

Practica 12

1. RESPONDA LAS SIGUIENTES PREGUNTAS:

a) LISTE ALGUNAS DE LAS VENTAJAS QUE OFRECE UN ÁRBOL DE DECISIÓN Y QUÉ LOS HACE TAN POPULARES CON RESPECTO A OTROS MÉTODOS DE PREDICCIÓN.

- Los árboles de decisión son muy populares ya que para su construcción no se requiere ningún conocimiento de dominio o establecimiento de parámetros, por lo que son recomendados para hacer un descubrimiento de conocimiento.

Ventajas

- Debido a su forma de construcción son fáciles de asimilar.
- Los pasos de aprendizaje y clasificación son simples y rápidos, además de que en general tienen buena exactitud.
- Su éxito depende de los datos sobre los que se aplique.
- Son utilizados en varias áreas de aplicación: medicina, manufactura y producción, análisis financiero, astronomía, biología molecular, etc.

b) EN UN ÁRBOL DE DECISIÓN ¿UN ATRIBUTO PUEDE SER REVISADO EN MÁS DE UNA OCASIÓN?, ¿EXISTEN ATRIBUTOS QUE, POR EL CONTRARIO, SEAN REVISADOS UNA SOLA VEZ? (JUSTIFIQUE).

- En un árbol de decisión un atributo puede ser revisado más de una vez para determinar el criterio de partición que mejor divida un conjunto de datos

c) ¿CÓMO Y POR QUÉ PUEDE UTILIZARSE UN ÁRBOL DE DECISIÓN COMO MÉTODO DE REDUCCIÓN DE DATOS?

- Un árbol de decisión busca minimizar el error de resustitución (probabilidad de equivocarse en la clasificación de una muestra).

d) DADO UN CONJUNTO DE DATOS ¿EL ÁRBOL DE DECISIÓN CONSTRUIDO A PARTIR DE ESTE ES ÚNICO?

- Si ya que se genera un árbol de decisión a partir de un conjunto de tuplas de entrenamiento de una partición

2.- REALIZA UNA EXPLORACIÓN DEL CONJUNTO DE DATOS WINE COMO EN LA PRÁCTICA 9.

1- Class Identifier

Descripción: Hace referencia al tipo de clase que pertenecen los vinos

Dominio: Cualquier valor entero entre el 1 y el 3

Tipo Dato: Polinomial

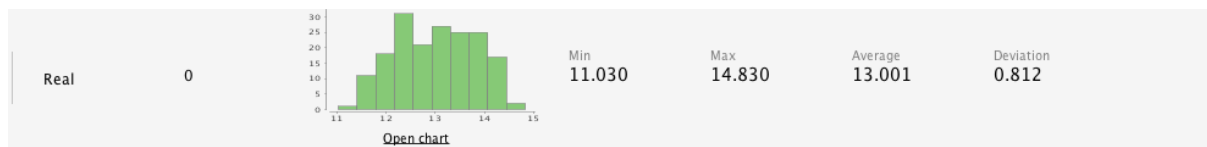


2- Alcohol

Descripción: Hace referencia al grado de alcohol en el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

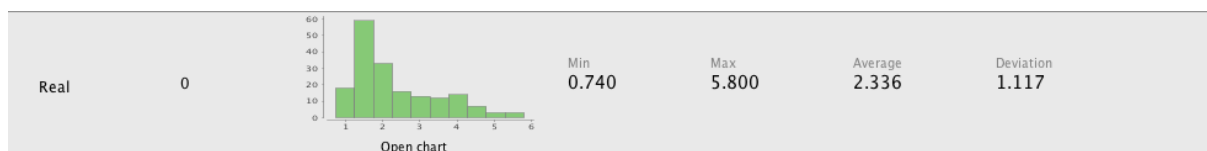


3- Malic Acid

Descripción: Hace referencia a la acides del vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

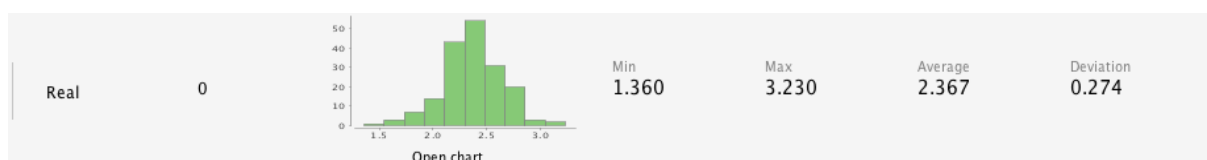


4- Ash

Descripción: Hace referencia a que tan cenizo es el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

