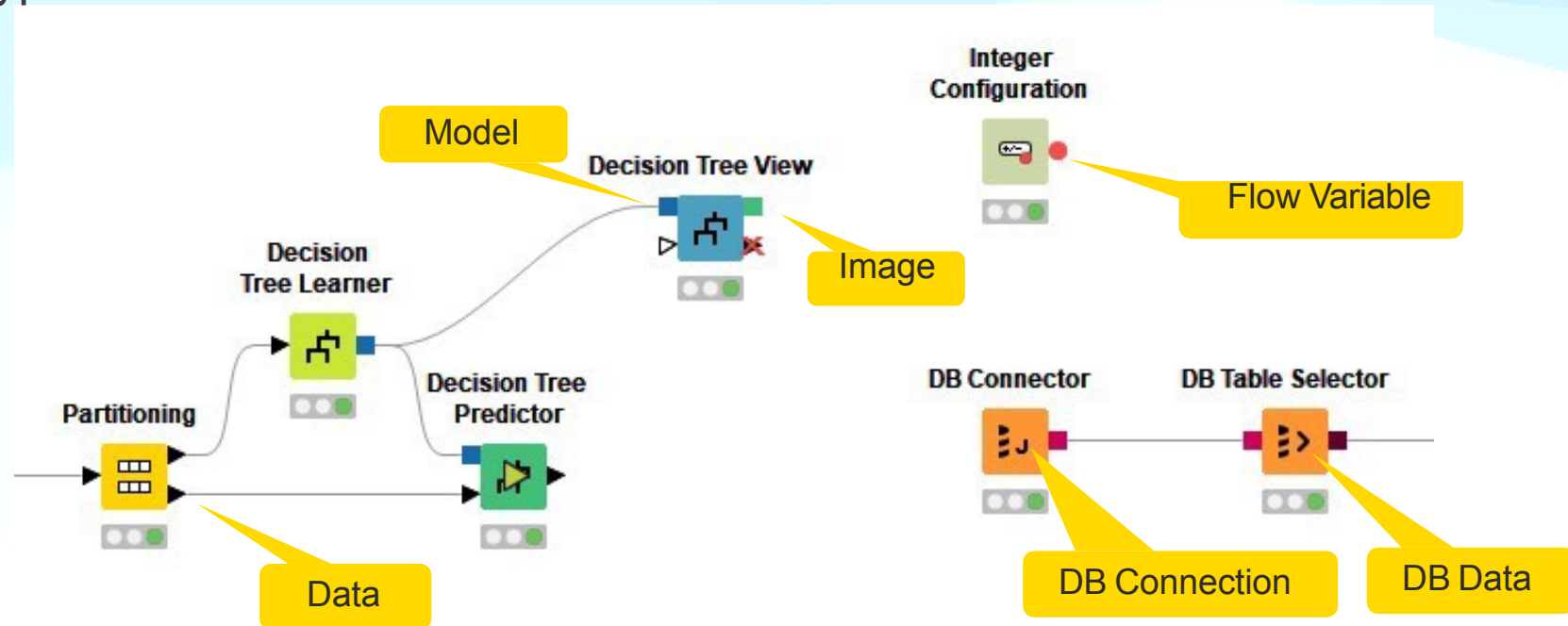
```
KNIME Console
*****
*** Welcome to the KNIME Analytics Platform v3.3.2.v201704041558 ***
*** Copyright by KNIME GmbH, Konstanz, Germany ***
*****
Log file is located at: /Users/kathrinmelcher/knime-workspace/.metadata/knime/knime.log
WARN File Reader 2:232 Can't access 'knime://knime.workflow/../../data/SAS_Scores.csv'. /User
WARN File Reader 2:236 Can't access 'knime://knime.workflow/../../data/Webdata%201d%20System.
WARN PMML Reader 2:244 Input file '/Users/kathrinmelcher/knime-workspace/New_ETL_Course/data/U
WARN Database Reader 2:238 Could not determine table spec from database query: [SQLITE_ERROR] SQL
```



- Console view prints out error and warning messages about what is going on under the hood
- Click on View and select Other... to add different views
 - Node Monitor, Licenses, etc.

Inserting and Connecting Nodes

- Insert nodes into workspace by dragging them from Node Repository or by double-clicking in Node Repository
- Connect nodes by left-clicking output port of Node A and dragging the cursor to (matching) input port of Node B
- Common port types:



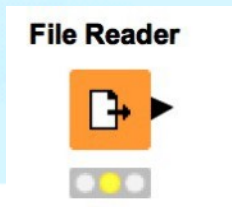
More on Nodes...

- A node can have 4 states:



Not Configured:

The node is waiting for configuration or incoming data.



Configured:

The node has been configured correctly and can be executed.



Executed:

The node has been successfully executed. Results may be viewed and used in downstream nodes.

