

Reconocimiento de géneros musicales utilizando técnicas de Machine Learning y Deep Learning

Daniel Aguayo

daniel.aguayo2001@alumnos.ubiobio.cl

Abstract

Este proyecto se centra en la clasificación de géneros musicales mediante Machine Learning. Utilizando el conjunto de datos GTZAN, que proporciona funciones de audio extraídas de un conjunto diverso de canciones, implementamos un proceso para preparar, procesar, analizar y clasificar canciones en géneros predefinidos. La implementación se lleva a cabo en Python, aprovechando bibliotecas populares como Librosa y Scikit-aprende. Este trabajo es una aplicación práctica desarrollada para el curso de Introducción a las aplicaciones de algoritmos de Machine Learning y Deep Learning.

Keywords: Reconocimiento de género musical, Machine Learning, GTZAN Dataset, Procesamiento de audio

Metadata

1. Descripción del software

1.1. Arquitectura del software

El proyecto está implementado en Python, organizado en módulos específicos para el preprocesamiento de datos de audio, la extracción de características y la clasificación de géneros musicales. El software utiliza las siguientes bibliotecas:

- **Librosa:** Para el procesamiento de audio, extracción de características como espectrogramas y mel-frequency cepstral coefficients (MFCC).
- **NumPy:** Para la manipulación de datos numéricos y cálculos matriciales.
- **Matplotlib y Seaborn:** Para la visualización de resultados y análisis exploratorio de datos.

Nr.	Descripción de los metadatos del código	Detalles
C1	Versión actual del código	v1.0.0
C2	Enlace permanente al código/repositorio utilizado para esta versión	https://github.com/Dantrotel/Clasificacion_generos_musicales
C3	Enlace permanente a una cápsula reproducible	https://colab.research.google.com/drive/1XNsHoY300LLpsq10kjMYYpreRoCwVA8H?usp=sharing
C4	Licencia del código	MIT License
C5	Sistema de control de versiones utilizado	git
C6	Lenguajes de programación, herramientas y servicios utilizados	Python, Jupyter Notebook, librosa, NumPy, Matplotlib, Seaborn, soundfile, scikit-learn, TensorFlow, yt-dlp, pydub
C7	Requisitos de compilación, entornos operativos y dependencias	Python 3.8 o superior, Jupyter Notebook, librosa, NumPy, Matplotlib, Seaborn, soundfile, scikit-learn, TensorFlow, yt-dlp, pydub
C8	Enlace a la documentación o manual del desarrollador	https://github.com/Dantrotel/Clasificacion_generos_musicales/blob/main/README.md
C9	Correo de soporte para consultas	Daniel.aguayo2001@alumnos.ubiobio.cl

Table 1: Metadatos del código (obligatorio)

- **Scikit-learn:** Para la preprocesamiento de datos, codificación de etiquetas y evaluación de modelos.
- **TensorFlow/Keras:** Para el desarrollo y entrenamiento de la red neuronal utilizada en la clasificación.
- **Pydub:** Para la manipulación y conversión de archivos de audio.
- **yt-dlp:** Para la descarga de canciones desde YouTube, en caso de requerir nuevas muestras de audio.
- **Soundfile:** Para la lectura y escritura de archivos de audio en diferentes formatos.

El diseño modular permite una fácil extensión y personalización para incluir nuevos géneros musicales o mejorar el rendimiento del modelo de clasificación.

Nr.	Descripción de los metadatos del software ejecutable	Detalles
S1	Versión actual del software	1.0.0
S2	Enlace permanente a los ejecutables de esta versión	https://github.com/Dantrotel/Clasificacion_generos_musicales
S3	Enlace permanente a una cápsula reproducible	https://colab.research.google.com/drive/1XNsHoY300LLpsq10kjMYYpreRoCwVA8H?usp=sharing
S4	Licencia del software	MIT License
S5	Plataformas de computación/sistemas operativos	Windows, macOS, Linux
S6	Requisitos de instalación y dependencias	Python 3.8 o superior, Jupyter Notebook, librosa, NumPy, Matplotlib, Seaborn, soundfile, scikit-learn, TensorFlow, yt-dlp, pydub
S7	Enlace al manual del usuario	https://github.com/Dantrotel/Clasificacion_generos_musicales/blob/main/README.md
S8	Correo de soporte para consultas	Daniel.aguayo2001@alumnos.ubiobio.cl

