

Universidad de Los Andes
Curso *Análisis no supervisado*- Entrega 1 proyecto

Grupo 19

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

Septiembre – 2024

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

1. Resumen

En la era digital, la música se ha convertido en un elemento omnipresente, con plataformas como Spotify que ofrecen acceso a millones de canciones. Sin embargo, la vastedad de opciones puede dificultar la personalización de la experiencia del usuario. Este proyecto propone un enfoque basado en Aprendizaje No Supervisado para analizar una base de datos de 30.000 canciones de Spotify, utilizando características como energía, locuacidad, acústica e instrumentalidad. Mediante el uso de Análisis de Componentes Principales (PCA), se reducirá la dimensionalidad del conjunto de datos, permitiendo la identificación de patrones subyacentes. Posteriormente, se aplicarán técnicas de clustering para agrupar las canciones según géneros emergentes y patrones de popularidad. Finalmente, se diseñará un sistema de recomendación a partir de filtrado colaborativo que aprovechará los resultados obtenidos del análisis no supervisado para ofrecer sugerencias musicales personalizadas. Este proyecto busca no solo descubrir nuevas tendencias musicales, sino también mejorar la experiencia de usuario mediante recomendaciones más precisas y adaptadas a los gustos individuales.

2. Introducción

En el vasto universo de la música digital, plataformas como Spotify, Amazon Music, YouTube Music y Apple Music compiten por ofrecer a sus usuarios la experiencia más personalizada posible. Sin embargo, 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. La diversidad de gustos musicales y la constante aparición de nuevos géneros hacen que las recomendaciones tradicionales basadas en la popularidad o en las preferencias pasadas del usuario resulten insuficientes para satisfacer la demanda de personalización.

El cliente potencial para este proyecto son las plataformas de streaming musical, específicamente los equipos de desarrollo y analítica de datos de empresas como Spotify. Estos equipos enfrentan la necesidad constante de mejorar sus sistemas de recomendación, no solo para aumentar la satisfacción del usuario, sino también para diferenciarse de la competencia en un mercado saturado. En este contexto, 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.

El uso del aprendizaje no supervisado se presenta como una solución prometedora para este desafío. A diferencia de los métodos supervisados, que requieren datos etiquetados, el aprendizaje no supervisado permite descubrir estructuras subyacentes en los datos sin la necesidad de etiquetas predefinidas. Esto es particularmente relevante en la industria musical, donde los gustos y tendencias cambian rápidamente, y las etiquetas de género tradicionales pueden no capturar la complejidad de la música moderna.

Este proyecto se enmarca dentro del área del aprendizaje no supervisado, con un enfoque en la reducción de dimensionalidad mediante Análisis de Componentes Principales (PCA), la agrupación de datos a través de técnicas de clustering, y el filtrado colaborativo para el desarrollo de recomendaciones personalizadas. La combinación de estas técnicas permitirá no solo reducir la complejidad del análisis, sino también identificar nuevos patrones y géneros emergentes que pueden ser utilizados para mejorar las recomendaciones musicales. En última instancia, se espera que este enfoque contribuya a mejorar la experiencia del usuario en plataformas como Spotify, proporcionando recomendaciones más personalizadas y precisas, adaptadas a los gustos individuales y a las tendencias emergentes.

3. Revisión de la literatura

Diversos estudios, tanto nacionales como internacionales, han explorado algoritmos para el análisis musical y el diseño de sistemas de recomendación personalizada. En esta revisión, se comparan estos enfoques con el propuesto en este proyecto, así como diferencias en la aplicación y los resultados obtenidos, con el fin de situar este trabajo en el contexto de la investigación existente.

En cuanto a la agrupación o segmentación de canciones, Cilibrasi, Vitanyi y de Wolf (2004) proponen un método de clustering jerárquico para música, basado en la compresión de cadenas musicales y la teoría de la complejidad de Kolmogorov, aplicable a varias áreas como la lingüística y la genómica. Utilizan la "distancia de compresión normalizada (NCD)" para agrupar objetos similares y su herramienta CompLearn muestra que este método puede diferenciar géneros musicales y agrupar piezas por compositor. Li, Ogihara et al. (2009) se enfocan en la identificación de artistas utilizando características de diversas fuentes y un algoritmo de clustering bimodal combinado con el clustering de restricciones generalizado. Sus pruebas con un dataset de 570 canciones de 53 álbumes muestran mejoras en la precisión de identificación. Jondya e Iswanto (2017) aplican PCA a segmentos de audio para seleccionar características óptimas y luego usan k-Means para agrupar 101 canciones de 18 provincias indonesias, obteniendo cuatro clusters distintos y realizando un clustering jerárquico aglomerativo para comparación.