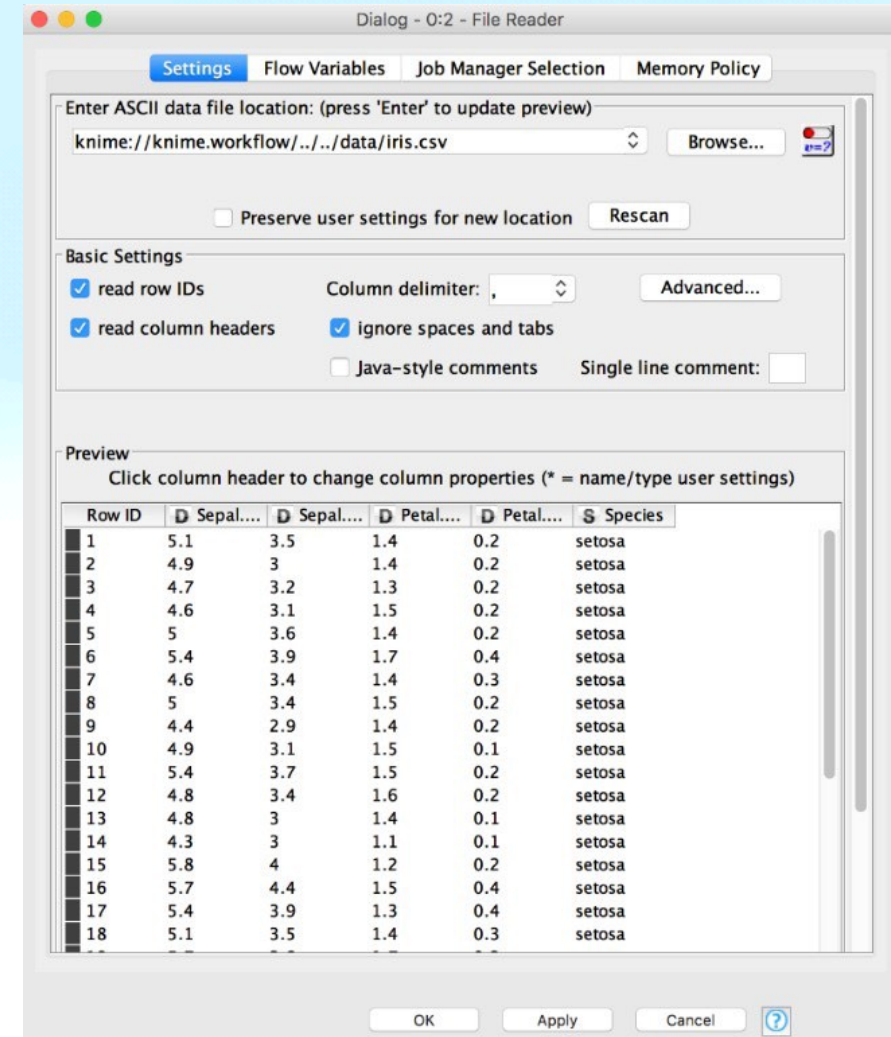


Error:

The node has encountered an error during execution.

