

Convert data into arff file Session 3

In this session, we will learn how to create an arff file from a spreadsheet.



A.- ORIGIN POINT

We start from the basis of having our data in the Google spreadsheet, which we left done at the end of the previous session.

c	く ち さ 母 宥	100% • £ % ·	0 , .00 123 Defaul.	• - 10 +	B <u>I</u> ≎ <u>A</u> ♦	• ⊞ 53 • ≣ • 3	<u>↓</u> • \$ • <u>A</u> • \$	^
A1	✓ fx Timestam	p						
	А	В	С	D	E	F	G	
1	Timestamp	Score	Select your course or occ	Insert your current age w	What is your gender?	1. How many deputies an	2. Which of these program	3. How mar
2	28/03/2024 11:47:17	1 / 20	1ESO	13	Female	100	Galileo	
3	28/03/2024 11:48:20	4 / 20	2ESO	15	Male	705	Erasmus+	
4	28/03/2024 11:49:38	4 / 20	1ESO	15	Nonbinary	751	Eurovisión	
5	28/03/2024 11:50:30	3 / 20	1ESO	13	Male	775	Frontex	
6	28/03/2024 11:51:42	11 / 20	Teacher	53	Male	751	Eurovisión	
7								
8								

B.- DATA PROCESSING AND DOWNLOAD

In Google's own spreadsheet we are going to process the data according to our needs with the following steps. It is important to check how many columns we have in the form. According to the Spanish form, that we are using, we have from column A to column Y, that is, 25 columns. 20 questions plus date, score, course or occupation, age and gender. We will eliminate the Timestamp column, so we will finally have 24 columns and, in the case of Spain, only 12 of them will have a score.

Our goal is to enclose all data in single quotes to convert it to the string data type and avoid problems with whitespace, accents, and control characters. We will do this in a new spreadsheet.

- **1.** We will move the Score column to the last position, column Y.
- 2. Let's create a new spreadsheet in the plus icon in the lower left area.
- 3. In the new spreadsheet we go to cell *A1* and we will write the following function """&'Form responses 1'!B1&""" to copy the data from cell *B1* and enclose it in single quotes.
- **4.** We copy and drag the relative reference A1 so that it does the same with the rest of the cells up to column X, one less than Y since Timestamp does not interest us, and up to as many records as we have. It should be like this.

٥	、 5 c 合 〒 100% -	£ % .0	.00 123 [Defaul 🔻 🚽	- 10 + 6	в <u>г</u> ÷ <u>А</u>	<u>è.</u> 🖽 🗄	3 - ≣ - ↓	▼ A ▼ \$: ^
A1:X6	▼ fx ="'"&'Form respo	onses 1'!B1&"'"								
	А	в	С	D	E	F	G	н	I.	J
1	'Select your course or occupation'	Insert your curre	'What is your gei	'1. How many de	2. Which of thes	'3. How many sta	'4. Which of the t	'5. How many of	'6. What is the E	'7. Which EU coi '8
2	'1ESO'	'13'	'Female'	'100'	'Galileo'	'6'	'Mark Zuckerber	'One (English)'	'Things can only	'Slovenia' 'T
3	'2ESO'	'15'	'Male'	'705'	'Erasmus+'	'12'	'Bono'	'Three (English,	'Ode to joy'	'Germany' 'E
4	'1ESO'	'15'	'Nonbinary'	'751'	'Eurovisión'	'28'	'Pope Francisco'	'Ten'	'We are the char	'France' 'C
5	'1ESO'	'13'	'Male'	'775'	'Frontex'	'50'	'All of them'	'Twenty-four'	'One nation unde	'Greece' 'L
6	'Teacher'	'53'	'Male'	'751'	'Eurovisión'	'12'	'All of them'	'Twenty-four'	'Ode to joy'	'Germany' 'C
7										
8										
9										
10										
11										
12										
13										
14										
10										
10										4 1
	+ ≡ 🗏 Form respor	ises 1 👻 She	eet1 👻							Count: 144
										jueves, 28 de ma

5. Now we can only download this new spreadsheet by clicking on *File/Download/Comma-separated* values (csv)

It is common that when downloading date-like columns in xlsx format, such as Score, it changes them to date format. To solve it we will follow the following steps:

A.- We create a new sheet and paste only the data to be converted into it.