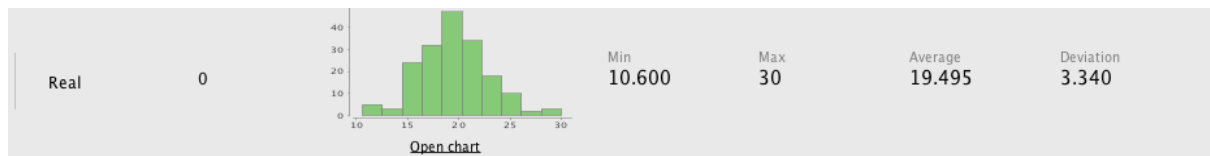


5- Alkalinity of Ash

Descripción: Hace referencia al grado de alcalinidad del vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

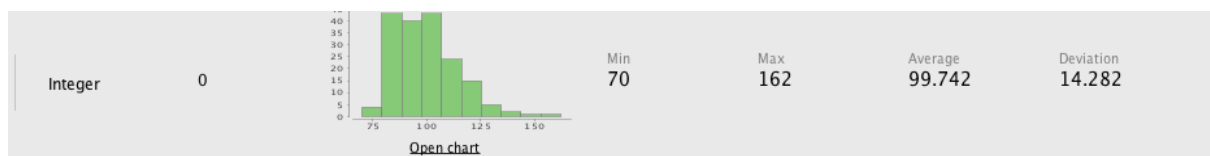


6- Magnesium

Descripción: Hace referencia al grado de magnesio en el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Integer

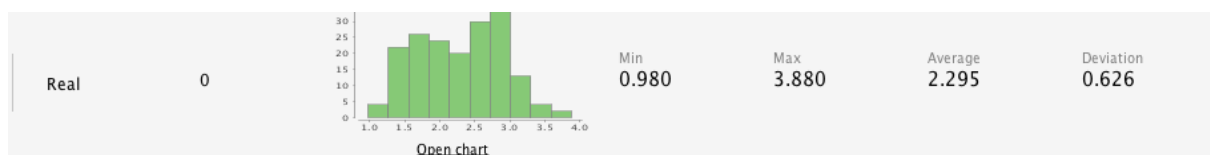


7- Total phenols

Descripción: Hace referencia al total de fenoles en el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

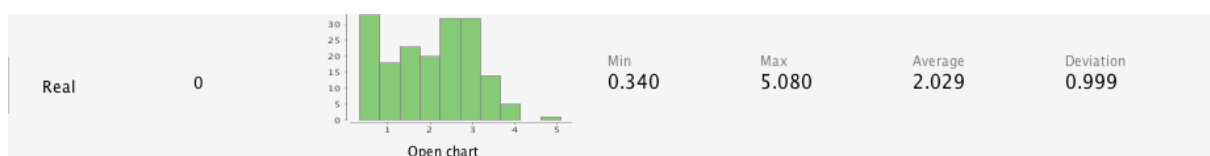


8- Flavanoids

Descripción: Hace referencia al numero de favanoides en el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

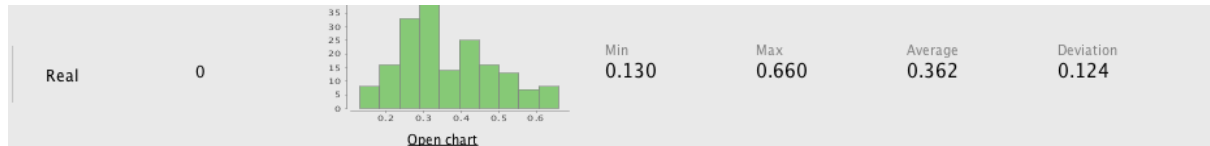


9- NonFlavanoids phenols

Descripción: Hace referencia al numero de flavanoides fenoles en el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

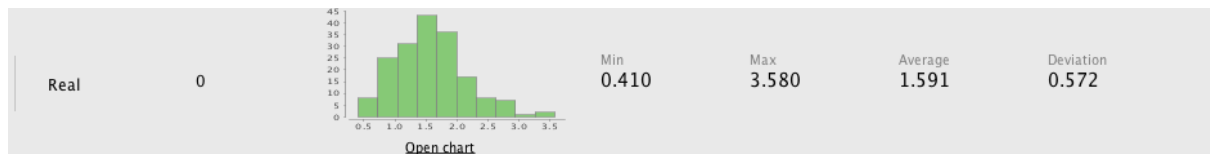


10- Proanthocyanins

Descripción: Hace referencia al numero de Proanthocyanins en el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

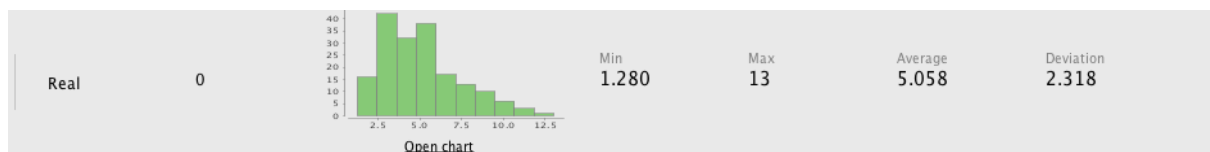


11- Color intensity

Descripción: Hace referencia a la intensidad de color en el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

