Análitica de Datos: Proyecto Final
Sistema de Clasificación Musical

Moisés David Quintero
Analítica de Datos

Pontificia Universidad Javeriana
Bogotá, Colombia
moisesd.quintero@javeriana.edu.co

Gabriel Alejandro Martín
Analítica de Datos ga_martin@javeriana.edu.co
Pontificia Universidad Javeriana
Bogotá, Colombia
ga_martin@javeriana.edu.co

Resumen: El presente proyecto se enfoca en la clasificación de géneros musicales mediante el uso de varios modelos de aprendizaje automático: K-Nearest Neighbors (KNN), redes neuronales profundas (DNN) y Random Forest. Se utilizará el dataset GTZAN, que contiene fragmentos de audio etiquetados en diez géneros diferentes: Blues, Classical (Clásica), Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae y Rock. A través del preprocesamiento de las canciones, se extraerán características relevantes que alimentarán los modelos de aprendizaje. El proyecto evaluará la precisión y eficiencia de cada modelo, buscando mejorar la categorización automática de música y establecer una base para sistemas de recomendación más avanzados.

Palabras Claves: Clasificación de Géneros Musicales, Aprendizaje Automático, K-Nearest Neighbors (KNN), Redes Neuronales Profundas (DNN), Random Forest, Dataset GTZAN, Procesamiento de Señales de Audio, Extracción de Características, Espectrogramas, Coeficientes Cepstrales de Frecuencia Mel (MFCC), Centroide Espectral, Ancho de Banda Espectral, Tempo (BPM), Armónicos, Percusión, Sistemas de Recomendación Musical.

Abstract: This project aims to classify music into ten genres (Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae, Rock) using machine learning models including K-Nearest Neighbors (KNN), deep neural networks (DNN), and Random Forest. Utilizing the GTZAN dataset, which consists of 1000 audio fragments labeled by genre, the audio signals will be preprocessed to extract significant features. These features will then be used to train and evaluate the performance of the different models. The goal is to enhance the accuracy and efficiency of automatic music genre classification, providing a foundation for developing sophisticated music recommendation systems.

Keywords: Music Genre Classification, Machine Learning, K-Nearest Neighbors (KNN), Deep Neural Networks (DNN), Random Forest, GTZAN Dataset, Audio Signal Processing, Feature Extraction, Spectrograms, Mel-frequency Cepstral Coefficients (MFCC), Spectral Centroid, Spectral Bandwidth, Tempo (BPM), Harmonics, Percussion, Music Recommendation Systems.

I. Introducción

La música es una de las formas de multimedia más populares y ha jugado un papel fundamental en la sociedad a lo largo de la historia. Con el crecimiento explosivo de los servicios de transmisión y descarga de música en línea y la disponibilidad de almacenamiento a bajo costo, nuestra manera de escuchar y consumir música ha cambiado drásticamente. La música no solo conecta a personas con intereses similares, sino que también actúa como un vínculo que une a grupos y comunidades. La revolución digital ha transformado la industria musical, facilitando el acceso a millones de canciones a través de plataformas como Spotify, SoundCloud, y Apple Music, lo que ha creado una base de consumidores más amplia y diversa.

Para navegar, buscar y organizar la inmensa cantidad de información musical de manera eficiente, se han desarrollado numerosos sistemas de recuperación de información musical basados en contenido (MIR, por sus siglas en inglés). Estos sistemas aplican técnicas de procesamiento de señales y aprendizaje automático para extraer información semántica significativa de las señales musicales. Aunque la investigación musical es relativamente nueva en comparación con los campos maduros del habla y la imagen/video, en los últimos años se han desarrollado algoritmos innovadores que explotan las propiedades particulares de la música, como sus estructuras armónicas y rítmicas.

Una pieza musical puede descomponerse en señales de audio que contienen información rica y compleja. Esta descomposición implica el análisis de características tanto en el dominio del tiempo como en el dominio de la frecuencia de la señal de audio. Métodos tradicionales como los coeficientes de frecuencia mel (MFCC) han sido utilizados para caracterizar texturas y ritmos, logrando un cierto grado de éxito en tareas de clasificación de géneros musicales. Sin embargo, la clasificación de géneros musicales sigue siendo un desafío debido a los límites difusos entre diferentes géneros y la variabilidad inherente de las señales de audio.

