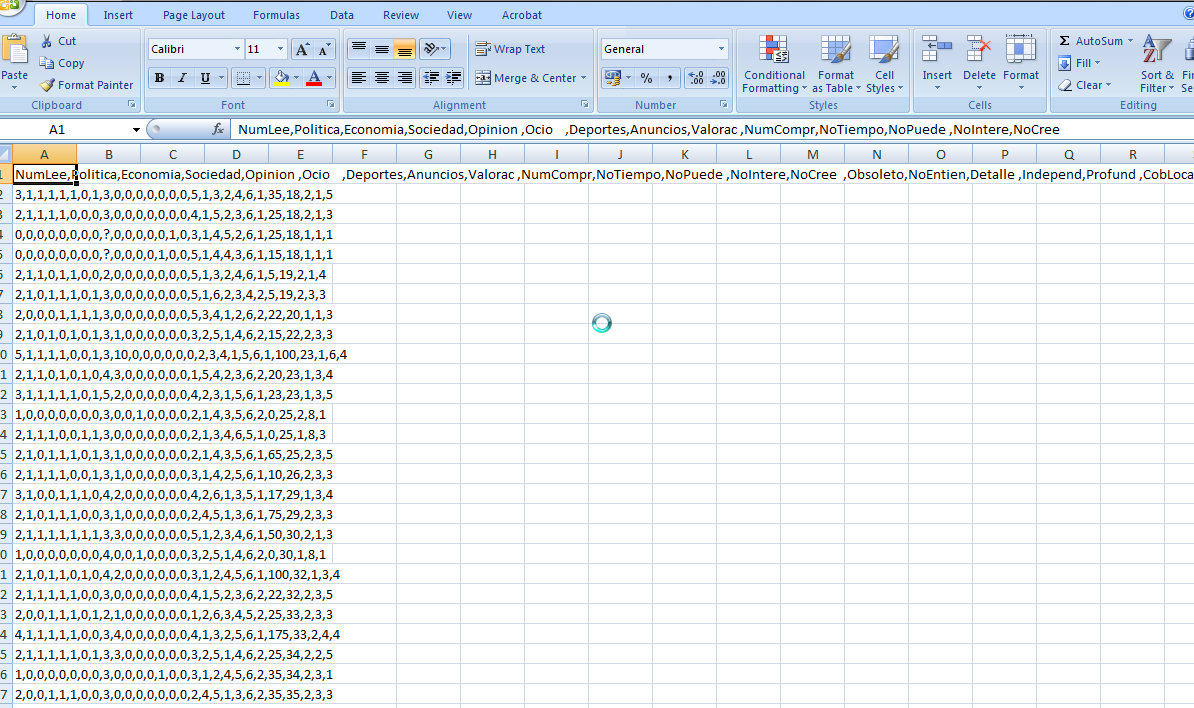
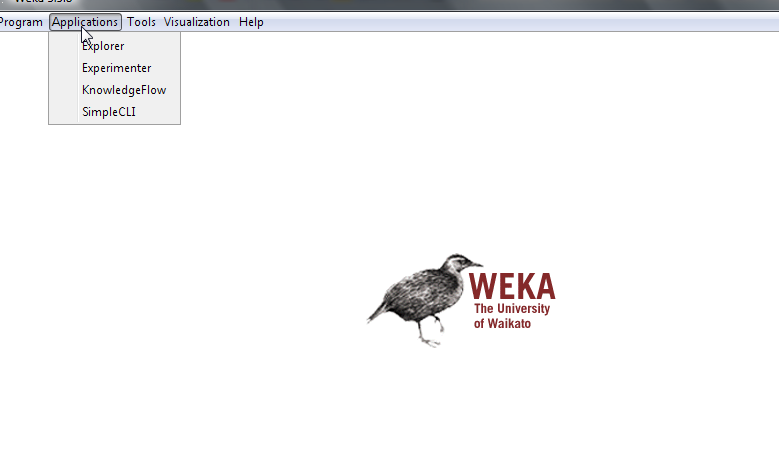
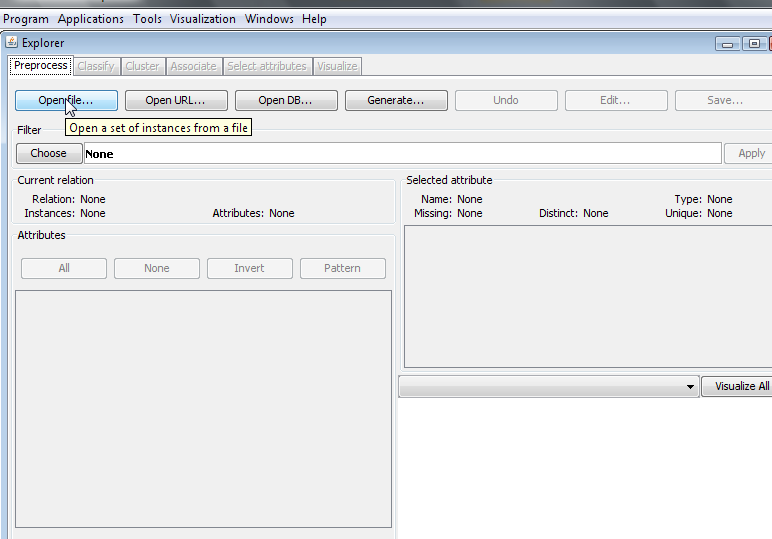
**NOTAS SOBRE ENCUESTA PERIÓDICO**

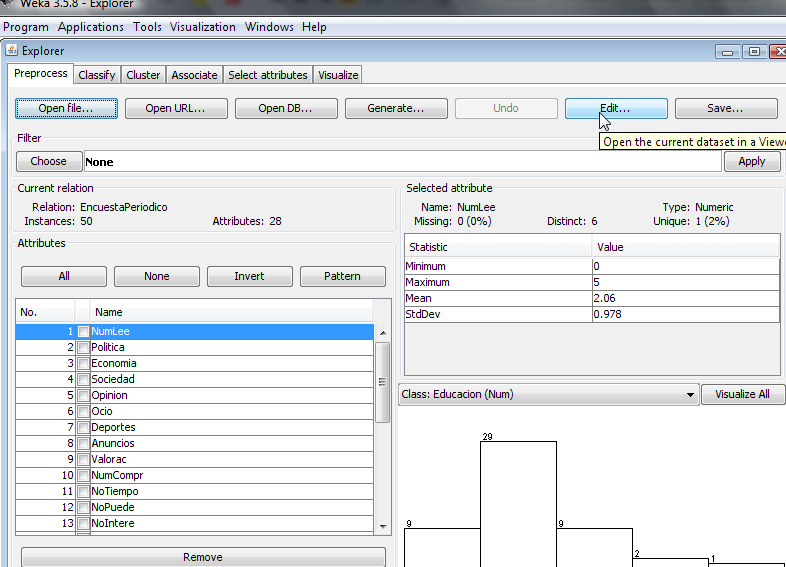
1. Abrir el fichero ordenadores desde Excel y salvarlo con el formato *CVS* ( Comma delimited) con el nombre *encuesta.cvs*.



1. Volverlo abrirlo, ahora con Notepad
2. Cambiar el LIST SEPARATOR que es ; por , usando EDIT -> REPLACE
3. Cambiar el DECIMAL SYMBOL que es ,(una coma) por . (un punto) En mi caso ya lo cambié en *Cloc, Language and Region*
4. Si hay palabras que no estén en la primera línea, cambiarlas por esas mismas palabras entre comillas.
5. La utilización de cualquier técnica de aprendizaje de tipo supervisado (con una variable respuesta) requiere definir en el fichero qué variable va actuar como tal. En la clase del 28/11 dice que utilizaremos la variable 23 que es *New Diary* como variable respuesta (Weka la denomina class). Lo primero es abrir el fichero y para ello vamos a *Aplications* y seleccionamos

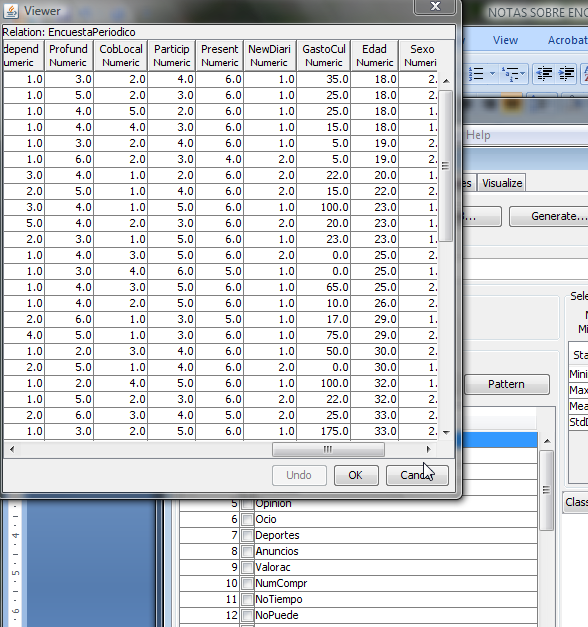
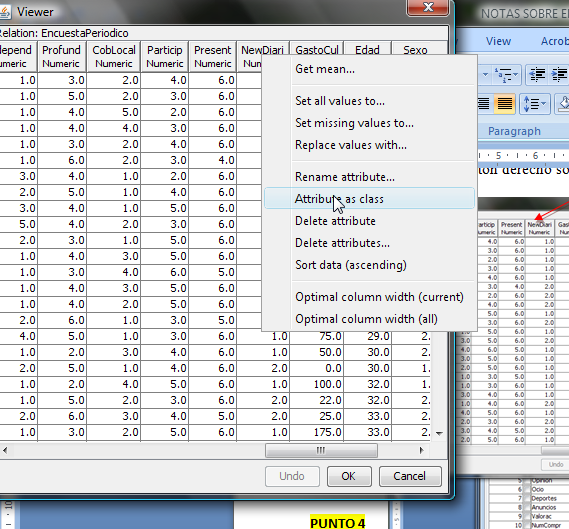
*Explorer* y vamos a *openfile*



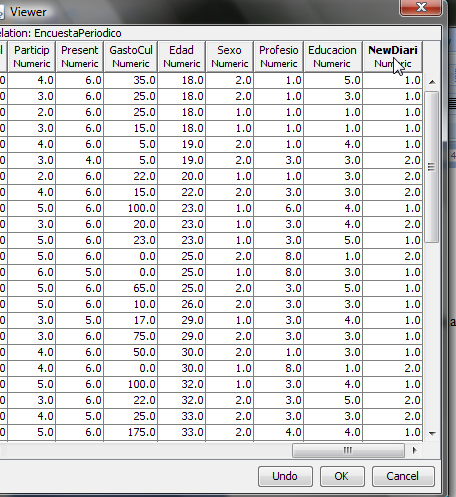


Después de abrir el fichero lo editamos. Para esto vamos a la tecla *Edit* y definimos la variable respuesta.