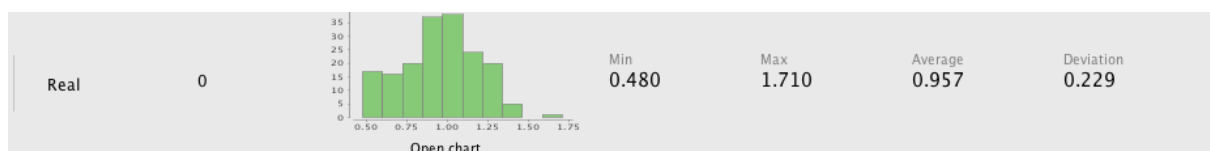


12- Hue

Descripción: Hace referencia al grado de matiz en el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

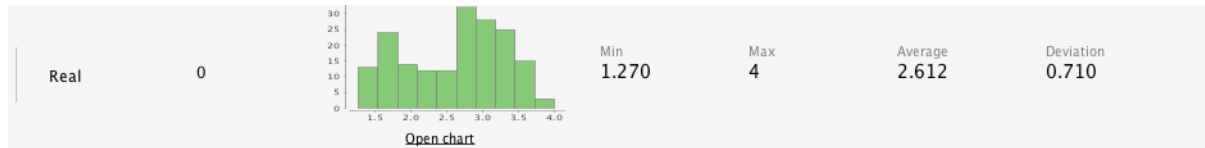


13- OD28/OD315 of diluted wine

Descripción: Hace referencia al grado de dilución en el vino

Dominio: Cualquier valor entero mayor a 0

Tipo Dato: Real

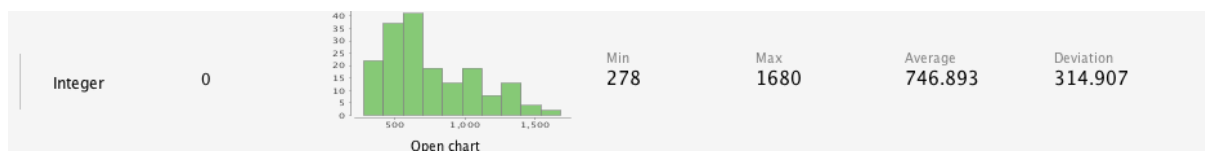


14- Proline

Descripción: Hace referencia al grado de proline en el vino

Dominio: Cualquier valor entero mayor a 0

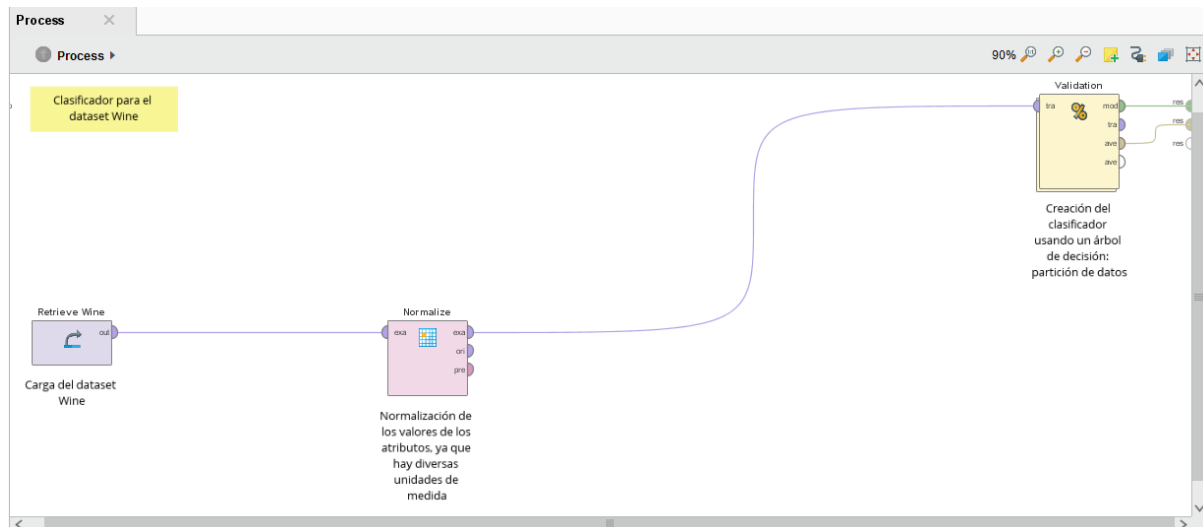
Tipo Dato: Integer



3.- ELABORA UN CLASIFICADOR PARA EL CONJUNTO DE DATOS WINE EN RAPIDMINER UTILIZANDO UN ÁRBOL DE DECISIÓN. EXPLIQUE BREVEMENTE LA CONFIGURACIÓN Y CONSTRUCCIÓN DE SU PROCESO.