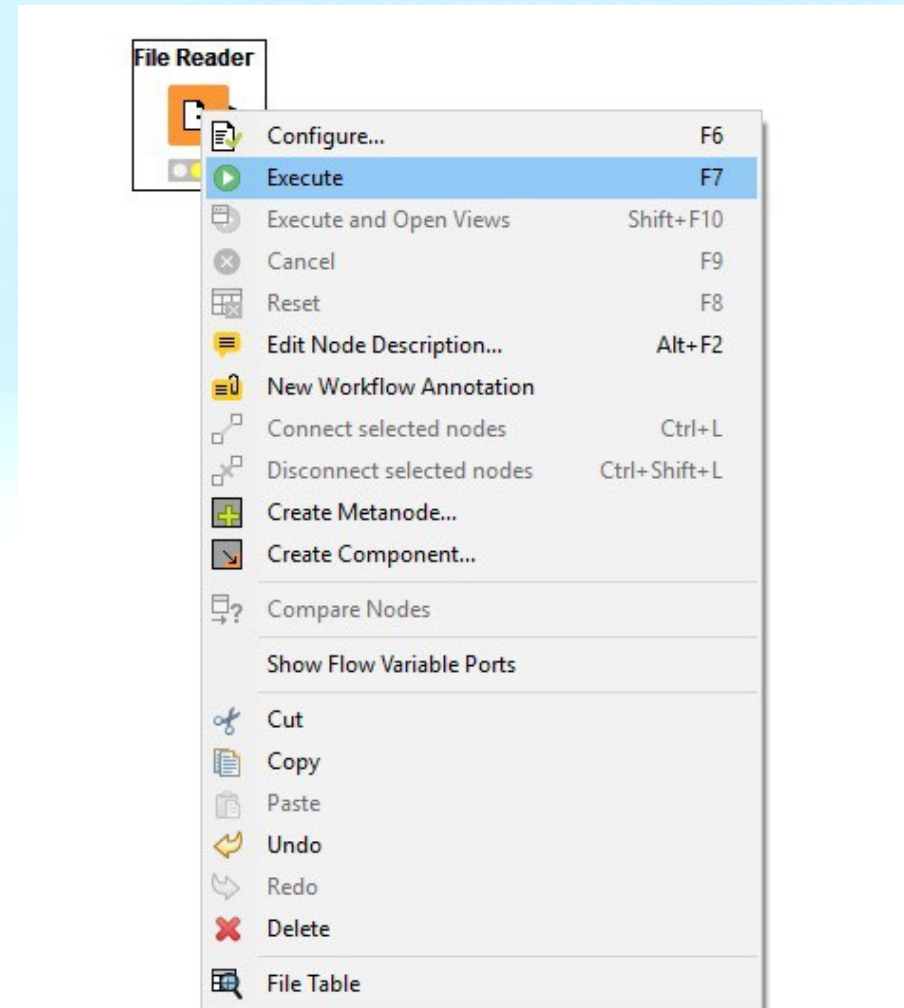
Node Configuration

- Most nodes require configuration
- To access a node configuration window:
 - Double-click the node
 - Right-click -> Configure

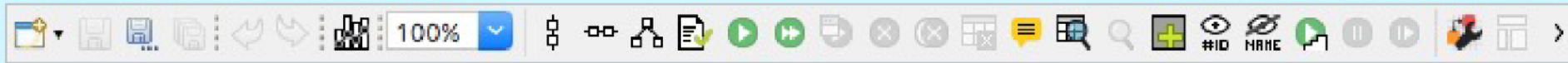


Node Execution







- Right-click node
- Select Execute in the context menu
- If execution is successful, status shows green light
- If execution encounters errors, status shows red light



Tool Bar



The buttons in the toolbar can be used for the active workflow. The most important buttons are:

-  Execute selected and executable nodes (F7)
-  Execute all executable nodes
-  Execute selected nodes and open first view
-  Cancel all selected, running nodes (F9)
-  Cancel all running nodes
- 

Node Views

- Right-click node to inspect the execution results by
 - selecting output ports (last option in the context menu) to inspect tables, images, etc.
 - selecting Interactive View to open visualization results in a browser

The diagram illustrates the process of inspecting a node's execution results. On the left, a 'Scatter Plot' node is shown with an input arrow on the left and two output arrows on the right. A context menu is open over the node, listing various actions. A yellow callout box labeled 'Plot View' points to the 'Interactive View: Scatter Plot' option in the context menu. On the right, a larger context menu is shown, with a yellow callout box labeled 'Data View' pointing to the 'Input data and view selection' option at the bottom.

Action	Shortcut
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 Metanode...	
Create Component...	
Interactive View: Scatter Plot	
Compare Nodes	
Show Flow Variable Ports	
Cut	
Copy	
Paste	
Undo	
Redo	
Delete	
Image	
Input data and view selection	

KNIME File Extensions

Dedicated file extensions for workflows and workflow groups associated with KNIME Analytics Platform