El presente estudio explora la aplicación de algoritmos de aprendizaje profundo para la identificación y clasificación de

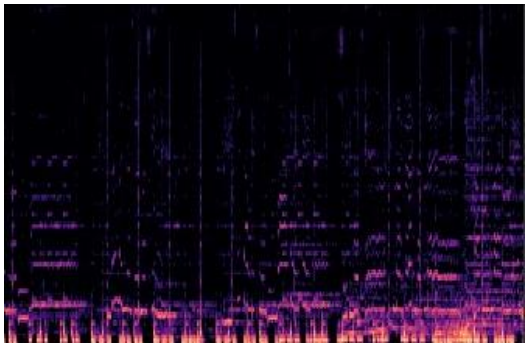
géneros musicales. Utilizando el espectro de la señal de audio, se entrenan modelos complejos de aprendizaje de máquina para categorizar automáticamente las canciones. Este enfoque no solo mejora la precisión de la clasificación, sino que también proporciona una base para desarrollar sistemas de recomendación musical más sofisticados.

II. Hipótesis

¿Es posible descomponer una canción en diferentes componentes y encontrar patrones consistentes entre canciones de un mismo género musical utilizando técnicas avanzadas de procesamiento de señales y modelos de aprendizaje automático?

III. Marco Teórico

A. *Espectograma:*



El espectrograma es una representación visual que muestra cómo varía la intensidad de las diferentes frecuencias de sonido a lo largo del tiempo en un archivo de audio [1]. En esencia, es como una imagen en la que el eje horizontal representa el tiempo y el eje vertical representa la frecuencia. Los colores o tonos en el espectrograma indican la intensidad o amplitud de las frecuencias en diferentes momentos. Por ejemplo, las regiones más brillantes u oscuras pueden indicar frecuencias más fuertes o más débiles respectivamente. Esta representación visual es útil porque permite observar patrones y características específicas del sonido que pueden ser difíciles de percibir auditivamente. En el contexto del procesamiento de datos para el aprendizaje automático, convertir archivos de audio en espectrogramas es fundamental, ya que proporciona una forma estructurada y comprensible para que los algoritmos de inteligencia artificial analicen y aprendan de los datos de audio. Específicamente, para el dataset en cuestión, la conversión a espectrogramas de Mel permite que los archivos de audio sean procesados por redes neuronales convolucionales (CNN), que son modelos efectivos para la clasificación de imágenes y, en este caso, de espectrogramas.

B. *Forma de Onda*

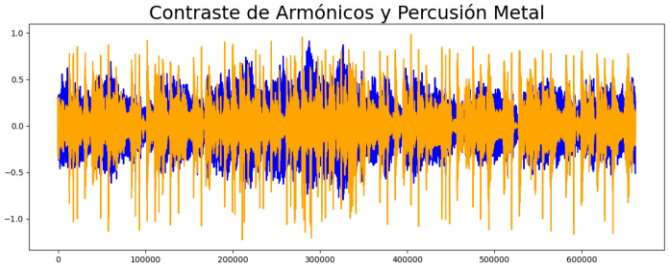
La forma de onda en el procesamiento de audio se refiere a la representación gráfica de la amplitud de la señal de audio en función del tiempo. Denota cómo varía la presión del sonido en el tiempo, con el eje horizontal representando el tiempo y el eje vertical representando la amplitud de la señal. Llega a ser fundamental en el procesamiento de audio al proporcionar información visual sobre la señal de audio.

C. *BPM*

El tempo, representado comúnmente como BPM (beats per minute), es una medida esencial en música que define la velocidad o ritmo de una composición. Este ritmo, expresado en pulsos por minuto, varía significativamente entre diferentes géneros musicales, y su percusión en ondas y frecuencias juega un papel crucial en la identidad sonora de cada estilo. En el rock y el pop, la batería marca un ritmo enérgico y prominente, mientras que en el jazz, la percusión es más compleja y flexible, permitiendo la improvisación y la interacción entre músicos.

