Universitat Oberta de Catalunya

TFC – Minería de datos

Estudio sobre las causas del abandono de los estudiantes de economía de la Universidad Oberta de Catalunya entre los años 1998 y 2008

Autor: Víctor Aguilera Arranz

Consultor: Ramón Caihuelas Quiles

24 de Junio de 2013

Índice de contenido

1. Objetivo del proyecto

Cada año, muchos estudiantes se matriculan en las diferentes carreras que ofrece la UOC, sin embargo, no todos acaban los estudios que comienzan, muchos de estos estudiantes los abandonan antes de finalizarlos. Al igual que pasaría en cualquier empresa, a la Universidad le interesaría saber si existen razones objetivas que no estén basadas en el azar para la perdida de estos estudiantes.

Por medio de diferente técnicas y algoritmos de minería de datos se va a intentar extraer el máximo conocimiento de los datos con los que cuenta la Universidad en sus bases de datos para intentar establecer los motivos que llevan a un estudiante a abandonar sus estudios o que patrones de conducta tienen los estudiantes que no llegan a finalizarlos.

2. Otros objetivos

En el desarrollo del proyecto posiblemente se obtengan otros conocimientos en los que en un principio no se ha reparado y que a posteriori pudieran ser interesantes. Se irán descubriendo a medida que se avanza en el desarrollo del proyecto.

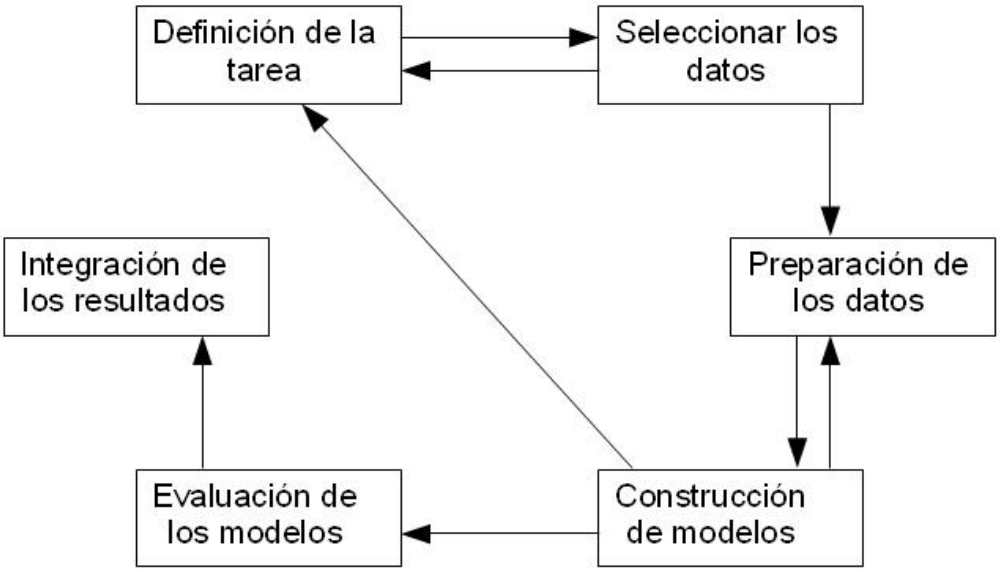
3. Planificación

1-6 - Objetivo del proyecto e introducción al TFC		01/03/13 - 11/04/13
7 - Comienzo de las fases de minería de datos		11/04/13
7.1 - Selección y preparación de los datos	9 dias	11/04/13 - 20/04/13
7.2 - Descripción de los datos	10 dias	21/04/13 – 30/04/13
7.3 - Elaboración del modelo	24 dias	01/05/13 - 23/05/13
7.4 – Extracción del conocimiento	10 dias	24/05/13 – 02/06/13
7.5 – Elaboración del modelo encontrado para resolver el problema planteado.	11 dias	03/06/13 – 13/06/13

8 - Revisión del trabajo realizado	5 dias	14/06/13 – 18/06/13
9 – Referencias y bibliografía	2 dias	19/06/13 – 20/06/13
10 - Anexos	4 dias	21/06/13 – 24/06/13
ENTREGA del TFC		25/06/13

4. Aspectos metodológicos

La metodología que se va a seguir es la del modelo de referencia CRISP-DM, este modelo cuenta con ciclo de vida compuesto de las siguientes fases:



Las flechas indican el sentido de las relaciones habituales aunque en la realidad podrían darse relaciones entre cualquiera de las fases. A continuación se explica cada una de las fases:

Definición de la tarea

El objetivo de esta fase es decidir cual será el objetivo del proyecto:

- Al final de esta fase tendremos que haber aproximado el objetivo a las tareas genéricas que hayamos decidido (encontrar similitudes, clasificar objetos, predecir un comportamiento, describir asociaciones significativas entre diferentes variables, explicar un comportamiento.
- Decidir el modelo que necesitamos o cual es que mejor se adapta a

nuestro proyecto

- Seleccionar el método necesario para construirlo

Seleccionar los datos

En esta fase hay que localizar la fuente de datos de entre las bases de datos dispersas y las bases de datos transaccionales.

Preparación de los datos

La preparación de los datos requiere de los siguientes tratamientos: limpieza de los datos, transformación de los datos y reducción de su dimensionalidad.

Construcción de modelos

En el libro de la asignatura también se refiere a esta fase como “Data mining”.

En esta fase disponemos de los resultados de las fases anteriores y tendremos que elegir el método de construcción del modelo que nos interesa. Realizaremos para ello un proceso de búsqueda que finalizará con la construcción del modelo que mejor se adapta a nuestro proyecto.

Evaluación e interpretación del modelo

En esta fase se evalúa la calidad del modelo, se valora si es el mejor modelo que se podría hacer o como mejorarlo. Y una vez que se ha obtenido el mejor modelo posible es el momento de interpretar el conocimiento que nos muestra, teniendo especial cuidado en no sacar falsas interpretaciones.

Integración de los resultados

Esta es la última fase, llegados a este punto se integraran los resultados obtenidos en el sistema de información en el que se este aplicando.

5. Estado del arte de la minería de datos

Debido a los avances tecnológicos y a la digitalización de la información cada vez las empresas acaparan mas datos sobre sus clientes o su negocio en general. Esta información acaba en grandes bases de datos que la empresa no duda en utilizar en su provecho, pero ¿que puede hacer una empresa con esta información?

En lineas generales esta actividad mejora el conocimiento sobre el entorno de negocio y puede hacer la empresa mas productiva, porque venda mas, porque gaste menos o porque en definitiva aproveche mejor los recursos con los que cuenta.

El reto que se plantea es como extraer conocimiento de entre los millones de datos que una empresa puede almacenar. De esta necesidad nace la minería de datos, que consiste en extraer la máxima información procesable que se encuentra en bases de datos o almacenes de datos (Data Warehouse) y lo mas importante, convertirla en conocimiento.

Para ello se usan diferente técnicas en función de los objetivos del proyecto que pueden ser desde encontrar similitudes y agrupar objetos parecidos hasta clasificar los objetos o predecir o describir o explicar un comportamiento o unos resultados)

La minería de datos esta estrechamente relacionada con la estadística, con la computación, con la inteligencia artificial y con la documentación entre otros campos.

Se trata de usar diversas técnicas y algoritmos para la extracción de este conocimiento. Este trabajo sería impensable que se pudiera realizar por parte de un analista, sin el uso de las nuevas tecnologías.

Desde los comienzos de la informática hasta nuestros días han ido apareciendo nuevas aplicaciones, algoritmos y técnicas especializadas en la minería de datos cada vez mas complejos y eficientes. Algunos ejemplos de aplicaciones son:

- RapidMiner
- R
- Orange
- Weka
- JhepWork
- Knime

estas aplicaciones son de código abierto, pero también compañías como Microsoft tienen soluciones de minería de datos como Microsoft SQL Server Analysis Services.

Por otro lado muchas universidades ofrecen masters en minería de datos como por ejemplo la Universidad de Sevilla: MDA: Minería de Datos Aplicada (Máster Universitario en Ingeniería y Tecnología del Software (R.D.1393/07) o la UOC en el master Business Intelligence.

Y por supuesto todas las grandes empresas hacen minería de datos para poder ser mas competitivas eliminando en lo posible sus debilidades y potenciando sus capacidades.

Por todo ello se puede estar seguro de que la minería de datos es una actividad que esta en auge y que tendrá cada vez mayor peso en la toma de decisiones de muchos negocios.

6. Estado de la minería de datos en el entorno educativo

Objetivos similares al planteado en este TFC han sido abordados con mayor o menor fortuna en otras ocasiones, estos son algunos de los ejemplos mas recientes:

En la edición de noviembre de 2012 de IEEE-RITA . (http://rita.det.uvigo.es/201208/uploads/IEEE-RITA.2012.V7.N3.A1.pdf) profesores de las universidades de Zacatecas (México) y Córdoba (España) publican un estudio llamado "Predicción del fracaso escolar mediante técnicas de minería de datos". Este estudio pretende buscar los factores que hacen que estudiantes de enseñanza media o secundaria fracasen. Las conclusiones a las que llegan son bastante interesantes obteniendo como resultados:

- Algoritmos de clasificación capaces de predecir el rendimiento académico de los estudiantes.
- Utilización de algoritmos de clasificación de tipo "caja-blanca" que permiten obtener modelos comprensibles por usuarios no expertos en minería.
- Y con respecto a los factores que influyen en el abandono de los estudiantes el principal ha sido las notas de las asignaturas.

La Universidad Politécnica de Valencia realiza un estudio llamado "Una propuesta metodológica de estudio de rendimiento académico en una titulación universitaria" (http://redaberta.usc.es/aidu/index2.php?option=com_docman&task=doc_view&gid=371&Itemid=8) en él trata de analizar los datos de rendimiento académico de los estudiantes de Informática de ocho universidades.

Expone que las técnicas empleadas podrían realizarse con aplicaciones comerciales como por ejemplo Clementine o libres como Weka.

Finalmente se alude a la privacidad de los resultados obtenidos y no son publicados en Internet.

Por último mencionar que el semestre pasado un alumno de la UOC realizó un estudio similar. La herramienta que utilizó para ello, fue principalmente WEKA y sus resultados se encuentran publicados en la biblioteca de la Universidad.

7. Minería de datos

7.1 Definición de la tarea de minería de datos

El objetivo que se persigue es saber cual es el motivo del abandono de ciertos estudiantes de la carrera de Ciencias Empresariales.

Para llevar a cabo el objetivo planteado lo primero que se necesita es tener una idea sobre los datos que se están utilizando, como se distribuyen o agrupan.

Los métodos de agregación o clusters suelen ser útiles cuando se tiene un relativo desconocimiento del dominio que se esta tratando con lo que a priori no sabemos como separar los datos en categorías o grupos de una forma eficiente.
Los métodos de agregación permiten en base a unos criterios de distancia y medidas de similitud separar los datos en grupos lo mas heterogéneos posible con la mayor homogeneidad entre los componentes internos de cada uno.

Por tanto el primero de los modelos que se van a realizar será un modelo de clusters.

Por otro lado se realizarán otros modelos en los que se va a utilizar la información aportada por los clusters para la realización de arboles de decisión. En un modelo se realizarán los arboles sobre los resultados obtenidos de los clusters y en otro se utilizará la información obtenida en la generación de los cluster solamente para la selección de los atributos y se generará de un árbol de decisión sobre los datos originales.

Los árboles de decisión son estructuras formadas por nodos en las cuales cada uno de ellos se pregunta sobre un atributo determinado y en función de la respuesta toman una rama u otra. Siguiendo las ramas desde la raíz a las hojas se pueden obtener una serie de condiciones que permiten clasificar las nuevas observaciones. Son uno de los modelos de decisión mas utilizados y permiten una

traducción rápida y natural a listas de reglas de decisión, por lo tanto son una opción muy adecuada para llevar a cabo el objetivo planteado.

7.2 Selección de los datos

Se va a trabajar con la base de datos proporcionada por la UOC. Esta base de datos almacena información sobre los alumnos que se han matriculado en la carrera de Ciencias Empresariales entre los años 1998 y 2008.

Los datos se encuentran almacenados en un fichero *.dat* donde los datos están separados por espacios.