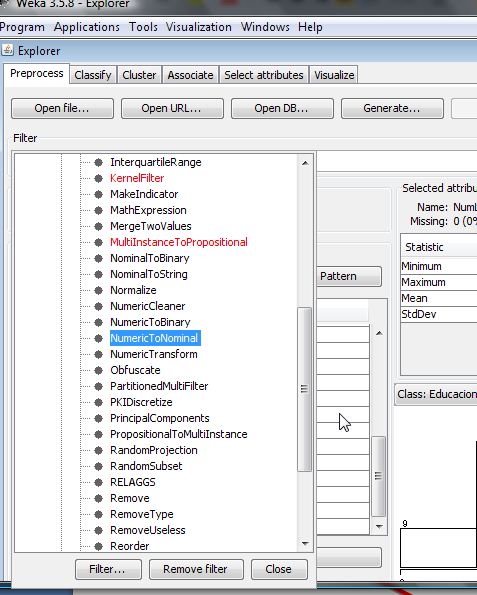
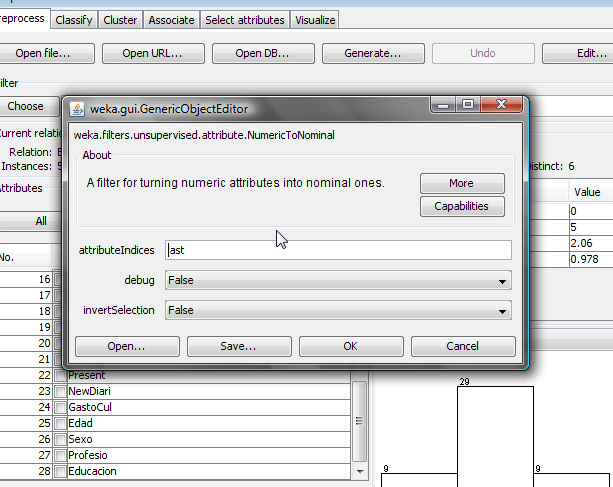
1. Para ello, con el botón derecho sobre *New Diary* seleccionamos *Attribute as class*.



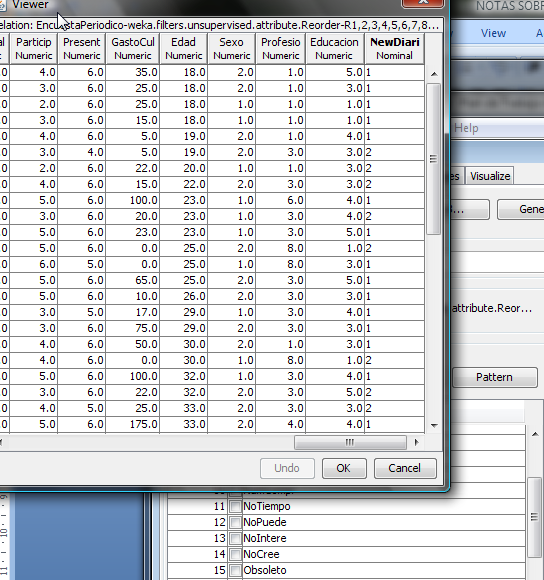
La variable se coloca en último lugar de la tabla (es un convenio del formato *arff*) y aparece en negrita



1. En la clase seleccionó el filtro *NumericToNominal* (fui a ver la encuesta y es la pregunta 7 que tiene dos respuestas (sí y no) con la clasificación 1 y 2) y en *atributeIndices* colocó *last* (pienso se refiere a la variable respuesta que es la última)

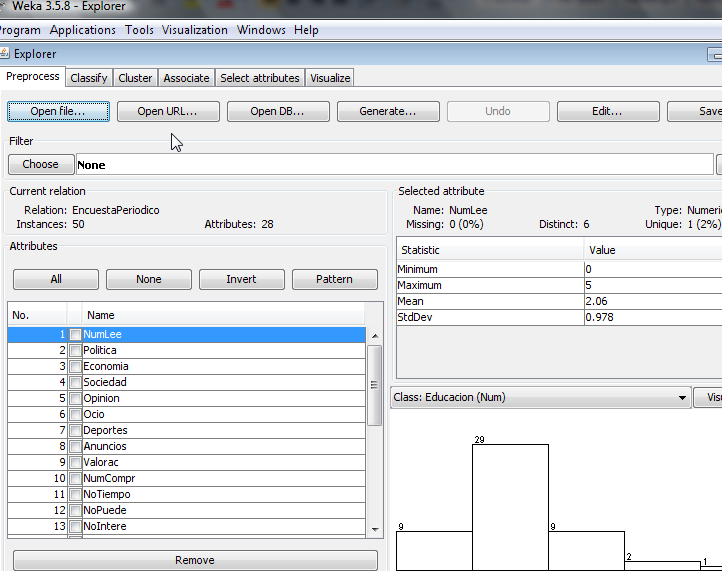


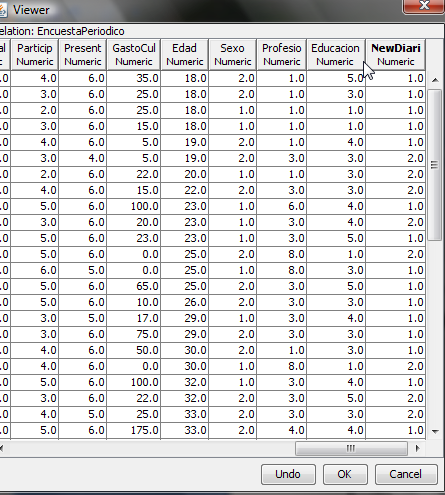
Se ve que ahora la variable aparece como Nominal o Categórica

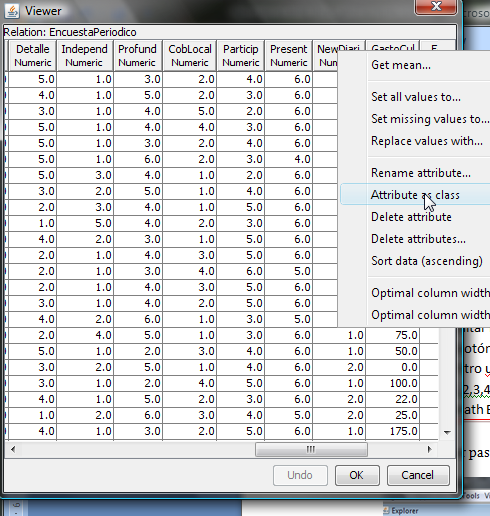


Ahora hay que pasar la clasificación de las respuestas 1 y 2 a 1 y 0. ¿Cómo hizo esta transformación? El día 1/12 volvió a explicar este punto porque el día 28/11 no se acordaba. Explicó:

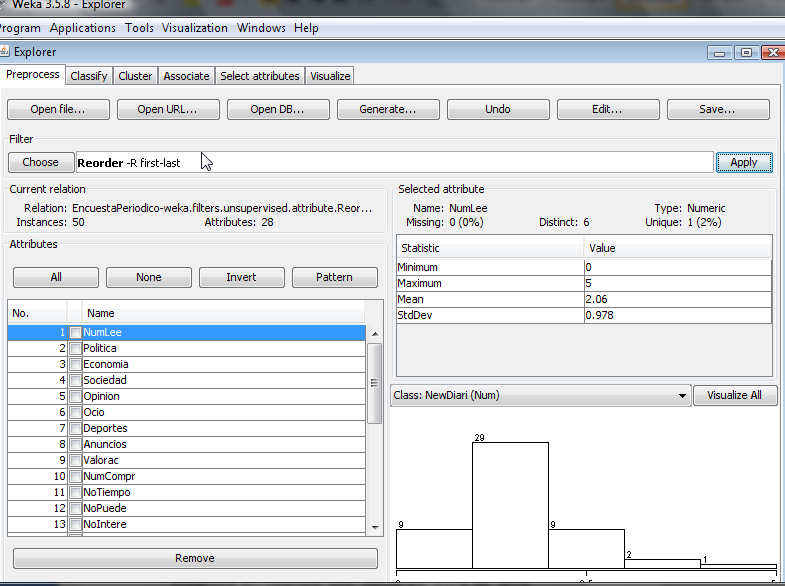
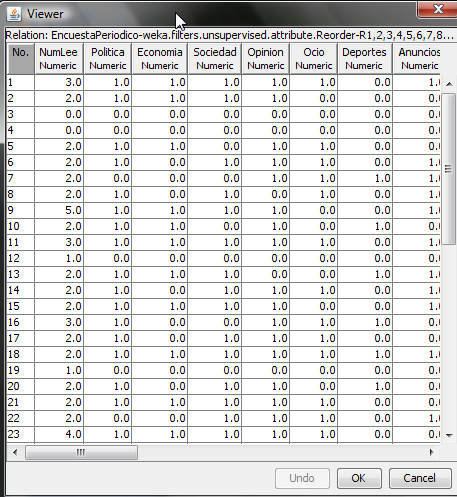
1. Seleccionar el fichero Encuesta Periódico.cvs
2. Editar los datos y pasar New diari as class (botón derecho)
3. Filtro unsupervised atribute. Reorder R 1,2,3,4,5,…
4. Math Expression
5. Este es el primer paso: Seleccionar el fichero Encuesta Periódico.cvs



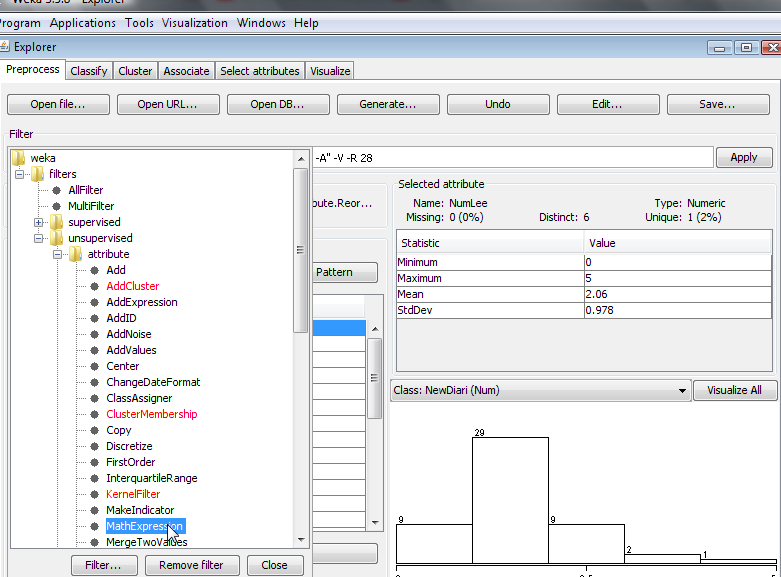
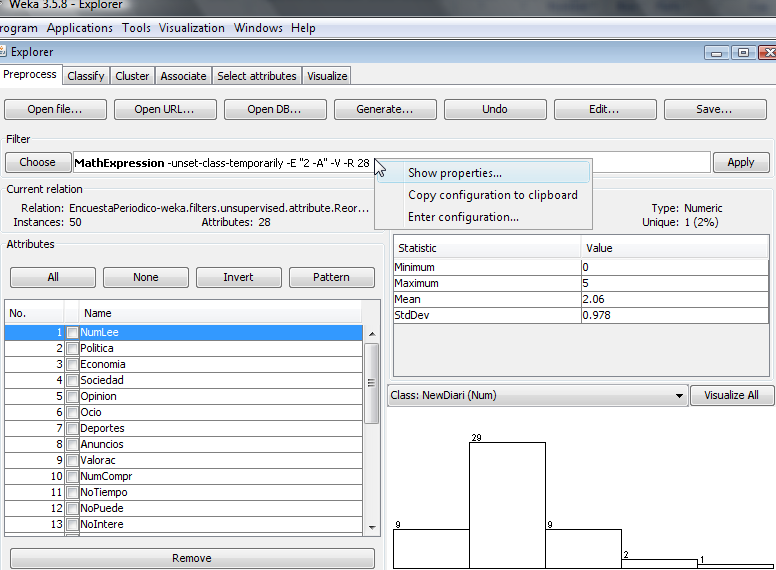
1. Editar los datos y pasar New diari as class (botón derecho)