D. *Percusión y armónicos*

La percusión y los armónicos son componentes fundamentales en el análisis de señales de audio y juegan un papel crucial en la caracterización musical de una pieza [2]. La percusión se refiere a los elementos rítmicos que proporcionan el pulso de la música, mientras que los armónicos son componentes de frecuencia que enriquecen el timbre de los sonidos. Librerías como librosa, ampliamente utilizadas en el procesamiento de audio, facilitan la extracción y el análisis de estas características mediante algoritmos avanzados. Estos algoritmos incluyen la normalización a media cero y desviación estándar unitaria, y la agregación de características en ventanas de clasificación, útiles para entrenar modelos de aprendizaje automático. Las características continuas originales pueden discretizarse utilizando análisis percentil o histogramas, simplificando así el tratamiento de los datos.



Librosa permite la extracción de características como la percusión y los armónicos utilizando una serie de técnicas y transformaciones de señal[3]. Para la percusión, se emplean métodos de detección de eventos de inicio (onset detection), que identifican los momentos en los que comienzan los golpes

rítmicos en la señal de audio. Esto se logra mediante la aplicación de transformadas de Fourier de corta duración (STFT) para convertir la señal en el dominio del tiempo en una representación en el dominio de la frecuencia, donde los picos correspondientes a los eventos percutivos son más fácilmente identificables. Los armónicos, por otro lado, se extraen analizando el contenido espectral de la señal. Librosa utiliza la transformada de Fourier y la transformada de Fourier de tiempo corto (STFT) para descomponer la señal en sus componentes de frecuencia, permitiendo la identificación y el análisis de los armónicos y su evolución a lo largo del tiempo.

IV. DATASET (GTZAN)

A. Contextualización

El dataset GTZAN es uno de los más conocidos y utilizados en la investigación de la recuperación de información musical (MIR), especialmente en la tarea de reconocimiento automático de géneros musicales [4]. Fue creado por Tzanetakis y Cook en 2002 y contiene 1000 fragmentos de música, cada uno con una duración de 30 segundos, categorizados en 10 géneros diferentes: Blues, Classical (Clásica), Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae y Rock. Cada fragmento de audio tiene una frecuencia de muestreo de 22,050 Hz y está en formato WAV.

B. Limitaciones

1) Repetición de Fragmentos

El dataset presenta una significativa limitación en términos de repetición de fragmentos. Estas repeticiones pueden ocurrir en diversos niveles de especificidad: fragmentos exactamente iguales, fragmentos de la misma grabación, diferentes versiones de la misma canción o incluso diferentes canciones del mismo artista. Esta repetición puede llevar a que los modelos de aprendizaje automático aprendan patrones específicos de un artista o una grabación particular, en lugar de características generales del género musical. Esto afecta la capacidad de generalización de los modelos, reduciendo su eficacia en la clasificación precisa de géneros musicales.

2) Etiquetado Incorrecto

Otra limitación crucial del dataset es el etiquetado incorrecto de los fragmentos de audio. Existen numerosos casos donde el género etiquetado no corresponde claramente con las características musicológicas del fragmento, como la composición, instrumentación, métrica, ritmo y otros. Además, algunos etiquetados son contenciosos, es decir, aunque el fragmento podría encajar en el género etiquetado, no lo hace completamente según los criterios musicológicos. Por ejemplo, un fragmento etiquetado como Hip Hop que en su mayoría contiene una muestra de música cubana. Estos problemas de etiquetado pueden llevar a los modelos de clasificación a aprender asociaciones incorrectas, disminuyendo la precisión de la clasificación.

