



# **Seminario 2: KNIME y Weka**

Jorge Casillas  
Inteligencia de Negocio  
Grado en Ingeniería Informática  
Curso 2019-2020

# KNIME

- KNIME –se pronuncia [naim]– es una herramienta de minería de datos basada en *software* libre (<http://www.knime.com/>)
- Presenta algunas características interesantes para IN pues integra otras plataformas como BIRT (para creación de informes), WEKA (para minería de datos), R y Python (para análisis estadísticos y visualización) además de otras extensiones para lectura de datos, ETL, generación de informes, visualización e incluso *big data*
- Aunque no se trate de una plataforma originalmente diseñada para IN\*, resulta muy adecuada como herramienta docente de iniciación por el tamaño comedido del *software* y su facilidad de instalación y configuración

\* No obstante, fue nombrado por Gartner como «Cool Vendor 2010» en *Analytics, Business Intelligence, and Performance Management*

# Instalación

- Instalar la versión de 64 bits y con acceso al máximo de RAM
- Extensiones útiles para la asignatura (v. 4.0.1 – 15/08/2019):  
Repositorio: “Work with: --All Available Sites--”

- KNIME & Extensions

- KNIME Data Generation
- KNIME Distance Matrix
- KNIME File Handling Notes
- KNIME HTML/PDF Writer
- KNIME Interactive R Statistics Integration
- KNIME Itemset Mining
- KNIME JavaScript Views
- KNIME JFreeChart
- KNIME Optimization extension
- KNIME Python Integration
- KNIME Report Designer
- KNIME Textprocessing
- KNIME Weka Data Mining Integration (3.7)

- KNIME Community Contributions – Other

- KNIME Python Scripting extension

- KNIME Labs Extensions

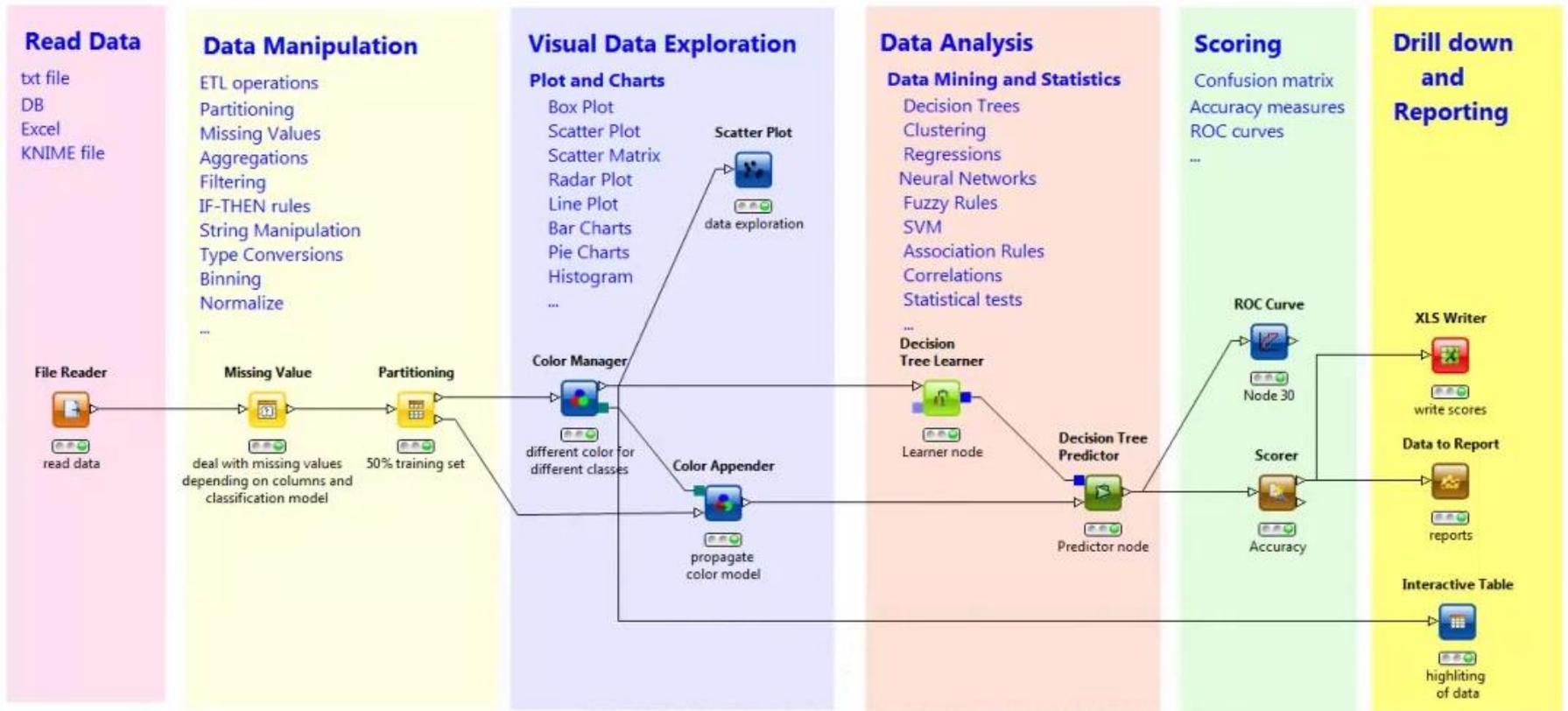
- KNIME JavaScript Views (Labs)
- KNIME Machine Learning Interpretability Extension
- KNIME Plotly
- KNIME Rule Viewer
- KNIME Statistics Nodes (Labs)
- KNIME XGBoost Integration