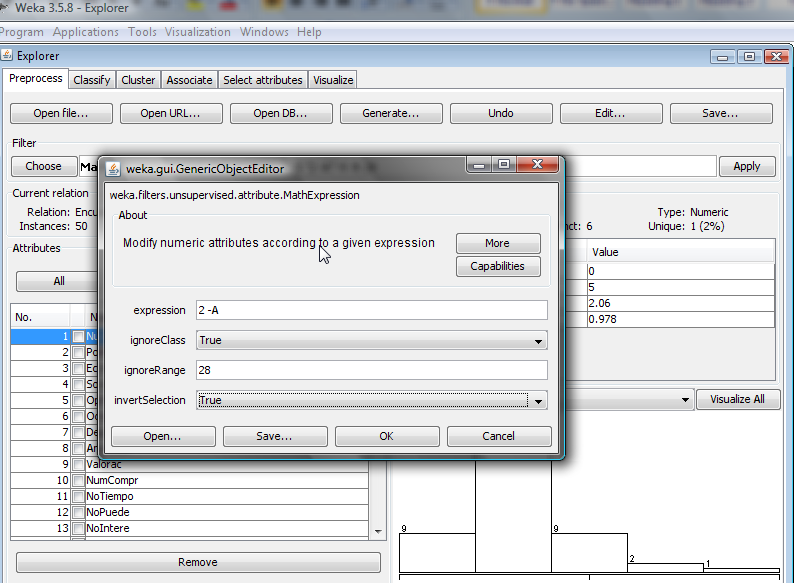


1. Filtro unsupervised atribute. Reorder R 1,2,3,4,5,…



1. Se aplica el filtro Math Expression y con el botón derecho seleccionar *Show properties* y seleccionar lo que está en el cuadro de texto y después aplicar:





Colocar:

Expression : 2-A

Ignore class: True

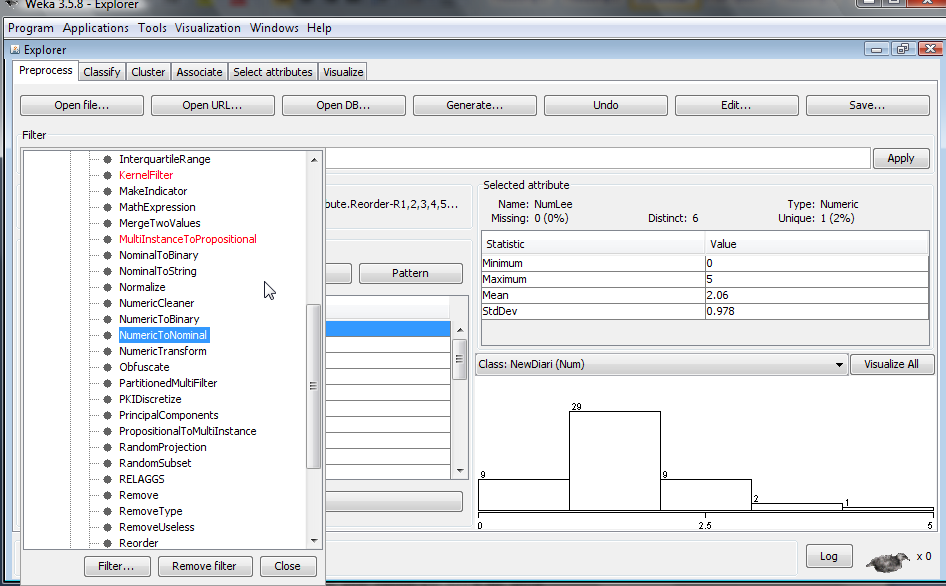
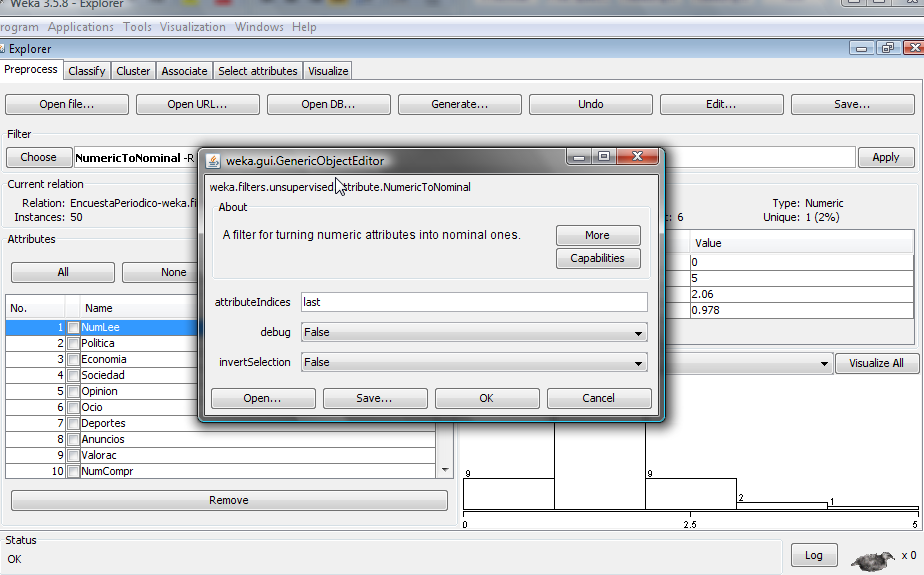
Ignore range: 28

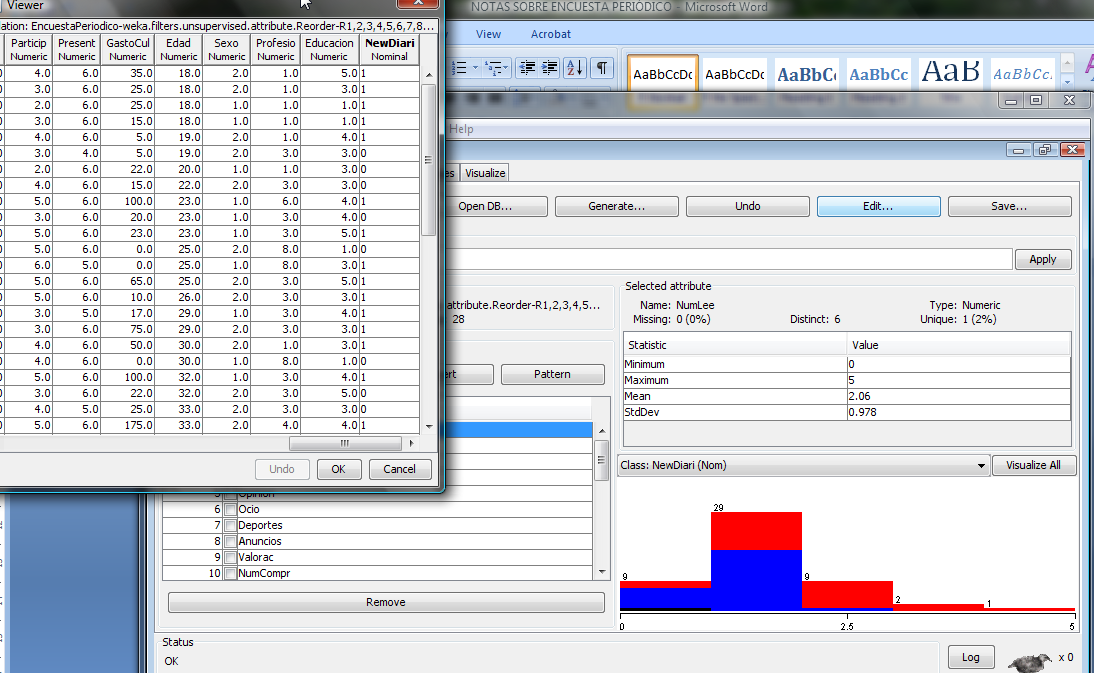
Invert Selection: True

Son proposiciones

negativas

1. Aplicar el filtro *NumericToNominal* seleccionado el *last* en *atributeIndice* (esto es para convertir la variable *New Diari* en Categórica). Recordar que para hacer esto tiene que ser con el botón derecho





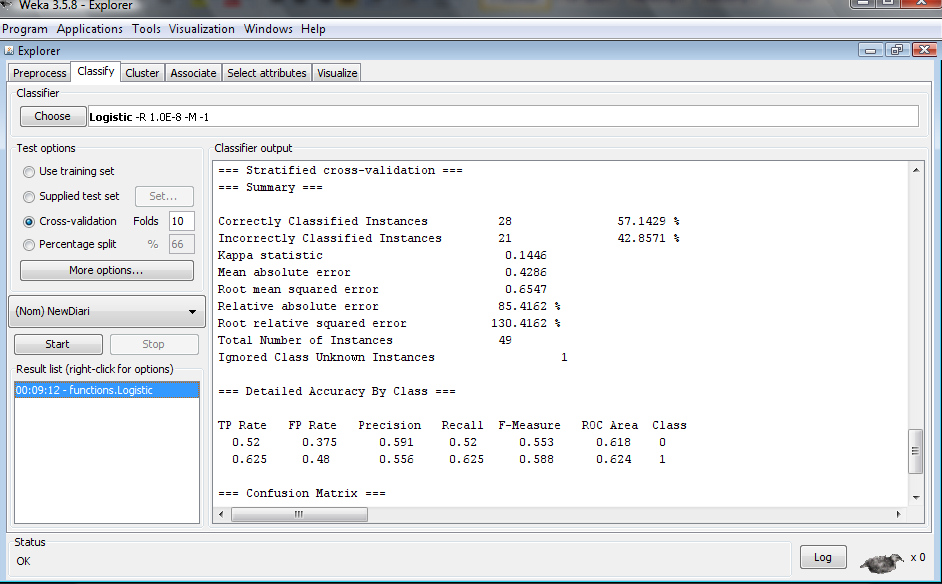
Se puede observar que la variable *New Diari* es la última, que es categórica y que ya se cambió la clasificación 1 y 2 en 0 y 1.

Ahora hay que realizar todo el trabajo anterior realizado en el fichero ordenadores. Mediante la pestaña *Classify* utilizar los siguientes métodos de clasificación para la variable *New Diari* utilizando 10*-fold-cv* y con las opciones propuestas por defecto.

