M.Sami SİVRİ

Overview

What is KNIME ?

- Konstanz Information Miner
  - January 2004
  - University of Konstanz
- An open source platform for data analysis, predictive analytics and modelling
- Is not based on a scripting language. Instead, it has a graphical interface.
- www.knime.com
- https://www.youtube.com/watch?v=ft7Ksgss3Tc

KNIME Usage Statistics

- 2015

> 2,000,000	**Downloads** annually (in 2015)
> 70,000	**Individuals** using KNIME
> 5,000	**Organizations** using KNIME
> 700	**Customers** paying for KNIME

- 2013

>25,000	**Individuals** using KNIME
>3,000	**Organizations** using KNIME
>300	**Customers** paying for KNIME

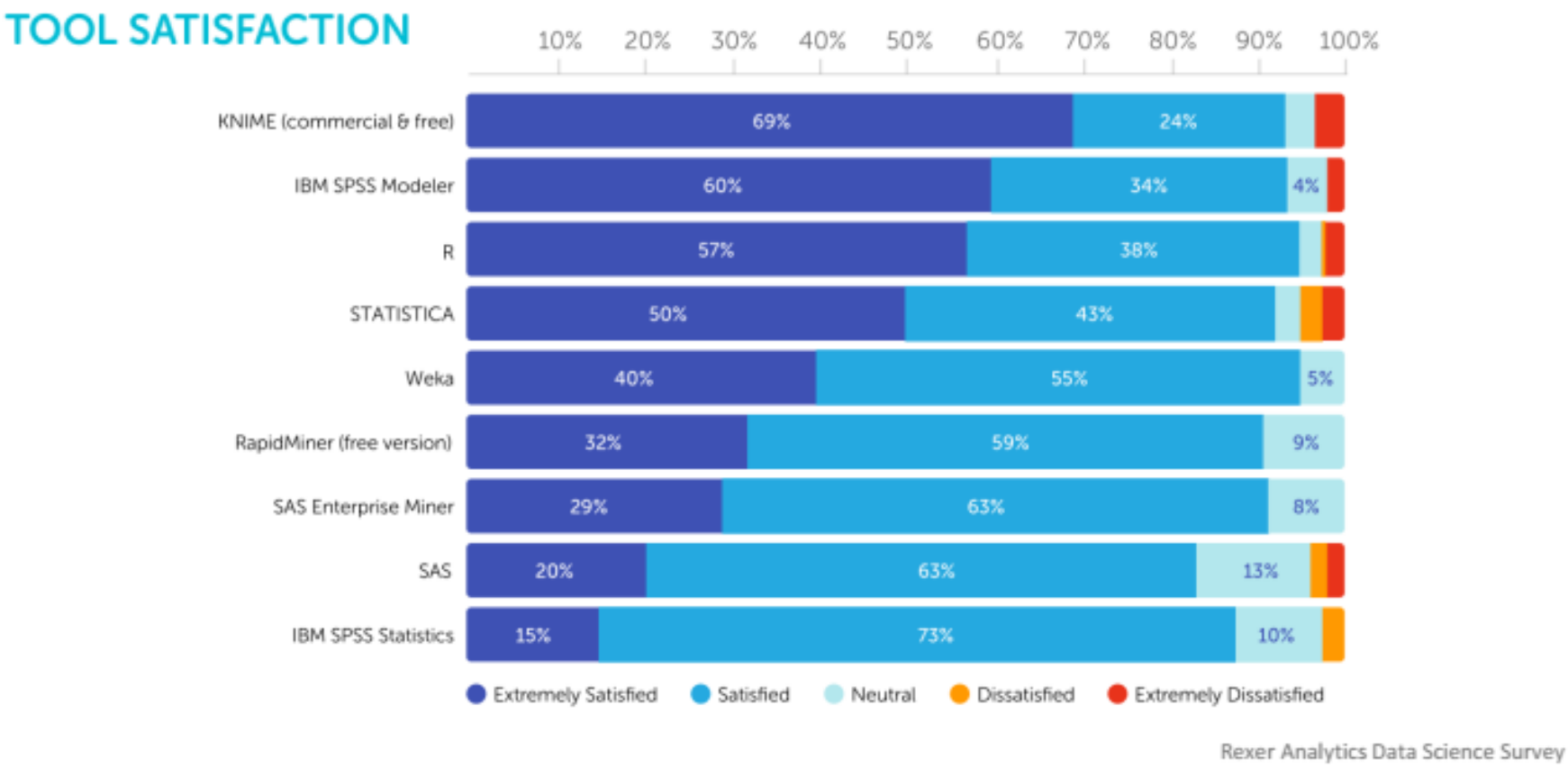
Application Areas

User Satisfaction

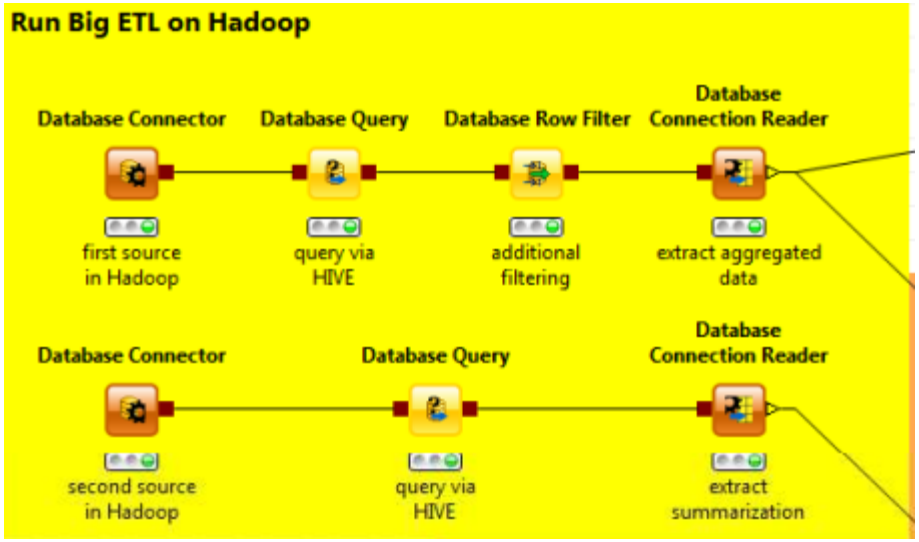
	Overall	IBM SPSS Statistics	IBM SPSS Modeler	KNIME	R	Rapid Miner	SAS	SAS Enterprise Miner	STATISTICA (StatSoft)	Weka
Software or platform Type:		Proprietary	Proprietary	Open Source	Open Source	Open Source	Proprietary	Proprietary	Proprietary	Open Source
Quality and accuracy of model performance	4,34	3,93	4,32	4,32	4,43	4,37	4,34	4,26	4,69	4,21
Dependability/Stability of software	4,25	4,19	4,05	4,43	4,34	4,22	4,32	4,44	4,56	3,73
Variety of available algorithms	4,20	3,69	4,30	4,48	4,72	4,54	4,01	4,00	4,63	4,33
Ease of use	4,19	4,28	4,60	4,76	3,58	4,47	3,69	4,00	4,49	4,06
Ability to automate repetitive tasks	4,17	3,75	3,96	4,39	4,39	4,40	4,31	4,00	4,45	3,71
Data manipulation capabilities	4,15	4,00	4,32	4,53	4,10	4,27	4,45	3,82	4,41	3,52
Quality of output / Ease of interpretation	4,10	3,91	4,04	4,39	4,04	4,36	3,69	4,00	4,53	3,66
Good metrics of model quality	4,10	3,85	3,96	4,05	4,13	4,28	4,08	4,18	4,50	3,85
Good variable discovery, profiling and selection	4,03	3,70	4,06	4,17	3,98	4,33	3,81	4,35	4,44	3,69
Quality of user interface	4,03	4,11	4,53	4,62	3,36	4,45	3,58	3,91	4,49	3,59
Ease of model deployment (scoring models to other data sets)	4,03	3,61	4,13	4,43	3,82	4,20	3,90	4,21	4,46	3,77
Speed	4,02	3,84	4,13	4,12	3,58	3,90	4,08	3,97	4,48	3,53
Data quality assessment and data preparation capabilities	4,00	3,94	4,22	4,33	3,76	4,20	4,05	3,68	4,38	3,53
Ability to handle very large data sets	3,99	3,82	4,21	4,35	2,95	3,74	4,41	4,44	4,58	3,03
Ability to modify algorithm options to fine-tune analyses	3,95	3,17	3,59	3,98	4,33	4,23	3,97	3,97	4,33	3,88
Enables mining within one's database	3,94	3,54	4,26	4,12	3,75	4,10	3,92	4,00	4,19	3,61
Ability to easily incorporate data at different levels of granularity (e.g. transaction data and customer data)	3,90	3,56	4,06	4,24*	3,77	3,99	4,14	3,94	4,24*	3,29
Useful help menu, demos and tutorials	3,87	3,83	3,99	3,93	3,68	3,90	3,76	3,79	4,35	3,61
Strong graphical visualization of models	3,83	3,24	3,68	3,88	4,14	4,28	3,02	3,88	4,62	3,28
Cost of software	3,79	3,16	3,00	4,93	4,90	4,82	2,33	2,74	3,90	4,88
Overall	4,04	3,76	4,07	4,32	3,99	4,25	3,89	3,98	4,44	3,74