# Workbench

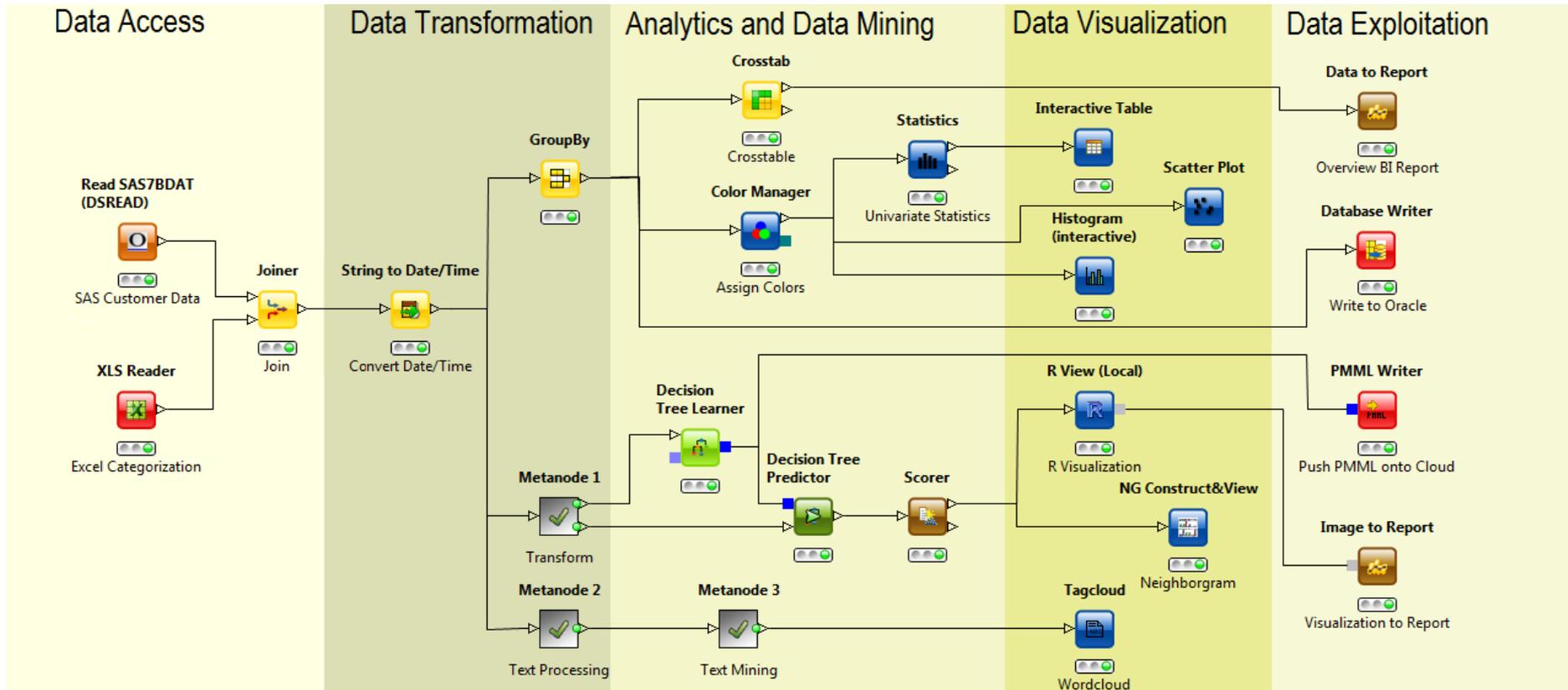
The screenshot shows the KNIME Workbench interface with several callout boxes explaining key components:

- Workflow Projects:** Each workflow refers to a workflow project. All projects are displayed here. Import and export of workflows is supported. Status (closed, idle, executing and executed) is indicated by an icon.
- Favorite Nodes:** Manage your favorite, most frequently and last used nodes. Nodes are added to favorites by dragging them from Node Repository into favorite nodes category.
- Node Repository:** Find here all KNIME nodes ordered by categories. Help for selected nodes are displayed in the Node Description. Drag them onto the workbench in order to use them.
- Workflow Editor:** Here are the workflows assembled by dragging nodes onto it, connecting, configuring and executing them.
- Node Description:** Provides help about the selected node, its dialog options, views, expected input data, and resulting output.
- Outline:** Overview over the workflow and navigation help for large workflows.
- Console:** Status information, warning and error messages are logged here. This information is also written to a log file. Change level of detail under Preferences/KNIME/KNIME GUI.

# Ejemplo 1 de Workflow



# Ejemplo 2 de Workflow



# Introducción a KNIME

<http://youtu.be/ft7Ksgss3Tc> (6'09'')

B  
U  
S  
I  
N  
E  
S  
S  
I  
N  
T  
E  
L  
L  
I  
G  
E  
N  
C  
E

The screenshot displays the KNIME software interface. On the left, a node repository lists various categories such as IO, Database, Data Manipulation, Data Views, Statistics, Mining, Distance Matrix, Meta, Flow Control, Misc, KNIME Labs, Time Series, Quick Form, R, Reporting, Weka, and XML. The main workspace shows a workflow with nodes for 'File Reader', 'Missing Value', 'Decision Tree Predictor', 'ROC Curve', 'Scorer', 'XLS Writer', 'Data to Report', and 'Interactive Table'. A central window titled 'Scatter Plot - 0:27 - Scatter Plot(data exploration)' is open, showing a scatter plot of 'age' (X-axis, 17 to 90) versus 'hours-per-week' (Y-axis, 1 to 99). The plot contains numerous data points colored in red and cyan. The bottom of the interface features a console window with the text: '\*\*\* Copyright. 2003 - 2012. Uni Konstanz and KNIME GmbH, Germany \*\*\*' and a page number '133M of 213M'.

# Material de Apoyo

- <http://www.youtube.com/user/KNIMETV>
- [http://tech.knime.org/files/KNIME\\_quickstart.pdf](http://tech.knime.org/files/KNIME_quickstart.pdf)
- <http://www.knime.org/knimepress/beginners-luck>
- <http://www.slideshare.net/itapia/manual-bsico-knime>
- [http://didawiki.cli.di.unipi.it/lib/exe/fetch.php/dm/knime\\_slides.pdf](http://didawiki.cli.di.unipi.it/lib/exe/fetch.php/dm/knime_slides.pdf)

# Ejercicios

1. Instalación
2. Probar distintos ejemplos de uso y bajarlos del servidor
3. Crear un flujo para obtener un árbol de decisión para clasificación con partición de datos
4. Exploración visual de datos (JFreeChart) y practicar con «hilting» (destacar grupos de datos con determinadas características)
5. Visualizar el conocimiento obtenido
6. Probar con otros algoritmos de Weka