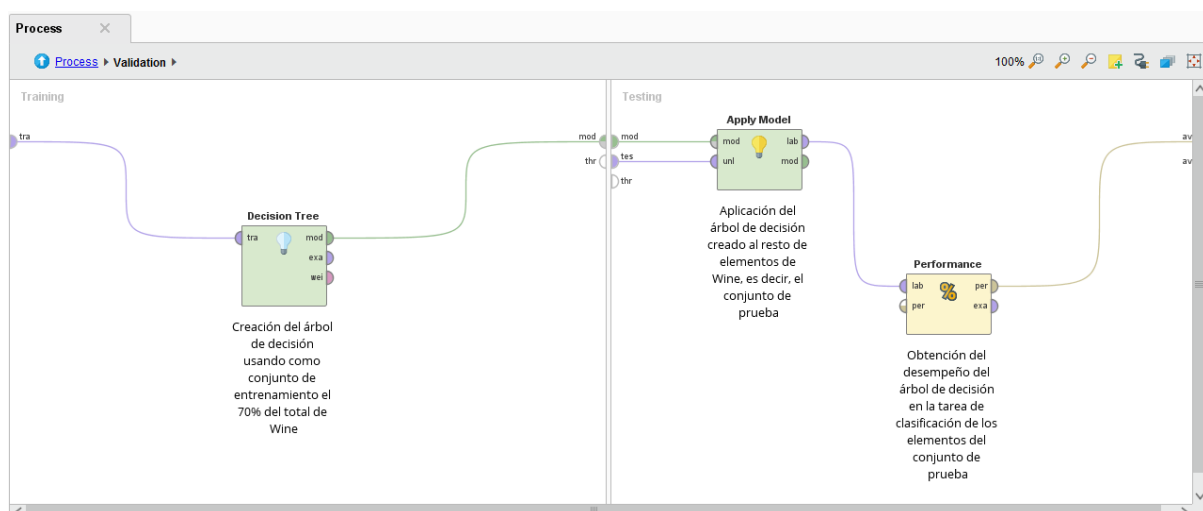
Para crear el árbol de decisión en RapidMiner, cargamos los datos y creamos un proceso. En ese proceso creamos un modelo con los datos.

Una vez cargado el csv deseado, lo primero que tenemos que hacer es normalizar los datos, ya que muchos de estos valores no los podemos ocupar sin normalizar.



Una vez normalizados hacemos la validación, la cual consiste de 3 pasos:

- 1.- Crear el árbol de decisión usando el conjunto de entrenamiento
- 2.- Aplicamos el modelo al conjunto de prueba, para ver como se comporta.
- 3- Verificamos el rendimiento de nuestro clasificador.

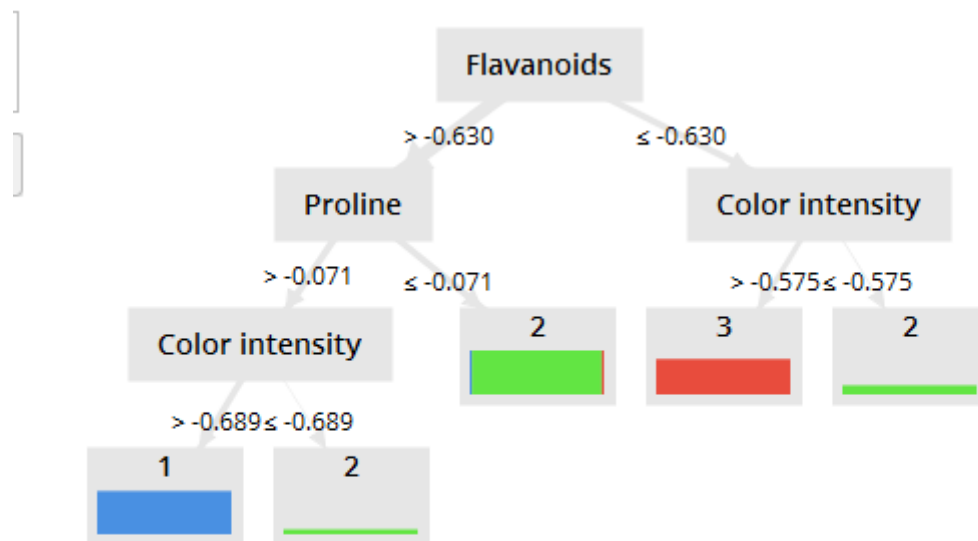


Una vez obtenido el modelo, ejecutamos y nos regresa la siguiente tabla, la cual nos dice como clasifica.

accuracy: 88.68%

	true 1	true 2	true 3	class precision
pred. 1	16	2	1	84.21%
pred. 2	0	18	0	100.00%
pred. 3	0	3	13	81.25%
class recall	100.00%	78.26%	92.86%	

Por otro lado nos da el árbol de decisión, de forma gráfica.