- ***.knwf** for KNIME Workflow Files

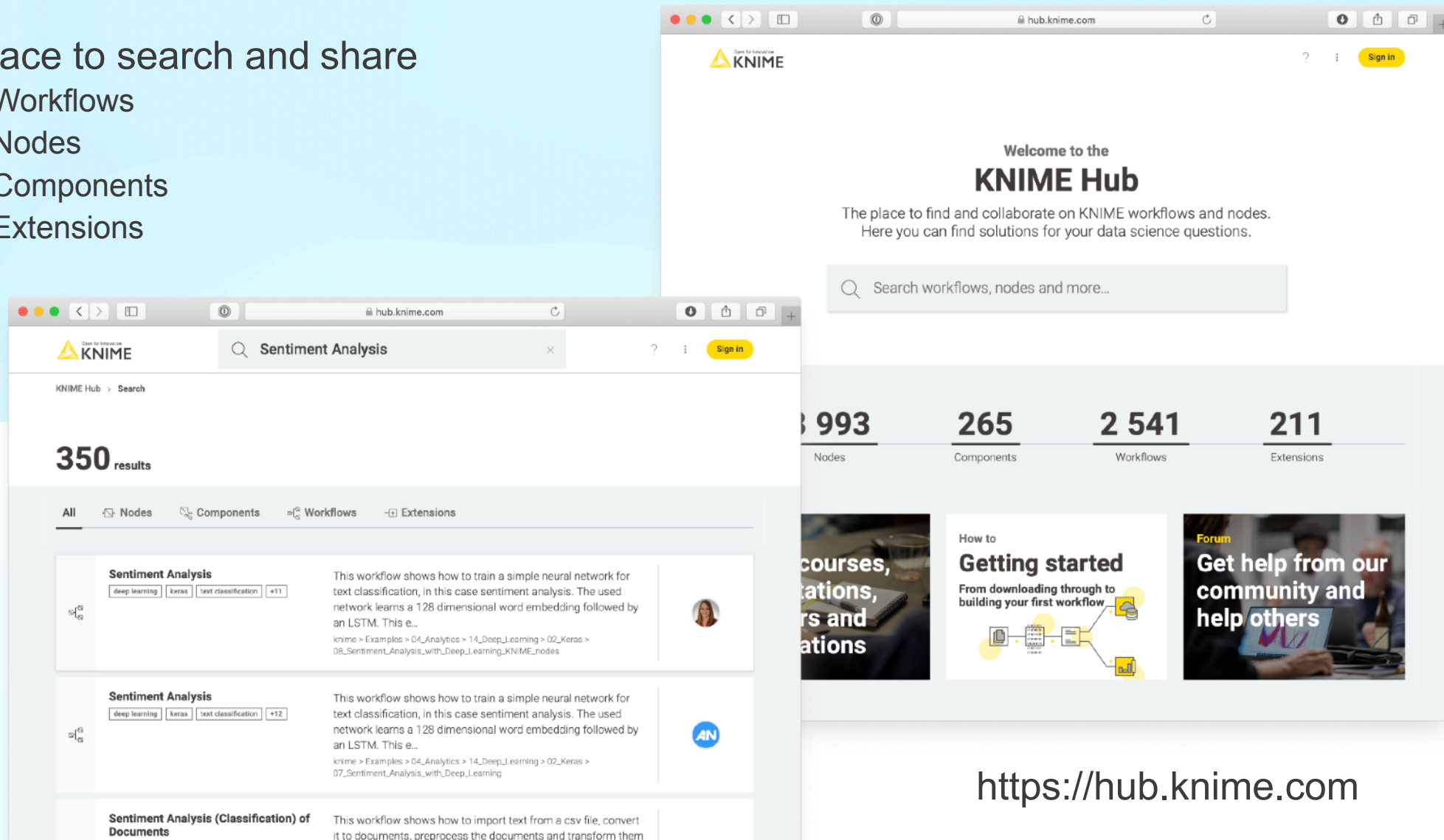


- ***.knar** for KNIME Archive Files



Getting Started: KNIME Hub

- Place to search and share
 - Workflows
 - Nodes
 - Components
 - Extensions



<https://hub.knime.com>

Getting Started: KNIME Example Server

- Connect via KNIME Explorer to a public repository with large selection of example workflows for many, many applications

The image displays two windows from the KNIME software. The left window is the KNIME Explorer, showing a tree view of a public repository. The right window is the KNIME Analytics Platform, showing a workflow titled "Simple Preprocessing Example".

KNIME Explorer (Left Window):

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
 - 00_Components
 - 01_Data_Access
 - 02_ETL_Data_Manipulation
 - 00_Basic_Examples
 - 00_Visual_Analysis_of_Sales_Data
 - 01_Example_for_Standard_Preprocessing
 - 02_ETL_Basics
 - 01_Filtering
 - 02_Aggregations
 - 03_Joining_and_Concatenating
 - 04_Transformation
 - 05_Indexing_Searching
 - 06_Date_and_Time_Manipulation
 - Random Forest Model Workflow 15 Plates
 - 03_Visualization
 - 04_Analytics
 - 05_Reporting
 - 06_Control_Structures
 - 07_Scripting
 - 08_Other_Analytics_Types
 - 09_Enterprise
 - 10_Big_Data
 - 20_Strange_but_Educational
 - 40_Partners
 - 50_Applications
 - 99_Community
 - LOCAL (Local Workspace)

KNIME Analytics Platform (Right Window):

