

Exploración de géneros y recomendación musical basada en Aprendizaje No Supervisado a partir de canciones de Spotify

Curso *Análisis no supervisado*--MIAD

Grupo 19

Carlos Galindo Aguilera, Daniel Enrique Granados Iglesias,
Sebastian Barrera Bernal, Mayra Alejandra Neisa Valero



Problema y contexto



En un entorno donde el catálogo de canciones disponible supera los millones, ofrecer recomendaciones musicales precisas y relevantes se ha convertido en un desafío crítico.



El problema a resolver es cómo identificar patrones ocultos y tendencias emergentes en un conjunto masivo de datos musicales, con el fin de mejorar la precisión y relevancia de las recomendaciones.



Se busca hacer uso de distintas técnicas de Análisis no supervisado para explorar este problema. En este caso, se tiene un listado con información solo de las canciones, no se tiene información de los usuarios por lo que se debe recurrir a métodos como el filtrado basado en contenidos para hacer una aproximación a las recomendaciones.

Algoritmo y datos (I)

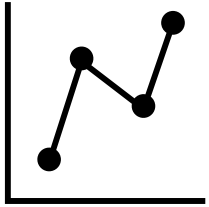
El dataset utilizado para este proyecto, corresponde a *30.000 Spotify Songs* un dataset de Kaggle* creado por Joakim Arvidsson. En este dataset se encuentran casi 30.000 canciones recuperadas desde la API de Spotify

- Variables categóricas: nombre de canción, del álbum, de listados donde se encuentran las canciones.
- Variables numéricas: índices sobre características de las canciones como bailabilidad, energía, acústica, entre otros. Además, se cuenta con un índice que representa la popularidad de la canción.

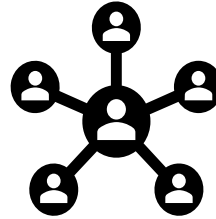
*https://www.kaggle.com/datasets/joebeachcapital/30000-spotify-songs/data?select=spotify_songs.csv

Algoritmo y datos (II)

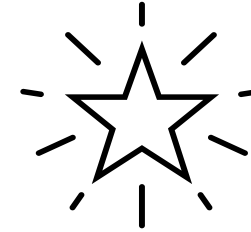
- Análisis inicial propuesto:



PCA: para la
reducción de
dimensión



Clusterización: para
encontrar patrones
subyacentes



Sistemas de
recomendación

Resultados y discusión (I)

Sistema de recomendación base: se toma como referencia la canción *Call You Mine - Keanu Silva Remix* de *The Chainsmokers* (genero pop) y se usaran las características de sonido para la clasificación.

- 1) Recomendación basada en las variables numéricas, sin considerar otras características como el género o el artista

	track_id	track_name	track_artist	playlist_genre
31439	6XtOpnP7gitKdibHXylwsC	I Don't Like You - Nicky Romero Remix	Eva Simons	edm
27757	79Uq59YH1T7EJ16uWxK8Oc	Bathing Suit Body	Badpojken	edm
2129	6YZdkObH88npeKrrkb8Ggf	DUELE EL CORAZON	Enrique Iglesias	pop
27333	5NwWGVjLvbxJZrJLhcRqm1	Music Has the Power - Radio Edit	!deladap	edm
3014	48QmG1dfvMuYLxMPt7KSRA	No Lie	Sean Paul	pop

Resultados y discusión (II)

Sistemas de recomendación base

2) Recomendación incluyendo como variable de filtro el género de la canción

	track_id	track_name	track_artist	playlist_genre
2129	6YZdkObH88npeKrrkb8Ggf	DUELE EL CORAZON	Enrique Iglesias	pop
3014	48QmG1dfvMuYLxMPt7KSRA	No Lie	Sean Paul	pop
3825	5m5TRf7ggoguXtcXQGprN4	Written in the Stars (feat. Eric Turner)	Tinie Tempah	pop
2497	3cFsmM51N5zMAuu1nNY4Tw	No Lie	Sean Paul	pop
3755	1jLsirPDkUS2g4gnkYua58	Ignite	Alan Walker	pop

Resultados y discusión (III)

Sistema de recomendación base:

3) Recomendación incluyendo como variable de filtro el artista

	track_id	track_name	track_artist	playlist_genre
29718	4B3NahKokb7ILkvaLFZFzn	Sick Boy - Owen Norton Remix	The Chainsmokers	edm
10915	33ecyF2OHCK4HFUSp1hltE	Closer - T-Mass Remix	The Chainsmokers	rap
30142	4Sdp10fc5XgHxL1VaiMVYN	Inside Out - DubVision Remix	The Chainsmokers	edm
24	4sqABRRGU7CzcHXCyxUzFw	Do You Mean (feat. Ty Dolla \$ign & bülow) - My...	The Chainsmokers	pop
30110	1JI70I1IE5IF2tgJm5TnMD	Don't Let Me Down - Hardwell & Sephyx Remix	The Chainsmokers	edm

Resultados y discusión (IV)

Sistema de recomendación incluyendo información del usuario: se plantea interactuar con el usuario, preguntando su edad y un género de preferencia.

Prueba con usuario de 44 años y genero seleccionado rap

Edad: 44

Género:

Recomendar Canci...

	track_name	track_artist	album_year	playlist_genre
9191	Straight Outta Compton - Extended Mix / Edit	N.W.A.	1996	rap
7794	Let Me Clear My Throat - Old School Reunion Re...	DJ Kool	1996	rap
7068	Down For My Niggaz	C-Murder	2005	rap
7374	Let Me Clear My Throat - Old School Reunion Re...	DJ Kool	1996	rap
9645	Natural Born Killaz	Dr. Dre	2001	rap

Conclusiones

- **Capacidad de personalización:** El modelo base, que toma como referencia únicamente características del sonido de las canciones podría tomarse como más objetivo en sus resultados, en comparación con el modelo de interacción con el usuario que tiene un enfoque más personal.
- **Bondad de similitud del coseno:** La similitud de coseno es una técnica sólida y eficaz para sistemas de recomendación basados en contenido, especialmente cuando se trabaja con características numéricas como las contenidas en el dataset de canciones. La simplicidad de la técnica, su independencia de la magnitud y su escalabilidad la convierten en una excelente opción para comparar canciones en términos de similitud musical.
- **Ventaja del aprendizaje supervisado:** estos modelos nos permiten explorar desde diferentes puntos de vista y características la información, sin estar sujetos a una variable respuesta.