



Propuesta de proyecto: clasificación de género musical

Instituto Politécnico Nacional.
Escuela Superior de Cómputo.
Licenciatura en ciencia de datos.

Nombre de la materia: Desarrollo de Aplicaciones para el Análisis de Datos
Grupo: 4AV1
Profesora: Sandra Luz Morales Guitron

Lopez Mendez Emiliano
Hinostroza Loera Leonardo

INDICE

INTRODUCCION	3
DESAROLLO	3
CONCLUSIONES	6

INTRODUCCION

En este proyecto, nos enfocamos en desarrollar un sistema de clasificación automática de géneros musicales utilizando técnicas de Data Science. La clasificación de música por géneros es una tarea compleja que requiere identificar patrones en las características auditivas y metadatos disponibles. El propósito principal de este proyecto es explorar cómo algoritmos de machine learning pueden aprender y predecir el género de una canción a partir de atributos extraídos de los datos musicales, como el ritmo, tono, timbre y otras propiedades estadísticas.

El análisis y clasificación de música tiene aplicaciones significativas en la industria del entretenimiento, como en los sistemas de recomendación en plataformas de streaming. En este proyecto, aplicaremos técnicas de preprocesamiento, extracción de características y algoritmos supervisados de machine learning para mejorar la precisión en la predicción de los géneros musicales.

DESAROLLO

Siguiendo la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), nuestro proyecto se divide en seis fases principales para garantizar un desarrollo estructurado y eficaz:

1. Entendimiento del negocio

El objetivo de este proyecto es proporcionar una recomendación musical personalizada basada en las preferencias del usuario. La idea es que el usuario ingrese sus tres canciones favoritas, y a partir de esas selecciones, nuestro sistema recomendará cinco canciones adicionales que podrían gustarle, utilizando un dataset previamente recopilado. Esta solución tiene aplicaciones prácticas en plataformas de streaming o sistemas de recomendación musical.

2. Entendimiento de los datos

Buscaremos datasets disponibles en línea con características musicales relevantes para la recomendación, como el dataset **GTZAN** o el API de **Spotify**, que nos permiten extraer atributos musicales como:

- Género.
- BPM (tempo).
- Tono y energía.
- Popularidad de la canción.

Además, utilizaremos **web scraping** para complementar la información del dataset, obteniendo detalles adicionales de canciones desde plataformas como **Last.fm** o **Spotify**, utilizando herramientas de Python como BeautifulSoup y requests.

3. Preparación de los datos

En esta fase, realizaremos las siguientes tareas:

- **Limpieza:** Eliminamos datos faltantes o duplicados.
- **Preprocesamiento:** Normalizamos valores numéricos para mejorar la precisión del modelo.
- **Enriquecimiento:** Integramos datos obtenidos por web scraping, como comentarios o etiquetas adicionales.
- **Extracción de características:** Usamos librosa para calcular atributos del audio si es necesario.

4. Modelamiento

Para desarrollar el sistema de recomendación, seguiremos estos pasos:

- **Selección del algoritmo:** Usaremos un algoritmo basado en similitud de vectores, como **K-Nearest Neighbors (KNN)**, para encontrar canciones similares según las características musicales.
- **Entrenamiento del modelo:**
 - Dividiremos el dataset en datos de entrenamiento y prueba.

- Calcularemos la distancia entre canciones basándonos en sus atributos musicales.
- **Proceso de recomendación:**
 - El usuario ingresará sus tres canciones favoritas.
 - Buscaremos estas canciones en el dataset y extraeremos sus características.
 - Calcularemos la similitud entre estas canciones y otras en el dataset.
 - Recomendaremos las cinco canciones más cercanas utilizando KNN.

5. Evaluación

- **Evaluación de resultados:** Calcularemos la precisión del sistema de recomendación utilizando métricas como:
 - **Precision:** Para evaluar qué porcentaje de las canciones recomendadas eran relevantes para el usuario.
 - **Mean Average Precision (MAP):** Promedio de las precisiones en diferentes conjuntos de prueba.
- **Pruebas adicionales:** Compararemos los resultados obtenidos con diferentes configuraciones del modelo, ajustando parámetros como el número de vecinos en KNN.

6. Publicación

El proyecto final se entregará en un notebook de Python con el código documentado, listo para su ejecución. Además, elaboraremos una presentación con los resultados más relevantes, incluyendo gráficas y ejemplos de recomendaciones obtenidas. El sistema de recomendación estará disponible para pruebas, permitiendo a los usuarios ingresar sus canciones y recibir recomendaciones en tiempo real.

CONCLUSIONES

Este proyecto aplicará habilidades fundamentales de **Data Science** en Python, integrando extracción de datos mediante **web scraping**, análisis musical y recomendación personalizada mediante **KNN**. Al seguir la metodología **CRISP-DM**, nos aseguramos de que cada fase se ejecute de manera estructurada y eficiente.

El sistema de recomendación ofrecerá una experiencia personalizada, ayudando al usuario a descubrir música basada en sus gustos personales. Además, la integración de herramientas como BeautifulSoup, scikit-learn y librosa fortalecerá nuestras competencias en programación y análisis de datos, preparando el camino para futuros proyectos más complejos.

Este proyecto no solo refleja la importancia de la ciencia de datos en la industria musical, sino también el impacto que puede tener en mejorar la experiencia del usuario mediante la personalización. A futuro, podríamos ampliar la solución incorporando APIs más robustas como **Spotify API** para sugerencias aún más precisas.