



PRESENTACION

# CLASIFICADOR DE MUSICA POR GENEROS

VANESSA LOZANO  
PAOLA GALLEGO PINTO

# INTEGRANTES:



**Paola Gallegos  
Pinto**



**Vanessa Lozano**

# DATOS DE AUDIO

## INTRODUCCION

Trabajar con datos de audio ha sido un problema relativamente menos explorado en el machine learning. En la mayoria de los casos, los puntos de referencia para el trabajo fundamental en Deep learning se miden en el rendimiento de datos de texto e imagenes.



# ¿QUÉ ES LA CLASIFICACIÓN DE GÉNEROS MUSICALES?

## FORMAS DE CONVERTIR DATOS DE AUDIO EN NUMÉRICOS

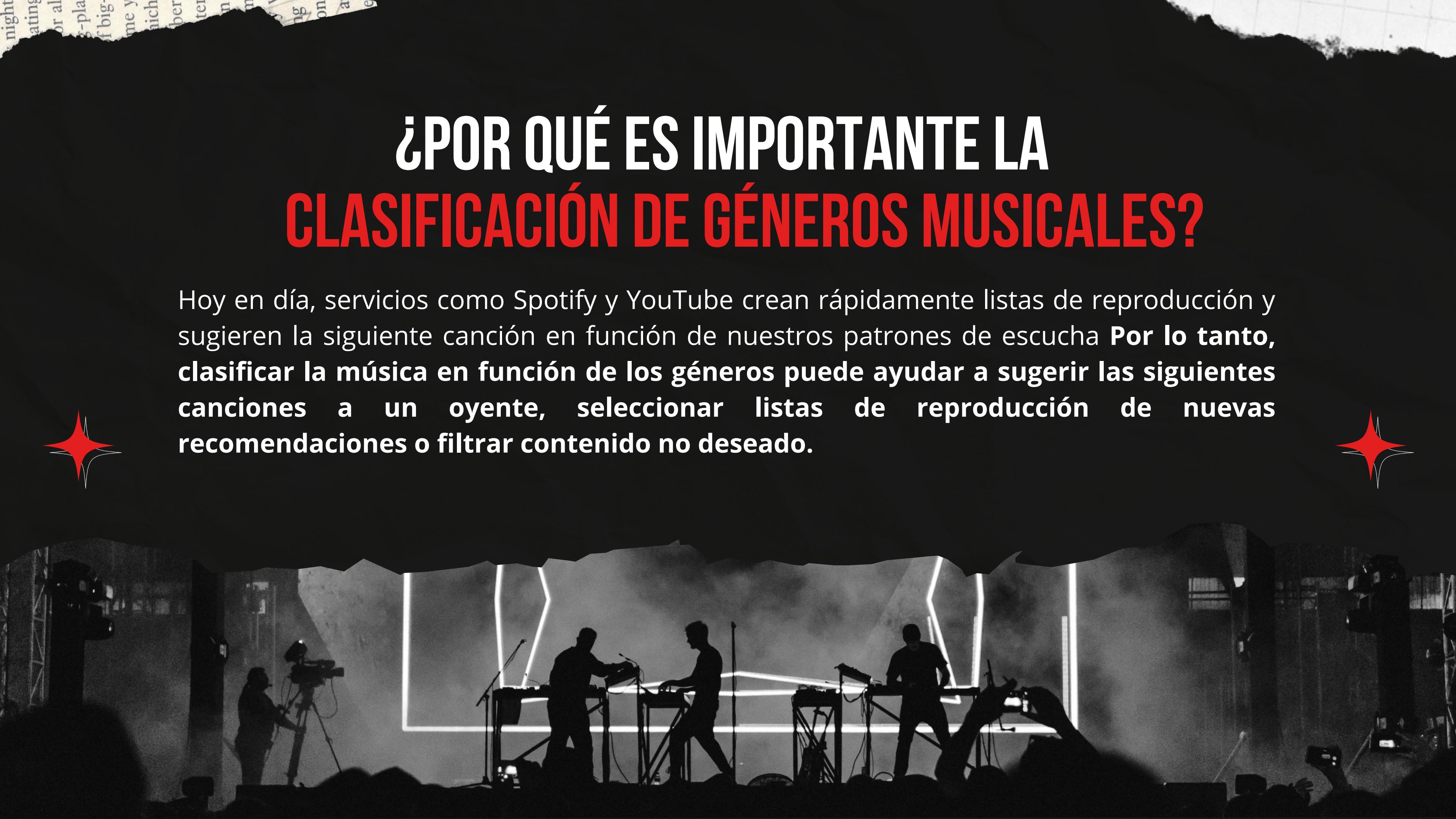
Clasificar la música en varios géneros (hip hop, rock, jazz, folk, pop, etc.) implica extraer características valiosas de los datos de audio, preprocesarlos y entrenar un modelo clasificador de aprendizaje automático.

