

# **UNIVERSIDAD NACIONAL DE LA MATANZA**

## **INTELIGENCIA DE NEGOCIOS**

---

**Tecnologías Inteligentes para Explotación de Información**

### **Introducción a KNIME**

**Docentes: ING. LORENA R. MATTEO**



# INTRODUCCIÓN A KNIME

**KNIME = KoNstanz Information MinEr**

<http://www.knime.org/>

- es un entorno totalmente gratuito para el desarrollo y ejecución de técnicas de minería de datos.
- fue desarrollado en 2004 el Dpto de Bioinformática y Minería de Datos de la Universidad de Konstanz (Constanza), Alemania, bajo la supervisión del profesor Michael Berthold.
- en la actualidad, la empresa KNIME.com GmbH, radicada en Zúrich, Suiza, continúa su desarrollo, además de prestar servicios de formación y consultoría.

KNIME:

- desarrollado sobre la plataforma Eclipse y programado, esencialmente, en Java.
- su uso se basa en el diseño de un flujo de ejecución que plasme las distintas etapas de un proyecto de minería de datos.
- el proceso de análisis de datos, utilizado por KNIME, consiste en un pipeline de nodos, conectados a través de puentes que transportan datos o modelos.
- cada nodo procesa la llegada de datos y/o modelo(s) y produce resultados como salida.
- es posible explorar los datos en forma visual, a través de vistas.



# INTRODUCCIÓN A KNIME

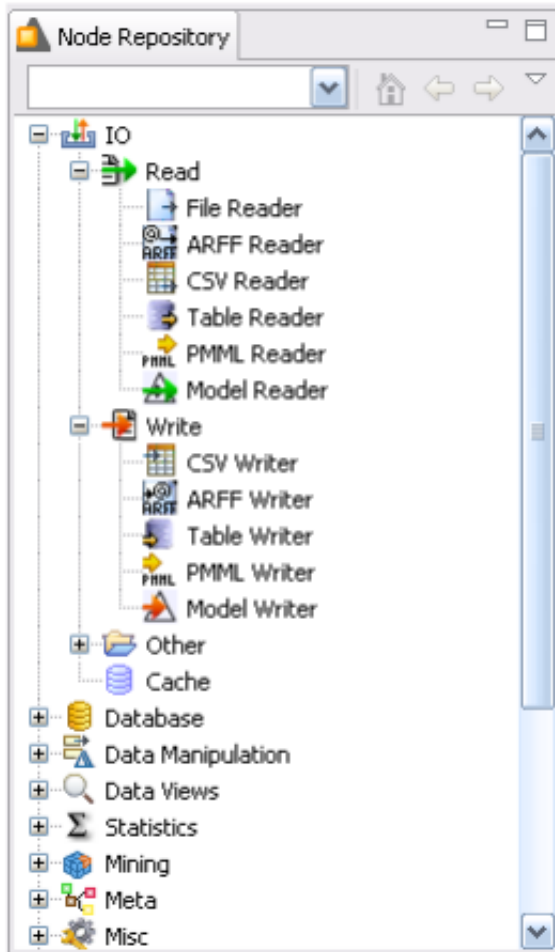
- Es una herramienta para análisis, manipulación, visualización y generación de informes de datos.
- Basado en el paradigma de programación gráfica.
- Proporciona una amplia gama de extensiones:

- Text Mining
- Network Mining
- Cheminformatics
- Many integrations, such as Java, R, Python, Weka, Keras, Plotly, H2O, etc.

The screenshot shows the KNIME website header with navigation links: Software, Pricing, Solutions, Community, Customers, Resources, and About us. There are also buttons for Contact us, Download, and Sign in. The main content area features a large text block stating: "KNIME offers a complete platform for end to end data science, from creating analytic models, to deploying them and sharing insights within the organization, through data apps and services." Below this text are two diagrams. The left diagram, labeled "KNIME Analytics Platform", shows a workflow starting with a data source (represented by a funnel), followed by a processing node (a yellow square with dots), and then a visualization node (a pie chart). The right diagram, labeled "KNIME Hub", shows a network of people icons connected by dotted lines, with a central server icon and a cloud icon, representing a collaborative environment. On the far right of the screenshot, there are three yellow circular icons: a magnifying glass, a heart, and a plus sign.