Table 2: Metadatos del software (opcional)

1.2. Funcionalidades del Software

- **Preprocesamiento de Datos de Audio:** El software preprocesa los archivos de audio del dataset GTZAN, asegurándose de que todos los audios tengan una tasa de muestreo consistente y eliminando silencios o ruidos no deseados. También se normalizan los volúmenes de las canciones para evitar sesgos debido a diferencias de volumen entre ellas.
- **Extracción de Características de Audio Relevantes (como MFCCs):** Se extraen las características más relevantes de las canciones, tales como los coeficientes cepstrales en la frecuencia Mel (MFCCs), que capturan los aspectos acústicos fundamentales de cada género musical. Además de MFCCs, también se pueden extraer características adicionales como espectrogramas, Chroma features o tasa de cruce por cero (ZCR).
- **Entrenamiento y Evaluación de Modelos de Machine Learning para Clasificación de Géneros:** Se entrena modelos de machine learning (como máquinas de soporte vectorial, bosques aleatorios, o re-

des neuronales profundas) para clasificar las canciones según su género. Durante el entrenamiento, se divide el dataset en conjuntos de entrenamiento y validación. El rendimiento del modelo se evalúa utilizando métricas como precisión, recall, F1-score y matrices de confusión.

- **Visualización de Resultados de Clasificación:** Una vez entrenado el modelo, se generan gráficos para visualizar los resultados de clasificación, tales como matrices de confusión para observar la distribución de las predicciones de géneros y gráficos de precisión para evaluar el rendimiento del modelo en cada clase. Si se usan redes neuronales profundas, también se pueden generar gráficos de aprendizaje, como la evolución de la pérdida durante las épocas de entrenamiento.

2. Resultados

Después de entrenar el modelo de clasificación de géneros musicales utilizando el dataset GTZAN y un clasificador Random Forest, se evaluó el rendimiento del modelo utilizando un conjunto de datos de validación. Los resultados de la clasificación se resumen a continuación:

Género	Precisión	Recall	F1-Score	Soporte
Blues	0.75	0.83	0.79	18
Classical	0.86	1.00	0.93	19
Country	0.53	0.67	0.59	12
Disco	0.60	0.50	0.55	24
HipHop	0.47	0.53	0.50	15
Jazz	0.83	0.74	0.78	27
Metal	0.75	0.86	0.80	21
Pop	0.78	0.90	0.84	20
Reggae	0.44	0.36	0.40	22
Rock	0.41	0.32	0.36	22
Promedio Macro	0.64	0.67	0.65	200
Promedio Ponderado	0.65	0.67	0.65	200