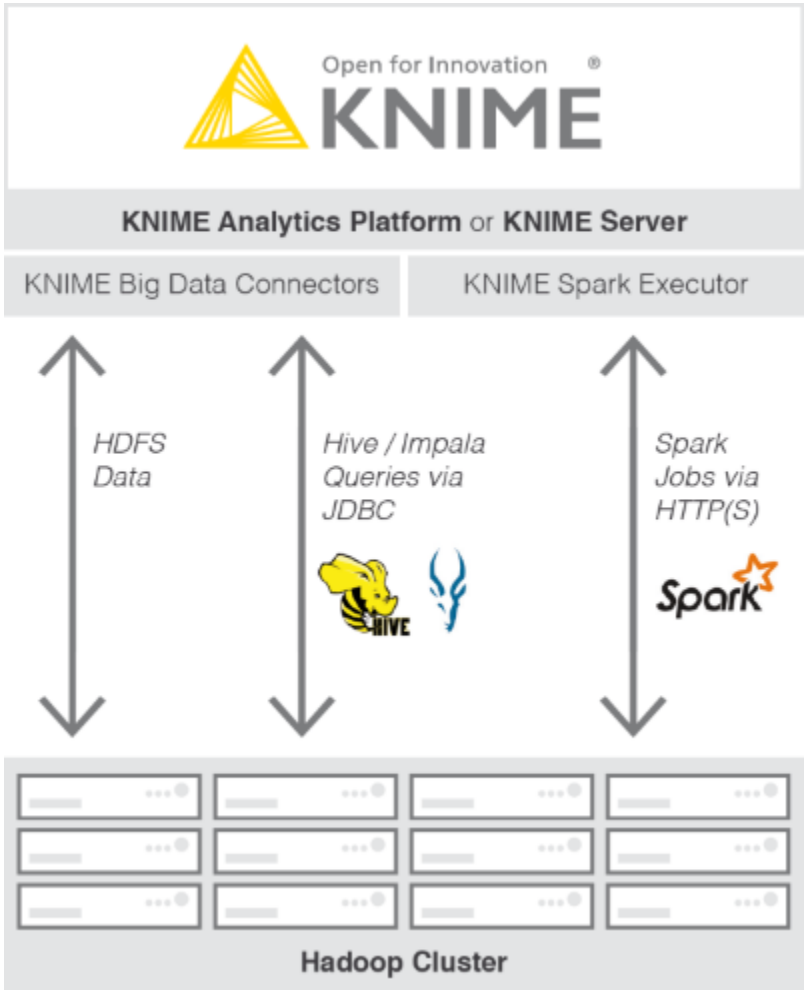
2012 & 2013 Rexer Analytics Survey

User Satisfaction

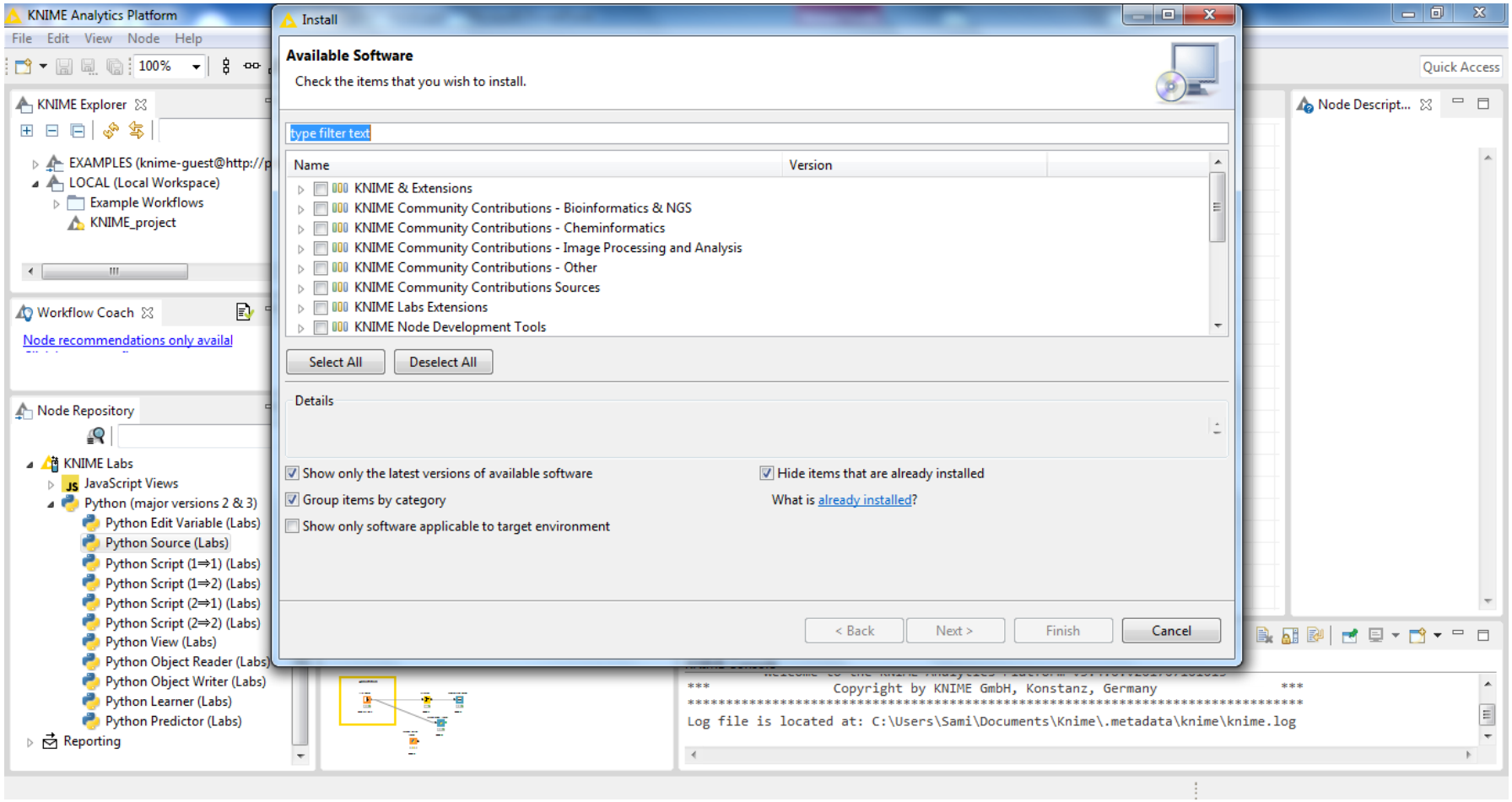


KNIME and Big Data



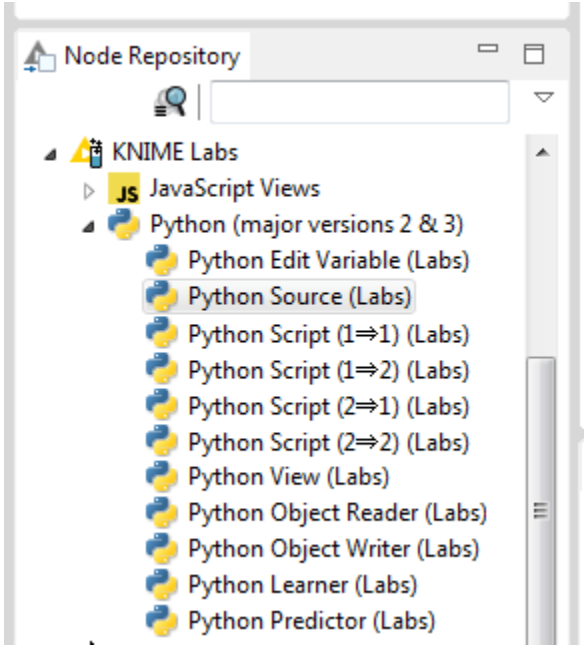


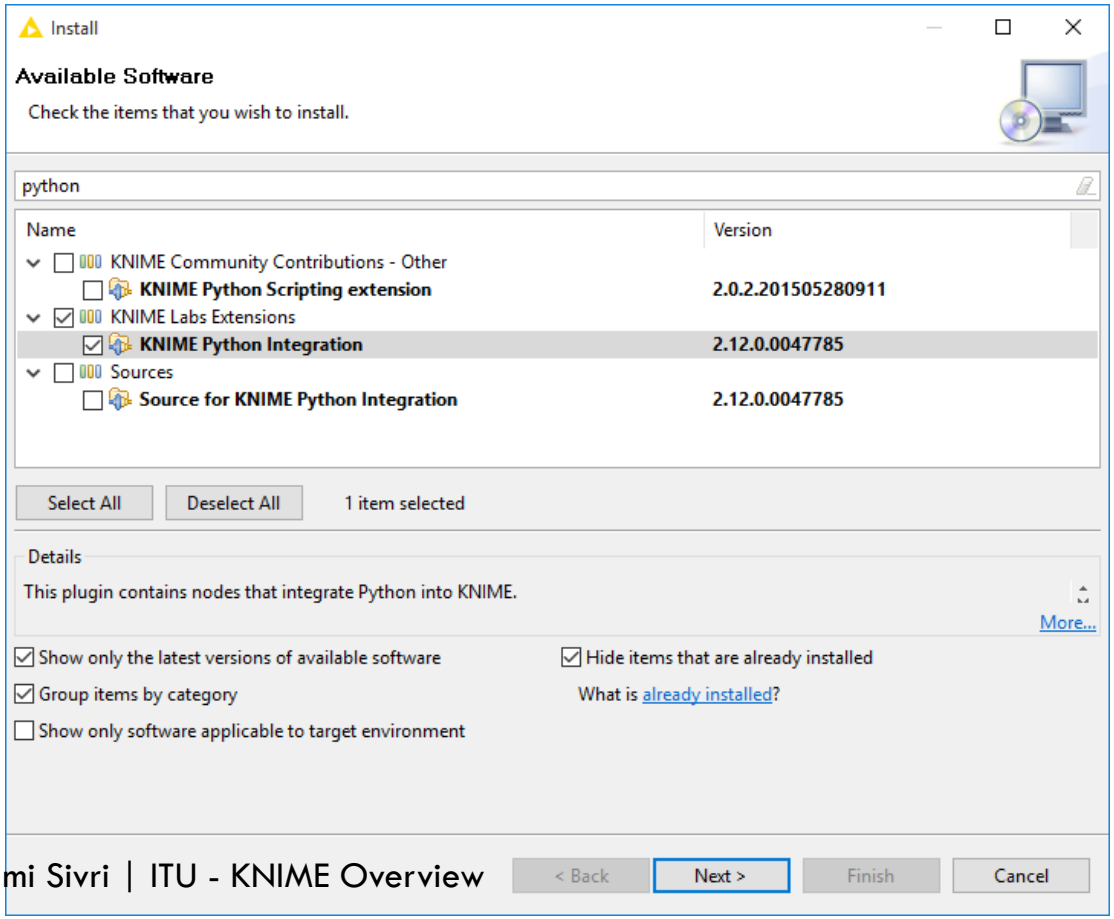
KNIME Extensions



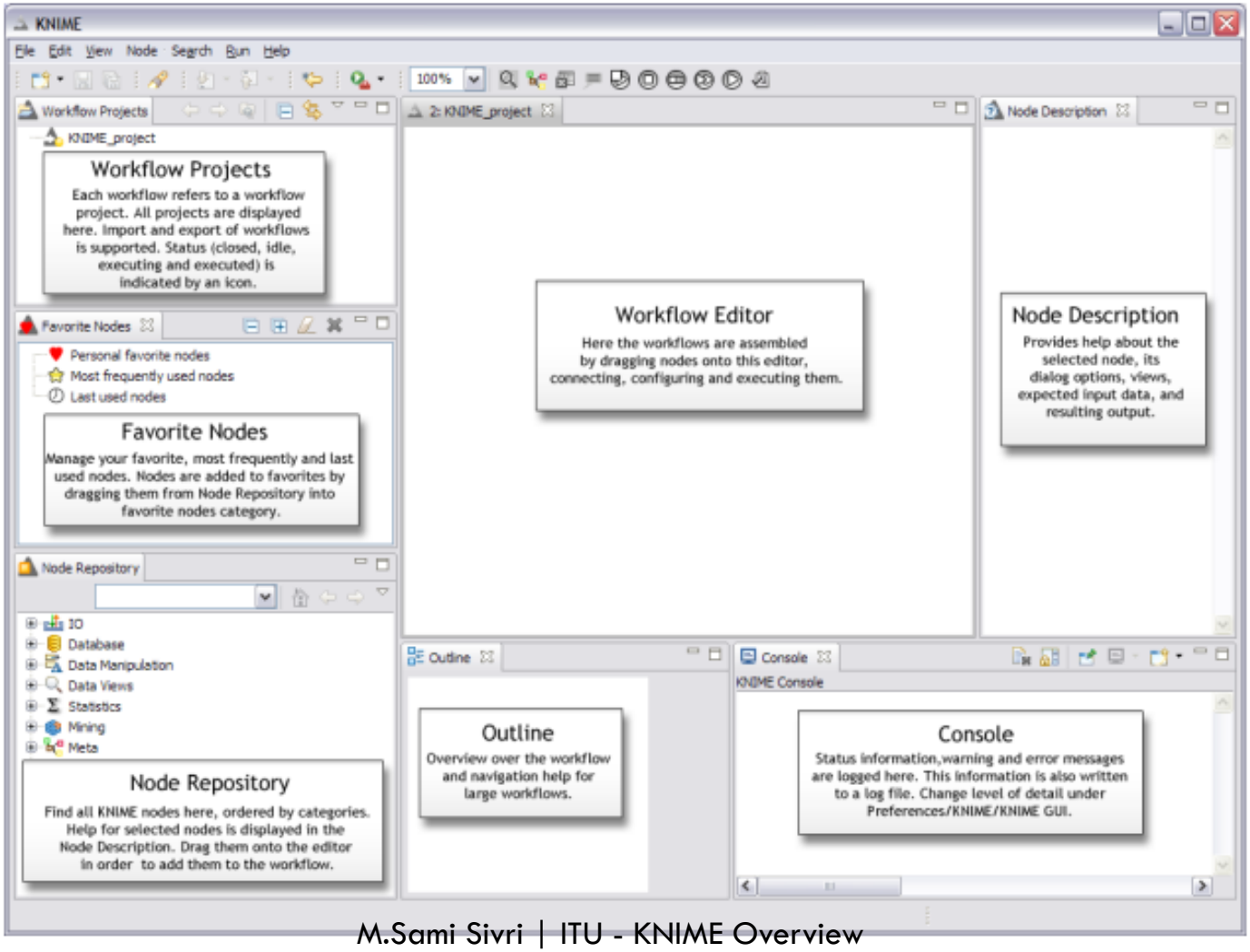
KNIME Python Integration

- https://www.knime.com/blog/how-to-setup-the-python-extension

