



Bases de Datos Masivas (11088)
Departamento de Ciencias Básicas

TRABAJO PRÁCTICO V: Minería de datos

PARTE 03: 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 **R**¹ 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. **Soporte & Confianza.** Calcule el soporte y la confianza (cuando corresponda) de los ítemsets del siguiente fragmento del dataset:

	A	B	C
1	0	1	0
2	1	0	1
3	0	0	1
4	1	0	0
5	1	1	1
6	0	1	1
7	1	1	0
8	1	0	1
9	1	0	0
10	1	1	1

En función de la tabla anterior, realice las siguientes operaciones:

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



Bases de Datos Masivas (11088)
Departamento de Ciencias Básicas

- a. Calcule el soporte para todos los ítemsets del dataset.
 - b. Arme todas las reglas resultantes considerando que se solicita un **soporte mínimo de 0,3**.
 - c. ¿Cuál es el soporte de A? ¿Cómo es el soporte de AB, AC y ABC con respecto al de A? ¿Por qué?
 - d. ¿Cuáles son las reglas de asociación resultantes si establecemos una **confianza mínima de 0,7**?
2. **Apriori**². Incorpore en una herramienta de *data mining* el dataset sobre la cesta de compras³ y responda:
- a. ¿Qué parámetros puedo modificar previo a ejecutar el algoritmo sobre el dataset? ¿Qué permite cada uno?
 - b. ¿Es posible ejecutar el algoritmo *apriori* con el dataset tal como se encuentra? Realice las operaciones necesarias para permitirlo.
 - c. Ejecute el algoritmo *Apriori* sobre los datos y detalle cuáles son las mejores reglas encontradas. ¿Cómo determina cuáles son las mejores?
 - d. ¿Qué nota al ejecutar el algoritmo con el dataset actual? ¿Cuál es la complejidad computacional del mismo? ¿Cómo puede resolverse?
 - e. Si tuviera que analizar los resultados desde el punto de vista de un analista especializado en marketing, ¿Cuáles son las asociaciones encontradas que le parecen más interesantes? ¿Qué políticas podría implementar a partir de estas?
 - f. Explore alguna visualización⁴ para el análisis de reglas generadas, explique brevemente.
 - g. Utilizando el mismo punto de vista, ¿Cuáles son los ítems marcan la presencia de cerveza? ¿Encuentra una relación lógica en estas asociaciones?

² Para trabajar con el algoritmo **Apriori** en **R**, deberá instalarse previamente la librería **arules**.

³ Dataset de tipo *transactions* denominado **Groceries** de la librería **arules** de **R**.

⁴ Se sugiere trabajar con la librería **arulesViz** de **R**.



Bases de Datos Masivas (11088)
Departamento de Ciencias Básicas

- h. ¿Qué parámetros ajustaría a efectos de modificar la cantidad de cantidad de reglas de asociación generadas? ¿Qué efecto generan esos parámetros? Ejemplifique en función del dataset actual.
 - i. Documente todas las actividades desarrolladas y exprese sus conclusiones en cada caso.
3. Incorpore el dataset del Banco de Portugal utilizado en el TP de árboles de decisión y realice las siguientes operaciones:
- a. Aplique las transformaciones necesarias a efectos de poder correr el algoritmo *apriori* sobre el dataset.
 - b. Ejecute el algoritmo *apriori* y explique los resultados más importantes.
 - c. ¿Cuáles son las asociaciones más importantes para determinar si el cliente aceptó o no el producto bancario?
 - d. Compare los resultados obtenidos en este punto con respecto a los resultados obtenidos a través de la aplicación de árboles de decisión.
 - e. Documente todas las actividades desarrolladas y exprese sus conclusiones en cada caso.

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>