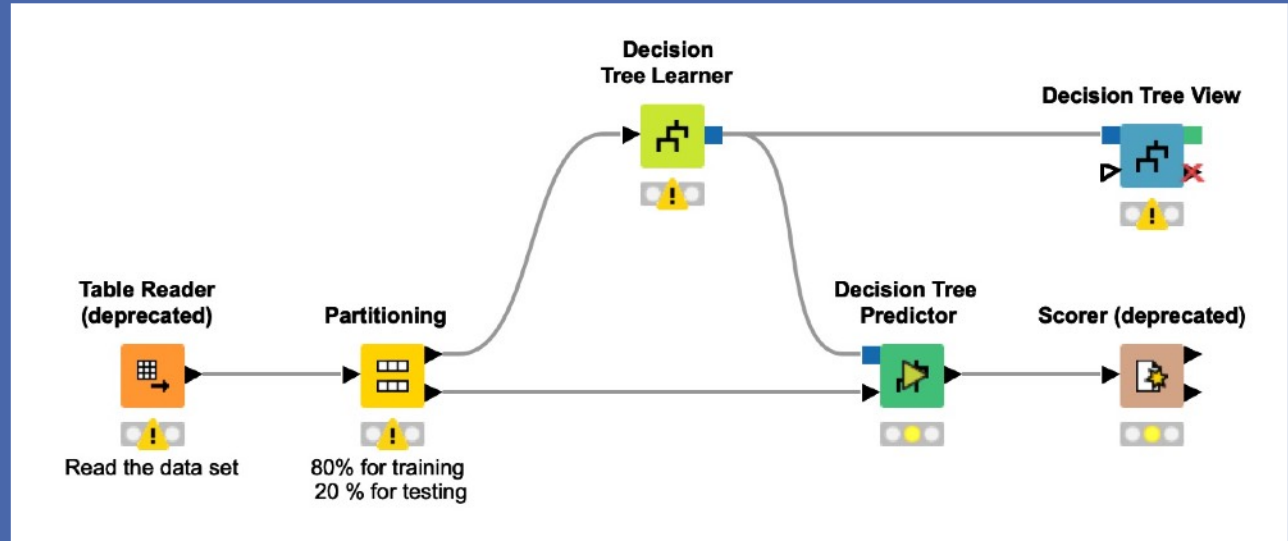
Simple Preprocessing Example
Shows the use of different filter nodes.

This is a temporary copy of "knime://EXAMPLES/Users/knime/Examples/02_ETL_Data_Manipulation/00_Basic_Examples/01_Example_for_Standard_Preprocessing.knime". Use "Save As..." to save a permanent copy of the workflow to your local workspace, or a mounted KNIME Server.

The workflow consists of the following nodes:

- File Reader
- Row Filter (keep rows where "native-country" is missing)
- Column Filter
- Reference Column Filter
- Numeric Binner (Replace numeric column age with an attribute column)
- Nominal Value Row Filter (Filter by this new attribute)
- Reference Row Filter (Exclude those rows from the original table)
- Row Filter (filter example with a regular expression)
- Concatenate