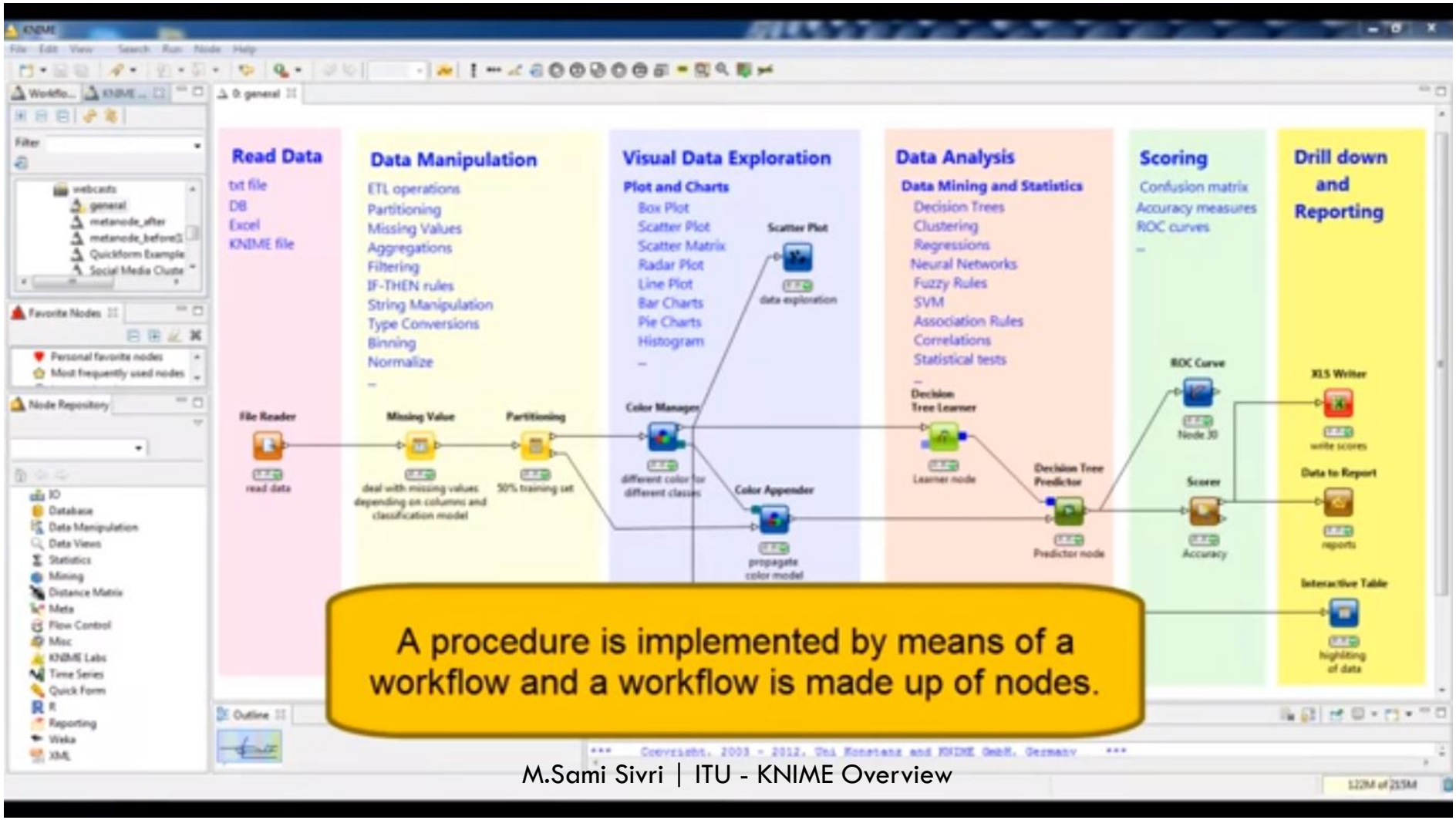




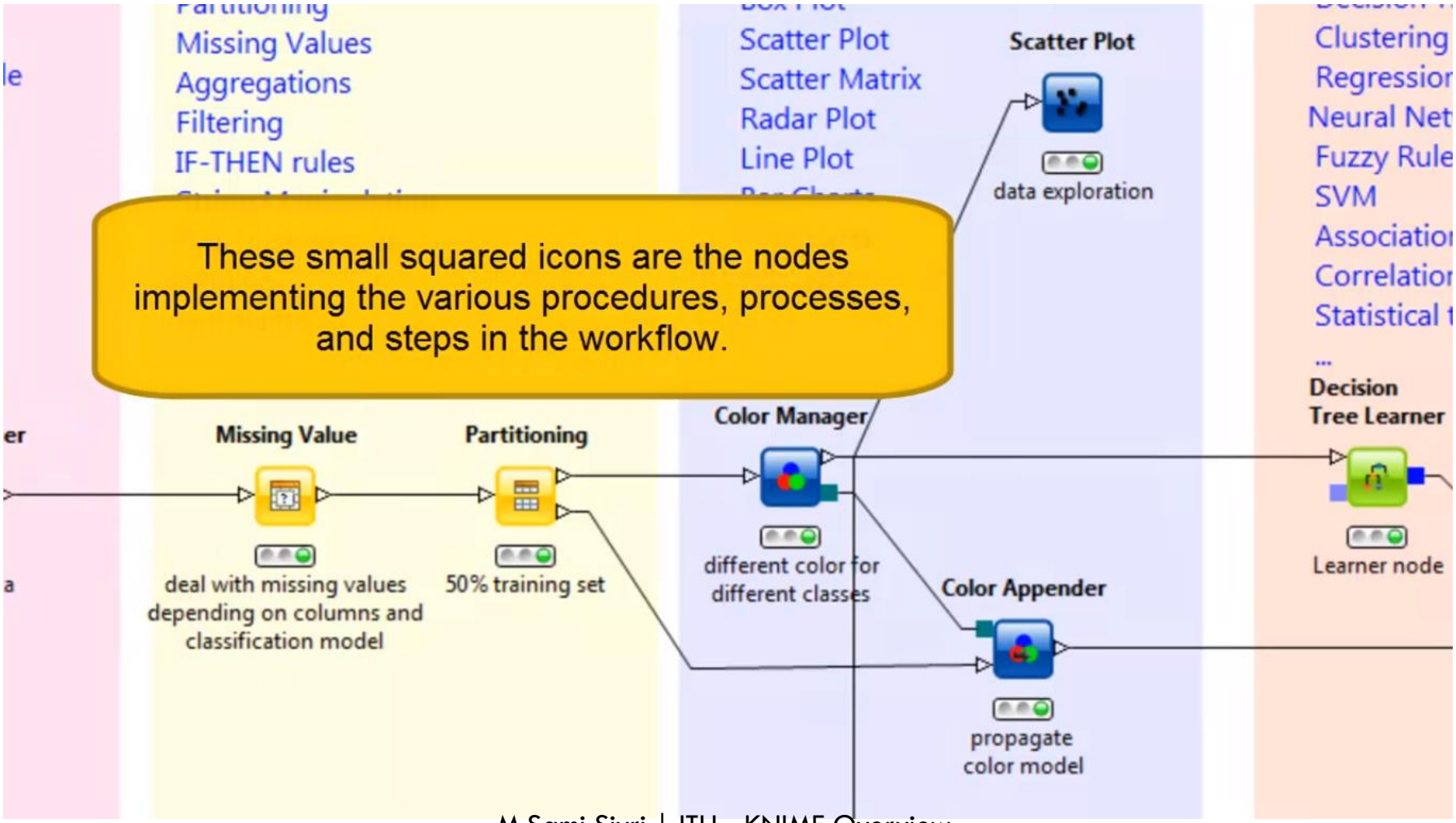
KNIME Workbench



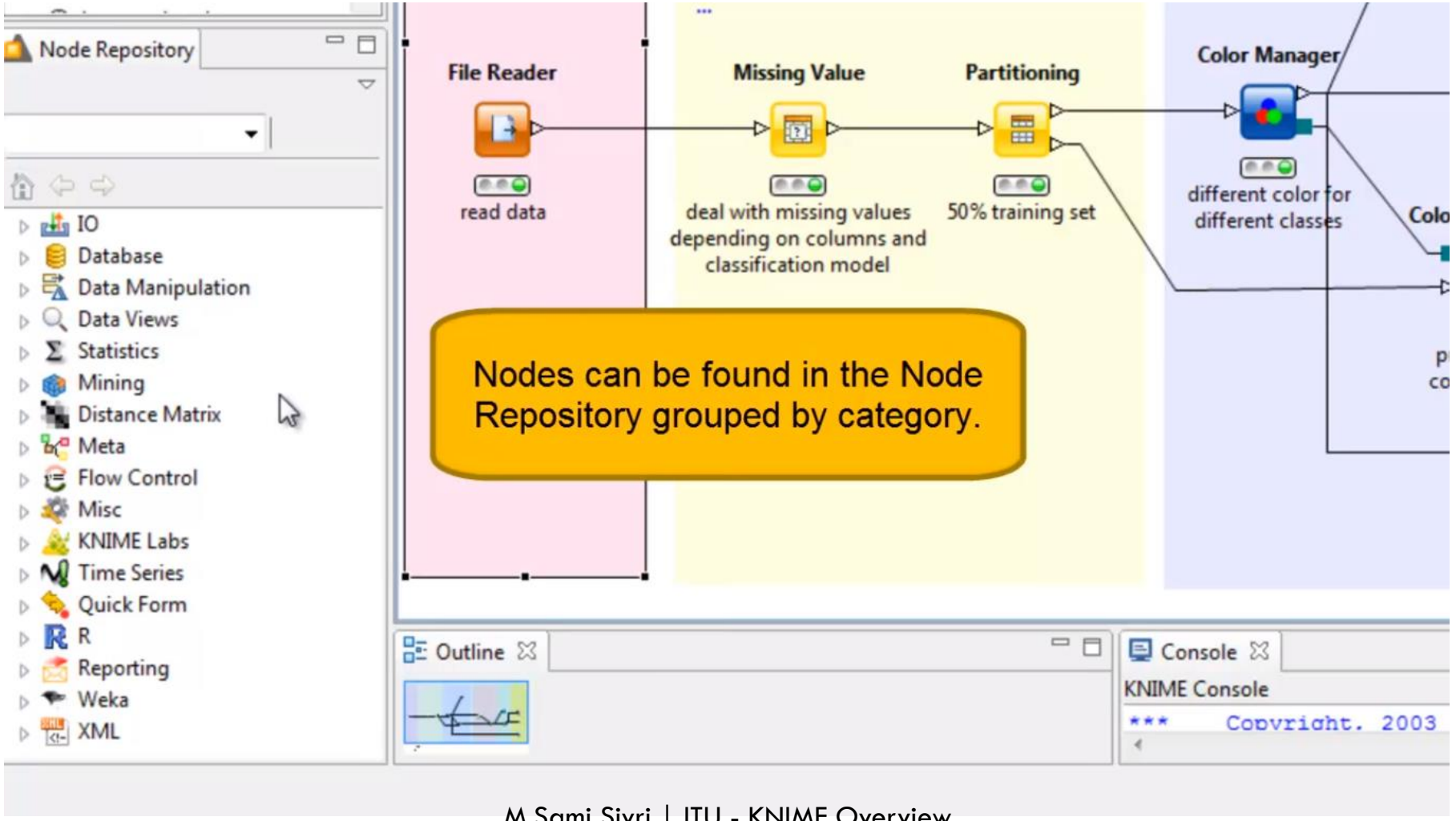
KNIME Workflow



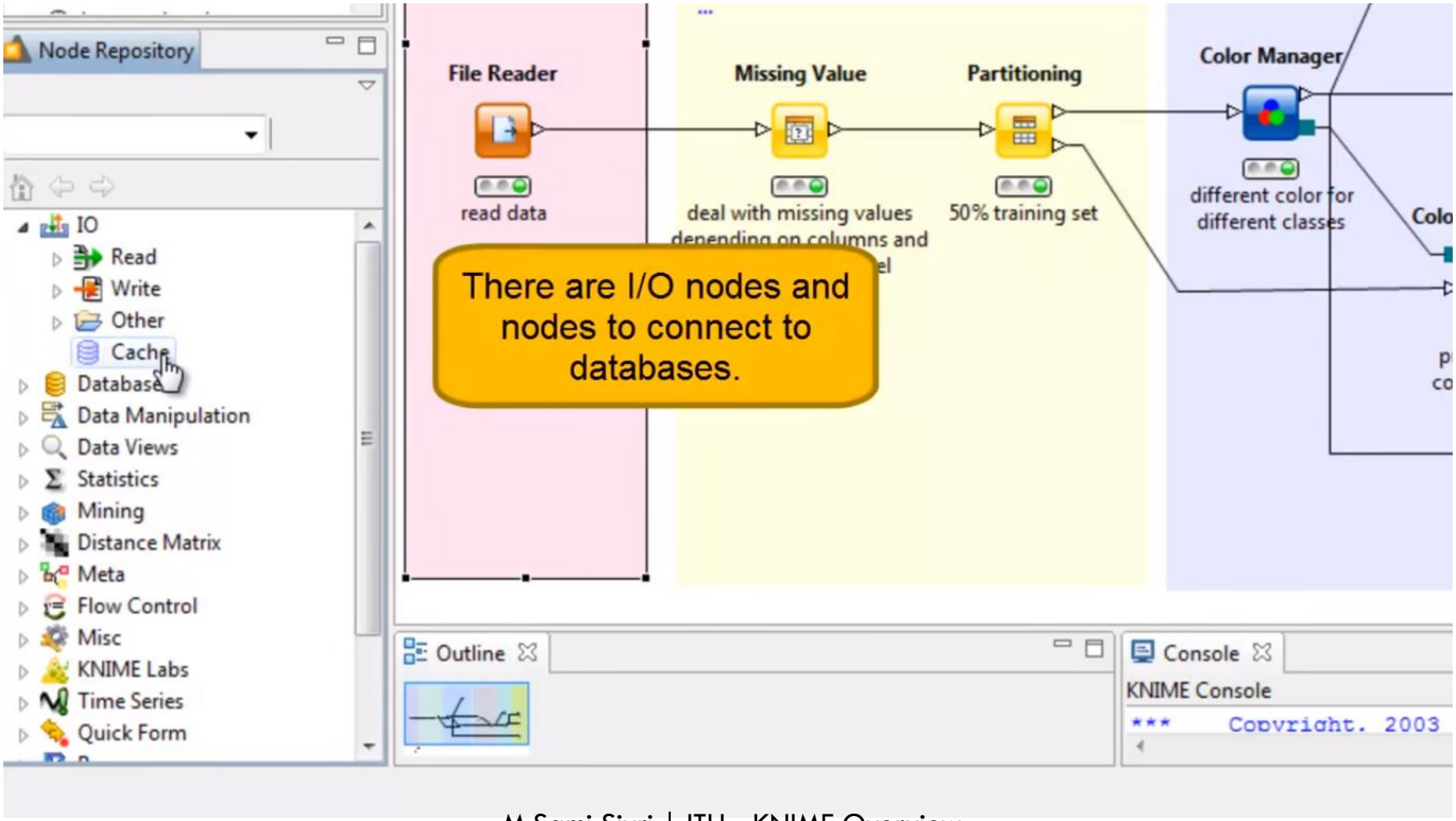
Nodes


Missing Values
Aggregations
Filtering
IF-THEN rules
Scatter Plot
Scatter Matrix
Radar Plot
Line Plot
Scatter Plot
data exploration
Clustering
Regressior
Neural Net
Fuzzy Rule
SVM
Associatio
Correlatio
Statistical
...
Decision
Tree Learner
Learner node
These small squared icons are the nodes implementing the various procedures, processes, and steps in the workflow.
Missing Value
deal with missing values depending on columns and classification model
Partitioning
50% training set
Color Manager
different color for different classes
Color Appender
propagate color model