4.- ELABORA UN CLASIFICADOR PARA EL CONJUNTO DE DATOS WINE EN R UTILIZANDO UN ÁRBOL DE DECISIÓN. EXPLIQUE BREVEMENTE LA CONFIGURACIÓN Y CONSTRUCCIÓN DE SU PROCESO.

Para crear el árbol de decisión vamos a ocupar R, para esto cargamos la librería rattle y ejecutamos el comando.

```
R Console

R version 3.5.0 (2018-04-23) -- "Joy in Playing"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

[Previously saved workspace restored]

> library(rattle)
Rattle: A free graphical interface for data science with R.
Versión 5.1.0 Copyright (c) 2006-2017 Togaware Pty Ltd.
Escriba 'rattle()' para agitar, sacudir y rotar sus datos.
> rattle()
> |
```

Una vez que la instrucción se halla ejecutado, nos abrirá una interfaz grafica en la cual podremos implementar nuestro árbol.
Lo primero que vamos hacer es cargar nuestro archivo csv.

Minero de datos R - [Rattle]

Proyecto Herramientas Configuración Ayuda

Ejecutar Nuevo Abrir Guardar Informe Exportar Detener Salir Connect R

Dato: Explorar Prueba Transformar Clúster Asociada Modelo Evaluar Registro

Origen: ☒ File ☐ ARFF ☐ ODBC ☐ Conjunto de datos R ☐ Archivo de datos R ☐ Librería ☐ Corpus ☐ Rutina

Archivo: wine.data.csv Separador: , Decimal: . ☒ Encabezado

☐ Partición 70/15/15 Semilla: 42 Ver Editar

☒ Entrada ☐ Ignorar Calculadora de peso: Tipo de datos de destino: ☒ Automática ☐ Categórica ☐ Numérica ☐ Supervivencia

No. Variable	Tipo de datos	Entrada	Destino	Riesgo	Ident	Ignorar	Weight	Comentario
--------------	---------------	---------	---------	--------	-------	---------	--------	------------

Ya cargado csv, ejecutamos y nos regresa este conjunto de datos.

Proyecto Herramientas Configuración Ayuda

Ejecutar Nuevo Abrir Guardar Informe Exportar Detener Salir Connect R

Datos: Explorar Prueba Transformar Clúster Asociada Modelo Evaluar Registro

Origen: ☒ File ☐ ARFF ☐ ODBC ☐ Conjunto de datos R ☐ Archivo de datos R ☐ Librería ☐ Corpus ☐ Rutina

Archivo: wine.data.csv Separador: , Decimal: . ☒ Encabezado

☒ Partición 70/15/15 Semilla: 42 Ver Editar

☒ Entrada ☐ Ignorar Calculadora de peso:

Tipo de datos de destino: ☒ Automática ☐ Categórica ☐ Numérica ☐ Supervivencia

No. Variable	Tipo de datos	Entrada	Destino	Riesgo	Ident	Ignorar	Weight	Comentario
1 Class.identifier	Numérica	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 3
2 Alcohol	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 126
3 Malic.acid	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 133
4 Ash	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 79
5 Alcalinity.of.ash	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 63
6 Magnesium	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 53
7 Total.phenols	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 97
8 Flavanoids	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 132
9 Nonflavanoid.phenols	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 39
10 Proanthocyanins	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 101
11 Color.intensity	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 132
12 Hue	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 78
13 OD280.OD315.of.diluted.wines	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 122
14 Proline	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 121

Como vemos el TIPO DE DATOS DE DESTINO esta en opción de AUTOMATICO, la cual no nos conviene, así que lo pasamos a CATEGORICA, una vez seleccionado esta opción ejecutamos, a simple vista parece que no hace nada, pero no es así, rattle de forma interna cambia los datos a forma categórica para poderlos transformar.

Proyecto Herramientas Configuración Ayuda

Ejecutar Nuevo Abrir Guardar Informe Exportar Detener Salir Connect R

Datos: Explorar Prueba Transformar Clúster Asociada Modelo Evaluar Registro

Origen: ☒ File ☐ ARFF ☐ ODBC ☐ Conjunto de datos R ☐ Archivo de datos R ☐ Librería ☐ Corpus ☐ Rutina

Archivo: wine.data.csv Separador: , Decimal: . ☒ Encabezado

☒ Partición 70/15/15 Semilla: 42 Ver Editar

☒ Entrada ☐ Ignorar Calculadora de peso: Tipo de datos de destino: ☐ Automática ☒ Categórica ☐ Numérica ☐ Supervivencia

No. Variable	Tipo de datos	Entrada	Destino	Riesgo	Ident	Ignorar	Weight	Comentario
1 Class.identifier	Numérica	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 3
2 Alcohol	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 126
3 Malic.acid	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 133
4 Ash	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 79
5 Alcalinity.of.ash	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 63
6 Magnesium	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 53
7 Total.phenols	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 97
8 Flavanoids	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 132
9 Nonflavanoid.phenols	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 39
10 Proanthocyanins	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 101
11 Color.intensity	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 132
12 Hue	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 78
13 OD280.OD315.of.diluted.wines	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 122
14 Proline	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 121

Ahora nos vamos a la opción de TRANSFORMAR, en esta sección se hace el paso de NORMALIZACION la cual nos sirve para transformar los datos en una forma categórica. Se seleccionan todos los datos, menos la etiqueta de clase, una vez seleccionados usamos los parámetros que nos da rattle y ejecutamos otra vez.

