

19 Mayo 2020. Tarea Support Vector Machines. Knime. Parte2

Archivo: empresas. txt

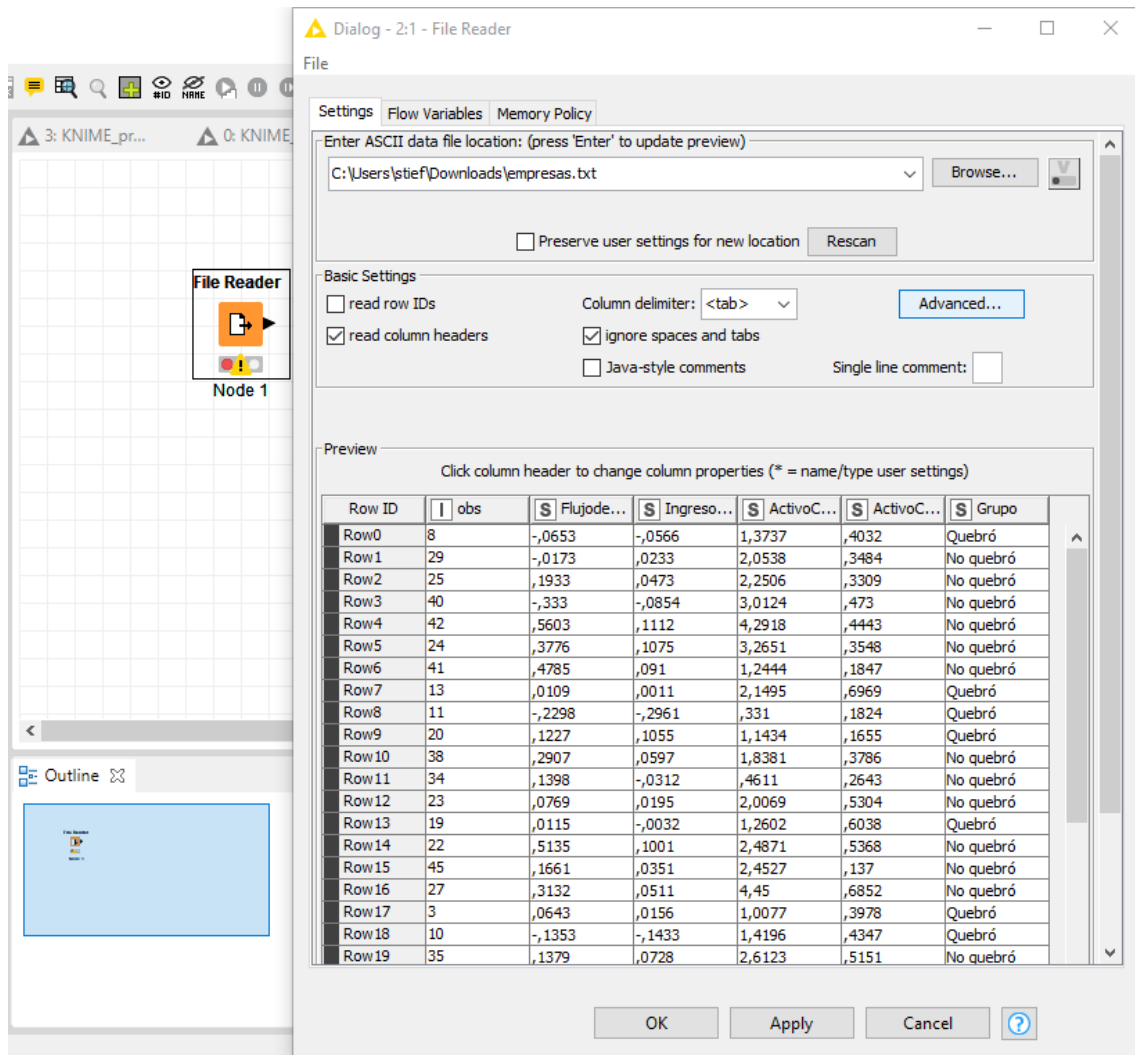
- 1. Cargar los datos con un nodo File Reader (descartar el atributo obs al cargar los datos)**
- 2. Mediante un nodo Partitioning separe el 80% aleatoriamente para construir el modelo**
- 3. Agregue un nodo SVM Learner considere class column Grupo (no cambie los parámetros por defecto)**
- 4. Agregue un nodo SVMPredictor. ¿En classification, qué representan las dos últimas columnas?**
- 5. Agregue un nodo Scorer, ¿Qué obtiene al ejecutarlo?.**

1.