En cuanto a los sistemas de recomendación musical (MRS), Song, Dixon y Pearce (2012) destacan la creciente importancia de la gestión y búsqueda de música digital. A pesar de los avances en la recuperación de información musical (MIR), los sistemas de recomendación están aún en desarrollo. Encuentran que el filtrado colaborativo (CF) y el modelo basado en contenido (CBM) son efectivos, pero sugieren que se preste más atención a modelos centrados en el usuario, como los basados en el contexto y la emoción. Examinaron tres componentes clave en sistemas de recomendación y propusieron un modelo basado en la motivación para abordar las limitaciones actuales. Schedl (2019) señala que el aprendizaje profundo (DL) está siendo cada vez más relevante en los sistemas de recomendación musical (MRS), utilizando redes neuronales para extraer factores latentes y patrones secuenciales en señales de audio y metadatos. Su revisión detalla el uso del aprendizaje profundo en diferentes enfoques de recomendación (filtrado colaborativo, basado en contenido, o ambos) y discute los desafíos actuales en este campo.

Tal y como se observa, existen diferencias y similitudes entre la metodología y algoritmos a utilizar en el presente proyecto, y las empleadas en las distintas investigaciones que se han realizado en torno al tema. Sin embargo, se valida el hecho de que los algoritmos de reducción de dimensionalidad, clustering y recomendación, pueden integrarse con el ánimo de lograr los objetivos propuestos.

4. Descripción de los datos

4.1. Fuente de datos

El dataset utilizado para este proyecto, corresponde a '30.000 Spotify Songs' un dataset de Kaggle creado por Joakim Arvidsson que puede ser consultado a través de la siguiente URL: [enlace](#). En este dataset se encuentran casi 30.000 canciones recuperadas desde la API de Spotify, y 23 columnas o variables que las describen.

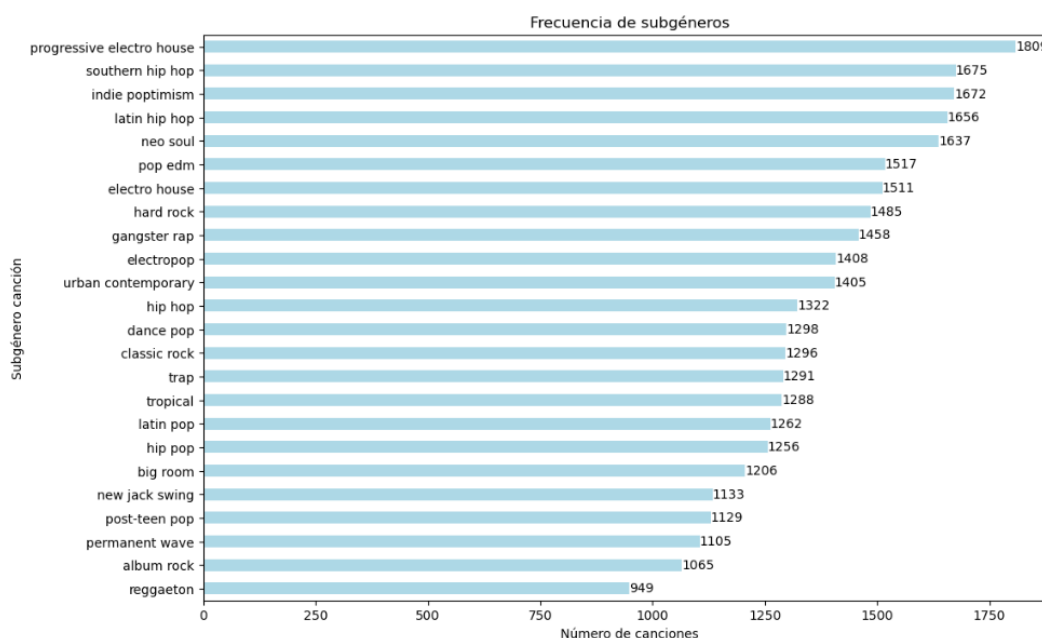
4.2. Descripción de variables

En el dataset podemos encontrar columnas identificadoras tales como el ID y nombre de la canción, el nombre del artista, el ID y nombre del álbum, y el ID y nombre de la playlist. De igual forma, encontramos la fecha en que fue lanzado el álbum, y variables categóricas como el género y el subgénero al que pertenece la playlist.