B.- In an adjacent column we are going to use the **DAY** function, in Spanish DIA, which allows us to extract the day from a date, in our case the score, and converts it to a number in the first cell. Then we drag to do it in all of them. We see that it is already a number since it aligns it to the right.

	TEXTO -	≺ ✓ <i>f</i> _x =DIA(A1)
	А	В
1	05-dic	=DIA(A1)
2	06-dic	DIA(núm_de_serie)
3	06-dic	6
4	03-dic	3
5	05-dic	5
6	07-dic	7
7	05-dic	5

C.- Now we will go to our original sheet, we select the range of the data, we delete it and put the number format.

Pegar		0 ▼ A* A* <u>③</u> ▼ <u>A</u> ▼		≫ <mark>]</mark> 律律 國	Ajustar texto	Número	▼ ⇒0
P.C		13		Aimeacion	19	Numero	1.00
	A		В			С	
1	Timestamp	Score			Selecciona	a tu curso u	In
32	09/04/2024 12:57				3º ESO		
33	09/04/2024 12:58				3º ESO		
34	09/04/2024 12:58				3º ESO		
35	09/04/2024 12:58				3º ESO		
36	09/04/2024 12:59				3º ESO		
37	09/04/2024 12:59				3º ESO		
38	09/04/2024 12:59				3º ESO		

D.- Then we copy the range of formatted numbers and on the original sheet with the selected range, we select the **special paste** with the right button to click on **paste only values**. They will appear with decimals, we remove them in the indicated button.

🗶 🛃	▶) • (≌ - =						spair	n.xlsx - Microso	oft Exce	1				
Archivo	Inicio Inser	tar Dise	eño de págin	a Fórmulas	Datos	Revisar	Vista							
Ê	🔏 Cortar	Arial	* 10	ĂĂ	= = =	≫	Ajustar texto	Número			5			+
Pegar	Copiar *	N K	s - 🖸 -	<u>></u> - <u>A</u> -	≣ ≣ <mark>∎</mark>		🛾 Combinar y centrar 👻	ഈ ~ % 00	0 58	Form	iato ional *	Dar formato como tabla *	Estilos de celda *	Inserta *
Po	rtapapeles 🕞		Fuente	Gr.		Alineació	n G	Número		1 _M		Estilos		
	B32 -	· (=	<i>f</i> _x 5							Disminuir de	ecimale	es		
		А			В			С		Muestra va Ios decima	lores r les.	menos preciso	os disminuyer	ndo
1	Timestam	р		Score			Seleccion	a tu cur	so u	Inser	ta t	u edad	l actua	ıl ¿(
32	09/04	/2024	12:57			5,00	3º ESO						15	5 Fe
33	09/04	/2024	12:58			6,00	3º ESO						14	l Fe
34	09/04	/2024	12:58			6,00	3º ESO						15	δ Fe
35	09/04	/2024	12:58			3,00	3º ESO						15	i Fe
				1						-				

Teacher Mr. Carlos Alberto López Laínez – IES Ítaca Alcorcón – Spain – 2024

E.- If we wanted to keep the Score / 12 format then we have to select the entire range of cells, click on the right button and then on Cell Format, within the category select Custom and overwrite the Type value with 0'' / 12'' to keep the starting number, it would be 0, and add a literal with "/12"