Table 3: Matriz de Evaluación de Resultados del Clasificador Random Forest

- **Precisión global:** 67%
- **Matriz de Evaluación por Género:**

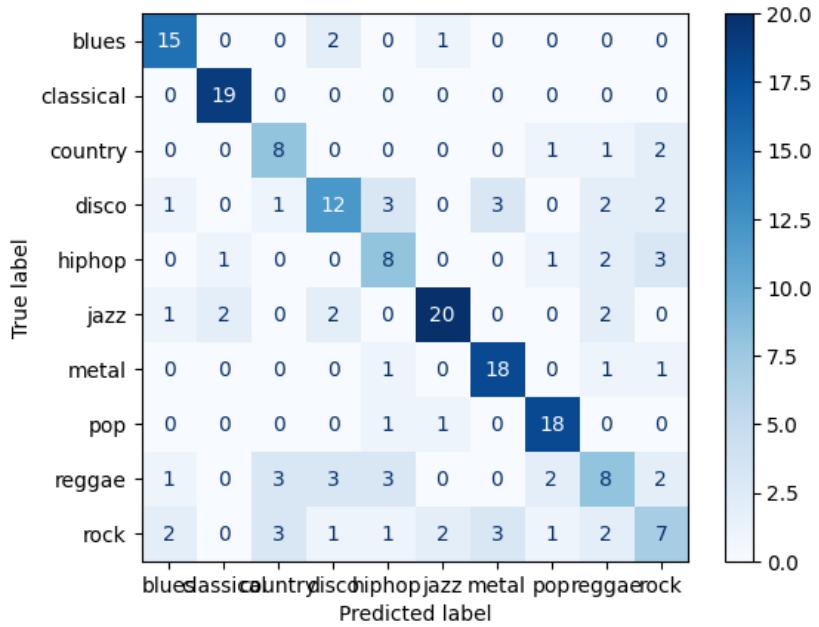


Figure 1: Matriz de Confusión de los Resultados de Clasificación

3. Errores en la Clasificación de Canciones

Durante las pruebas del modelo con canciones de los grupos AC/DC, KISS y algunos artistas pop, se observaron varios errores en la clasificación. Las canciones de AC/DC y KISS fueron clasificadas incorrectamente como música clásica o en algunos casos muy específicos, como hip-hop. Además, se probó con canciones de artistas pop como "My Way" de Calvin Harris con Rihanna y "Mirror" de Justin Timberlake, las cuales fueron clasificadas erróneamente como *Reggae* y *Blues*, respectivamente. Es importante señalar que el modelo nunca clasificó estas canciones como *rock*, a pesar de que tanto AC/DC como KISS pertenecen a este género, y las canciones de los artistas pop tampoco fueron clasificadas correctamente dentro de su género correspondiente.

Possible causa de los errores:

- Similitudes y superposición de características acústicas entre géneros:** Las canciones de AC/DC y KISS, aunque pertenecen al género *rock*, pueden compartir algunas características acústicas o patrones en los MFCCs con otros géneros, como *música clásica* o *hip-hop*. Elementos como las guitarras eléctricas, los cambios en las frecuencias y los ritmos pueden tener similitudes con los de estos géneros, especialmente en canciones menos intensas o con estructuras rítmicas poco convencionales. De manera similar, las canciones pop, como "My Way"

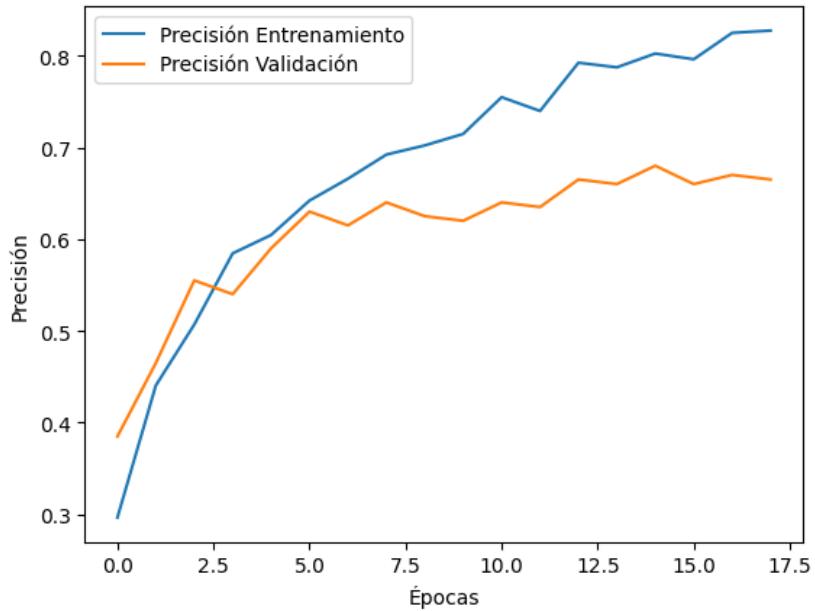


Figure 2: Gráfico de Precisión del Clasificador

de Calvin Harris ft Rihanna y "Mirror" de Justin Timberlake, presentan una fusión de géneros y el uso de efectos electrónicos y cambios rítmicos comunes en géneros como *Reggae*, *Blues* y *electronica*. Esta superposición de características acústicas entre géneros podría haber confundido al modelo y causado la clasificación errónea de canciones de AC/DC, KISS y artistas pop en géneros no relacionados.