Utilizando componentes de frecuencia, características acústicas o espectrogramas

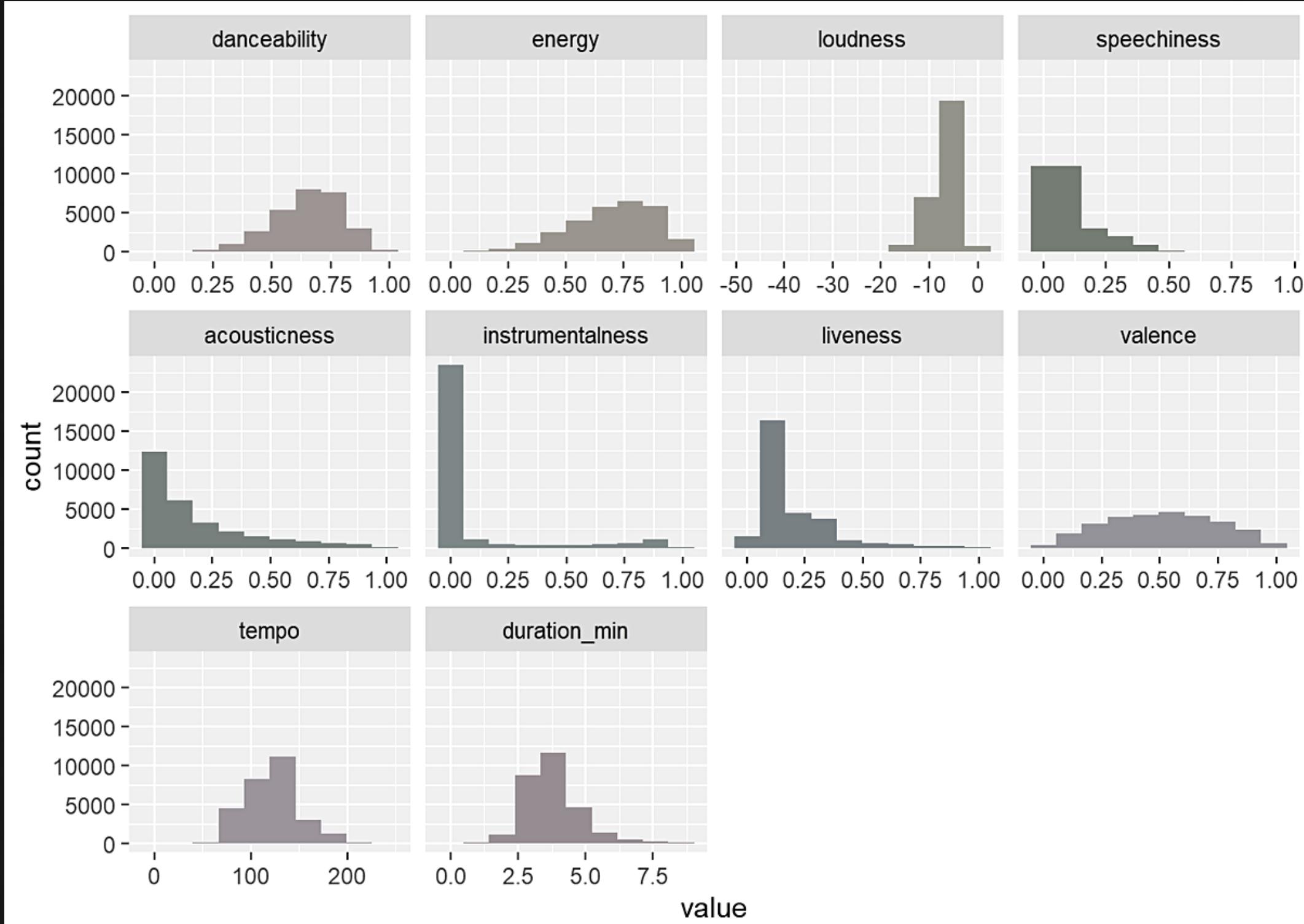


# ¿POR QUÉ ES IMPORTANTE LA CLASIFICACIÓN DE GÉNEROS MUSICALES?

Hoy en día, servicios como Spotify y YouTube crean rápidamente listas de reproducción y sugieren la siguiente canción en función de nuestros patrones de escucha. **Por lo tanto, clasificar la música en función de los géneros puede ayudar a sugerir las siguientes canciones a un oyente, seleccionar listas de reproducción de nuevas recomendaciones o filtrar contenido no deseado.**