😨 I 🛄	k 0 • (31 • 1 =			radio viev - Microsoft Even
Archivo	Inicio Insertar Diseño de página	Fórmulas Datos Revisar Vi	spe	spani,xisk - Iniclosoft Excel
Pegar	A Cortar Arial > 10	· A [*] A [*] ≡ ≡ ≡ ≫·· → A [*] A [*] ≡ ≡ ≡ → · A [*] = ≡ ≡ ‡ ‡ ₩	Ajustar texto Combinar y centrar	Personalizada Personalizada Formato Dar formato Estilos de Insertar Eliminar Formato
Po	rtapapeles 5 Fuente	Alineación		s Número s Estilos Celdas
	B39 ▼ (* <i>f</i> _x 4			
	А	В	1	Formato de celdas ? X
1	Timestamp	Score	Seleccio	O Número Alineación Fuente Bordes Relleno Proteger
35	09/04/2024 12:58	3 / 12	3º ESO	Categoría: General Muestra
36	09/04/2024 12:59	5 / 12	3º ESO	Moneda Contabilidad 200:
37	09/04/2024 12:59	7 / 12	3º ESO	Hora Porcentaje
38	09/04/2024 12:59	5 / 12	3º ESO	Traccon * # # # 0 ∈ _', * # , # # 0 ∈ _', - * € _', - @ _ Gentifica _* # , # # 0 € ∈ _', * # , # # 0 € ∈ _', - * € _', - @ _ Texto _* # , # # 0,00 € _', * # , # # 0,00 € _', - * * . ?? €', @ _
39	09/04/2024 12:59	4 / 12	3 200	Cspecial
40	09/04/2024 12:59	3 / 12	3º ESO	"Si"; "No" "Verdadero"; "Falso" "Activado"; "Activado"; "Desactivado"
41	09/04/2024 12:59	3 / 12	3º ESO	[[\$E-2] #.##0,00_;;[Rojo]((\$E-2] #.##0,00)
42	09/04/2024 13:00	8 / 12	3º ESO	Escriba el código de formato de número, usando como punto de partida uno de los códigos existentes.
43	09/04/2024 13:00	4 / 12	3º ESO	
44	09/04/2024 13:00	8 / 12	3º ESO	Aceptar Cancelar
4.5	00/04/0004 40:00	1 1 4 0	20 500	

F.- Then we would delete the new sheet and we could convert to the csv format with Excel itself using *Save As* and selecting *CSV* (*Comma Separated Value*) (*.csv) and continuing with the steps on the following pages.

🔣 Guardar como						×
$\leftarrow \rightarrow \land \uparrow$	> Es	te equi	po > Documentos >	~ Õ	Buscar en Docur	mentos 🔎
Organizar 🔻 Nu	eva ci	arpeta				
💻 Este equipo	^	No	mbre		Fecha de modifica	Тіро
👆 Descargas			Arduino		13/08/2023 11:54	Carpeta de archiv
Documentos		۲	Mis sitios Web		13/08/2023 11:54	Carpeta de archiv
Escritorio			Plantillas personalizadas de Office		13/08/2023 11:54	Carpeta de archiv
			Python Scripts		13/08/2023 11:54	Carpeta de archiv
inagenes			WindowsPowerShell		13/08/2023 11:54	Carpeta de archiv
Música			Wondershare		13/08/2023 11:54	Carpeta de archiv
🧊 Objetos 3D			Wondershare Filmora		13/08/2023 11:54	Carpeta de archiv
📑 Vídeos			Wondershare Filmora 9		13/08/2023 11:54	Carpeta de archiv
🏪 SISTEMA (C:)	~	<	-			>
Nombre de archivo:	Libro	1.csv				~
Tipe	CSV (delimi	ado por comas) (*.csv)			~
Autores:	Carlo	os	Etiquetas: A	gregar u	na etiqueta	
∧ Ocultar carpetas			<u>H</u> errami	ientas	▼ <u>G</u> uardar	Cancelar



This is not our case, but it is possible that in certain datasets we find the absence of null values. For example, if we do not know the age, occupation, etc. In these cases, Weka has a special character to indicate that there is a null, it is the ?, which we must insert as is.

Once the data has been downloaded in csv format, we must associate the file format with the notepad. To do this, we will right-click and choose *Open with*, and select *Notepad*. This way, forever, all csv files will be opened directly with notepad.

Our data will appear as follows.