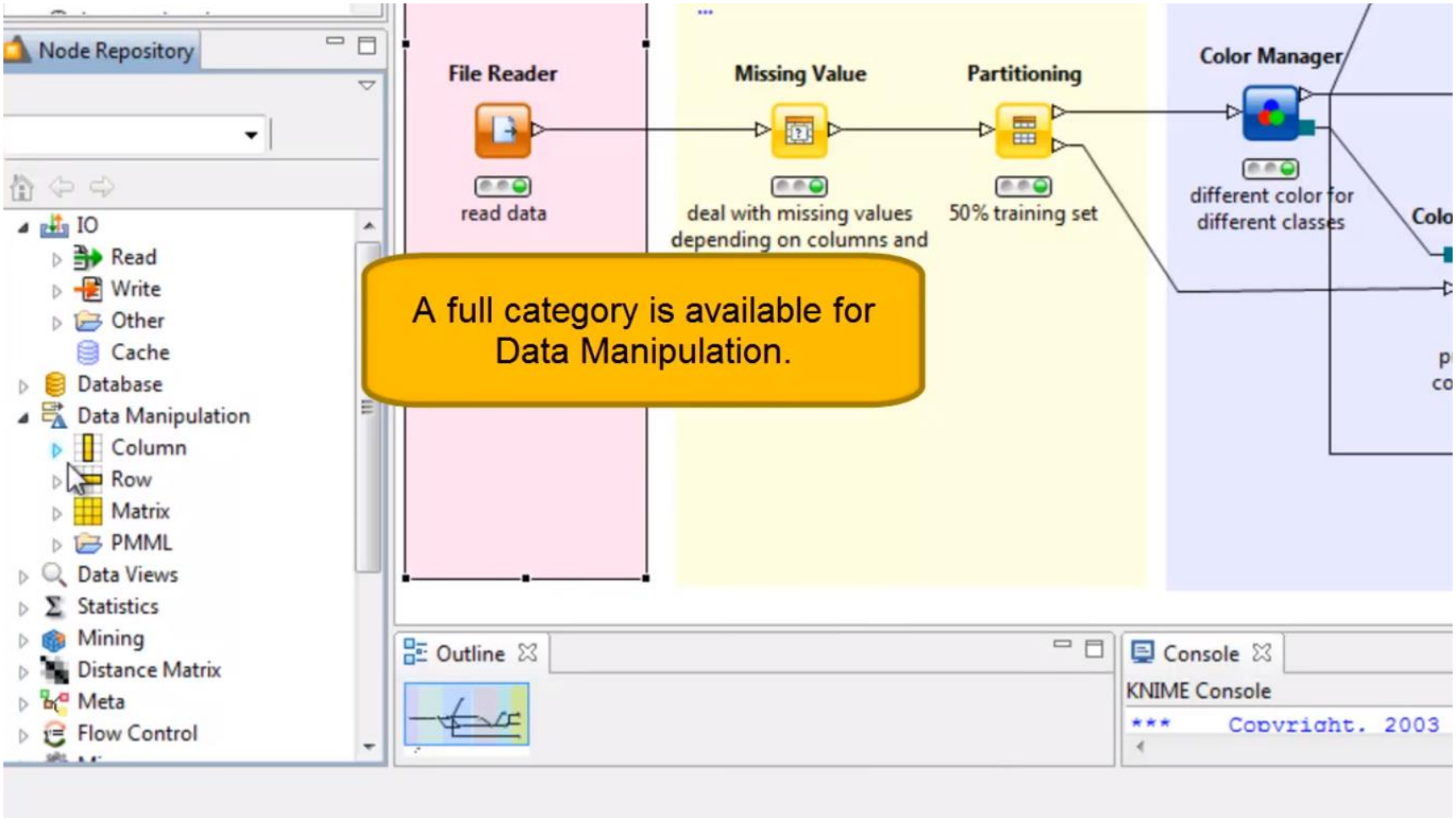
Node Repository



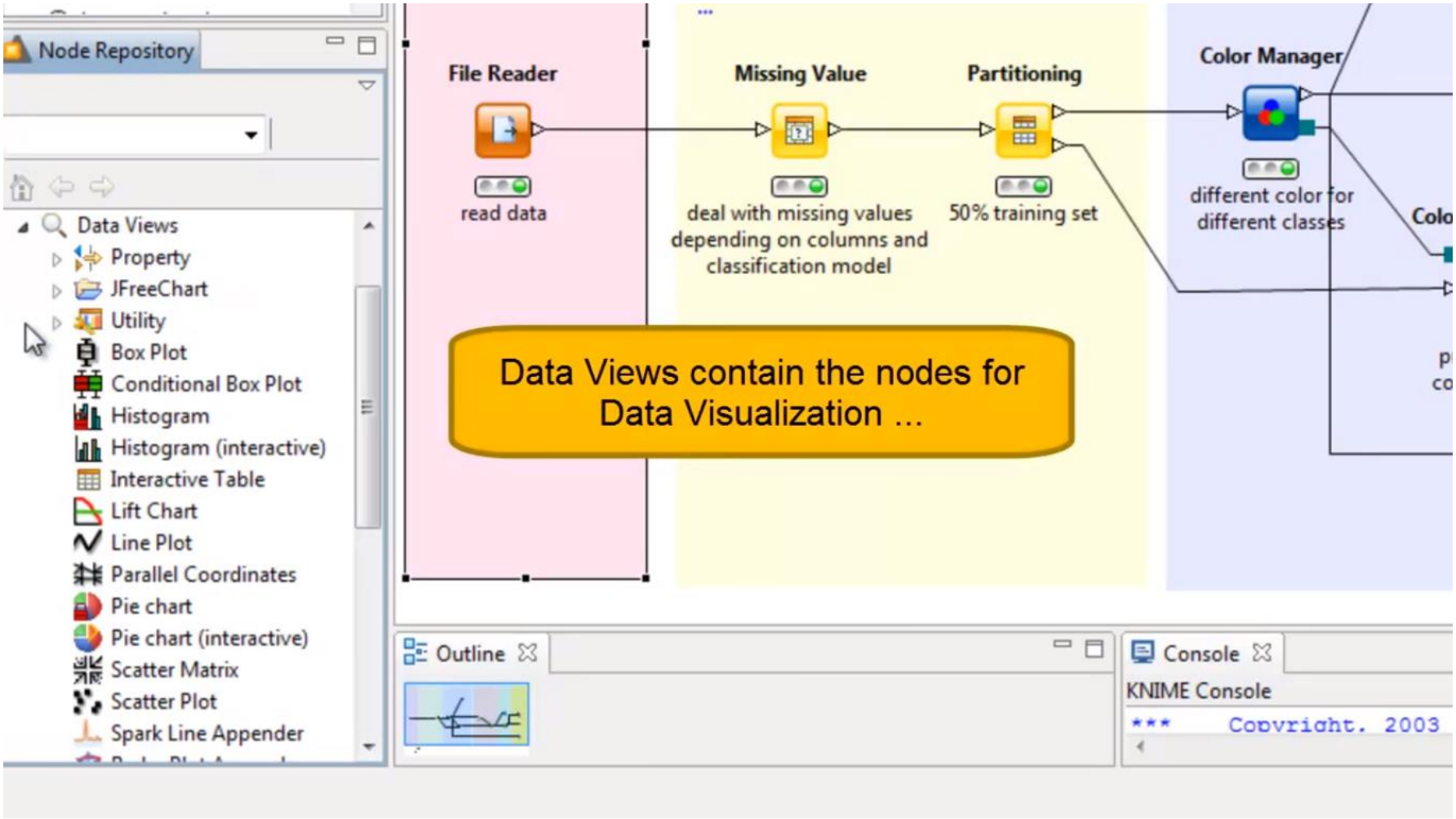
Node Categories



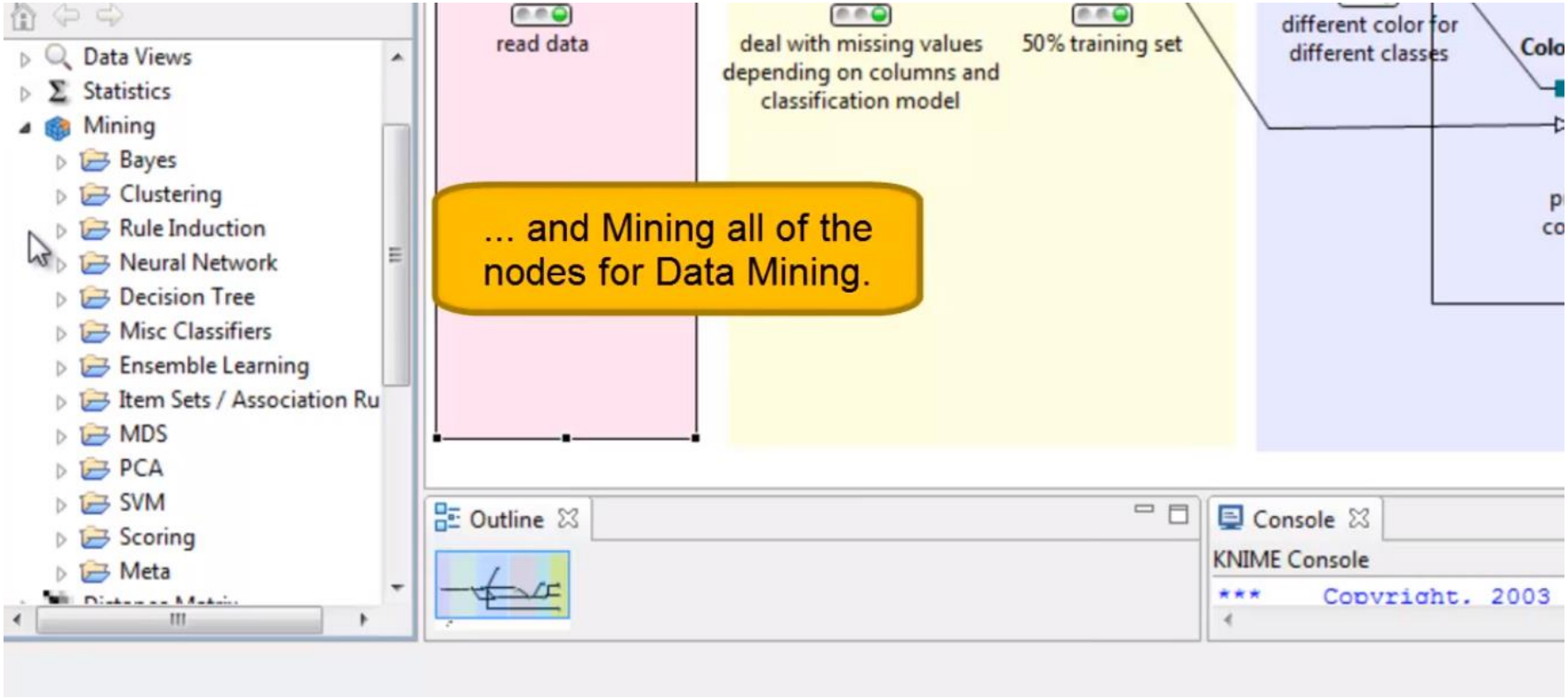
Node Categories



Node Categories

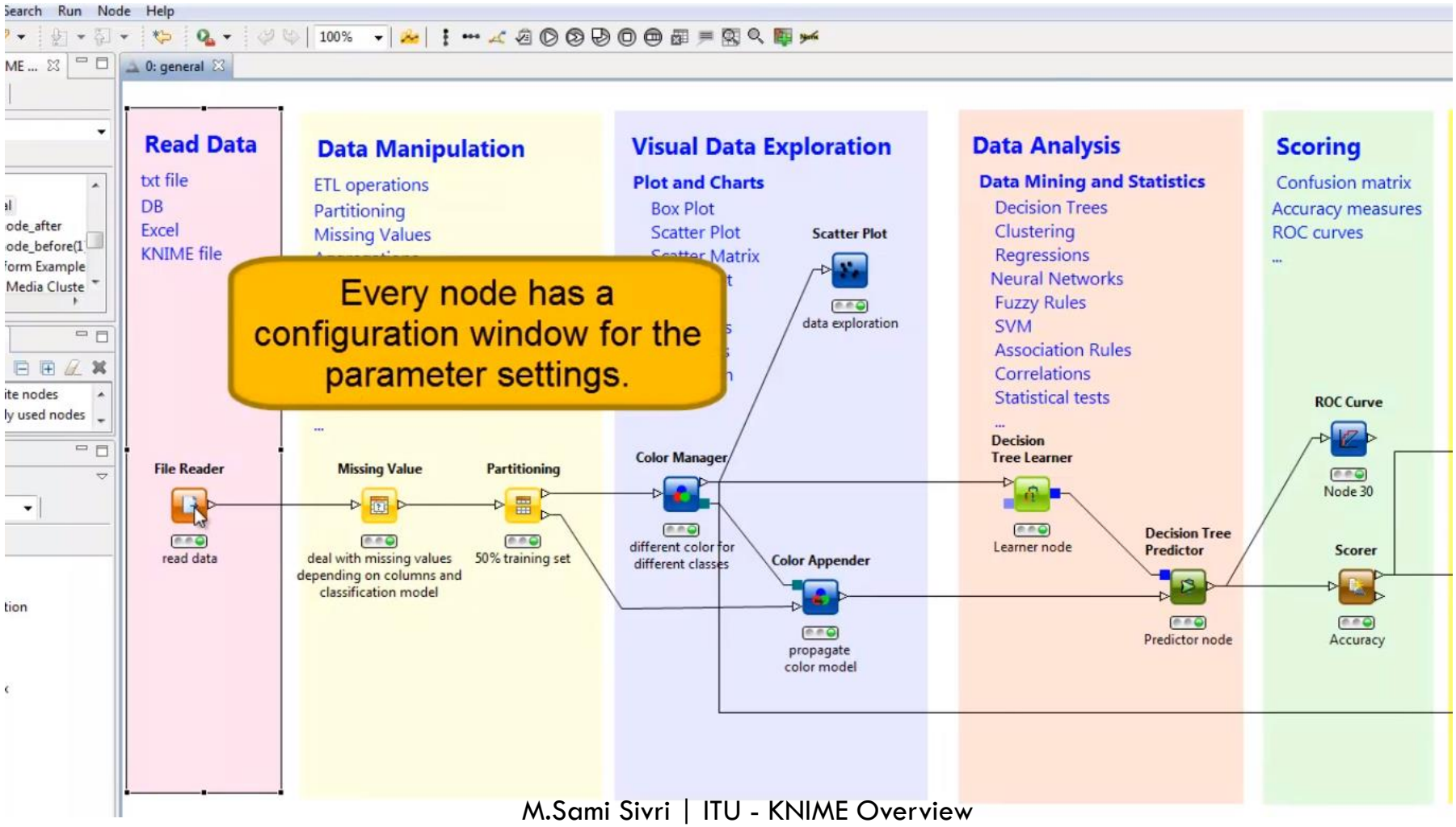


Node Categories



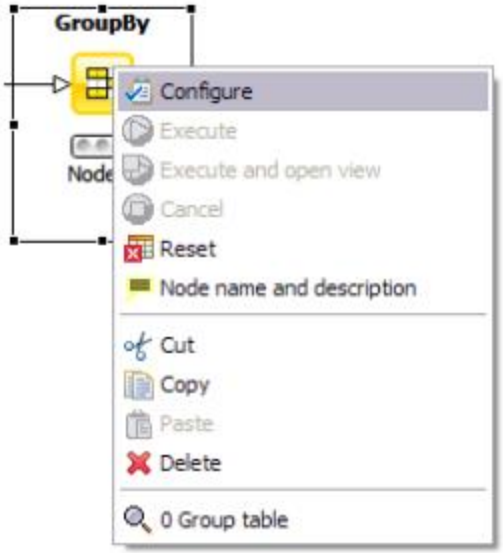
- https://www.knime.com/features

Node Configuration



Node Status

- When a node is dragged onto the workflow editor the status light shows red
  - Means that the node has to be configured in order to be able to be executed.
- A node is configured by right clicking it, choosing ‘Configure’, and adjusting the necessary settings in the node's dialog