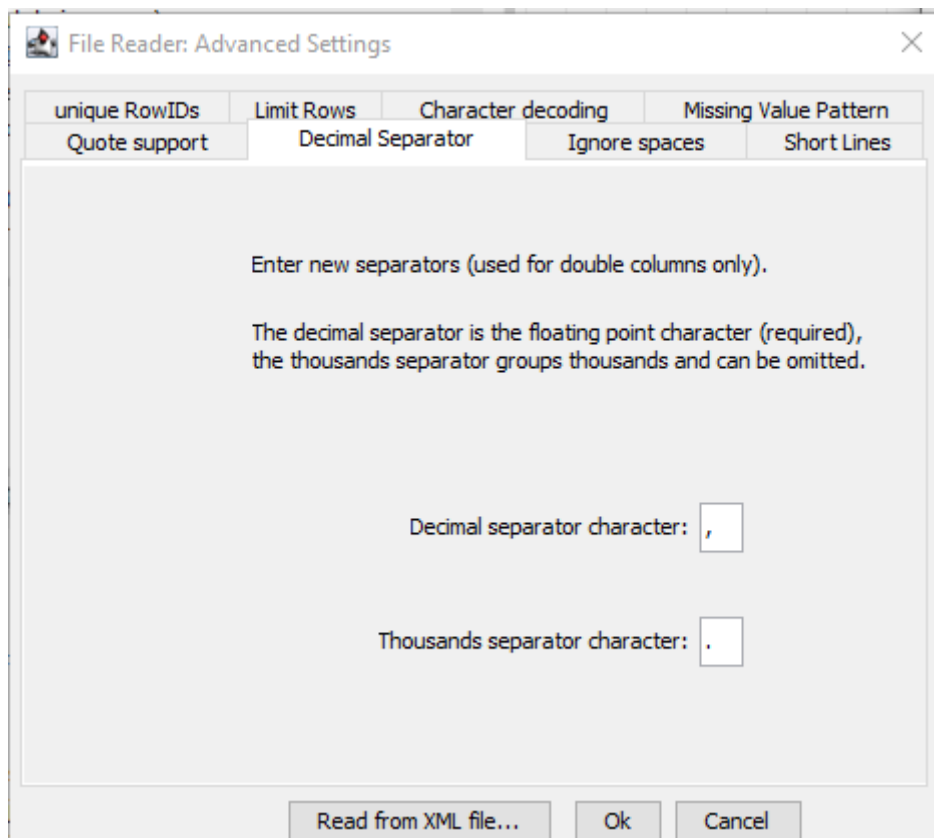
Comenzamos abriendo la herramienta Knime y creando un nuevo espacio de trabajo (File→New→New KNIME Workflow).

Buscamos en el “Node Repository” el nodo “File Reader” y lo arrastramos al espacio de trabajo.

Con el nodo seleccionado, pulsamos F6 para configurar el nodo. En la ventana siguiente buscaremos el archivo empresas.txt y lo abriremos. Lo más probable es que los datos nos salgan mal, como si estuviesen todos entremezclados. Si es el caso, debemos mirar que la opción “Column Delimeter” tenga señalada “<tab>”.

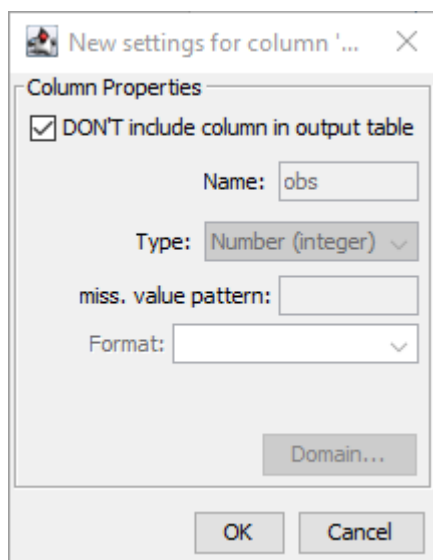


Como vemos a nosotros aún nos salen datos erróneos. Esto se debe a que el programa está confundiendo los separadores decimales con los separadores de millares. Por ello, pulsamos “Advanced” y en la pestaña “decimal separator” ponemos en el primer recuadro “,” y en el segundo “.”.