Se muestra un fragmento del fichero:

```
TFC_MD.dat
1 ID ABANDONA TITULAT SEXE EDAT FRANJA SEMESTRE NSEM NA NC NASUP NCSUP PCTAS PCTCS VIA NACMAT NACPRE NACSU
2 12445859 0 0 0 28 3 20011 11 3 16.5 3 16.5 1 1 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 12355702 0 0 1 36 4 19992 8 6 31.5 5 25.5 0.833333 0.809524 4 6 6 5 1 1 1 1 1 1 1 1 1 -1 0 0 1 1 0 0 0 0
4 12355902 0 0 1 48 5 19982 6 2 10.5 2 10.5 1 1 4 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5 12450601 0 0 0 43 5 20062 22 2 10.5 2 10.5 1 1 1 2 2 2 1 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
6 12450792 1 0 0 27 2 20031 15 2 10.5 0 0 0 0 3 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

Junto con el fichero de datos se proporciona también un fichero con la descripción de cada uno de los campos. Estos datos son los siguientes:

ID: identificador único por cada estudiante

ABANDONA: abandona después del primer semestre (0 falso, 1 verdadero)

TITULAT: se titula despues del primer semestre (0 fals, 1 cert)

SEXE: 0 muje, 1 hombre

EDAT: edad en años en el momento de entrar

FRANJA: franja de edad <=24 años (1), <=27 (2), <=30 (3), <=36 (4), > 36 (5)

SEMESTRE: semestre en el que inicia los estudios

NSEM: numero de semestre relativo desde que se iniciaron los estudios

NA: numero de asignaturas matriculadas en el primer semestre

NC: numero de créditos matriculados

NASUP: numero de asignaturas superadas

NCSUP: numero de créditos superados

PCTAS: porcentaje de asignaturas superadas (NASUP/NA)

PCTCS: porcentaje de créditos superados (NCSUP/NC)

VIA: vía de acceso 1 no cou, 2 cou, 3 estudios inacabados, 4 titulado

NACMAT: numero de asignaturas matriculadas del conjunto de les 12 mas comunes del 1er semestre.

NACPRE: numero de asignaturas a las cuales se presenta del conjunto de las 12 mas comunes del 1er semestre

NACSUP: numero de asignaturas superadas del conjunto de las 12 mas comunes del 1er semestre

A1M: 0 si no es matricula de l'assignatura 1, 1 si es matricula

A1S: -1 si no supera la asignatura 1, 0 si no la matricula o no es presenta, 1 si la supera

A2M A2S: idem para la asignatura 2

A3M A3S A4M A4S A5M A5S A6M A6S A7M A7S A8M A8S A9M A9S A10M A10S A11M A11S A12M A12S: idem para el resto

y la lista de las asignaturas que componen el plan de estudios de la carrera que estamos estudiando:

Lista de asignaturas:

00.010: multimedia i comunicacio

00.002: angles I

01.001: introduccio al dret

01.079: introduccio a la macroeconomia

01.005: introduccio a la comptabilitat

01.003: matematiques I

01.006: organitzacio i administracio d'empreses I

01.004: estadistica I

01.078: introduccio a la microeconomia

01.009: direccio de la produccio I

00.004: angles III

00.003: angles II

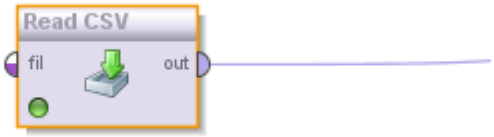
Se importan estos datos a un repositorio de *RapidMiner* para poder analizarlos con la aplicación.

7.3 Preparación de los datos

Es necesario definir cada uno de los campos de la forma mas adecuada a su tipología y significado. De esta forma pasaremos de tener la información en un fichero de texto separado por espacios a una estructura mas compleja y mas útil a la hora de trabajar con ella.

7.3.1 Importación de los datos

Para leer el contenido del fichero *TFC_MD.dat* se hace uso del objeto *Read CSV* que habrá que configurar para que tome como entrada este fichero y lo sepa interpretar como corresponde, para ello se especifica que tome como separador el carácter espacio.



A continuación se definen cada uno de los campos de la siguiente forma:

ID

Este campo es un identificador único de un estudiante. La información que proporciona es la de un nombre que representa de forma univoca a un estudiante. Esta formado por una cadena de 8 dígitos pero no tiene ningún valor numérico ya que no es un campo con el que tenga sentido realizar ningún tipo de operación

matemática, por tanto se define este atributo como nominal.

ABANDONA

Este campo toma los valores 0 y 1 que representa los valores falso o verdadero respectivamente, con lo que no se va a poder hacer un tratamiento matemático de este campo.

Este atributo se define como nominal.

TITULAT

Toma dos valores 0 y 1 representando a falso o verdadero respectivamente.

Este atributo se define como nominal.

SEXE

Este campo toma dos valores 0 y 1 en función de que sea mujer u hombre respectivamente.

Este atributo se define como nominal.

EDAT

En este campo se almacena la edad del estudiante en años. Por tanto es un campo numérico.

Puesto que este atributo es numérico y no tiene decimales, se define como integer.

FRANJA

Este campo toma cinco posibles valores, del 1 al 5 indicando la categoría de edad en la que esta el estudiante. No es un campo numérico por tanto.

Este atributo se define como nominal.

SEMESTRE

Este campo esta formado por el año en formato de cuatro dígitos y el semestre,

1 dígito. No es un campo fecha y tampoco un número.

Se toma este atributo como nominal.

NSEM

Este campo representa el numero de semestres, con lo que el atributo es numérico y entero.

Se define este atributo como integer.

NA

Este campo representa el número de asignaturas matriculadas, con lo que se trata de un campo numérico entero.

Se define este atributo como integer.

NC

Este campo representa el número de créditos matriculados, con lo que se trata de un campo numérico que puede tomar valores decimales.

Se define este atributo como real.

NASUP

Este campo representa el número de asignaturas superadas, con lo que se trata de un campo numérico entero.

Se define este atributo como integer.

NCSUP

Este campo representa el número de créditos superados, con lo que se trata de un campo numérico que puede tomar valores decimales.

Se define este atributo como real.

PCTAS

Este campo representa el porcentaje de asignaturas superadas, con lo que se

trata de un campo numérico que puede tomar valores decimales.

Se define este atributo como real.

PCTCS

Este campo representa el porcentaje de créditos superados, con lo que se trata de un campo numérico que puede tomar valores decimales.

Se define este atributo como real.

NACMAT

Este campo representa el número de asignaturas matriculadas, con lo que se trata de un campo numérico entero.

Se define este atributo como integer.

NACPRE

Este campo representa el número de asignaturas a las que se presenta, con lo que se trata de un campo numérico entero.

Se define este atributo como integer.

NACSUP

Este campo representa el número de asignaturas superadas, con lo que se trata de un campo numérico entero.

Se define este atributo como integer.

A1M – A12M

Estos campos representan las asignaturas de las que se pueden matricular, toman los valores 1 y 0 dependiendo de si el estudiante se ha matriculado de esa asignatura o no respectivamente. Por lo tanto su valor no es numérico.

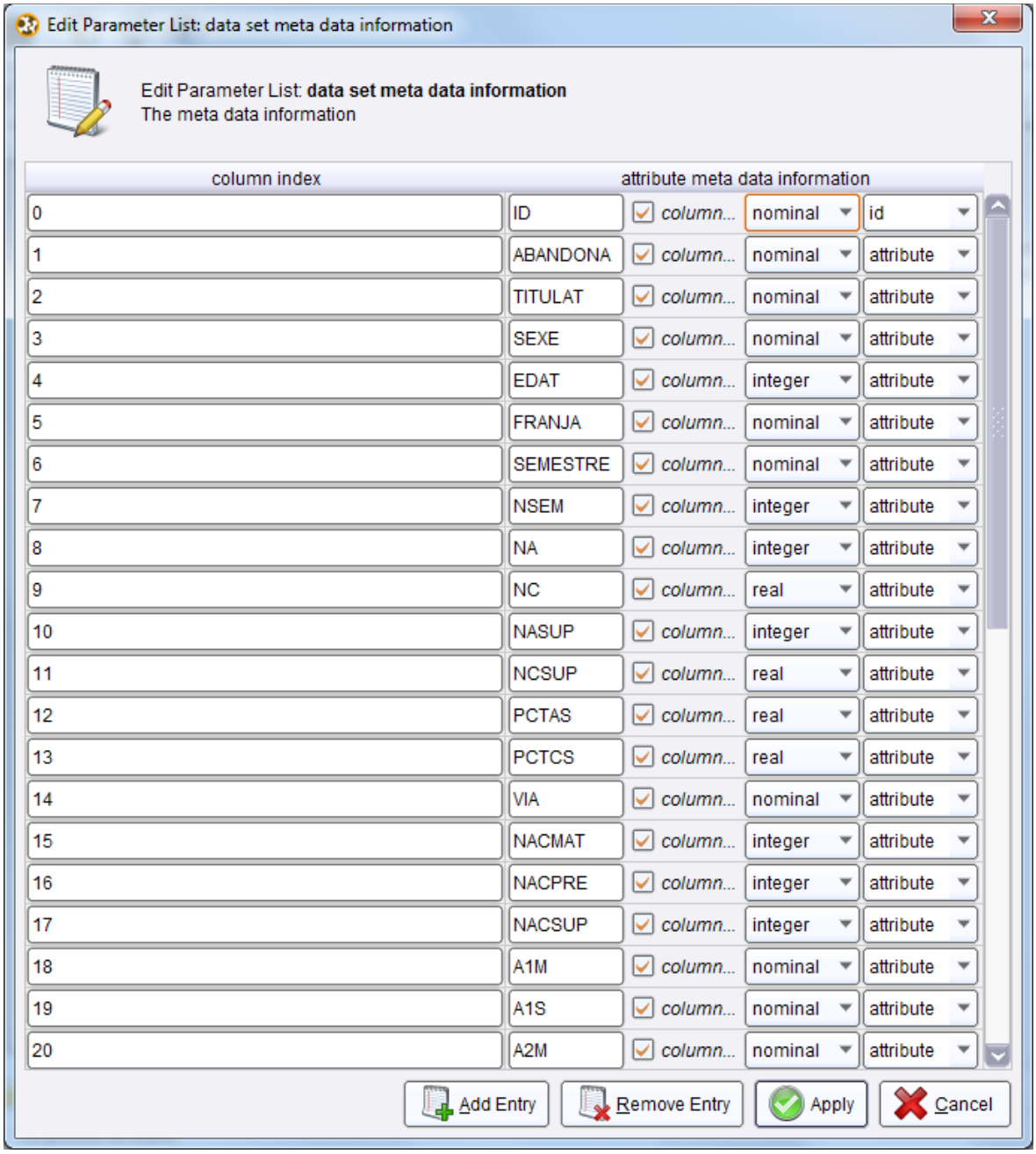
Estos atributos se definen como nominales.

A1S – A12S

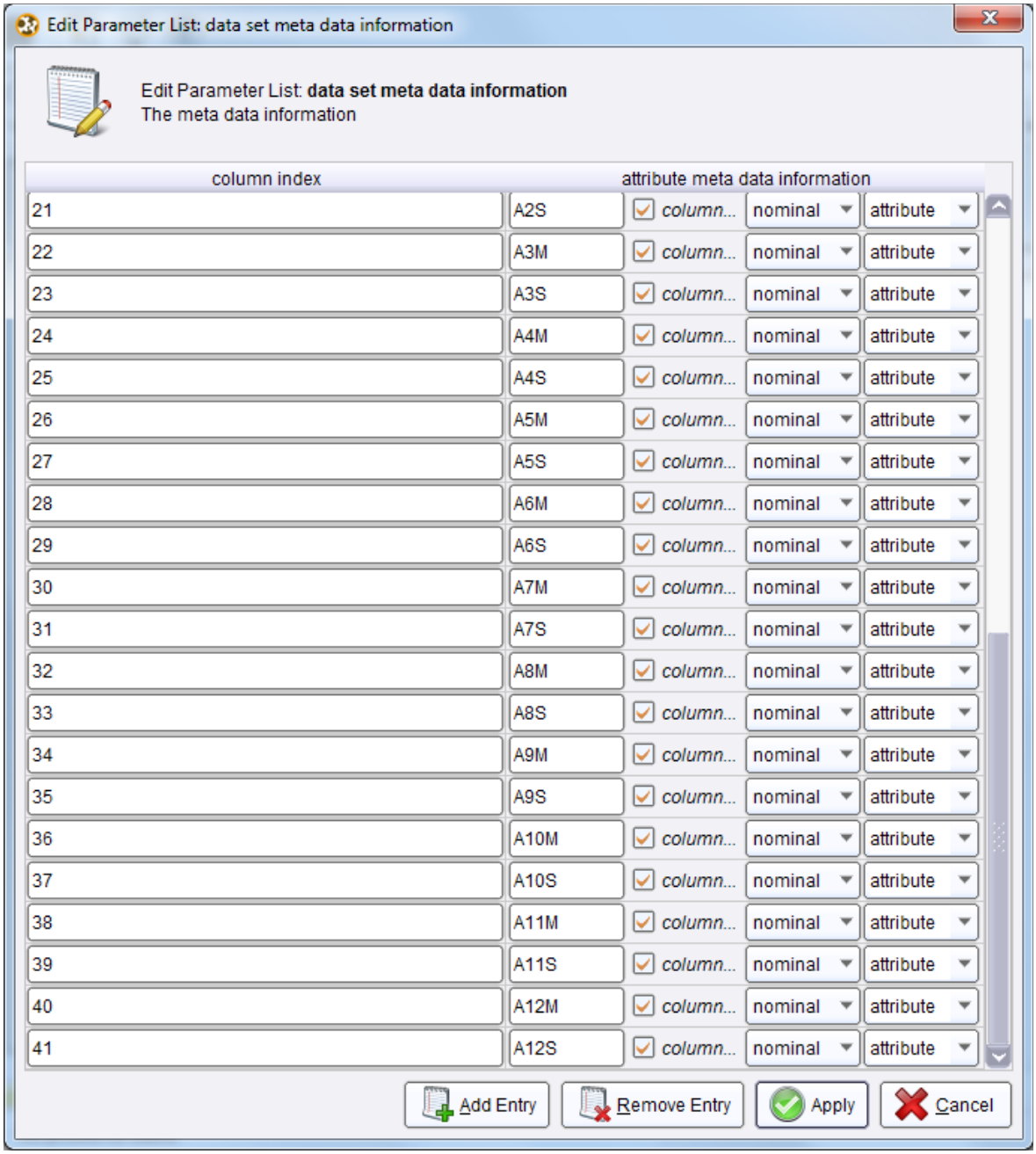
Estos campos representan las asignaturas de las que se pueden matricular, toman los valores -1, 0 y 1 dependiendo de si el estudiante no supera la asignatura, si no se matricula o no se presenta, o si la supera. Por lo tanto su valor no es numérico.

Estos atributos se definen como nominales.

Este sería el resultado:



column index				
0	ID	column...	nominal	id
1	ABANDONA	column...	nominal	attribute
2	TITULAT	column...	nominal	attribute
3	SEXE	column...	nominal	attribute
4	EDAT	column...	integer	attribute
5	FRANJA	column...	nominal	attribute
6	SEMESTRE	column...	nominal	attribute
7	NSEM	column...	integer	attribute
8	NA	column...	integer	attribute
9	NC	column...	real	attribute
10	NASUP	column...	integer	attribute
11	NCSUP	column...	real	attribute
12	PCTAS	column...	real	attribute
13	PCTCS	column...	real	attribute
14	VIA	column...	nominal	attribute
15	NACMAT	column...	integer	attribute
16	NACPRE	column...	integer	attribute
17	NACSUP	column...	integer	attribute
18	A1M	column...	nominal	attribute
19	A1S	column...	nominal	attribute
20	A2M	column...	nominal	attribute



En la salida del proceso *ReadCSV* se tendrá incorporados los datos del fichero *.dat* a *RapidMiner*, que una vez se haya comprobado que están correctamente se puede asignar la salida del *ReadCSV* a la entrada de *Store* y almacenar estos datos en un repositorio.

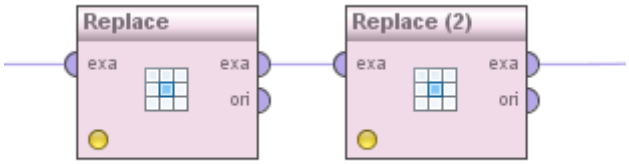
El resultado es el siguiente:

Row No.	ID	ABANDONA	TITULAT	SEXE	EDAT	FRANJA	SEMESTRE	NSEM	NA	NC	NASUP	NCSUP
1	12445859	0	0	0	28	3	20011	11	3	16.500	3	16.500
2	12355702	0	0	1	36	4	19992	8	6	31.500	5	25.500
3	12355902	0	0	1	48	5	19982	6	2	10.500	2	10.500

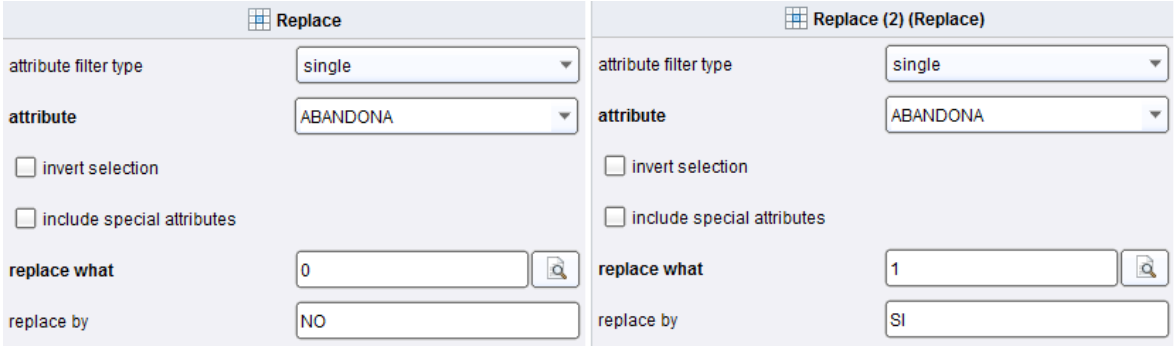
7.3.2 Transformaciones

En este punto se van a realizar diversas transformaciones que serán útiles para el posterior manejo de los datos.

El campo ABANDONA tiene dos valores el 0 para indicar que NO abandona y el 1 indica que SI abandona. Se sustituyen estos valores numéricos por NO y SI respectivamente.



Se encadenan dos objetos *Replace* para sustituir los dos valores como se acaba de indicar y se configuran de la siguiente manera:

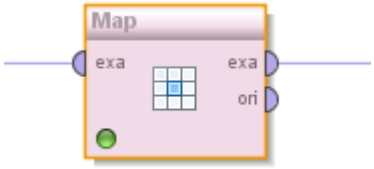


A la salida de este proceso ya tendremos todos los valores de este campo sustituidos.

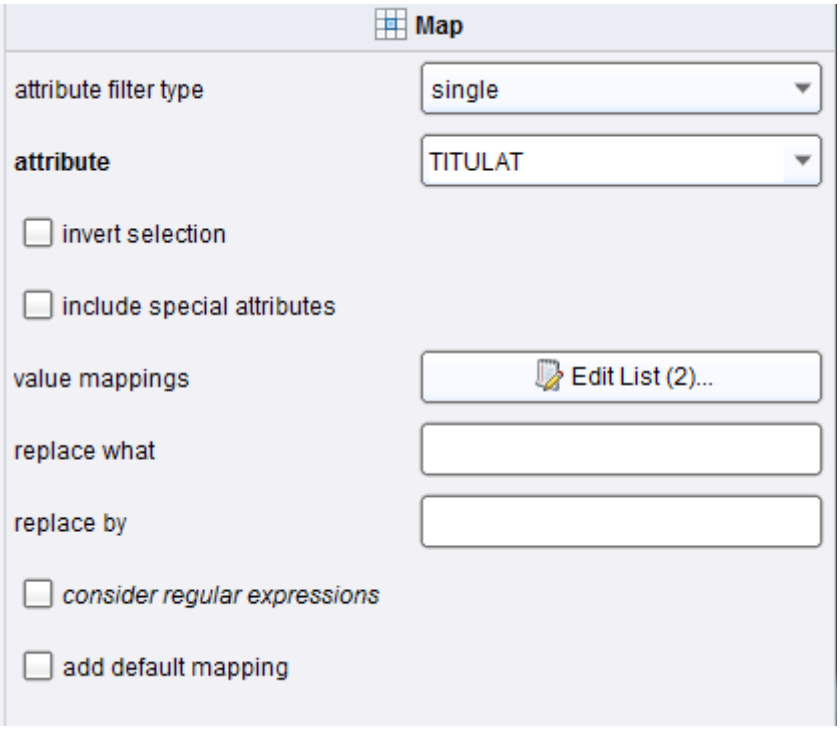
TITULAT

Este campo tiene dos valores, el 0 indica falso y el 1 verdadero

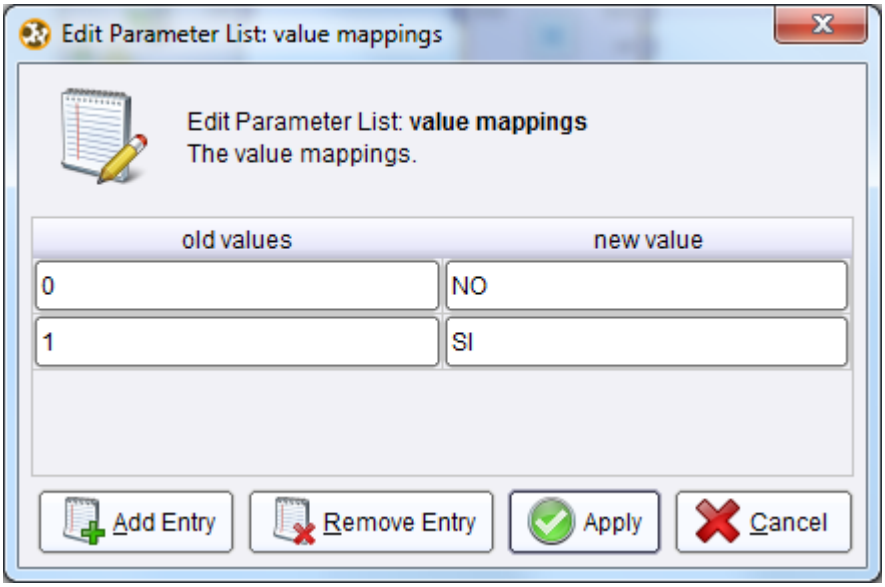
En este caso para sustituir los valores se usa el operador *MAP* que sera el que se use a partir de ahora para estos procedimientos:



Y se configurara de la siguiente manera:



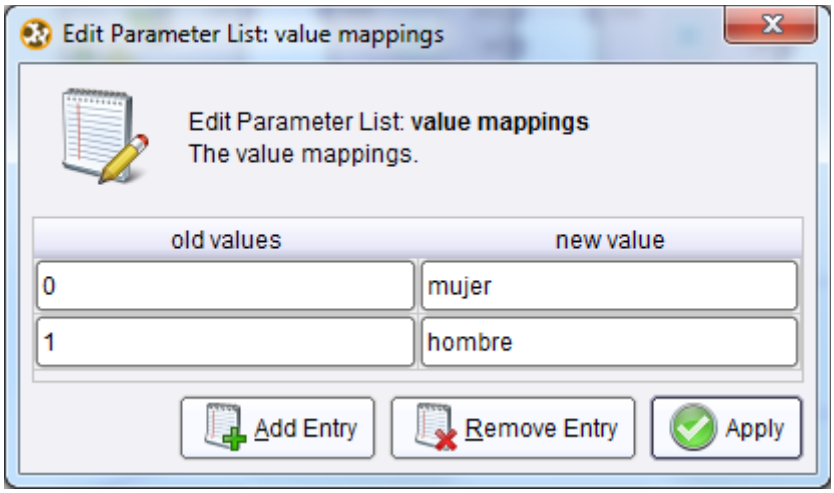
Editando el campo *value mappings* se consigue sustituir los valores de este campo en un solo paso:



SEXE

Este atributo toma dos valores 0 para mujer y 1 para hombre.

Se sustituyen los valores numéricos por el nombre de su categoría correspondiente.

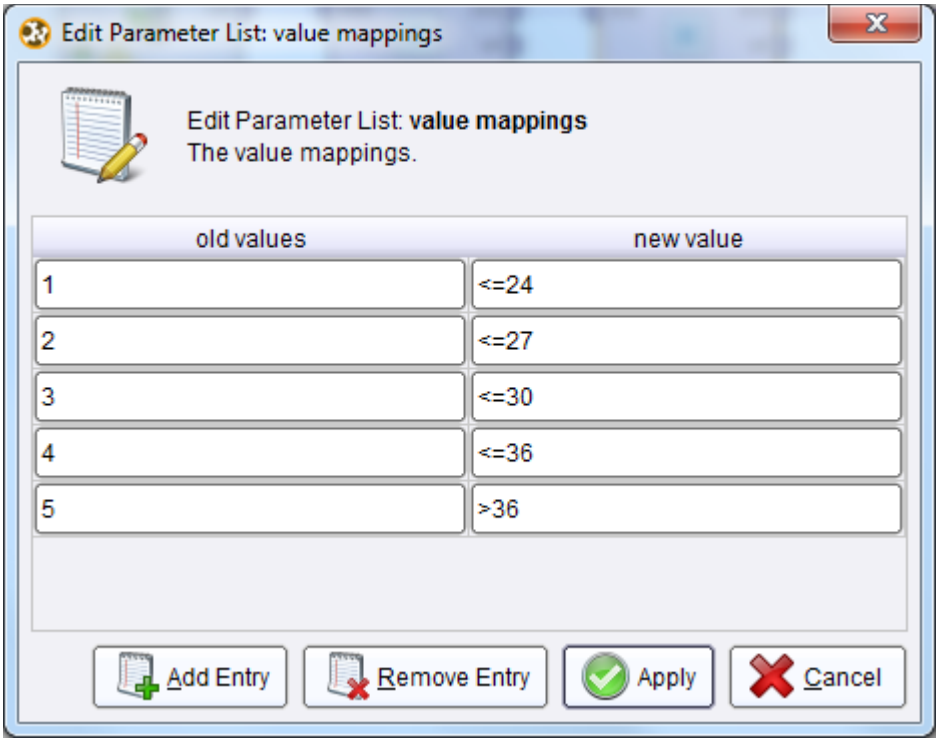


EDAT

Este campo sería susceptible de categorizar, dividiendo su rango en varias categorías. Como tenemos el campo FRANJA que es este mismo campo categorizado, de momento no vamos a hacer ninguna transformación con él.

FRANJA

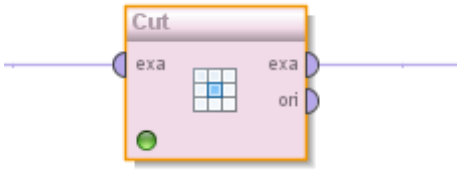
Se sustituyen los valores 1, 2, 3, 4, 5 por los nombres de sus categorías correspondientes, para ello se usará de nuevo el operador *MAP*:



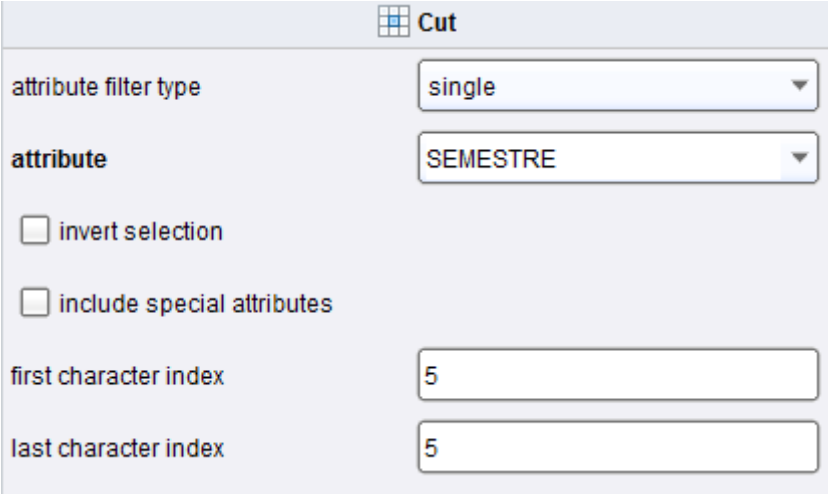
SEMESTRE

Este campo esta formado realmente por dos datos, que son el año y el semestre. En principio solo interesa la información sobre semestre con lo que se elimina de este campo la parte correspondiente al año. Si posteriormente se viera que la información sobre el año fuera interesante se volvería sobre este paso.

Para realizar esta transformación se usará el operador *CUT*:



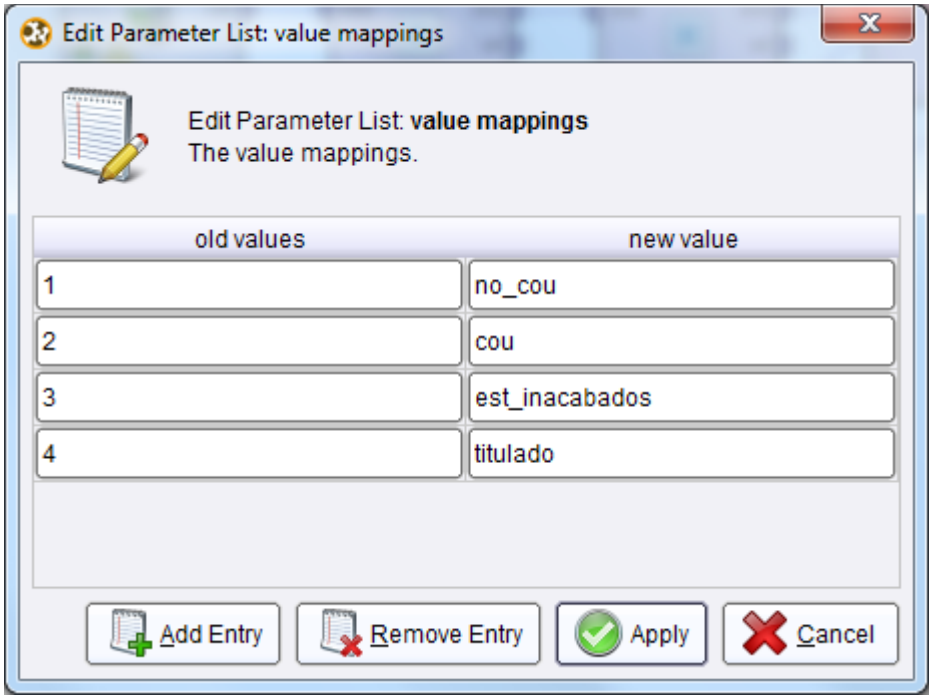
y se configurará de la siguiente manera:



De esta forma este campo a partir de ahora solo almacena la información sobre el número de semestre.

VIA

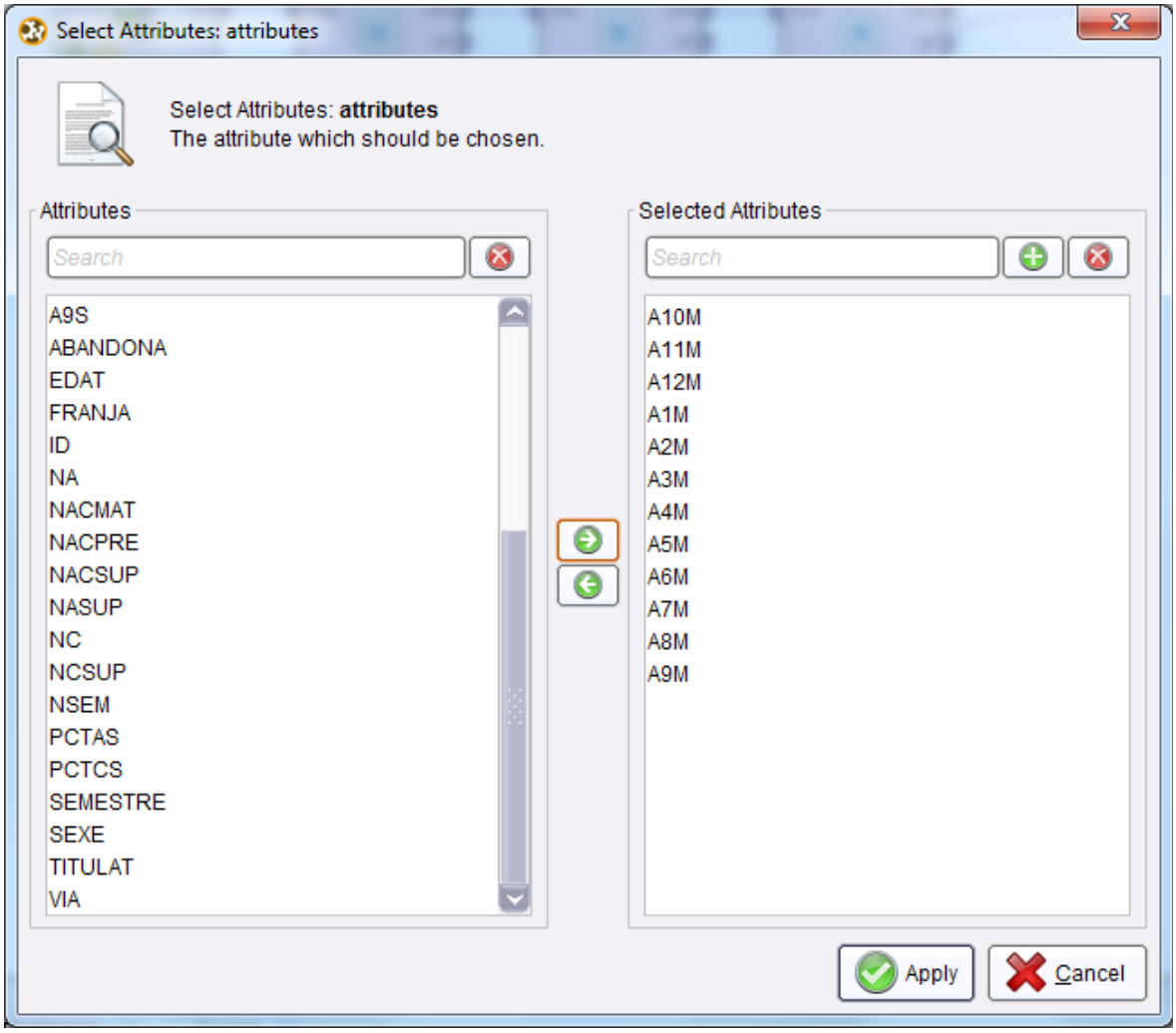
Este campo toma los valores 1, 2, 3 y 4 y se sustituyen estos valores por los de los nombres de las categorías que representan:



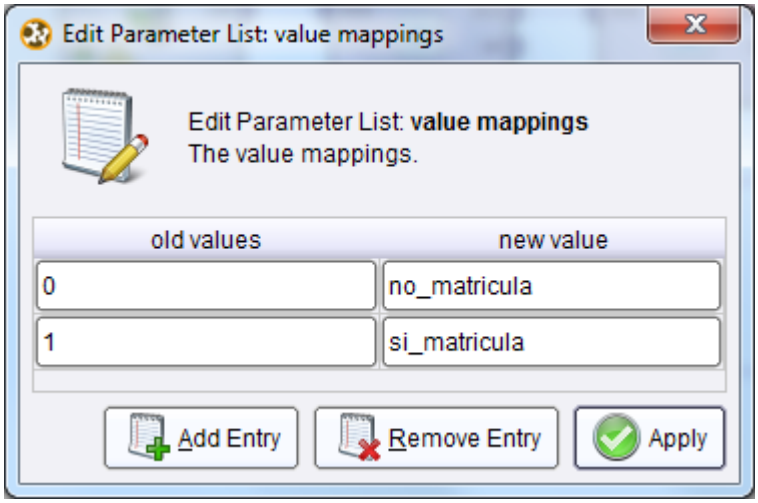
A1M – A12M

Los atributos de A1M a A12M pueden tomar los valores 0 o 1 dependiendo de si el alumno se matricula de la asignatura o no, se sustituyen eso valores por su valor correspondiente:

Para realizar este cambio en todos los campos a la vez se seleccionaran todos los atributos afectados por este cambio:

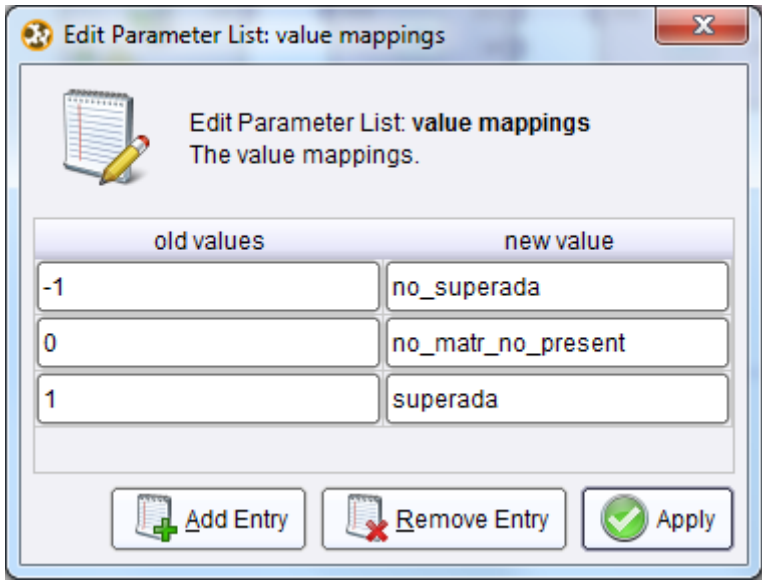


y se hace la siguiente modificación:



A1S – A12S

Los atributos de A1S a A12S pueden tomar los valores -1, 0 o 1 dependiendo si se supera la asignatura, no se presenta o no se supera, se sustituyen eso valores por su valor correspondiente:



Esto es un fragmento de como quedaría después de todas las transformaciones:

A11M	A12M	VIA	SEMESTRE	FRANJA	SEXE	TITULAT	ABANDONA	EDAT
no_matricula	no_matricula	titulado	1	<=30	mujer	NO	NO	28
no_matricula	no_matricula	titulado	2	<=36	hombre	NO	NO	36
no_matricula	no_matricula	titulado	2	>36	hombre	NO	NO	48
no_matricula	no_matricula	no_cou	2	>36	mujer	NO	NO	43
no_matricula	no_matricula	est_inacaba	1	<=27	mujer	NO	SI	27
si_matricula	no_matricula	est_inacaba	2	<=30	mujer	NO	NO	30

7.3.3 Descripción de los datos

En este apartado se van a realizar diversas visualizaciones y estadísticas de los datos del repositorio de manera que se pueda tener una idea mejor de la información con la que se cuenta.

A continuación se muestran los datos estadísticos de algunos de los valores:

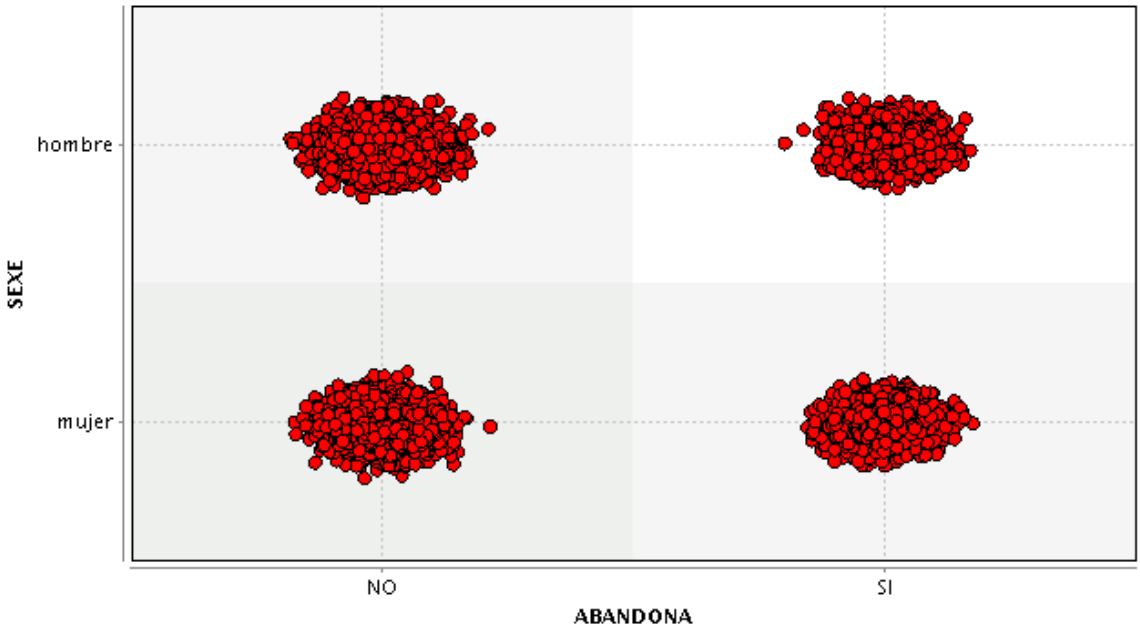
Name	Statistics
VIA	mode = est_inacabados (7609), least = cou (3111)
SEMESTRE	mode = 2 (9842), least = 1 (8710)
FRANJA	mode = <=27 (4393), least = >36 (2398)
SEXE	mode = hombre (9352), least = mujer (9200)
TITULAT	mode = NO (18548), least = SI (4)
ABANDONA	mode = NO (13865), least = SI (4687)
EDAT	avg = 29.245 +/- 6.241
NSEM	avg = 15.977 +/- 5.752
NA	avg = 3.098 +/- 1.034
NC	avg = 16.071 +/- 5.559
NASUP	avg = 1.786 +/- 1.430
NCSUP	avg = 9.173 +/- 7.486
PCTAS	avg = 0.590 +/- 0.423
PCTCS	avg = 0.586 +/- 0.425
NACMAT	avg = 2.768 +/- 1.113
NACPRE	avg = 1.793 +/- 1.390
NACSUP	avg = 1.624 +/- 1.366

En la siguiente tabla se pueden ver los rangos en los que se mueven los datos:

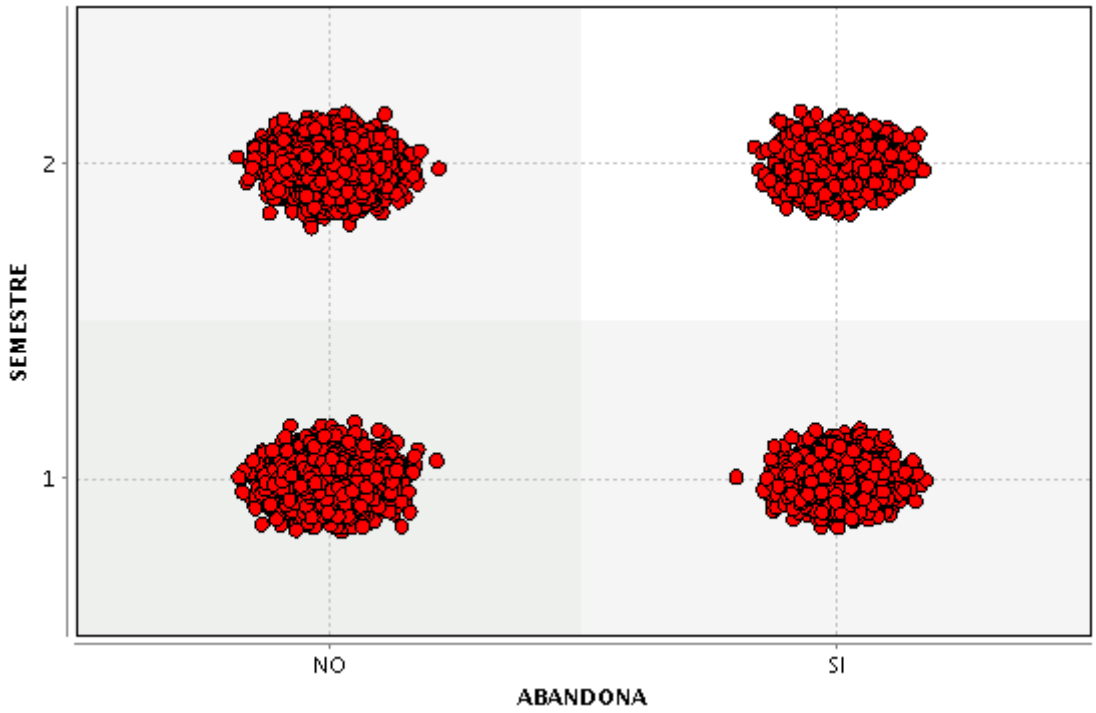
Name	Range
VIA	titulado (4045), no_cou (3787), est_inacabados (7609), cou (3111)
SEMESTRE	1 (8710), 2 (9842)
FRANJA	<=30 (3483), <=36 (3968), >36 (2398), <=27 (4393), <=24 (4310)
SEXE	mujer (9200), hombre (9352)
TITULAT	NO (18548), SI (4)
ABANDONA	NO (13865), SI (4687)
EDAT	[17.000 ; 75.000]
NSEM	[6.000 ; 25.000]
NA	[1.000 ; 13.000]
NC	[4.500 ; 69.000]
NASUP	[0.000 ; 10.000]
NCSUP	[0.000 ; 52.500]
PCTAS	[0.000 ; 1.000]
PCTCS	[0.000 ; 1.000]
NACMAT	[0.000 ; 10.000]
NACPRE	[0.000 ; 8.000]
NACSUP	[0.000 ; 7.000]

Con las siguientes visualizaciones se tiene una idea mas gráfica de los datos que se están estudiando.

Esta gráfica muestra la distribución de los abandonos en cuanto al sexo del alumno:

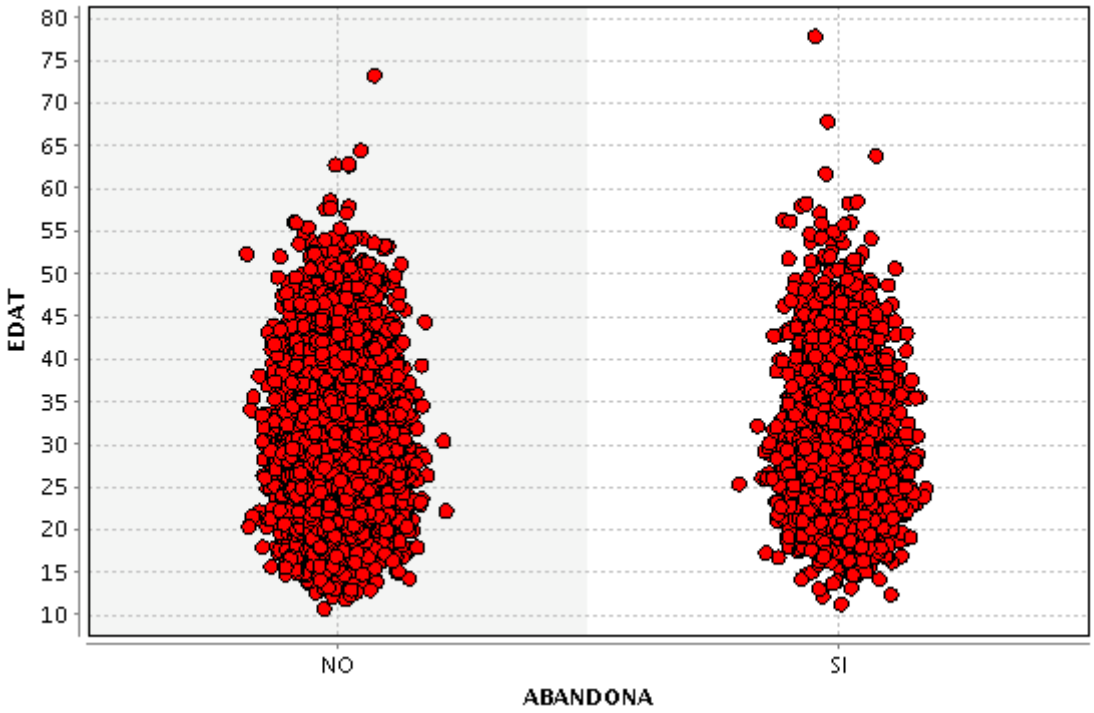


Puede apreciarse que la distribución es bastante homogénea.



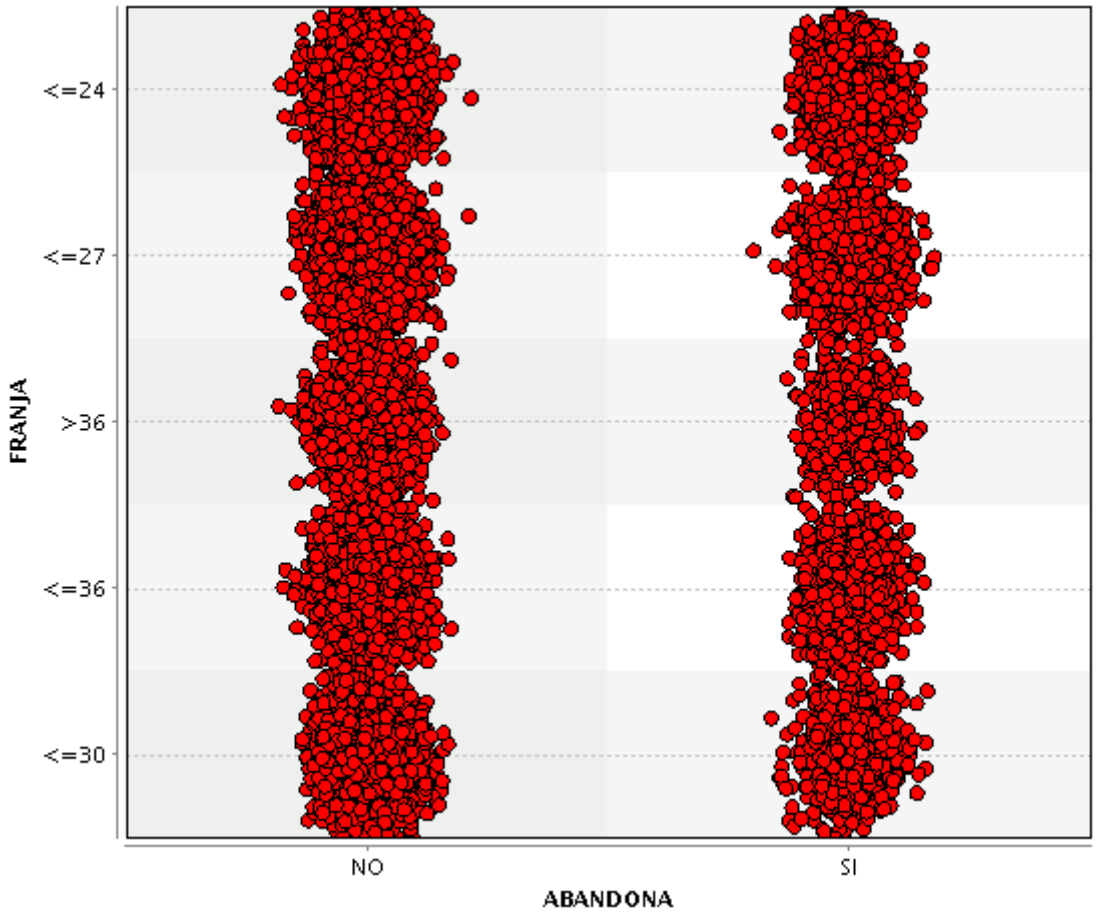
La distribución por semestres es muy similar a la anterior, también es muy homogénea.

