Análisis predictivo con KNIME

SISTEMAS INTELIGENTES PARA LA GESTIÓN DE LA EMPRESA

CURSO 2016-2017

Plataforma **KNIME**

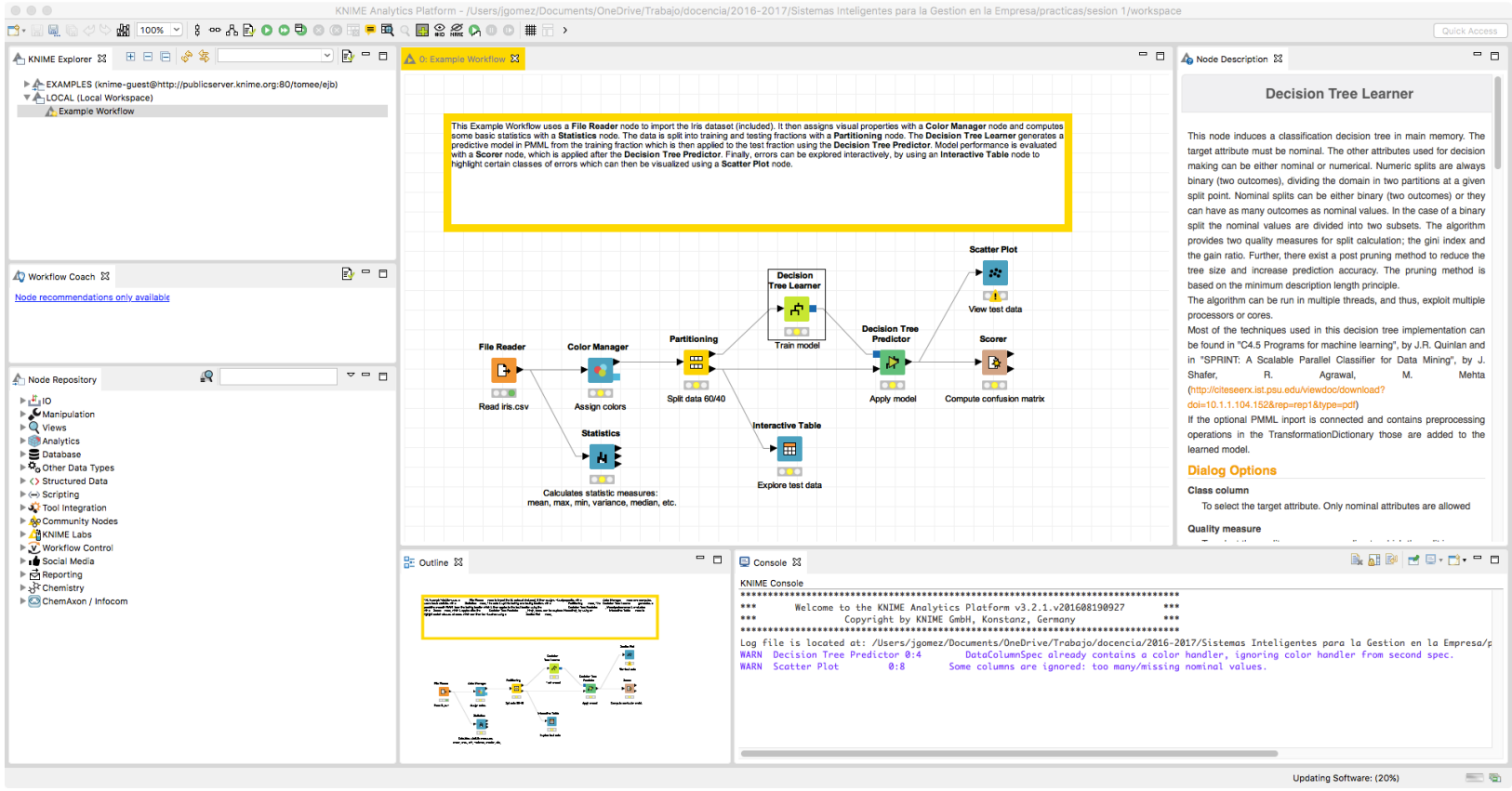
KNIME es una herramienta software para manipulación, análisis y visualización de datos

Características

- Basada en programación visual
- *Open source*
- Ecosistema de módulos (+1500)
  - Análisis descriptivo
  - Predicción
  - Análisis de redes y grafos
  - Minería de textos
- Puede utilizarse en combinación con otras plataformas: Weka, Java, Scala, R, Python, Hadoop/Spark, etc.

https://www.knime.org

El entorno de programación KNIME



Ejemplo: Predicción con *Iris Dataset*

Dataset clásico en aprendizaje automático – R. Fisher (1936) *The use of multiple measurements in taxonomic problems*