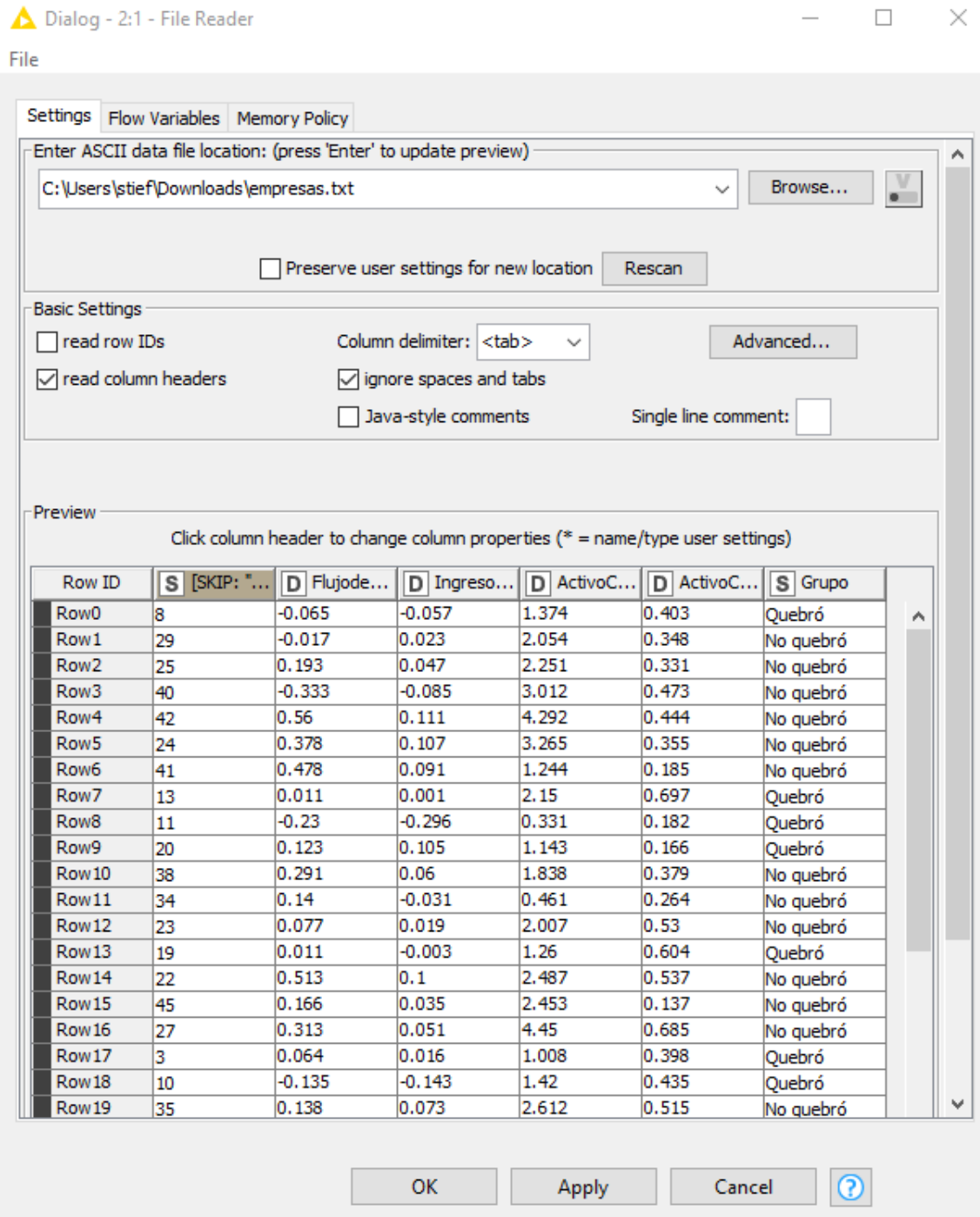


De esta manera ya nos saldrán correctamente representados los datos.