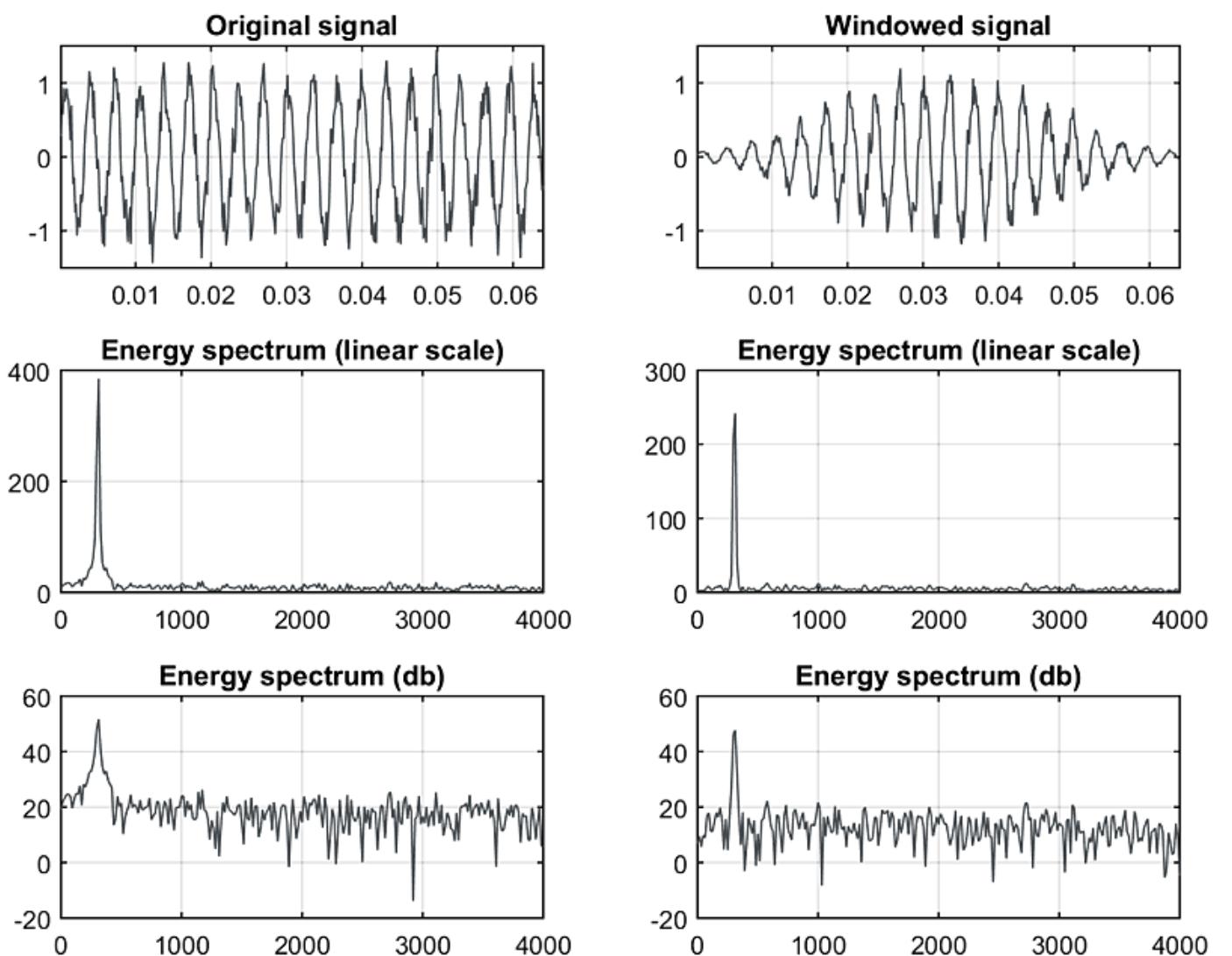


# ¿CÓMO CLASIFICAR LOS GÉNEROS MUSICALES?



La conversión de datos de audio a formato numérico o vectorial determinará cuánta información fundamental se retiene cuando se pierde la forma de audio. Por ejemplo, si un formato de datos no puede representar el volumen y el ritmo de una canción de rock, sería difícil incluso para los mejores modelos de aprendizaje automático aprender el género y clasificar la muestra. De los diversos métodos para representar datos de audio:

# MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCCS)



MFCCs son coeficientes para la representación del habla basados en la percepción auditiva humana. Estos surgen de la necesidad, en el área del reconocimiento de audio automático, de extraer características de las componentes de una señal de audio que sean adecuadas para la identificación de contenido relevante.



*Veremos uno de esos enfoques para construir un modelo clasificador de género musical utilizando un conjunto de datos de código abierto GTZAN.*



# **DATASET: GTZAN**

El conjunto de datos consta de 1000 pistas de audio cada una de 30 segundos de duración. Contiene 10 géneros, cada uno representado por 100 pistas. Las pistas son todos archivos de audio mono de 16 bits de 22050 Hz en formato .wav.

- Jazz
- Classical
- Country
- Rock
- Pop
- Hip-Hop
- Blues
- Metal
- Reggae
- Disco



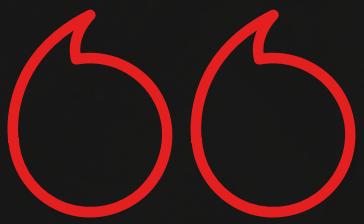


# DATASET: GTZAN

El conjunto de datos tiene las siguientes carpetas:



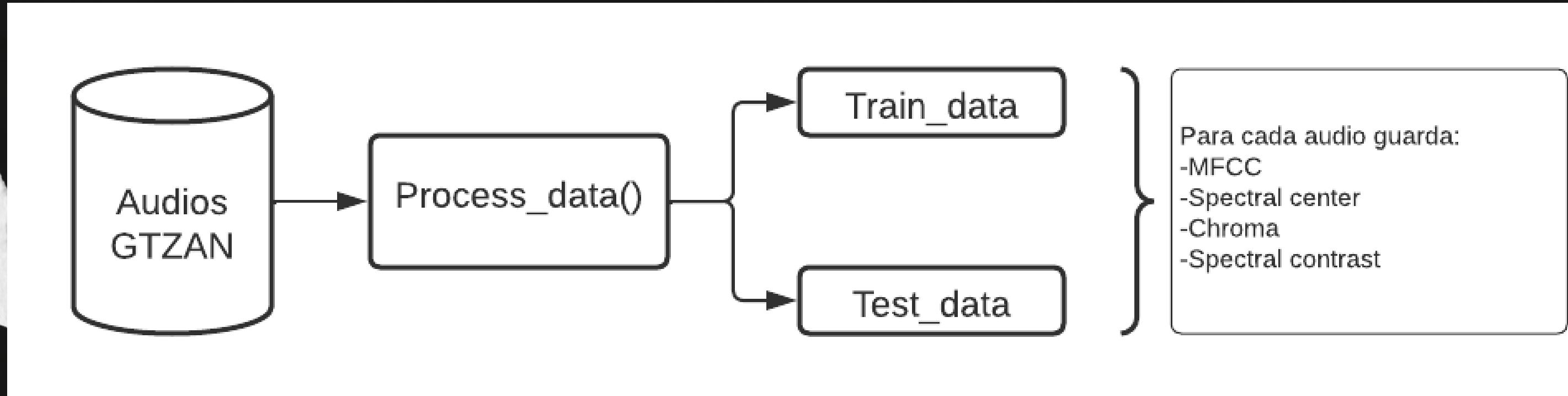
- **Géneros originales:** una colección de 10 géneros con 100 archivos de audio cada uno, todos con una duración de 30 segundos.
- **Imágenes originales:** una representación visual de cada archivo de audio. Una forma de clasificar los datos es a través de redes neuronales porque las NN generalmente toman algún tipo de representación de imagen.
- **2 archivos CSV:** contienen características de los archivos de audio. Un archivo tiene para cada canción (30 segundos de duración) una media y una varianza calculadas sobre múltiples funciones que se pueden extraer de un archivo de audio. El otro archivo tiene la misma estructura, pero las canciones se dividen antes en archivos de audio de 3 segundos.