Un vez ejecutado nos aparecerán los mismos atributos con las iniciales RRC_, lo cual significa que ya están normalizados.

Proyecto Herramientas Configuración Ayuda

Ejecutar Nuevo Abrir Guardar Informe Exportar Detener Salir Connect R

Datos Explorar Prueba Transformar Clúster Asociada Modelo Evaluar Registro

Tipo: ☒ Cambiar escala ☐ Atribuir ☐ Recode ☐ Limpiar

Normalizar: ☒ Recentralizar ☐ Escala [0-1] ☐ -Mediana/MAD ☐ Logaritmo natural ☐ Log 10 ☐ Matriz

Orden: ☐ Intervalo ☐ Interval Groups: 100

No. Variable	Faltan tipo de datos y número
1 Class.identifier	Númerica [1 a 3; única=3; media=1; mediana=2].
2 Alcohol	Númerica [11.03 a 14.83; única=126; media=13.00; mediana=13.05; ignorado].
3 Malic.acid	Númerica [0.74 a 5.80; única=133; media=2.34; mediana=1.87; ignorado].
4 Ash	Númerica [1.36 a 3.23; única=79; media=2.37; mediana=2.36; ignorado].
5 Alcalinity.of.ash	Númerica [10.60 a 30.00; única=63; media=19.49; mediana=19.50; ignorado].
6 Magnesium	Númerica [70 a 162; única=53; media=99; mediana=98; ignorado].
7 Total.phenols	Númerica [0.98 a 3.88; única=97; media=2.30; mediana=2.35; ignorado].
8 Flavanoids	Númerica [0.34 a 5.08; única=132; media=2.03; mediana=2.13; ignorado].
9 Nonflavanoid.phenols	Númerica [0.13 a 0.66; única=39; media=0.36; mediana=0.34; ignorado].
10 Proanthocyanins	Númerica [0.41 a 3.58; única=101; media=1.59; mediana=1.56; ignorado].
11 Color.intensity	Númerica [1.28 a 13.00; única=132; media=5.06; mediana=4.69; ignorado].
12 Hue	Númerica [0.48 a 1.71; única=78; media=0.96; mediana=0.96; ignorado].
13 OD280.OD315.of.diluted.wines	Númerica [1.27 a 4.00; única=122; media=2.61; mediana=2.78; ignorado].
14 Proline	Númerica [278 a 1680; única=121; media=746; mediana=673; ignorado].
15 RRC_Alcohol	Númerica [-2.43 a 2.25; única=126; media=-0.00; mediana=0.06].
16 RRC_Malic.acid	Númerica [-1.43 a 3.10; única=133; media=-0.00; mediana=-0.42].
17 RRC_Ash	Númerica [-3.67 a 3.15; única=79; media=0.00; mediana=-0.02].
18 RRC_Alcalinity.of.ash	Númerica [-2.66 a 3.15; única=63; media=-0.00; mediana=0.00].
19 RRC_Magnesium	Númerica [-2.08 a 4.36; única=53; media=-0.00; mediana=-0.12].

Nos regresamos a datos y nos daremos cuenta que ahora ignorara los datos que no están normalizados, y tomara en cuenta los que ya normalizamos.

Proyecto Herramientas Configuración Ayuda

Ejecutar Nuevo Abrir Guardar Informe Exportar Detener Salir Connect R

Datos: Explorar Prueba Transformar Clúster Asociada Modelo Evaluar Registro

Origen: ☒ File ☐ ARFF ☐ ODBC ☐ Conjunto de datos R ☐ Archivo de datos R ☐ Librería ☐ Corpus ☐ Rutina

Archivo: wine.data.csv Separador: , Decimal: . ☒ Encabezado

☒ Partición 70/15/15 Semilla: 42 Ver Editar

☒ Entrada ☐ Ignorar Calculadora de peso:

Tipo de datos de destino
☒ Automática ☐ Categórica ☐ Numérica ☐ Supervivencia

No. Variable	Tipo de datos	Entrada	Destino	Riesgo	Ident	Ignorar	Weight	Comentario
12 Hue	Numérica	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 78
13 OD280.OD315.of.diluted.wines	Numérica	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 122
14 Proline	Numérica	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 121
15 RRC_Alcohol	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 126
16 RRC_Malic.acid	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 133
17 RRC_Ash	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 79
18 RRC_Alcalinity.of.ash	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 63
19 RRC_Magnesium	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 53
20 RRC_Total.phenols	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 97
21 RRC_Flavonoids	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 132
22 RRC_Nonflavonoid.phenols	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 39
23 RRC_Proanthocyanins	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 101
24 RRC_Color.intensity	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 132
25 RRC_Hue	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 78
26 RRC_OD280.OD315.of.diluted.wines	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 122
27 RRC_Proline	Numérica	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 121

Ahora creamos el modelo del árbol, con un conjunto de aprendizaje y otro de prueba, así que ocupamos los valores por defecto de rattle. Así que ejecutamos y nos regresa lo siguiente.

Proyecto Herramientas Configuración Ayuda

Ejecutar Nuevo Abrir Guardar Informe Exportar Detener Salir Connect R

Datos Explorar Prueba Transformar Clúster Asociada Modelo Evaluar Registro

Tipo: ☒ Árbol ☐ Bosque ☐ Potenciar ☐ SVM ☐ Lineal ☐ Red neural ☐ Supervivencia ☐ Todos

Destino: Class.identifier Algoritmo: ☒ Tradicional ☐ Condicional

División mínima: 20 Profundidad máxima: 3 Previos:

Cubo mínimo: 7 Complejidad: 0.0100 Matriz de pérdida:

Resumen del modelo Árbol de decisión de Clasificación (construido con 'rpart'):

n= 124

```
node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 124 73 2 (0.30645161 0.41129032 0.28225806)
2) RRC_Proline>=0.6560871 33 0 1 (1.00000000 0.00000000 0.00000000) *
3) RRC_Proline< 0.6560871 91 40 2 (0.05494505 0.56043956 0.38461538)
6) RRC_Color.intensity< -0.5318973 44 0 2 (0.00000000 1.00000000 0.00000000) *
7) RRC_Color.intensity>=-0.5318973 47 12 3 (0.10638298 0.14893617 0.74468085)
14) RRC_Flavanoids>=-0.6450058 13 6 2 (0.38461538 0.53846154 0.07692308) *
15) RRC_Flavanoids< -0.6450058 34 0 3 (0.00000000 0.00000000 1.00000000) *
```

Classification tree:

```
rpart(formula = Class.identifier ~ ., data = crs$dataset[crs$train,
c(crs$input, crs$target)], method = "class", parms = list(split = "information"),
control = rpart.control(usesurrogate = 0, maxsurrogate = 0))
```

Variables actually used in tree construction:

```
[1] RRC_Color.intensity RRC_Flavanoids RRC_Proline
```

Root node error: 73/124 = 0.58871

n= 124

Despues de que el modelo este terminado, evaluaremos con el conjunto de prueba.

Proyecto Herramientas Configuración Ayuda

Ejecutar Nuevo Abrir Guardar Informe Exportar Detener Salir Connect R

Datos Explorar Prueba Transformar Clúster Asociada Modelo Evaluar Registro

Tipo: ☒ Matriz de error ☐ Riesgo ☐ Curva de costo ☐ Hand ☐ Elevación ☐ ROC ☐ Precisión ☐ Sensibilidad ☐ Obj prev ☐ Calificación

Modelo: ☒ Árbol ☐ Potenciar ☐ Bosque ☐ SVM ☐ Lineal ☐ Red neural ☐ Supervivencia ☐ KMeans ☐ HClust

Datos: ☐ Entrenamiento ☐ Convalidación ☒ Prueba ☐ Completo ☐ Ingresar ☐ Archivo CSV ☐ Conjunto de datos R

Variable de riesgo: Informe: ☒ Clase ☐ Probabilidad Incluir: ☒ Identificadores ☐ Todos

Matriz de error para el modelo Árbol de decisión en wine.data.csv [prueba] (cuentas):

