

Sesión 2 - Introducción a los Flujos de Trabajo con KNIME

Elabora un informe con el trabajo realizado para los ejercicios que se proponen a continuación. Empaqueta junto con el informe los flujos de trabajo exportados en un fichero comprimido. Incorpora capturas de pantalla de la configuración de los nodos o exporta imágenes de los flujos de trabajo en formato SVG para ilustrar el trabajo.

Ejercicio 1

1. Prueba a actualizar la aplicación y verifica que están instaladas las siguientes extensiones:
 - KNIME Math Expression Extension (JEP)
 - KNIME External Tool Node
 - KNIME Report Designer
 - KNIME SVG Support
2. Crea un grupo de flujos de trabajo que se llame sesión 2.
3. Busca todos los nodos de tipo “Row Filter” en el repositorio. A continuación, explica para qué sirven analizando las diferencias entre los nodos de tipo “Row Filter”, “Reference Row Filter” y “Nominal Value Row Filter”. Para ello, usa las siguientes tablas de datos que puedes crear con un nodo de tipo “Table Creator”:

Original Table

Position	name	team
1	The Black Rose	4
2	Cynthia	4
3	Tinkerbelle	4
4	Mother	4
5	Augusta	3
6	The Seven Seas	3

Reference Table

Ranking	scores
1	22
3	14
4	10

Ejercicio 2

Crea un flujo de trabajo que realice las siguientes operaciones:

1. Leer el fichero “data1.txt” y verificar que la columna “ranking” es de tipo String. Cambiar el nombre a “marcas”.
2. Eliminar el comentario inicial del fichero. Puedes intentar usar la funcionalidad correspondiente del nodo de lectura, aunque puede no ser de utilida.
3. Eliminar la columna “class”.
4. Escribir los datos finales en formato CSV en un fichero de nombre “data1Nuevo.csv” usando el carácter “;” como separador. ¿Se puede leer o importar bien el fichero desde Excel?
5. Introduce una pequeña descripción para todos los nodos del flujo de trabajo y documenta el flujo de trabajo con etiquetas (“New Workflow annotation”).
6. Verifica que el flujo de trabajo se ejecuta bien (todos los nodos en verde y no hay errores en el log) y expórtalo para adjuntarlo con la memoria.

Ejercicio 3

Crea un flujo de trabajo, reutilizando el anterior si lo ves necesario, que realice las siguientes operaciones:

1. Leer el fichero CSV creado en el ejercicio de la práctica anterior “data1Nuevo.csv” y cambiar de nuevo el nombre de la columna “marcas” a “ranking”.
2. Filtrar todas las filas con el campo comments = “average”.
3. Excluir la columna “ranking”.

4. Prueba a añadir datos al final del fichero en modo “Append” y usa el tabulador como carácter separador.
5. Documenta las descripciones para todos los nodos del flujo de trabajo (“Edit node description”) y el flujo de trabajo con etiquetas.
6. Verifica que el flujo de trabajo se ejecuta bien (todos los nodos en verde) y expórtalo para adjuntarlo en la memoria.

Ejercicio 4

El conjunto de datos “wine.data” contiene los resultados del análisis químico de un conjunto de vinos italianos de tres tipos distintos (clase 1, clase 2 y clase 3). El análisis determinó las cantidades de trece constituyentes encontrados en cada uno de los tipos de vinos:

- 1) Alcohol
- 2) Malic acid
- 3) Ash
- 4) Alcalinity of ash
- 5) Magnesium
- 6) Total phenols
- 7) Flavanoids
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity
- 11) Hue
- 12) OD280/OD315 of diluted wines
- 13) Proline

Descarga el fichero “wine.data” (si no lo tienes ya) de la página web <https://archive.ics.uci.edu/ml/datasets/Wine>. A continuación, crea un flujo de trabajo reutilizando el del ejemplo sobre el árbol de decisión que realice las siguientes operaciones:

1. Entrena un árbol de decisión (con el 80 % de los datos) que reconozca cada una de las clases de vino a las que pertenece cada muestra.
2. Ejecuta el árbol de decisión y mide su rendimiento. Busca cuántos falsos positivos para la clase 2.

3. Identifica si hay algún registro que perteneciendo a la clase 2 se ha clasificado como de clase 3.
4. ¿Qué ocurre si entrenas el árbol sólo con un 50 % de los datos? ¿Y con un 20 %? Prueba a realizar varias ejecuciones (entre 5 y 10) y calcula un promedio de las mismas antes de dar tu respuesta.
5. ¿Qué parámetros puedes modificar en el algoritmo? Averigua para qué sirven y prueba a cambiarlos. ¿En qué se ve afectado el resultado?

Ejercicio 5

Reutiliza y extiende el flujo de trabajo desarrollado para el ejercicio anterior y realiza las tareas que se proponen a continuación:

1. Además del árbol de decisión, entrena un perceptrón multicapa (MLP) con el 80 %, 50 % y 20 % de los datos. Prueba a realizar varias ejecuciones (entre 5 y 10) y calcula un promedio como resultado de las mismas.
2. ¿Qué parámetros puedes modificar en el algoritmo? Averigua para qué sirven y prueba a cambiarlos. ¿En qué se ve afectado el resultado?
3. Aplica el modelo y compara los resultados obtenidos con el modelo de árbol de decisión. ¿Qué modelo se comporta mejor? ¿En qué circunstancias?

