



Evaluación 3b – Reglas de Asociación

Fecha de Entrega: domingo 20 de agosto

El objetivo de esta evaluación es utilizar y analizar reglas de asociación. La base de datos contiene registros de 15.000 listas de reproducción musical (i.e. “transacciones”) de personas de distintos países. En cada lista de reproducción pueden estar más de 1.000 artistas de distintos géneros musicales.

Usted deberá generar reglas de asociación a partir de los datos, para realizar recomendaciones.

Las tareas por desarrollar consisten en:

1. Realizar un análisis previo de los datos, que le permita obtener información útil para definir los parámetros del algoritmo Apriori para obtener reglas de asociación.
2. Aplicar el algoritmo Apriori para generar reglas de asociación a partir de los datos, definiendo y justificando claramente los parámetros utilizados. Describa a grandes rasgos los resultados obtenidos.
3. Realizar recomendaciones de artistas a diez amigos distintos, donde a cada uno le gusta uno de los siguientes artistas (un artista por amigo): Arctic Monkeys, Beyonce, Coldplay, Jimi Hendrix, Metallica, Muse, Slayer, The Killers, The Shins y Wilco. Justifique sus recomendaciones.

Debe generar un breve reporte y entregarlo al correo: mineria.datos.PUC@gmail.com. El asunto del correo y el nombre del archivo deben comenzar con [Evaluación 3b] seguido con los apellidos de los estudiantes. Ejemplo: “[Evaluación 3b] Gutiérrez y Soto”.

El objetivo de la tarea es demostrar que entiende lo que está haciendo R al obtener reglas de asociación. Por favor presente sólo la información relevante, sin llenar múltiples páginas con gráficos, códigos y estadísticas. Extensión recomendada: 8 planas de contenido (i.e. sin contar portada, índices o anexos, los cuales no son obligatorios).



Pontificia Universidad Católica de Chile
Educación Profesional - Escuela de Ingeniería
Diplomado en Big Data y Ciencias de Datos
Minería de Datos
Relator: Sebastián Raveau

Su entrega será evaluada de acuerdo con los siguientes aspectos:

Claridad [1,5 puntos]

Se espera que el reporte sea claro y bien redactado, con las ideas debidamente desarrolladas. Las tablas y figuras deben estar debidamente presentadas y explicadas. El lenguaje utilizado debe ser apropiado para un reporte técnico.

Análisis Previo [1,5 puntos]

El análisis previo debe ser entendido por cualquier lector, independiente de su grado de familiaridad con los datos o el algoritmo a utilizar. La información proporcionada debe ser relevante para la aplicación del algoritmo.

Aplicación de Algoritmo Apriori [1,5 puntos]

Los criterios utilizados al aplicar el algoritmo (e.g. soporte, confianza, lift) deben ser explícitos y justificados adecuadamente, en función del análisis previo. Se presentan y describen los resultados más relevantes de la aplicación del algoritmo (e.g. cantidad de reglas, artistas más recurrentes, tamaño de itemsets, análisis crítico de los resultados, entre otros).

Recomendaciones [1,5 puntos]

En base a los resultados obtenidos se realizan y justifican recomendaciones para los diez amigos. Las recomendaciones se explican adecuadamente. Se explicita y justifica cualquier potencial modificación a los criterios utilizados al aplicar el algoritmo.