En cuanto a las variables de contenido de tipo numéricas, tenemos las siguientes:

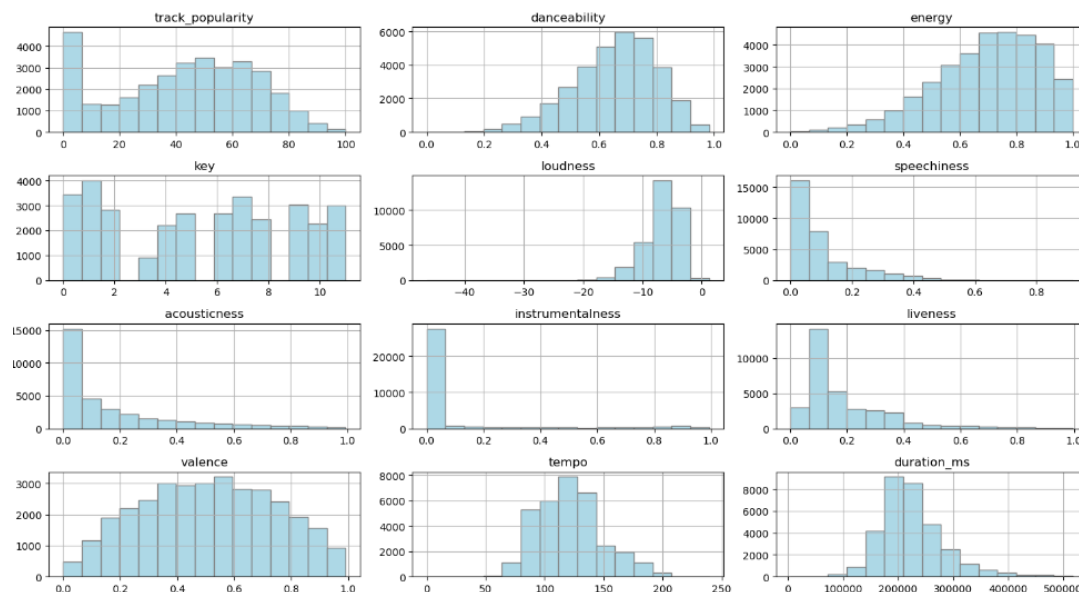
Variable	Descripción
Track popularity	Popularidad de la canción entre 0 y 100, donde entre más grande el número, mayor es la popularidad.
Danceability	Describe cuán adecuada es la pista para bailar basada en una combinación de elementos musicales que incluyen tempo, estabilidad de ritmo, fuerza de ritmo y regularidad general. Un valor de 0.0 es menos bailable y 1.0 es el más bailable.
Energy	La energía es una medida de 0.0 a 1.0 y representa una medida perceptiva de intensidad y actividad. Por lo general, las pistas enérgicas se sienten rápidas, fuertes y ruidosas. Por ejemplo, el death metal tiene alta energía, mientras que un preludio de Bach obtiene un puntaje bajo en la escala. Las características perceptivas que contribuyen a este atributo incluyen el rango dinámico, el volumen percibido, el timbre, la tasa de inicio y la entropía general.
Key	La clave general estimada de la pista. Se asignan enteros a lanzamientos utilizando notación de clase de tono estándar. P.ej. 0 = C, 1 = C#/D b, 2 = D, y así sucesivamente. Si no se detectó ninguna clave, el valor es -1.
Loudness	El volumen general de una pista en Decibels (DB). Los valores de volumen se promedian en toda la pista y son útiles para comparar el volumen relativo de las pistas. El volumen es la calidad de un sonido que es el correlato psicológico primario de la fuerza física (amplitud). Los valores rango típico se encuentran entre -60 y 0 dB.
Mode	El modo indica la modalidad (mayor o menor) de una pista, el tipo de escala de la que se deriva su contenido melódico. Mayor está representado por 1 y menor es 0.
Speechiness	El habla detecta la presencia de palabras habladas en una pista. Cuanto más exclusivamente se graba la grabación (por ejemplo, programa de entrevistas, audiolibro, poesía), más cerca de 1.0 el valor del atributo. Los valores superiores a 0.66 describen pistas que probablemente se hagan completamente de palabras habladas. Los valores entre 0.33 y 0.66 describen pistas que pueden contener música y habla, ya sea en secciones o en capas, incluidos casos como la música rap. Los valores por debajo de 0.33 probablemente representan la música y otras pistas sin voz.
Acousticness	Una medida de confianza de 0.0 a 1.0 de si la pista es acústica. 1.0 representa una alta confianza, la pista es acústica.
Instrumentalness	Predice si una pista no contiene voces. Los sonidos "OOH" y "AAH" se tratan como instrumentales en este contexto. Las pistas de rap o palabras habladas son claramente "vocales". Cuanto más cerca sea el valor de instrumentalidad a 1.0, mayor es la probabilidad de que la pista no contenga contenido vocal. Los valores superiores a 0.5 están destinados a representar pistas instrumentales, pero la confianza es mayor a medida que el valor se acerca a 1.0.
Liveness	Detecta la presencia de una audiencia en la grabación. Los valores de liveness más altos representan una mayor probabilidad de que la pista se haya realizado en vivo. Un valor superior a 0.8 proporciona una fuerte probabilidad de que la pista esté en vivo.
Valence	Una medida de 0.0 a 1.0 que describe la positividad musical transmitida por una pista. Las pistas con alta valencia suenan más positivas (por ejemplo, felices, alegres, eufóricas), mientras que las pistas con baja valencia suenan más negativas (por ejemplo, triste, deprimida, enojada).
Tempo	El tempo estimado general de una pista en ritmos por minuto (BPM). En la terminología musical, el tempo es la velocidad o el ritmo de una pieza dada y se deriva directamente de la duración promedio del latido.
Duration	Duración de la canción en milisegundos.