Descripción

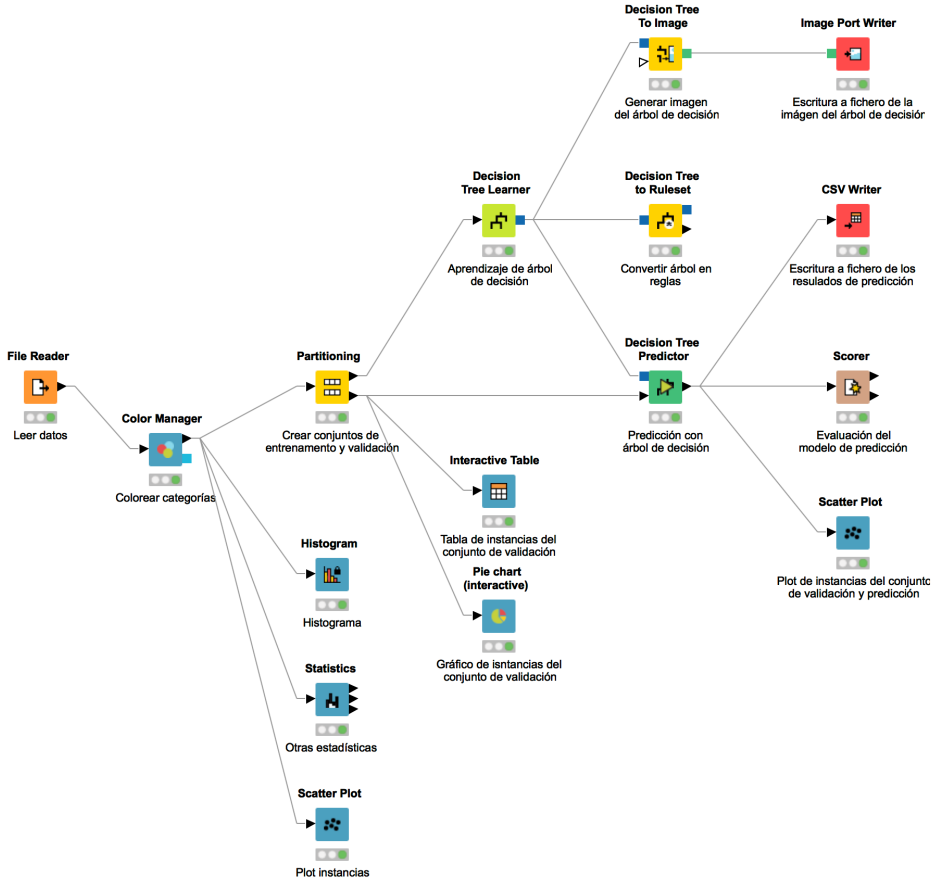
- Tres especies de flores: iris setosa, iris virginica, iris versicolor)
- 50 muestras de cada especie
- 4 características: longitud y anchura de pétalos y sépalos

Objetivo

- Obtener la clase de una flor dados los valores de las 4 carácterísticas

sepal length	sepal width	petal length	petal width	class
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3.0	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5.0	3.6	1.4	0.2	Iris-setosa
7.0	3.2	4.7	1.4	Iris-versicolor
6.4	3.2	4.5	1.5	Iris-versicolor
6.9	3.1	4.9	1.5	Iris-versicolor
5.5	2.3	4.0	1.3	Iris-versicolor
6.5	2.8	4.6	1.5	Iris-versicolor
6.7	2.5	5.8	1.8	Iris-virginica
7.2	3.6	6.1	2.5	Iris-virginica
6.5	3.2	5.1	2.0	Iris-virginica
6.4	2.7	5.3	1.9	Iris-virginica
6.8	3.0	5.5	2.1	Iris-virginica
...				

Ejemplo: Predicción con *Iris Dataset*



Ejemplo: Predicción de crimen

Challenge organizado por el National Institute of Justice con datos reales sobre eventos criminales en Portland [link]