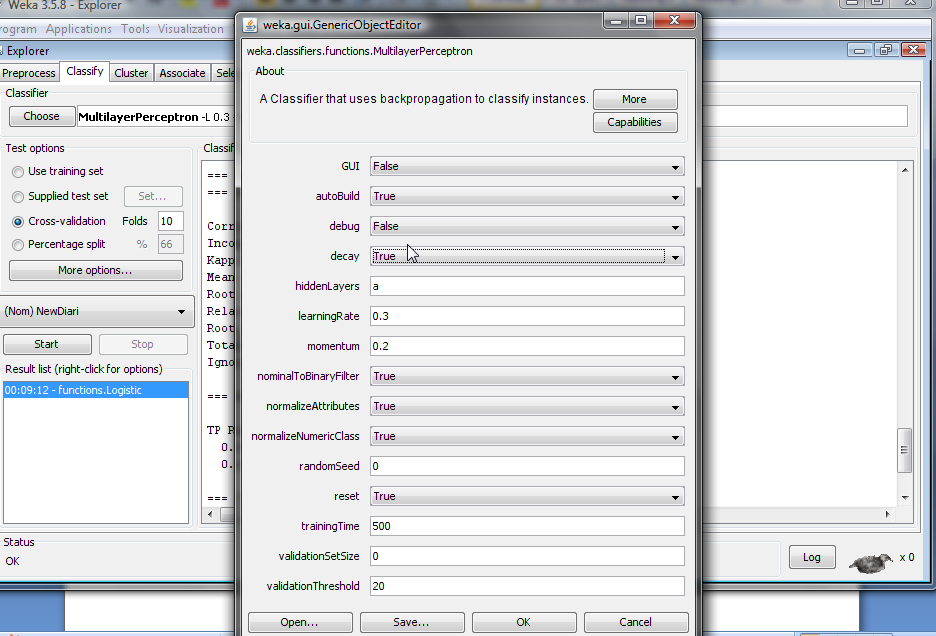
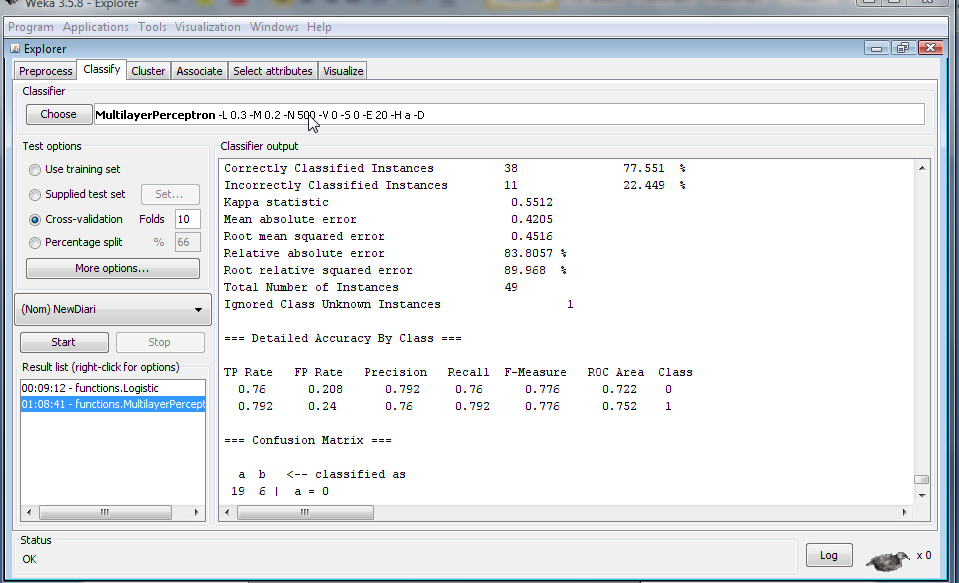
CON TODOS LOS ATRIBUTOS

De la carpeta *functions*: *logistic, LibSVM, SMO* y *Multilayer Perceptron.*

*Logistic*

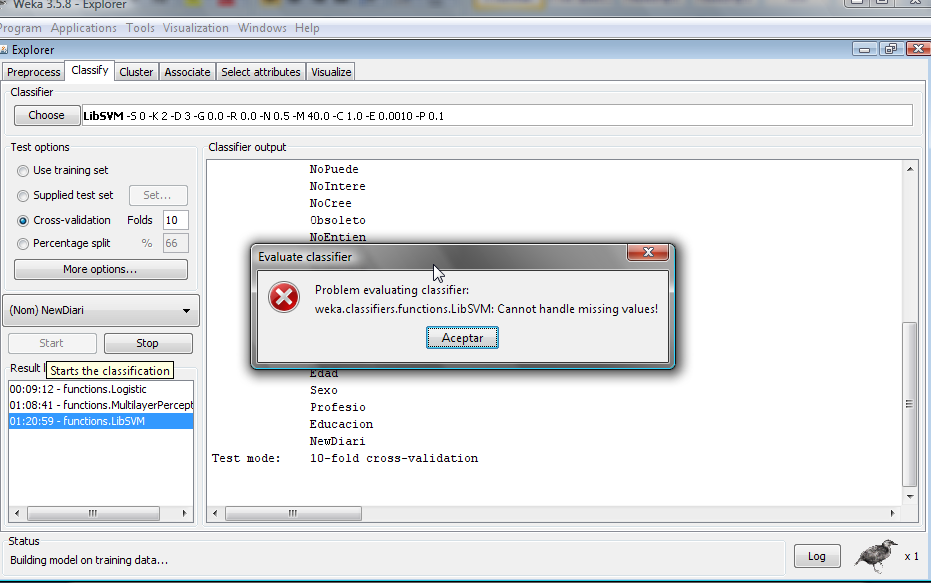


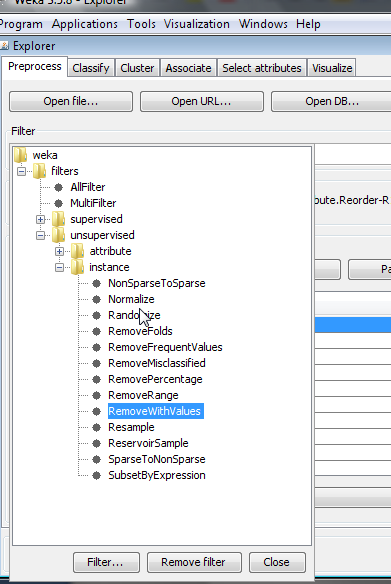
*Multilayer Perceptron* (MLP)con *decay True*

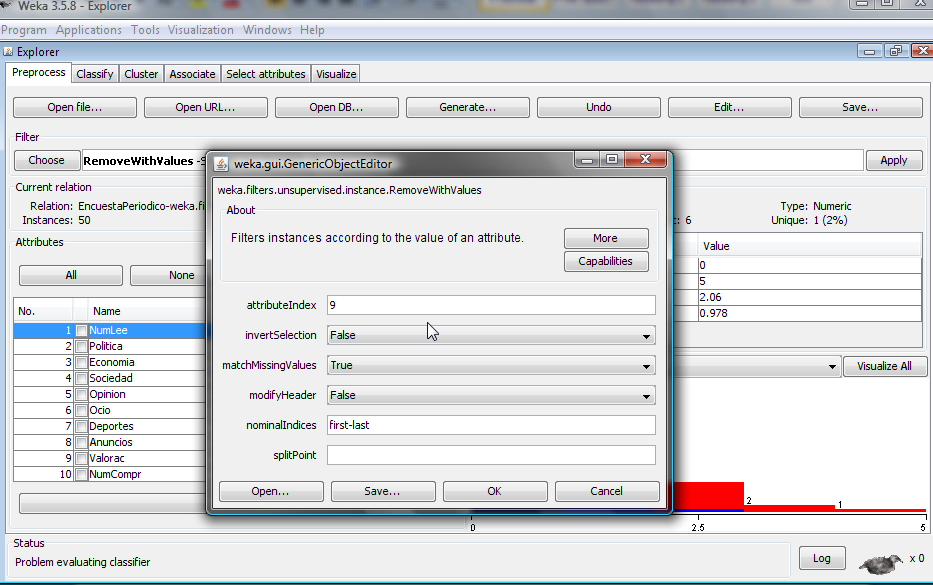


*LibSVM*

Como hay datos faltantes *LibSVM* no funciona.



En la clase del 1/12 explicó que hay datos faltantes en la variable Valoración que es la pregunta. Aparecen datos faltantes porque la respuesta no tiene sentido. Para poder usar *LibSVM* eliminó los 5 datos faltantes. Lo hizo aplicando el *filter unsupervised Instance* – *Removewithvalues*. Para esto fue a la pestaña *Process*