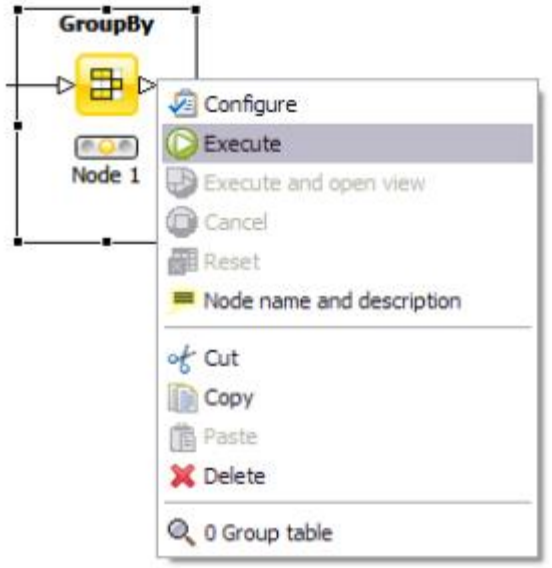


Node Status

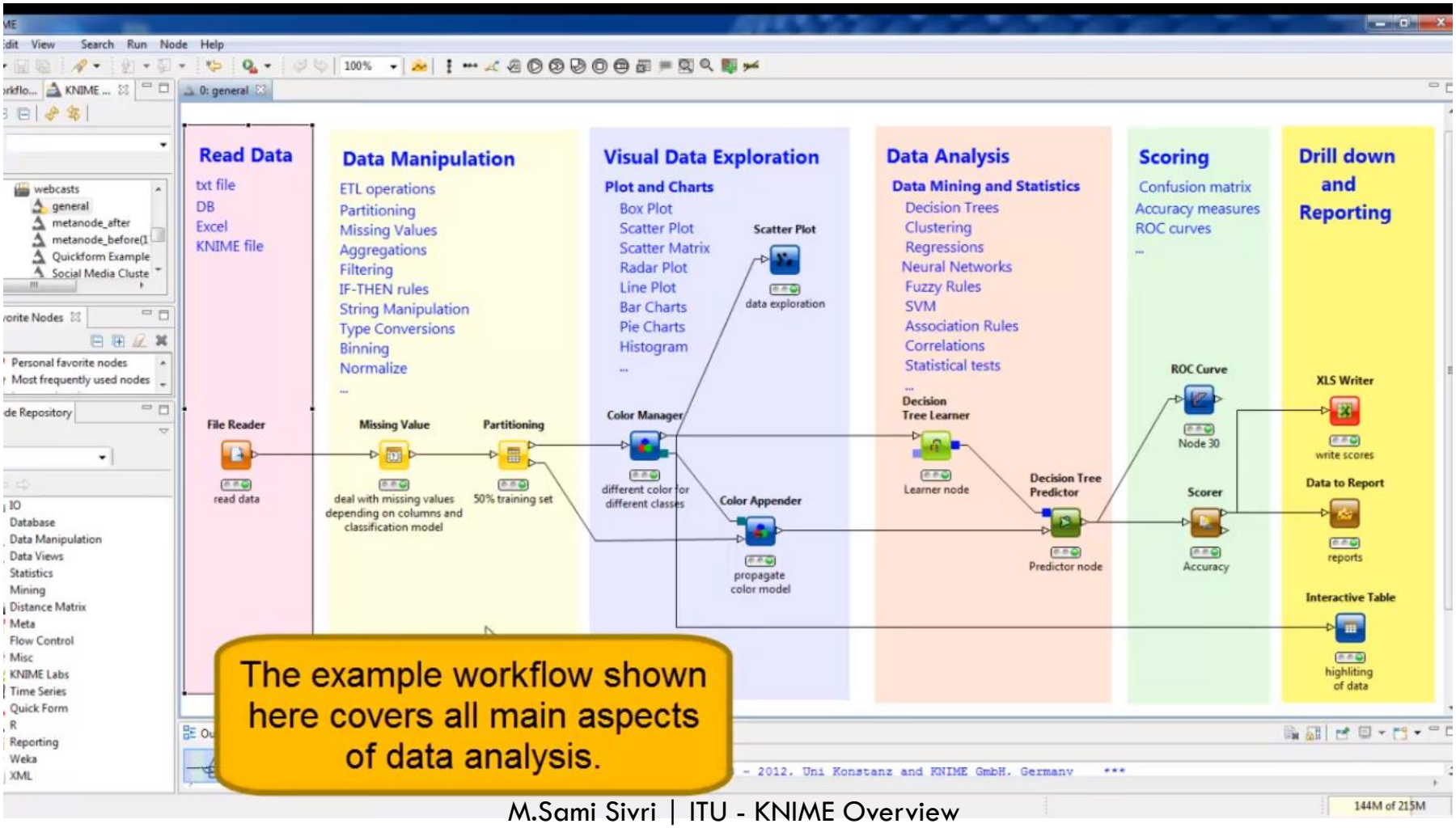
- When the dialog is closed by pressing the ‘OK’ button, the node is configured and the status light changes to yellow
  - The node is ready to be executed.

Node Status

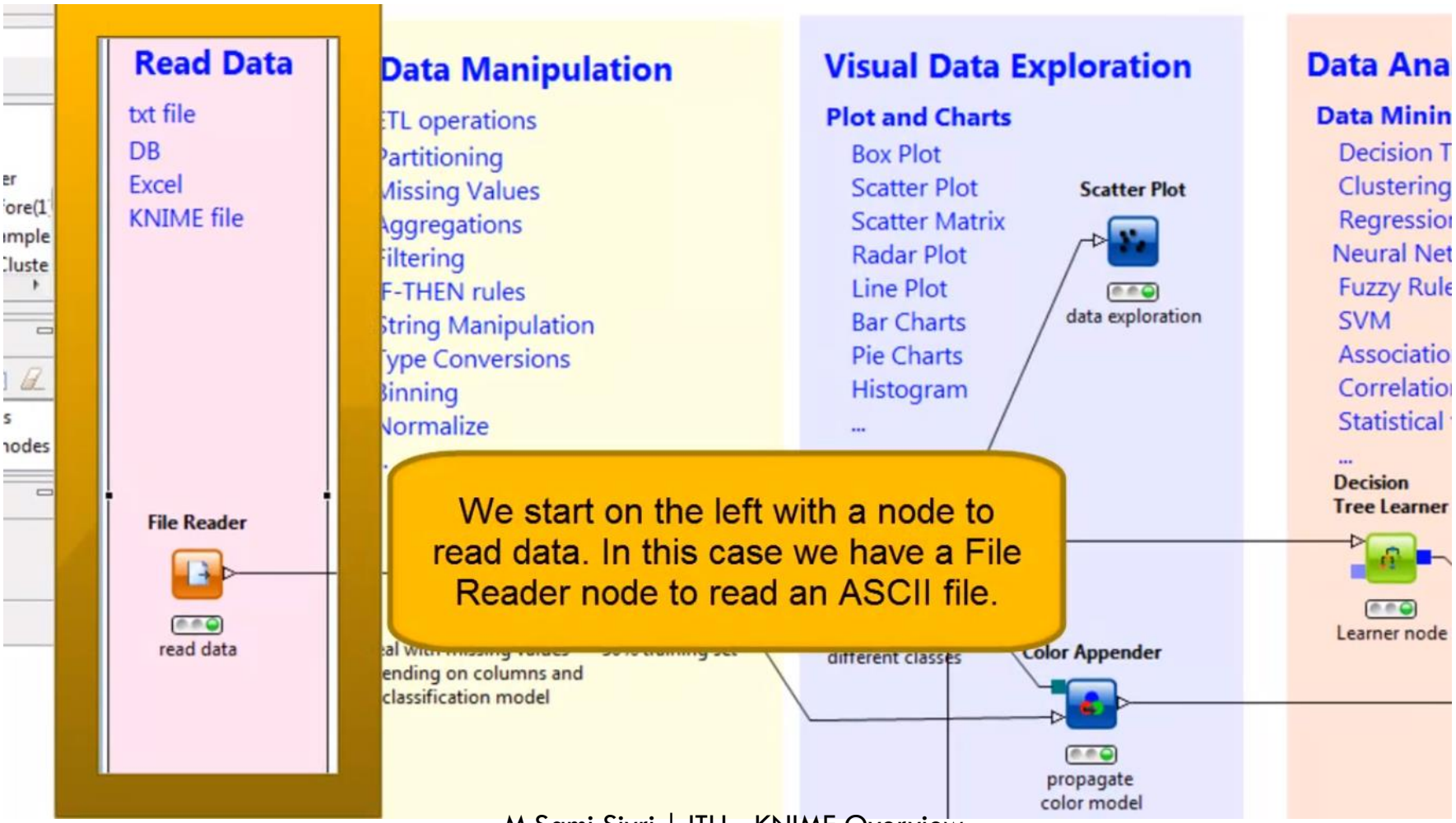
- Right-click on the node again shows an enabled ‘Execute’ option; pressing it will execute the node and the result of this node will be available at the out-port.
- After a successful execution the status light of the node is green.



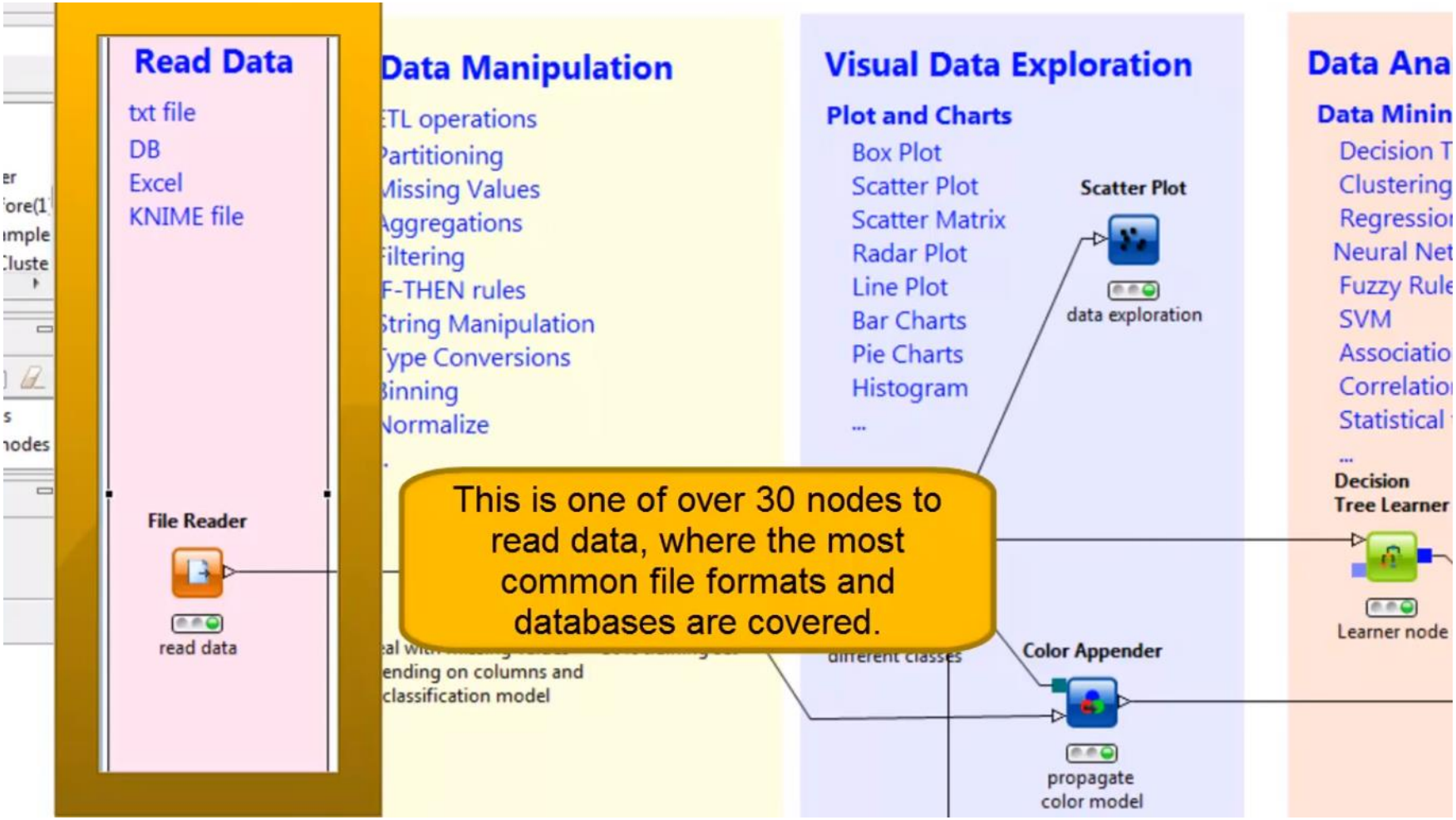
Main Steps


Read Data
txt file
DB
Excel
KNIME file
Data Manipulation
ETL operations
Partitioning
Missing Values
Aggregations
Filtering
IF-THEN rules
String Manipulation
Type Conversions
Binning
Normalize
...
Visual Data Exploration
Plot and Charts
Box Plot
Scatter Plot
Scatter Matrix
Radar Plot
Line Plot
Bar Charts
Pie Charts
Histogram
...
Data Analysis
Data Mining and Statistics
Decision Trees
Clustering
Regressions
Neural Networks
Fuzzy Rules
SVM
Association Rules
Correlations
Statistical tests
...
Scoring
Confusion matrix
Accuracy measures
ROC curves
...
Drill down and Reporting
File Reader
read data
Missing Value
deal with missing values depending on columns and classification model
Partitioning
50% training set
Color Manager
different color for different classes
Scatter Plot
data exploration
Color Appender
propagate color model
Decision Tree Learner
Learner node
Decision Tree Predictor
Predictor node
ROC Curve
Node 30
Scorer
Accuracy
XLS Writer
write scores
Data to Report
reports
Interactive Table
highliting of data
The example workflow shown here covers all main aspects of data analysis.

