



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*, *apyori* y gráfica *Matplotlib*
- Entorno Anaconda o Google Colab
- Editor Jupyter o Google Colab

3. Práctica 1 (reglas de Asociación)

Objetivo (5 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.

I) (5 puntos) Usaremos la librería con el Dataset **Market_Basket_Optimisation.csv** que se ha proporcionado.

Prueba al menos tres configuraciones de soporte y frecuencia para cada tipo de puntuación. (1 punto)

¿Qué diferencias hay entre usar soporte y frecuencia? Respalda la respuesta con datos (1 punto)

¿Qué tipo reglas desaparecen según la configuración usada y el tipo de puntuación? ¿Por qué? (1 punto)

Para una de las configuraciones, interpreta algunas de las reglas que te hayan resultado curiosas obtenidas usando la clase `association_rules` y un par de configuraciones cambiando la variable `metric` y `min_threshold`. (1 punto)

Dadas las mejores configuraciones. ¿Existen reglas o patrones que se repitan? ¿Podemos generalizar de alguna manera como se comportan los clientes del supermercado? (1 punto)

Para obtener los datos pedidos anteriormente, se hará uso de la librería de la siguiente manera:

- 1) Empezaremos obteniendo los itemsets frecuentes para $k=1$. En este punto necesitaremos obtener el soporte de los itemset. Por lo tanto, se tendrá que usar un método que dado un itemset devuelva su soporte.
- 2) A partir de $k=2$.
 - a. Mostrar los itemset frecuentes candidatos y su soporte.
- 3) Repetir el proceso 2 hasta que no se generen nuevos itemsets frecuentes.
- 4) Mostrar todas las posibles reglas con la confianza de cada una de ellas.
- 5) Listar todas las reglas que sean de alta confianza.
- 6) Usar los siguientes métodos. Dado un antecesor, devolver todas las reglas que contengan a dicho antecesor. Dado un umbral mínimo devolver todas las reglas que cumplan con dicha confianza.
- 7) Utiliza al menos dos representaciones gráficas para representar las reglas obtenidas e interpretar los datos.



4. Práctica 2 (Patrones Secuenciales)

Objetivo (5 puntos)

El conjunto de datos “**Online Retail.xlsx**” contiene productos (definidos por la columna StockCode y Description) comprados por diferentes clientes (definidos por la columna CustomerID) en diferentes momentos (definidos por el campo InvoiceDate. Habrá que procesar el fichero para crear una lista de los distintos clientes con las secuencias de las transacciones de las compras realizadas en distintos instantes (2 puntos). Una vez hecho esto y analizando esta información se pueden extraer conclusiones sobre en que 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 *gsppy*.

Prueba al menos dos configuraciones de soporte diferentes. (1 punto)

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

Saca conclusiones de negocio en virtud de algunas de las reglas elegidas por vosotros que mayor confianza tengan(1 punto)

Librería *gsppy*:

Podéis encontrar su documentación en el siguiente enlace: <https://pypi.org/project/gsppy/>

5. Forma de entrega del laboratorio:

La entrega consistirá en un enlace al github donde se encuentra el jupyter notebook.

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.