HANDS-ON I: SENTIMENT ANALYSIS



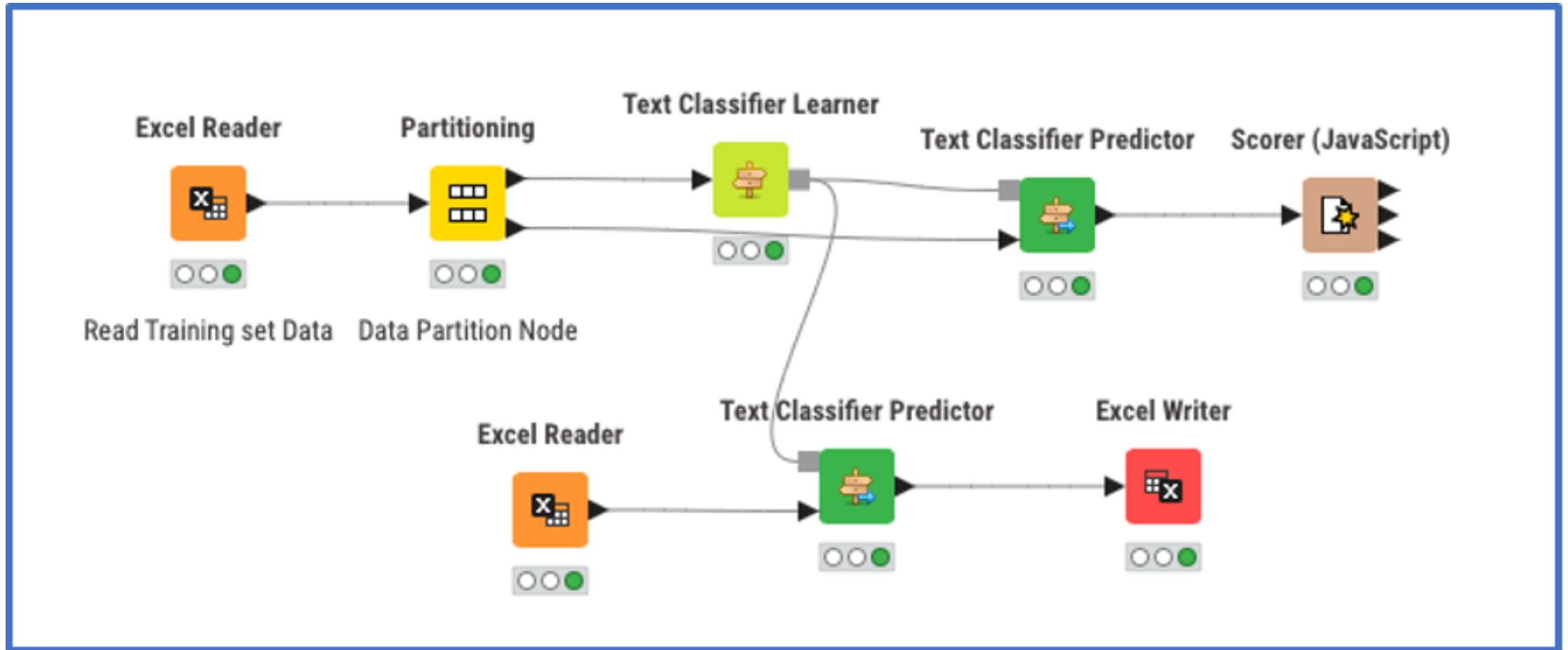
Dataset

Data Attribute	Explanation
No	Serial number of discussion text
<u>Discussion_text</u>	Text from the discussion forum for 1 thread
Label	Sentiment Label: positive, negative, neutral. (dependent variable)
<u>ModelName</u>	Name of the model. In this case, since we are building sentiment analysis, the value is the 'Sentiment'.

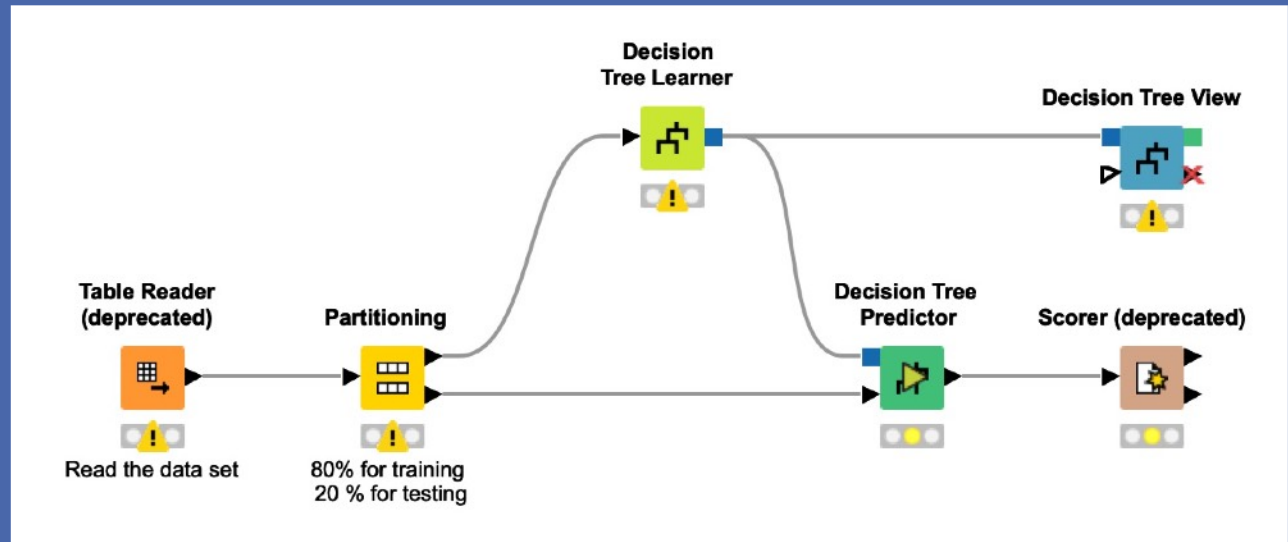
Dataset

	A	B	C	D	E
1	No	Discussion_text	Label	ModelName	TextId
2	1	...been on a low-carb diet my blood glucose has been pretty normal although I did have hypo-type symptoms when I briefly tried taking ...associated Type 2 diabetes it's something I am always...	Neutral	Sentiment	508e98fc-8db4-4283-a2a4-63c3b70acf86
3	2	A major problem with the American diet is too much refined grains and added sugar which are associated with the rise in obesity and type 2 diabetes [Source](http://www.joslin.org/news/Research-affirms-good-nutrition-can-help-prevent-and-control-type2-diabetes.html). This article is from the results of a systematic review of randomized controlled studies which represents high-tier evidence.. Edit: I'll add wheat bread can be great and is usually very high in fiber (which is absolutely critical to digestive health).. Potatoes have very few redeeming qualities (high in potassium and a few minerals are the only ones I can think of).	Negative	Sentiment	1349404b-4192-4da2-89ed-346d17351c6b
4	3	A major problem with the American diet is too much refined grains and added sugar which are associated with the rise in obesity and type 2 diabetes [Source](http://www.joslin.org/news/Research-affirms-good-nutrition-can-help-prevent-and-control-type2-diabetes.html). This article is from the results of a systematic review of randomized controlled studies which represents high-tier evidence.. Edit: I'll add wheat bread can be great and is usually very high in fiber (which is absolutely critical to digestive health).. Potatoes have very few redeeming qualities (high in potassium and a few minerals are the only ones I can think of).	Negative	Sentiment	1349404b-4192-4da2-89ed-346d17351c6b