attributeIndex: 9

invertSeleccion:False

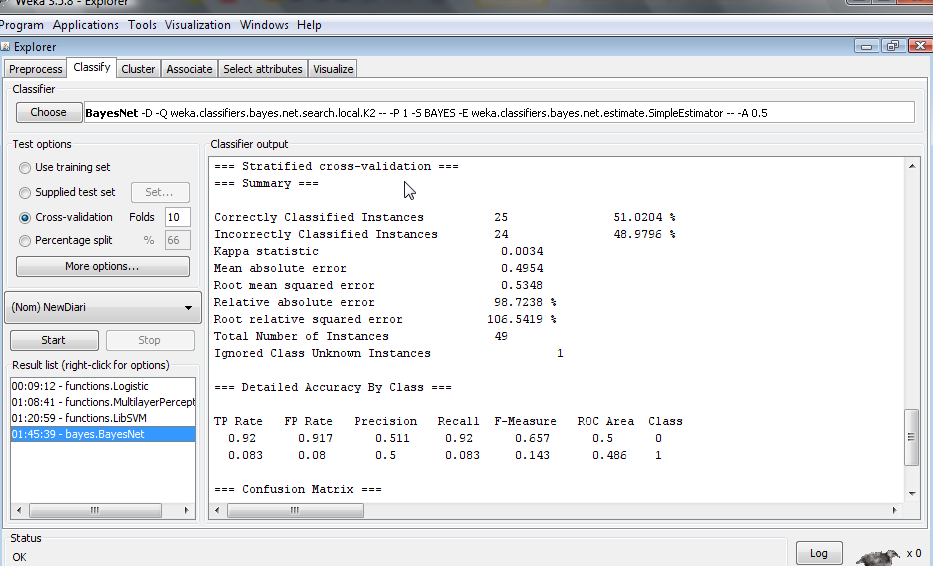
matchMisssingvalues:True

splitPoint: Lo dejó en blanco

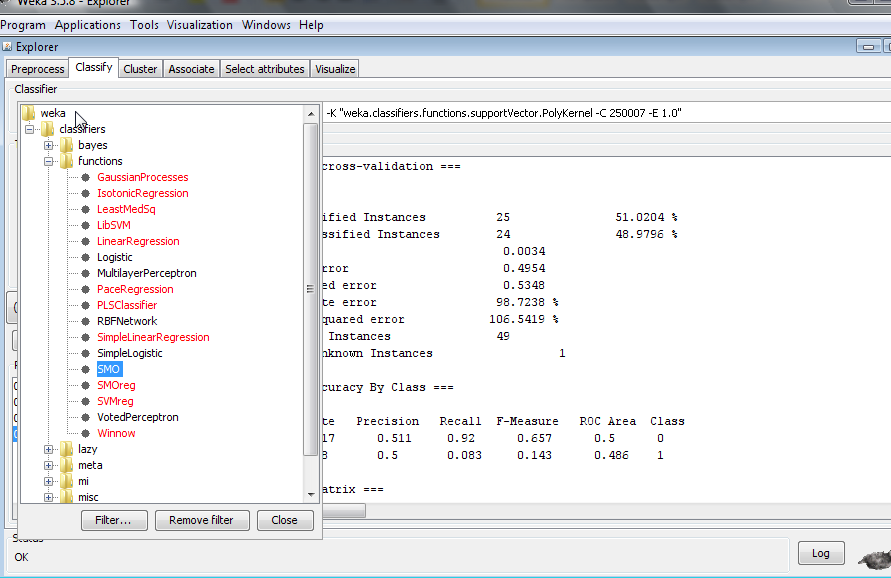
matchMissingValues

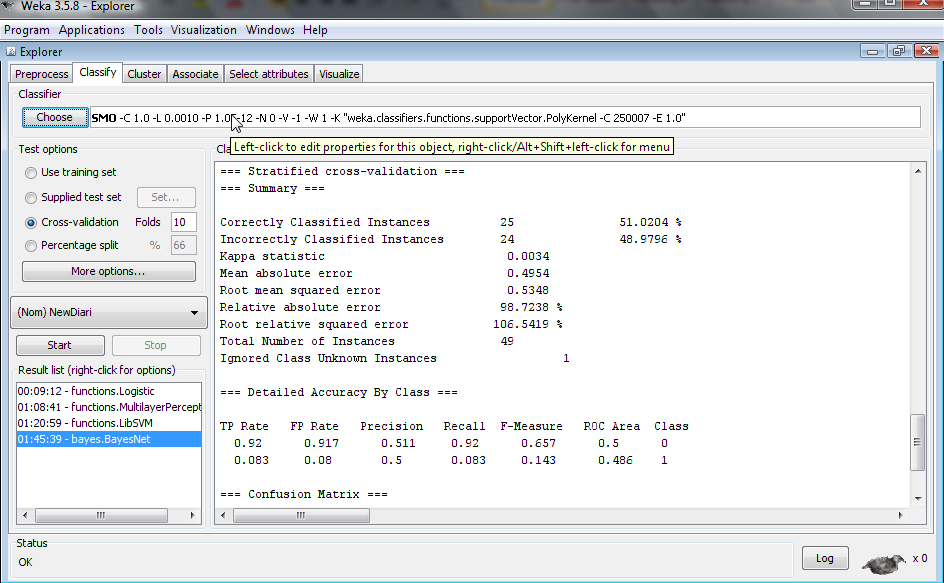
elimina los datos faltantes

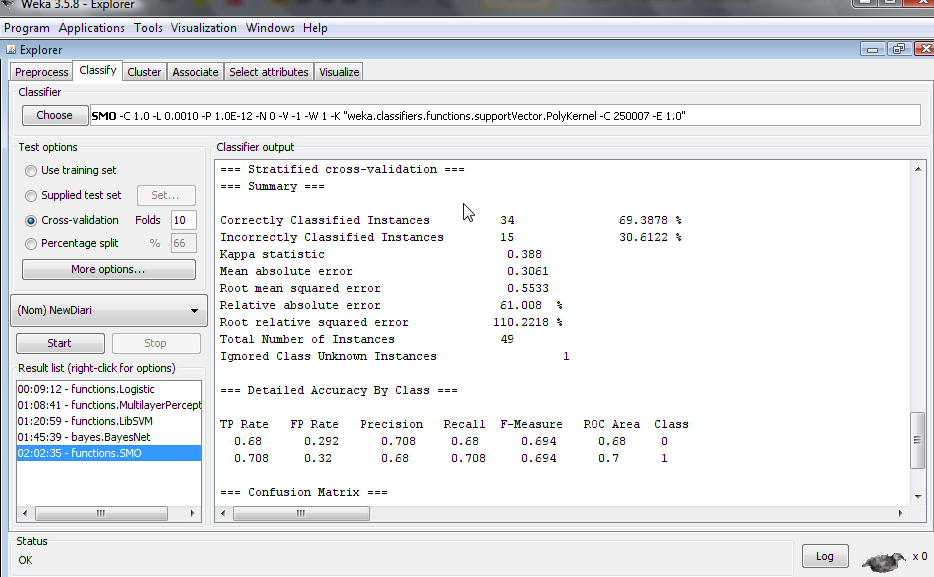
Al aplicar el filtro se ve que quedan 45 datos en vez de 50:

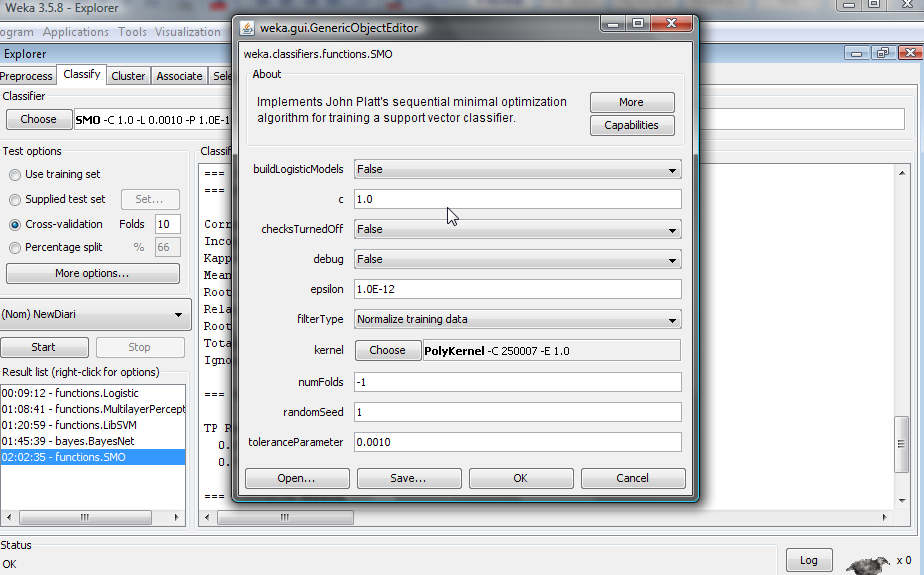


*BayesNet*

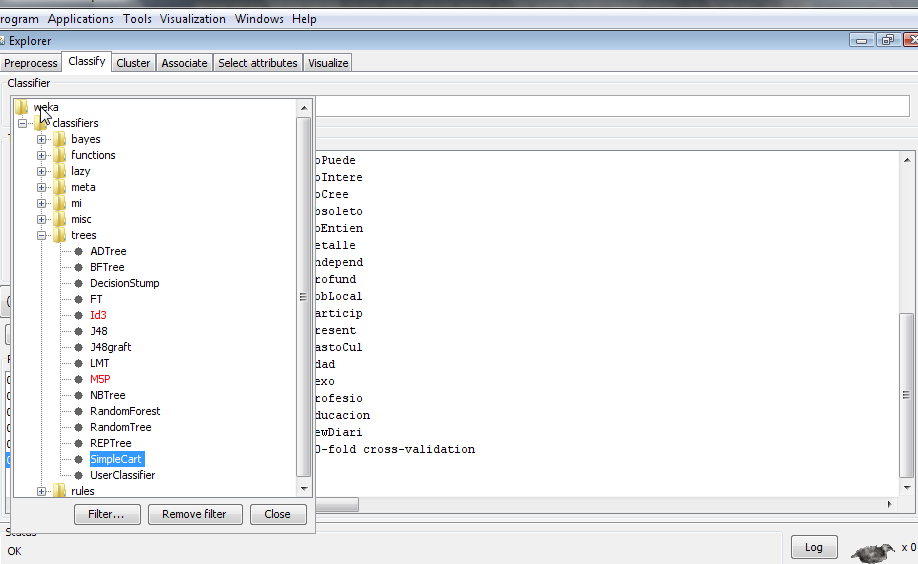
*SMO*



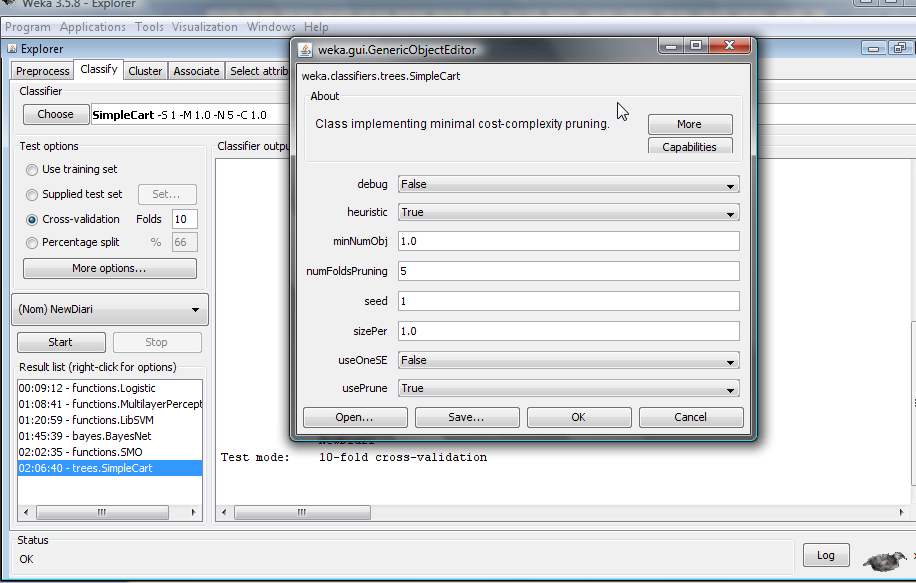




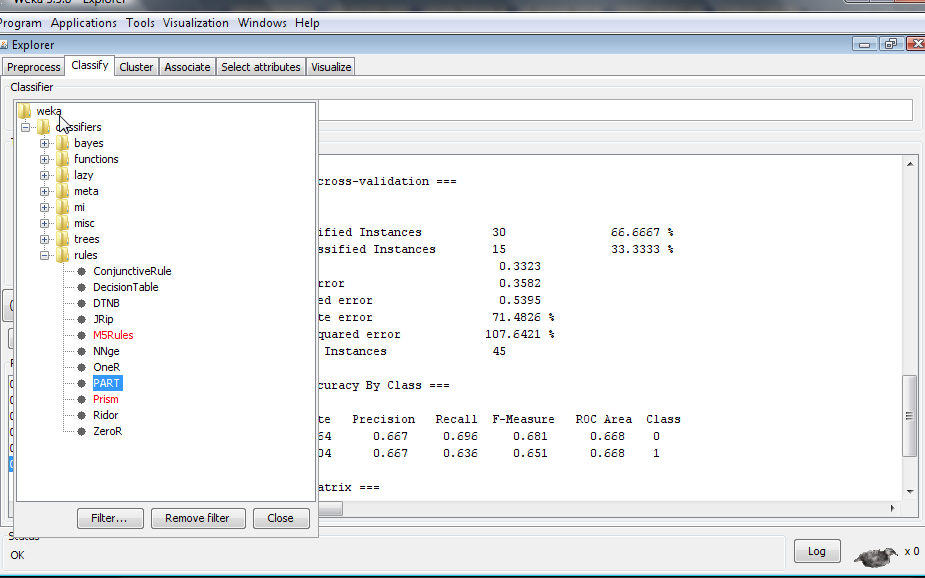
De la carpeta *Tress* seleccionar *SimpleCart*