UE TEST FINAL COPY (Responses) - Sheet1.csv: Bloc de notas	;		-		\times
<u>A</u> rchivo <u>E</u> dición F <u>o</u> rmato <u>V</u> er <u>A</u> yuda					
'Select your course or occu ers and video games?'",'04 '1ESO','13','Female','100' '2ESO','15','Male','705','f '1ESO','15','Nonbinary','7 '1ESO','13','Male','775','f 'Teacher','53','Male','751	upation,"')n . Select your ,'Galileo','6 Erasmus+','12 51','Eurovisi Frontex','50' ','Eurovisión	sert your favorite ','Mark Zu ','Bono'," ón','28',' ,'All of t ','12','A]	curre activ ucker ''Thre Pope chem' ll of	ent /ity perg ee (Fra ,'Tw the	a E n e m
<					>



The data appears raw apparently, but within the data we have the following order,

- \checkmark The first row is the question headings, the *columns*.
- \checkmark The following *rows* are the data itself, although we will only have a single row now.

Now we are ready to convert the csv file to arff format.

Be careful with double quotes

When downloading the file in csv format, it is very likely that some phrase will be enclosed in double quotes. Knowing that we do not have any double quotes that we want to keep in our data, we can then go to the top menu **Edit/Replace** and where we put double quotes we will put a blank space, then clicking on **Replace all**.

Reemplazar	×
Buscar:	<u>B</u> uscar siguiente
por:	Reemplazar
Coincidir <u>m</u> ayúsculas y minúsculas	Cancelar
Ajuste automático	

C.- CONVERT INTO ARFF FILE

Our goal in this section is to achieve the following structure compatible with arff files, which allows us to open the file with the Weka program.

*weather.numeric.arff: Bloc de notas	
<u>A</u> rchivo <u>E</u> dición F <u>o</u> rmato <u>V</u> er <u>A</u> yuda	
@relation weather	^
Mattribute outlook string	
attribute temperature num	anic
wattribute numidity numeri	
<pre>@attribute windy {TRUE, FA</pre>	LSE}
<pre>@attribute play {yes, no}</pre>	
@data	
Sullity, 85, 85, FALSE, 110	
sunny,80,90,TRUE,no	
overcast,83,86,FALSE,yes	
rainv.70.96.FALSE.ves	
<	>
Línea 3, columna 26 100% UNIX (LF) I	JTF-8:

The arff file is divided into 3 sections, each identified by a type of tag that always begins with @.

@relation name: every arff file must begin with this declaration in its first line, which will name the file.

@attribute name data_type: then we will include a line for each attribute or column that we have in our data set, indicating its name and the corresponding data type that it may be,

- ✓ Numerical Attributes: they can be real numbers, taking the point as a decimal separator, with numeric or real. We can also express whole numbers with integers. For example, *numeric*, *real*, *integer*.
- ✓ *Text attributes*: takes the string data type of the Java language for textual expressions. For example, *string*.
- ✓ Date attributes: we put date and, optionally, we will indicate the date format optionally, which can be of the type "yyyy-MM-dd HH:mm:ss", taking the Java parameters. For example, date "yyyy-MM-dd"
- ✓ Nominal attributes: These are data types defined by ourselves using a list of values that are separated by commas and enclosed in braces. They are the ones we should use whenever we have textual values. We can write them ourselves, if there are few values, or much better, let Weka do it with an attribute filter, we will see it in another session. For example, {sunny, cloudy, rainy}

@data: are the data records, each one on one line. We'll make sure that all rows have the same number of columns and that that number matches the number of *@attribute* statements we added earlier.



Usually we must assign the correct data type to each attribute, however the A Priori algorithm only allows attributes that are textual and nominal. Therefore, we will put all the attributes as string.

We will follow the following steps for the conversion.

- 1. We will write the name of the file putting @*relation 'test EU'* as the first line
- **2.** We will separate the different sections with carriage returns.
- **3.** We will put each column in a row.



- 4. We must prepend @*attribute* in front of all columns and remove the final commas from each question.
- 5. We will change the text of the first 3 general questions about the course or occupation, age and gender to *course_occupation*, *age*, *gender*.
- 6. We must also simplify the names of the columns to avoid long names by q1, q2, oq1, oq2, ...
- 7. All columns must be textual due to *A Priori algorithm* so we will mark them as *string*.
- 8. We will put the @*data* section before all the data but before we must delete the row with the column names.

