Escuela Superior de Cómputo

Instituto Politécnico Nacional

Escuela Superior de Cómputo

Examen práctico de árboles - KNIME

Unidad de aprendizaje: Data Mining

Grupo: 3CV6

Alumno(a):
Ramos Diaz Enrique

Profesor(a):
Ocampo Botello Fabiola

1 de noviembre 2019

Índice

1. Diccionario de datos

La base de datos a utilizar tiene como nombre *Evaluación de Autos.csv*. Para fines prácticos, se renombro como *cars.csv*

buying: Precio de compra		
Dominio de datos	**Significado**	**Tipo de dato**
vhigh	Muy alto	String
high	Alto	String
med	Medio	String
low	Bajo	String

maint: Costo de mantenimiento		
Dominio de datos	**Significado**	**Tipo de dato**
vhigh	Muy alto	String
high	Alto	String
med	Medio	String
low	Bajo	String

doors: Número de puertas		
Dominio de datos	**Significado**	**Tipo de dato**
2	Dos	Int
3	Tres	Int
4	Cuatro	Int
5more	Cinco o más	String

persons: Capacidad de personas a bordo		
Dominio de datos	**Significado**	**Tipo de dato**
2	Dos	Int
4	Cuatro	Int
more	Cinco o más	String

lug_boot: Tamaño del maletero		
Dominio de datos	**Significado**	**Tipo de dato**
small	Pequeño	String
med	Mediano	String
big	Grande	String

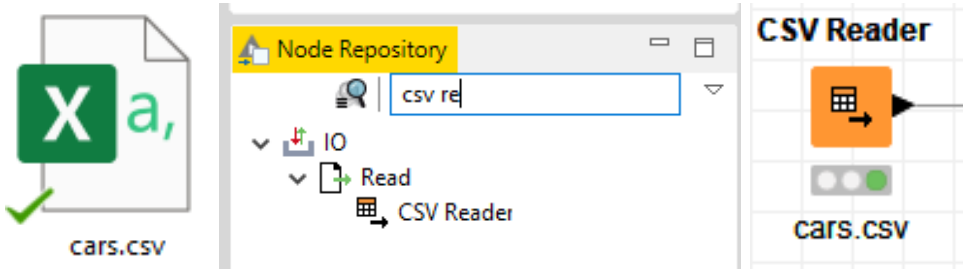
safety: Seguridad		
Dominio de datos	**Significado**	**Tipo de dato**
low	Baja	String
med	Media	String
big	Alta	String

class: Estado del auto		
Dominio de datos	**Significado**	**Tipo de dato**
unacc	Inaceptable	String
acc	Aceptable	String
good	Bueno	String
vgood	Muy bueno	String

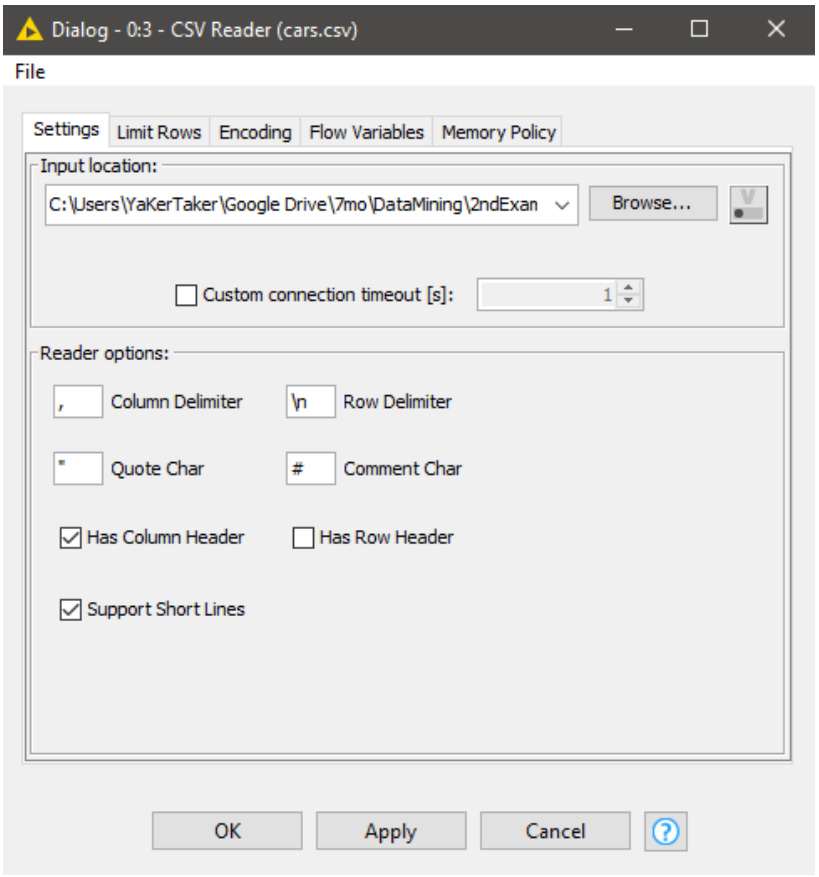
2. Desarrollo: Proceso KDD

2.1. Carga de la base de datos

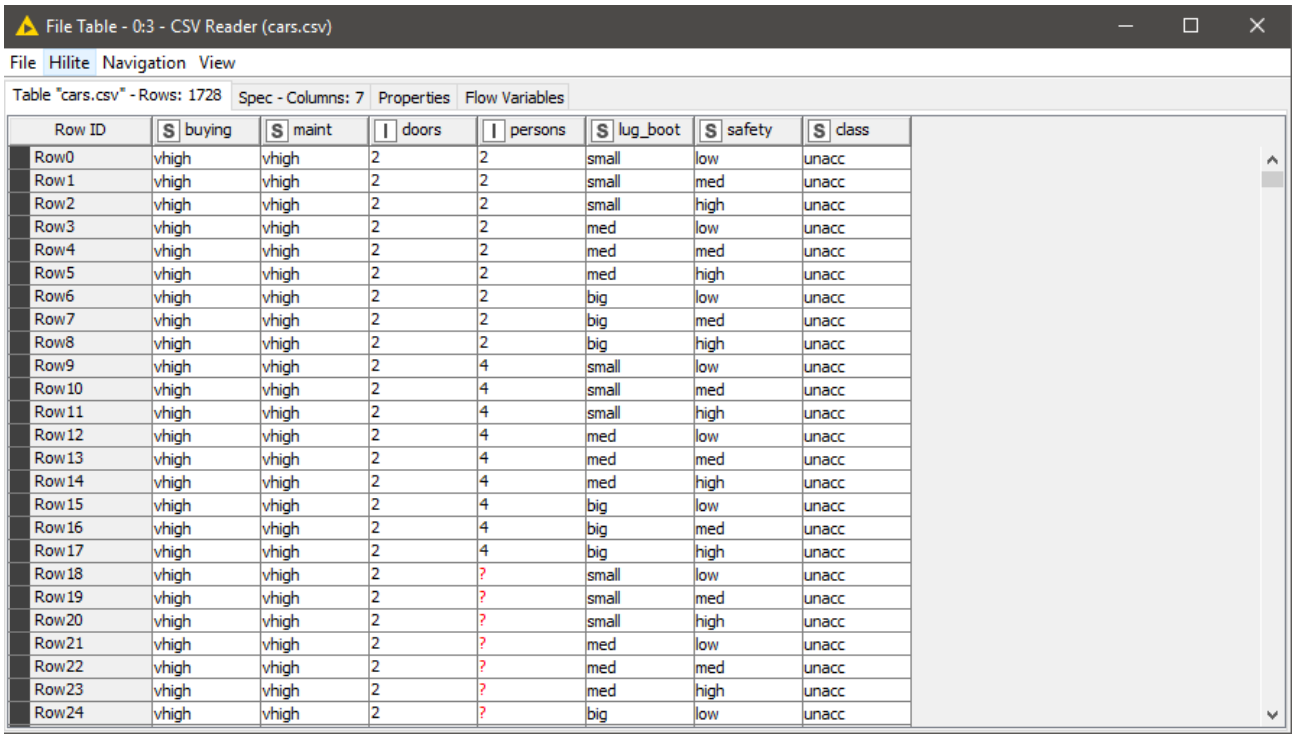
Dentro de KNIME, cargamos la base de datos con el nodo **CSV Reader**.



A continuación se muestra la configuración de esta pantalla para el propósito de esta práctica:



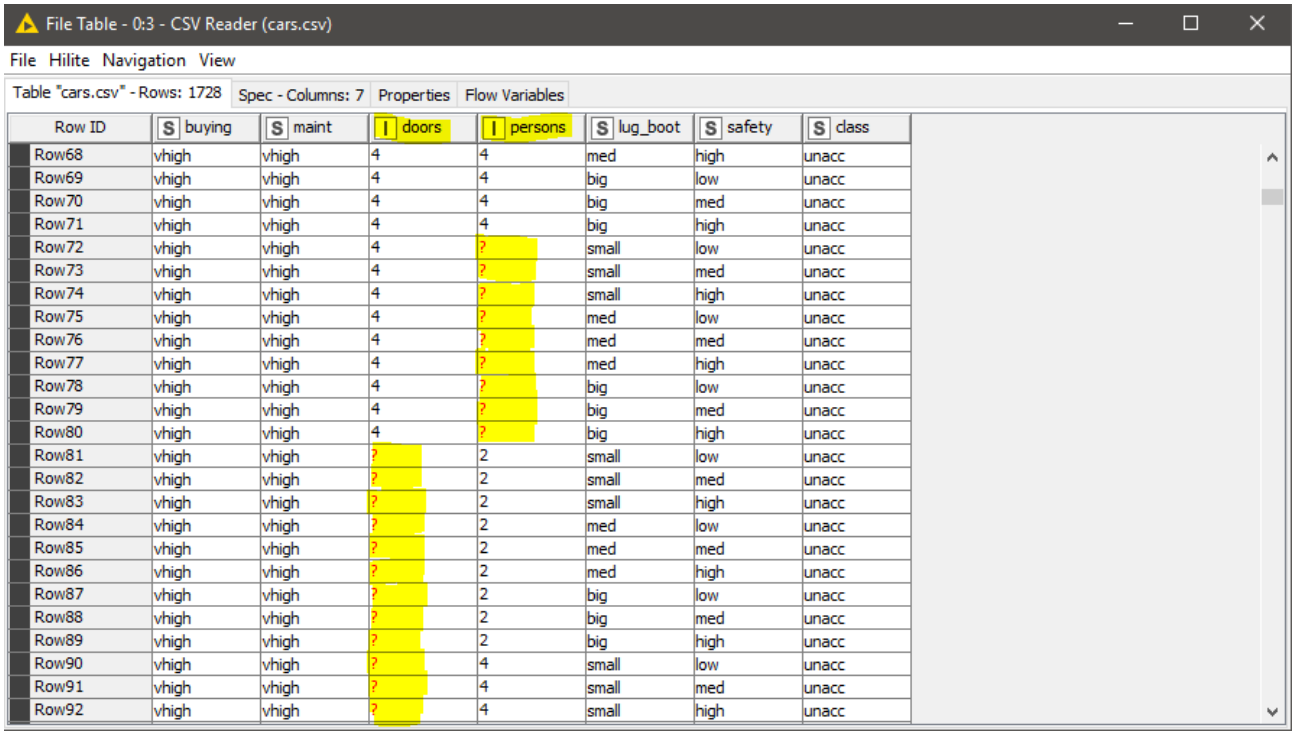
Podemos verificar la carga correcta de la base de datos seleccionando la opción *File Table*:



Row ID	buying	maint	doors	persons	lug_boot	safety	class
Row0	vhigh	vhigh	2	2	small	low	unacc
Row1	vhigh	vhigh	2	2	small	med	unacc
Row2	vhigh	vhigh	2	2	small	high	unacc
Row3	vhigh	vhigh	2	2	med	low	unacc
Row4	vhigh	vhigh	2	2	med	med	unacc
Row5	vhigh	vhigh	2	2	med	high	unacc
Row6	vhigh	vhigh	2	2	big	low	unacc
Row7	vhigh	vhigh	2	2	big	med	unacc
Row8	vhigh	vhigh	2	2	big	high	unacc
Row9	vhigh	vhigh	2	4	small	low	unacc
Row10	vhigh	vhigh	2	4	small	med	unacc
Row11	vhigh	vhigh	2	4	small	high	unacc
Row12	vhigh	vhigh	2	4	med	low	unacc
Row13	vhigh	vhigh	2	4	med	med	unacc
Row14	vhigh	vhigh	2	4	med	high	unacc
Row15	vhigh	vhigh	2	4	big	low	unacc
Row16	vhigh	vhigh	2	4	big	med	unacc
Row17	vhigh	vhigh	2	4	big	high	unacc
Row18	vhigh	vhigh	2	?	small	low	unacc
Row19	vhigh	vhigh	2	?	small	med	unacc
Row20	vhigh	vhigh	2	?	small	high	unacc
Row21	vhigh	vhigh	2	?	med	low	unacc
Row22	vhigh	vhigh	2	?	med	med	unacc
Row23	vhigh	vhigh	2	?	med	high	unacc
Row24	vhigh	vhigh	2	?	big	low	unacc

2.2. Limpieza de datos

Al revisar los registros de la base de datos, notaremos que aparentemente existen datos faltantes o nulos. Sin embargo, no lo son; se tratan de valores que poseen un significado real en las columnas *persons* y *doors*, pero que la herramienta de KNIME detecta como nulos. En la sección de **Transformación de los datos** se hará un tratamiento específico para estos.