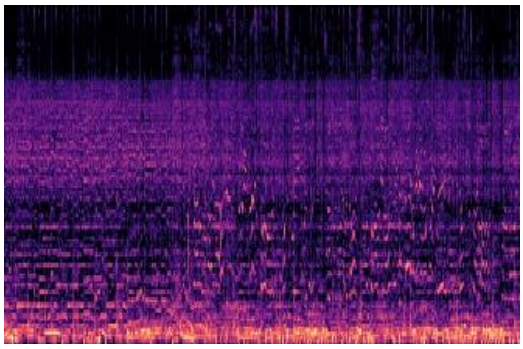
3) Distorsión de audio

La distorsión en los fragmentos de audio es otra limitación significativa del dataset GTZAN. Algunos fragmentos contienen distorsiones notables, como estática, clipping digital y saltos, que afectan su calidad y utilidad para el análisis. En casos extremos, la distorsión puede hacer que un fragmento sea prácticamente inutilizable para el análisis musical. Estas distorsiones no solo comprometen la calidad del audio, sino que también pueden afectar la extracción de características y la precisión de los modelos de clasificación de géneros, ya que el ruido y las alteraciones pueden ser mal interpretados por los algoritmos de aprendizaje automático.

C. Preprocesamiento (espectrogramas)

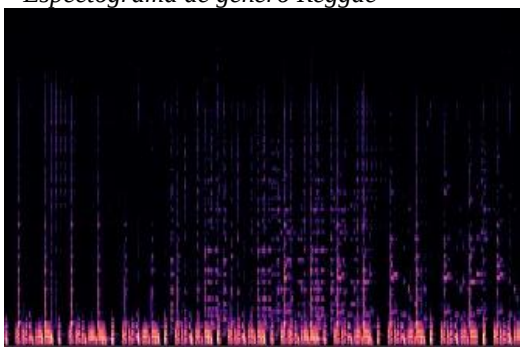
Las canciones del dataset han presentado un preprocesamiento que encontramos de gran utilidad para comprender el manejo de las canciones, como son los espectogramas. Gracias a estos podemos entender ciertos patrones o características propias de cada género. Podemos encontrarnos una carpeta donde se encuentran estas representaciones gráficas por cada canción de todos los géneros.

1) Espectograma de género Metal



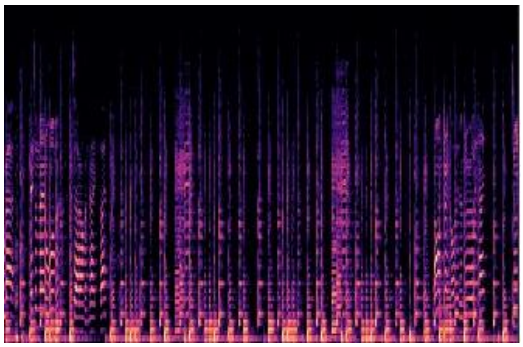
Predominio de frecuencias bajas, armónicos prominentes manifestados como bandas verticales, y una textura granulada indicando ruido debido a la distorsión y amplificación intensa.

2) Espectograma de género Reggae



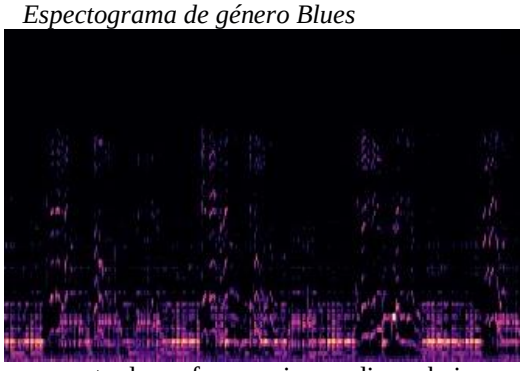
Predominio de frecuencias medias, ritmo constante representado por líneas horizontales, y patrones repetitivos de acordes y melodías.

