



### 1. Objetivo del laboratorio

*Aprender el uso de la herramienta de Data Mining Weka.*

### 2. Elementos a utilizar:

- Weka

### 3. Práctica 1 (Reglas de asociación)

#### Objetivo

- 1) Abrir weka
- 2) Seleccionar la opción Explorer
- 3) Abrir archivo "weather.arff"
- 4) Pulsa en la pestaña "Associate"
- 5) Pulsa en el cuadro que pone A priori
- 6) Configura los parámetros. Aquí el algoritmo empieza con un soporte mínimo en 1 y va bajando hasta el lowerBoundMinSupport marcado (0.1 por defecto). Esa bajada se hace en relación a la delta que por defecto es 0.05. metricType nos indica como se van a evaluar las reglas. Por último, outputsItemSets permite obtener todos los itemsets frecuentes cuando su valor está a True (cambiarlo a este valor). Realiza los siguientes pasos (para ejecutar el algoritmo pulsa Start) y copia los resultados en un documento Word.
  - a. Prueba con una configuración con soporte mínimo por defecto y confianza 0.6.
  - b. Prueba con una configuración con soporte mínimo 0.6 y confianza igual que la anterior.
  - c. Repite el proceso anterior con métrica para evaluar reglas Lift.
- 7) Escribe después de los resultados que diferencias hay entre cada uno de los pasos anteriores. Razona la respuesta.

### 4. Práctica 2 (Árboles de decisión)

#### Objetivo

- 1) Abrir archivo "bank.csv"
- 2) Seleccionamos la clase para clasificar. En este caso mortgage. Para decidir si darle una hipoteca o no.
- 3) Pulsa en la pestaña "Classify"
- 4) Pulsa el botón "Choose"
- 5) Selecciona el algoritmo J48 (C 4.5) en la carpeta trees.
- 6) Pulsa en el cuadro que pone J48. Realiza los siguientes pasos (para ejecutar el algoritmo pulsa Start) y copia los resultados en un documento Word.
  - a. Prueba con la confianza por defecto (0.25).
  - b. Aumenta la confianza a (0.75)
- 7) ¿Cuáles son sus accuracies? ¿Cuáles son los atributos de las reglas que más especializa? ¿Cuántas hojas tienen ambos arboles? Razona a que se deben dichas diferencias.
- 8) Obten los árboles de ambos experimentos. Botón derecho en cada experimento dentro del cuadro Result list. Después pulsar "Visualiza tree". Adjunta dichas imágenes al documento Word.



## 5. Práctica 3 (Regresión)

### Objetivo

- 1) Abrir archivo “icecream.csv”
- 2) Pulsa en la pestaña “Classify”
- 3) Pulsa el botón “Choose”
- 4) Selecciona LinearRegression en la carpeta functions.
- 5) Pulsa “Start”. Soluciona el problema diviendo training y test en 80/20. Para ello usa la caja Test options. Copia los resultados en el documento Word.
- 6) Interpreta el modelo obtenido.
- 7) Abrir archivo “house.arff”
- 8) Repite los pasos anteriores.
- 9) Realiza el mismo experimento, pero con la opción Use training set. Copia los resultados en el documento Word.
- 10) Mira los “Summary” de ambos experimentos e interpreta las métricas.
- 11) Interpreta el mejor modelo obtenido.

## 6. Práctica 4 (Clustering)

### Objetivo (2 puntos)

- 1) Abrir archivo “iris.arff”
- 2) Pulsa en la pestaña “Clustering”
- 3) Pulsa el botón “Choose”
- 4) Selecciona el algoritmo SimpleKMeans
- 5) Pulsa en el cuadro que pone SimpleKMeans. Realiza los siguientes pasos (para ejecutar el algoritmo pulsa Start) y copia los resultados en un documento Word.
  - a. Prueba con distancia Euclidea con 3 clusters y 4 clusters
  - b. Prueba con distancia de Manhattan con 3 clusters y 4 clusters
- 6) Que puedes interpretar de los diferentes clusters generados.
- 7) Visualízalos pulsando con el botón derecho en cada experimento y con la opción “Visualize cluster assignments”. ¿Qué ocurre? Adjunta dichas gráficas al documento Word.

## 7. Forma de entrega del laboratorio:

La entrega consistirá en un fichero comprimido RAR con nombre **LAB05-GRUPOxx.ZIP** subido a la tarea **LAB5** que contenga únicamente

1. La memoria del laboratorio en .pdf.

**Las entregas que no se ajusten exactamente a esta norma NO SERÁN EVALUADAS.**

## 8. Rúbrica de la Práctica:

### 1. IMPLEMENTACIÓN: Multiplica la nota del trabajo por 0/1

Siendo una práctica de Data Mining, todos los aspectos de programación se dan por supuesto. La implementación será:

- Original: Código fuente no copiado de internet. Grupos con igual código fuente serán suspendidos
- Correcta: El programa funciona y ejecuta correctamente todo lo planteado en los apartados de cada práctica.
- Comentada: Inclusión (**obligatoria**) de comentarios.
- En las gráficas que se realicen proporciona todos los datos que creas necesarios.



## **2. MEMORIA DEL LABORATORIO**

Obligatorio redacción clara y correcta ortográfica/gramaticalmente. Cada paso que se haga tiene que estar justificado.