File Table - 0:3 - CSV Reader (cars.csv)

File Hilite Navigation View

Table "cars.csv" - Rows: 1728 | Spec - Columns: 7 | Properties | Flow Variables

Row ID	buying	maint	doors	persons	lug_boot	safety	class
Row68	vhigh	vhigh	4	4	med	high	unacc
Row69	vhigh	vhigh	4	4	big	low	unacc
Row70	vhigh	vhigh	4	4	big	med	unacc
Row71	vhigh	vhigh	4	4	big	high	unacc
Row72	vhigh	vhigh	4	?	small	low	unacc
Row73	vhigh	vhigh	4	?	small	med	unacc
Row74	vhigh	vhigh	4	?	small	high	unacc
Row75	vhigh	vhigh	4	?	med	low	unacc
Row76	vhigh	vhigh	4	?	med	med	unacc
Row77	vhigh	vhigh	4	?	med	high	unacc
Row78	vhigh	vhigh	4	?	big	low	unacc
Row79	vhigh	vhigh	4	?	big	med	unacc
Row80	vhigh	vhigh	4	?	big	high	unacc
Row81	vhigh	vhigh	?	2	small	low	unacc
Row82	vhigh	vhigh	?	2	small	med	unacc
Row83	vhigh	vhigh	?	2	small	high	unacc
Row84	vhigh	vhigh	?	2	med	low	unacc
Row85	vhigh	vhigh	?	2	med	med	unacc
Row86	vhigh	vhigh	?	2	med	high	unacc
Row87	vhigh	vhigh	?	2	big	low	unacc
Row88	vhigh	vhigh	?	2	big	med	unacc
Row89	vhigh	vhigh	?	2	big	high	unacc
Row90	vhigh	vhigh	?	4	small	low	unacc
Row91	vhigh	vhigh	?	4	small	med	unacc
Row92	vhigh	vhigh	?	4	small	high	unacc

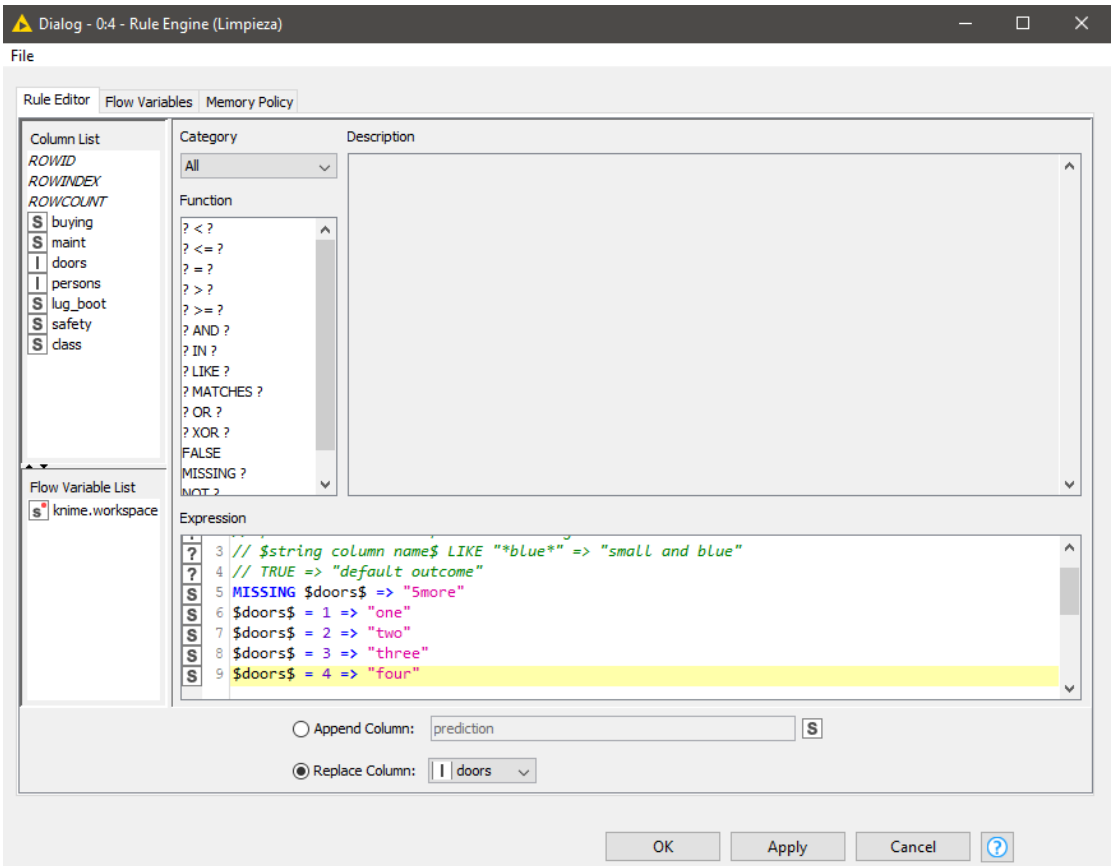
2.3. Integración y selección de los datos

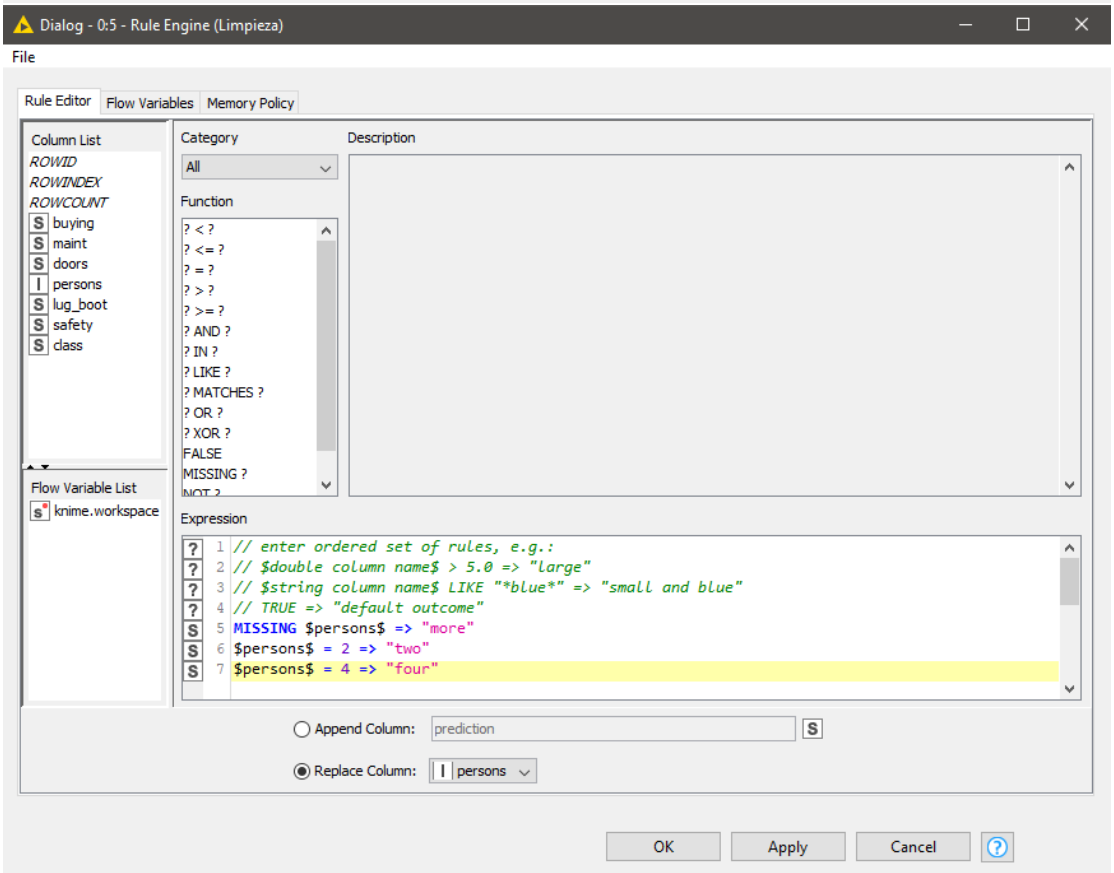
Solo nos basta con la base de datos principal *cars.csv*. Se utilizarán todos los atributos disponibles en la base de datos, así como todos sus registros, por lo que no es necesario hacer alguna selección ni integración adicional específica.

2.4. Transformación de los datos

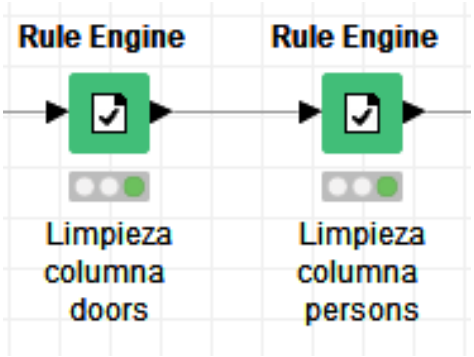
Utilizando nodos **Rule Engine**, debemos reemplazar los aparentes datos faltantes en los atributos *persons* y *doors*. Para ambos, se reemplazarán estos por el valor real de puertas y personas, el cual es "5more" y "more" respectivamente.

Los otros valores de estos dos atributos se deben convertir de numéricos a nominales, ya que el tipo de árbol que se utilizará en KNIME es **ID3**, por lo que los reemplazamos los números por su nombre en texto, como se muestra a continuación:


Dialog - 0:4 - Rule Engine (Limpieza)
File
Rule Editor
Flow Variables
Memory Policy
Column List
ROWID
ROWINDEX
ROWCOUNT
buying
maint
doors
persons
lug_boot
safety
class
Flow Variable List
knime.workspace
Category
All
Description
Function
? < ?
? <= ?
? = ?
? > ?
? >= ?
? AND ?
? IN ?
? LIKE ?
? MATCHES ?
? OR ?
? XOR ?
FALSE
MISSING ?
Expression
// $string column name$ LIKE "*blue*" => "small and blue"
// TRUE => "default outcome"
MISSING $doors$ => "5more"
$doors$ = 1 => "one"
$doors$ = 2 => "two"
$doors$ = 3 => "three"
$doors$ = 4 => "four"
Append Column:
prediction
Replace Column:
doors
OK
Apply
Cancel


Dialog - 0:5 - Rule Engine (Limpieza)
File
Rule Editor
Flow Variables
Memory Policy
Column List
ROWID
ROWINDEX
ROWCOUNT
buying
maint
doors
persons
lug_boot
safety
class
Flow Variable List
knime.workspace
Category
All
Description
Function
? < ?
? <= ?
? = ?
? > ?
? >= ?
? AND ?
? IN ?
? LIKE ?
? MATCHES ?
? OR ?
? XOR ?
FALSE
MISSING ?
Expression
// enter ordered set of rules, e.g.:
// $double column name$ > 5.0 => "large"
// $string column name$ LIKE "*blue*" => "small and blue"
// TRUE => "default outcome"
MISSING $persons$ => "more"
$persons$ = 2 => "two"
$persons$ = 4 => "four"
Append Column:
prediction
Replace Column:
persons
OK
Apply
Cancel

Se necesitan dos nodos, uno para cada atributo.



Verificamos que se han reemplazado correctamente los registros de los atributos *doors* y *persons* seleccionando la opción *Classified values* del último nodo, siendo ahora cadenas de texto (nominales):

Classified values - 0:5 - Rule Engine (Limpieza)

File Hilite Navigation View

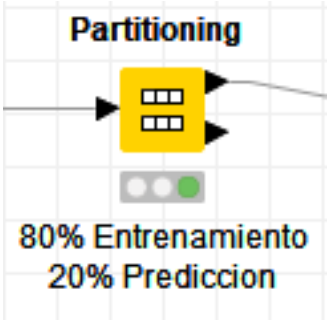
Table "default" - Rows: 1728 | Spec - Columns: 7 | Properties | Flow Variables

Row ID	buying	maint	doors	persons	lug_boot	safety	class
Row194	vhigh	high	5more	two	med	high	unacc
Row195	vhigh	high	5more	two	big	low	unacc
Row196	vhigh	high	5more	two	big	med	unacc
Row197	vhigh	high	5more	two	big	high	unacc
Row198	vhigh	high	5more	four	small	low	unacc
Row199	vhigh	high	5more	four	small	med	unacc
Row200	vhigh	high	5more	four	small	high	unacc
Row201	vhigh	high	5more	four	med	low	unacc
Row202	vhigh	high	5more	four	med	med	unacc
Row203	vhigh	high	5more	four	med	high	unacc
Row204	vhigh	high	5more	four	big	low	unacc
Row205	vhigh	high	5more	four	big	med	unacc
Row206	vhigh	high	5more	four	big	high	unacc
Row207	vhigh	high	5more	more	small	low	unacc
Row208	vhigh	high	5more	more	small	med	unacc
Row209	vhigh	high	5more	more	small	high	unacc
Row210	vhigh	high	5more	more	med	low	unacc
Row211	vhigh	high	5more	more	med	med	unacc
Row212	vhigh	high	5more	more	med	high	unacc
Row213	vhigh	high	5more	more	big	low	unacc
Row214	vhigh	high	5more	more	big	med	unacc
Row215	vhigh	high	5more	more	big	high	unacc
Row216	vhigh	med	two	two	small	low	unacc
Row217	vhigh	med	two	two	small	med	unacc
Row218	vhigh	med	two	two	small	high	unacc

2.5. Minería de datos

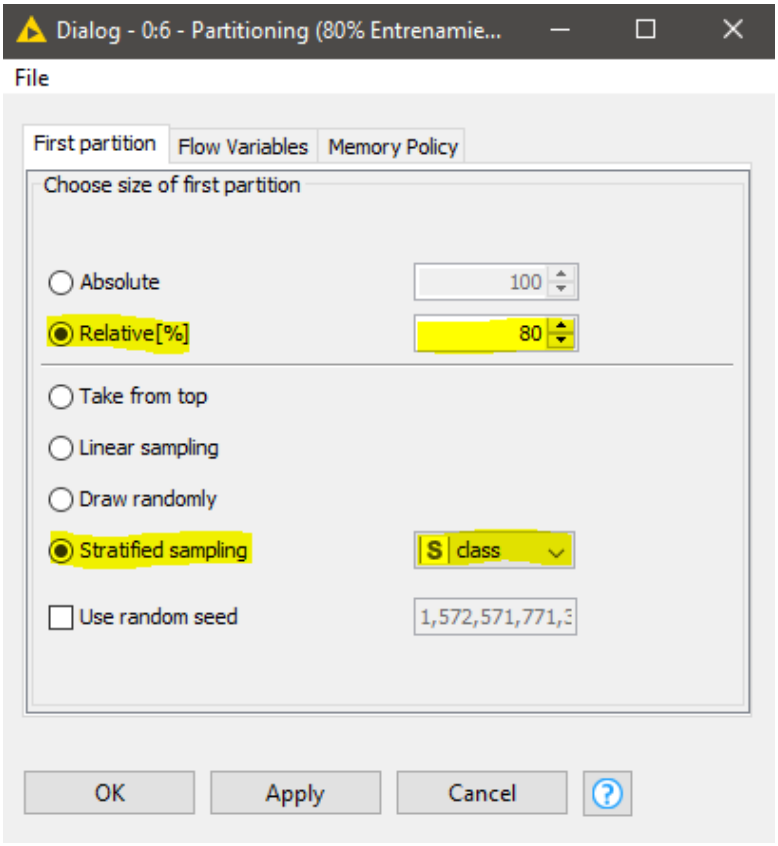
2.5.1. Partición de los datos

Al ser los árboles de decisiones algoritmos de aprendizaje supervisados, es necesario dividir el conjunto de datos para generar el modelo. Una parte será para el entrenamiento o aprendizaje, que será el 80 % de todos los datos. La otra parte servirá para ser evaluada por el árbol de predicción, que por el momento no será utilizada en ésta práctica (esta parte corresponde al 20 % restante).



Para este fin utilizamos el nodo **Partitioning**. Tendrá como entrada la salida del último nodo **Rule Engine** con los registros ya reemplazados.

La pantalla de configuración deberá quedar como se muestra a continuación:



En la opción *Stratified sampling* se selecciona el atributo de interés para la futura clasificación, en este caso **class**.