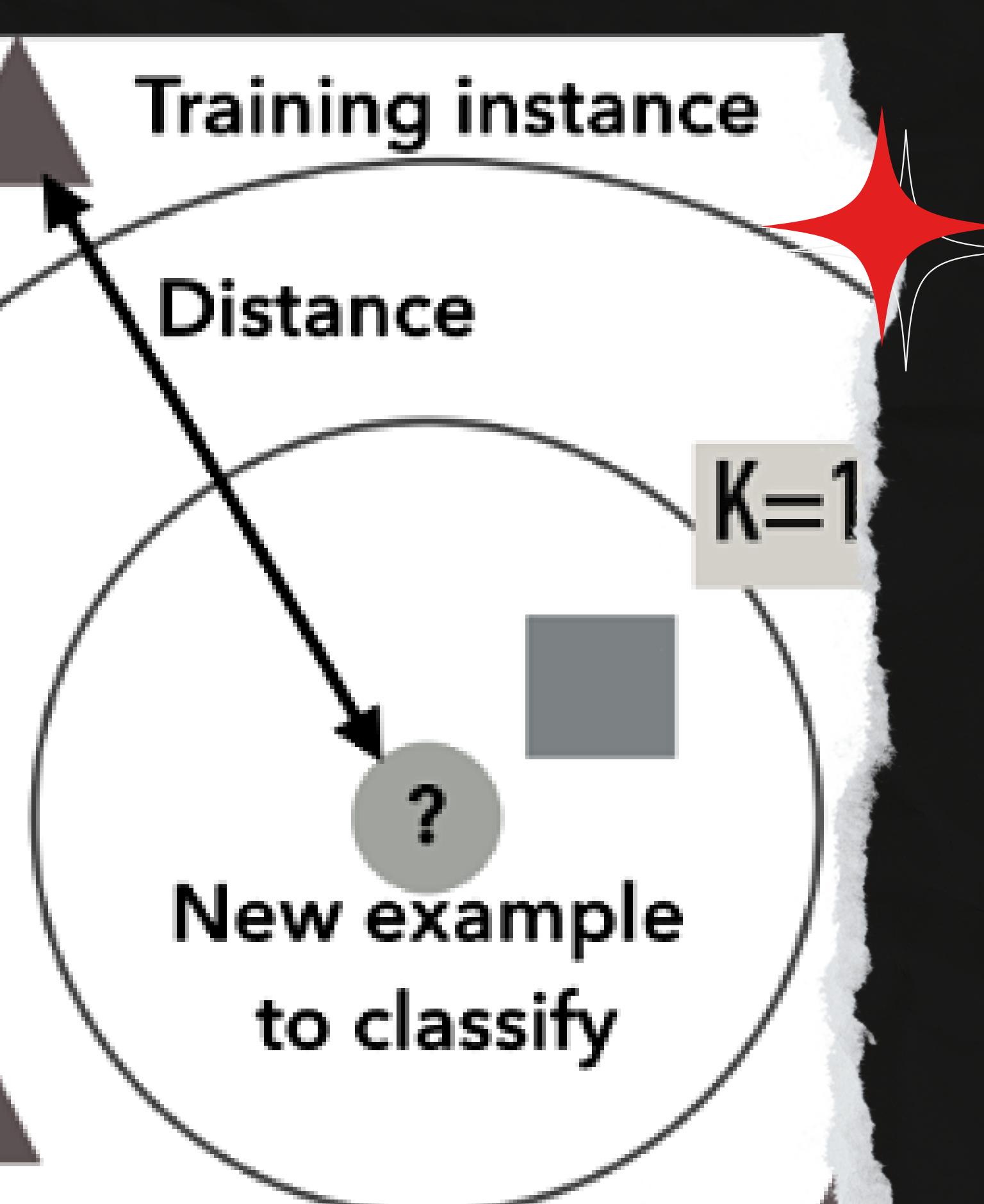


*Por lo tanto, con el conjunto de datos GTZAN, podemos realizar una clasificación de géneros musicales usando KNN (K nearest neighbors) u otros modelos de visión por computadora o usar datos de series temporales como MFCC en el entrenamiento de un LSTM (Long short-term memory) o CNN (Convolutional neural network).*