The result that we should obtain will be the following.

```
٥
*UE TEST FINAL COPY (Responses) - Sheet1.csv: Bloc de notas
Archivo Edición Formato Ver Ayuda
@relation 'test EU'
Wattribute 'course_occupation' string
Wattribute 'age' string
Wattribute 'gender' string
Wattribute q1 string
@attribute q2 string
@attribute q3 string
@attribute q4 string
@attribute q5 string
@attribute q6 string
@attribute q7 string
@attribute q8 string
@attribute q9 string
@attribute q10 string
@attribute q11 string
@attribute q12 string
@attribute oq1 string
@attribute oq2 string
@attribute oq3 string
@attribute oq4 string
@attribute oq5 string
@attribute oq6 string
@attribute og7 string
@attribute oq8 string
@attribute 'Score' string
@data
 .
1ESO','13','Female','100','Galileo','6','Mark Zuckerberg','One (English)','Things can only get better','Slovenia','Türkiye','The Euro
                                                                                                                             Línea 5, columna 27 100% Windows (CRLF)
                                                                                                                                                                     UTF-8
```

9. Now we have to save the file as arff. To do this we will go to *File/Save as...* and we will choose the name that we have with the arff extension, for example *finaltestEU.arff*, and in *type* we will put *All files* (*.*) and click on *Save*.

🗐 Guardar como							×
← → ~ ↑ 🛱 > E	ste equipo > Documentos >			~ (ອ Buscar en Docum	ientos	9
Organizar 🔻 Nueva	carpeta					•	?
🗸 🛄 Este equipo 🔥	Nombre	Fecha de modifica	Тіро	Tamaño			
> 🕂 Descargas	Arduino	13/08/2023 11:54	Carpeta de archivos				
> Documentos	Mis sitios Web	13/08/2023 11:54	Carpeta de archivos				
> Escritorio	📙 Plantillas personalizadas de Office	13/08/2023 11:54	Carpeta de archivos				
> Imáganas	Python Scripts	13/08/2023 11:54	Carpeta de archivos				
> imagenes		13/08/2023 11:54	Carpeta de archivos				
> J) Musica	Wondershare	13/08/2023 11:54	Carpeta de archivos				
> 🧊 Objetos 3D	Wondershare Filmora	13/08/2023 11:54	Carpeta de archivos				
> 📑 Vídeos	Wondershare Filmora 9	13/08/2023 11:54	Carpeta de archivos				
> 🏪 SISTEMA (C:)	Zoom	13/08/2023 11:54	Carpeta de archivos				
> _ DATOS (D:)	Database1.accdb	27/02/2024 12:17	Microsoft Access	368 KB			
-	DSC_0273.JPG	13/04/2019 18:36	Archivo JPG	1.407 KB			
> 💣 Red 🗸 🗸							
Nombre: FIN	ALtestEU.arff						~
Tere Ted	a las sustriums (# #)						
Tibo: Togo	os los archivos ()						~
 Ocultar carpetas 			Codi <u>f</u> icación: UTF-8		 <u>G</u>uardar 	Cancel	ar

10. In the event that we have the test divided into several forms, we will only have to copy and paste the new data after the last existing ones since the columns will be the same.

D.- OPEN ARFF FILE WITH WEKA

Once the file is finished, we will open Weka and press the *Explorer* button to enter into graphical interface. In the graphical interface we will load the path file where it was saved by clicking on the *Open File* ... button, and the image that will appear will be the following.



Weka Explorer Beautiment Cluster Associate Select attributes Visualize Interactive Parallel Coordinates Plot Vis	ualize 3D Forecast Projection Plot	- 6 ×
Open file_ Open URL. Open DB_ Gen	rateUndoEdit	Save
Chose None		Apply Stop
Current relation	Selected attribute	
Relation: test EU Attributes: 23 Instances: 1 Sum of weights: 1	Name: course_occupation Type: String Missing: 0 (0%) Distinct: 1 Unique: 1 (100%)	
Attributes		
All None Inust Dallars		
An None Invent Patern		
No. Name		
2 399		
4 _ q1		
5 _ q2 6 _ q3	Class: oq8 (Str)	Visualiz
7 _ q4 8 _ q5		
9 q6		
11 98		
12 d9 13 q10	Attribute is neither numeric nor nominal.	
14 q11 15 q12		
16 0 0 0 1		
Remove		
Status		
ОК		Log

However, we do not get any histogram of each attribute. This is because attributes are required to be nominal, allowing only written values, and enclosed in curly braces. We will apply an attribute filter.