El siguiente gráfico muestra la distribución por edades:



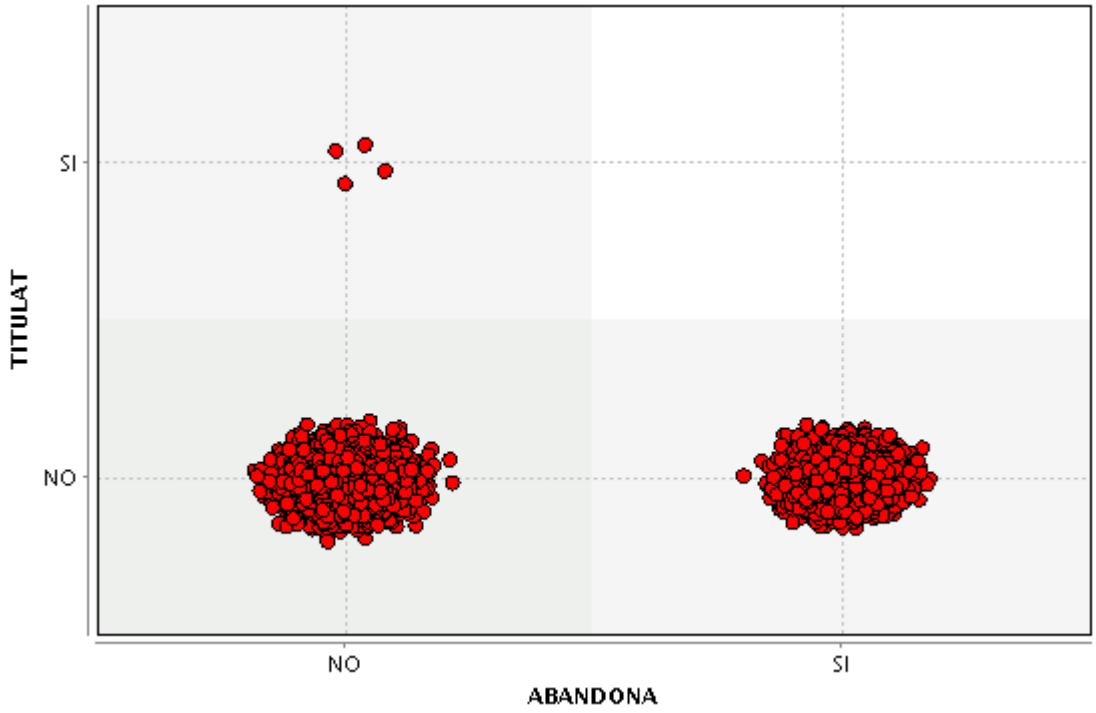
En este caso se descubre que atendiendo a la edad hay varios elementos extremos principalmente entre los máximos. La mayor parte de los alumnos tienen una edad inferior a 55 años, por encima de estas edades hay muy pocos estudiantes, con lo que para realizar nuestro estudio se pueden eliminar estos elementos, mejorando la claridad de los resultados.

El siguiente gráfico muestra la distribución por franja de edades:



En este caso la distribución es bastante homogénea.

El siguiente gráfico muestra la distribución de los abandonos en función de si se titulan después del primer semestre.



En este caso puede verse que los que se titulan después del primer semestre son una gran minoría y ademas, por supuesto, no abandonan, por lo tanto para el estudio se puede prescindir de ellos.

7.4 Elaboración del modelo

Después de realizar los apartados anteriores se ha conseguido tener una idea de los datos con los que se cuentan para realizar el estudio.

Como se ha visto los datos son de diversa naturaleza, en este apartado se van a desarrollar varios modelos con el fin de alcanzar los objetivos propuestos.

7.4.1 Modelo de clusters sobre los datos de matriculación

Los datos con los que se cuenta son bastante limitados. Aun así, a simple vista no es fácil dar con la causa que motiva el abandono de los estudiantes si es que la hay (y se puede llegar a ella a partir de los datos de las matrículas).

En este primer modelo que se va a desarrollar se van a tener en cuenta cada una de las asignaturas de las que un alumno se ha matriculado, tratando de encontrar similitudes entre asignaturas de las que se matricula cada alumno. Se va a desarrollar un modelo de clusters con el que se pretenden agrupar aquellos comportamientos similares de los alumnos en cuanto a matriculación de asignaturas se refiere.

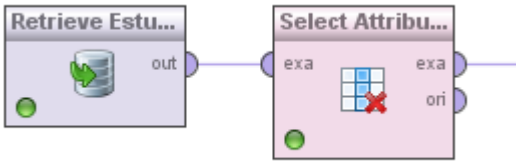
Una vez hechas las agrupaciones, se quiere ver si los alumnos agrupados en cada uno de los clusters tienen un comportamiento similar también en cuanto al abandono de los estudios

Para realizar este agrupamiento se va utilizar el algoritmo k-means. Este algoritmo calculará la medida de similitud que hay entre cada uno de los alumnos y en función de ella pondrá en el mismo grupo aquellos que sean mas parecidos.

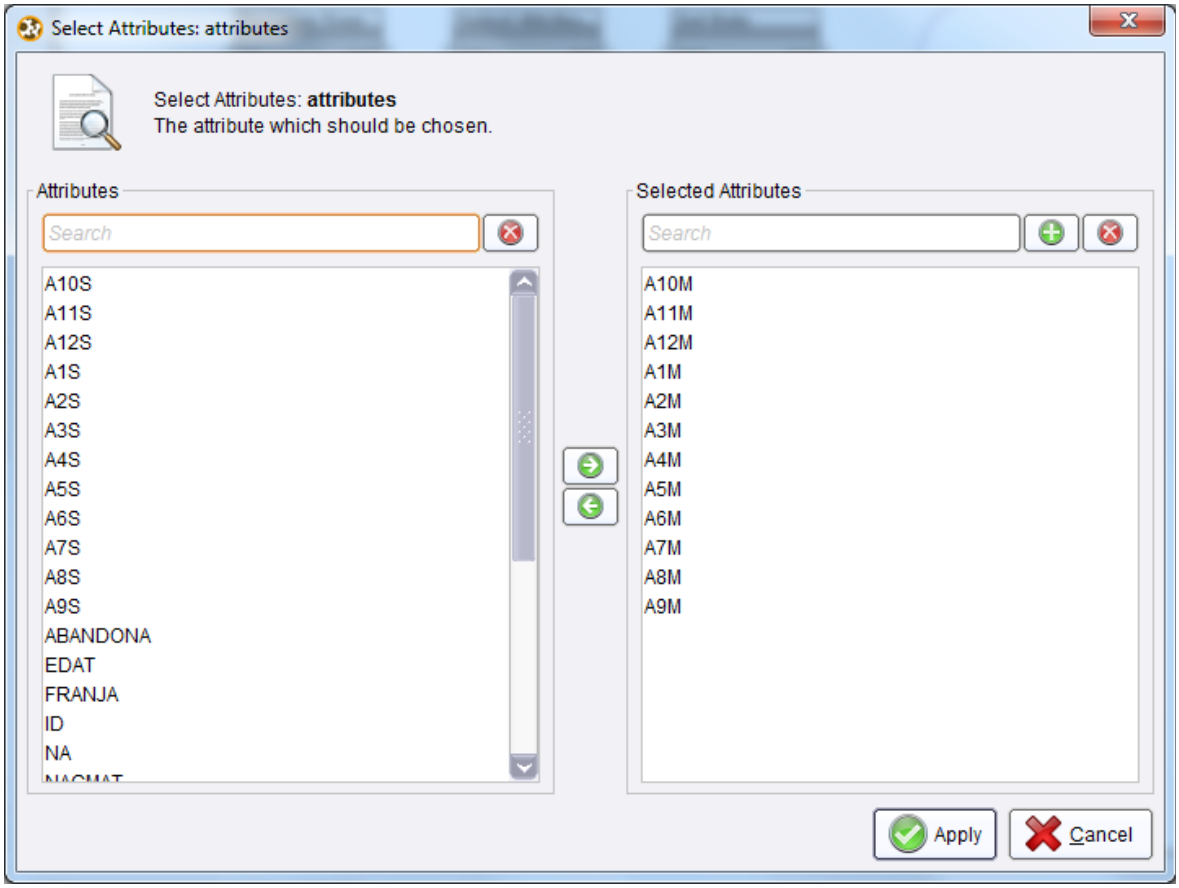
La medida de similitud entre los valores se hace a partir del cálculo de las distancias que hay entre los mismos. En el caso que nos ocupa, al tratarse de valores nominales este cálculo se hace por medio de distancias Hamming, que consiste en contar el número de atributos diferentes que hay entre los objetos, a mayor número de atributos diferentes entre dos objetos, mayor será su distancia.

Se eligen los campos A1M, A2M, A3M, A4M, A5M, A6M, A7M, A8M, A9M, A10M, A11M, A12M, que son los atributos que representan las doce asignaturas mas comunes del primer semestre. Estos campos tenían como valores iniciales 0 o 1 pero en la fase de transformación se han cambiado por valores nominales si_matricula/no_matricula.

Los datos se encuentran almacenados en un repositorio a cuya salida se pone un objeto *Select* para seleccionar los atributos citados anteriormente.



Y se eligen los atributos a seleccionar:



El resultado del *Select* es el siguiente:

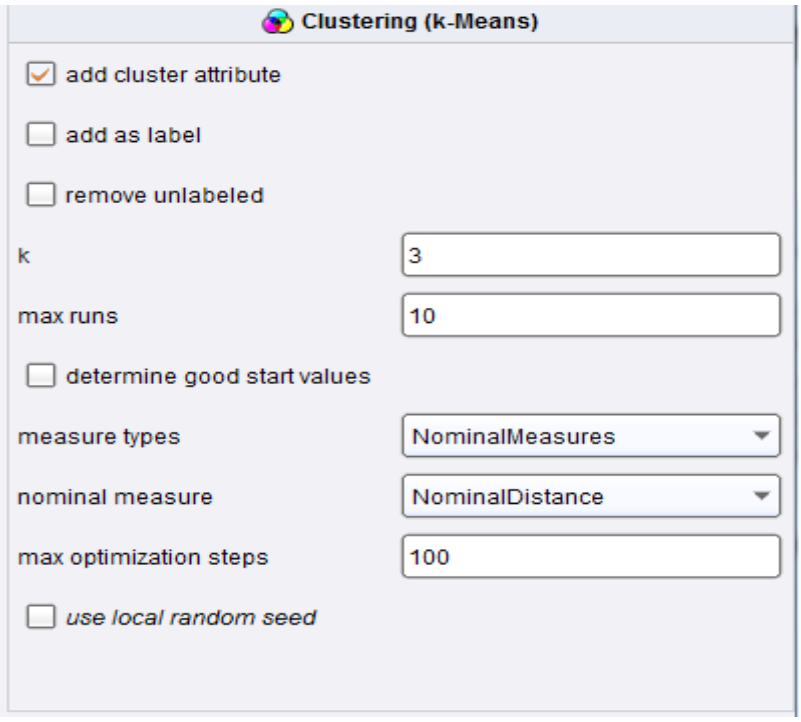
Row No.	ID	A1M	A2M	A3M	A4M	A5M	A6M	A7M	A8M	A9M	A10M	A11M	A12M
1	12445859	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
2	12355702	si_matricula	si_matricula	si_matricula	si_matricula	si_matricula	no_matricula	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
3	12355902	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
4	12450601	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
5	12450792	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
6	12450837	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	si_matricula	no_matricula
7	12450851	no_matricula	no_matricula	no_matricula	no_matricula	si_matricula	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
8	12451313	no_matricula	no_matricula	si_matricula	no_matricula	si_matricula	no_matricula	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula

Este resultado es el que se pasará como origen de datos al objeto *Cluster*.

Se elige el objeto cluster *k-Means* que es uno de los algoritmos mas populares de agrupación.

Se fija un número de clusters a calcular de 3, se han realizado varias pruebas y se ha llegado a la conclusión de que con 3 clusters es como mejor se realizan las agrupaciones. Por encima de este número en los resultados se encontraban clusters vacíos.

Como los datos con los que se van a trabajar son de tipo nominal las distancias como ya se comentó anteriormente no pueden ser euclidianas. Se seleccionan las opciones *NominalMeasures* y *NominalDistance*:



A continuación se muestran los resultados obtenidos:

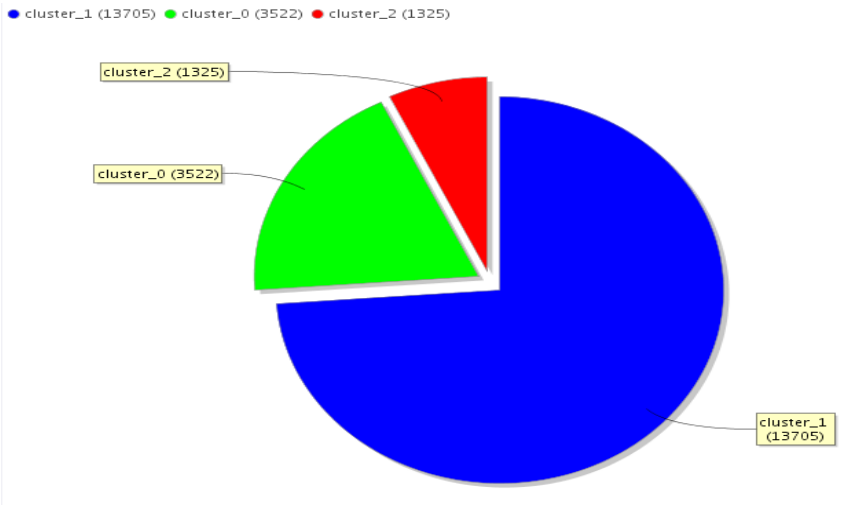
Row No.	ID	cluster	A1M	A2M	A3M	A4M	A5
15522	12862700	cluster_0	si_matricula	no_matricula	no_matricula	no_matricula	no_matri
15523	12862708	cluster_0	si_matricula	no_matricula	no_matricula	no_matricula	si_matri
15524	12862720	cluster_2	no_matricula	no_matricula	no_matricula	no_matricula	no_matri
15525	12862721	cluster_2	no_matricula	no_matricula	no_matricula	no_matricula	no_matri
15526	12862723	cluster_2	si_matricula	no_matricula	no_matricula	no_matricula	no_matri
15527	12862725	cluster_1	si_matricula	si_matricula	si_matricula	si_matricula	no_matri
15528	12862728	cluster_1	si_matricula	si_matricula	no_matricula	no_matricula	no_matri
15529	12862736	cluster_1	si_matricula	no_matricula	si_matricula	no_matricula	si_matri

En la tabla se ve como se ha añadido una nueva columna con el título cluster en la que aparece el nombre del cluster al que pertenece cada fila después de la agrupación.

Name	Range
ID	12355702 (1), 12355902 (1), 12365689 (1), 12365709
cluster	cluster_1 (13705), cluster_0 (3522), cluster_2 (1325)
A1M	no_matricula (4327), si_matricula (14225)
A2M	no_matricula (14424), si_matricula (4128)
A3M	no_matricula (11906), si_matricula (6646)
A4M	no_matricula (12670), si_matricula (5882)
A5M	no_matricula (12618), si_matricula (5934)
A6M	no_matricula (14702), si_matricula (3850)
A7M	no_matricula (14625), si_matricula (3927)
A8M	no_matricula (16217), si_matricula (2335)
A9M	no_matricula (17160), si_matricula (1392)
A10M	no_matricula (17479), si_matricula (1073)
A11M	no_matricula (17487), si_matricula (1065)
A12M	no_matricula (17655), si_matricula (897)

En el cuadro anterior se ve el numero de alumnos que se ha agrupado en cada cluster. En el cluster_1 se han agrupado 13.705 alumnos mientras que en el cluster_0 son 3.522 y en el cluster_2 1.325.

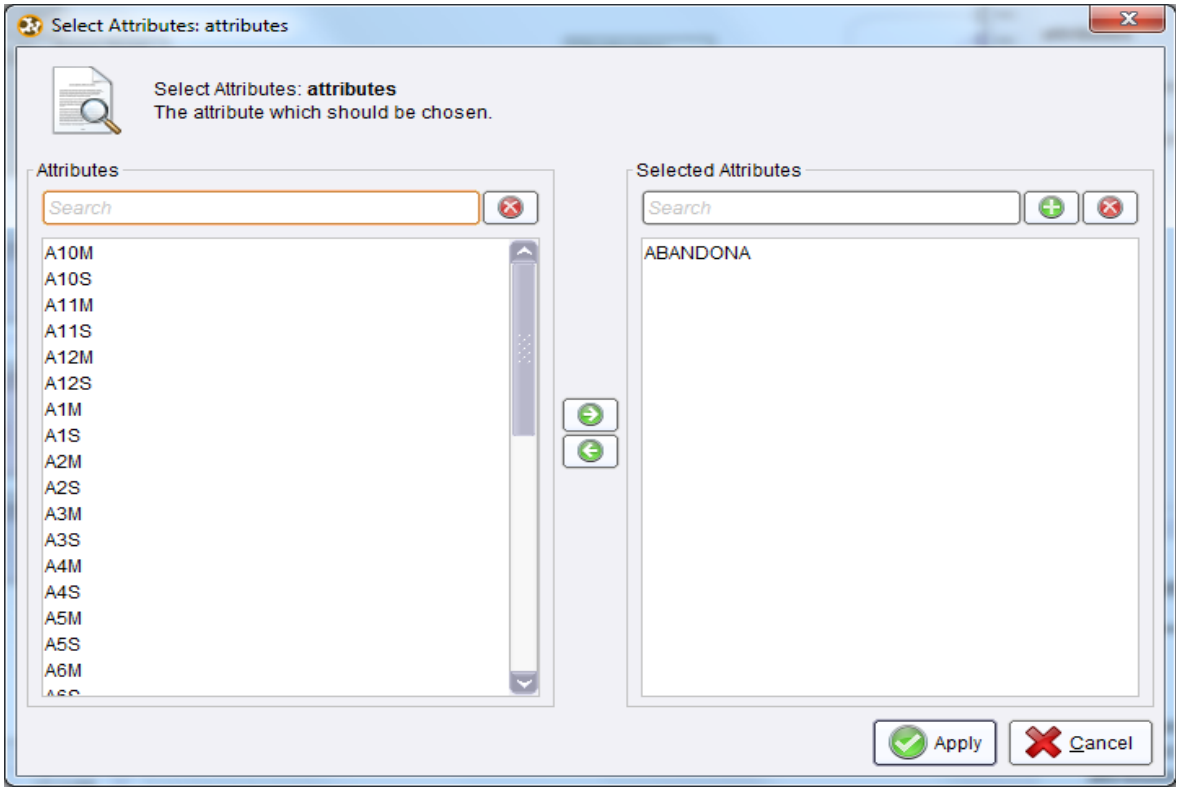
Las distribuciones de las agrupaciones se pueden ver mas claramente en el siguiente gráfico:



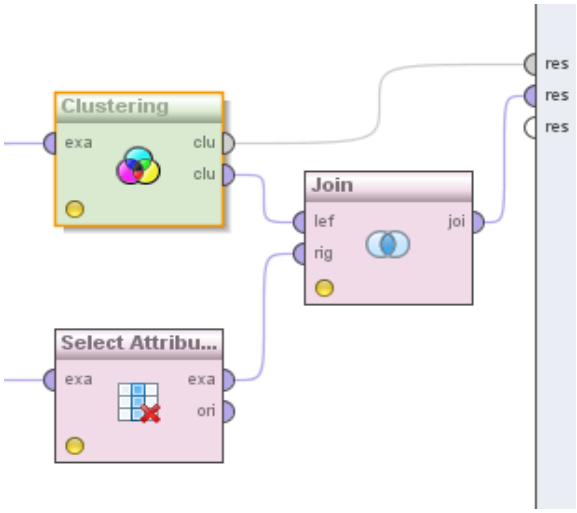
Se aprecia que el cluster_1 es mucho mas grande que los otros dos, lo que puede llevar a pensar que las agrupaciones no están muy bien hechas, pero si se miran las estadísticas mostradas en la fase de descripción de los datos se ve que el número de alumnos que abandona los estudios es de 4.687 y el de los alumnos que no abandonan es 13.685 estos datos son muy parecidos a los obtenidos en los clusters con lo que se va a seguir adelante por este camino.

Lo que interesa saber es como se distribuyen los abandonos entre los estudiantes en cada uno de los clusters. Si se viera una clara tendencia al abandono o al no abandono de los estudiantes de alguno de los clusters se podría ver que características comunes les une para encontrar el motivo por el cual el abandono o no abandono de ese grupo es mas marcado.

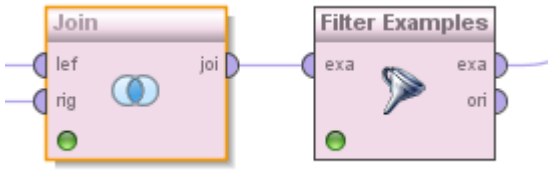
Es necesario añadir a los resultados obtenidos de los cluster el campo ABANDONA para hacerlo es necesario seleccionar del repositorio inicial el campo ABANDONA con un *Select*:



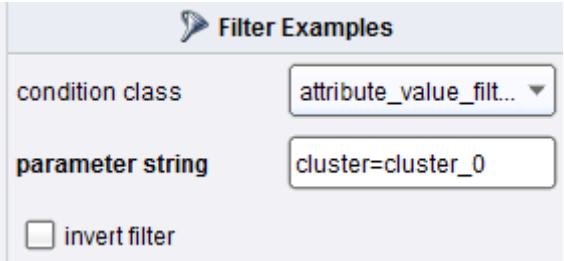
Y posteriormente usar un *Join* para añadir los valores a los resultados:



Para poder ver los resultados por cada cluster se usa un *filtro*:



Por cada cluster hay que indicar en el filtro el valor a filtrar, en la imagen se filtra por el cluster_0.

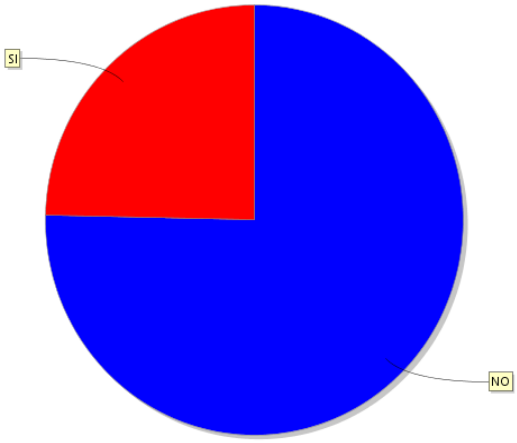


Para el resto de clusters se haría de la misma forma.

A continuación se muestran los valores de cada cluster:

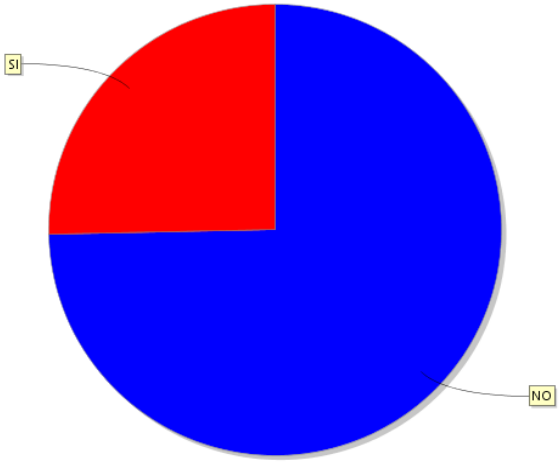
Cluster_0

Role	Name	Type	Statistics	Range
regular	ABANDONA	nominal	mode = NO (2654), least = SI (868)	NO (2654), SI (868)



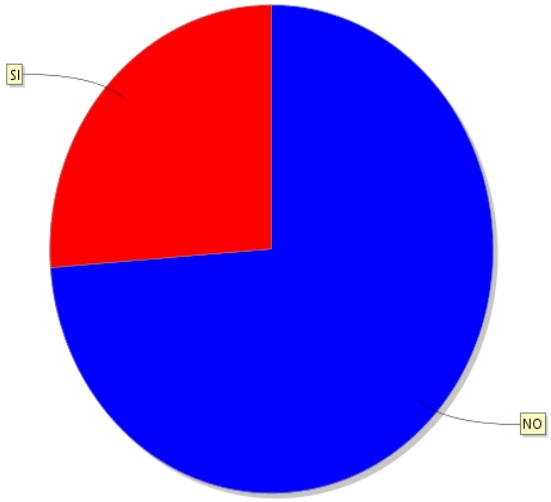
Cluster_1

Role	Name	Type	Statistics	Range
regular	ABANDONA	nominal	mode = NO (10233), least = SI (3472)	NO (10233), SI (3472)



Cluster_2

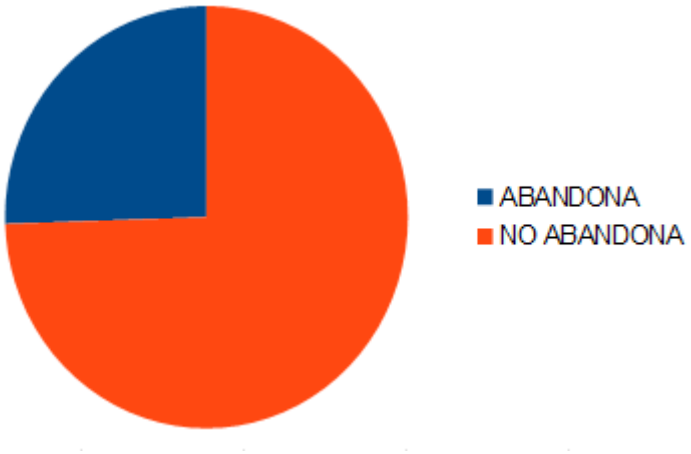
Role	Name	Type	Statistics	Range
regular	ABANDONA	nominal	mode = NO (978), least = SI (347)	NO (978), SI (347)



7.4.2 Conclusiones sobre el modelo obtenido

Los resultados obtenidos muestran que en el cluster_0 los alumnos que abandonan son aproximadamente un cuarto de los alumnos matriculados. En el resto de clusters se aprecia lo mismo, aproximadamente en todos ellos la cuarta parte de los alumnos abandona y las tres cuartas partes de ellos no abandona.

En los datos que se tienen de origen puede verse que los alumnos que abandonan los estudios son 4.687 y los alumnos que no abandonan son 13.685. Gráficamente quedaría así:



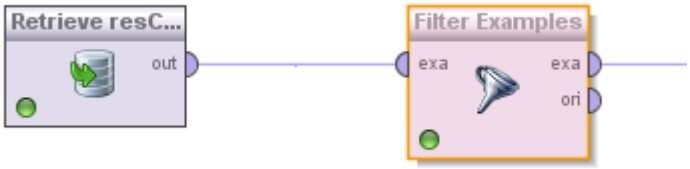
Al ver estos datos se descubre que la distribución obtenida en cada uno de los clusters es prácticamente la misma que había en los datos originales, con lo que las similitudes en cuanto a la matriculación de asignaturas que ha propiciado el agrupamiento de los estudiantes no parece que de momento tengan una clara relación con el abandono de los mismos.

7.4.3 Modelo de cluster y árbol de decisión sobre los datos de matriculación

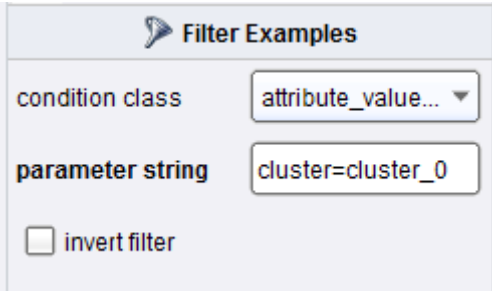
Este modelo va a partir de los datos calculados del modelo anterior. Teniendo los datos de los alumnos agrupados en clusters se va a construir un árbol de decisión por cada uno de ellos para intentar descubrir si existe algún patrón de comportamiento que indique, por ejemplo, que los alumnos que se matriculan de una asignatura A y de otra B tienen una probabilidad de abandonar muy alta y en cambio los alumnos que se matriculan de una asignatura B y otra C sus probabilidades de abandonar son bajas.

Teniendo como punto de partida los datos de las asignaturas agrupadas en clusters:

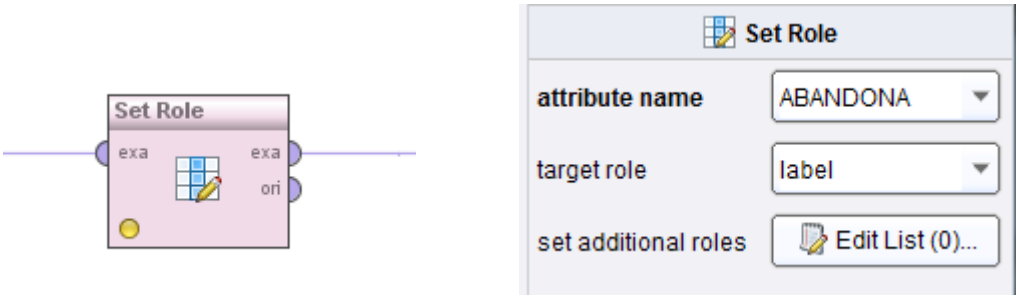
ID	cluster	ABANDONA	A1M	A2M	A3M	A4M	A5M	A6M	A7M	A8M	A9M	A10M	A11M	A12M
12445859	cluster_1	NO	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
12355702	cluster_1	NO	si_matricula	si_matricula	si_matricula	si_matricula	si_matricula	no_matricula	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
12355902	cluster_1	NO	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
12450601	cluster_0	NO	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
12450792	cluster_1	SI	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
12450837	cluster_1	NO	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	si_matricula	no_matricula
12450851	cluster_0	NO	no_matricula	no_matricula	no_matricula	no_matricula	si_matricula	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
12451313	cluster_1	NO	no_matricula	no_matricula	si_matricula	no_matricula	si_matricula	no_matricula	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula
12451317	cluster_1	NO	si_matricula	si_matricula	no_matricula	si_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula	no_matricula



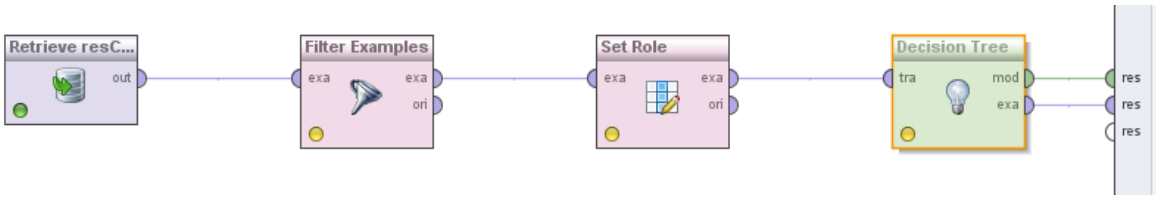
Se coloca un filtro a continuación para seleccionar los datos de cada uno de los clusters:



La finalidad es crear un modelo de clasificación que prediga el valor del atributo objetivo, en este caso este atributo es ABANDONA, así que para que el árbol de decisión se genere teniendo esto en cuenta hay que cambiar el role del atributo ABANDONA y ponerlo como *label*. Esto se hace por medio de *Set Role*:



Finalmente se añade el árbol de decisión:



El árbol que se va a generar tendrá unas características concretas, a continuación se describe cada una de ellas:

Criterio: se selecciona “accuracy” para que los atributos se seleccionen de forma que la exactitud del árbol sea la máxima posible.

El tamaño mínimo del subconjunto que se esta evaluando para realizar una división se establece en 8. Quiere decir que si un nodo tiene menos instancias no se dividirá mas.

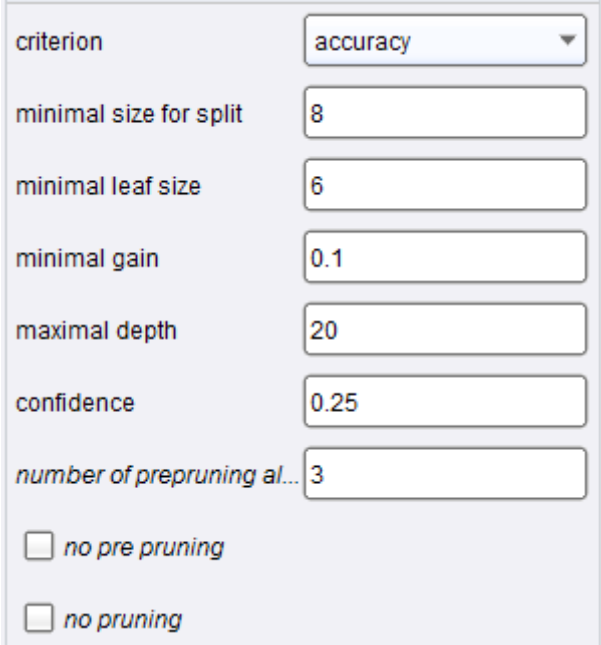
El tamaño mínimo de la hoja se establece en 6. Todas las hojas del árbol tendrán como mínimo 6 instancias.

La ganancia mínima se sitúa en 0,1. La ganancia de un nodo se calcula antes de su división y solo se dividirá si su ganancia es mayor que 0,1.

La profundidad del árbol se deja en 20, dado que no se llega a tener tantos niveles en el árbol que se va a generar este valor no le afecta, también se podría

poner el valor -1 para indicar que no se aplicarán limites en cuanto a la profundidad del árbol.

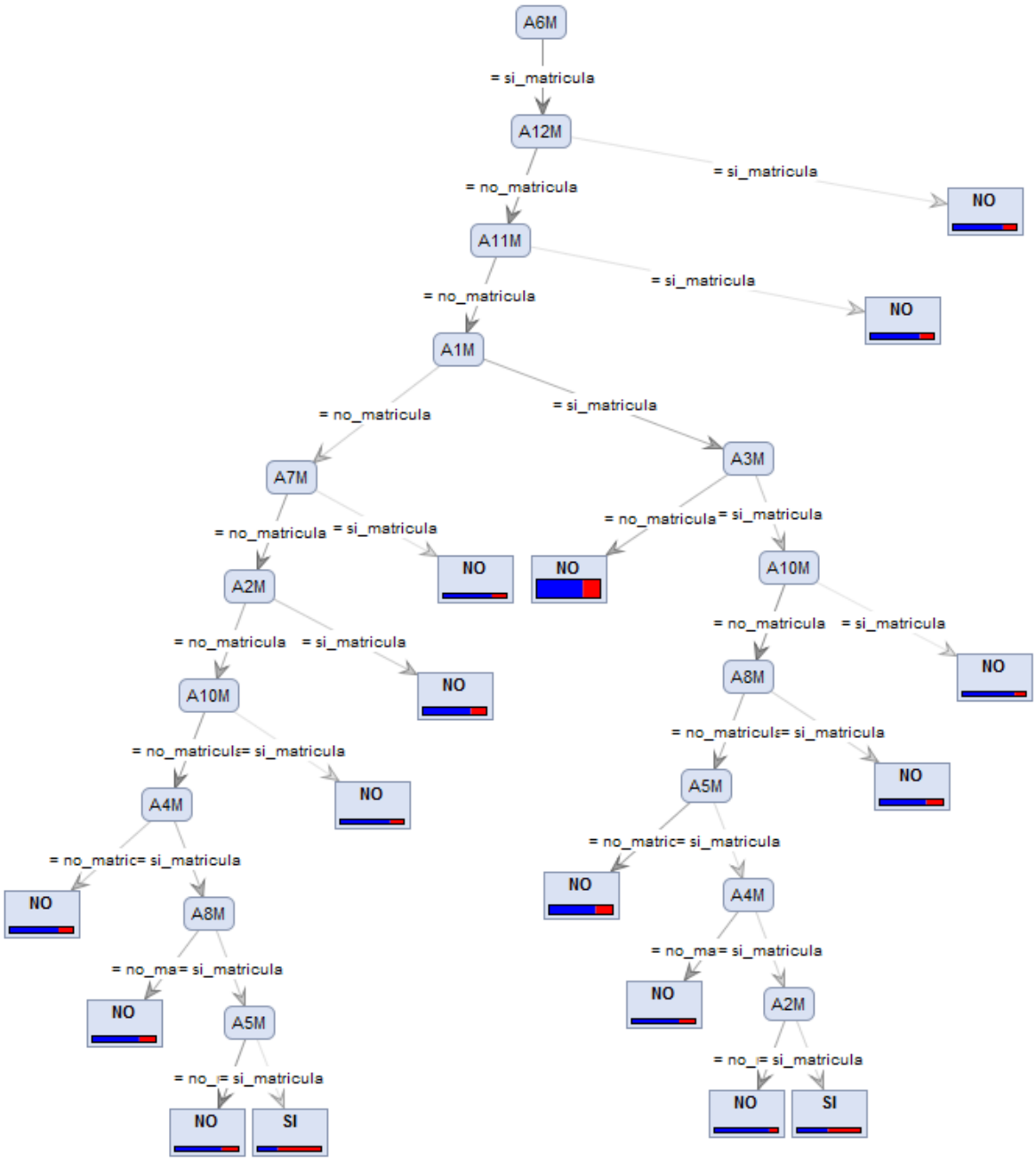
El nivel de confianza se establece en 0,25, este valor se usa como el mínimo nivel de confianza en el caso mas desfavorable a la hora de hacer la división.



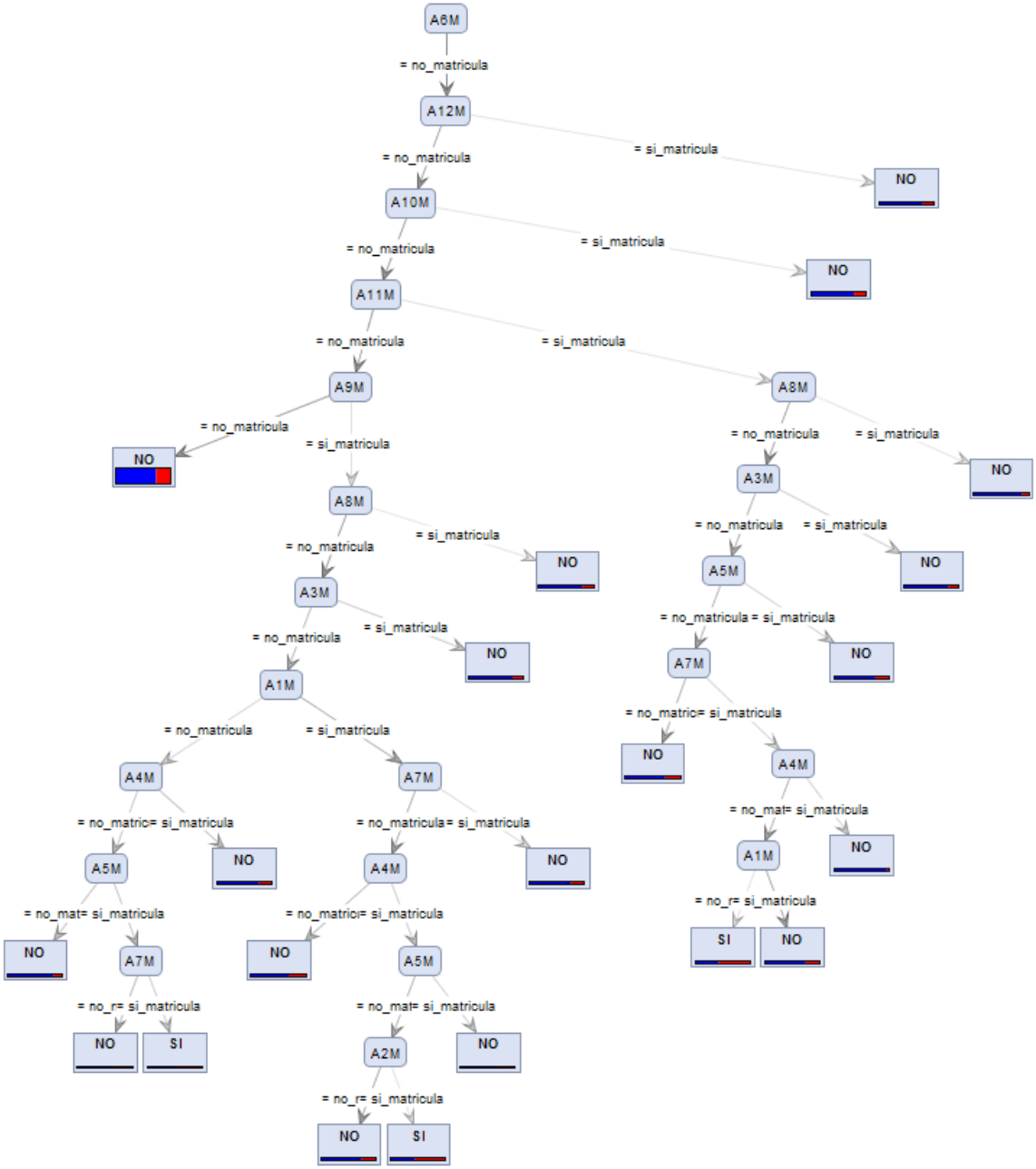
Todos estos valores se tienen en cuenta a la hora de generar el árbol y en función de ello se realizará la poda del mismo para que los valores alcanzados no exceda de los valores prefijados. Esto garantiza que el árbol generado no tengan mas niveles ni particiones de los que sean necesarios para alcanzar un buen nivel de predicción, lo que facilita ademas su posterior interpretación.

A continuación se muestran los resultados obtenidos:

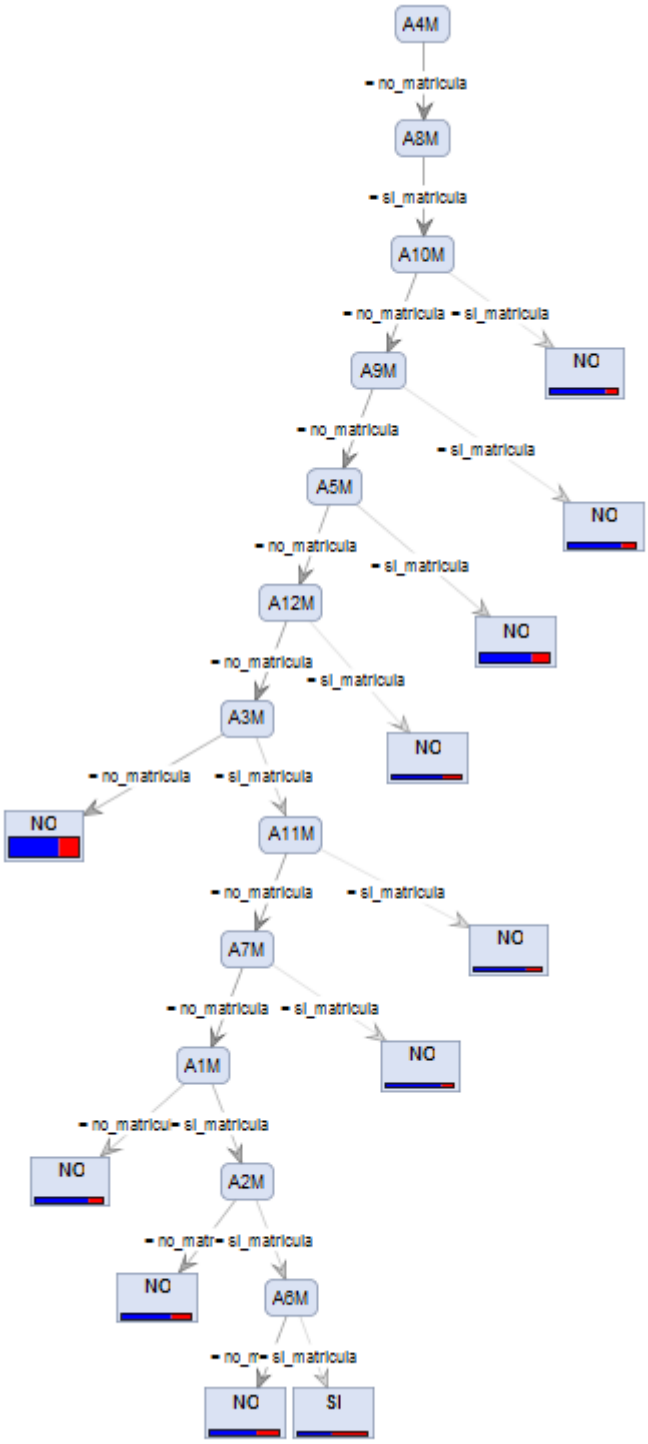
Árbol obtenido con el subconjunto de datos del Cluster_0



Árbol obtenido con el subconjunto de datos del Cluster_1



Árbol obtenido con el subconjunto de datos del Cluster_2



7.4.4 Conclusiones sobre el modelo obtenido

A continuación se muestran los arboles en modo texto junto con el error obtenido en la clasificación:

Cluster_0