3) Espectograma de género Rock



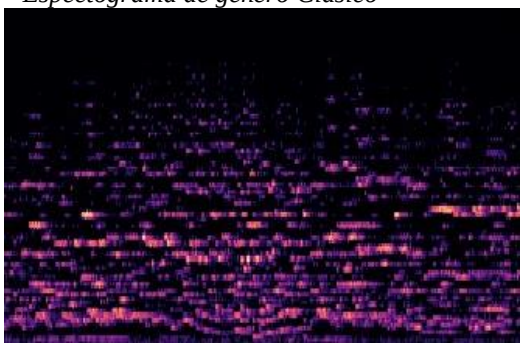
Energía en frecuencias medias y bajas, armónicos prominentes visibles como bandas verticales, y una textura granulada que indica un nivel moderado de ruido.

4) *Espectograma de género Blues*



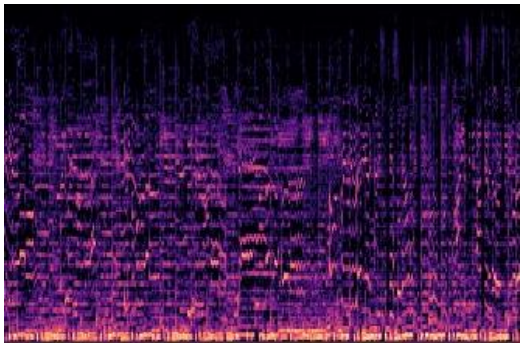
Energía concentrada en frecuencias medias y bajas, pequeñas variaciones de frecuencia indicando notas "blues", y una estructura repetitiva reflejada como bloques de color.

5) *Espectograma de género Clásico*



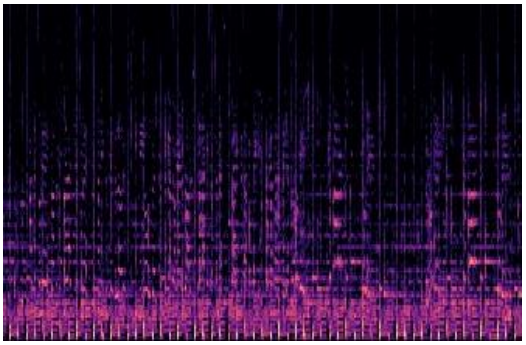
Amplio rango de frecuencias, patrones complejos de color indicando melodías intrincadas, y capas superpuestas que reflejan armonías ricas.

6) *Espectrograma de género Country*



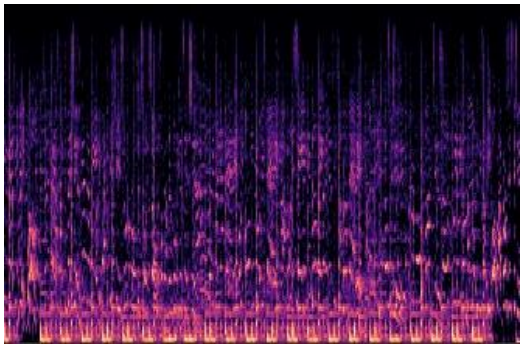
Predominio de frecuencias medias, líneas de color delgadas y espaciadas por la instrumentación simple, y bloques de color repetitivos que muestran la estructura repetitiva.

7) *Espectograma de género Disco*



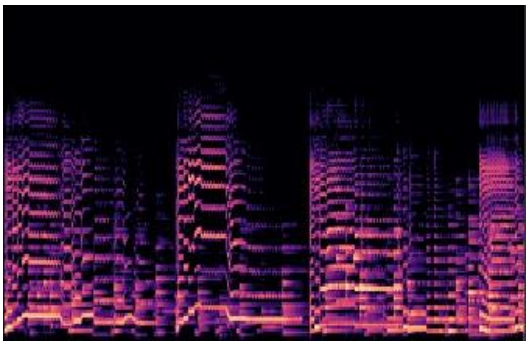
Energía en frecuencias medias y altas, líneas horizontales constantes que indican un ritmo marcado, y bloques de color repetitivos que reflejan patrones de acordes y melodías.

