



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

UANL

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas.

Minería de datos
Primer avance del proyecto integrador

Equipo: 08
Gpo: 003
Viernes: 6:00 pm – 9:00 pm

Integrantes

Manuel Joseph Romero Pascacio. 1811177
Manuel de Jesús Vázquez Bocanegra. 1823593
Sergio Velázquez Rivera. 1805244

1) Título de la base de datos

A) Spotify - All Time Top 2000s Mega Dataset

B) URL: <https://www.kaggle.com/iamsumat/spotify-top-2000s-mega-dataset>

2) Descripción de los datos:

A) Son tablas que contienen estadísticas de audio de las 2000 pistas principales en Spotify. Los datos contienen aproximadamente 15 columnas, cada una de las cuales describe la pista y sus cualidades. Las canciones lanzadas desde 1956 hasta 2019 se incluyen de algunos artistas notables y famosos como Queen, The Beatles, Guns N 'Roses, etc.

B) La Base de datos contiene 14 columnas:

- **Index:** ID
- **Title:** Nombre de la pista
- **Artist:** Nombre del artista
- **Top Genre:** género de la pista
- **Year:** Año de lanzamiento de la pista
- **Beats per Minute (BPM):** el tempo de la canción
- **Energy:** la energía de una canción; cuanto mayor es el valor, más enérgico. canción
- **Danceability:** cuanto mayor sea el valor, más fácil será bailar esta canción.
- **Loudness:** cuanto mayor sea el valor, más fuerte será la canción.
- **Valence:** cuanto mayor sea el valor, más positivo será el estado de ánimo de la canción.
- **Length:** la duración de la canción.
- **Acoustic:** cuanto mayor sea el valor, más acústica será la canción.
- **Speechiness:** cuanto mayor sea el valor, más palabras habladas contiene la canción.
- **Popularity:** cuanto mayor sea el valor, más popular será la canción.

3) Justificación del uso de datos:

A) Una de las características que más nos llamó la atención y que nos hizo escoger esta base de datos es la gran cantidad de datos y la cantidad de categorías que nos aportan para aplicar diferentes tipos de modelos para alcanzar nuestros objetivos y tener varias conclusiones de los resultados obtenidos.

B) Los beneficios de trabajar con esta base de datos es que al momento de utilizar los métodos de las técnicas que ocuparemos, no tendremos la problemática de tener pocos datos para aplicarlos ya que cuenta con 1994 canciones diferentes con sus datos correspondientes desglosados en las categorías ya presentadas en el punto 2.

4) Planteamiento del problema:

¿Playlist de recomendación automática? ¿Búsquedas avanzadas de canciones?

Imagina que solamente con un clic tengas la posibilidad de crear una lista de recomendaciones en base a tus gustos previamente seleccionadas, esto podría ser muy sencillo con este algoritmo ya que tomando en cuenta las diferentes características de cada canción que nos brinda la base de datos. En base a esto diseñaremos un algoritmo que con una simple selección pueda crearte una playlist que sea lo más certera posible, para que el usuario se sienta convencido que es una buena selección. Ahora también podrás buscar canciones con palabras clave como (115 BPM, bailable, relajar, estudiar, etc.) todo esto lo haremos tomando en cuenta ciertas características que definan esas palabras clave.

Estas problemáticas surgieron debido a que los reproductores de música actualmente como Spotify las playlist de recomendación que crean solo toman en cuenta el género y generan listas de recomendación solo con el género, sabemos que existen más variaciones que pueden ayudarnos a mejorar esta experiencia dentro de la aplicación. Las búsquedas son un plus que pueden ayudar a esos usuarios que son fanáticos de la música y que se fijan en más características y no solamente en el género. Todo esto podrían ser grandes herramientas que potencien la experiencia de los usuarios y que puedan tener una gran ventaja sobre las otras plataformas de música.

5) Objetivo principal

Como objetivo principal nos enfocaremos en crear una lista de recomendación de las canciones que se encuentran en la base de datos de Spotify, el usuario tomará una muestra de esta y el programa le devolverá una lista de recomendación en base a su selección, tomando en cuenta factores que influyan en la decisión como lo es el género, BPM, que tan bailable es la canción etc. Ofreciendo una respuesta más certera a partir de una búsqueda detallada en secuencia. Hablando de búsquedas como objetivo secundario lo que se podía implementar en este programa es desarrollar un algoritmo que pueda colocar una extensión de búsqueda inteligente a través del BPM de las canciones que están en la base de datos de Spotify.

6) Planeación de la herramienta a utilizar

Revisando nuestros objetivos podemos darnos cuenta de que estaríamos trabajando inicialmente dos técnicas:

Clasificación: Realizar modelos con los datos que tenemos de las categorías para predecir que música es la que le podría llegar a gustar al usuario como, por ejemplo: arboles de decisión, redes neuronales, etc.

Clustering por parte de la categoría descriptiva y patrones secuenciales por parte de la categoría predictiva.