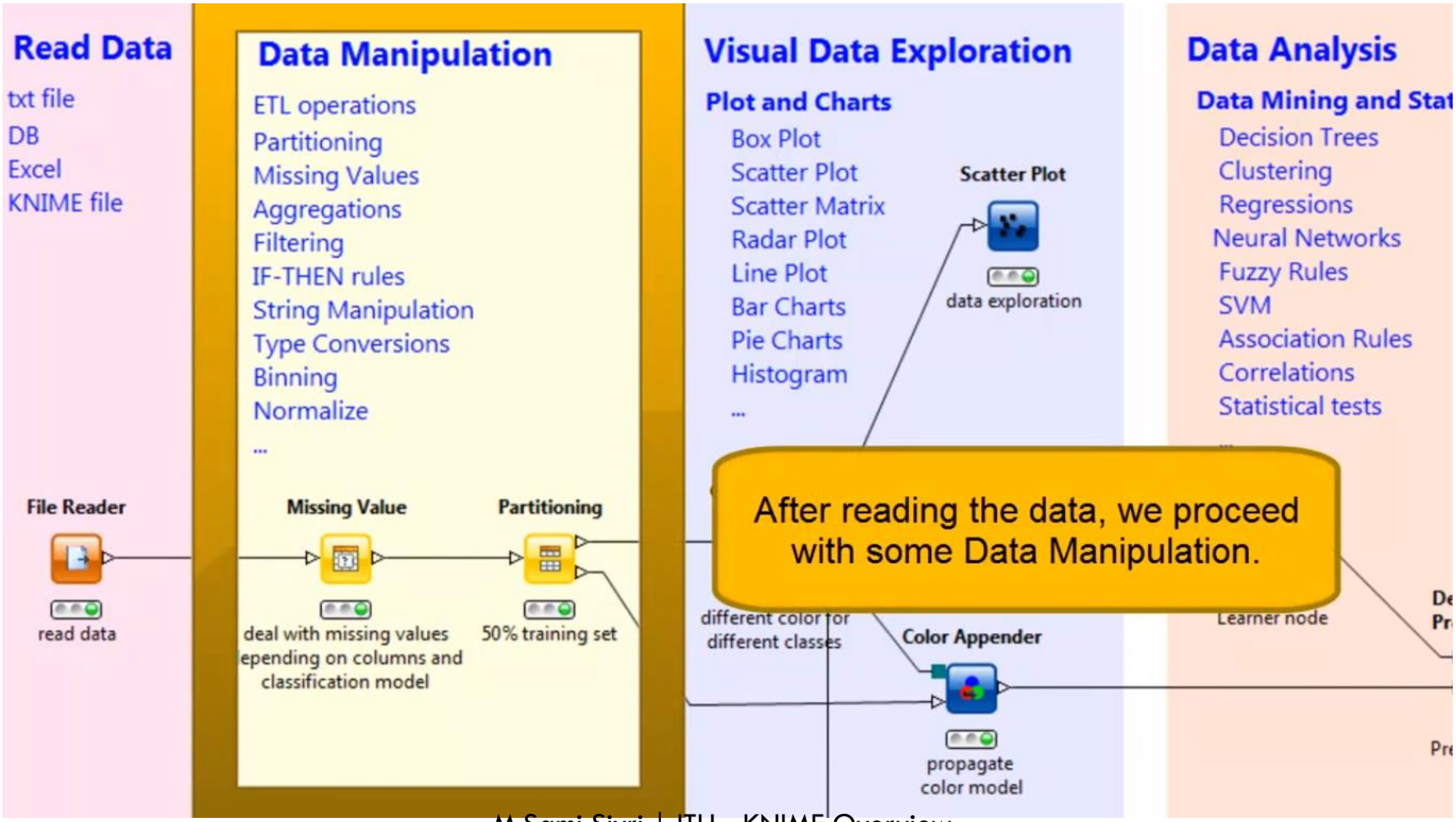
Read Data


Read Data
txt file
DB
Excel
KNIME file
File Reader
read data
Data Manipulation
TL operations
artitioning
Missing Values
ggregations
iltering
F-THEN rules
tring Manipulation
ype Conversions
inning
ormalize
ending on columns and
classification model
Visual Data Exploration
Plot and Charts
Box Plot
Scatter Plot
Scatter Matrix
Radar Plot
Line Plot
Bar Charts
Pie Charts
Histogram
...
Scatter Plot
data exploration
different classes
Color Appender
propagate
color model
Data Ana
Data Minin
Decision T
Clustering
Regressio
Neural Net
Fuzzy Rule
SVM
Associatio
Correlatio
Statistical
...
Decision
Tree Learner
Learner node
We start on the left with a node to read data. In this case we have a File Reader node to read an ASCII file.