KNIME Workflow for Sentiment Analysis



HANDS-ON II: TOPIC MODELING



Dataset

Data Attribute	Explanation
Index	Review Index
URL	Old link to the movie review (ignore the link)
Text	Review text
Sentiment	Sentiment Label: POS, NEG. (dependent variable or target label)

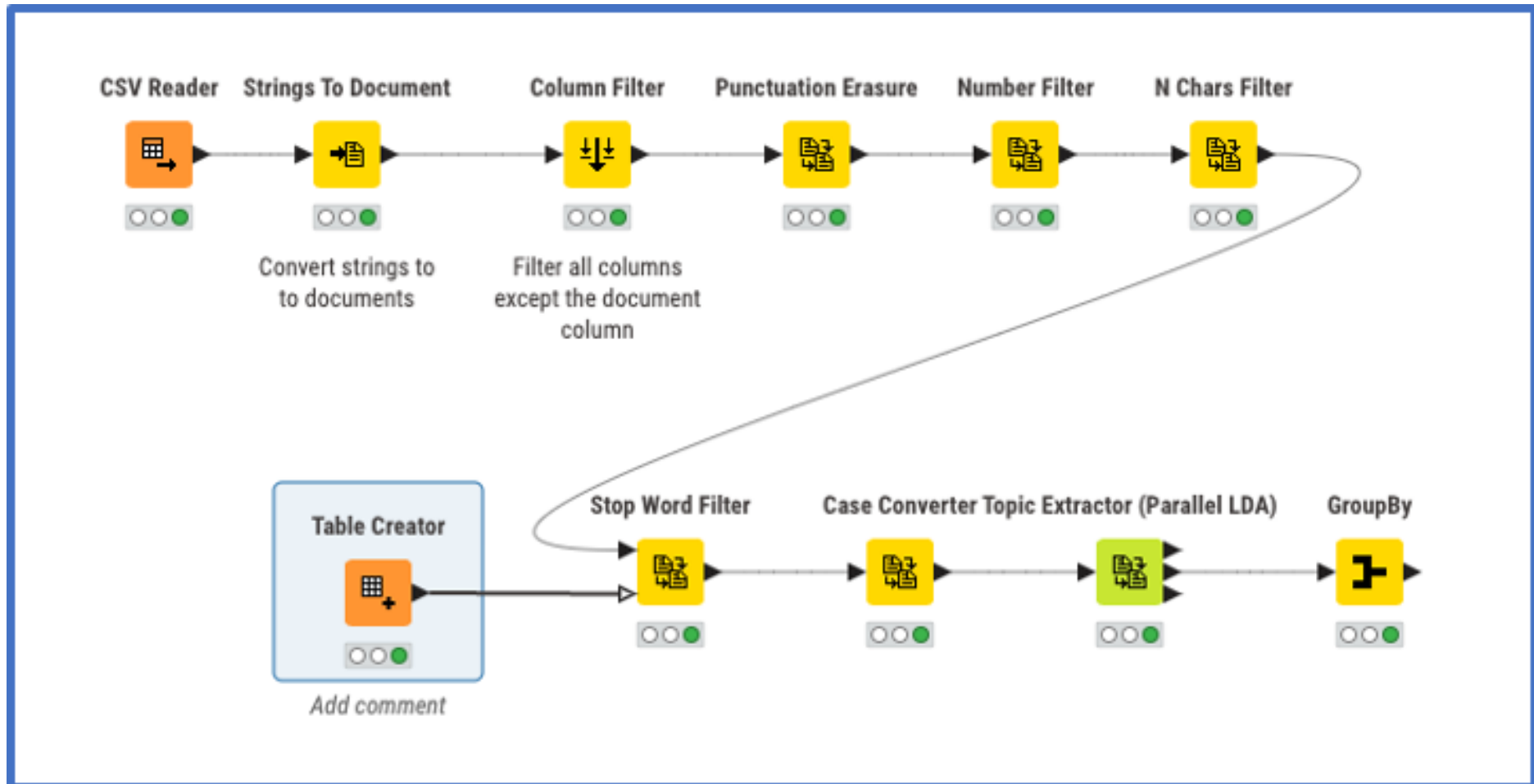
Dataset

```
IMDb-sample.csv
IMDb-sample.csv
1 "Index","URL","Text","Sentiment"
2 "3617","http://www.imdb.com/title/tt0210075/
usercomments","Girlfight follows a project dwelling New York high
school girl from a sense of futility into the world of amateur
boxing where she finds self esteem, purpose, and much more.
Although the film is not about boxing, boxing is all about the
film. So much so you can almost smell the sweat. Technically and
artistically a good shoot with an sense of honesty and reality
about it, Girlfight is no chick flick and no Rocky. It is, rather,
a very human drama which even viewers who don't know boxing will
be able to connect with.Girlfight follows a project dwelling New
York high school girl from a sense of futility into the world of
amateur boxing where she finds self esteem, purpose, and much
more.","POS"
3 "3671","http://www.imdb.com/title/tt0337640/
usercomments","Hollywood North is an euphemism from the movie
industry as they went to Canada to make movies because of tax
breaks and cheaper costs in a civilized city like Toronto, in this
case, later in Vancouver. Peter O'Brian, the director, probably
saw a lot of the invaders from California that this movie seems to
be the right way to deal with the arriving personalities trying to
capitalize on the economics that Canada presented.Needless to say,
Moon Lantern, the successful novel written by a Canadian author is
turned into Flight to Bogota, which has nothing to do with the
original film. A great egotistical has-been, Michael Baytes, who
is obsessed with what is happening in Iran, is offered the lead
part, which turns to be a disaster.The film seems to be saying
that too many cooks have spoiled the broth, which seems to be the
```