Queremos eliminar el atributo “obs”, por lo que hacemos click derecho en su columna y en el recuadro que nos saldrá seleccionamos la opción “Don’t include column in output table”.



De esta manera veremos que ahora en la columna donde se encuentra “obs” nos encontraremos que pone “SKIP”.



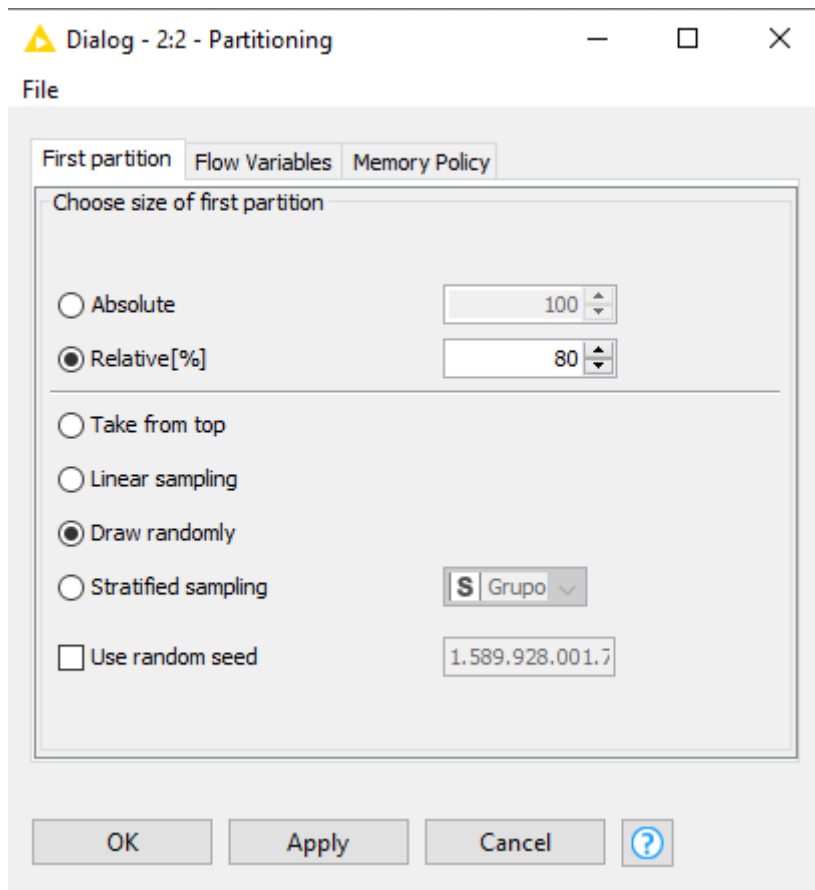
2.

En segundo lugar, vamos a dividir los datos aleatoriamente en el 80 y el 20%. El 80% lo usaremos para crear un modelo y el 20% restante lo usaremos para testear dicho modelo.

Por tanto, vamos a “Node Repository” y buscamos el nodo “Partitioning”, seleccionándolo y arrastrándolo hasta nuestro espacio de trabajo.

Es hora de configurarlo, por lo que, con el nodo seleccionado, pulsamos F6.

En la ventana emergente seleccionaremos como queremos tomar los datos y que cantidad queremos dividir. Nosotros dividiremos en el 80% de los datos, y los seleccionaremos al azar.