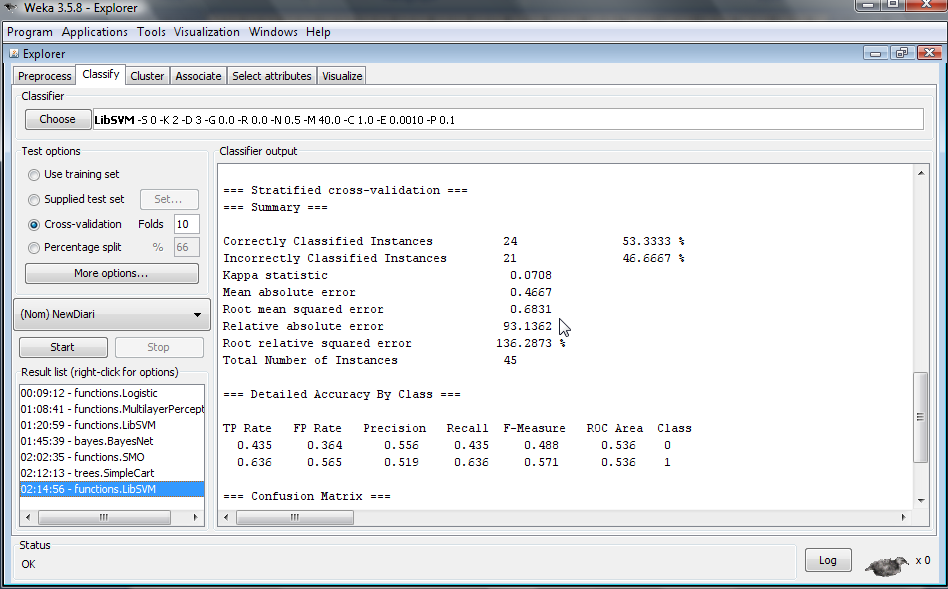
Poner en *minNumObj* 1



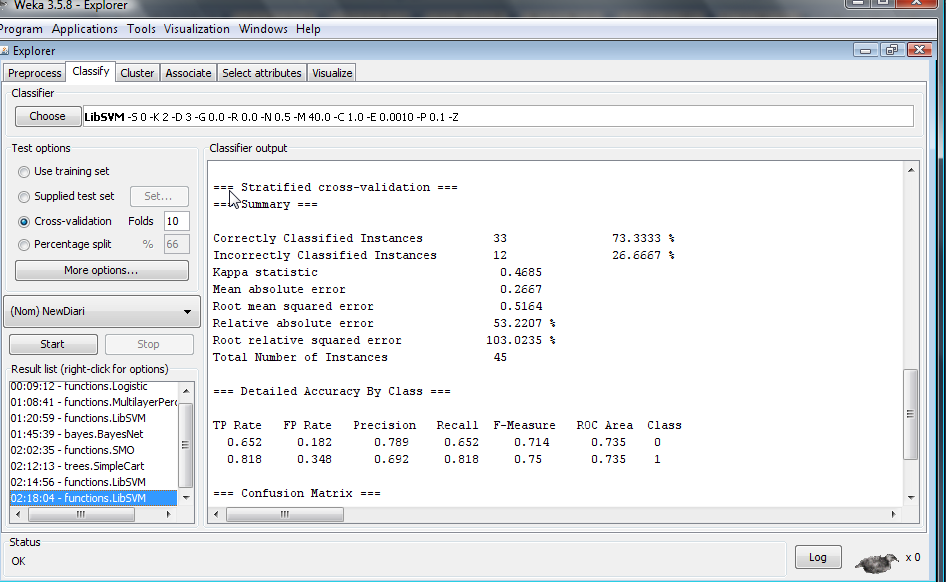
De la carpeta *Rules* seleccionar *PART*



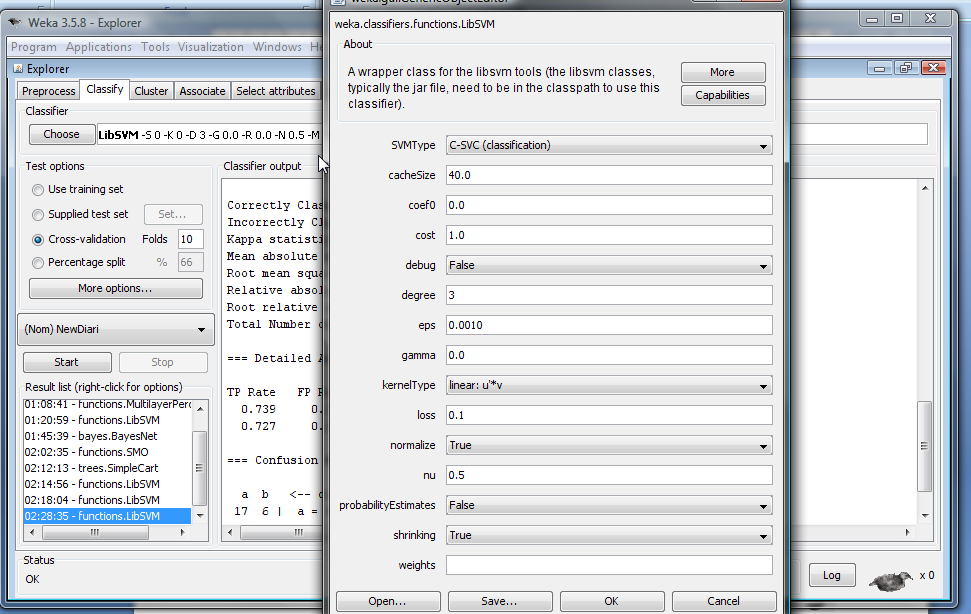
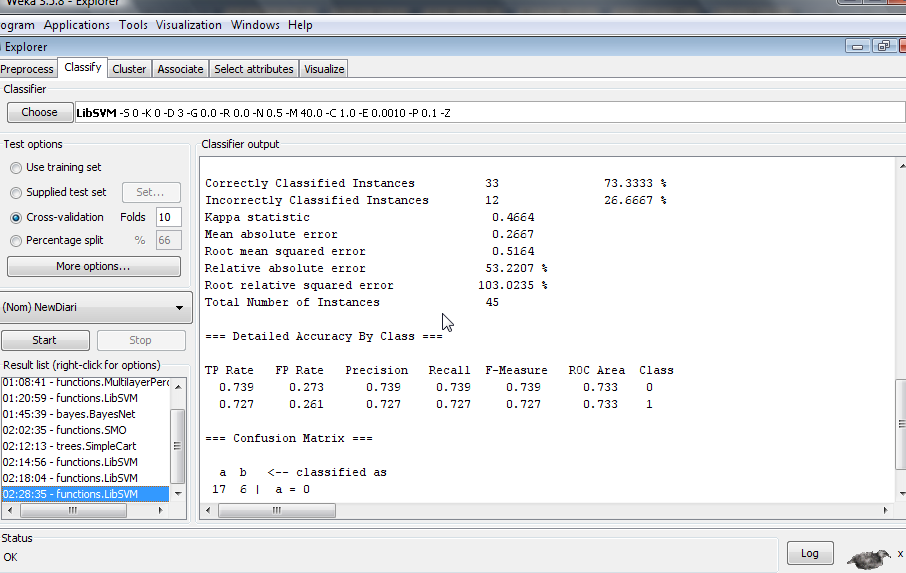
*LibSVM*

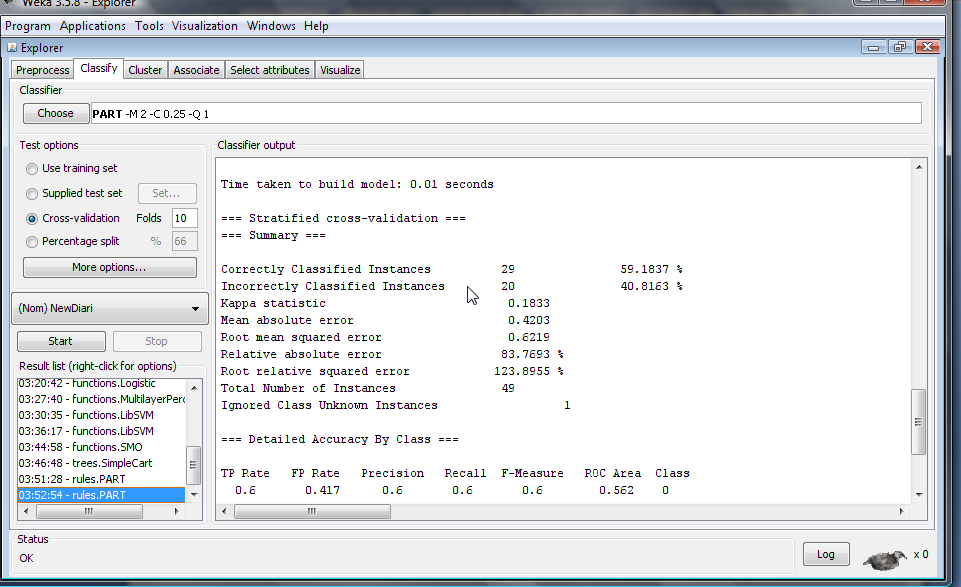
Aplicando *LibSVM* ahora después de eliminar los datos faltantes. Primero lo hice sin normalizar

Después lo hice normalizando y mejoró

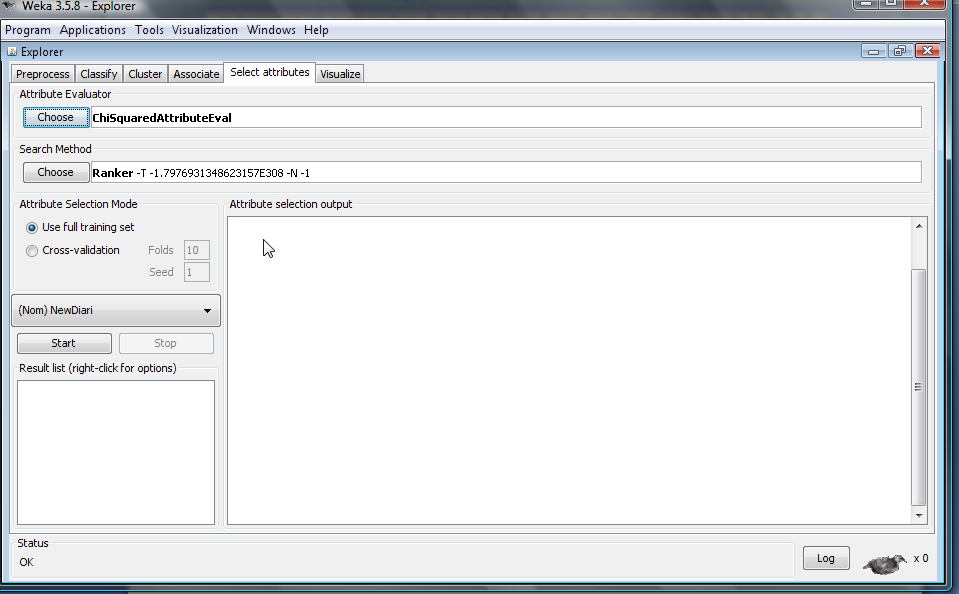


En la clase dijo que *normalize linear* 1. No encontré esto. Lo que hice fue en *kernelType* seleccioné *linear*

