



### 1. Objetivo del laboratorio

Desarrollar de forma autónoma **un Notebook** que permitan usar algoritmos para obtener reglas de asociación y patrones secuenciales.

### 2. Elementos a utilizar:

- Lenguaje Python
- Librería numérica *NumPy*, *pandas*, *scikit-learn*, *mlxtend*, *apriori*, *gsppy* y gráfica *Matplotlib*
- Entorno Anaconda o Google Colab
- Editor Jupyter o Google Colab

### 3. Práctica 1 (reglas de Asociación)

#### Objetivo (6 puntos)

Usa la librería *mlxtend* que nos permita solucionar todos los problemas relacionados con las reglas de asociación. Para ello tendremos que usar el algoritmo Apriori y una serie de métodos para obtener la distinta información que este genera. Usaremos la librería con el Dataset **Supermercado.csv** que se ha proporcionado. Prueba al menos tres configuraciones de soporte y da respuesta a las siguientes preguntas. Estas deben estar sustentadas con datos obtenidos durante la práctica y razonadas debidamente.

Nota: También se pueden usar otras librerías como Orange3-Associate o Efficient-A priori.

Con un soporte mínimo bajo explica que itemsets frecuentes se generan. Interpreta al menos dos de las reglas que se obtienen con dicho soporte. (1 punto)

¿Qué tipo reglas desaparecen según aumentamos el valor de los soportes usados? ¿Por qué? (1 punto)

Establece ahora el soporte mínimo a 0,3 y juega con la confianza. Primero establécela en un valor bajo y luego en uno alto. Explica que cambios ves en las reglas y porque crees que ocurren dichos cambios. Interpreta la regla que veas más significativa. (1 punto)

Obtén una configuración de soporte y confianza que funcione mejor según tu criterio. Razona y explica cómo has tomado la decisión. (1 punto)

Dadas la que consideres mejor configuración de soporte y confianza. ¿Existen reglas o patrones que se repitan? ¿Podemos generalizar de alguna manera como se comportan los clientes del supermercado? (1 punto)

Con los resultados del apartado anterior, interpreta que ocurre con el lift para al menos dos instancias. (1 punto)

Con la configuración anterior. Primero elimina un número de instancias del producto que crea que es más crítico para generar reglas (deja al menos una instancia). Segundo, realiza un proceso parecido, pero eliminando transacciones. ¿Qué ves que ocurre? Razona porque sucede esto. (1 punto)



#### 4. Práctica 2 (Patrones Secuenciales)

##### Objetivo (4 puntos)

El conjunto de datos “**Navegación\_Web.csv**” contiene clientes y las acciones que hace en una Web, teniendo estas asociadas un timestamp.

Habrás que procesar el fichero para crear una estructura de datos de los distintos clientes con las secuencias de las acciones realizadas en distintos instantes (1 punto).

Una vez hecho esto y analizando esta información se pueden extraer conclusiones sobre en qué orden compran los clientes los productos y así tomar decisiones de negocio. Este estudio se puede llevar a cabo aplicando el algoritmo Generalized Sequential Patterns utilizando la implementación de éste disponible en la librería *gspppy*.

Prueba al menos tres configuraciones de soporte diferentes. ¿Qué diferencias hay en los resultados y a que se debe? (1 punto)

Para una de ellas, interpreta algunos de los patrones secuenciales que te resulten curiosos. (1 punto)

¿Qué transacción es clave eliminar para cambiar los patrones obtenidos en el punto anterior? ¿Y cuál de las secuencias? (1 punto)

#### 5. Forma de entrega del laboratorio:

La entrega consistirá en un fichero comprimido RAR con nombre **LAB02-GRUPOxx.RAR** subido a la tarea **LAB1** que **contenga únicamente**

1. **Por cada práctica** un notebook de Jupyter (archivos con extensión **.ipynb**).
2. La **memoria del laboratorio** que se irá construyendo en el Notebook de manera que se explique todo lo que se hace.

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

#### 6. 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.