4.3. Análisis de variables categóricas



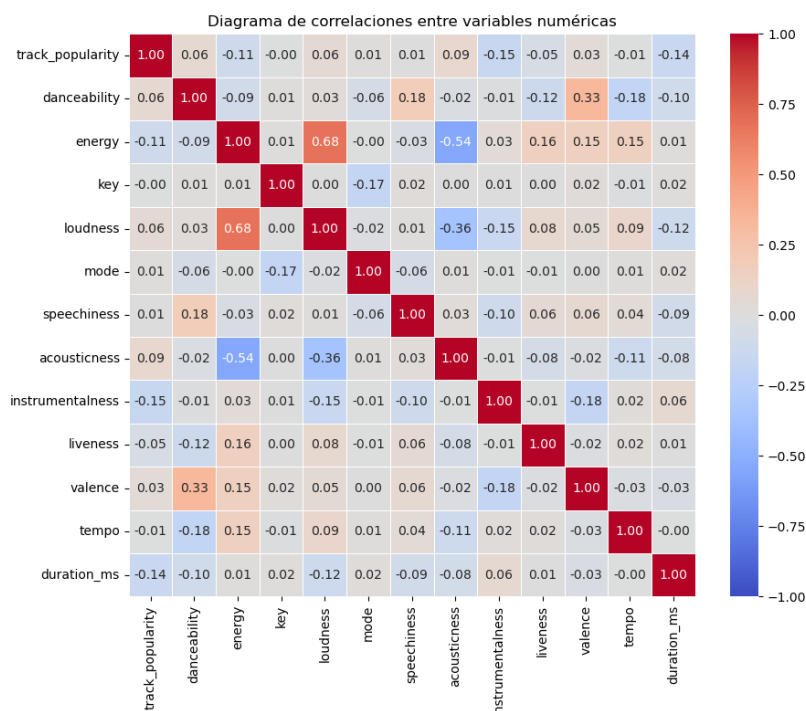
Tal y como observamos, los subgéneros con más canciones en el dataset son progressive electro house, southern hip hop, indie popitism, latin hip hop y neo soul. Mientras que los tres subgéneros con menos canciones son, permanent wave, álbum rock y reggaetón. Sin embargo, las categorías parecen estar lo suficientemente balanceadas, con un rango de 949 canciones a 1.809 canciones entre las 24 categorías.

