



CURSO: MINERÍA DE DATOS
MAESTRÍA EN EXPLOTACIÓN DE DATOS Y DESCUBRIMIENTO DE CONOCIMIENTO

LABORATORIO IX: Reglas de Asociación

INTRODUCCIÓN

En este trabajo se abordará el aprendizaje de reglas de asociación, las cuales permiten encontrar patrones de comportamiento dentro de un conjunto de datos. En particular, se trabajará con el algoritmo Apriori y los conceptos de soporte y confianza, relacionados a este algoritmo.

Para ello, se utilizará el software R¹ con el objetivo de resolver problemas de la disciplina, los cuales son una combinación ejercicios clásicos de minería de datos complementados con ejercicios propuestos por el equipo docente.

CONSIGNAS

1. **Exploración de datos.** Incorpore el dataset *Groceries* y a partir de las herramientas de exploración de transacciones abordadas en clase, caracterice el conjunto de datos en función de la siguiente información:
 - a. Cantidad de transacciones,
 - b. Cantidad de ítems diferentes,
 - c. Dominio del dataset
 - d. Ítems más frecuentes,
 - e. Distribución de la cantidad de elementos observada en los itemsets.
2. **Generación de reglas.** A partir del dataset anterior, genere reglas en función de diferentes parámetros de soporte y confianza.
 - a. ¿Qué sucede conforme aumentan los valores para estos dos parámetros?
 - b. Deje fijos los parámetros de *soporte* y *confianza* en *0.01* y *0.1* respectivamente y explore las reglas generadas:
 - i. ¿Cuántas reglas se generan con esos parámetros?
 - ii. Explore las reglas, elija 5 reglas que resulten interesantes en función de las métricas estudiadas y explique el conocimiento que aportan.

¹ Para trabajar con el algoritmo **Apriori** en **R**, deberá instalarse previamente la librería **arules**.



CURSO: MINERÍA DE DATOS

MAESTRÍA EN EXPLOTACIÓN DE DATOS Y DESCUBRIMIENTO DE CONOCIMIENTO

- iii. ¿Cuál es el valor observado para *lift*, *confianza* y *soporte*? Interprete estos valores.
 - iv. Ordene las reglas generadas por *lift*:
 - 1. ¿Qué sucede para las reglas con *lift* demasiado altos?
 - 2. ¿Qué indica un *lift* cercano a 1?
 - v. Escoja un ítem de su interés y filtre las reglas en las cuales aparece en el antecedente. Interprete el conocimiento obtenido.
 - vi. Con el mismo ítem, filtre las reglas en las cuales aparece en el consecuente. Interprete el conocimiento obtenido.
3. **Generación de reglas a partir de términos.** En función del dataset con las letras de las canciones en español con el que se trabajó en el LAB08:
- a. Genere transacciones con los términos que aparece en cada letra, pensando a las canciones como transacciones y los términos como ítems.
 - b. A partir de las transacciones del punto anterior, genere reglas con el algoritmo *apriori()* y explore las mismas.

Referencias sugeridas:

Data Mining: Concepts and Techniques. Jiawei Han & Micheline Kamber. Morgan Kaufmann. Third Edition. 2011. Chapter 5.

Arules (R): <https://cran.r-project.org/web/packages/arules/arules.pdf>