

Práctica 2 SATD

Pedro Allué Tamargo

7 de octubre de 2020

Índice

1. Parte 1	1
2. Parte 2	1
3. Parte 3	1
4. Parte 4	1

1. Parte 1

Para diferenciar entre los diferentes nodos de tipo *Row Filter*, se han creado dos tablas tal y como se indica en el gui3n.

- *Row Filter*: Permite filtrar elementos de la tabla a partir de una cadena o una expresi3n regular.
- *Nominal Row Filter*: Se puede filtrar manualmente a partir de los atributos de la tabla o mediante expresiones regulares.
- *Reference Row Filter*: Permite filtrar el contenido de una tabla a partir de una segunda tabla. Para ello, se selecciona la columna deseada de cada tabla para realizar *Join* entre ambas. Tambi3n permite *Join* con la condici3n inversa (*exclude*).

2. Parte 2

Para verificar que la columna *R3nking* es de tipo *String*, se selecciona esa columna desde la pre visualizaci3n del fichero y se escoge el tipo de dato correspondiente. Desde ah3 tambi3n se puede modificar el nombre de la columna. Para poder eliminar los comentarios, se deben de modificar las columnas correspondientes a3nadiendo `'/'` al principio de las mismas para que sean de tipo comentario *Java*. Desde la previsualizaci3n de *Knime*, se selecciona la casilla *Java-Style comments* para poder ignorar los mismos.

La columna *class* se puede eliminar a3nadiendo un elemento de *Column Filter* con el cual se introduce esta columna en el apartado de *exclude*.

Si se desea escribir un nuevo fichero CSV, se debe a3adir un nuevo componente *CSV writer*. En configuraci3n del m3dulo se selecciona el tipo de separador, la ruta donde debe escribir el archivo y que se incluya la cabecera con la opci3n *write column header*.

3. Parte 3

En este ejercicio, la modificaci3n del nombre de la columna "marcas" se ha realizado a3nadiendo un nodo *column rename* a continuaci3n de *CSV reader* (mediante el cual se lee el archivo "data1Nuevo.csv").

Si se desea filtrar las filas con el campo `comments = .average`, se debe a3adir un nuevo nodo *Nominal Value Row Filter*. Se selecciona la columna *comments* y en la secci3n de *include* se busca el valor *average*. La columna *ranking* se elimina con el nodo *column filter*.

Para la escritura final, se introduce un nodo *CSV writer* y se configura con el modo *append* (se debe indicar que el archivo ya existe), y tabulador como car3cter delimitador.

4. Parte 4

Bas3ndonos en el ejemplo de 3rbol de decisiones de *Knime*, se ha creado un *Workflow* de 3rbol de decisi3n para clasificar las muestras de vinos. Para leer el archivo se utiliza el nodo *File Reader*. Se a3ade un nodo *Partitioning* con el que se selecciona la cantidad de datos de entrenamiento (el 80Seguidamente se a3ade un nodo *Decision Tree Learner*. Para evitar errores con este nodo, se debe modificar la columna *class* desde el nodo lector de archivos e indicar que es de tipo *String*. Posteriormente se incorpora un nodo *Decision Tree Predictor* el cual "predice" los valores.

Por 3ltimo, se a3ade el nodo *Scorer* para poder observar las predicciones realizadas. Los resultados obtenidos son los siguientes:

Row ID	I 1	I 2	I 3
1	12	1	0
2	1	10	1
3	0	1	10

Figura 1: Clasificación del árbol de decisión