# INTRODUCCIÓN A KNIME



KNIME proporciona distintos nodos agrupados en fichas, como por ejemplo:

- a) Entrada de datos [*IO > Read*].
- b) Salida de datos [*IO > Write*].
- c) Preprocesamiento [*Data Manipulation*], para filtrar, discretizar, normalizar, filtrar, seleccionar variables...
- d) Minería de datos [*Analytics/Mining*], para construir modelos (reglas de asociación, clustering, clasificación, MDS, PCA...).
- e) Salida de resultados [*Data Views*] para mostrar resultados en pantalla (ya sea de forma textual o gráfica).
- f) **WEKA**: Bayes, Trees, Rules.

**USO: Drag & Drop sobre el Editor Windows**



# INTRODUCCIÓN A KNIME – ENTORNO DE TRABAJO DE LA PLATAFORMA DE ANÁLISIS

The screenshot shows the KNIME Analytics Platform interface. The main workspace is the **Workflow Editor**, which contains a workflow titled "My first Workflow". The workflow consists of 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.

Other components labeled in the interface include:

- KNIME Explorer**: Located on the left, it shows the project structure, including "My-KNIME-Hub", "EXAMPLES", and "LOCAL (Local Workspace)".
- Workflow Coach**: Located on the left, it provides recommendations for nodes based on the current workflow.
- Node Repository**: Located on the left, it lists various nodes available in the platform, categorized by function (e.g., IO, Manipulation, Views, Analytics, DB, etc.).
- Node Description**: Located on the right, it provides detailed information about the selected node, including its purpose, configuration options, and dialog options.
- KNIME Hub Search**: Located on the right, it allows users to search for workflows, nodes, and more.
- Console & Node Monitor**: Located at the bottom, it displays the execution status of the selected node and provides a table of the output data.
- Outline**: Located at the bottom left, it shows a small overview of the workflow structure.

The **Console & Node Monitor** section shows the following 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

# INTRODUCCIÓN A KNIME

## INTERFASE CLÁSICA DE USUARIO



KNIME Analytics Platform

File Edit View Node Help

75%

KNIME Explorer

- My-KNIME-Hub (api.hub.knime.com)
- EXAMPLES (knime@api.hub.knime.com)
- LOCAL (Local Workspace)
  - Example Workflows
  - KNIME\_project
  - KNIME\_project2
  - KNIME\_project3

Workflow Coach

Node recommendations only available with

Node Repository

- Connectors
- File Folder Utility
- Other
- File Handling (legacy)
- Cache
- Manipulation
- Views
- Analytics

Credit Scoring

Credit scoring is a technique used to determine whether or not to extend credit (and if so, how much) to a borrower. This workflow illustrates how to create and choose a credit scoring model based on both historical data and on the application of different machine learning algorithms.

Task Create a credit scoring model based on historical data. Select the best machine learning algorithm to be applied. Use cross-validation to evaluate model performance.

A use case is described at URL: <https://www.knime.org/knime-applications/credit-scoring>

Data Reading	Pre-processing	Model Training and Evaluation	Model Selection	Save the Model
The data are German Credit data, including credit status, demographic data, and customer history. The file is located in TheData/Credit	Learners such as neural network or SVM can only handle numeric attributes. Nominal columns are converted into numerical columns.	1) The following algorithms are trained and evaluated with cross-validation: <ul style="list-style-type: none"><li>- Neural Network</li><li>- SVM</li><li>- Decision Tree</li></ul> 2) Double-click on the metanode to see the subworkflow	All results, i.e. accuracies and respective models, are combined in one single table. Rows are then sorted by descending accuracy and only first row (best performing model) is kept.	<ul style="list-style-type: none"><li>- Convert the model cell back to PMML</li><li>- Save the model.</li></ul> KNIME Analytics Platform writes out the model in the official PMML format, so that other applications can use the model.

