



UNIVERSIDAD DE CÓRDOBA
ESCUELA POLITÉCNICA SUPERIOR DE CÓRDOBA

INGENIERÍA INFORMÁTICA
ESPECIALIDAD: COMPUTACIÓN
CUARTO CURSO. PRIMER CUATRIMESTRE

INTRODUCCIÓN A LA MINERÍA DE DATOS

Práctica 3: Reglas de asociación

Valentín Gabriel Avram Aenachioei

03524931C

p92avavv@uco.es

Curso académico 2022-2023
Córdoba, 9 de junio de 2023

Índice

Índice de tablas	II
1. Preguntas práctica 3	1
1.1. Utilizando el conjunto de datos “store_data.csv” ejecute el programa “assoc.py” para comprobar su funcionamiento. Intente interpretar las reglas obtenidas e indicar cuáles de ellas son importantes.	1
1.2. Usando los conjuntos “titanic.csv” y “bank-data-final.arff” ejecute de nuevo el programa de generación de reglas, ordenando los resultados según su valor de lift. Interprete las reglas que se obtienen indicando su evaluación objetiva de interés	1
1.3. Seleccione las 5 mejores reglas usando las medida de confianza y lift. Compare las reglas obtenidas y comente qué información puede obtener de algunas de las reglas	2

Índice de tablas

1.	Mejores reglas según Confianza	2
2.	Mejores reglas según Lift	2

En este documento se recogerán los resultados obtenidos realizando las pruebas indicadas en el guión de la práctica 3, así como el respectivo análisis de esos resultados. No se hará mención a las implementaciones en código necesarias para resolver cada problema.

1. Preguntas práctica 3

1.1. Utilizando el conjunto de datos “store_data.csv” ejecute el programa “assoc.py” para comprobar su funcionamiento. Intente interpretar las reglas obtenidas e indicar cuáles de ellas son importantes.

Podemos considerar como importantes aquellas reglas que tengan un valor de soporte alto y a la vez, un valor de confianza alto, significando esto que aunque la regla tenga poca frecuencia de aparición, cuando aparece es muy probable que se cumpla.

Podemos destacar como regla mas importante *Spaghetti* \rightarrow *Ground Beef*, con un valor de soporte de *0.006* y un valor de confianza del 52 %. Otra regla que podría ser considerada importante sería *Mineral Water* \rightarrow *Ground Beef*, con un valor de soporte de *0.006* y un valor de confianza del 52 %.

1.2. Usando los conjuntos “titanic.csv” y “bank-data-final.arff ” ejecute de nuevo el programa de generación de reglas, ordenando los resultados según su valor de lift. Interprete las reglas que se obtienen indicando su evaluación objetiva de interés

Primeramente, para el dataset *Titanic*, se obtienen numerosas reglas de asociación. Según el valor de lift, se pueden destacar la regla *S.O.C14879* \rightarrow 73,5, con un valor de *Lift* = 178,2, *Soporte* = 0,0056 y *Confianza* = 100 %. Otro regla que se puede considerar como importante sería la regla *15.5* \rightarrow 3, con un valor de *Confianza* = 100 %, y un *Soporte* = 0,0089, aunque solo se tenga un valor de *Lift* = 12,3749. Para este dataset se han encontrado numerosas reglas de asociaciones, muchas con altos valores de confianza y split, pero solo se han mostrado dos de las varias que hay.

Para el dataset *bank-data* se han encontrado de nuevo numerosas reglas de asociación. Entre ellas destacan la regla $52_max \rightarrow 1$, con un valor de *Confianza*=100 %, *Soporte*=0.02 y *Lift*=3.1413. Al igual que esta, se encuentran numerosas reglas de asociación con una *Confianza*=100 %, que pueden ser consideradas como importantes. Otra regla a destacar puede ser $3 \rightarrow 43759_max$, con un valor de *Lift*=17.0454, *Soporte*=0.0083 y *Confianza*=0.625.

1.3. Seleccione las 5 mejores reglas usando las medida de confianza y lift. Compare las reglas obtenidas y comente qué información puede obtener de algunas de las reglas

Para comparar las mejores reglas según su medida de confianza, estas se recogen ordenadas en la tabla 1. Para comparar las mejores reglas por su medida de Lift, estas están recogidas en la tabla 2

Regla de Asociación	Confianza
$vgood \rightarrow 2$	100 %
$vgood \rightarrow high$	100 %
$3 \rightarrow high$	66.66 %
$3 \rightarrow more$	66.66 %
$small \rightarrow high$	66.66 %

Tabla 1: Mejores reglas según Confianza

Regla de Asociación	Lift
$vgood \rightarrow high$	6.0
$more \rightarrow high$	5.5384
$high \rightarrow 4$	4.5
$high \rightarrow small$	4.0
$3 \rightarrow high$	4.0

Tabla 2: Mejores reglas según Lift