Dataset

Index <i>Number (integer)</i>	URL <i>String</i>	Text <i>String</i>	Sentiment <i>String</i>
6168	http://www.imdb.com/title/tt0217...	Hmm, IMDb rating of 7.5, good comments, bla, bla ... okay, two of my friends and me, we ord...	NEG
3963	http://www.imdb.com/title/tt0433...	How many fricken' times do we have to see a spook walking by in the background & peaking...	NEG
7173	http://www.imdb.com/title/tt0113...	I HATE plane crash movies...ALL of them! In fact, I hate them all with a passion! First of all, t...	NEG
5118	http://www.imdb.com/title/tt0805...	I am a Christian and I say this movie had terrible acting, unreal situations and a completely f...	NEG
8693	http://www.imdb.com/title/tt0840...	I am from the Dallas/Fort Worth area and lived in Arlington for a few years. This movie was ...	NEG
12305	http://www.imdb.com/title/tt0494...	I am watching this movie right now on WTN because that was the channel that the TV was t...	NEG
4438	http://www.imdb.com/title/tt0253...	I borrowed this movie from library think it might be delightful. How wrong am I!It is such a b...	NEG
6091	http://www.imdb.com/title/tt0273...	I bought this movie exciting a gloriously gratuitous, over the top, entertaining bloodbath. I go...	NEG
6117	http://www.imdb.com/title/tt0083...	I can't say this is the worst movie ever made, but personally I think of it that way because wh...	NEG
4198	http://www.imdb.com/title/tt0309...	I caught this Cuban film at at an arthouse film club. It was shown shortly after the magisteri...	NEG
30	http://www.imdb.com/title/tt0127...	I did not like the idea of the female turtle at all since 1987 we knew the TMNT to be four brot...	NEG
2820	http://www.imdb.com/title/tt0108...	I don't know much about Tobe Hooper, or why he gets his name in the title, but maybe he sh...	NEG
8129	http://www.imdb.com/title/tt0071...	I dunno sometimes...you try and try and try to be charitable towards all the B thru Z grade m...	NEG
3118	http://www.imdb.com/title/tt0448...	I felt that the movie was dry... very disappointing no plot..kept waited for something to happ...	NEG
12277	http://www.imdb.com/title/tt0074...	I found The Arab Conspiracy in a bargain bin and thought I'd uncovered a lost treasure. Folks...	NEG
9192	http://www.imdb.com/title/tt0317...	I found this movie in the 'horror' section of my video store. That seems to make sense as m...	NEG
7270	http://www.imdb.com/title/tt0783...	I gave this loooooooooooooong film a 2 because of the attractive actors and semi-sexy love sc...	NEG
10283	http://www.imdb.com/title/tt0480...	I hate to sound like an 'old person', but frankly I haven't seen too many movies that I like that ...	NEG
8668	http://www.imdb.com/title/tt0086...	I have never read the Bradbury novel that this movie is based on but from what I've gathered,...	NEG
11689	http://www.imdb.com/title/tt0454...	I have never seen any of Spike Lee's prior films, as their trailers never caught my interest. I h...	NEG

KNIME Workflow for Topic Modeling



Thank you!

Raghava Mukkamala

rrm.digi@cbs.dk

<https://www.cbs.dk/staff/rrmdigi>

<https://raghavamukkamala.github.io/>

<https://cbsbda.github.io/>