# USO DE LA BASE DE DATOS GTZAN





Training instance

Distance

$K=1$

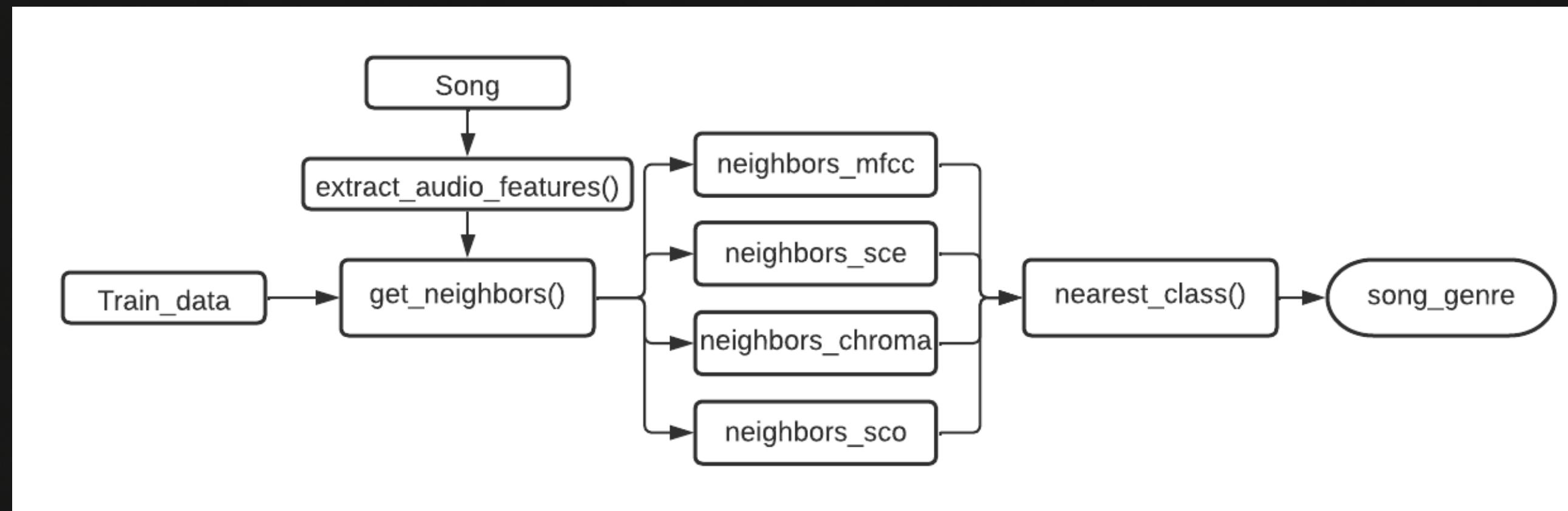
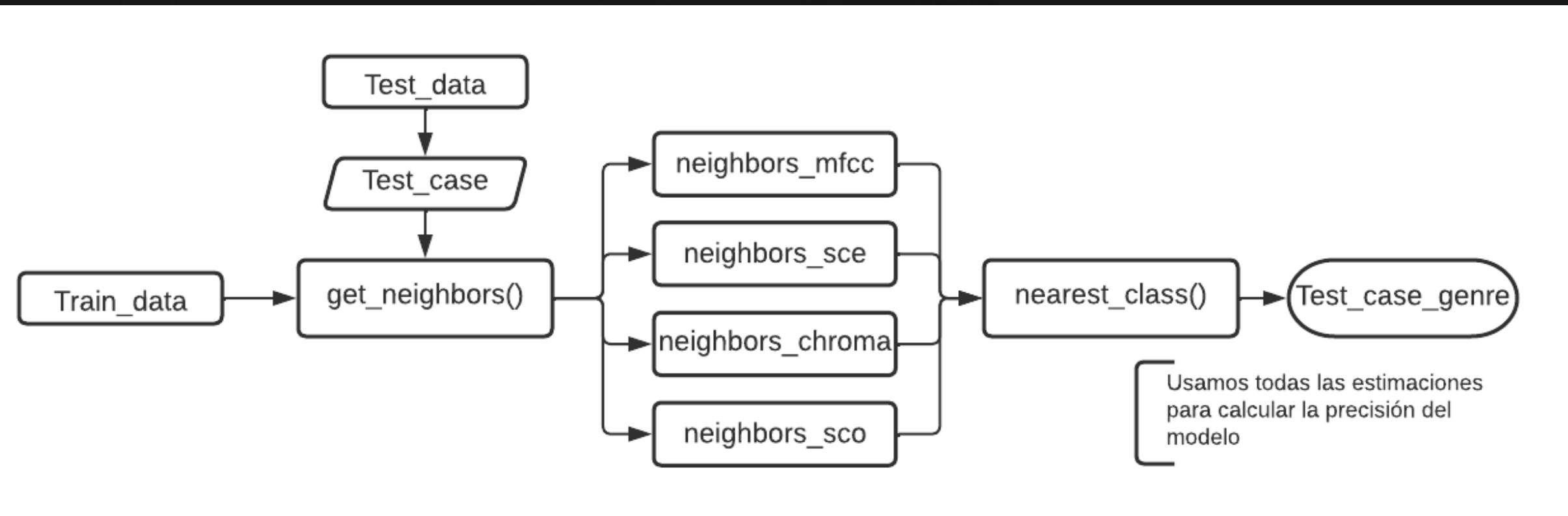
New example  
to classify