Verificamos la partición de entrenamiento (80 %), que es la única que usaremos:

First partition (as defined in dialog) - 0:6 - Partitioning (80% Entrenamiento)

File Hilite Navigation View

Table "default" - Rows: 1382 Spec - Columns: 7 Properties Flow Variables

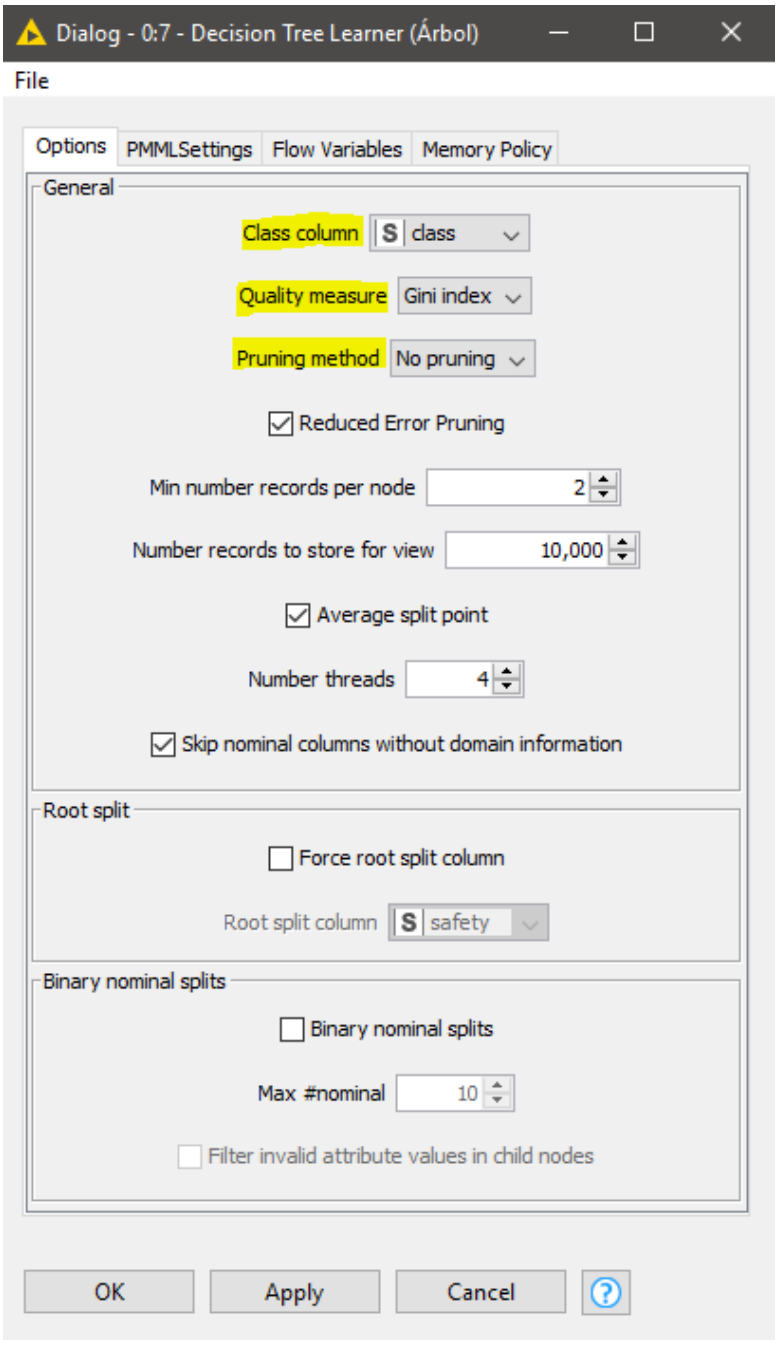
Row ID	buying	maint	doors	persons	lug_boot	safety	class
Row0	vhigh	vhigh	two	two	small	low	unacc
Row1	vhigh	vhigh	two	two	small	med	unacc
Row2	vhigh	vhigh	two	two	small	high	unacc
Row3	vhigh	vhigh	two	two	med	low	unacc
Row4	vhigh	vhigh	two	two	med	med	unacc
Row5	vhigh	vhigh	two	two	med	high	unacc
Row6	vhigh	vhigh	two	two	big	low	unacc
Row7	vhigh	vhigh	two	two	big	med	unacc
Row8	vhigh	vhigh	two	two	big	high	unacc
Row9	vhigh	vhigh	two	four	small	low	unacc
Row10	vhigh	vhigh	two	four	small	med	unacc
Row12	vhigh	vhigh	two	four	med	low	unacc
Row14	vhigh	vhigh	two	four	med	high	unacc
Row15	vhigh	vhigh	two	four	big	low	unacc
Row16	vhigh	vhigh	two	four	big	med	unacc
Row17	vhigh	vhigh	two	four	big	high	unacc
Row18	vhigh	vhigh	two	more	small	low	unacc
Row20	vhigh	vhigh	two	more	small	high	unacc
Row21	vhigh	vhigh	two	more	med	low	unacc
Row22	vhigh	vhigh	two	more	med	med	unacc
Row23	vhigh	vhigh	two	more	med	high	unacc
Row25	vhigh	vhigh	two	more	big	med	unacc
Row26	vhigh	vhigh	two	more	big	high	unacc
Row27	vhigh	vhigh	three	two	small	low	unacc
Row28	vhigh	vhigh	three	two	small	med	unacc

2.5.2. Árbol de decisión ID3: Aprendizaje

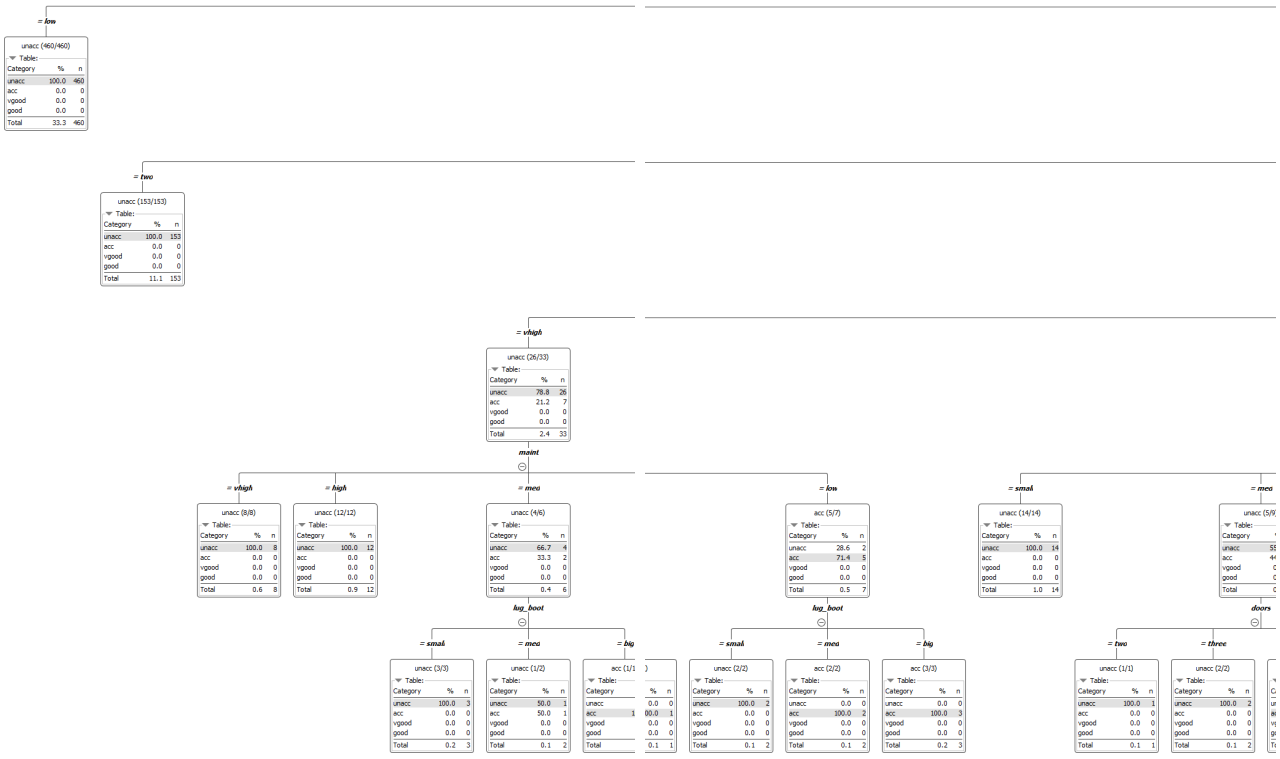
Es necesario entrenar al árbol de aprendizaje ID3 con un conjunto de datos, que posteriormente, podrá ser usado como referencia para clasificar nuevos datos entrantes desconocidos.

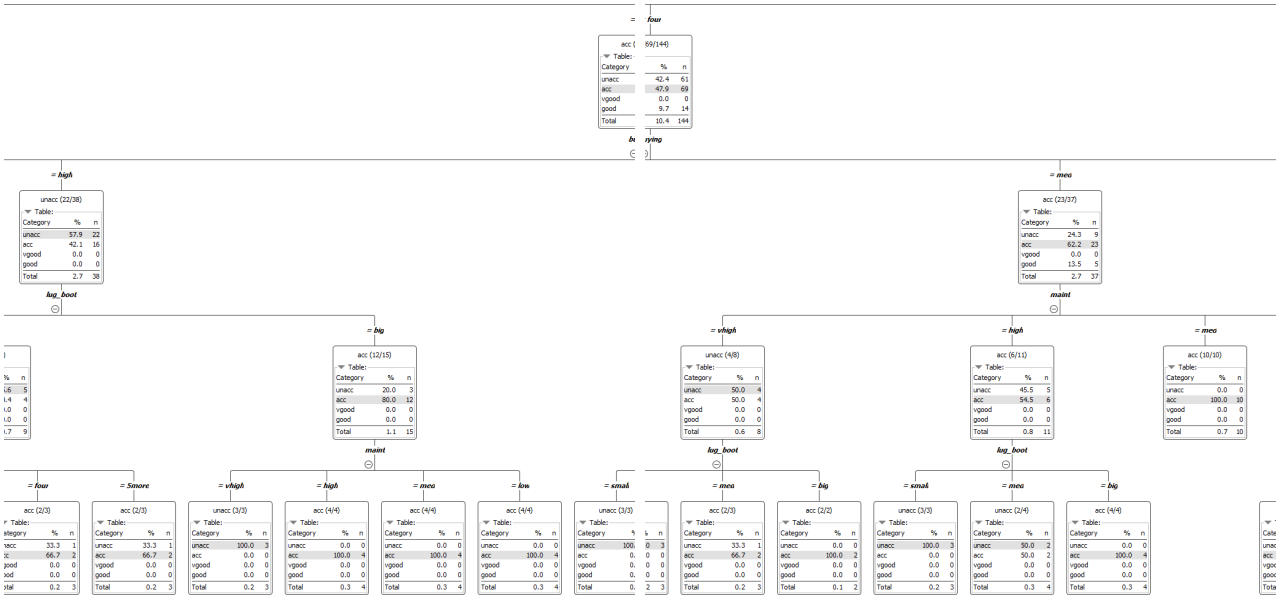
Agregamos al espacio de trabajo el nodo **Decision Tree Learner**. A la entrada conectamos la salida del nodo **Partitioning** correspondiente al 80 % del total de registros de la base de datos.

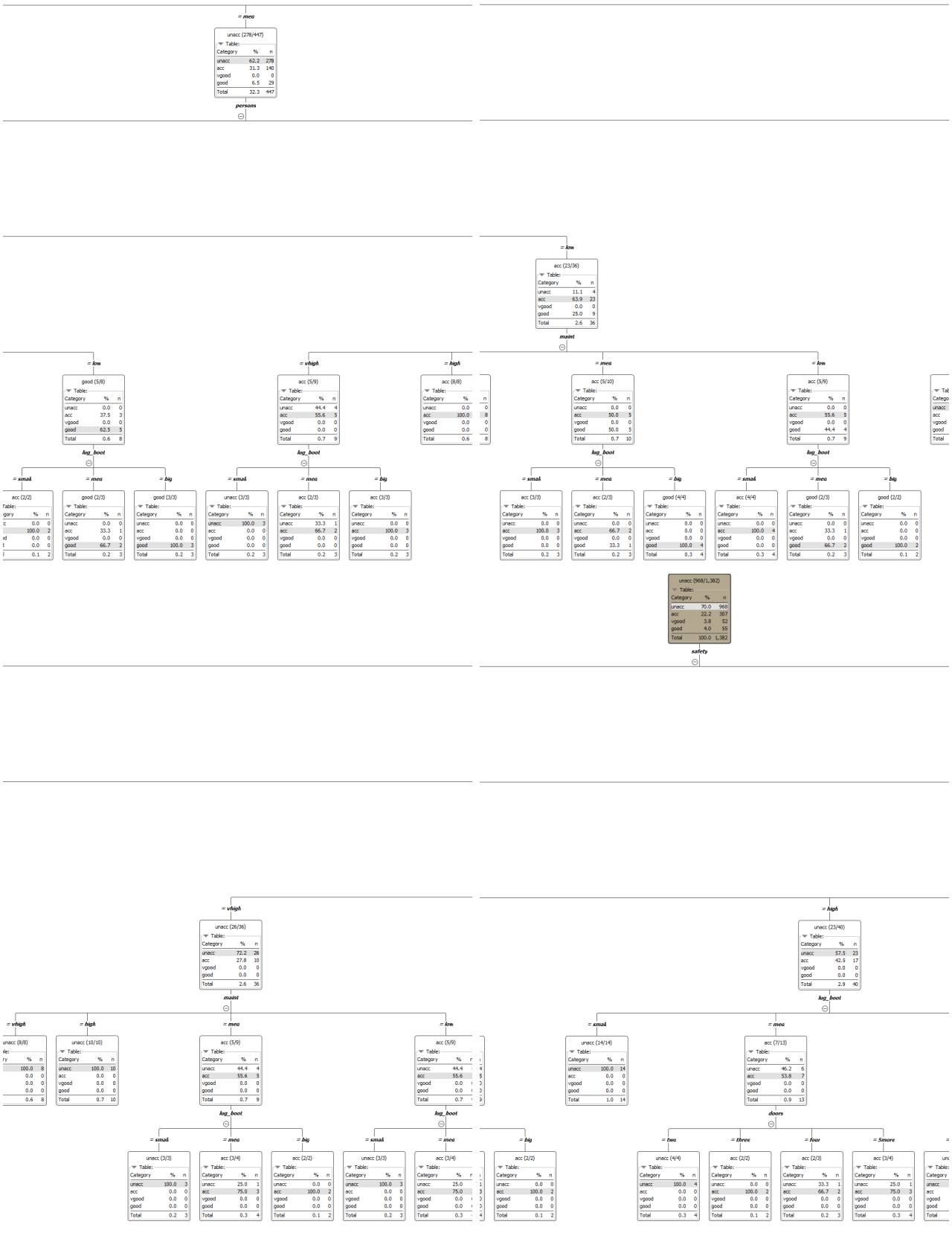
La pantalla de configuración queda de la siguiente manera:

