Logo

Description automatically generatedLogo

Description automatically generated with medium confidence

*Proyecto avance 02*

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

Leonardo Hinostroza Loera

INDICE

[INTRODUCCION 3](#_Toc179453627)

[DESAROLLO 3](#_Toc179453628)

[LISTAS 3](#_Toc179453629)

[DICCIONARIOS 3](#_Toc179453630)

[SETS 3](#_Toc179453631)

[FICHEROS 3](#_Toc179453632)

[CONCLUSIONES 3](#_Toc179453633)

# INTRODUCCION

El proyecto busca desarrollar un sistema de recomendación musical que, basado en las tres canciones favoritas del usuario, sugiera cinco canciones adicionales con características similares. Para ello, se aplican técnicas de **ciencia de datos**, como la extracción de características musicales con librosa y la implementación del algoritmo **K-Nearest Neighbors (KNN)** para el modelamiento.

Este proyecto sigue la metodología **CRISP-DM**, que estructura el desarrollo en fases como la comprensión del problema, la preparación de datos y la evaluación del modelo. Además, se utiliza **web scraping** para enriquecer los datos obtenidos con información adicional desde plataformas como Last.fm. El resultado será un sistema capaz de mejorar la experiencia del usuario en plataformas de streaming mediante recomendaciones musicales personalizadas.

# DESAROLLO

En este código, realizamos un **web scraping** para extraer información de canciones y artistas de la página específica de Last.fm, y posteriormente almacenamos esta información en un formato estructurado mediante un **DataFrame** de pandas. A continuación, se describe brevemente el flujo del proceso:

1. **Obtención de los Datos (Web Scraping):**
   * Utilizamos la librería requests para realizar una solicitud HTTP y acceder al contenido HTML de la página web de Last.fm.
   * Analizamos el contenido HTML con BeautifulSoup para identificar y extraer los elementos correspondientes a las canciones y artistas.
2. **Procesamiento de los Datos:**
   * Los nombres de las canciones y los artistas se extraen utilizando selectores específicos (class='chartlist-name' y class='chartlist-artist').
   * Guardamos cada canción y su artista en una lista de diccionarios, donde cada diccionario representa una fila con las claves "Canción" y "Artista".
3. **Estructuración de los Datos:**
   * Convertimos la lista de diccionarios en un **DataFrame** de pandas, lo que permite manipular los datos de forma más fácil y ordenada.
4. **Obtención de los datos**

Para cada pagina, necesitábamos buscar en donde se guardan las y los nombres de la canción y del artista, como en los tags no viene el genero de la canción, mejor solo hacemos la diferencia por pagina para ver donde están

A screenshot of a computer program

Description automatically generated

Primero obtenemos el callback del sitio web, 200, significa que si funciona y obtenemos el html.parser

Hicimos cada eso para cada pagina de genero de música

Rock: A screen shot of a computer

Description automatically generated

Hip-hop: A computer screen shot of a program code

Description automatically generated

Indie: A screen shot of a computer screen

Description automatically generated

Reggae: A computer screen shot of a program code

Description automatically generated

Dance: A computer screen shot of a program code

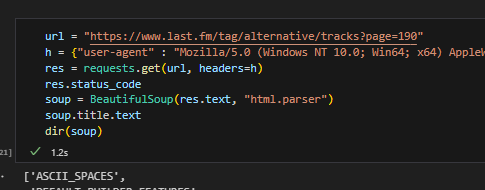
Description automatically generated

Rnb: A computer screen with text and images

Description automatically generated

Rap: A screen shot of a computer screen

Description automatically generated

Alt: 

Electronica: A screen shot of a computer

Description automatically generated

1. **Procesamiento**

Luego necesitamos ver las etiquetas para cada canción

El nombre esta en chartlist-name A screenshot of a computer

Description automatically generated

El autor esta en chartlist-artist

A screenshot of a computer

Description automatically generated

De aquí podemos crear un ciclo for que nos extrae las canciones de la pagina A screenshot of a music album

Description automatically generated A computer screen shot of text

Description automatically generated

Extrae, la canción y el nombre del artista, con chartlist-row, se obtiene todas las canciones de esa pagina web y termina hasta que ya no encuentre mas canciones. Hacemos eso para cada genero musical.

1. **Extracción**

Para mas facilidad de manejar los datos**,** los manejaremos como dataframes, entonces por cada genero hacemos una lista del genero: por ejemplo rock  y en el ciclo for, se añade a la lista con append usando pandas

Finalmente convertimos la lista en un dataframe y lo mostramosA screenshot of a computer

Description automatically generated

Así podemos manejar mejor el uso de las canciones

A screen shot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screen shot of a computer program

Description automatically generatedA screen shot of a computer

Description automatically generatedA screenshot of a computer program

Description automatically generated

Se hizo para cada genero de musica

# CONCLUSIONES

Durante el desarrollo de aproximadamente 46 programas, se trabajó principalmente con métodos básicos como append para listas y lower y strip para manipulación de cadenas de texto. Además, se utilizaron métodos como istitle para verificar la capitalización de las cadenas y path.exists para verificar rutas de archivos. También se implementaron funciones relacionadas con el manejo de rutas (path.abspath, path.join) y entorno (environ.get), así como funciones de visualización (rich.print) y manejo de módulos (runpy.run\_path). Estos ejercicios no introdujeron conceptos nuevos, sino que se enfocaron en reforzar el conocimiento existente de estos métodos comunes. Fue un excelente repaso después de las vacaciones, permitiendo retomar la práctica de la programación tras un periodo de inactividad.