Práctica cero. Primeros pasos con WEKA. Tras seguir estas instrucciones enviar un documento con los resultados obtenidos.

Tras descargar el programa y echarle un vistazo .

1.- Desde el panel EXPLORER abrir el archivo adult.csv

Marcar la opción invoke options dialog

Cambiar filedSeparator por ; ( por defecto es , Sin embargo si el archivo csv procede de Excel ( en español) usará ;

2.- Visualizar los datos en el panel de visualización:

Contestar a preguntas del tipo:

¿ Cuál es la media de la edad ¿Qué edad tiene la persona de más edad? ¿Qué empleo es el más frecuente?

Seleccionando la edad y viendo el gráfico ¿qué dirías de los ingresos de los más jóvenes 17-19 años?

Notar las diferencias en la visiualización de variables numéricas y de variables nominales.

Etc.

3. – Llevar a cabo algunas clalsificaciones planel Classify :

* ¿qué variable es la que se va a clasificar por defecto? ¿Cuáles son las opciones de validación test/ training por defecto)?
* Hacer una clasificación con ZeroR

El clasificador por defecto es ZeroR: usarlo ( con todas las opciones por defecto).

ZeroR consiste en clasificar todas las instancias como la más frecuentes.

Curiosamente no funciona mal ¿Por qué?

* Usar el clasificador OneR

One R usa sólo la información de una variable para clasificar ¿cuál?

También parace funcionar bastante bien

4.- Clasificar los individuos según ingresos con dos procedimiestos cualqesquiera ( de dos grupos distintos ) Y comparar con ZeroR y OneR.

5.- Cambiar las condiciones de validación a

* SPLIT ( por defecto) reservar una parte de los datos para test. Y repetir el proceso.
* Usar el conjunto de datos de entrenamiento

6.- Volver al panel de prepoceso.

Elimina dos atributos ( MARCAR Y TECLA REMOVE) que sospeches que importan poco y repite el proceso 4 y 5

¿Cómo sales parados los dos clasificadores de antes ( no ZeroR y OneR, los otros dos)?

7.- Volver al panel de prepoceso.

* Deshacer la eliminación de 6 Tecla UNDO.
* Usa el filtro : filters >unsupervised.>instance.> Resample

Este filtro muestrea aleatoriamente las instancias

Edita sus propiedades ( marcar con botón derecho del ratón: show properties)

Modifica: la semilla aleatoria y el tamaño relativo de la muestra ( % ) por ejemplo a 3 y 50 u otros valores.

 

Aplica el filtro : tecla apply: observa cómo se reducen las instancias.

 8.- Repetir la clasificación como en 6

Comentar resultados.

Finalmente, deshacer el ultimo filtro usado ( date cuenta que siempre se puede)