✓ In the same *Preprocess tab* we have *Filter section* we will press the *Choose button* and within the *filter/unsupervised/attribute* options to choose the *StringToNominal filter*, which should be written in the text box.

eprocess	Classify	Cluster	Associate	Select attr	ibutes		Preprocess	Classify	Cluster	Associate
						-	Treprocess	Oldabiliy	oldater	Associate
Ope	en file		Op	en URL			0	an file		
r							U Op	en file		
							Filtor			
🗎 weka										
🔻 🚞 filter	rs						Choose	StringToble	minal D	act
A	llFilter						Cilous	Sangrona		สอเ
🕒 N	lultiFilter						Current relativ			
F	RenameRel	ation					Current relation	211		
	unervised						Relation:	test EU		
V 🔤 u	Insupervise	d					Instances:	1		
· ·						l				
		uster								
	AddEx	pression								
	AddID									
	AddNo	ise								
	🕒 AddUs	erFields								
	📄 AddVa	lues								
	Cartes	ianProdu	ict							
	Center	r 								
	Chang	jeDateFo	rmat							
- I.	Cluste	rMember	chin							
	Conv		amp							
	DateT	Numerio								
	Discre	tize								
	FirstO	rder								
•	📄 FixedD	lictionary	StringToWord	Vector						
	📄 Interqu	lartileRar	ige		V					
	Eit	tor	Remove filte							
			Comove line							
		cemissin	gvalues		_					
	📄 Repla	ceMissin	gWithUserCo	nstant						
	📄 Repla	ceWithMi	ssingValue							
	📄 SortLa	abels								
	Stand	ardize								
	String	ToNomin	al							
	Sung	rowordv	ector							
	Swap	/alues	-							
	Times	eriesDel	ia noloto							
	Trans	eries i ra	islate							
	rrans	pose								
	inctonec.									

Teacher Mr. Carlos Alberto López Laínez – IES Ítaca Alcorcón – Spain – 2024

✓ Next we will click on the text box that says StringToNominal to configure the filter and the following window will appear.

🥥 weka.gui.GenericObject	Editor	×
weka.filters.unsupervised.a	ttribute.StringToNominal	
About		_
Converts a range of stri values) to nominal (set attributeRange	ng attributes (unspecified number of More number of values). Capabilities	
debug	False	•
doNotCheckCapabilities	False	•
Open	Save OK Cancel	

✓ In the *attributeRange* text box we must write the numbers of attributes that we want to nominalize. They are all the ones that we put as string in the arff file. Therefore, we must write in the text box 1-24 to indicate the interval between attribute 1 and 24, both inclusive. And then press the *OK* button. The filter will change to the image below.

	🜍 weka.gui.GenericObjectEditor	×
	weka.filters.unsupervised.attribute.StringToNominal	
	About Converts a range of string attributes (unspecified number of values) to nominal (set number of values). Capabilities attributeRange 1-24 debug False doNotCheckCapabilities False Open OK Cancel	
Filter Choose StringToNominal -R 1-24		Apply

✓ Then we will press the *Apply* button to the right of the filter and the histograms will appear.

F.- SAVE DATASET

After making all the previous changes and before passing the algorithm, we must save the dataset with all the changes made. To do this, in the *Preprocess* tab we will click on the *Save...* button and we will save it with the name we want. This file will be our final dataset.