Workflow Diagram:

```
graph LR; CSVReader[CSV Reader] --> CategoryToNumber[Category To Number]; CategoryToNumber --> TrainCrossValidateNN[Train and Cross Validate a Neural Network]; CategoryToNumber --> TrainCrossValidateSVM[Train and Cross Validate a SVM]; CategoryToNumber --> TrainCrossValidateDT[Train and Cross Validate a Decision Tree]; TrainCrossValidateNN --> Concatenate[Concatenate]; TrainCrossValidateSVM --> Concatenate; TrainCrossValidateDT --> Concatenate; Concatenate --> Sorter[Sorter]; Sorter --> RowFilter[Row Filter]; RowFilter --> CellToPMML[Cell To PMML]; CellToPMML --> PMMLWriter[PMML Writer];
```

Try this: Bar Chart

KNIME Console

```
*****
*** Welcome to KNIME Analytics Platform v4.7.3.v202305100921 ***
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
```

# INTRODUCCIÓN A KNIME

## NUEVA INTERFAZ DE USUARIO



KNIME Modern UI

KNIME Analytics Platform

Home

KNIME Analytics Platform 5

Help

Preferences

Menu

### Get started with KNIME Analytics Platform 5



#### 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 Community Hub ([hub.knime.com](https://hub.knime.com))

Sign in

3/6/2025

IN2025

# INTRODUCCIÓN A KNIME

## NUEVA INTERFAZ DE USUARIO



KNIME Modern UI

### Application tabs

Entry page tab and all opened workflows tabs.

### Side panel navigation

#### Description

Description of the displayed workflow or component.

#### Node repository

All available nodes in KNIME Analytics Platform to build your workflows.

#### Space explorer

Navigate local or KNIME Hub spaces and access your workflows, components and files.

### Workflow Editor

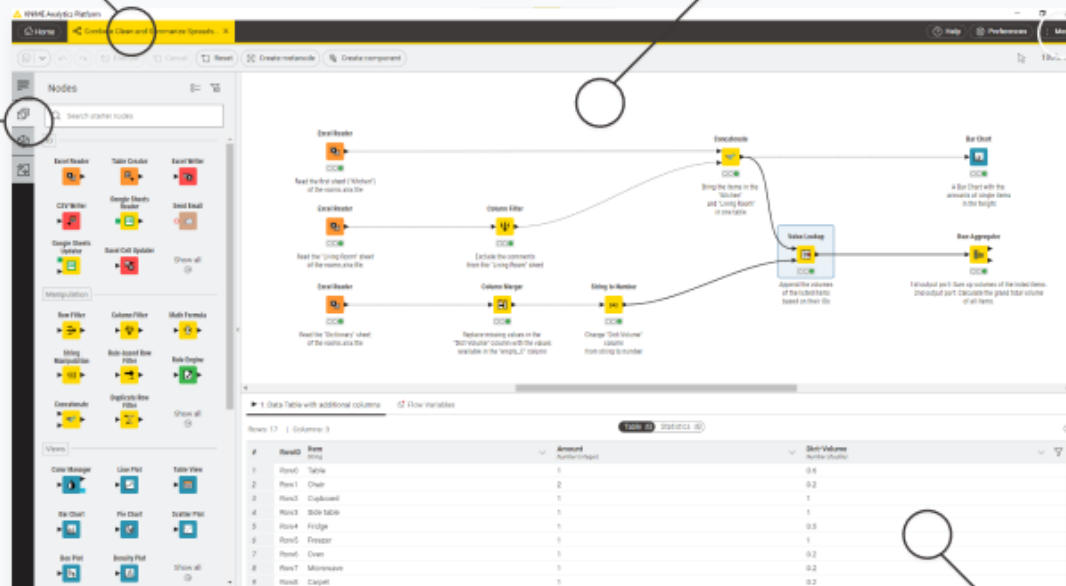
Canvas for editing the currently active workflow.