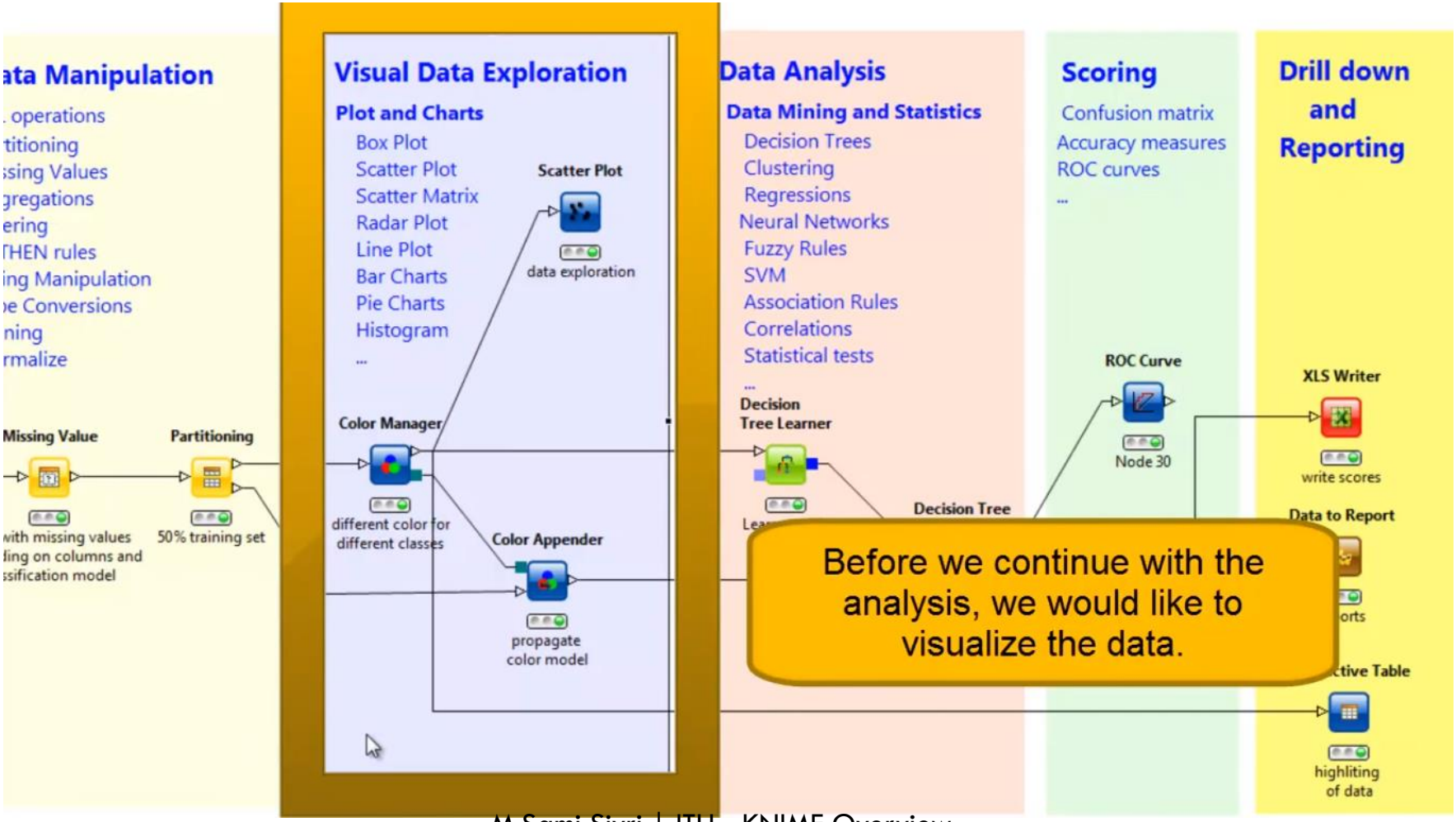
Data Sources



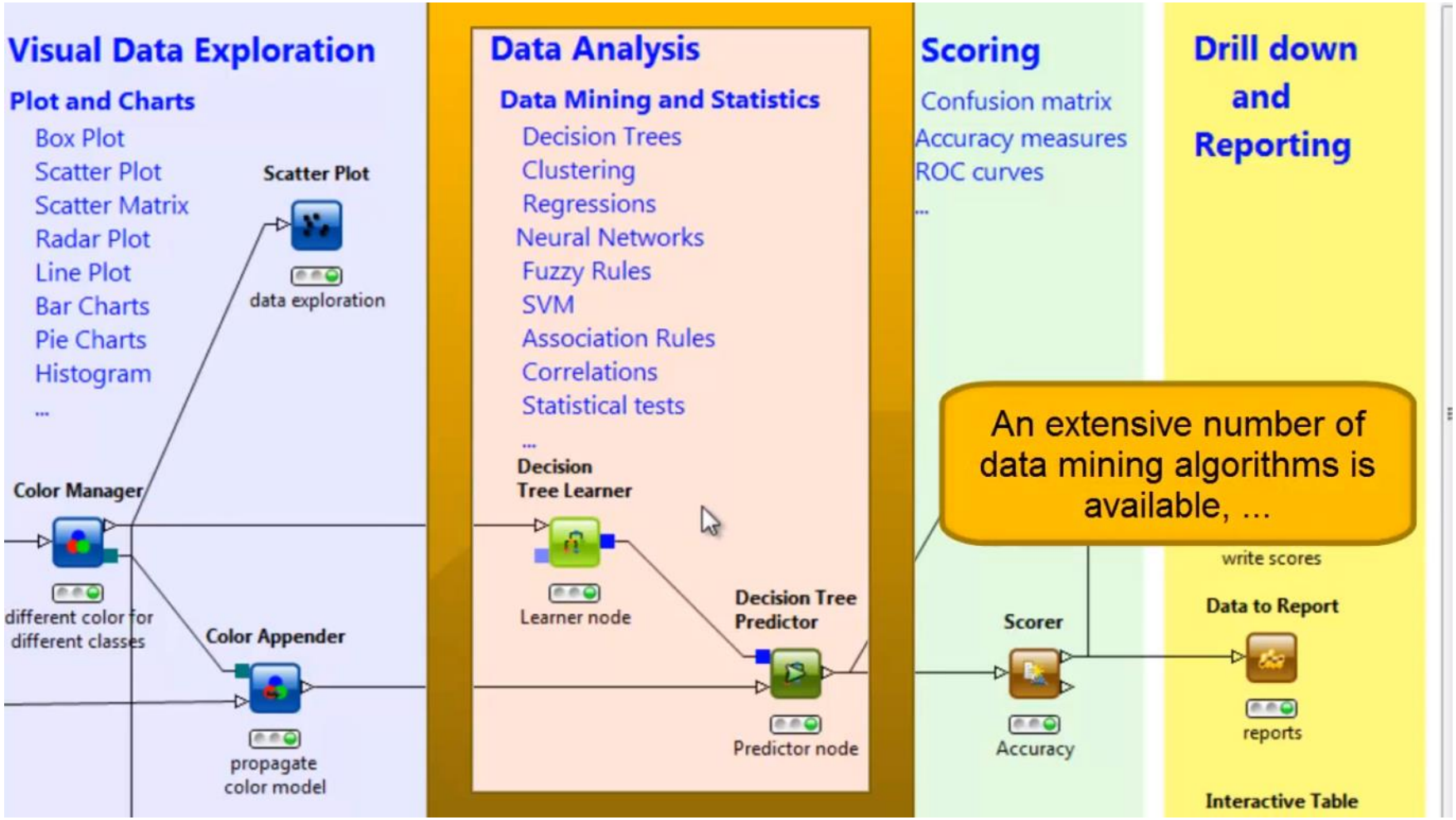
Data Manipulation



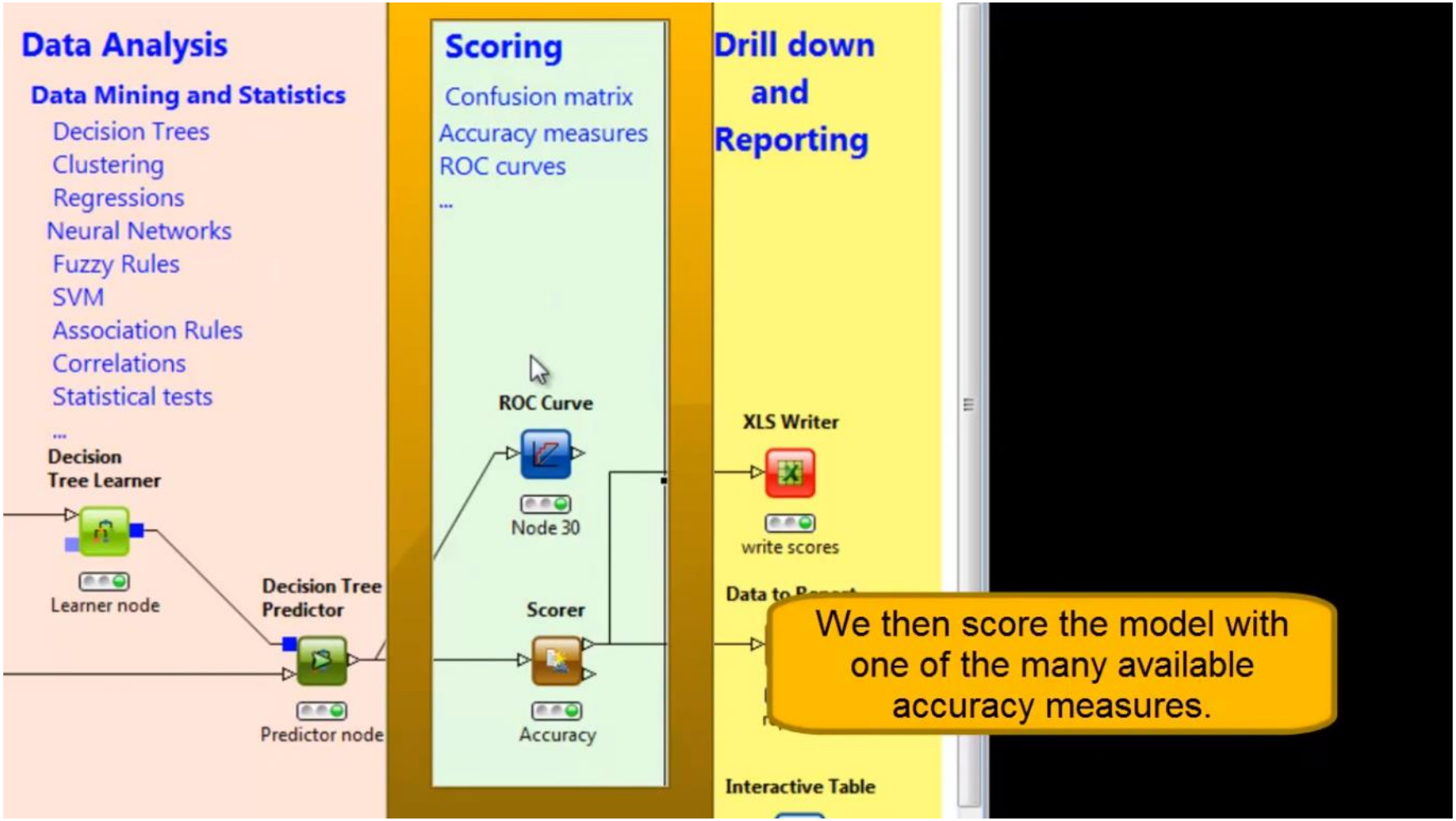
Visualisation



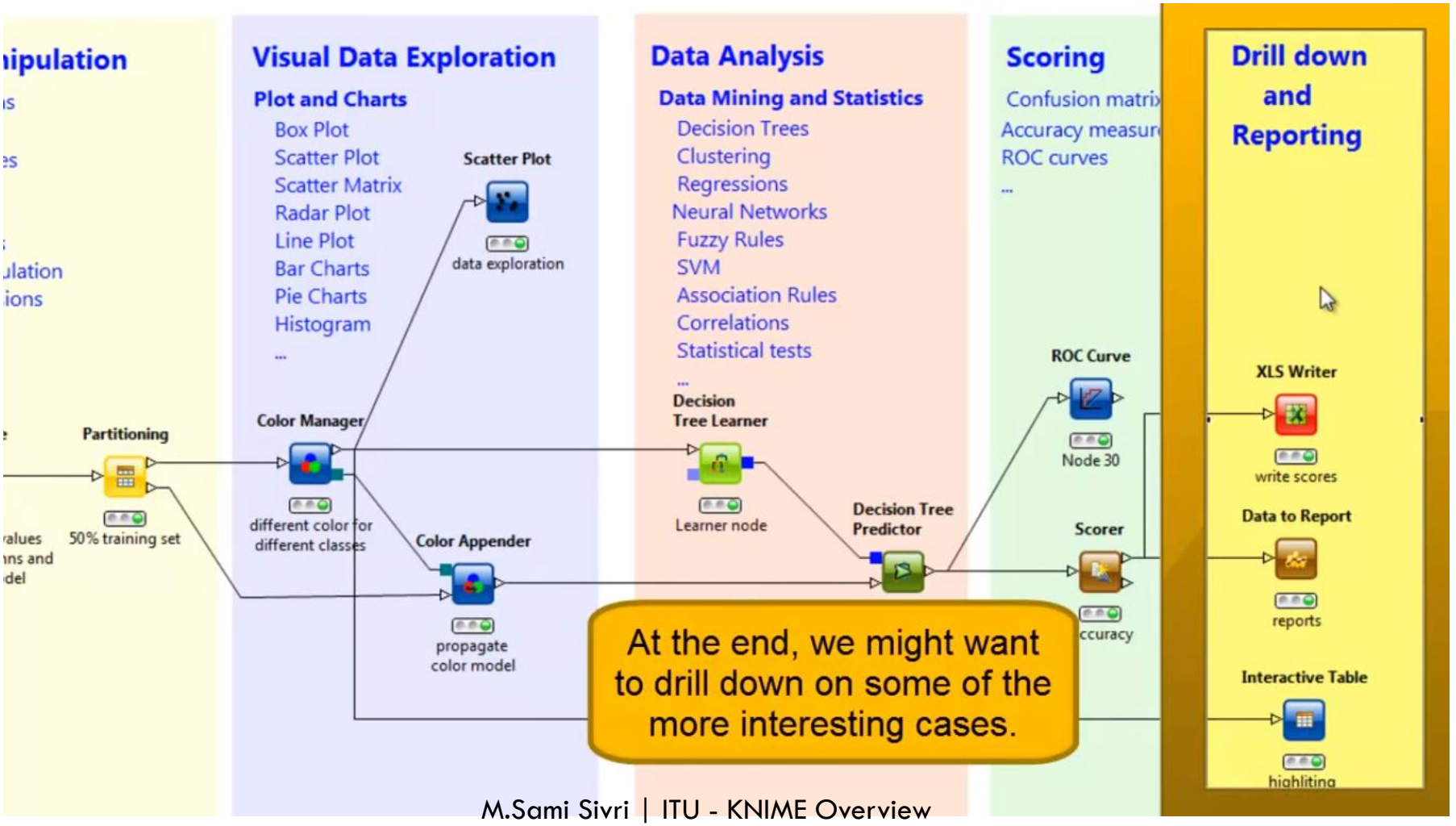
Algorithms


Visual Data Exploration
Plot and Charts
Box Plot
Scatter Plot
Scatter Matrix
Radar Plot
Line Plot
Bar Charts
Pie Charts
Histogram
...
Scatter Plot
data exploration
Color Manager
different color for different classes
Color Appender
propagate color model
Data Analysis
Data Mining and Statistics
Decision Trees
Clustering
Regressions
Neural Networks
Fuzzy Rules
SVM
Association Rules
Correlations
Statistical tests
...
Decision Tree Learner
Learner node
Decision Tree Predictor
Predictor node
Scoring
Confusion matrix
Accuracy measures
ROC curves
...
Scorer
Accuracy
Drill down and Reporting
An extensive number of data mining algorithms is available, ...
write scores
Data to Report
reports
Interactive Table

Accuracy Measurement



Report & Export



Hotkeys

Task	Hotkey	Description
Node Configuration	F6	opens the configuration dialog of a node
	F7	executes selected nodes
Node Execution	Shift + F7	executes all configured nodes
	Shift + F10	executes configured nodes and opens all views
	F9	cancels selected running nodes
	Shift + F9	cancels all running nodes
Move Nodes and Annotations	Ctrl + Shift + Arrow	moves a selected node in the workflow editor
	Ctrl + Shift + PgUp/PgDown	Moves the selected up or down in z order
	F8	resets selected nodes
Workflow Operations	Ctrl + S	Saves the workflow
	Ctrl + Shift + S	Saves all open workflows
	Ctrl + Shift + W	Closes all open workflows
Meta-node	Shift + F12	Opens meta-node wizard