## VEAMOS UN POCO DEL MODELO KNN

- Se utiliza para encontrar soluciones para problemas de clasificación y regresión.
- Se basa en datos de entrada etiquetados para procesar datos no etiquetados en el futuro

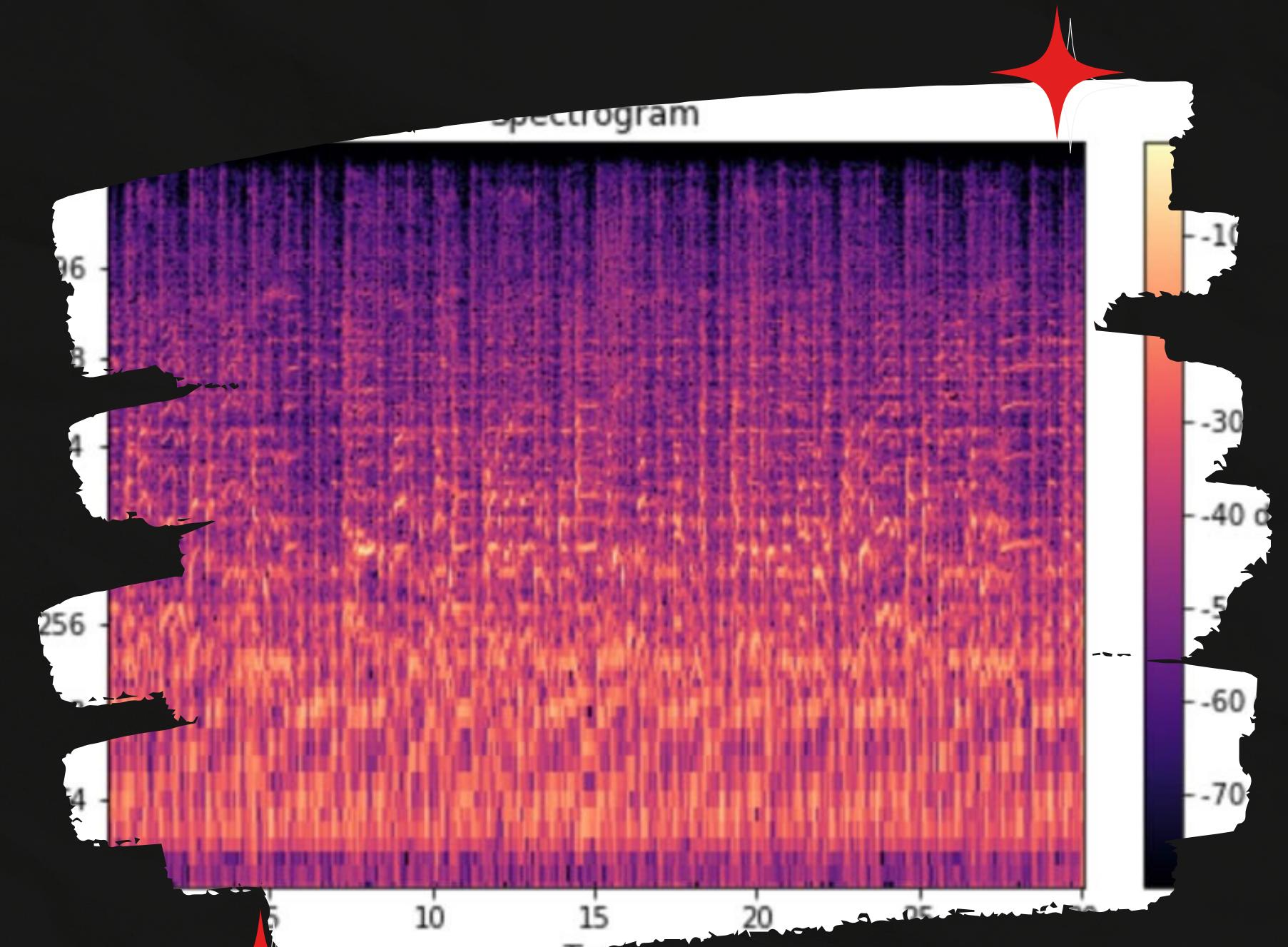
" The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other. " KNN Algorithm - ML