```
A6M = si_matricula
|   A12M = no_matricula
|   |   A11M = no_matricula
|   |   |   A1M = no_matricula
|   |   |   |   A7M = no_matricula
|   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   A4M = no_matricula: NO {NO=175, SI=49}
                                                        ERROR 21,88%
|   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula: NO {NO=158, SI=52}
                                                        ERROR 24,76%
|   |   |   |   |   |   |   |   A8M = si_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=72, SI=24}
                                                        ERROR 25%
|   |   |   |   |   |   |   |   |   A5M = si_matricula: SI {NO=2, SI=4}
                                                        ERROR 33,33%
|   |   |   |   |   |   A10M = si_matricula: NO {NO=20, SI=5} ERROR 20%
|   |   |   |   |   A2M = si_matricula: NO {NO=231, SI=70} ERROR 23,26%
|   |   |   |   A7M = si_matricula: NO {NO=74, SI=21} ERROR 22,11%
|   |   |   A1M = si_matricula
|   |   |   |   A3M = no_matricula: NO {NO=1040, SI=368} ERROR 26,14%
|   |   |   |   A3M = si_matricula
|   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   A5M = no_matricula: NO {NO=371, SI=128}
                                                        ERROR 25,65%
|   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   A4M = no_matricula: NO {NO=56, SI=18}
                                                        ERROR 24,32%
|   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   A2M = no_matricula: NO {NO=14, SI=2}
                                                        ERROR 12,50%
|   |   |   |   |   |   |   |   |   A2M = si_matricula: SI {NO=3, SI=3}
                                                        ERROR 50%
|   |   |   |   |   |   A8M = si_matricula: NO {NO=144, SI=49}
                                                        ERROR 25,39%
|   |   |   |   |   A10M = si_matricula: NO {NO=5, SI=1} ERROR 16,67%
|   |   A11M = si_matricula: NO {NO=152, SI=40} ERROR 20,83%
|   A12M = si_matricula: NO {NO=137, SI=34} ERROR 19,88%
```

Cluster_1