		Predicted			
Actual	1	2	3	Error	
1	9	5	0	35.7	
2	0	9	0	0.0	
3	0	0	5	0.0	

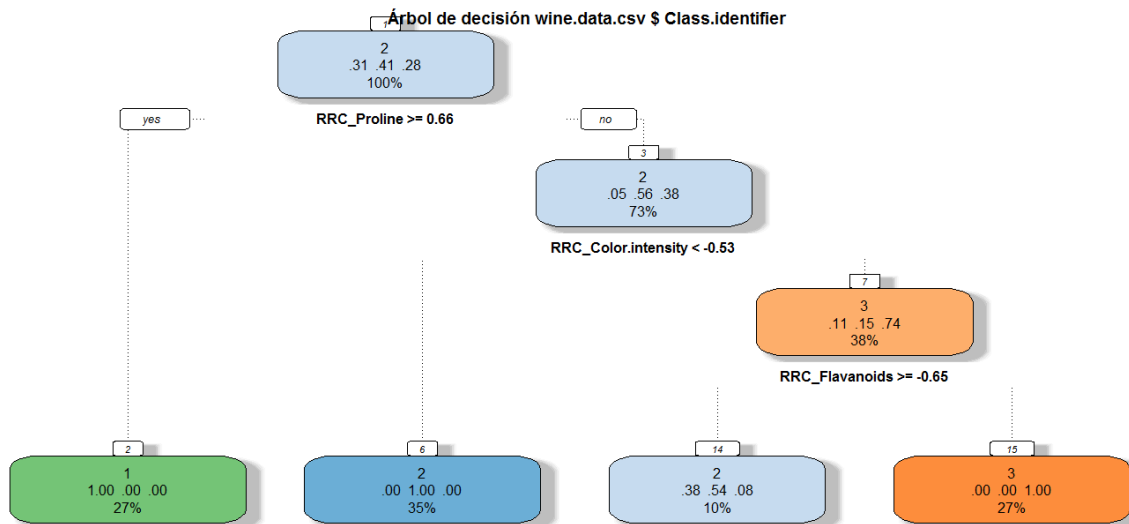
Error matrix for the Árbol de decisión model on wine.data.csv [prueba] (proportions):

		Predicted			
Actual	1	2	3	Error	
1	32.1	17.9	0.0	35.7	
2	0.0	32.1	0.0	0.0	
3	0.0	0.0	17.9	0.0	

Overall error: 17.9%, Averaged class error: 11.9%

Rattle marca de tiempo: 2018-05-15 18:50:00 Rogelio Alcantar

Por otro lado veremos el árbol que nos da R.



5.- COMPARA LOS DOS ÁRBOLES DE DECISIÓN OBTENIDOS Y JUSTIFICA CUÁL DE LOS DOS ES “MEJOR”.

Para RapidMiner:

Con la matriz de error podemos saber que RapidMiner acierta en un 88.68% lo cual su promedio de aciertos es bueno.

Ahora interpretamos la tabla:

Para clasificar el valor 1, tiene un 84.21% de acierto a la hora de clasificarlo, esto se debe a que algunos valores que pertenecen a la clasificación 2 y 3 los clasifica como 1.

Para clasificar el valor 2, tiene un 100% de acierto a la hora de clasificarlo.

Para clasificar el valor 3, tiene un 81.25% de acierto a la hora de clasificarlo.

accuracy: 88.68%

	true 1	true 2	true 3	class precision
pred. 1	16	2	1	84.21%
pred. 2	0	18	0	100.00%
pred. 3	0	3	13	81.25%
class recall	100.00%	78.26%	92.86%	

Para R:

Con la matriz de error podemos saber que R acierta en un 82.10% lo cual su promedio de aciertos es bueno.

Ahora interpretamos la tabla:

Para clasificar el valor 1, tiene un 64.3% de acierto a la hora de clasificarlo, esto se debe a

que algunos valores que pertenecen a la clasificación 2 los clasifica como 1.
 Para clasificar el valor 2, tiene un 100% de acierto a la hora de clasificarlo.
 Para clasificar el valor 3, tiene un 100% de acierto a la hora de clasificarlo.

The screenshot shows the RapidMiner interface with the 'Evaluar' (Evaluate) tab selected. The 'Modelo' (Model) dropdown is set to 'Árbol' (Tree). The 'Datos' (Data) dropdown is set to 'Prueba' (Test). The 'Informe' (Report) dropdown is set to 'Clase' (Class). The 'Incluir' (Include) dropdown is set to 'Identificadores' (Identifiers). The output window displays the following text:

```
Matriz de error para el modelo Árbol de decisión en wine.data.csv [prueba] (cuentas):
```

	Predicted			
Actual	1	2	3	Error
1	9	5	0	35.7
2	0	9	0	0.0
3	0	0	5	0.0

```
Error matrix for the Árbol de decisión model on wine.data.csv [prueba] (proportions):
```

	Predicted			
Actual	1	2	3	Error
1	32.1	17.9	0.0	35.7
2	0.0	32.1	0.0	0.0
3	0.0	0.0	17.9	0.0

```
Overall error: 17.9%, Averaged class error: 11.9%
Rattle marca de tiempo: 2018-05-15 18:50:00 Rogelio Alcantar
```

A pesar de que RapidMiner tiene más aciertos generales, R es mejor clasificador de clase, ya que este tiene dos clases con 100%.

Nosotros deducimos que R es mejor clasificador, ya que solo se puede equivocar a la hora de clasificar la clase 1 tomando tuplas de la clase 2, pero en el caso de que clasifique primero la clase 2, no tendríamos 100% de aciertos el cual sería el mejor de los casos, y el por sería de 17.9 (checha la imagen y ponle el porcentaje de error bien).

En cambio RapidMiner no importa cual clasifique primero, siempre tiene un porcentaje de error en la clase.