### Help, Preferences, Menu

Access to more material, install additional extensions and change settings for the node repository.

### Node Monitor

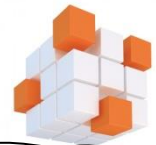
Shows the output of the current selected node and also the flow variable values.





# INTRODUCCIÓN A KNIM

## NUEVA INTERFAZ DE USUARIO



KNIME Modern UI

KNIME Analytics Platform

Home 01\_Performing\_a\_k-Means\_Clustering Help Preferences Menu

Execute Cancel Reset Create metanode Create component

Nodes

Search starter nodes

IO

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

String Cleaner Table Cropper

Workflow Diagram:

Table Reader (deprecated) Partitioning k-Means Cluster Assigner

Search for three clusters

Visualization

Color Manager Shape Manager Scatter Plot (legacy)

1: Assigned Data

Rows: 20 | Columns: 9

#	RowID	DATO	0	1	2	3	4	5	6
		Number (inte...	String	String	String	String	String	String	String
1	Row1	1	0,090625	0,629687	0,015625	0,7375	-0,185937	0,669531	0,0625
2	Row13	13	-0,064062	0,517187	0,690625	0,678906	0,367188	0,111719	0,278125
3	Row17	17	0,7	0,806641	0,186719	0,242969	0,676562	0,1375	0,592188
4	Row18	18	0,30625	-0,113281	0,170312	-0,026563	0,68125	0,203125	0,414062
5	Row22	22	0,620313	0,480859	0,355469	-0,080469	0,341406	0,109375	0,221875

# INTRODUCCIÓN A KNIME

## NUEVA INTERFAZ DE USUARIO



KNIME Modern UI

KNIME Analytics Platform

Home KNIME\_project5KMeans X Help Preferences Menu

Execute Cancel Reset Create metanode Create component

### CSV Reader

Reads CSV files. To auto-guess the structure of the file click the Autodetect format button. If you encounter problems with incorrect guessed data types disable the Limit data rows scanned option in the Advanced Settings tab. If the input file structure changes between different invocations, enable the Support changing file schemas option in the Advanced Settings tab. For further details see the KNIME File Handling Guide [File Handling Guide](#).

**Note:** If you find that this node can't read your file, try the **File Reader** node. It offers more options for reading complex files.

*This node can access a variety of different file systems. More information about file handling in KNIME can be found in the official [File Handling Guide](#).*

**Parallel reading:** Individual files can be read in parallel if

- They are located on the machine that is running this node.
- They don't contain any quotes that contain row delimiters.
- They are not gzip compressed.
- No lines or rows are limited or skipped.
- The file index is not prepended to the RowID.

Statistics View

Scatter Plot

k-Means

Search compatible nodes

Table Splitter

Table Cropper

Row to Column Names

Joiner

Column Filter

Concatenate

GroupBy

Row Filter

Missing Value

Partitioning

Number to String

String Manipulation

1: File Table

Flow Variables

Rows: 120 | Columns: 8

#	RowID	DATO	0
		Number (inte...	Number (dou...
1	Row0	0	0.916
2	Row1	1	0.091
3	Row2	2	0.456
4	Row3	3	0.355
5	Row4	4	-0.172
6	Row5	5	0.287