```
A6M = no_matricula
|   A12M = no_matricula
|   |   A10M = no_matricula
|   |   |   A11M = no_matricula
|   |   |   |   A9M = no_matricula: NO {NO=7840, SI=2766} ERROR 26,08%
|   |   |   |   A9M = si_matricula
|   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   A1M = no_matricula
|   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=29, SI=6}
                                                          ERROR 17,14%
|   |   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO {NO=16,
                                                      SI=8} ERROR 33,33%
|   |   |   |   |   |   |   |   |   |   A7M = si_matricula: SI {NO=7,
                                                      SI=7} ERROR 50%
|   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=54, SI=16}
                                                          ERROR 22,86%
|   |   |   |   |   |   |   A1M = si_matricula
|   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   A4M = no_matricula: NO {NO=135,
                                                      SI=56} ERROR 29,32%
|   |   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A2M = no_matricula: NO
                                          {NO=65, SI=22}ERROR 25,29%
|   |   |   |   |   |   |   |   |   |   |   A2M = si_matricula: SI {NO=5,
                                                      SI=6} ERROR 45,45%
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=33,
                                                      SI=15} ERROR 31,25%
|   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=55, SI=18}
                                                          ERROR 24,66%
|   |   |   |   |   |   A3M = si_matricula: NO {NO=250, SI=64}
                                                          ERROR 20,38%
|   |   |   |   |   A8M = si_matricula: NO {NO=24, SI=6}
                                                          ERROR 20%
|   |   |   A11M = si_matricula
|   |   |   |   A8M = no_matricula
|   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   A7M = no_matricula: NO {NO=209, SI=79}
                                                          ERROR 27,43%
|   |   |   |   |   |   |   A7M = si_matricula
|   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   A1M = no_matricula: SI {NO=3, SI=4}
                                                          ERROR 42,86%
|   |   |   |   |   |   |   |   |   A1M = si_matricula: NO {NO=44, SI=15}
                                                          ERROR 25,42%
|   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=20, SI=1}
                                                          ERROR 4,76%
|   |   |   |   |   |   A5M = si_matricula: NO {NO=99, SI=34}
                                                          ERROR 25,56%
```