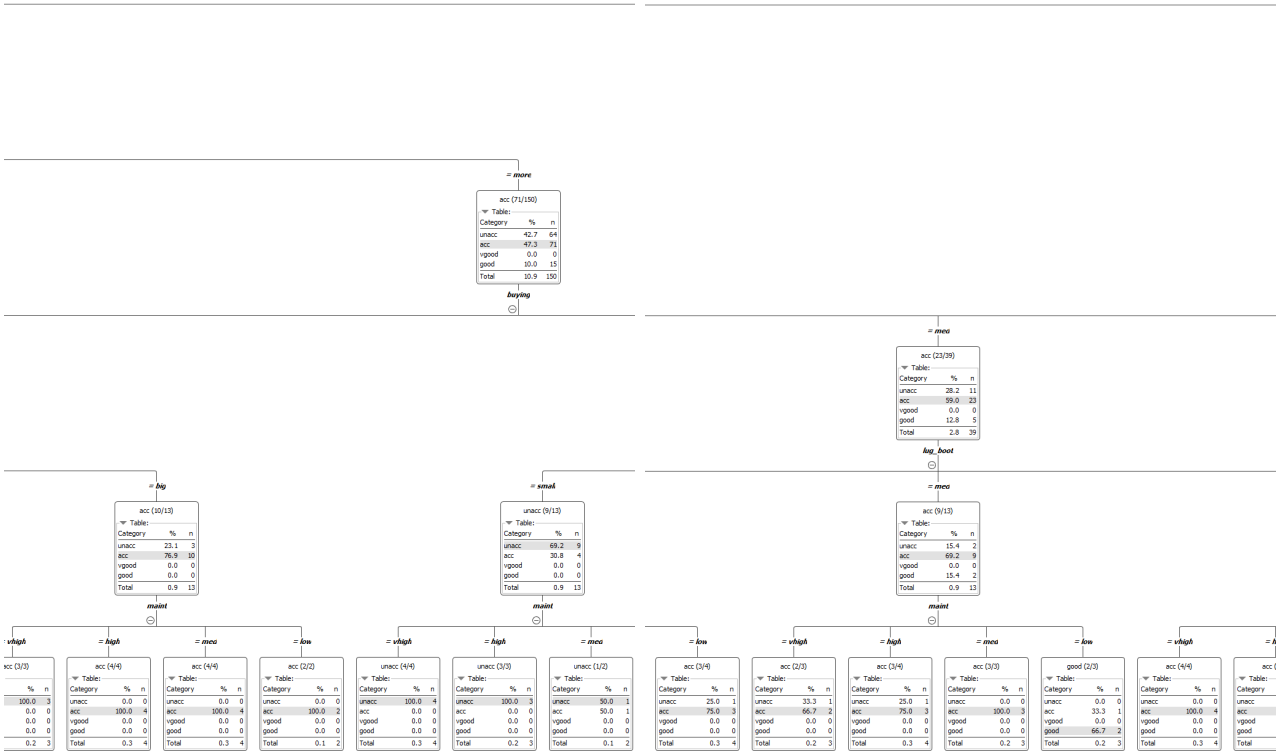


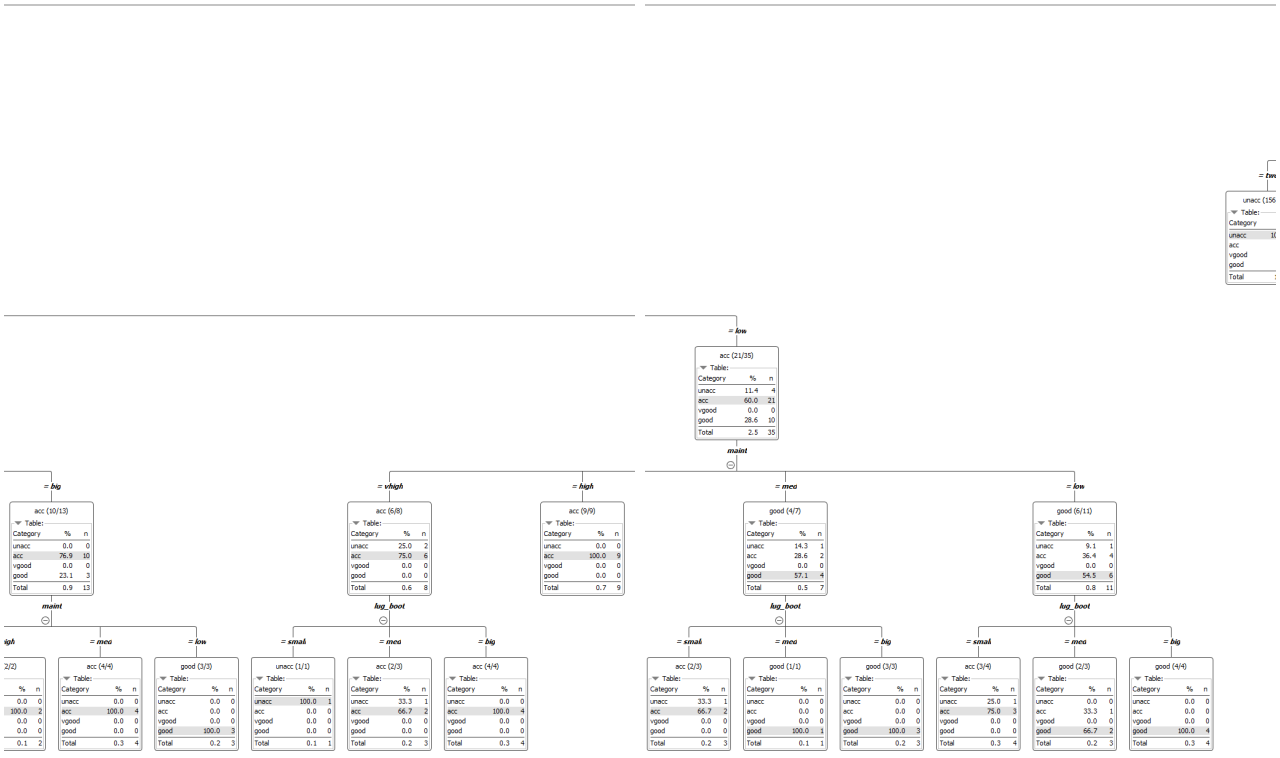
El árbol de decisión ID3 generado es el siguiente:

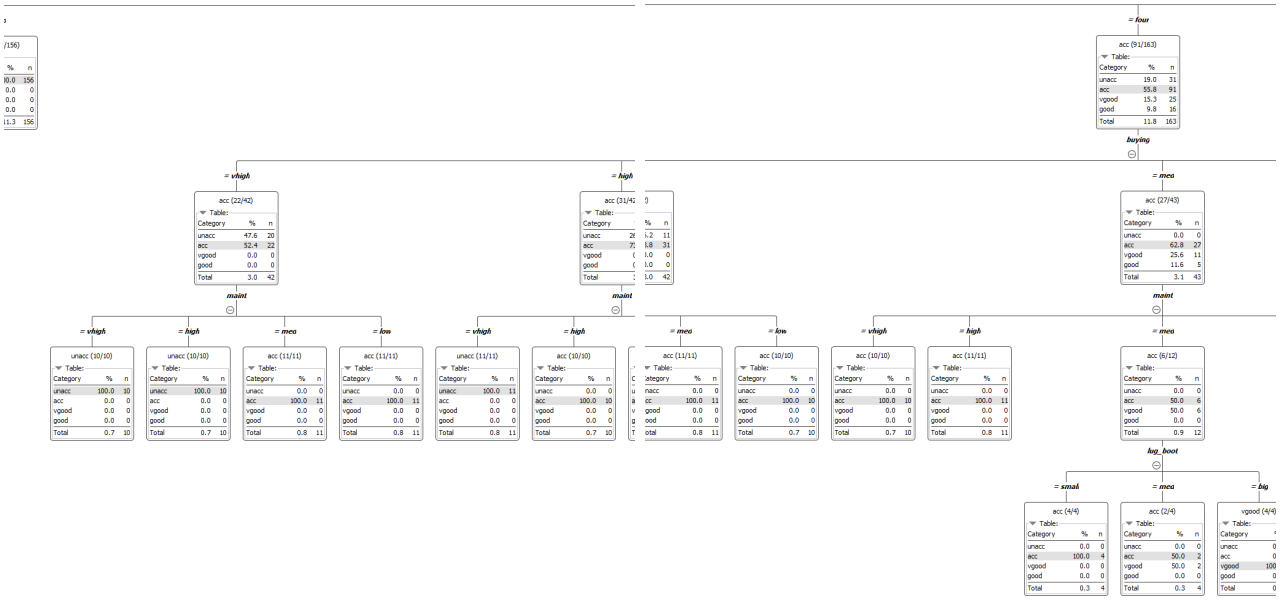

= low
unacc (460/460)
= two
unacc (153/153)
= vhigh
unacc (26/33)
maint
= vhigh
unacc (8/8)
= high
unacc (12/12)
= mea
unacc (4/6)
lug_boot
= small
unacc (3/3)
= mea
unacc (1/2)
= big
acc (1/1
= low
acc (5/7)
lug_boot
= small
unacc (2/2)
= mea
acc (2/2)
= big
acc (3/3)
= small
unacc (14/14)
= mea
unacc (5/9
doors
= two
unacc (1/1)
= three
unacc (2/2)

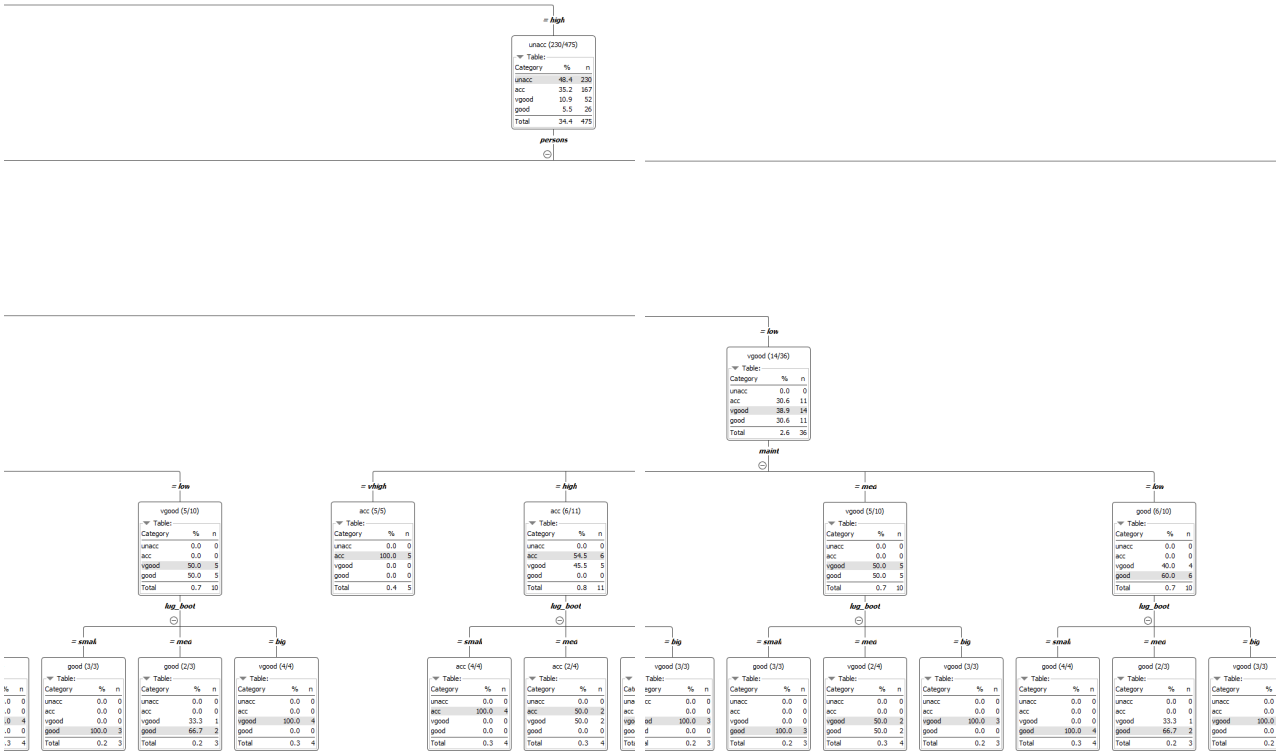

= four
acc (69/144)
= high
unacc (22/38)
lug_boot
= big
acc (12/15)
maint
= four
acc (2/3)
= 5more
acc (2/3)
= vhigh
unacc (3/3)
= high
acc (4/4)
= mea
acc (4/4)
= low
acc (4/4)
= small
unacc (3/3)
= mea
acc (23/37)
maint
= vhigh
unacc (4/8)
= high
acc (6/11)
= mea
acc (10/10)
lug_boot
= mea
acc (2/3)
= big
acc (2/2)
= small
unacc (3/3)
= mea
unacc (2/4)
= big
acc (4/4)

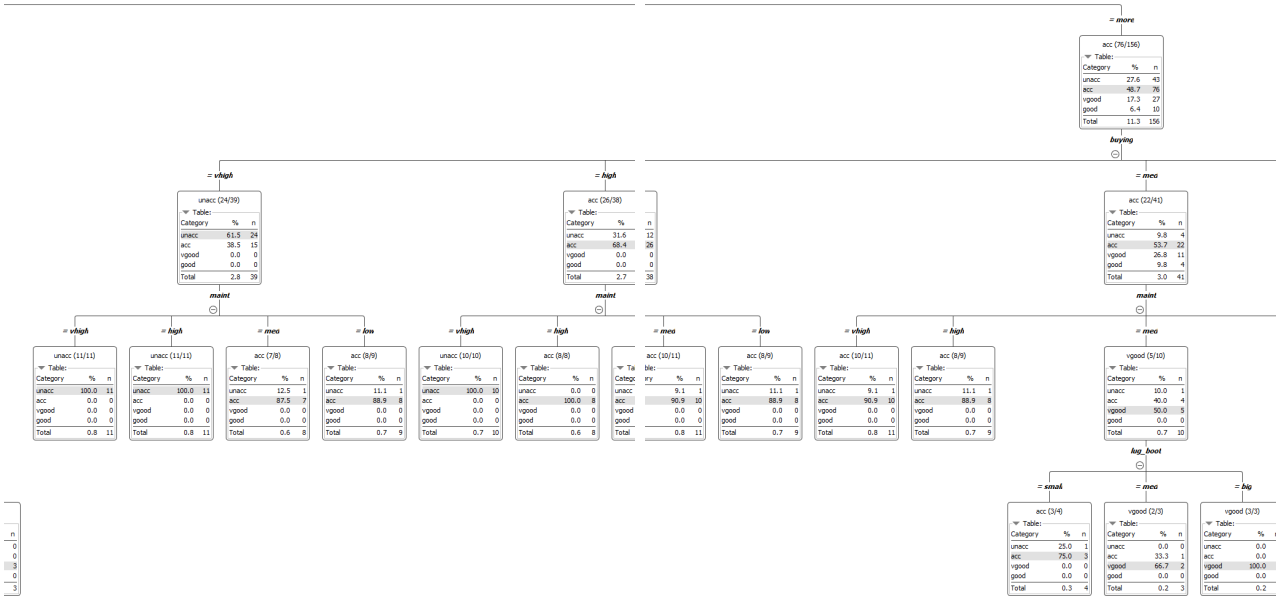

= meo
unacc (278/447)
persons
= low
= vhigh
= high
maint
lug_boot
= small
= med
= big
safety
unacc (968/1,382)
unacc (26/36)
unacc (23/40)
doors
= two
= three
= four
= 5more