Silhouette Coefficient

Scorer

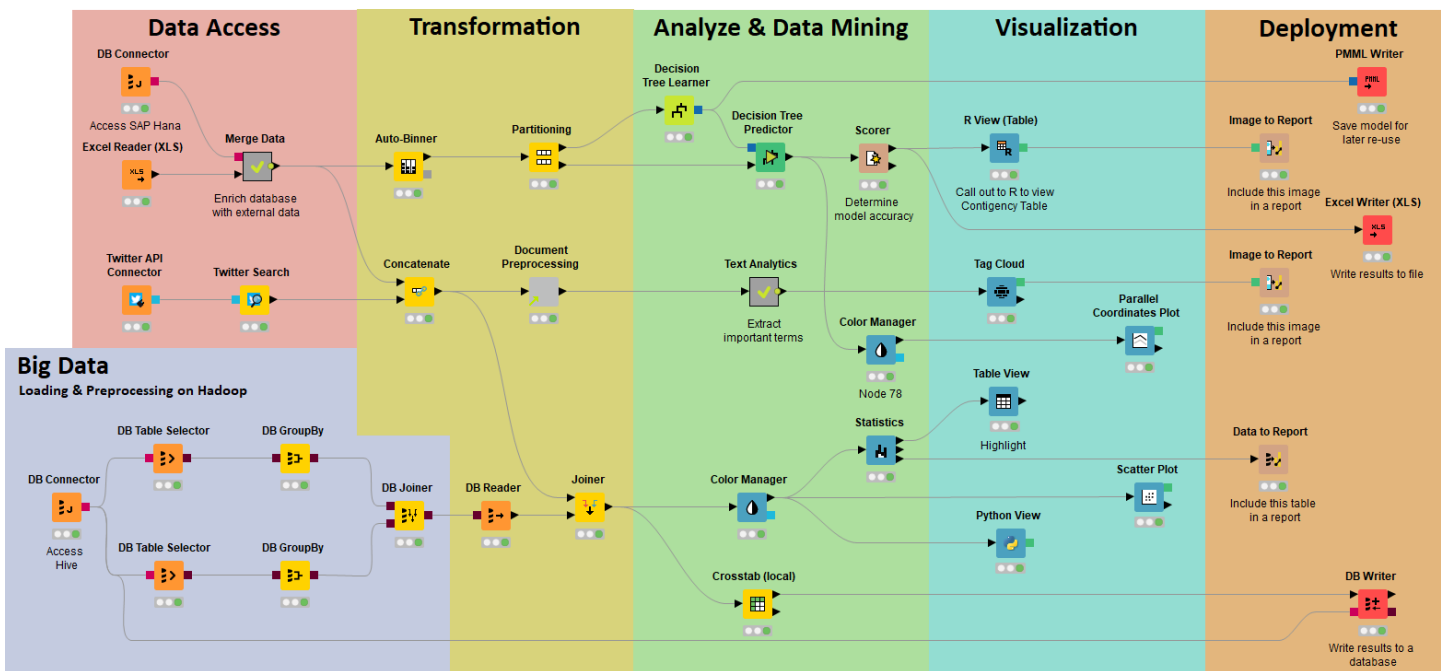
	5	6
	Number (dou...	Number (dou...
dou...	-0.144	0.616
	0.67	0.062
	0.283	0.424
	0.993	-0.07
	0.459	0.756
	0.052	0.501

3/6/2025 IN2025



# INTRODUCCIÓN A KNIME

Provee más de 4000 rutinas de análisis de datos, tanto nativas como a través de Python, R o Weka, algunos de los nodos se muestran en la siguiente imagen:



## 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



# INTRODUCCIÓN A KNIME

- Instalar la versión de 64 bits configurando # acceso a la RAM.
- Extensiones útiles:

## ❑ **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

## ❑ **KNIME Community Contributions – Other**

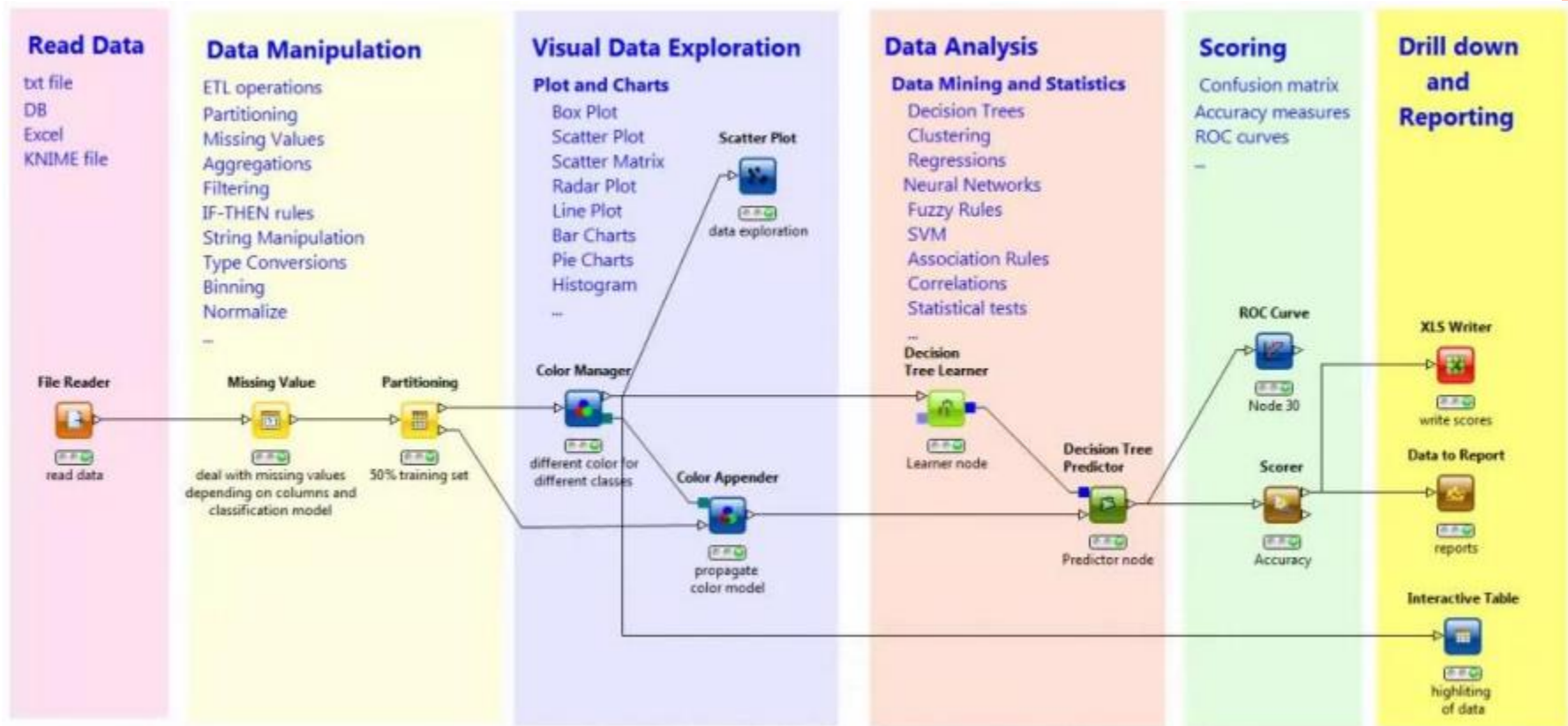
- KNIME Python Scripting extension

## ❑ **KNIME Labs Extensions**

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



# INTRODUCCIÓN A KNIME – EJEMPLO WORKFLOW

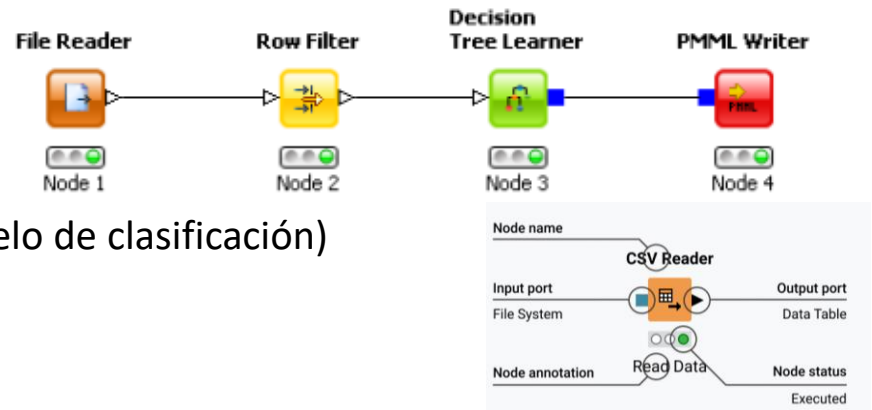


- Un flujo de trabajo es una secuencia de nodos, cada uno de los cuales se puede configurar para realizar una tarea específica.
- Los datos fluyen a través de los nodos de izquierda a derecha.