🖉 Weka Explorer					~
Preprocess Classify Cluster Associate Select attribut	es Visualize Interactive Parallel Coor	dinates Plot Visualize 3D Forecast	Projection Plot		
Open file Open URL	Open DB	Generate	Undo	Edit	Save
Filter Gua	rdar			×	
Choose StringToNominal -R 1-23 Buscar	en: 🔲 Descargas				Apply Stop
Current relation	SCARGAS		Invoke options dialog		
Relation: test EU-weka.filters.unsupervised.attribu	4j-materiales-master stEU.arff			Typ Uniqu	e: Nominal e: 1 (100%)
Attributes				1	Weight
					1
All None					
No. Name					
2 age					
3 gender					
4 _ q1					
6 q3					Visualize A
7 🗌 q4					
8 95				L	
9 q0				2	
11 g8 Nombr	e de archive EINALtestELLaff				
12 q9	The archite				
13 q10 Archivo	s de tipo: Arff data files (*.arff)				
15 012					
16 oq1			Guardar Cancelar		

G.- A PRIORI ALGORITHM

Finally, we can now go to the *Associate* tab, choose the A Priori algorithm in the *Choose* button, click it in the name of text box and configure it.

Weka Explore	er				
Preprocess	Classify	Cluster	Associate	Select attributes	Visua
Associator Choole	Apriori -N	р-то-с	0.9 -D 0.05 -	U 1.0 -M 0.1 -S -1.0	-c -1
Start			Stop	Associator out	put

- ✓ *minMetric*: we must indicate the minimum percentage, by one, that we want the rules it finds to comply with. For example, we will put 0.5.
- ✓ *numRules*: we indicate the number of rules that we want to appear. For example, we will put 200.

🥥 weka.gui.GenericObject	Editor	;
weka.associations.Apriori		
Class implementing a	n Apriori-type algorithm.	More Capabilities
car	False	•
classIndex	-1	
delta	0.05	
doNotCheckCapabilities	False	•
IowerBoundMinSupport	0.1	
metricType	Confidence	•
minMetric	0.5	
numRules	200	
outputitemSets	False	
removeAllMissingCols	False	•
significanceLevel	-1.0	
treatZeroAsMissing	False	
upperBoundMinSupport	1.0	
verbose	False	/
Open	Save OK	Cancel

Then we click on *OK*, and we execute it on the *Start* button, we will automatically get 200 rules with a minimum of 50% probability, ordered by highest to lowest probability.

```
Best rules found:
 1. age=13 2 ==> course_occupation=1ESO 2
                                             <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
 2. Score=4 2 ==> age=15 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
 3. age=15 2 ==> Score=4 2
                            <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
  4. q3=12 2 ==> gender=Male 2 <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
 5. q4=All of them 2 ==> gender=Male 2 <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
                                        <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
 6. q5=Twenty-four 2 ==> gender=Male 2
 7. g6=Ode to joy 2 ==> gender=Male 2 <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
 8. q7=Germany 2 ==> gender=Male 2 <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
 9. q12=The three previous countries 2 ==> gender=Male 2
                                                           <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
10. oq2=None of them 2 ==> gender=Male 2 <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
11. q2=EurovisiÃ<sup>3</sup>n 2 ==> q1=751 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
12. ql=751 2 ==> q2=EurovisiÃ<sup>3</sup>n 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
13. q8=Croatia 2 ==> q1=751 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
 14. q1=751 2 ==> q8=Croatia 2
                                <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
 15. q9=The European Union as a whole 2 ==> q1=751 2
                                                       <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
16. q1=751 2 ==> q9=The European Union as a whole 2
                                                      <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
17. ql0=Every five years 2 ==> ql=751 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
18. ql=751 2 ==> ql0=Every five years 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
19. oq5=Laptop 2 ==> q1=751 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
20. ql=751 2 ==> oq5=Laptop 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
21. q8=Croatia 2 ==> q2=EurovisiÃ<sup>3</sup>n 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
                                       <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
22. g2=EurovisiÃ'n 2 ==> g8=Croatia 2
 23. q9=The European Union as a whole 2 ==> q2=EurovisiÃ'n 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
 24. q2=EurovisiÃ'n 2 ==> q9=The European Union as a whole 2
                                                              <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
 25. gl0=Every five years 2 ==> g2=EurovisiÃ'an 2 <conf:(1)> lift:(2.5) lev:(0.24) [1] conv:(1.2)
```

a 11