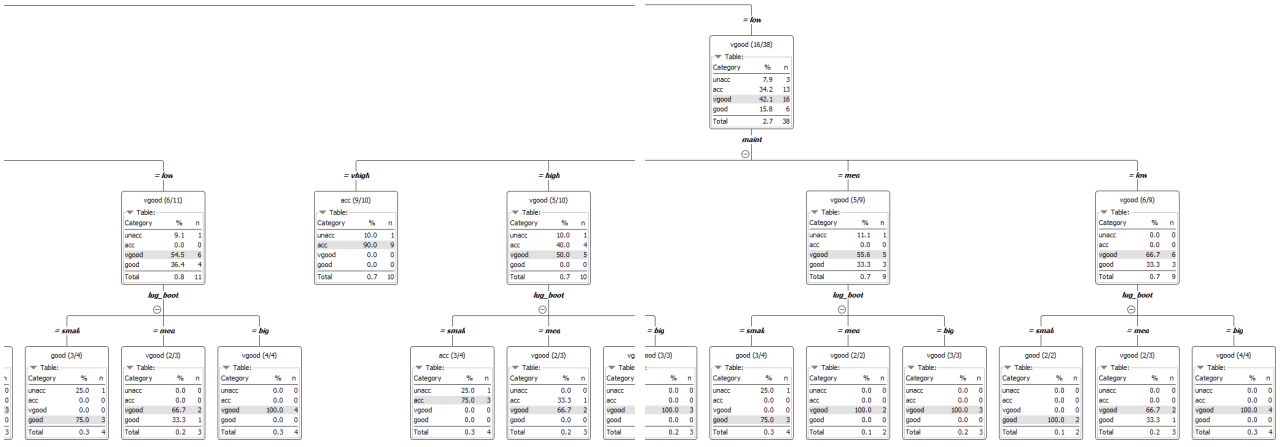

= more
acc (71/150)
Table:
Category % n
unacc 42.7 64
acc 47.3 71
vgood 0.0 0
good 10.0 15
Total 10.9 150
buying
= mea
acc (23/39)
Table:
Category % n
unacc 28.2 11
acc 59.0 23
vgood 0.0 0
good 12.8 5
Total 2.8 39
lug_boot
= big
acc (10/13)
Table:
Category % n
unacc 23.1 3
acc 76.9 10
vgood 0.0 0
good 0.0 0
Total 0.9 13
maint
= small
unacc (9/13)
Table:
Category % n
unacc 69.2 9
acc 30.8 4
vgood 0.0 0
good 0.0 0
Total 0.9 13
maint
= mea
acc (9/13)
Table:
Category % n
unacc 15.4 2
acc 69.2 9
vgood 0.0 0
good 15.4 2
Total 0.9 13
maint


= low
acc (21/35)
Table:
Category % n
unacc 11.4 4
acc 60.0 21
vgood 0.0 0
good 28.6 10
Total 2.5 35
maint
= big
acc (10/13)
Table:
Category % n
unacc 0.0 0
acc 76.9 10
vgood 0.0 0
good 23.1 3
Total 0.9 13
maint
= vhigh
acc (6/8)
Table:
Category % n
unacc 25.0 2
acc 75.0 6
vgood 0.0 0
good 0.0 0
Total 0.6 8
lug_boot
= high
acc (9/9)
Table:
Category % n
unacc 0.0 0
acc 100.0 9
vgood 0.0 0
good 0.0 0
Total 0.7 9
= mea
good (4/7)
Table:
Category % n
unacc 14.3 1
acc 28.6 2
vgood 0.0 0
good 57.1 4
Total 0.5 7
lug_boot
= low
good (6/11)
Table:
Category % n
unacc 9.1 1
acc 36.4 4
vgood 0.0 0
good 54.5 6
Total 0.8 11
lug_boot



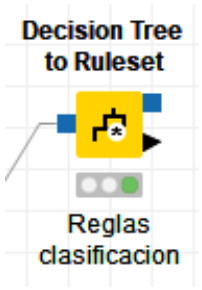




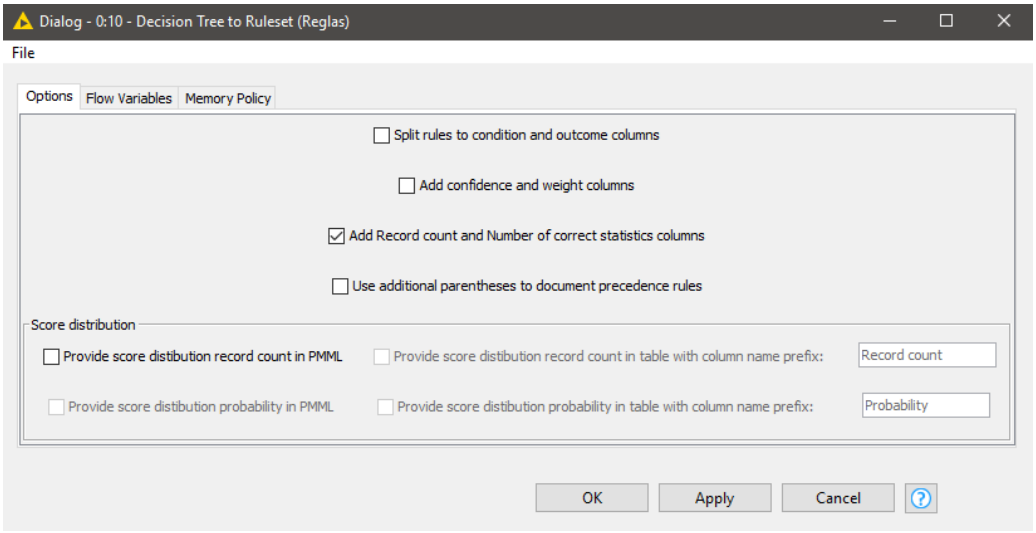


2.5.3. Reglas de clasificación

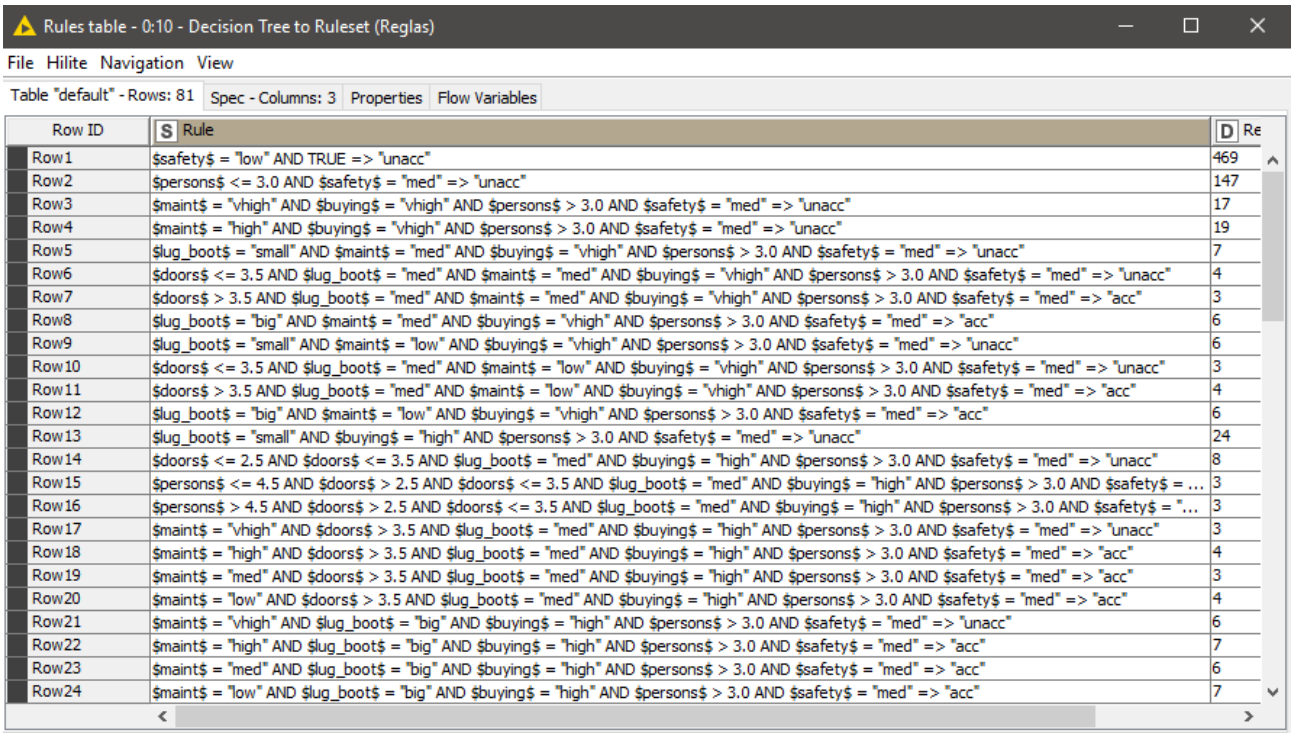
Las reglas de clasificación de un árbol simplemente son la representación en texto del modelo obtenido. Para esto, utilizamos el nodo llamado **Decision Tree To Ruleset**, cuya entrada es la salida PMML del nodo **Decision Tree Learner**.



La pantalla de configuración debe lucir de la siguiente forma, no hay mucho que comentar sobre los campos de la misma:



Para visualizar las reglas del árbol ID3 generado, seleccionamos la opción *Rules table*:


Rules table - 0:10 - Decision Tree to Ruleset (Reglas)

File Hilite Navigation View

Table "default" - Rows: 81 | Spec - Columns: 3 | Properties | Flow Variables

Row ID	S Rule	D Re
Row1	$safety$ = "low" AND TRUE => "unacc"	469
Row2	$persons$ <= 3.0 AND $safety$ = "med" => "unacc"	147
Row3	$maint$ = "vhigh" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	17
Row4	$maint$ = "high" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	19
Row5	lug_boot = "small" AND $maint$ = "med" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	7
Row6	$doors$ <= 3.5 AND lug_boot = "med" AND $maint$ = "med" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	4
Row7	$doors$ > 3.5 AND lug_boot = "med" AND $maint$ = "med" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	3
Row8	lug_boot = "big" AND $maint$ = "med" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	6
Row9	lug_boot = "small" AND $maint$ = "low" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	6
Row10	$doors$ <= 3.5 AND lug_boot = "med" AND $maint$ = "low" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	3
Row11	$doors$ > 3.5 AND lug_boot = "med" AND $maint$ = "low" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	4
Row12	lug_boot = "big" AND $maint$ = "low" AND $buying$ = "vhigh" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	6
Row13	lug_boot = "small" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	24
Row14	$doors$ <= 2.5 AND $doors$ <= 3.5 AND lug_boot = "med" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	8
Row15	$persons$ <= 4.5 AND $doors$ > 2.5 AND $doors$ <= 3.5 AND lug_boot = "med" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = ...	3
Row16	$persons$ > 4.5 AND $doors$ > 2.5 AND $doors$ <= 3.5 AND lug_boot = "med" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "...	3
Row17	$maint$ = "vhigh" AND $doors$ > 3.5 AND lug_boot = "med" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	3
Row18	$maint$ = "high" AND $doors$ > 3.5 AND lug_boot = "med" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	4
Row19	$maint$ = "med" AND $doors$ > 3.5 AND lug_boot = "med" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	3
Row20	$maint$ = "low" AND $doors$ > 3.5 AND lug_boot = "med" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	4
Row21	$maint$ = "vhigh" AND lug_boot = "big" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "unacc"	6
Row22	$maint$ = "high" AND lug_boot = "big" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	7
Row23	$maint$ = "med" AND lug_boot = "big" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	6
Row24	$maint$ = "low" AND lug_boot = "big" AND $buying$ = "high" AND $persons$ > 3.0 AND $safety$ = "med" => "acc"	7