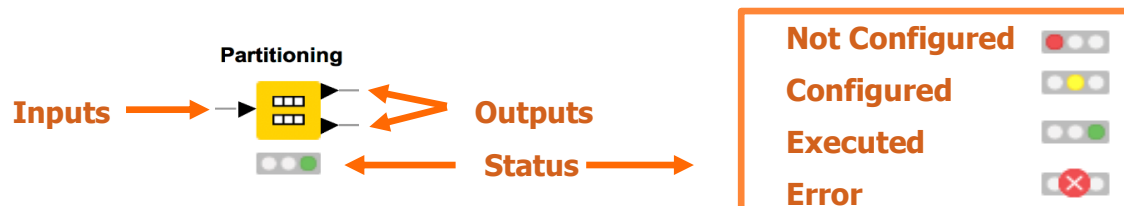


# INTRODUCCIÓN A KNIME – CREAR WORKFLOWS

- Para crear un flujo de ejecución, las salidas de unos nodos se utilizan como entradas de otros. Por ejemplo, un flujo básico podría ser de la forma:



- Nodo de lectura de datos
  - Nodo de preprocesamiento
  - Nodo de salida de resultados.
  - Nodo de modelado (por ejemplo, modelo de clasificación)
- Flujo de datos
  - Se construye arrastrando y soltando los nodos desde el repositorio de nodos al Editor Windows y conectándolos entre sí.
  - Los datos se transportan entre los nodos a través de los puertos de entrada y salida.
  - Después de colocar los nodos en el Editor Windows, es necesario conectar la entrada de cada nodo con la salida del nodo predecesor.





# INTRODUCCIÓN A KNIME – RESUMEN

## Funcionalidades:

- Software libre bajo licencia GNU.
- Combinación de datos y herramientas.
- Analíticas poderosas.
- Más de 4000 módulos y creciendo.
- Conectores para todos los formatos de archivos y bases de datos más utilizadas.
- Soporte para una gran variedad de tipos de datos.
- Combinación y transformación de datos nativa y en la base de datos.
- Funciones matemáticas y estadísticas.
- Algoritmos de predicción avanzados y de Machine Learning.
- Control de flujo.
- Herramienta de unión para R, Python, SQL, Java, Weka, etc.
- Vistas de datos y reportes interactivos.
- Licenciamiento: GNU GPL 3

## Para más información consultar los siguientes enlaces:

- Página oficial: <https://www.knime.com/>
- Documentación: <https://www.knime.com/documentation>
- Comunidad: <https://www.knime.com/knime-community>
- Tutoriales: <https://www.knime.com/resources>
- Videos: <https://www.youtube.com/user/KNIMETV>

# CUADRANTE MÁGICO DE GARTNER PARA DATA SCIENCE Y ML PLATFORMS (ABR 2025)



Gartner.

Fuente: [The 2025 Gartner Magic Quadrant for Data Science & Machine Learning Platforms](#)



# CUADRANTE MÁGICO DE GARTNER PARA DATA SCIENCE Y ML PLATFORMS (ENE 2021)



Fuente: <https://www.gartner.com/reviews/market/data-science-and-machine-learning-platforms>

