

Control 2 - IIC2433

Jueves 09 de Septiembre de 2020
Contenidos: Reglas de asociación

Indicaciones

- El control es **individual**. La copia será castigada con nota 1.1 al curso, además de las sanciones disciplinarias correspondientes.
 - El control tiene cuatro preguntas, para obtener el 7 deben contestarlas todas de forma correcta. Cada una tiene igual ponderación.
 - La entrega de este control se realizará a través de un buzón en Canvas, que permanecerá abierto hasta las 23:59 del día Jueves 20 de Agosto. **No se permitirán entregas atrasadas.**
 - El control debe ser entregado en formato **PDF**. En caso de entregar con otro formato, este no se corregirá.
 - Las preguntas o sub preguntas que requieran fundamentación, se espera que esta no exceda las 5 líneas.
-

Preguntas Reglas de Asociación (6.0 pts)

Para las siguientes 3 preguntas considere el siguiente dataset de transacciones.

TID	Items
1	A, B, E
2	A, B, D, E
3	B, C, D, E
4	B, D, E
5	A, B, D
6	B, E
7	A, E

1. (1.7 pts.) Utilizando el dataset anterior, ejecute el algoritmo Apriori con un soporte mínimo de 3/7. Finalmente muestre los *itemsets* frecuentes obtenidos. Debe ir registrando cada paso e iteración del algoritmo con el conjunto de transacciones.
2. (1.7 pts.) Utilizando el dataset anterior, ejecute el algoritmo Fp-Growth con un soporte mínimo de 3/7. Finalmente muestre los *itemsets* frecuentes obtenidos. Recuerde que debe construir el FP-Tree y luego minarlo para obtener los itemsets. Debe registrar cada paso del algoritmo en su respuesta.

3. (1.0 pts.) Usando los resultados de las preguntas anteriores, calcule el soporte, la confianza y el *lift* de las siguientes reglas de asociación:

- $(B, E) \rightarrow (D)$
- $(B, D) \rightarrow (E)$
- $(D, E) \rightarrow (B)$

¿Para cuál(es) de las reglas la confianza es mayor? ¿Qué significa que la confianza sea mayor para dicha(s) regla(s)?

4. (1.6 pts.) Suponga que tiene un conjunto de datos de usuarios y ratings dados a libros en una biblioteca digital, con ratings del 1 al 10. Las filas de este conjunto de datos son de la forma (id_usuario, id_libro, rating). Suponga que tiene otra tabla con los géneros de cada libro, donde cada fila es de la forma (id_libro, genero). ¿Cómo podría utilizar algún algoritmo de itemset frecuentes para generar reglas de asociación entre los géneros de los libros? Describa el paso a paso de como implementaría su solución.