```
|   |   |   |   |   A3M = si_matricula: NO {NO=167, SI=41} ERROR 19,71%
|   |   |   |   A8M = si_matricula: NO {NO=22, SI=3}       ERROR 12%
|   |   A10M = si_matricula: NO {NO=646, SI=174}           ERROR 21,22%
|   A12M = si_matricula: NO {NO=510, SI=131}               ERROR 20,44%
```

Cluster_2

```
A4M = no_matricula
|   A8M = si_matricula
|   |   A10M = no_matricula
|   |   |   A9M = no_matricula
|   |   |   |   A5M = no_matricula
|   |   |   |   |   A12M = no_matricula
|   |   |   |   |   |   A3M = no_matricula: NO {NO=396, SI=159}
                                                          ERROR 28,65%
|   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   A11M = no_matricula
|   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   A1M = no_matricula: NO {NO=74, SI=21}
                                                          ERROR 22,11%
|   |   |   |   |   |   |   |   |   A1M = si_matricula
|   |   |   |   |   |   |   |   |   |   A2M = no_matricula: NO {NO=69,
                                                    SI=28} ERROR 28,87%
|   |   |   |   |   |   |   |   |   |   A2M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A6M = no_matricula: NO
                                          {NO=29, SI=14} ERROR 35,56%
|   |   |   |   |   |   |   |   |   |   |   A6M = si_matricula: SI {NO=4,
                                                    SI=4} ERROR 50%
|   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=13, SI=3}
                                                          ERROR 18,75%
|   |   |   |   |   |   |   A11M = si_matricula: NO {NO=10, SI=3}
                                                          ERROR 23,08%
|   |   |   |   |   A12M = si_matricula: NO {NO=47, SI=17} ERROR 25,56%
|   |   |   |   A5M = si_matricula: NO {NO=195, SI=63} ERROR 24,42%
|   |   |   A9M = si_matricula: NO {NO=69, SI=19} ERROR 21,59%
|   |   A10M = si_matricula: NO {NO=72, SI=16} ERROR 18,18%
```

Analizando los árboles obtenidos se pueden observar algunos casos singulares:

En el caso del árbol del cluster_0 pueden verse al menos tres casos en los que el error es inferior al 20%. Observando como se han seleccionado las asignaturas para generar el árbol vemos que el 100% de los alumnos de este cluster se matricularon de A6M y de estos:

- por un lado casi el 5% se matricula de A12M y no abandonan los estudios con una probabilidad de éxito en la clasificación de mas del 80%.
- por otro camino se ve que los alumnos que eligen: no matricularse de

A12M, no matricularse de A11M, si matricularse de A1M, si matricularse de A3M y si matricularse de A10M son 6 alumnos que suponen un porcentaje del total del cluster de 0,17% pero que clasificándolos como alumnos que no abandonan se tendría un error del 16,67%.

- en último lugar se ve el caso mas favorable que es el obtenido por los alumnos que se matriculan de A12M, no se matriculan de A11M, si se matriculan de A1M, si se matriculan de A3M, no se se matriculan de A8M, si se matriculan de A5M, si se matriculan de A4M y no se matriculan de A2M, en este caso se clasifican 16 alumnos (0,45% del total) como que no abandonan con un error del 12,5%

En el caso del cluster_1 se da la situación inversa a la del caso anterior, el 100% de los alumnos clasificados en este cluster no se han matriculado de la asignatura A6M y en el resultado del mismo se puede ver que hay tres casos en los que el error es inferior al 20%

- por un lado los alumnos que eligen no matricularse de A6M, no matricularse de A12M, no matricularse de A10M, no matricularse de A11M, si matricularse de A9M, no matricularse de A8M, no matricularse de A3M, no matricularse de A1M, no matricularse de A4M y no matricularse de A5M son 35, representando el 0,25% del total y clasificados como alumnos que no abandonan se obtiene un error del 17,14%.

- por otro camino, lo alumnos que eligen no matricularse de A12M, no matricularse de A10M, si matricularse de A11M, no matricularse de A8M, no matricularse e A3M, no matricularse de A5M, si matricularse de A7M y si matricularse de A4M son en este caso 21 alumnos que representan el 0,15% del total de alumnos de este grupo y que al clasificarlos como que no abandonan se incurre en un error del 4,76%.

- por último en este cluster los alumnos que eligieron no matricularse de A12M, no matricularse de A10M, si matricularse de A11M y si matricularse de A8M fueron 25 que suponen un 0,18% y se clasificaron como que no abandonan con un error en la predicción del 12%.

En el caso del cluster_2 existen dos asignaturas que no influyen a la hora de predecir el abandono, ya que independientemente del resultado (si abandona o no abandona) el 100% de los alumnos agrupados en este cluster no se han matriculado de A4M y si se han matriculado de A8M.

Analizando el resto del árbol se observan dos resultados en los que se obtiene un error en la clasificación de menos del 20%:

- uno de ellos es para los alumnos que si se matriculan de A10M, en este caso son 88 y suponen el 6,6% del total de alumnos de este grupo que se ha clasificado como que no abandona con un error del 18,18%

- por otro lado los alumnos que eligen no matricularse de A10M, no matricularse de A8M, no matricularse de A5M, no matricularse de A12M, si matricularse de A3M, no matricularse de A11M y si matricularse de A7M son 16 que suponen un 1,2% y se comete un error del 18,75% al clasificarlos como alumnos que no abandonan.

Viendo todos los resultados en conjunto se observa que el error cometido en la clasificación de los alumnos en la mayoría de las hojas de los tres árboles es de alrededor del 25%, si bien en algunos casos es algo mejor, que son los casos mas favorables que se acaban de desarrollar, pero también hay casos de error del 50%.

Como se vio con anterioridad, la distribución del abandono entre los alumnos del estudio es de alrededor del 25%, con lo que la conclusión que se obtiene de este modelo es que la predicción que podría realizarse con él es claramente deficiente ya que el error cometido en la mayoría de los casos esta en torno a la distribución natural de los datos con lo que es incapaz de predecir nada en la mayoría de los casos.

7.4.5 Modelo de árbol de decisión sobre los datos de matriculación

En este modelo se tendrán en cuenta algunos de los datos que se obtuvieron en la creación de los clusters y esta información se aplicara directamente sobre los datos originales.

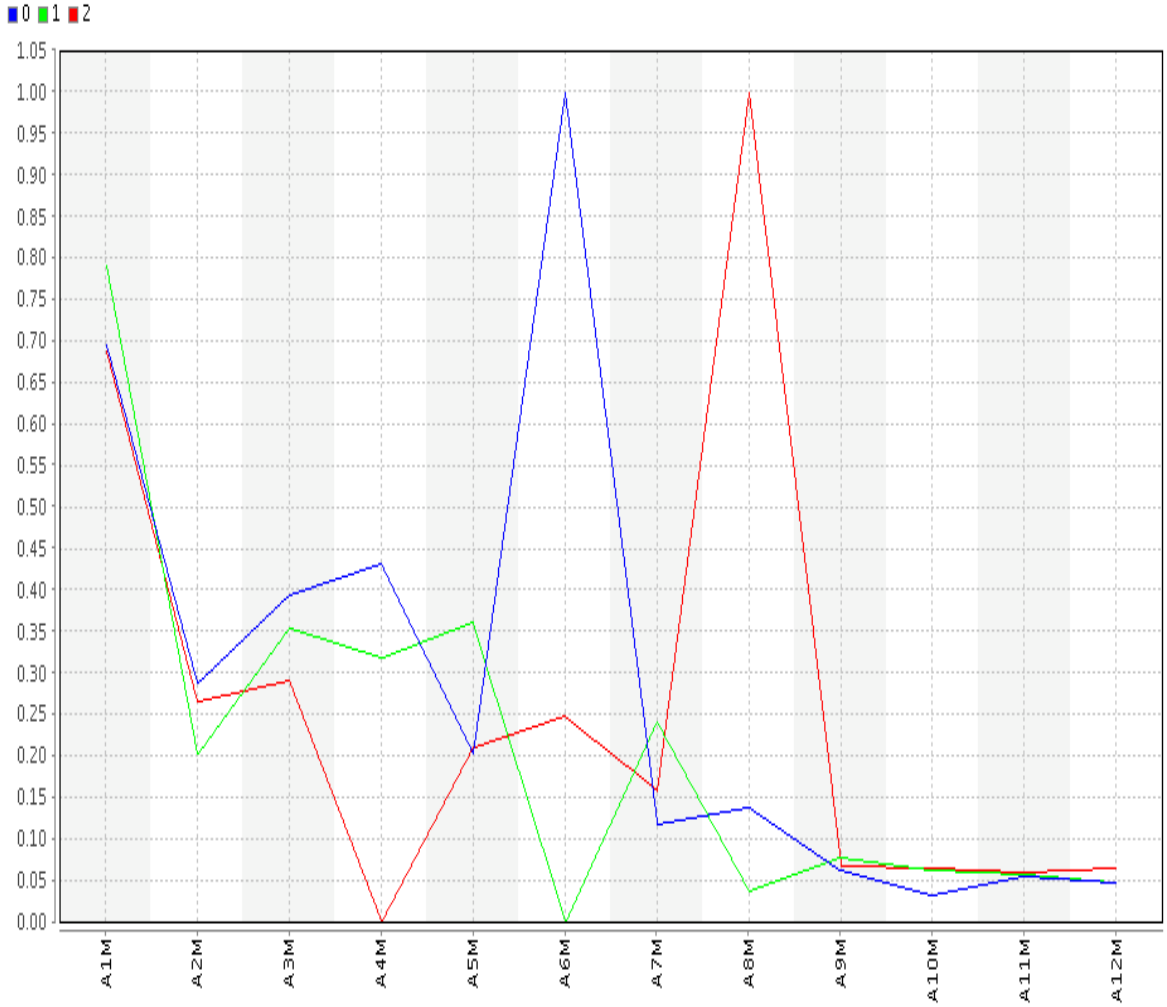
Entre los atributos que se seleccionaron en el modelo de clusters había algunos que tenían valores muy similares y otros que eran muy diferentes entre ellos. Lo que se pretende hacer en este modelo es aprovechar el conocimiento obtenido en la creación de los clusters para ayudar a seleccionar las asignaturas que hacen que unos alumnos se clasifiquen en un grupo y no en otro, es decir, aquellas asignaturas que son mas decisivas a la hora de clasificar a los alumnos

Este proceso se hace con el fin de descubrir que atributos son mas decisivos a la hora de agrupar a los alumnos y posteriormente utilizar esta información para seleccionar dichos atributos y realizar un árbol de decisión con el que se pueda predecir el fin último que nos interesa, que es el abandono de los estudios por parte de los estudiantes.

La información sobre los centroides de cada cluster ofrece una información muy útil para saber la homogeneidad o heterogeneidad de los atributos que hay agrupados

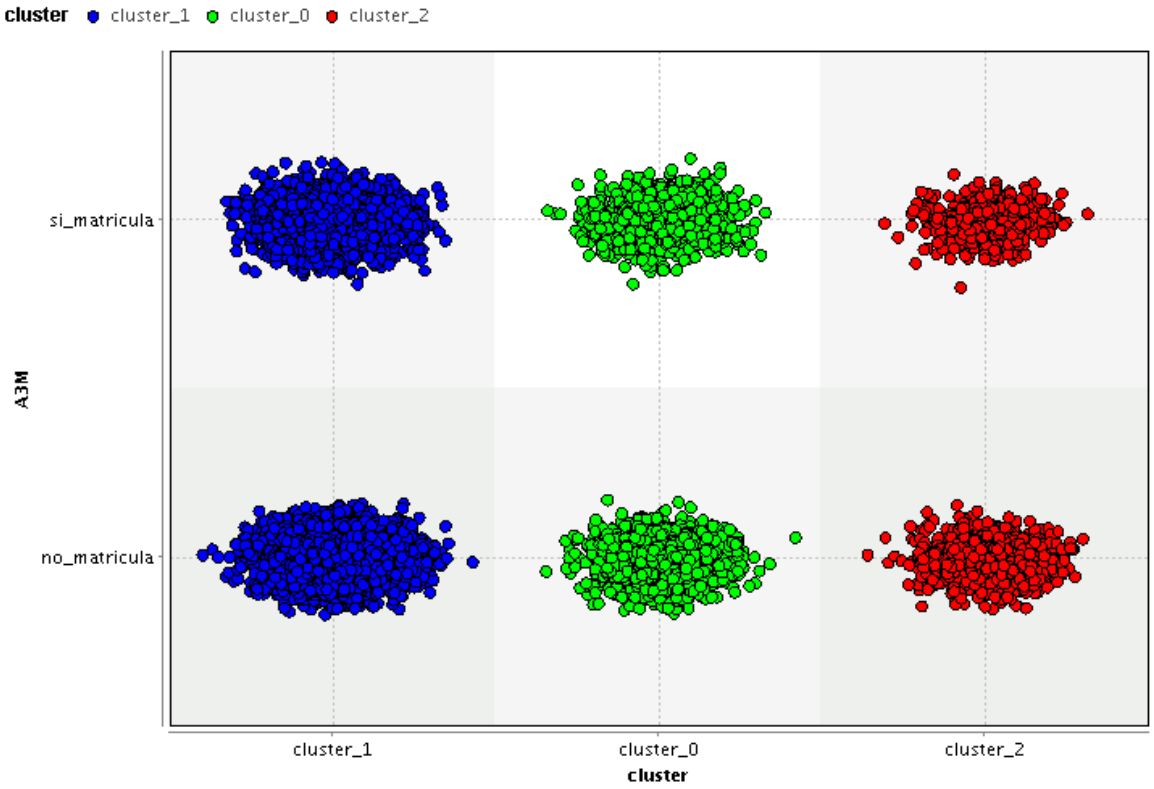
Attribute	cluster_0	cluster_1	cluster_2
A1M	0.697	0.792	0.689
A2M	0.288	0.202	0.265
A3M	0.394	0.355	0.291
A4M	0.432	0.318	0
A5M	0.203	0.360	0.211
A6M	1	0	0.248
A7M	0.119	0.241	0.158
A8M	0.139	0.038	1
A9M	0.062	0.079	0.069
A10M	0.033	0.063	0.066
A11M	0.055	0.058	0.060
A12M	0.049	0.047	0.064

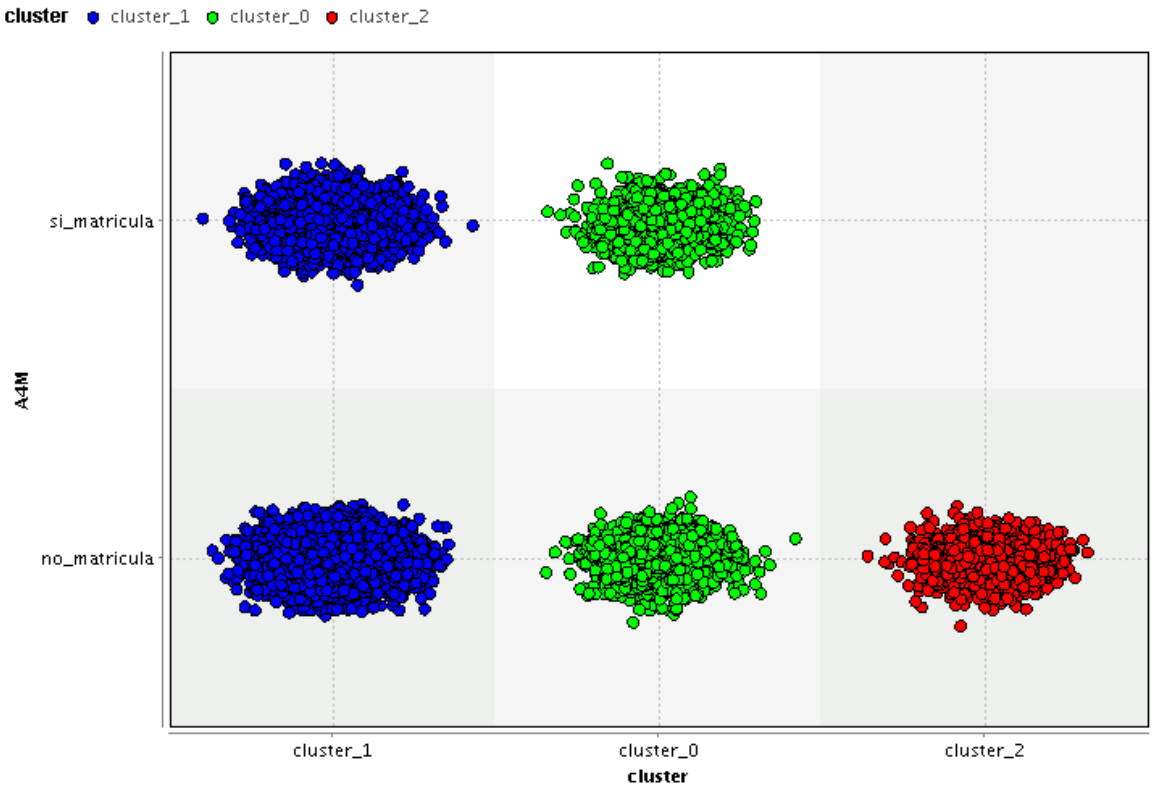
En este cuadro se ven los valores que tienen los centroides de cada atributo y en cada cluster. De forma gráfica se vería de la siguiente forma:

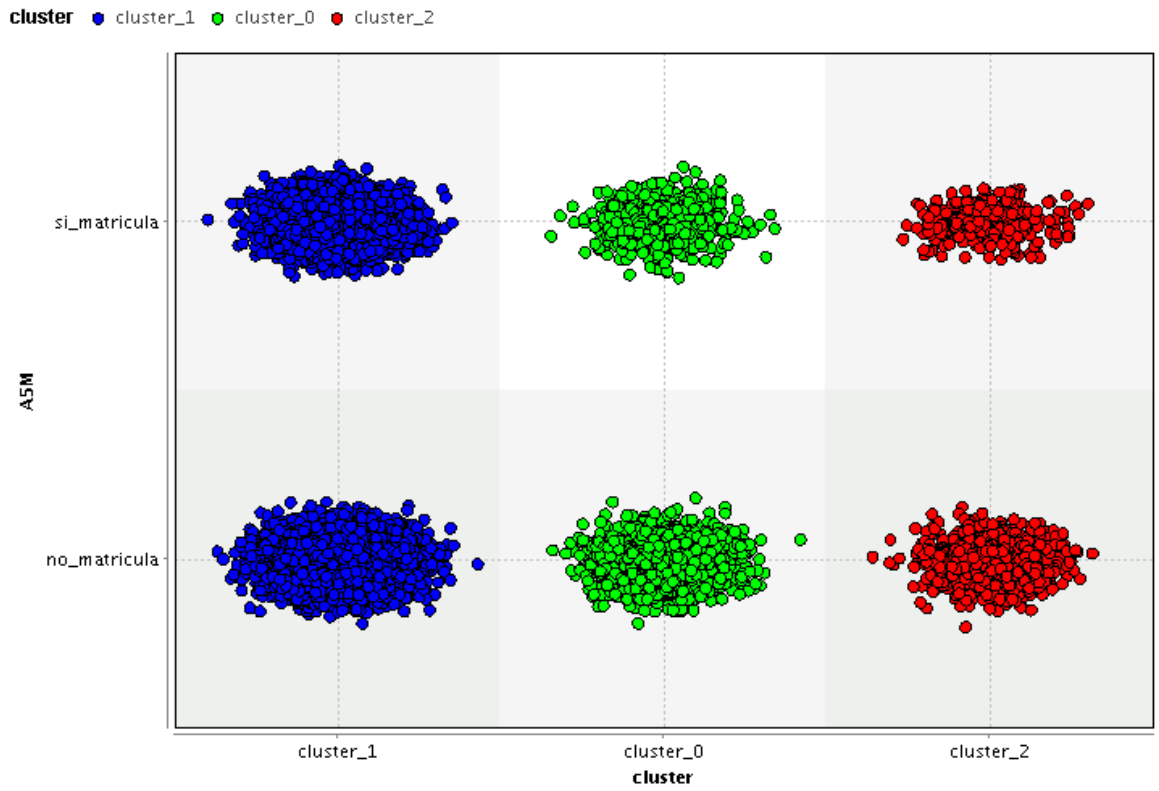


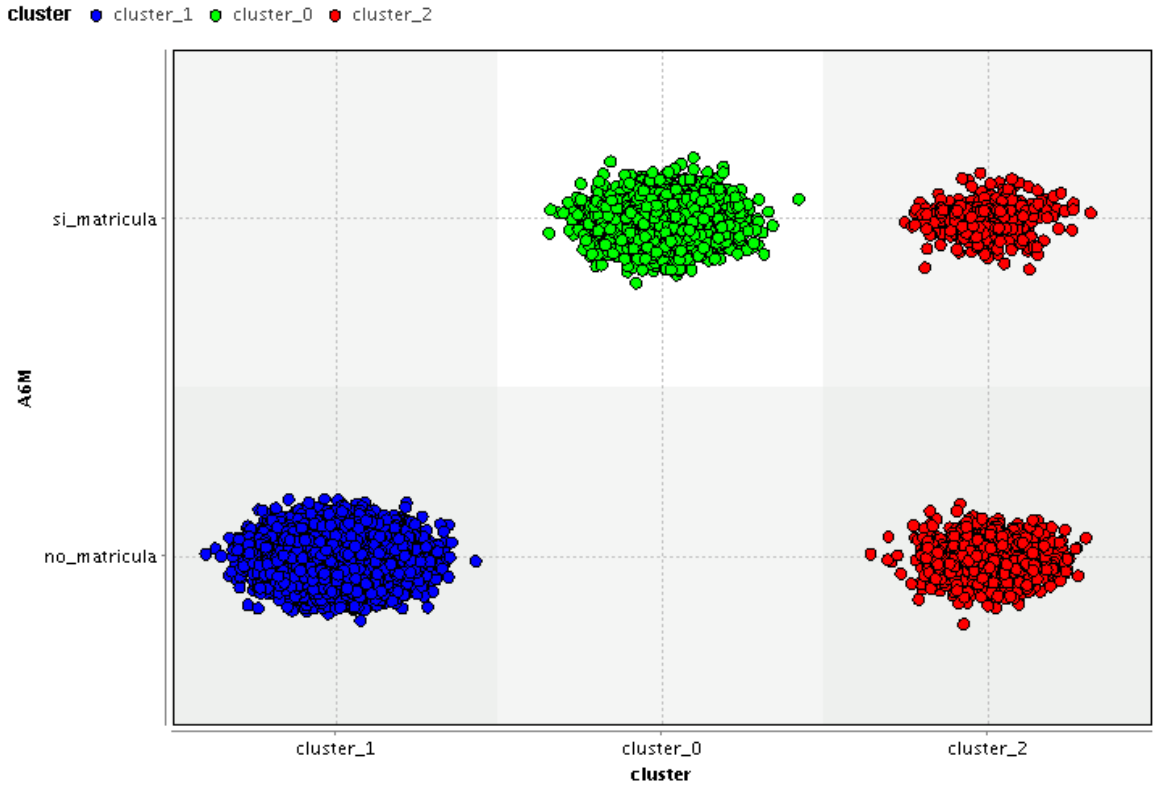
Se ve que los centroides de los atributos A3M, A4M, A5M, A6M, A7M y A8M están mas distantes entre ellos que el resto.

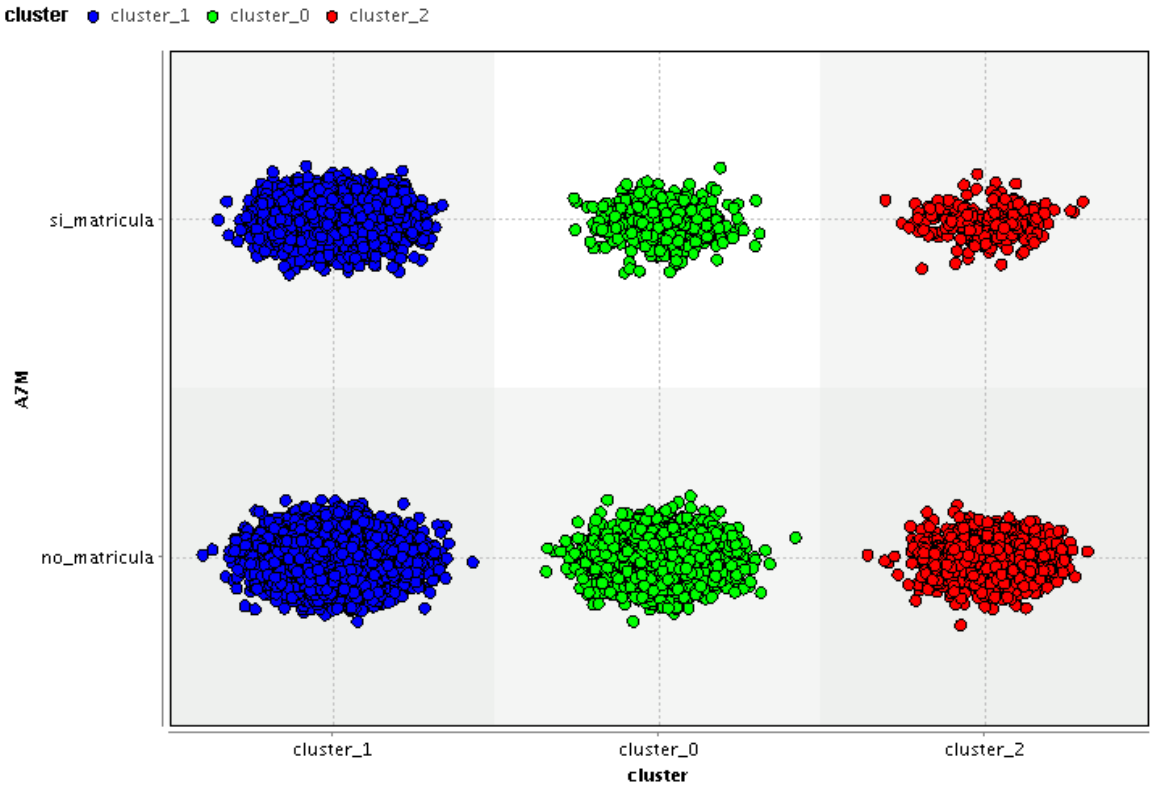
Se va a ver en detalle los valores de estos atributos y su clasificación en cada cluster:

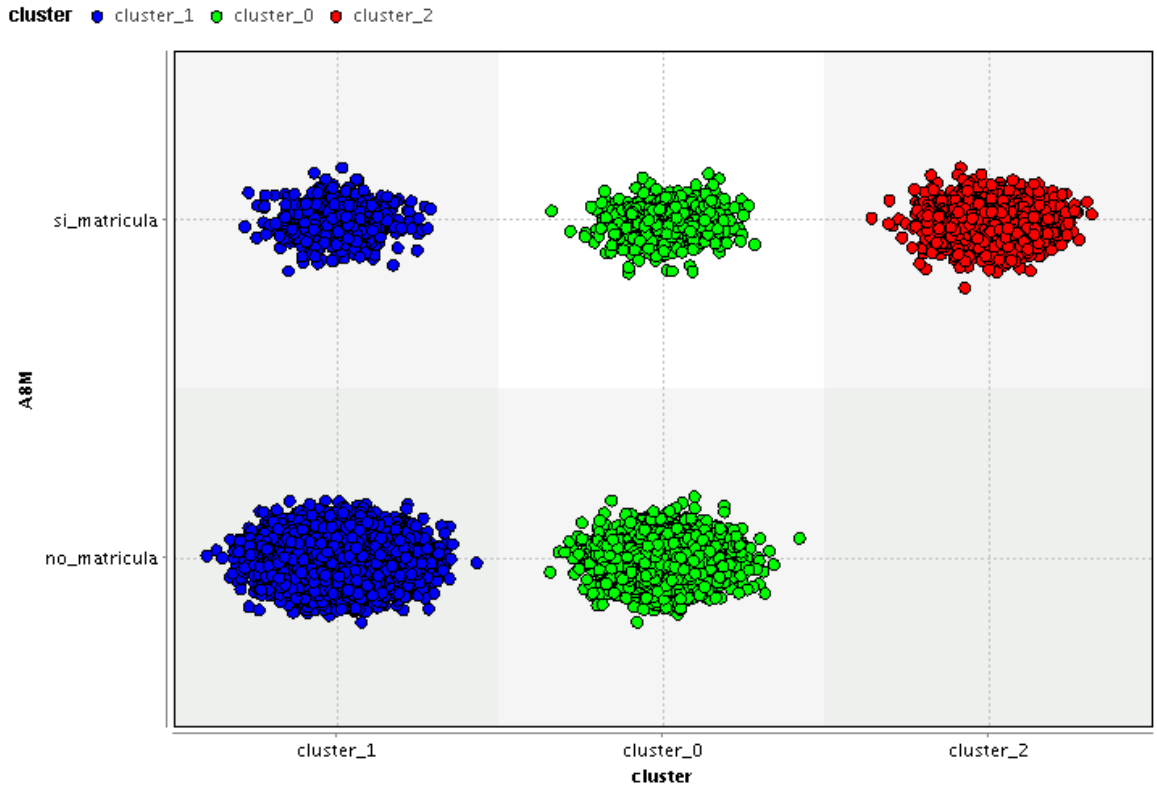





cluster
cluster_1
cluster_0
cluster_2
si_matricula
no_matricula
A5M
cluster_1
cluster_0
cluster_2
cluster


cluster
cluster_1
cluster_0
cluster_2
si_matricula
no_matricula
A6M
cluster_1
cluster_0
cluster_2
cluster


cluster
cluster_1
cluster_0
cluster_2
si_matricula
no_matricula
A7M
cluster_1
cluster_0
cluster_2
cluster


cluster
cluster_1
cluster_0
cluster_2
si_matricula
no_matricula
A8M
cluster_1
cluster_0
cluster_2
cluster

En el caso del atributo A3M, A5M y A7M las diferencias son mas sutiles y no se aprecia gráficamente tanta diferencia como en el resto de atributos seleccionados.

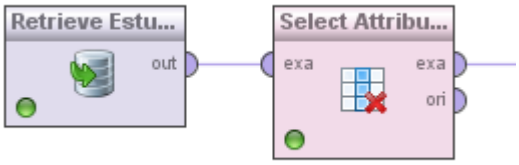
En la gráfica sobre el atributo A4M se ve claramente una diferencia y es que en el cluster_2 todos los individuos agrupados en el no se matricularon en esa asignatura.

En la gráfica del atributo A6M se ve que todos los individuos del cluster_1 no se matricularon de la asignatura y en el caso del cluster_0 todos se matricularon de la asignatura.

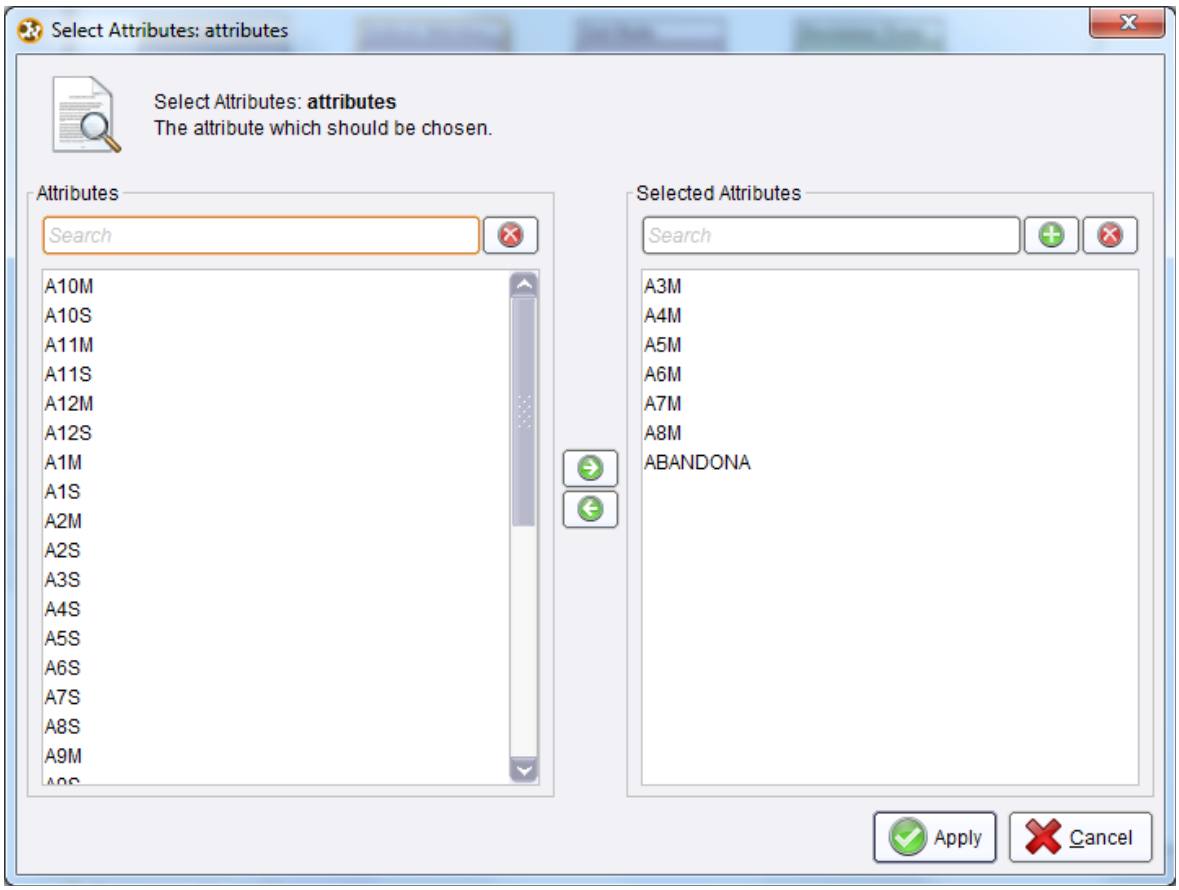
Por último en la gráfica del último atributo elegido, A8M, se ve que se ha agrupado de tal forma que en los cluster_1 y cluster_0 los individuos que si se matriculan son minoritarios y en el caso del cluster_2 todos los agrupados en este cluster son alumnos que si se matriculan de la asignatura.

Con esta información por tanto se va a realizar un modelo con un árbol de decisión en el que se van a seleccionar como atributos intervinientes los atributos A3M, A4M, A5M, A6M, A7M y A8M y por supuesto el atributo ABANDONA que es el dato que se trata de predecir.

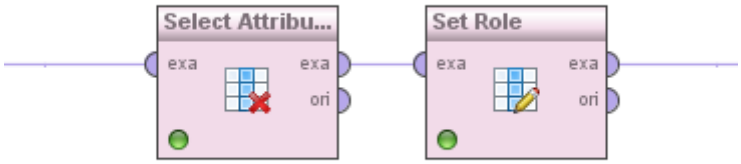
Partiendo de nuevo del repositorio original nuevamente se usará el objeto *Select* para seleccionar los datos que se van a utilizar:



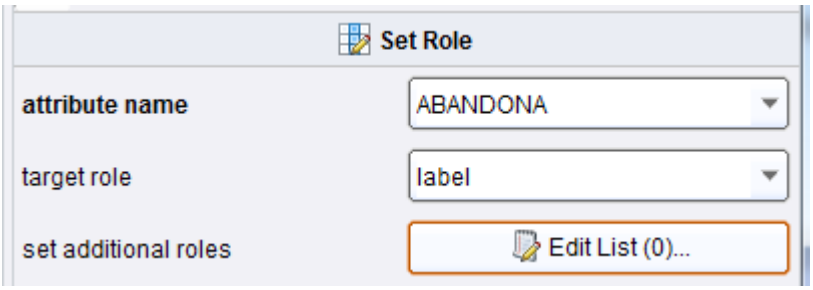
Se seleccionan los atributos:



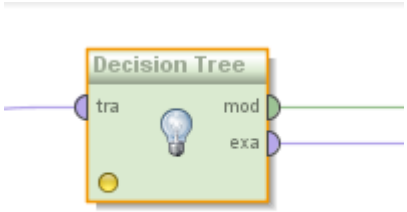
Se hace el cambio de rol nuevamente para el atributo ABANDONA:



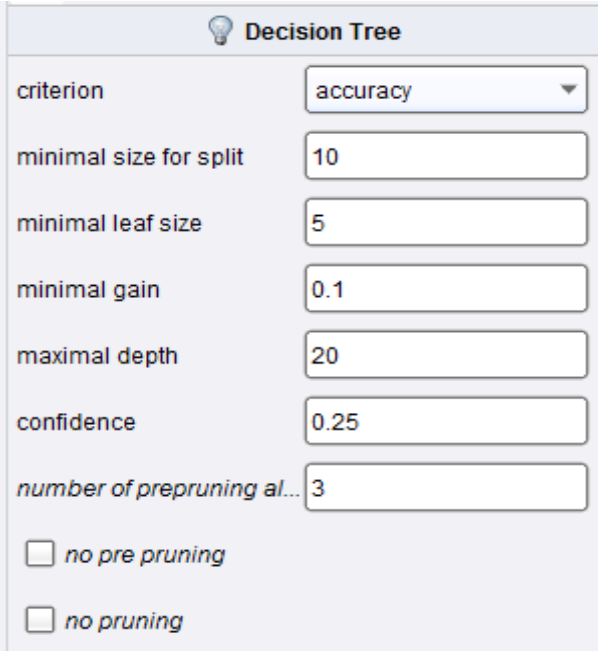
Y se configura de la siguiente forma:



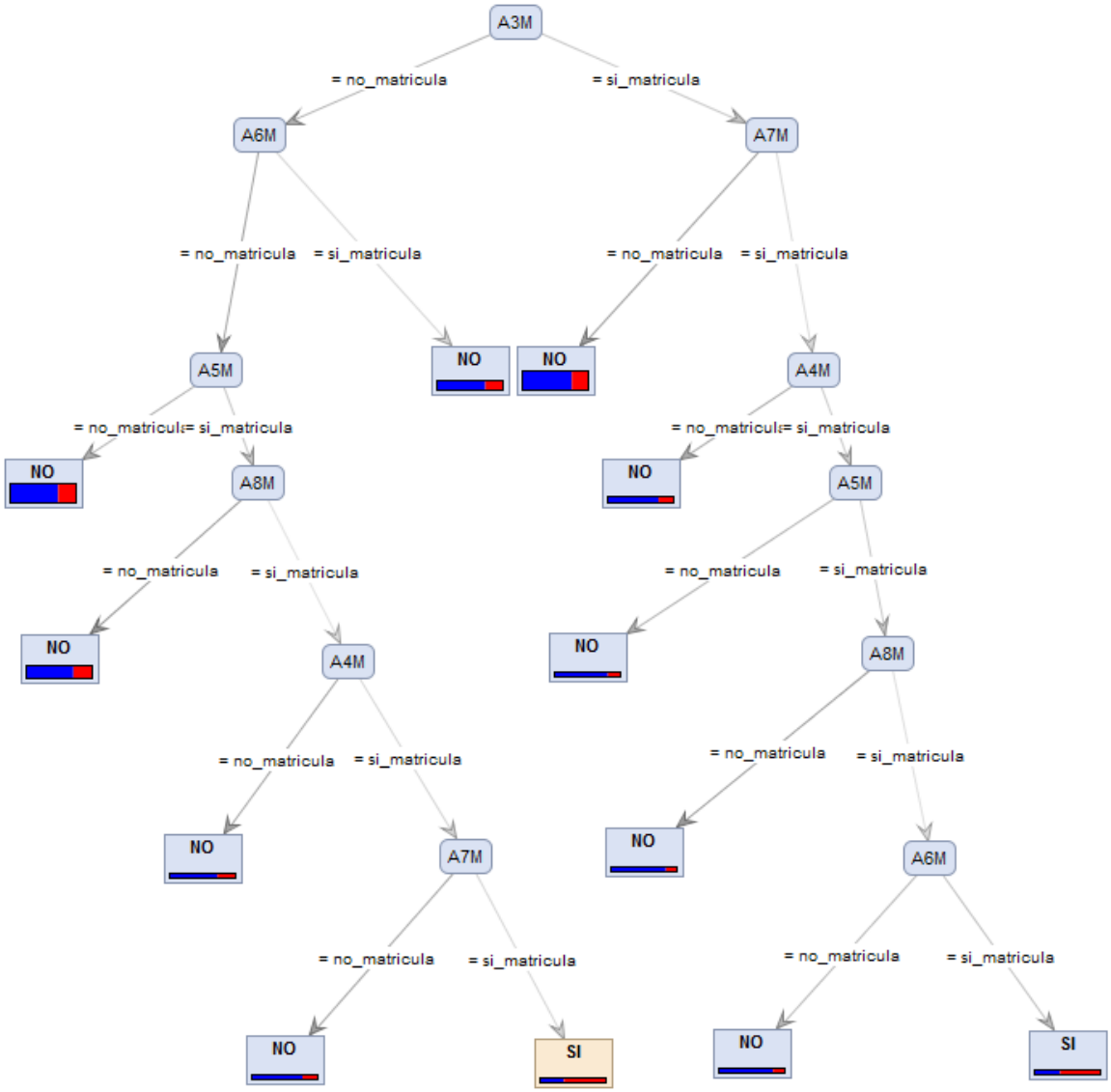
Finalmente incorporaremos el objeto árbol de decisión a la salida del objeto *Set Role.*



El árbol que se va a generar implementa métodos de poda siguiendo la siguiente configuración:



Este es el árbol resultante:



En formato texto pueden verse los valores obtenidos:

```
A3M = no_matricula
|   A6M = no_matricula
|   |   A5M = no_matricula: NO {NO=4469, SI=1578}
|   |   A5M = si_matricula
|   |   |   A8M = no_matricula: NO {NO=2393, SI=900}
|   |   |   A8M = si_matricula
|   |   |   |   A4M = no_matricula: NO {NO=115, SI=41}
|   |   |   |   A4M = si_matricula
|   |   |   |   |   A7M = no_matricula: NO {NO=25, SI=7}
|   |   |   |   |   A7M = si_matricula: SI {NO=3, SI=5}
|   A6M = si_matricula: NO {NO=1766, SI=604}
A3M = si_matricula
|   A7M = no_matricula: NO {NO=4357, SI=1361}
|   A7M = si_matricula
|   |   A4M = no_matricula: NO {NO=507, SI=141}
|   |   A4M = si_matricula
|   |   |   A5M = no_matricula: NO {NO=146, SI=32}
|   |   |   A5M = si_matricula
|   |   |   |   A8M = no_matricula: NO {NO=77, SI=14}
|   |   |   |   A8M = si_matricula
|   |   |   |   |   A6M = no_matricula: NO {NO=5, SI=1}
|   |   |   |   |   A6M = si_matricula: SI {NO=2, SI=3}
```

7.4.6 Conclusiones sobre el modelo obtenido

Si calculamos el error de cada uno de los caminos no salen las siguientes cifras:

```
A3M = no_matricula
|   A6M = no_matricula
|   |   A5M = no_matricula: NO {NO=4469, SI=1578} ERROR 26,09%
|   |   A5M = si_matricula
|   |   |   A8M = no_matricula: NO {NO=2393, SI=900} ERROR 27,33%
|   |   |   A8M = si_matricula
|   |   |   |   A4M = no_matricula: NO {NO=115, SI=41} ERROR 26,28%
|   |   |   |   A4M = si_matricula
|   |   |   |   |   A7M = no_matricula: NO {NO=25, SI=7} ERROR 21,87%
|   |   |   |   |   A7M = si_matricula: SI {NO=3, SI=5} ERROR 37,5%
|   A6M = si_matricula: NO {NO=1766, SI=604} ERROR 25,48%
A3M = si_matricula
|   A7M = no_matricula: NO {NO=4357, SI=1361} ERROR 23,80%
|   A7M = si_matricula
|   |   A4M = no_matricula: NO {NO=507, SI=141} ERROR 21,75%
|   |   A4M = si_matricula
|   |   |   A5M = no_matricula: NO {NO=146, SI=32} ERROR 17,98%
|   |   |   A5M = si_matricula
|   |   |   |   A8M = no_matricula: NO {NO=77, SI=14} ERROR 15,38%
|   |   |   |   A8M = si_matricula
|   |   |   |   |   A6M = no_matricula: NO {NO=5, SI=1} ERROR 16,66%
|   |   |   |   |   A6M = si_matricula: SI {NO=2, SI=3} ERROR 40%
```

Se ve a simple vista que el modelo predice no abandono en la mayoría de los caminos elegidos.

El error cometido es muy alto en todos los casos, solo predice que un alumno abandonara los estudios en dos de los caminos y en esos casos el porcentaje de alumnos supone menos del 0,1% y se comete un error de alrededor del 40% en ambos casos.

Por lo tanto el modelo desarrollado no sirve para conseguir los objetivos del proyecto, ya que se trataba de obtener un modelo que de alguna forma explicara cuales eran los motivos del abandono de los alumnos y en este caso no lo hace, o lo hace con un error tan grande que no es útil.

8. Revisión del trabajo - conclusiones finales

En el proyecto se ha estudiado la relación que pudiera haber entre las asignaturas de las que se matricula un alumno y el abandono de la carrera. Las conclusiones obtenidas de todo ello no han sido las esperadas, si bien es verdad que quedan abiertos otros caminos.

Durante las primeras fases del desarrollo del proyecto se han obtenido diversos tipos de datos procedentes de las bases de datos de la UOC y que han sido adaptados e integrados en un repositorio de Rapid Miner para poder ser tratados con esta aplicación. Además a la hora de la selección de los datos y de la preparación, se han utilizado todos los datos de los que se disponía de tal manera que han quedado en disposición para ser utilizados en otras lineas de investigación.

Las pruebas realizadas con diferentes datos y modelos han llevado todas ellas a unas conclusiones similares. Puede llegar a pensarse que los motivos por los que un alumno abandona sus estudios se encuentran en otros que no aparecen reflejados en la bases de datos de la UOC.

Es posible que para llegar a descubrir los motivos con una probabilidad de éxito superior al obtenido, se deberían apoyar los datos de matriculación de los alumnos en otros datos obtenidos de forma voluntaria en encuestas en las que se detallen otros aspectos de las personas tales como si trabaja o no, o cuanto tiempo libre tienen para dedicar al estudio, las cargas familiares o los ingresos por ejemplo.

Quizás completando los datos de una forma así pueda llegarse a obtener un conocimiento mejorado de la situación.

9. Bibliografía y referencias

Bibliografía

Material docente de la UOC

Ramón Sangüesa i Solé, Luis Carlos Molina Félix (2010)

Principles of Data Mining

David Hand, Heikki Mannila and Padhraic Smyth (2001)

RapidMiner 5 - Operator Reference

Fareed Akthar, Caroline Hahne

Rapid Miner - User Manual

Material en Internet consultado

Rapid-I Forum

https://rapid-i.com/rapidforum/

Mineria de Datos - Herramientas

http://www.authorstream.com/Presentation/anabelladg-1597016-mineria-de-datos-herramientas/

Generación de modelos

http://www.dataprix.com/734-generaci-n-modelos

Cluster analysis

http://en.wikipedia.org/wiki/Cluster_analysis

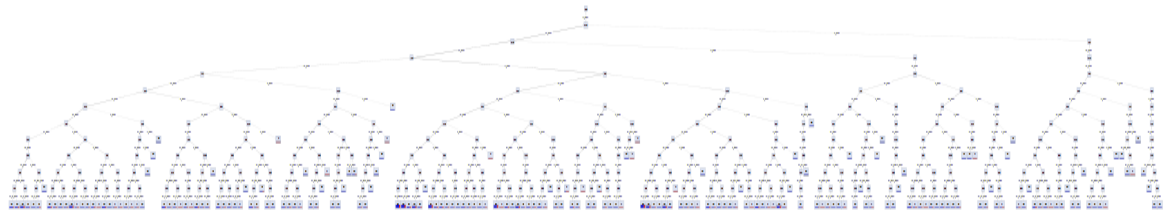
10. Anexos

Buscando la forma de mejorar la capacidad predictiva de los modelos trabajados se hicieron algunas pruebas con otros algoritmos y aquí se muestran los resultados.

Se hicieron pruebas en el caso del modelo del apartado 7.4.3 con algoritmos ID3 con una configuración que maximizara la precisión, se estableció el tamaño mínimo del nodo para la partición en 8 y el tamaño mínimo de la hoja en 4 con una ganancia de 0,1.

Los resultados obtenidos como puede verse a continuación nos devuelve árboles desmesuradamente grandes y sobrespecializados que si bien predicen mejor algunos datos su tamaño los hace inmanejables. En modo gráfico no se puede tener una visión en conjunto y en modo texto es realmente complicado poder seguirlos.

Se muestran los resultados del árbol del cluster_0, que son bastante ilustrativos sobre el tema que se acaba de exponer:



```
A6M = si_matricula
|   A12M = no_matricula
|   |   A11M = no_matricula
|   |   |   A1M = no_matricula
|   |   |   |   A7M = no_matricula
|   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO
{NO=42, SI=10}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO
{NO=27, SI=11}
|   |   |   |   |   |   |   |   |   |   A3M = si_matricula
```

```
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO
{NO=68, SI=22}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO
{NO=18, SI=3}
|   |   |   |   |   |   |   |   |   A9M = si_matricula
|   |   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=7,
SI=0}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=4,
SI=3}
|   |   |   |   |   |   |   |   |   |   A3M = si_matricula: NO {NO=9,
SI=0}
|   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=42, SI=14}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=9,
SI=6}
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=10, SI=2}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO
{NO=88, SI=26}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: SI {NO=2,
SI=2}
|   |   |   |   |   |   |   |   |   |   A9M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=3,
SI=0}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   A8M = si_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=23, SI=10}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: SI {NO=0,
SI=1}
|   |   |   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=48, SI=13}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: SI {NO=1,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: SI {NO=1,
SI=2}
|   |   |   |   |   |   A10M = si_matricula
```

```
|   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO {NO=8,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: SI {NO=1,
SI=2}
|   |   |   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: SI {NO=1,
SI=1}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=3, SI=0}
|   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=3, SI=0}
|   |   |   |   |   |   |   A5M = si_matricula: NO {NO=2, SI=0}
|   |   |   |   |   A2M = si_matricula
|   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=44, SI=14}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=3,
SI=1}
|   |   |   |   |   |   |   |   |   |   A10M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO {NO=5,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: SI {NO=0,
SI=1}
|   |   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=14, SI=5}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: SI {NO=0,
SI=1}
|   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=41, SI=14}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=3,
SI=1}
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO {NO=8,
SI=2}
|   |   |   |   |   |   |   |   |   A10M = si_matricula: NO {NO=1, SI=0}
|   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=38, SI=8}
```

```
|   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=1,
SI=1}
|   |   |   |   |   |   |   |   |   A8M = si_matricula
|   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=11, SI=3}
|   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   |   |   A9M = si_matricula: SI {NO=1,
SI=1}
|   |   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=35, SI=6}
|   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=1,
SI=1}
|   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   A8M = si_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO
{NO=17, SI=7}
|   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   A10M = si_matricula: SI {NO=0, SI=1}
|   |   |   A7M = si_matricula
|   |   |   |   A10M = no_matricula
|   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO
{NO=14, SI=4}
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=5,
SI=4}
|   |   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=6,
SI=2}
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula: SI {NO=1,
SI=1}
|   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=2, SI=0}
|   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=5,
SI=1}
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=3,
SI=0}
```

```
|   |   |   |   |   |   |   |   |   A8M = si_matricula: SI {NO=0, SI=1}
|   |   |   |   |   |   |   A9M = si_matricula
|   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=6,
SI=1}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   |   |   A3M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=1, SI=0}
|   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=2, SI=1}
|   |   |   |   |   |   A2M = si_matricula
|   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=5,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=6,
SI=2}
|   |   |   |   |   |   |   |   |   |   A3M = si_matricula: NO {NO=4,
SI=0}
|   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=5, SI=0}
|   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=2, SI=0}
|   |   |   |   |   |   |   A9M = si_matricula: SI {NO=0, SI=1}
|   |   |   |   |   A10M = si_matricula: NO {NO=3, SI=0}
|   |   |   A1M = si_matricula
|   |   |   |   A3M = no_matricula
|   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A8M = no_matricula: NO
{NO=168, SI=60}
|   |   |   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A8M = no_matricula: NO
{NO=159, SI=54}
|   |   |   |   |   |   |   |   |   |   |   A8M = si_matricula: NO
{NO=39, SI=17}
|   |   |   |   |   |   |   |   |   A7M = si_matricula
|   |   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO
{NO=72, SI=23}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO
{NO=15, SI=8}
|   |   |   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   A4M = no_matricula
```

```
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO
{NO=140, SI=42}
|   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO
{NO=22, SI=11}
|   |   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO
{NO=28, SI=13}
|   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=7,
SI=4}
|   |   |   |   |   |   |   |   A8M = si_matricula
|   |   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO {NO=4,
SI=1}
|   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   A10M = si_matricula
|   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO
{NO=21, SI=5}
|   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=3,
SI=2}
|   |   |   |   |   |   |   |   |   A7M = si_matricula
|   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO {NO=4,
SI=1}
|   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO {NO=4,
SI=0}
|   |   |   |   |   |   |   |   |   |   A7M = si_matricula: SI {NO=1,
SI=1}
|   |   |   |   |   |   |   A8M = si_matricula: SI {NO=0, SI=1}
|   |   |   |   |   A2M = si_matricula
|   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   A10M = no_matricula: NO
{NO=134, SI=50}
|   |   |   |   |   |   |   |   |   |   A10M = si_matricula: SI
{NO=1, SI=3}
|   |   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   |   A10M = no_matricula: NO
{NO=33, SI=13}
|   |   |   |   |   |   |   |   |   |   A10M = si_matricula: NO
{NO=1, SI=0}
|   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO
{NO=65, SI=21}
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=7,
SI=2}
|   |   |   |   |   |   |   |   |   A8M = si_matricula
```

```
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO
{NO=16, SI=4}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=3,
SI=3}
|   |   |   |   |   |   |   A7M = si_matricula
|   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO
{NO=23, SI=3}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: SI {NO=2,
SI=3}
|   |   |   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO {NO=4,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   |   |   A8M = si_matricula: SI {NO=1,
SI=2}
|   |   |   |   |   A9M = si_matricula
|   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO
{NO=23, SI=8}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=3,
SI=1}
|   |   |   |   |   |   |   |   |   |   A2M = si_matricula: SI {NO=0,
SI=1}
|   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=9,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   |   A2M = si_matricula
|   |   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=1,
SI=1}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   A10M = si_matricula
|   |   |   |   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO {NO=3,
SI=0}
```

```
|   |   |   |   |   |   |   |   |   |   A4M = si_matricula: SI {NO=0,
SI=1}
|   |   |   |   |   |   A7M = si_matricula
|   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula: NO {NO=10, SI=0}
|   |   |   |   |   |   |   |   A8M = si_matricula: SI {NO=0, SI=1}
|   |   |   |   |   |   |   A10M = si_matricula: SI {NO=0, SI=2}
|   |   |   |   A3M = si_matricula
|   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO
{NO=156, SI=59}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO
{NO=91, SI=29}
|   |   |   |   |   |   |   |   |   |   A9M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO {NO=5,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   |   A2M = si_matricula
|   |   |   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO
{NO=70, SI=20}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO
{NO=25, SI=8}
|   |   |   |   |   |   |   |   |   |   A9M = si_matricula: SI {NO=0,
SI=3}
|   |   |   |   |   |   |   |   A7M = si_matricula
|   |   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A2M = no_matricula: NO
{NO=13, SI=2}
|   |   |   |   |   |   |   |   |   |   |   A2M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A2M = no_matricula: NO {NO=5,
SI=1}
|   |   |   |   |   |   |   |   |   |   |   A2M = si_matricula: SI {NO=1,
SI=3}
|   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=1, SI=0}
|   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO
{NO=40, SI=13}
|   |   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=4,
SI=2}
|   |   |   |   |   |   |   |   |   |   A2M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO
{NO=10, SI=2}
```

```
|   |   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A2M = no_matricula: NO
{NO=10, SI=1}
|   |   |   |   |   |   |   |   |   |   |   A2M = si_matricula: SI {NO=3,
SI=3}
|   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   |   A7M = si_matricula
|   |   |   |   |   |   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO {NO=3,
SI=0}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: SI {NO=0,
SI=1}
|   |   |   |   |   |   A8M = si_matricula
|   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO
{NO=53, SI=12}
|   |   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A7M = no_matricula: SI {NO=1,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A7M = si_matricula: SI {NO=0,
SI=1}
|   |   |   |   |   |   |   |   |   A2M = si_matricula
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO
{NO=79, SI=33}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=3,
SI=1}
|   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=6, SI=0}
|   |   |   |   |   A10M = si_matricula
|   |   |   |   |   |   A2M = no_matricula
|   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A8M = no_matricula: NO {NO=3,
SI=1}
|   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=1, SI=0}
|   |   |   |   |   |   A2M = si_matricula: NO {NO=1, SI=0}
|   |   A11M = si_matricula
|   |   |   A2M = no_matricula
|   |   |   |   A1M = no_matricula
|   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   A4M = no_matricula
```

```
|   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A10M = no_matricula: NO
{NO=8, SI=2}
|   |   |   |   |   |   |   |   |   |   |   A10M = si_matricula: SI
{NO=0, SI=1}
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A10M = no_matricula: NO
{NO=3, SI=1}
|   |   |   |   |   |   |   |   |   |   |   A10M = si_matricula: NO
{NO=1, SI=0}
|   |   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=7,
SI=0}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: SI {NO=1,
SI=2}
|   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=3, SI=0}
|   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO {NO=8,
SI=2}
|   |   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   |   A8M = no_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   |   |   A8M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO {NO=3,
SI=0}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: SI {NO=2,
SI=2}
|   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=2, SI=0}
|   |   |   |   |   |   |   A10M = si_matricula: NO {NO=1, SI=0}
|   |   |   |   |   A7M = si_matricula
|   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=8,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=1, SI=0}
|   |   |   |   A1M = si_matricula
|   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   A8M = no_matricula
```

```
|   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A10M = no_matricula: NO
{NO=33, SI=7}
|   |   |   |   |   |   |   |   |   |   |   A10M = si_matricula: SI
{NO=2, SI=2}
|   |   |   |   |   |   |   |   |   |   A7M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A10M = no_matricula: NO
{NO=5, SI=2}
|   |   |   |   |   |   |   |   |   |   |   A10M = si_matricula: SI
{NO=0, SI=1}
|   |   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO
{NO=10, SI=4}
|   |   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A8M = no_matricula: NO
{NO=11, SI=3}
|   |   |   |   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A8M = no_matricula: NO {NO=4,
SI=1}
|   |   |   |   |   |   |   |   |   |   |   A8M = si_matricula: NO {NO=8,
SI=3}
|   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=1, SI=0}
|   |   |   |   |   |   A9M = si_matricula
|   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   A4M = no_matricula: NO {NO=4, SI=0}
|   |   |   |   |   |   |   |   A4M = si_matricula: SI {NO=0, SI=1}
|   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   A8M = no_matricula
|   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO {NO=8,
SI=1}
|   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO {NO=4,
SI=1}
|   |   |   |   |   |   |   |   |   |   |   A7M = si_matricula: SI {NO=1,
SI=1}
|   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=2, SI=0}
|   |   |   |   |   |   |   A8M = si_matricula: NO {NO=1, SI=0}
```

```
|   A12M = si_matricula
|   |   A2M = no_matricula
|   |   |   A11M = no_matricula
|   |   |   |   A8M = no_matricula
|   |   |   |   |   A9M = no_matricula
|   |   |   |   |   |   A3M = no_matricula
|   |   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   A1M = no_matricula
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: SI {NO=2,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=6,
SI=1}
|   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   |   A1M = si_matricula
|   |   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO
{NO=30, SI=16}
|   |   |   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=2,
SI=1}
|   |   |   |   |   |   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A7M = no_matricula: NO
{NO=12, SI=3}
|   |   |   |   |   |   |   |   |   |   |   A7M = si_matricula: SI {NO=0,
SI=2}
|   |   |   |   |   |   |   |   A10M = si_matricula
|   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A1M = no_matricula: NO {NO=2,
SI=0}
|   |   |   |   |   |   |   |   |   |   |   A1M = si_matricula: NO {NO=3,
SI=1}
|   |   |   |   |   |   |   A5M = si_matricula
|   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   |   A1M = no_matricula: NO {NO=4,
SI=0}
|   |   |   |   |   |   |   |   |   |   A1M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A4M = no_matricula: NO {NO=7,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=5,
SI=0}
|   |   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=4, SI=0}
|   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   |   A4M = no_matricula
|   |   |   |   |   |   |   |   |   |   A1M = no_matricula
|   |   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=9,
SI=2}
|   |   |   |   |   |   |   |   |   |   |   A5M = si_matricula: NO {NO=1,
SI=0}
|   |   |   |   |   |   |   |   |   |   A1M = si_matricula
```

```
|   |   |   |   |   |   |   |   |   |   A5M = no_matricula: NO {NO=8,
SI=0}
|   |   |   |   |   |   |   |   |   |   A5M = si_matricula: SI {NO=0,
SI=2}
|   |   |   |   |   |   |   |   |   A4M = si_matricula: NO {NO=12, SI=0}
|   |   |   |   |   |   |   |   A7M = si_matricula: NO {NO=5, SI=0}
|   |   |   |   |   A9M = si_matricula
|   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   A1M = no_matricula: NO {NO=2, SI=0}
|   |   |   |   |   |   |   |   A1M = si_matricula
|   |   |   |   |   |   |   |   |   A3M = no_matricula: NO {NO=4, SI=0}
|   |   |   |   |   |   |   |   |   A3M = si_matricula: SI {NO=0, SI=1}
|   |   |   |   |   |   |   A10M = si_matricula: NO {NO=1, SI=0}
|   |   |   |   A8M = si_matricula
|   |   |   |   |   A4M = si_matricula
|   |   |   |   |   |   A5M = no_matricula
|   |   |   |   |   |   |   A7M = no_matricula
|   |   |   |   |   |   |   |   A10M = no_matricula
|   |   |   |   |   |   |   |   |   A1M = no_matricula: NO {NO=3, SI=0}
|   |   |   |   |   |   |   |   |   A1M = si_matricula
|   |   |   |   |   |   |   |   |   |   A3M = no_matricula: NO {NO=3,
SI=0}
|   |   |   |   |   |   |   |   |   |   A3M = si_matricula
|   |   |   |   |   |   |   |   |   |   |   A9M = no_matricula: NO {NO=8,
SI=1}
|   |   |   |   |   |   |   |   |   |   |   A9M = si_matricula: NO {NO=2,
SI=0}
```