# TOP 10 HERRAMIENTAS LÍDERES ML & AI SEGÚN LLMs



Herramienta	Descripción	Plataformas Soportadas	Características Clave
<b>TensorFlow</b>	Ideal para redes neuronales y aprendizaje profundo, con soporte para múltiples plataformas	Servidores, Nube, Web, Android, iOS	APIs de alto nivel, ejecución distribuida, integración con Keras.
<b>PyTorch</b>	Popular en investigación por su flexibilidad, especialmente en aprendizaje profundo.	Linux, Windows, macOS	Ejecución inmediata, soporte GPU, comunidad activa, ideal para investigación.
<b>Scikit-learn</b>	Biblioteca de Python para aprendizaje automático tradicional, basada en SciPy y NumPy.	Windows, Linux, macOS	Algoritmos de clasificación, regresión, clustering, preprocesamiento de datos.
<b>Keras</b>	API de alto nivel para redes neuronales, ahora parte de TensorFlow, fácil de prototipar.	Nube (Google, Azure, AWS), Desktop, Móvil	Interfaz amigable, soporte multi-GPU, modularidad, ideal para principiantes.
<b>Microsoft Azure ML</b>	Plataforma en la nube para desarrollar, entrenar y desplegar modelos, con AutoML y MLOps.	Nube (Azure)	Interfaz drag-and-drop, soporte para Python, R, integración con servicios Azure.
<b>Amazon SageMaker</b>	Servicio de AWS para todo el ciclo de vida del aprendizaje automático, con AutoML y algoritmos integrados.	Nube (AWS)	Escalabilidad empresarial, integración con AWS, soporte para PyTorch, TensorFlow.
<b>Google Cloud Vertex AI</b>	Plataforma unificada para desarrollo y despliegue de modelos de IA, con AutoML y notebooks.	Nube (Google Cloud)	Integración con TensorFlow, escalabilidad, herramientas automatizadas para colaboración.
<b>H2O.ai</b>	Plataforma de código abierto para aprendizaje automático, conocida por AutoML y procesamiento distribuido.	Nube, Desktop, Integración con R, Python	AutoML, interpretabilidad de modelos, soporte para Jupyter, seguridad empresarial.
<b>XGBoost</b>	Biblioteca para gradient boosting, optimizada para modelos predictivos, con alta precisión.	Cross-platform	Alto rendimiento, manejo de datos diversos, ideal para clasificación y regresión.
<b>Apache Spark MLlib</b>	Biblioteca de aprendizaje automático para grandes volúmenes de datos, parte de Apache Spark.	Distribuido, Nube	Algoritmos escalables, integración con Spark SQL, ideal para big data.

**Fuente:** Grok 3 (2025). DeeperSearch (versión 31/03/2025) [LLM]. <https://grok.com/>



# MATERIAL DE CONSULTA RECOMENDADO

- Página oficial: <https://www.knime.com/>
  - Documentación: <https://www.knime.com/documentation>
  - Comunidad: <https://www.knime.com/knime-community>
  - Tutoriales: <https://www.knime.com/resources>
  - Videos: <https://www.youtube.com/user/KNIMETV>

## Recursos sugeridos:

- [Spaces of knime – KNIME Community Hub](#). [Teaching Materials | KNIME](#) *Previo registro en el sitio.*
  - [knime/Education – Courses – KNIME Community Hub](#)
  - [stervis/Public – E-Learning – KNIME Community Hub](#)
  - [knime/Education – L4-ML Machine Learning Algorithms – KNIME Community Hub](#)
  - [knime/Examples – 04\\_Analytics – KNIME Community Hub](#)
  - [knime/Educators Alliance – Guide to Intelligent Data Science – KNIME Community Hub](#)

# TRABAJO PRÁCTICO DE MINERÍA DE DATOS

## CASO: CRÉDITOS BANCARIOS

ENTREGA **18/06/2025**



- Utilice la herramienta KNIMNE para desarrollar los procesos de explotación de información identificados en el Caso de Estudio; incluyendo tareas de Preprocesamiento, Modelos Predictivos, Descriptivos y Evaluación.
- Entregue un informe que contenga resultados, conclusiones obtenidas, gráficos, una tabla comparativa de métodos aplicados, y las recomendaciones que daría, de acuerdo con lo requerido en el enunciado del trabajo práctico.