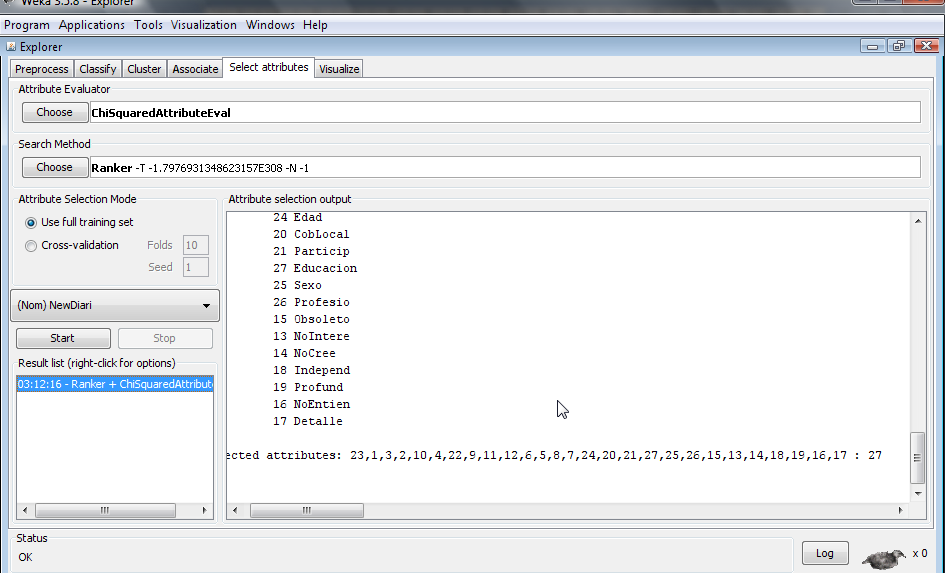


De la carpeta *Rules*: *PART*

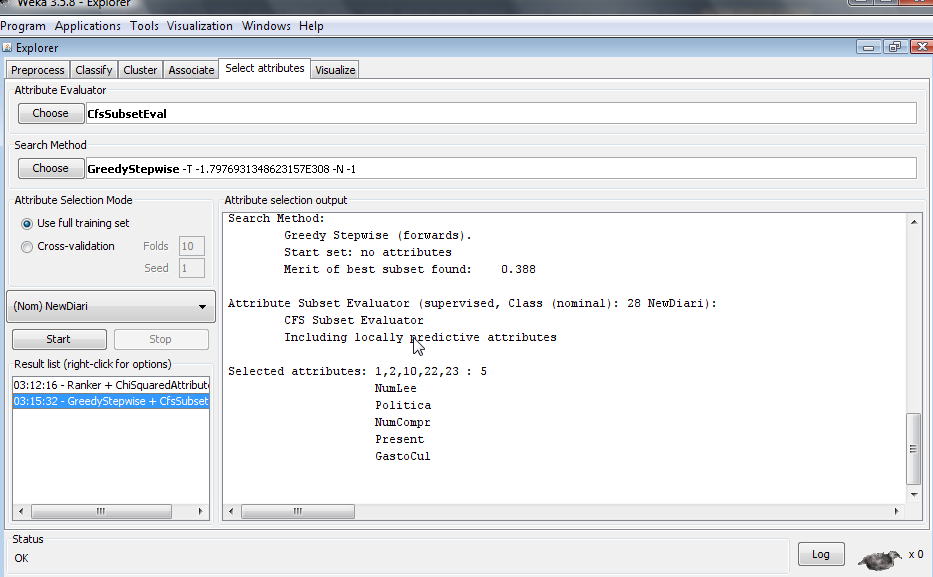
Seleccionando *Atributes*. Ir a la pestaña *Select attributes* y el primer *evaluator* que usó fue



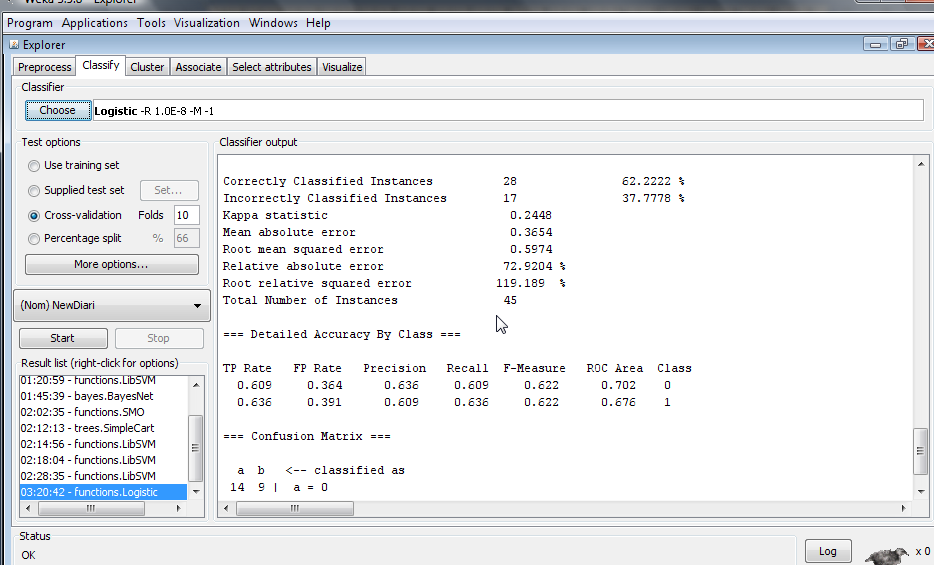
Seleccionó el ChiSquare Attribute Eval



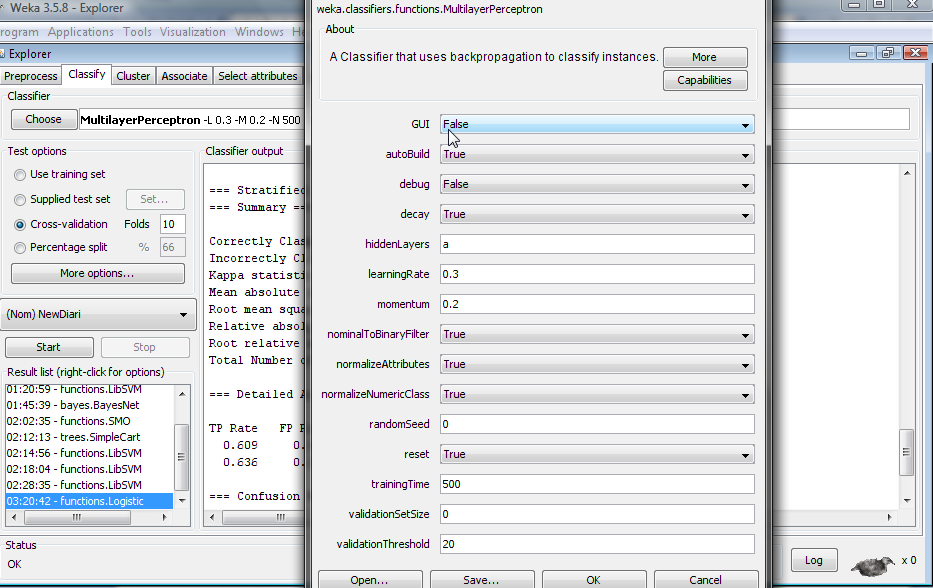
El *attribute* seleccionado fue Educación

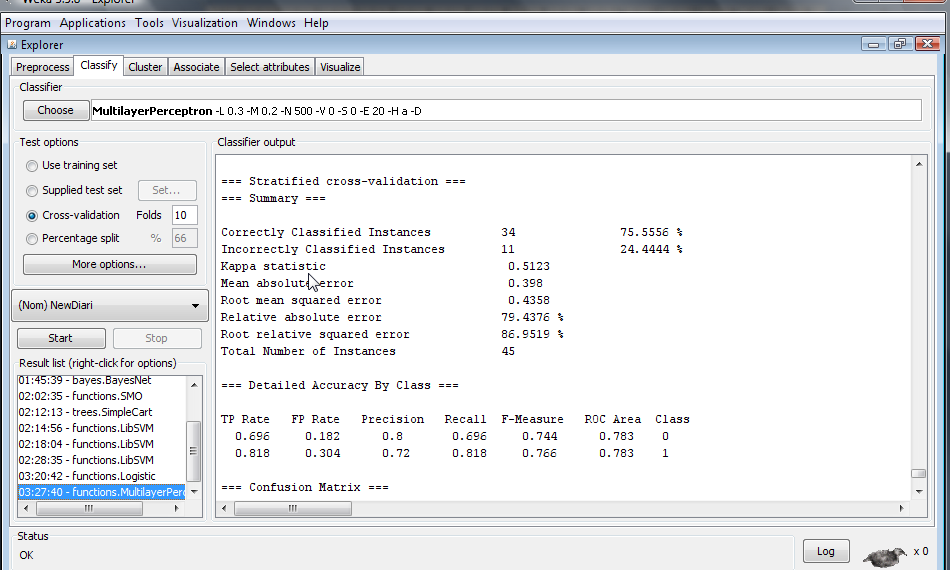
El segundo *evaluator* que usó fue *CfsSubsetEval* . En el rectángulo rojo están las variables seleccionadas