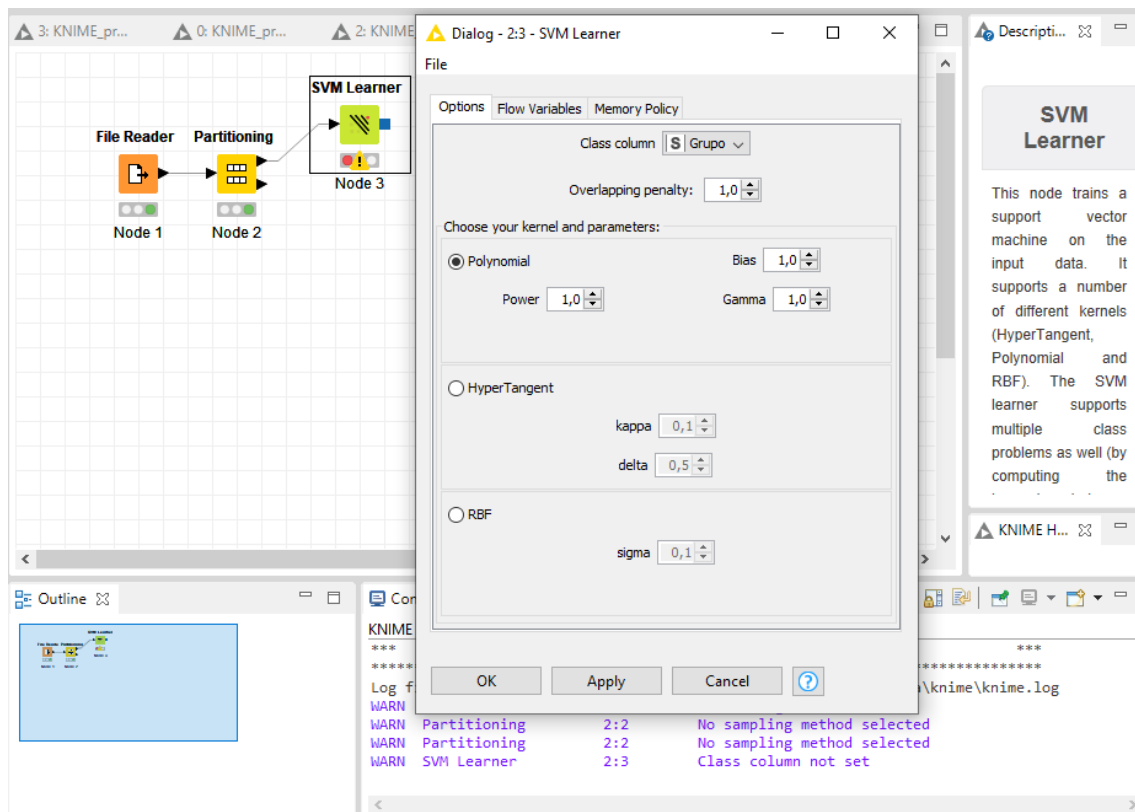


Pulsando F7 con el módulo seleccionado, lo ejecutamos.

3.

Buscamos en "Node Repository" el nodo "SVM Learner", lo seleccionamos y arrastramos al espacio de trabajo. Conectaremos la entrada del nodo con la salida superior del nodo "Partitioning", ya que es la salida del 80% de los datos que usaremos para crear nuestro modelo.

Con el nodo seleccionado, pulsaremos F6 ver el menú de configuración y asegurarnos que "Class Column" es "Grupo".



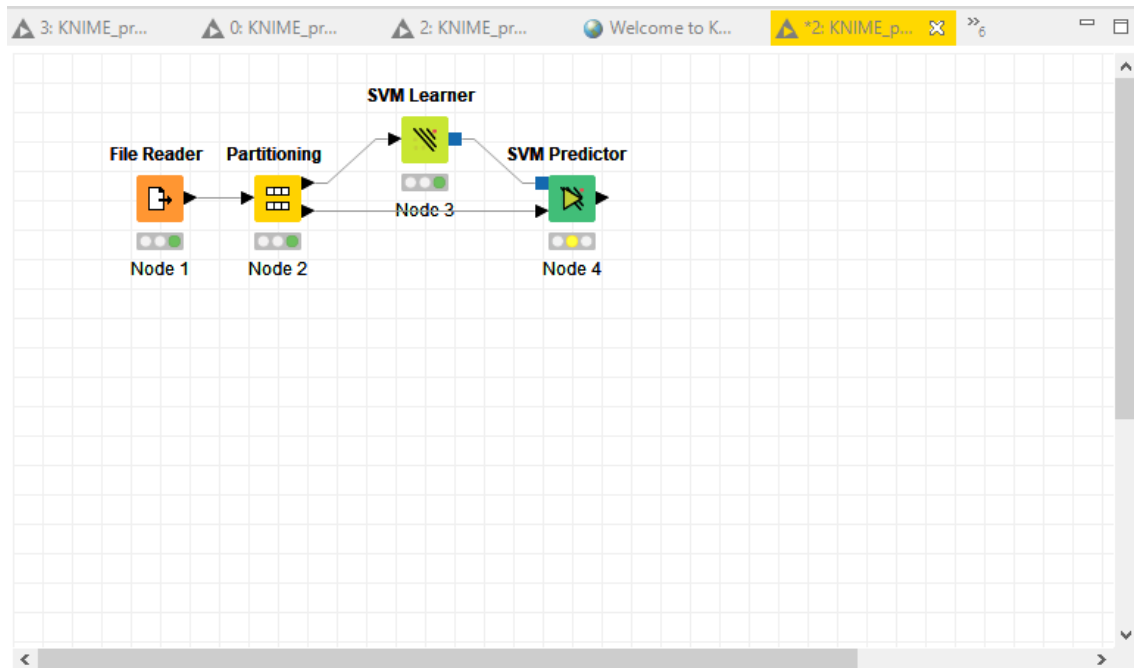
Pulsamos “Ok” y ejecutamos el nodo.