Reglas de clasificación del árbol ID3 generado expresadas en texto

```
"Rule","Record count","Number of correct"
"$safety$ = low AND TRUE => unacc",460,460
"$persons$ = two AND $safety$ = med => unacc",153,153
"$maint$ = vhigh AND $buying$ = vhigh AND $persons$ = four AND $safety$ = med => unacc",8,8
"$maint$ = high AND $buying$ = vhigh AND $persons$ = four AND $safety$ = med => unacc",12,12
"$lug_boot$ = small AND $maint$ = med AND $buying$ = vhigh AND $persons$ = four AND $safety$
 ↪  = med => unacc",3,3
"$lug_boot$ = med AND $maint$ = med AND $buying$ = vhigh AND $persons$ = four AND $safety$ =
 ↪  med => unacc",2,1
"$lug_boot$ = big AND $maint$ = med AND $buying$ = vhigh AND $persons$ = four AND $safety$ =
 ↪  med => acc",1,1
"$lug_boot$ = small AND $maint$ = low AND $buying$ = vhigh AND $persons$ = four AND $safety$
 ↪  = med => unacc",2,2
"$lug_boot$ = med AND $maint$ = low AND $buying$ = vhigh AND $persons$ = four AND $safety$ =
 ↪  med => acc",2,2
"$lug_boot$ = big AND $maint$ = low AND $buying$ = vhigh AND $persons$ = four AND $safety$ =
 ↪  med => acc",3,3
"$lug_boot$ = small AND $buying$ = high AND $persons$ = four AND $safety$ = med =>
 ↪  unacc",14,14
"$doors$ = two AND $lug_boot$ = med AND $buying$ = high AND $persons$ = four AND $safety$ =
 ↪  med => unacc",1,1
"$doors$ = three AND $lug_boot$ = med AND $buying$ = high AND $persons$ = four AND $safety$
 ↪  = med => unacc",2,2
"$doors$ = four AND $lug_boot$ = med AND $buying$ = high AND $persons$ = four AND $safety$ =
 ↪  med => acc",3,2
```

```
"$doors$ = 5more AND $lug_boot$ = med AND $buying$ = high AND $persons$ = four AND $safety$ = med => acc",3,2
"$maint$ = vhigh AND $lug_boot$ = big AND $buying$ = high AND $persons$ = four AND $safety$ = med => unacc",3,3
"$maint$ = high AND $lug_boot$ = big AND $buying$ = high AND $persons$ = four AND $safety$ = med => acc",4,4
"$maint$ = med AND $lug_boot$ = big AND $buying$ = high AND $persons$ = four AND $safety$ = med => acc",4,4
"$maint$ = low AND $lug_boot$ = big AND $buying$ = high AND $persons$ = four AND $safety$ = med => acc",4,4
"$lug_boot$ = small AND $maint$ = vhigh AND $buying$ = med AND $persons$ = four AND $safety$ = med => unacc",3,3
"$lug_boot$ = med AND $maint$ = vhigh AND $buying$ = med AND $persons$ = four AND $safety$ = med => acc",3,2
"$lug_boot$ = big AND $maint$ = vhigh AND $buying$ = med AND $persons$ = four AND $safety$ = med => acc",2,2
"$lug_boot$ = small AND $maint$ = high AND $buying$ = med AND $persons$ = four AND $safety$ = med => unacc",3,3
"$lug_boot$ = med AND $maint$ = high AND $buying$ = med AND $persons$ = four AND $safety$ = med => unacc",4,2
"$lug_boot$ = big AND $maint$ = high AND $buying$ = med AND $persons$ = four AND $safety$ = med => acc",4,4
"$maint$ = med AND $buying$ = med AND $persons$ = four AND $safety$ = med => acc",10,10
"$lug_boot$ = small AND $maint$ = low AND $buying$ = med AND $persons$ = four AND $safety$ = med => acc",2,2
"$lug_boot$ = med AND $maint$ = low AND $buying$ = med AND $persons$ = four AND $safety$ = med => good",3,2
"$lug_boot$ = big AND $maint$ = low AND $buying$ = med AND $persons$ = four AND $safety$ = med => good",3,3
"$lug_boot$ = small AND $maint$ = vhigh AND $buying$ = low AND $persons$ = four AND $safety$ = med => unacc",3,3
"$lug_boot$ = med AND $maint$ = vhigh AND $buying$ = low AND $persons$ = four AND $safety$ = med => acc",3,2
"$lug_boot$ = big AND $maint$ = vhigh AND $buying$ = low AND $persons$ = four AND $safety$ = med => acc",3,3
"$maint$ = high AND $buying$ = low AND $persons$ = four AND $safety$ = med => acc",8,8
"$lug_boot$ = small AND $maint$ = med AND $buying$ = low AND $persons$ = four AND $safety$ = med => acc",3,3
"$lug_boot$ = med AND $maint$ = med AND $buying$ = low AND $persons$ = four AND $safety$ = med => acc",3,2
"$lug_boot$ = big AND $maint$ = med AND $buying$ = low AND $persons$ = four AND $safety$ = med => good",4,4
"$lug_boot$ = small AND $maint$ = low AND $buying$ = low AND $persons$ = four AND $safety$ = med => acc",4,4
"$lug_boot$ = med AND $maint$ = low AND $buying$ = low AND $persons$ = four AND $safety$ = med => good",3,2
"$lug_boot$ = big AND $maint$ = low AND $buying$ = low AND $persons$ = four AND $safety$ = med => good",2,2
"$maint$ = vhigh AND $buying$ = vhigh AND $persons$ = more AND $safety$ = med => unacc",8,8
"$maint$ = high AND $buying$ = vhigh AND $persons$ = more AND $safety$ = med => unacc",10,10
"$lug_boot$ = small AND $maint$ = med AND $buying$ = vhigh AND $persons$ = more AND $safety$ = med => unacc",3,3
```

```
"$lug_boot$ = med AND $maint$ = med AND $buying$ = vhigh AND $persons$ = more AND $safety$ =
 ↪  med => acc",4,3
"$lug_boot$ = big AND $maint$ = med AND $buying$ = vhigh AND $persons$ = more AND $safety$ =
 ↪  med => acc",2,2
"$lug_boot$ = small AND $maint$ = low AND $buying$ = vhigh AND $persons$ = more AND $safety$
 ↪  = med => unacc",3,3
"$lug_boot$ = med AND $maint$ = low AND $buying$ = vhigh AND $persons$ = more AND $safety$ =
 ↪  med => acc",4,3
"$lug_boot$ = big AND $maint$ = low AND $buying$ = vhigh AND $persons$ = more AND $safety$ =
 ↪  med => acc",2,2
"$lug_boot$ = small AND $buying$ = high AND $persons$ = more AND $safety$ = med =>
 ↪  unacc",14,14
"$doors$ = two AND $lug_boot$ = med AND $buying$ = high AND $persons$ = more AND $safety$ =
 ↪  med => unacc",4,4
"$doors$ = three AND $lug_boot$ = med AND $buying$ = high AND $persons$ = more AND $safety$
 ↪  = med => acc",2,2
"$doors$ = four AND $lug_boot$ = med AND $buying$ = high AND $persons$ = more AND $safety$ =
 ↪  med => acc",3,2
"$doors$ = 5more AND $lug_boot$ = med AND $buying$ = high AND $persons$ = more AND $safety$
 ↪  = med => acc",4,3
"$maint$ = vhigh AND $lug_boot$ = big AND $buying$ = high AND $persons$ = more AND $safety$
 ↪  = med => unacc",3,3
"$maint$ = high AND $lug_boot$ = big AND $buying$ = high AND $persons$ = more AND $safety$ =
 ↪  med => acc",4,4
"$maint$ = med AND $lug_boot$ = big AND $buying$ = high AND $persons$ = more AND $safety$ =
 ↪  med => acc",4,4
"$maint$ = low AND $lug_boot$ = big AND $buying$ = high AND $persons$ = more AND $safety$ =
 ↪  med => acc",2,2
"$maint$ = vhigh AND $lug_boot$ = small AND $buying$ = med AND $persons$ = more AND $safety$
 ↪  = med => unacc",4,4
"$maint$ = high AND $lug_boot$ = small AND $buying$ = med AND $persons$ = more AND $safety$
 ↪  = med => unacc",3,3
"$maint$ = med AND $lug_boot$ = small AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => unacc",2,1
"$maint$ = low AND $lug_boot$ = small AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => acc",4,3
"$maint$ = vhigh AND $lug_boot$ = med AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => acc",3,2
"$maint$ = high AND $lug_boot$ = med AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => acc",4,3
"$maint$ = med AND $lug_boot$ = med AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => acc",3,3
"$maint$ = low AND $lug_boot$ = med AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => good",3,2
"$maint$ = vhigh AND $lug_boot$ = big AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => acc",4,4
"$maint$ = high AND $lug_boot$ = big AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => acc",2,2
"$maint$ = med AND $lug_boot$ = big AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => acc",4,4
"$maint$ = low AND $lug_boot$ = big AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  med => good",3,3
```

```
"$lug_boot$ = small AND $maint$ = vhigh AND $buying$ = low AND $persons$ = more AND $safety$ = med => unacc",1,1
"$lug_boot$ = med AND $maint$ = vhigh AND $buying$ = low AND $persons$ = more AND $safety$ = med => acc",3,2
"$lug_boot$ = big AND $maint$ = vhigh AND $buying$ = low AND $persons$ = more AND $safety$ = med => acc",4,4
"$maint$ = high AND $buying$ = low AND $persons$ = more AND $safety$ = med => acc",9,9
"$lug_boot$ = small AND $maint$ = med AND $buying$ = low AND $persons$ = more AND $safety$ = med => acc",3,2
"$lug_boot$ = med AND $maint$ = med AND $buying$ = low AND $persons$ = more AND $safety$ = med => good",1,1
"$lug_boot$ = big AND $maint$ = med AND $buying$ = low AND $persons$ = more AND $safety$ = med => good",3,3
"$lug_boot$ = small AND $maint$ = low AND $buying$ = low AND $persons$ = more AND $safety$ = med => acc",4,3
"$lug_boot$ = med AND $maint$ = low AND $buying$ = low AND $persons$ = more AND $safety$ = med => good",3,2
"$lug_boot$ = big AND $maint$ = low AND $buying$ = low AND $persons$ = more AND $safety$ = med => good",4,4
"$persons$ = two AND $safety$ = high => unacc",156,156
"$maint$ = vhigh AND $buying$ = vhigh AND $persons$ = four AND $safety$ = high => unacc",10,10
"$maint$ = high AND $buying$ = vhigh AND $persons$ = four AND $safety$ = high => unacc",10,10
"$maint$ = med AND $buying$ = vhigh AND $persons$ = four AND $safety$ = high => acc",11,11
"$maint$ = low AND $buying$ = vhigh AND $persons$ = four AND $safety$ = high => acc",11,11
"$maint$ = vhigh AND $buying$ = high AND $persons$ = four AND $safety$ = high => unacc",11,11
"$maint$ = high AND $buying$ = high AND $persons$ = four AND $safety$ = high => acc",10,10
"$maint$ = med AND $buying$ = high AND $persons$ = four AND $safety$ = high => acc",11,11
"$maint$ = low AND $buying$ = high AND $persons$ = four AND $safety$ = high => acc",10,10
"$maint$ = vhigh AND $buying$ = med AND $persons$ = four AND $safety$ = high => acc",10,10
"$maint$ = high AND $buying$ = med AND $persons$ = four AND $safety$ = high => acc",11,11
"$lug_boot$ = small AND $maint$ = med AND $buying$ = med AND $persons$ = four AND $safety$ = high => acc",4,4
"$lug_boot$ = med AND $maint$ = med AND $buying$ = med AND $persons$ = four AND $safety$ = high => acc",4,2
"$lug_boot$ = big AND $maint$ = med AND $buying$ = med AND $persons$ = four AND $safety$ = high => vgood",4,4
"$lug_boot$ = small AND $maint$ = low AND $buying$ = med AND $persons$ = four AND $safety$ = high => good",3,3
"$lug_boot$ = med AND $maint$ = low AND $buying$ = med AND $persons$ = four AND $safety$ = high => good",3,2
"$lug_boot$ = big AND $maint$ = low AND $buying$ = med AND $persons$ = four AND $safety$ = high => vgood",4,4
"$maint$ = vhigh AND $buying$ = low AND $persons$ = four AND $safety$ = high => acc",5,5
"$lug_boot$ = small AND $maint$ = high AND $buying$ = low AND $persons$ = four AND $safety$ = high => acc",4,4
"$lug_boot$ = med AND $maint$ = high AND $buying$ = low AND $persons$ = four AND $safety$ = high => acc",4,2
"$lug_boot$ = big AND $maint$ = high AND $buying$ = low AND $persons$ = four AND $safety$ = high => vgood",3,3
```

```
"$lug_boot$ = small AND $maint$ = med AND $buying$ = low AND $persons$ = four AND $safety$ =
 ↪  high => good",3,3
"$lug_boot$ = med AND $maint$ = med AND $buying$ = low AND $persons$ = four AND $safety$ =
 ↪  high => vgood",4,2
"$lug_boot$ = big AND $maint$ = med AND $buying$ = low AND $persons$ = four AND $safety$ =
 ↪  high => vgood",3,3
"$lug_boot$ = small AND $maint$ = low AND $buying$ = low AND $persons$ = four AND $safety$ =
 ↪  high => good",4,4
"$lug_boot$ = med AND $maint$ = low AND $buying$ = low AND $persons$ = four AND $safety$ =
 ↪  high => good",3,2
"$lug_boot$ = big AND $maint$ = low AND $buying$ = low AND $persons$ = four AND $safety$ =
 ↪  high => vgood",3,3
"$maint$ = vhigh AND $buying$ = vhigh AND $persons$ = more AND $safety$ = high =>
 ↪  unacc",11,11
"$maint$ = high AND $buying$ = vhigh AND $persons$ = more AND $safety$ = high =>
 ↪  unacc",11,11
"$maint$ = med AND $buying$ = vhigh AND $persons$ = more AND $safety$ = high => acc",8,7
"$maint$ = low AND $buying$ = vhigh AND $persons$ = more AND $safety$ = high => acc",9,8
"$maint$ = vhigh AND $buying$ = high AND $persons$ = more AND $safety$ = high =>
 ↪  unacc",10,10
"$maint$ = high AND $buying$ = high AND $persons$ = more AND $safety$ = high => acc",8,8
"$maint$ = med AND $buying$ = high AND $persons$ = more AND $safety$ = high => acc",11,10
"$maint$ = low AND $buying$ = high AND $persons$ = more AND $safety$ = high => acc",9,8
"$maint$ = vhigh AND $buying$ = med AND $persons$ = more AND $safety$ = high => acc",11,10
"$maint$ = high AND $buying$ = med AND $persons$ = more AND $safety$ = high => acc",9,8
"$lug_boot$ = small AND $maint$ = med AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  high => acc",4,3
"$lug_boot$ = med AND $maint$ = med AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  high => vgood",3,2
"$lug_boot$ = big AND $maint$ = med AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  high => vgood",3,3
"$lug_boot$ = small AND $maint$ = low AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  high => good",4,3
"$lug_boot$ = med AND $maint$ = low AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  high => vgood",3,2
"$lug_boot$ = big AND $maint$ = low AND $buying$ = med AND $persons$ = more AND $safety$ =
 ↪  high => vgood",4,4
"$maint$ = vhigh AND $buying$ = low AND $persons$ = more AND $safety$ = high => acc",10,9
"$lug_boot$ = small AND $maint$ = high AND $buying$ = low AND $persons$ = more AND $safety$
 ↪  = high => acc",4,3
"$lug_boot$ = med AND $maint$ = high AND $buying$ = low AND $persons$ = more AND $safety$ =
 ↪  high => vgood",3,2
"$lug_boot$ = big AND $maint$ = high AND $buying$ = low AND $persons$ = more AND $safety$ =
 ↪  high => vgood",3,3
"$lug_boot$ = small AND $maint$ = med AND $buying$ = low AND $persons$ = more AND $safety$ =
 ↪  high => good",4,3
"$lug_boot$ = med AND $maint$ = med AND $buying$ = low AND $persons$ = more AND $safety$ =
 ↪  high => vgood",2,2
"$lug_boot$ = big AND $maint$ = med AND $buying$ = low AND $persons$ = more AND $safety$ =
 ↪  high => vgood",3,3
"$lug_boot$ = small AND $maint$ = low AND $buying$ = low AND $persons$ = more AND $safety$ =
 ↪  high => good",2,2
```

```
131 "$lug_boot$ = med AND $maint$ = low AND $buying$ = low AND $persons$ = more AND $safety$ =
    ↪  high => vgood",3,2
132 "$lug_boot$ = big AND $maint$ = low AND $buying$ = low AND $persons$ = more AND $safety$ =
    ↪  high => vgood",4,4
```

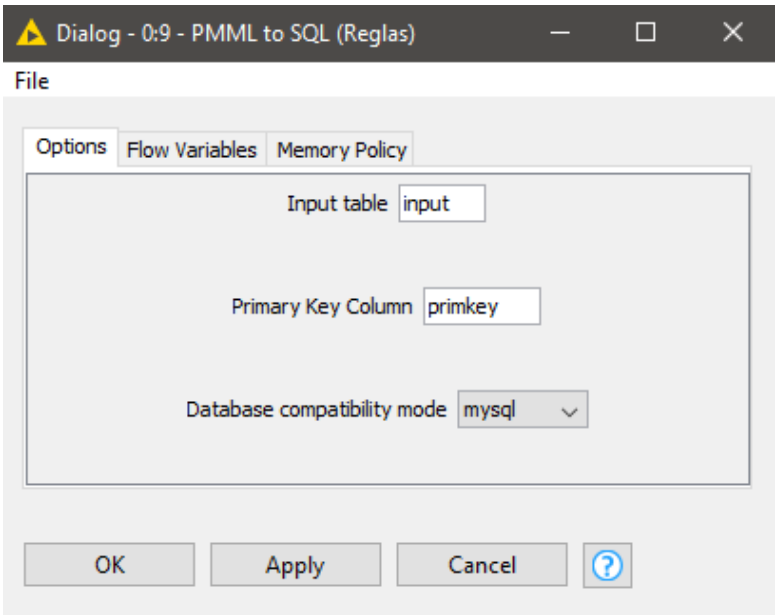
2.5.4. Importar reglas de clasificación a sentencia SQL

Para obtener una sentencia SQL a partir del modelo representado por las reglas de clasificación del árbol de decisión ID3 generado, utilizamos el nodo **PMML To SQL**. A su entrada va conectada la salida PMML del nodo **Decision Tree Learner**.

