Práctica 2 SATD

Pedro Allué Tamargo

7 de octubre de 2020

•				
т		1	•	
	n	a	1	CC
		l l		

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 guión.

- Row Filter: Permite filtrar elementos de la tabla a partir de una cadena o una expresión 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. También permite Join con la condición inversa (exclude).

2. Parte 2

Para verificar que la columna $R\'{a}nking$ es de tipo String, se selecciona esa columna desde la pre visualización del fichero y se escoge el tipo de dato correspondiente. Desde ahí también se puede modificar el nombre de la columna. Para poder eliminar los comentarios, se deben de modificar las columnas correspondientes añadiendo '//' al principio de las mismas para que sean de tipo comentario Java. Desde la previsualización de Knime, se selecciona la casilla Java-Style comments para poder ignorar los mismos.

La columna *class* se puede eliminar añadiendo 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 añadir un nuevo componente *CSV writer*. En configuración del módulo se selecciona el tipo de separador, la ruta donde debe escribir el archivo y que se incluya la cabecera con la opción *write column header*.

3. Parte 3

En este ejercicio, la modificación del nombre de la columna "marcas" se ha realizado añadiendo un nodo column rename a continuación 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 añadir un nuevo nodo *Nominal Value Row Filter*. Se selecciona la columna *comments* y en la seccion 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 carácter delimitador.

4. Parte 4

Basándonos en el ejemplo de árbol de decisiones de *Knime*, se ha creado un *Workflow* de árbol de decisión para clasificar las muestras de vinos. Para leer el archivo se utiliza el nodo *File Reader*. Se añade un nodo *Partitioning* con el que se selecciona la cantidad de datos de entrenamiento (el 80Seguidamente se añade 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 último, se añade el nodo *Scorer* para poder observar las predicciones realizadas. Los resultados obtenidos son los siguientes:

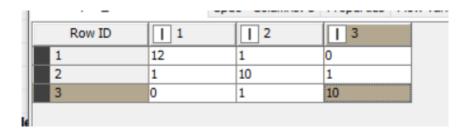


Figura 1: Clasificación del árbol de decisión