Probando otra vez los clasificadores

*Logistic*: Mejoró algo

*MLP*: con *decay True*

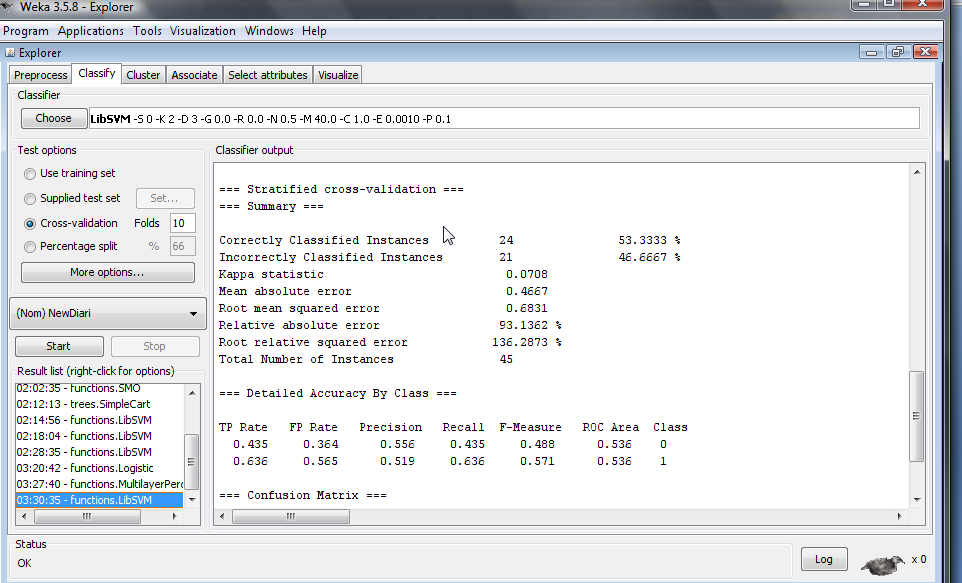




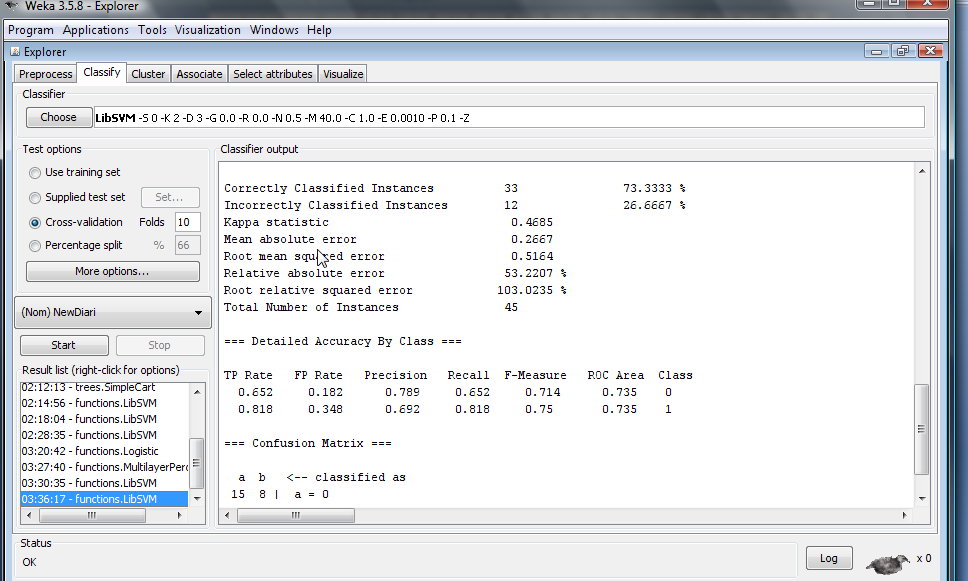
Mejoró la clasificación

*LibSVM*

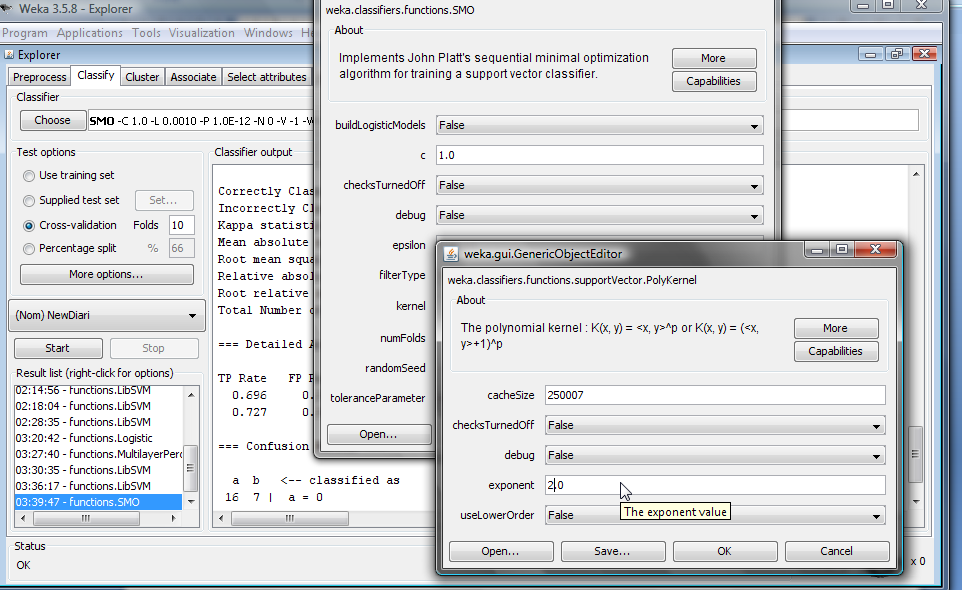
Sin normalizar



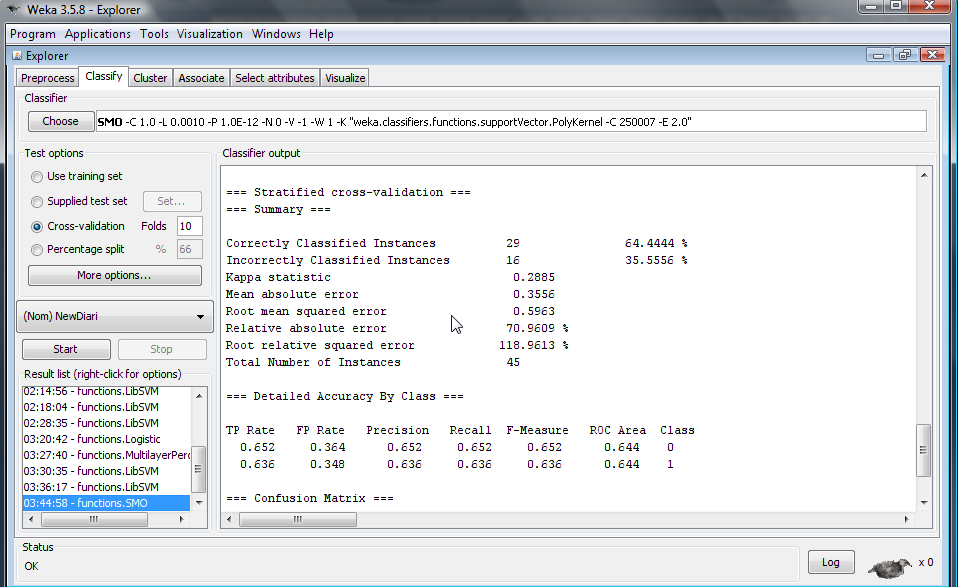
Normalizando. Mejoró la clasificación



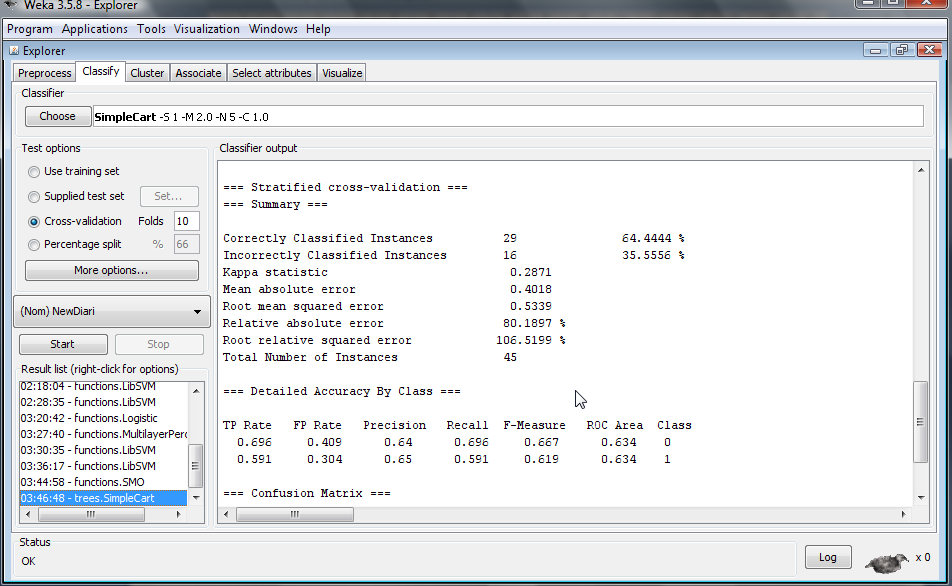
*SMO*. Con *polykernel 2*



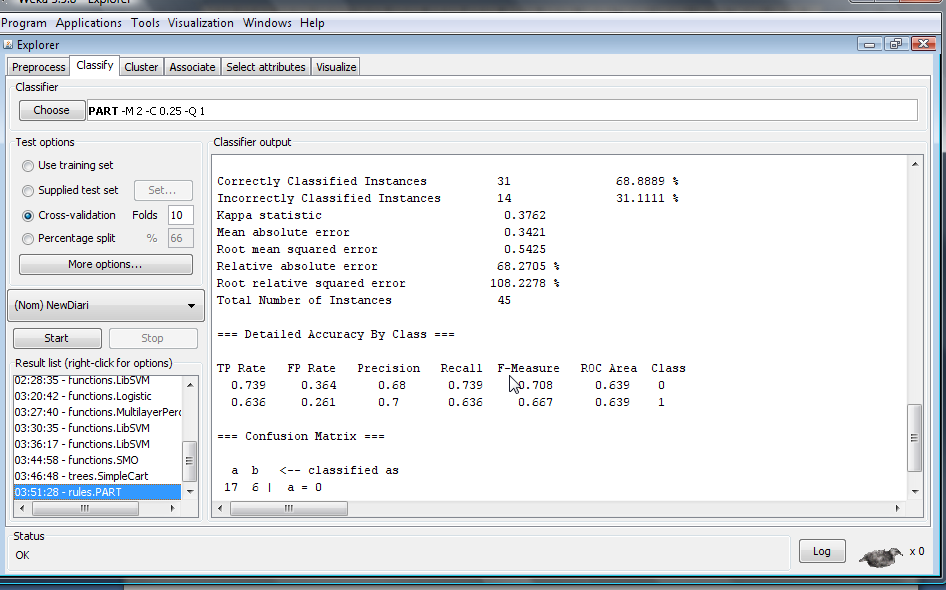
Mejoró



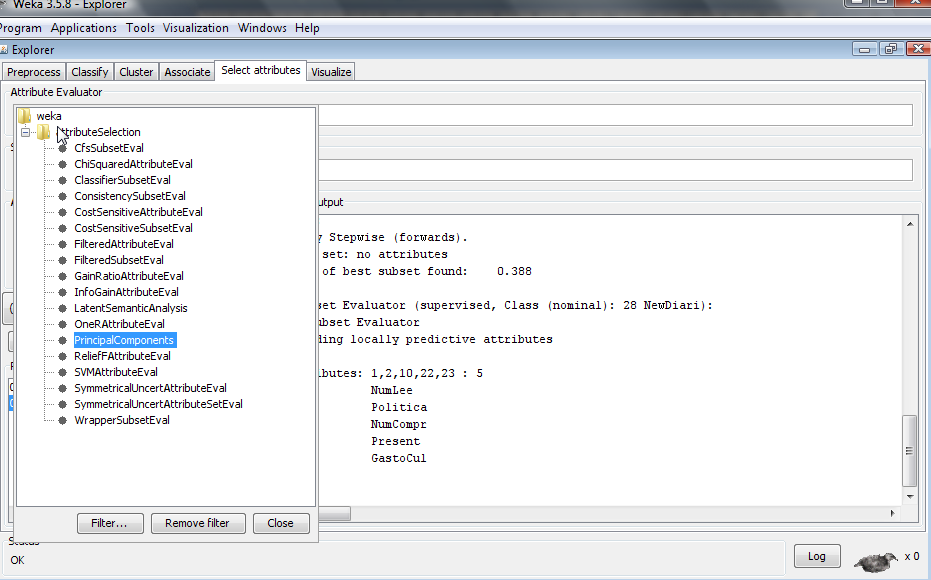
De la carpeta *Trees SimpleCart*. Mejoró



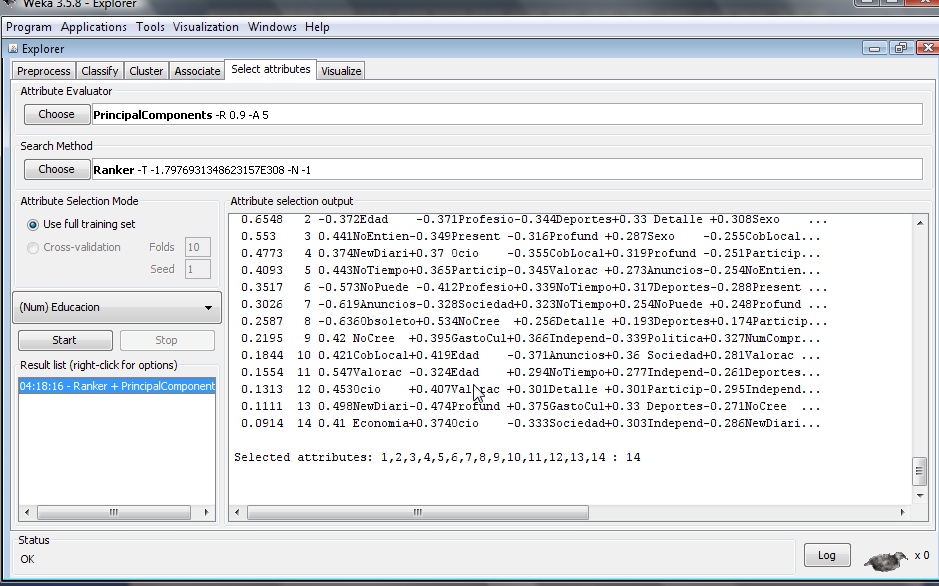
De la carpeta *Rules*: *PART*



Con el fichero original ir a la pestaña *Select Attributes* y seleccionar el *evaluator PrincipalComponents*



Selecciona el atributo 14 que es la Economía. Lo repetí, después de aplicar los filtros y me dio la 13



Aplicando clasificador *LibSVM y evaluator CfsSubsetEval*