Esto es de gran ayuda cuando se tenga la base de datos cargada en un gestor de base de datos, como MySQL, y queramos hacer consultas para obtener registros que cumplan una serie de condiciones, logrando clasificándolos en el proceso.

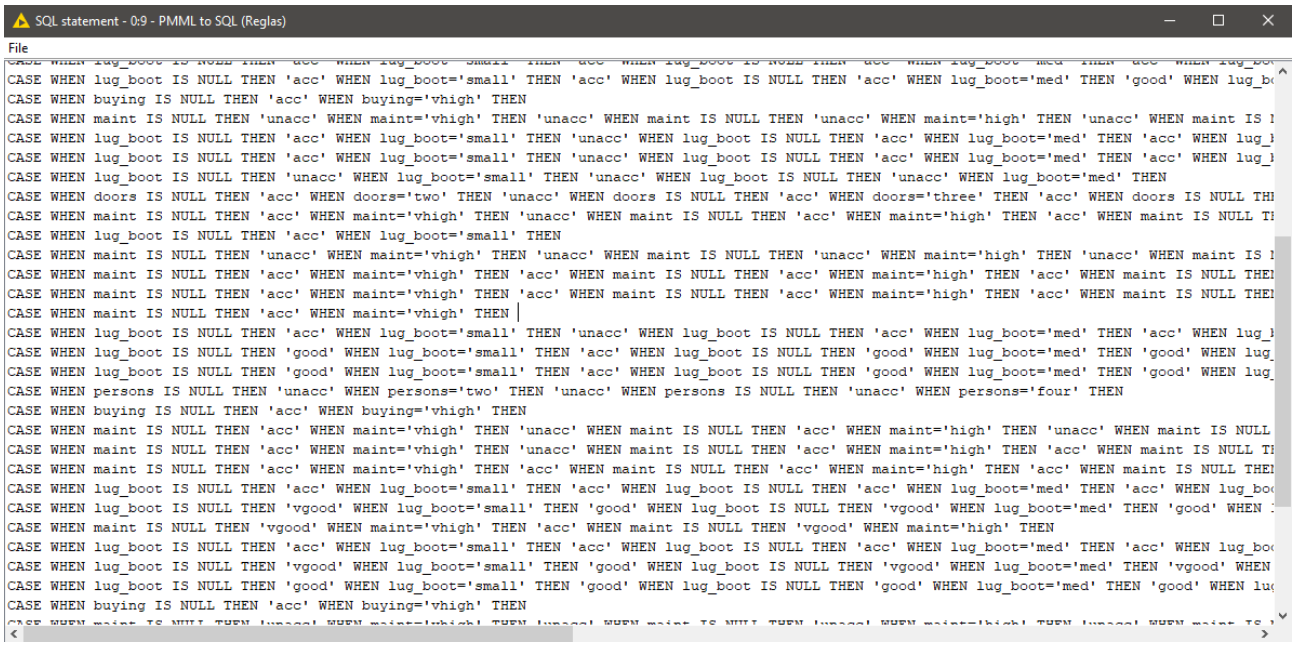
NOTA: Se debe instalar la extensión KNIME PMML Translations para hacer uso de este nodo.

La ventana de configuración queda de la siguiente forma:



En este caso, le indicamos que la sentencia SQL se quiere utilizar en el manejador MySQL.

Para visualizar la sentencia SQL generada, seleccionamos la opción *View: SQL Statement*:



Sentencia SQL obtenida

```
SELECT primkey, (
CASE WHEN safety IS NULL THEN 'unacc' WHEN safety='low' THEN 'unacc' WHEN safety IS NULL THEN 'unacc' WHEN safety='med' THEN
CASE WHEN persons IS NULL THEN 'unacc' WHEN persons='two' THEN 'unacc' WHEN persons IS NULL THEN 'unacc' WHEN persons='four' THEN
CASE WHEN buying IS NULL THEN 'acc' WHEN buying='vhigh' THEN
CASE WHEN maint IS NULL THEN 'unacc' WHEN maint='vhigh' THEN 'unacc' WHEN maint IS NULL THEN 'unacc' WHEN maint='high' THEN 'unacc' WHEN maint='med' THEN
CASE WHEN lug_boot IS NULL THEN 'unacc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS NULL THEN 'unacc' WHEN lug_boot='med' THEN 'unacc' WHEN lug_boot='big' THEN 'acc' ELSE NULL END WHEN maint IS NULL THEN 'unacc' WHEN maint='low' THEN
CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='med' THEN 'acc' WHEN lug_boot='big' THEN 'acc' ELSE NULL END ELSE NULL END WHEN buying IS NULL THEN 'acc' WHEN buying='high' THEN
CASE WHEN lug_boot IS NULL THEN 'unacc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS NULL THEN 'unacc' WHEN lug_boot='med' THEN
CASE WHEN doors IS NULL THEN 'unacc' WHEN doors='two' THEN 'unacc' WHEN doors IS NULL THEN 'unacc' WHEN doors='three' THEN 'unacc' WHEN doors='four' THEN 'acc' WHEN doors IS NULL THEN 'unacc' WHEN doors='5more' THEN 'acc' ELSE NULL END WHEN lug_boot IS NULL THEN 'unacc' WHEN lug_boot='big' THEN
CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN 'unacc' WHEN maint IS NULL THEN 'acc' WHEN maint='high' THEN 'acc' WHEN maint='med' THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='low' THEN 'acc' ELSE NULL END ELSE NULL END WHEN buying IS NULL THEN 'acc' WHEN buying='med' THEN
```

```
11 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN
12 CASE WHEN lug_boot IS NULL THEN 'unacc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS
    ↪ NULL THEN 'unacc' WHEN lug_boot='med' THEN 'acc' WHEN lug_boot IS NULL THEN 'unacc'
    ↪ WHEN lug_boot='big' THEN 'acc' ELSE NULL END WHEN maint IS NULL THEN 'acc' WHEN
    ↪ maint='high' THEN
13 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS
    ↪ NULL THEN 'acc' WHEN lug_boot='med' THEN 'unacc' WHEN lug_boot IS NULL THEN 'acc' WHEN
    ↪ lug_boot='big' THEN 'acc' ELSE NULL END WHEN maint IS NULL THEN 'acc' WHEN maint='med'
    ↪ THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='low' THEN
14 CASE WHEN lug_boot IS NULL THEN 'good' WHEN lug_boot='small' THEN 'acc' WHEN lug_boot IS
    ↪ NULL THEN 'good' WHEN lug_boot='med' THEN 'good' WHEN lug_boot IS NULL THEN 'good' WHEN
    ↪ lug_boot='big' THEN 'good' ELSE NULL END ELSE NULL END WHEN buying IS NULL THEN 'acc'
    ↪ WHEN buying='low' THEN
15 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN
16 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS
    ↪ NULL THEN 'acc' WHEN lug_boot='med' THEN 'acc' WHEN lug_boot IS NULL THEN 'acc' WHEN
    ↪ lug_boot='big' THEN 'acc' ELSE NULL END WHEN maint IS NULL THEN 'acc' WHEN maint='high'
    ↪ THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='med' THEN
17 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'acc' WHEN lug_boot IS NULL
    ↪ THEN 'acc' WHEN lug_boot='med' THEN 'acc' WHEN lug_boot IS NULL THEN 'acc' WHEN
    ↪ lug_boot='big' THEN 'good' ELSE NULL END WHEN maint IS NULL THEN 'acc' WHEN maint='low'
    ↪ THEN
18 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'acc' WHEN lug_boot IS NULL
    ↪ THEN 'acc' WHEN lug_boot='med' THEN 'good' WHEN lug_boot IS NULL THEN 'acc' WHEN
    ↪ lug_boot='big' THEN 'good' ELSE NULL END ELSE NULL END ELSE NULL END WHEN persons IS
    ↪ NULL THEN 'unacc' WHEN persons='more' THEN
19 CASE WHEN buying IS NULL THEN 'acc' WHEN buying='vhigh' THEN
20 CASE WHEN maint IS NULL THEN 'unacc' WHEN maint='vhigh' THEN 'unacc' WHEN maint IS NULL THEN
    ↪ 'unacc' WHEN maint='high' THEN 'unacc' WHEN maint IS NULL THEN 'unacc' WHEN maint='med'
    ↪ THEN
21 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS
    ↪ NULL THEN 'acc' WHEN lug_boot='med' THEN 'acc' WHEN lug_boot IS NULL THEN 'acc' WHEN
    ↪ lug_boot='big' THEN 'acc' ELSE NULL END WHEN maint IS NULL THEN 'unacc' WHEN
    ↪ maint='low' THEN
22 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS
    ↪ NULL THEN 'acc' WHEN lug_boot='med' THEN 'acc' WHEN lug_boot IS NULL THEN 'acc' WHEN
    ↪ lug_boot='big' THEN 'acc' ELSE NULL END ELSE NULL END WHEN buying IS NULL THEN 'acc'
    ↪ WHEN buying='high' THEN
23 CASE WHEN lug_boot IS NULL THEN 'unacc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS
    ↪ NULL THEN 'unacc' WHEN lug_boot='med' THEN
24 CASE WHEN doors IS NULL THEN 'acc' WHEN doors='two' THEN 'unacc' WHEN doors IS NULL THEN
    ↪ 'acc' WHEN doors='three' THEN 'acc' WHEN doors IS NULL THEN 'acc' WHEN doors='four'
    ↪ THEN 'acc' WHEN doors IS NULL THEN 'acc' WHEN doors='5more' THEN 'acc' ELSE NULL END
    ↪ WHEN lug_boot IS NULL THEN 'unacc' WHEN lug_boot='big' THEN
25 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN 'unacc' WHEN maint IS NULL THEN
    ↪ 'acc' WHEN maint='high' THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='med' THEN
    ↪ 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='low' THEN 'acc' ELSE NULL END ELSE NULL
    ↪ END WHEN buying IS NULL THEN 'acc' WHEN buying='med' THEN
26 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN
27 CASE WHEN maint IS NULL THEN 'unacc' WHEN maint='vhigh' THEN 'unacc' WHEN maint IS NULL THEN
    ↪ 'unacc' WHEN maint='high' THEN 'unacc' WHEN maint IS NULL THEN 'unacc' WHEN maint='med'
    ↪ THEN 'unacc' WHEN maint IS NULL THEN 'unacc' WHEN maint='low' THEN 'acc' ELSE NULL END
    ↪ WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='med' THEN
```