4.4. Análisis de variables numéricas



En general, tenemos más canciones con baja popularidad que con alta popularidad. La mayor parte de las canciones son relativamente bailables, y parecen reflejar una alta energía. También observamos que la mayoría de las canciones son poco habladas, es decir, están representadas principalmente por la pista. También tenemos canciones con baja acústica y baja instrumentalización, esto último coincide con el grado de habla de la canción.

4.5. Análisis de correlaciones de variables numéricas



De manera general, observamos correlaciones no muy altas entre las variables numéricas, lo cual pudiera sugerir la no viabilidad de emplear técnicas como Análisis de Componentes Principales al set de datos para reducir su dimensión, sin embargo, esto se determinará al momento de aplicar en detalle esta técnica. Las correlaciones más altas se evidencian entre la variable ‘energy’ y ‘loudness’ (0.68), y ‘energy’ y ‘acousticness’ (0.54).

5. Propuesta metodológica

Inicialmente, se hará uso de PCA (Análisis de Componentes Principales) para reducir la dimensionalidad de las 30.000 canciones de Spotify al transformar las características musicales como acústica, instrumentalidad y energía. Esto simplifica el análisis al retener la mayor variación posible en un espacio de menor dimensión, facilitando la identificación de patrones y relaciones en el dataset.

A través de clustering, agruparemos las canciones en función de sus características musicales reducidas por PCA. Algoritmos como k-Means y k-Medoides organizan las canciones en clusters basados en similitudes, revelando géneros emergentes y patrones de estilo. Esto nos ayudará a identificar grupos de canciones que comparten características similares.

Finalmente, a través de algoritmos como el filtrado colaborativo y el filtrado basado en contenido, se diseñará un sistema de recomendación de canciones a los usuarios basándonos en las preferencias de otros usuarios con gustos similares. Al combinar esto con los clusters identificados por PCA y clustering, el sistema puede proporcionar recomendaciones más personalizadas y precisas, considerando tanto las características musicales como el comportamiento de escucha de los usuarios. Se espera que géneros similares suenen similares y vengan de períodos de tiempo similares, mientras que lo mismo puede decirse de las canciones dentro de esos géneros. Podemos usar esta idea para construir un sistema de recomendación tomando los puntos de datos de las canciones que un usuario ha escuchado y recomendando canciones correspondientes a los puntos de datos cercanos. Para esto, nos apoyaremos de Spotipy, un cliente de Python para la API web de Spotify que facilitará obtener datos y consultar el catálogo de Spotify para canciones escuchadas y elegidas por nosotros mismos, y validar, en función de esto, qué canciones se recomiendan.

6. Bibliografía

- Cilibrasi, R., Vitanyi, P., & de Wolf, R. (2004). Algorithmic clustering of music. In *Proceedings of the Fourth International Conference on Web Delivering of Music (EDELMUSIC 2004)*. IEEE. Recuperado de: [enlace](#).
- Jondya, A. G., & Iswanto, B. H. (2017). Indonesian’s traditional music clustering based on audio features. *Procedia Computer Science*, 116, 174–181. ScienceDirect. Recuperado de: [enlace](#).
- Li, T., Ogihara, M., Peng, W., Shao, B., & Zhu, S. (2009). Music clustering with features from different information sources. *IEEE Transactions on Multimedia*, 11(3), 423-434. IEEE. Recuperado de: [enlace](#).
- Schedl, M. (2019). Deep learning in music recommendation systems. *Frontiers in Applied Mathematics and Statistics*. Institute of Computational Perception, Johannes Kepler University, Linz, Austria. Recuperado de: [enlace](#).
- Song, Y., Dixon, S., & Pearce, M. (2012). A survey of music recommendation systems and future perspectives. In *Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)* (pp. 19-22). Queen Mary University of London. Recuperado de: [enlace](#).