# CLASIFICAR GÉNEROS MUSICALES USANDO EL MODELO KNN

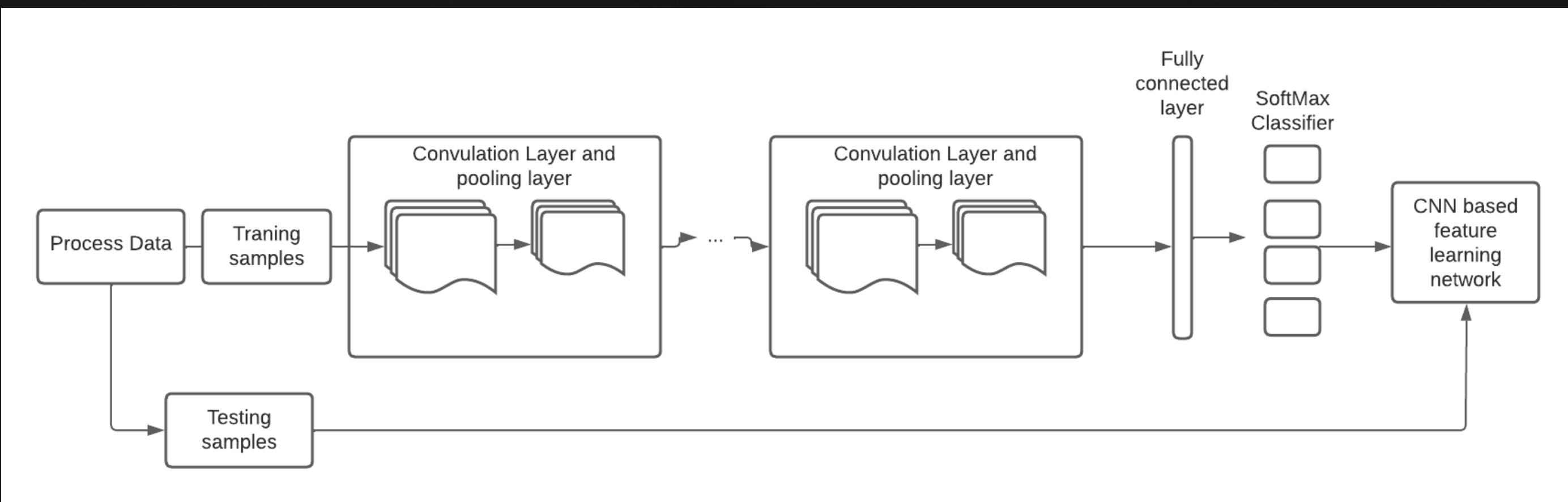


# VEAMOS UN POCO DEL MODELO CNN

- Algoritmo de aprendizaje profundo que puede tomar una imagen de entrada, asignar importancia (pesos) a varias características en la imagen y poder diferenciar una de otra.
- CNN trabaja con convolución y agrupación en secuencia alterna en la red neuronal.
- Las entradas para nuestro modelo CNN son los espectogramas asociados a cada audio



# CLASIFICAR GÉNEROS MUSICALES USANDO EL MODELO CNN



## ALGUNOS RESULTADOS

- **Efectividad:** Usando el modelo KNN obtuvimos una efectividad 40-60% y para K muy grandes de 10-20%, mientras que usando el modelo CNN la efectividad fue del 70-75%
- **Otros factores que afectan en la precisión:** La precisión de ambos modelos al clasificar un audio depende de que género sea, pues existen géneros con características muy similares o que carecen de propiedades distintivas, lo que hace que el modelo sea más propenso a confundirlo y clasificarlo incorrectamente.

**MUSIC**

soul blues rock & roll  
surf metal punk rock  
rave goth disco jazz trip hop  
noise garage new wave funk jazz  
blues goth disco jazz trip hop  
punk rock  
techno  
psychedelic  
glam rock

# THANK'S FOR YOUR ATTENTION