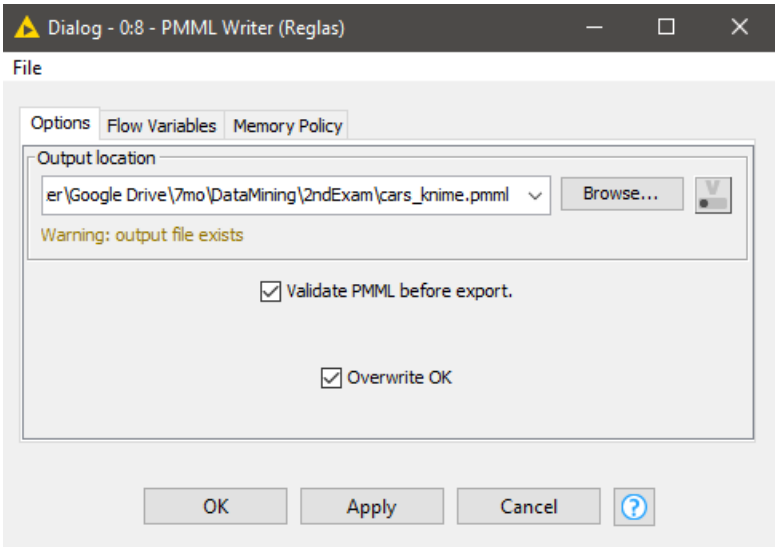
```
28 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN 'acc' WHEN maint IS NULL THEN
    ↪ 'acc' WHEN maint='high' THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='med' THEN
    ↪ 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='low' THEN 'good' ELSE NULL END WHEN
    ↪ lug_boot IS NULL THEN 'acc' WHEN lug_boot='big' THEN
29 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN 'acc' WHEN maint IS NULL THEN
    ↪ 'acc' WHEN maint='high' THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='med' THEN
    ↪ 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='low' THEN 'good' ELSE NULL END ELSE
    ↪ NULL END WHEN buying IS NULL THEN 'acc' WHEN buying='low' THEN
30 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN
31 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'unacc' WHEN lug_boot IS
    ↪ NULL THEN 'acc' WHEN lug_boot='med' THEN 'acc' WHEN lug_boot IS NULL THEN 'acc' WHEN
    ↪ lug_boot='big' THEN 'acc' ELSE NULL END WHEN maint IS NULL THEN 'acc' WHEN maint='high'
    ↪ THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='med' THEN
32 CASE WHEN lug_boot IS NULL THEN 'good' WHEN lug_boot='small' THEN 'acc' WHEN lug_boot IS
    ↪ NULL THEN 'good' WHEN lug_boot='med' THEN 'good' WHEN lug_boot IS NULL THEN 'good' WHEN
    ↪ lug_boot='big' THEN 'good' ELSE NULL END WHEN maint IS NULL THEN 'acc' WHEN maint='low'
    ↪ THEN
33 CASE WHEN lug_boot IS NULL THEN 'good' WHEN lug_boot='small' THEN 'acc' WHEN lug_boot IS
    ↪ NULL THEN 'good' WHEN lug_boot='med' THEN 'good' WHEN lug_boot IS NULL THEN 'good' WHEN
    ↪ lug_boot='big' THEN 'good' ELSE NULL END ELSE NULL END ELSE NULL END ELSE NULL END WHEN
    ↪ safety IS NULL THEN 'unacc' WHEN safety='high' THEN
34 CASE WHEN persons IS NULL THEN 'unacc' WHEN persons='two' THEN 'unacc' WHEN persons IS NULL
    ↪ THEN 'unacc' WHEN persons='four' THEN
35 CASE WHEN buying IS NULL THEN 'acc' WHEN buying='vhigh' THEN
36 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN 'unacc' WHEN maint IS NULL THEN
    ↪ 'acc' WHEN maint='high' THEN 'unacc' WHEN maint IS NULL THEN 'acc' WHEN maint='med'
    ↪ THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='low' THEN 'acc' ELSE NULL END WHEN
    ↪ buying IS NULL THEN 'acc' WHEN buying='high' THEN
37 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN 'unacc' WHEN maint IS NULL THEN
    ↪ 'acc' WHEN maint='high' THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='med' THEN
    ↪ 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='low' THEN 'acc' ELSE NULL END WHEN
    ↪ buying IS NULL THEN 'acc' WHEN buying='med' THEN
38 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN 'acc' WHEN maint IS NULL THEN
    ↪ 'acc' WHEN maint='high' THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='med' THEN
39 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'acc' WHEN lug_boot IS NULL
    ↪ THEN 'acc' WHEN lug_boot='med' THEN 'acc' WHEN lug_boot IS NULL THEN 'acc' WHEN
    ↪ lug_boot='big' THEN 'vgood' ELSE NULL END WHEN maint IS NULL THEN 'acc' WHEN
    ↪ maint='low' THEN
40 CASE WHEN lug_boot IS NULL THEN 'vgood' WHEN lug_boot='small' THEN 'good' WHEN lug_boot IS
    ↪ NULL THEN 'vgood' WHEN lug_boot='med' THEN 'good' WHEN lug_boot IS NULL THEN 'vgood'
    ↪ WHEN lug_boot='big' THEN 'vgood' ELSE NULL END ELSE NULL END WHEN buying IS NULL THEN
    ↪ 'acc' WHEN buying='low' THEN
41 CASE WHEN maint IS NULL THEN 'vgood' WHEN maint='vhigh' THEN 'acc' WHEN maint IS NULL THEN
    ↪ 'vgood' WHEN maint='high' THEN
42 CASE WHEN lug_boot IS NULL THEN 'acc' WHEN lug_boot='small' THEN 'acc' WHEN lug_boot IS NULL
    ↪ THEN 'acc' WHEN lug_boot='med' THEN 'acc' WHEN lug_boot IS NULL THEN 'acc' WHEN
    ↪ lug_boot='big' THEN 'vgood' ELSE NULL END WHEN maint IS NULL THEN 'vgood' WHEN
    ↪ maint='med' THEN
43 CASE WHEN lug_boot IS NULL THEN 'vgood' WHEN lug_boot='small' THEN 'good' WHEN lug_boot IS
    ↪ NULL THEN 'vgood' WHEN lug_boot='med' THEN 'vgood' WHEN lug_boot IS NULL THEN 'vgood'
    ↪ WHEN lug_boot='big' THEN 'vgood' ELSE NULL END WHEN maint IS NULL THEN 'vgood' WHEN
    ↪ maint='low' THEN
```

```
44 CASE WHEN lug_boot IS NULL THEN 'good' WHEN lug_boot='small' THEN 'good' WHEN lug_boot IS
    ↪ NULL THEN 'good' WHEN lug_boot='med' THEN 'good' WHEN lug_boot IS NULL THEN 'good' WHEN
    ↪ lug_boot='big' THEN 'vgood' ELSE NULL END ELSE NULL END ELSE NULL END WHEN persons IS
    ↪ NULL THEN 'unacc' WHEN persons='more' THEN
45 CASE WHEN buying IS NULL THEN 'acc' WHEN buying='vhigh' THEN
46 CASE WHEN maint IS NULL THEN 'unacc' WHEN maint='vhigh' THEN 'unacc' WHEN maint IS NULL THEN
    ↪ 'unacc' WHEN maint='high' THEN 'unacc' WHEN maint IS NULL THEN 'unacc' WHEN maint='med'
    ↪ THEN 'acc' WHEN maint IS NULL THEN 'unacc' WHEN maint='low' THEN 'acc' ELSE NULL END
    ↪ WHEN buying IS NULL THEN 'acc' WHEN buying='high' THEN
47 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN 'unacc' WHEN maint IS NULL THEN
    ↪ 'acc' WHEN maint='high' THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='med' THEN
    ↪ 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='low' THEN 'acc' ELSE NULL END WHEN
    ↪ buying IS NULL THEN 'acc' WHEN buying='med' THEN
48 CASE WHEN maint IS NULL THEN 'acc' WHEN maint='vhigh' THEN 'acc' WHEN maint IS NULL THEN
    ↪ 'acc' WHEN maint='high' THEN 'acc' WHEN maint IS NULL THEN 'acc' WHEN maint='med' THEN
49 CASE WHEN lug_boot IS NULL THEN 'vgood' WHEN lug_boot='small' THEN 'acc' WHEN lug_boot IS
    ↪ NULL THEN 'vgood' WHEN lug_boot='med' THEN 'vgood' WHEN lug_boot IS NULL THEN 'vgood'
    ↪ WHEN lug_boot='big' THEN 'vgood' ELSE NULL END WHEN maint IS NULL THEN 'acc' WHEN
    ↪ maint='low' THEN
50 CASE WHEN lug_boot IS NULL THEN 'vgood' WHEN lug_boot='small' THEN 'good' WHEN lug_boot IS
    ↪ NULL THEN 'vgood' WHEN lug_boot='med' THEN 'vgood' WHEN lug_boot IS NULL THEN 'vgood'
    ↪ WHEN lug_boot='big' THEN 'vgood' ELSE NULL END ELSE NULL END WHEN buying IS NULL THEN
    ↪ 'acc' WHEN buying='low' THEN
51 CASE WHEN maint IS NULL THEN 'vgood' WHEN maint='vhigh' THEN 'acc' WHEN maint IS NULL THEN
    ↪ 'vgood' WHEN maint='high' THEN
52 CASE WHEN lug_boot IS NULL THEN 'vgood' WHEN lug_boot='small' THEN 'acc' WHEN lug_boot IS
    ↪ NULL THEN 'vgood' WHEN lug_boot='med' THEN 'vgood' WHEN lug_boot IS NULL THEN 'vgood'
    ↪ WHEN lug_boot='big' THEN 'vgood' ELSE NULL END WHEN maint IS NULL THEN 'vgood' WHEN
    ↪ maint='med' THEN
53 CASE WHEN lug_boot IS NULL THEN 'vgood' WHEN lug_boot='small' THEN 'good' WHEN lug_boot IS
    ↪ NULL THEN 'vgood' WHEN lug_boot='med' THEN 'vgood' WHEN lug_boot IS NULL THEN 'vgood'
    ↪ WHEN lug_boot='big' THEN 'vgood' ELSE NULL END WHEN maint IS NULL THEN 'vgood' WHEN
    ↪ maint='low' THEN
54 CASE WHEN lug_boot IS NULL THEN 'vgood' WHEN lug_boot='small' THEN 'good' WHEN lug_boot IS
    ↪ NULL THEN 'vgood' WHEN lug_boot='med' THEN 'vgood' WHEN lug_boot IS NULL THEN 'vgood'
    ↪ WHEN lug_boot='big' THEN 'vgood' ELSE NULL END ELSE NULL END ELSE NULL END ELSE NULL
    ↪ END ELSE NULL END) AS class
55 FROM (SELECT primkey, buying AS buying, maint AS maint, doors AS doors, persons AS persons,
    ↪ lug_boot AS lug_boot, safety AS safety
56 FROM input AS ms_input) AS model_input
```

2.5.5. Importar reglas de clasificación a PMML

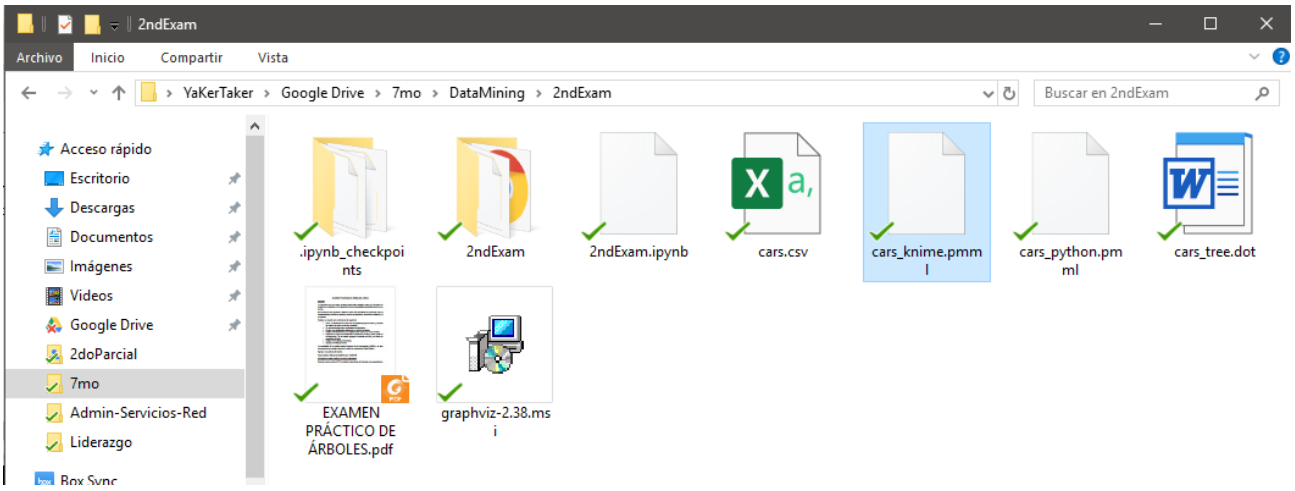
Para guardar el modelo del árbol ID3 generado representado por sus reglas de decisión en un archivo PMML, utilizamos el nodo **PMML Writer**, cuya entrada es la salida PMML del nodo **Decision Tree Learner**.

La ventana de configuración queda de la siguiente forma:



Le indicamos que valide el PMML antes de exportar, y que sobreescriba el archivo en caso de existir.

Al dirigirnos a la ruta indicada, observaremos un archivo con extensión .pmml



Cuando se requiera hacer una clasificación con otros datos de la base de datos, este archivo PMML es útil para no repetir todo el proceso para la creación del modelo de clasificación (árbol de decisión ID3) descrito en este reporte desde el principio.

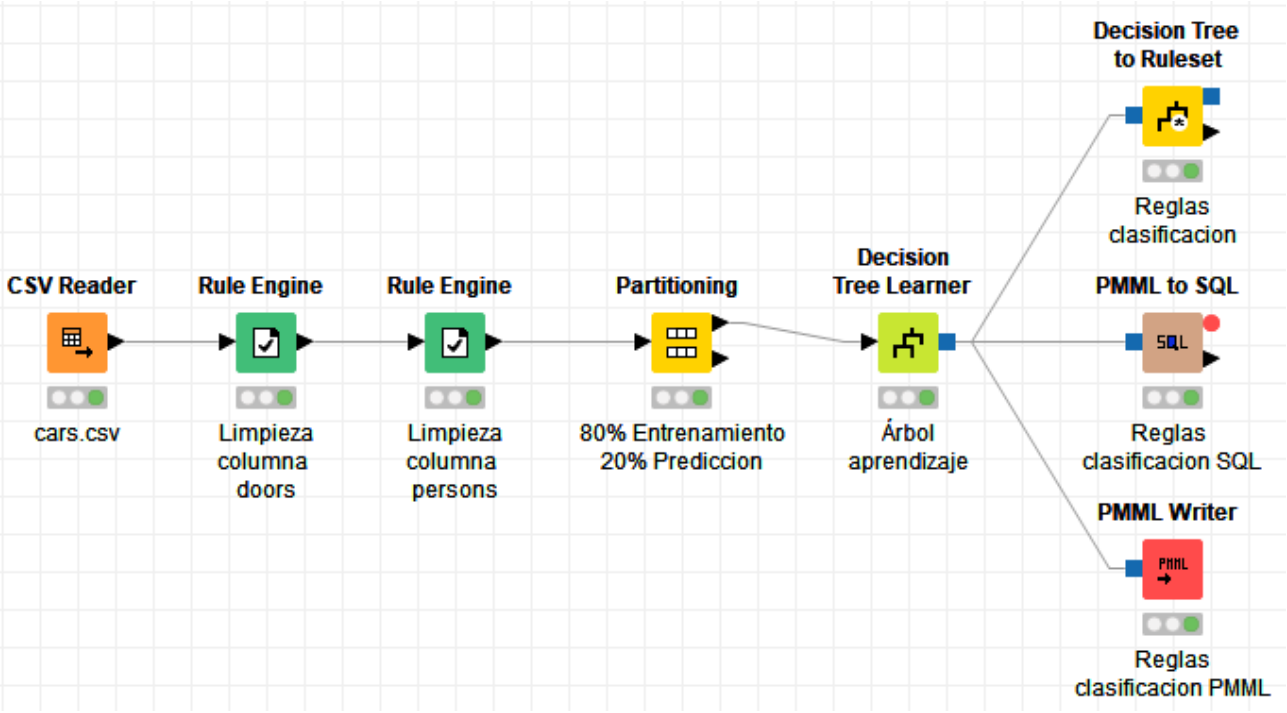
3. Resultados encontrados

El atributo más significativo de toda la base de datos para determinar el estado de un automóvil es su **seguridad**, ya que si esta es baja o media, casi siempre el auto se clasificará en un estado inaceptable.

Por otro lado, el **precio de venta** y el **costo de mantenimiento** también juegan un papel importante. Si el precio de venta es bajo, pero el costo de mantenimiento es alto, el auto cae en clasificaciones de tipo aceptable o bueno, es decir, un estado intermedio que no es ni bueno ni malo. Si ambos atributos tienen costos muy altos, se clasificará como un auto no aceptable, a menos que tenga un **maletero** de tamaño grande, un **número de puertas** mayor a 4, y/o una **capacidad de personas** de más de 5; si la **seguridad** esta en un rango medio alto, se puede clasificar incluso en un buen estado, a pesar de los altos costos.

Los mejores autos son los que tengan un **costo de mantenimiento** y **precio de venta** medios, un **maletero** de gran tamaño, **capacidad para más de 4 personas**, y una **seguridad** alta.

4. Diagrama final KNIME



5. Bibliografía

- H. Sahu, S. Shrma y S. Gondhalakar, *A Brief Overview on Data Mining Survey*, International Journal of Computer Technology and Electronics Engineering (IJCTEE), vol. 1, no. 3, pp. 114 - 115, 2011
- NodePit for KNIME. [Online] Disponible en: https://nodepit.com/nodepit-for-knime