# Práctica 2: Similarity

Contenido

[Práctica 2: Similarity 1](#_Toc459644712)

[Contenido 1](#_Toc459644713)

[Objetivos 1](#_Toc459644714)

[Propósito del curso al que apoya la práctica 1](#_Toc459644715)

[Actividades 1](#_Toc459644716)

[Ejercicios 1](#_Toc459644717)

[1 Filtrado Colaborativo para minar preferencias (basado en items) 1](#_Toc459644718)

[2 Filtrado Colaborativo para minar preferencias (basado en usuarios) 2](#_Toc459644719)

[Entregables 3](#_Toc459644720)

[Ubicación física de los datos 3](#_Toc459644721)

## Objetivos

1. Entender los conceptos de similitud.

## Propósito del curso al que apoya la práctica

Entender diversas técnicas de minería de datos, conocer sus aplicaciones y desarrollar su implementación utilizando R y Cassandra.

## Actividades

* Generar dos tipos de recomendaciones basadas en similitud: por ítem y por usuario.

## Ejercicios

## Filtrado Colaborativo para minar preferencias (basado en items)

1.- Descarga los datos de aquí:

<http://www.salemmarafi.com/wp-content/uploads/2014/04/lastfm-matrix-germany.csv>

2.- usa la función read.csv para cargar los datos.

3.- Visualiza algunos datos del dataset.

4.- Elimina la columna de usuarios

5.- Crea una función para calcular la similitud Coseno.

6.- Crea una matriz para guardar la similitud Coseno entre dos pares de canciones

7.- Calcula todas las distancias, despliega los primeros resultados.

8.- Crea una matriz para guardar los artistas más semejantes entre sí

9.- Encuentra a los 10 vecinos más cercanos y puebla la matriz creada anteriormente. Crea un ciclo para procesar todos los artistas, ordena la matriz de similitud por artista, obtén los primeros 11 valores y aplica la transpuesta de la matriz.

10.- Qué artistas son los más similares a los beatles?

## 2 Filtrado Colaborativo para minar preferencias (basado en usuarios)

1.- Usaremos los mismos datos que en el primer apartado.

2.- Crea una matriz de NAs con las dimensiones del dataset original

3.- Crea dos ciclos anidados: el exterior por usuario y el interior por artista.

4.- Dentro del ciclo interior obtén los valores del usuario y el producto.

5.- Crea una validación: si el artista ya es del gusto del usuario, coloca una cadena vacía (para evitar recomendar artistas que ya son del gusto del usuario)

6.- Obtén los 10 valores más similares (usando la matriz previamente creada) y crea dos variables, una con los nombres de los rows y otra con los valores de similitud. Remueve la primera columna.

7.- Obtén la historia de uso del usuario para los 10 valores más similares. Primero obtén un data.frame con las columnas de user y los 10 artistas más frecuentes, luego filtra esta matriz usando el usuario actual. Por último, obtén el vector con los valores que indican si el usuario ha escuchado al artista.

8.- Calcula el SCORE para ese producto y ese usuario. Nuestro SCORE estará basado en similitud e historia. Esta es la ecuación: SUM(H\*S)/SUM(S). Crea una función auxiliar y llámala dentro del ciclo interno.

9.- Corre el ciclo anidado, tardará un tiempo.

10.- Examina el resultado, como interpretas el primer registro?

11.- Cambia el formato de las recomendaciones para utilizar los nombres de los artistas a recomendar en orden.

## Entregables

* Un documento en formato Doc, Docx o PDF, que integre las respuestas a las preguntas expuestas en los ejercicios y a las preguntas de análisis posteriores a los ejercicios

### Ubicación física de los datos

Referido en el documento.