Conclusión:

Los resultados obtenidos del modelo de clasificación de géneros musicales muestran un rendimiento moderado, con una precisión global del 67%. Sin embargo, se observaron errores notables en la clasificación de canciones de varios géneros. Por ejemplo, canciones de AC/DC como "Back in Black" fueron correctamente clasificadas como *Metal*, mientras que otras como "Thunderstruck" fueron erróneamente clasificadas como música *Disco*. Este comportamiento ilustra que la clasificación depende significativamente de las características específicas de cada canción, lo que sugiere que el modelo tiene dificultades para captar las sutilezas dentro de un mismo grupo o género. En el caso de artistas pop, canciones como "My Way" de Calvin Harris con Rihanna y "Mirror" de Justin Timberlake fueron clasificadas erróneamente como *Reggae* y *Blues*, respectivamente. Este tipo de errores refleja la complejidad y variabilidad dentro de los géneros musicales, donde ciertas canciones pueden compartir características acústicas con géneros inesperados, lo que

confunde al modelo.

A pesar de estos errores, géneros como *Classical* (93% F1-score) y *Pop* (84% F1-score) mostraron buenos resultados durante el entrenamiento, mientras que géneros como *Reggae* (40% F1-score) y *Rock* (36% F1-score) presentaron un rendimiento inferior. Este desempeño sugiere que la representación insuficiente de algunos géneros en el dataset y las dificultades inherentes para capturar las complejidades de ciertos estilos musicales contribuyen a los errores observados.

Para mejorar el modelo, es necesario realizar ajustes en el preprocesamiento de datos y en la selección de características, así como explorar métodos de balanceo del dataset, especialmente para géneros menos representados como *Rock*. Además, el uso de modelos más complejos, como redes neuronales profundas, podría permitir al modelo aprender de manera más efectiva las características únicas de los géneros musicales y mejorar la precisión de la clasificación, reduciendo los errores en la clasificación de canciones dentro de géneros similares o complejos.

4. Impacto

Este software demuestra la aplicación práctica de técnicas de machine learning para resolver problemas del mundo real en el dominio de la recuperación de información musical. Sirve como una herramienta educativa para comprender la intersección del procesamiento de señales y el machine learning, al tiempo que proporciona conocimientos sobre los desafíos asociados con la clasificación de géneros musicales.

Además, este proyecto puede inspirar a otros a explorar la clasificación automática de música y expandir las capacidades de los sistemas de recomendación, mejorando la experiencia del usuario al interactuar con plataformas de música. También puede ser útil en el desarrollo de tecnologías que permitan a los músicos analizar y organizar sus propias composiciones, abriendo nuevas posibilidades en la música digital.

5. Conclusiones

La implementación de la clasificación de géneros musicales utilizando el dataset GTZAN resalta tanto las capacidades como las limitaciones de los algoritmos de deep learning en el análisis de datos de audio. Aunque las redes neuronales ofrecen un rendimiento prometedor, la clasificación de géneros aún puede mejorarse mediante ajustes en el modelo y en el preprocesamiento de los datos.

El trabajo futuro podría centrarse en la optimización de las redes neuronales empleadas, explorando arquitecturas más complejas como redes neuronales convolucionales (CNN) o recurrentes (RNN), para mejorar la capacidad del modelo en la detección de patrones más sutiles. Además, se podría investigar el uso de características acústicas adicionales y métodos de preprocesamiento avanzados para optimizar aún más el rendimiento del sistema.

References

- [1] G. Tzanetakis, P. Cook, "Musical Genre Classification of Audio Signals," IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293-302, July 2002.
- [2] Olteanu, A. "GTZAN Dataset - Music Genre Classification", 2019. Kaggle. Available: <https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification?resource=download-directory>.