4.

Buscamos en “Node Repository” el nodo “SVM Predictor” y lo arrastramos hasta nuestro espacio de trabajo.

Conectaremos la salida del nodo “SVM Learner” a la entrada superior del nodo “SVM Predictor”, ya que este es el modelo creado con el 80% de los datos.

A la entrada inferior del nodo “SVM Predictor” le conectaremos la salida inferior del nodo “Partitioning”, que contiene el 20% de los datos que usaremos para testear el modelo.



Ejecutamos el nodo y hacemos click con el botón derecho en el nodo y seleccionamos la opción “Clasification”.

Aquí podemos observar que las dos últimas columnas son “Grupo” y la predicción de grupo.

Classification - 2:4 - SVM Predictor

File Hilite Navigation View

Table "default" - Rows: 10 Spec - Columns: 6 Properties Flow Variables

Row ID	D Flujo de...	D Ingreso...	D ActivoC...	D ActivoC...	S Grupo	S Predicti...
Row0	-0.065	-0.057	1.374	0.403	Quebró	Quebró
Row1	-0.017	0.023	2.054	0.348	No quebró	No quebró
Row3	-0.333	-0.085	3.012	0.473	No quebró	No quebró
Row5	0.378	0.107	3.265	0.355	No quebró	No quebró
Row13	0.011	-0.003	1.26	0.604	Quebró	Quebró
Row16	0.313	0.051	4.45	0.685	No quebró	No quebró
Row31	0.581	0.037	5.059	0.127	No quebró	No quebró
Row32	0.149	0.056	2.235	0.556	No quebró	No quebró
Row40	-0.563	-0.311	1.513	0.164	Quebró	Quebró
Row41	-0.142	-0.065	0.707	0.279	Quebró	Quebró

5.

Por último, buscaremos un nodo “Scorer” y lo arrastraremos hasta nuestro espacio de trabajo.

Le conectamos a la entrada la salida del nodo “SVM Predictor” y pulsamos F6 para asegurarnos que tenemos correctamente seleccionado el atributo que queremos explicado y su predicción:

Dialog - 2:5 - Scorer

File

Scorer Flow Variables Memory Policy

First Column

S Grupo

Second Column

S Prediction (Grupo)

Sorting of values in tables

Sorting strategy: Insertion order Reverse order

Provide scores as flow variables

Use name prefix

Missing values

In case of missing values... Ignore Fail

OK Apply Cancel ?

Pulsamos "Ok" y ejecutamos el nodo. Haciendo click derecho sobre el mismo, seleccionamos "Confusion matrix". En la ventana emergente observaremos la matriz de confusión, que en este caso clasifico correctamente el 100% de los casos. Es decir, el modelo creado con el 80% de los datos se asemeja al 100% con la realidad en este caso.

Confusion matrix - 2:5 - Scorer

File Hilite Navigation View

Table "spec_name" - Rows: 2 Spec - Columns: 2 Properties Flow Variables

Row ID	Quebró	No que...
Quebró	4	0
No quebró	0	6