Descripción

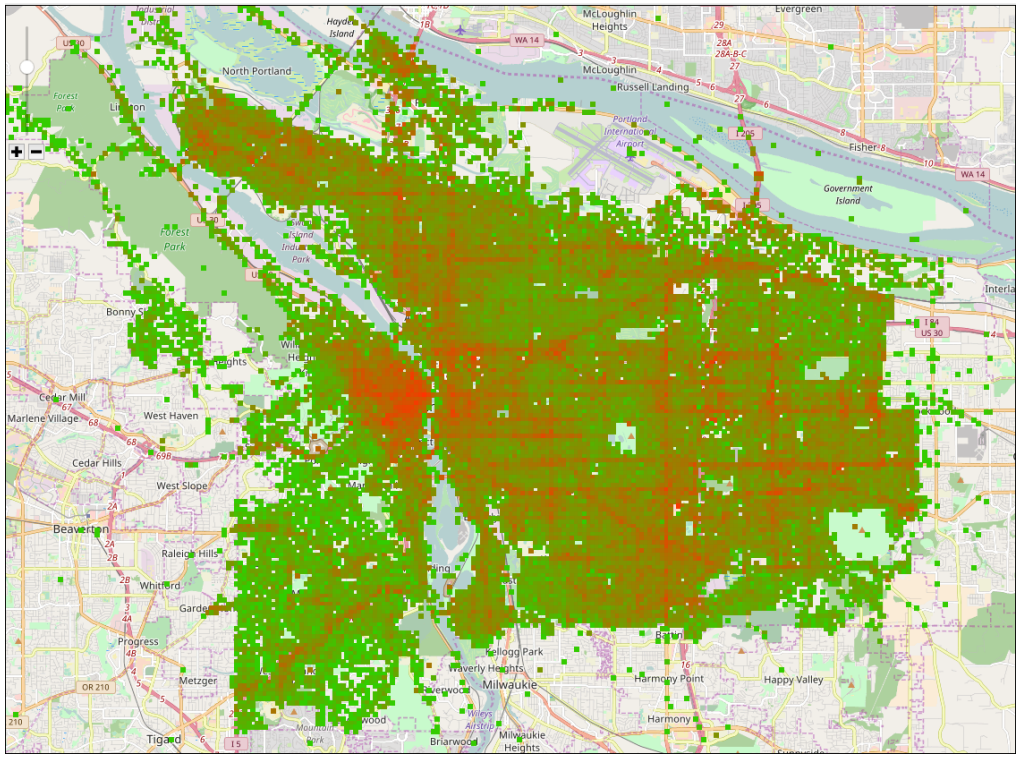
- Datos de eventos criminales de 2013 en adelante, geolocalizados
- 8 características: Categoría, Grupo de Llamada, Identificación Final del Caso, Descripción, Fecha, Coordenadas y Localización Censal
- Dependencias entre características

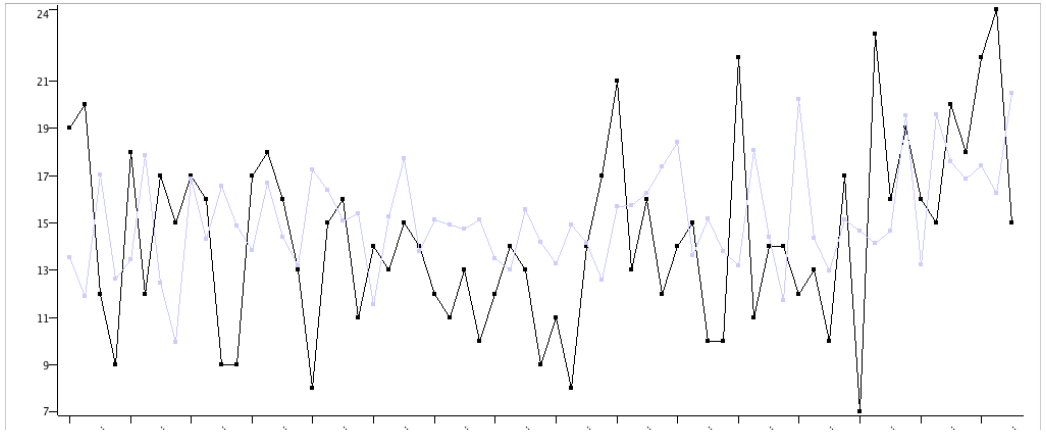
Objetivo

- Predecir zonas de concentración de crímenes

CATEGORY	CALL GROUPS	final_case	CASE DESC	occ_date	x_coordinate	y_coordinate	census_tract
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/12/16	7627009	710228	4102
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/12/16	7627109	710045	4102
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/12/16	7644761	690250	2303
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/12/16	7649826	680465	1101
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/12/16	7664343	696396	7500
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/12/16	7666986	663517	8600
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/12/16	7669275	668437	602

Ejemplo: Predicción de crimen





Documentación adicional

KNIME Learning Hub
https://www.knime.org/learning-hub

KNIME Online Self-Training
https://www.knime.org/knime-online-self-training

O'Reilly "Introduction to Data Analytics with KNIME"
https://www.safaribooksonline.com/library/view/introduction-to-data/9781491967546/

KNIME Beginner's Luck (Rosaria Filipo)
https://www.knime.org/knimepress/beginners-luck

KNIME Node Guide
https://www.knime.org/nodeguide

Coursera "Machine Learning with Big Data"
https://es.coursera.org/learn/big-data-machine-learning