8) *Espectograma de género HipHop*



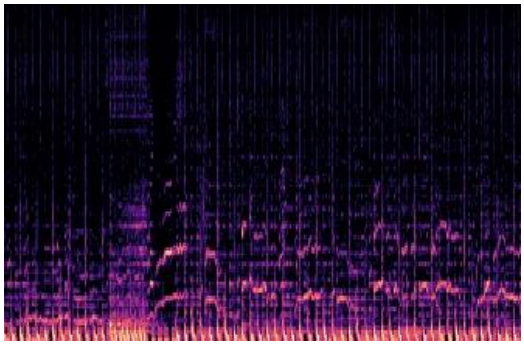
Predominio de frecuencias bajas, líneas horizontales que muestran un ritmo constante, y manchas de color extendidas que representan voces habladas o rapeadas.

9) *Espectograma de Jazz*



Amplio rango de frecuencias, patrones intrincados de color que cambian constantemente reflejando improvisación, y capas superpuestas que indican armonías complejas.

10) Espectograma de Pop



Predominio de frecuencias medias, bloques de color repetitivos que muestran la estructura de la canción, y patrones de color simples y repetitivos que reflejan melodías pegadizas.

D. Preprocesamiento (Características extraídas)

En el dataset también se encuentran presentes diferentes características que fueron extraídas con funcionalidades de la librería Librosa. Las características que llegan a resaltar en mayor medida son:

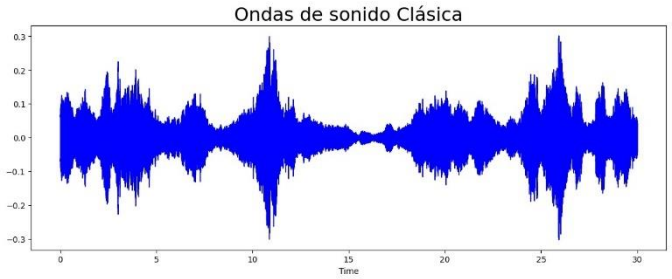
- length: La duración del archivo de audio en milisegundos.
- chroma_stft_mean: La media del espectrograma cromático calculado usando la Transformada de Fourier de Tiempo Corto (STFT). Esta característica representa la intensidad de las 12 diferentes clases de tonos (do, do#, re, etc.) en la pista de audio.
- chroma_stft_var: La varianza del espectrograma cromático. Esto indica la variabilidad en la intensidad de los tonos a lo largo de la pista.
- rms_mean: La media del valor cuadrático medio (RMS) de la amplitud de la señal de audio, que representa la energía promedio de la señal.
- rms_var: La varianza del RMS, indicando la variabilidad de la energía a lo largo de la pista.
- spectral_centroid_mean: La media del centroide espectral, que indica el "centro de masa" del espectro de frecuencias y se asocia con la percepción del brillo del sonido.
- spectral_centroid_var: La varianza del centroide espectral, mostrando cuán variable es el brillo del sonido.
- spectral_bandwidth_mean: La media del ancho de banda espectral, que mide la extensión de las frecuencias en la señal y se relaciona con la percepción de la riqueza del sonido.
- spectral_bandwidth_var: La varianza del ancho de banda espectral, indicando la variabilidad en la extensión de las frecuencias.
- rolloff_mean: La media del roll-off espectral, que es la frecuencia por debajo de la cual se encuentra un porcentaje específico del total de la energía espectral (normalmente el 85%).
- rolloff_var: La varianza del roll-off espectral, mostrando la variabilidad de esta medida a lo largo de la pista.
- zero_crossing_rate_mean: La media de la tasa de cruces por cero, que cuenta cuántas veces la señal de audio cruza el eje cero. Esta característica es un indicador de la actividad y la brusquedad de la señal.
- zero_crossing_rate_var: La varianza de la tasa de cruces por cero, indicando la variabilidad en la brusquedad de la señal.
- harmony_mean: La media de la armonía, que captura las propiedades armónicas de la señal de audio.
- harmony_var: La varianza de la armonía, mostrando cuán variable son las propiedades armónicas en la pista.
- perceptr_mean: La media de la percepción tonal, que mide la cantidad de información tonal presente en la señal.
- perceptr_var: La varianza de la percepción tonal, indicando la variabilidad de la información tonal.
- tempo: El tempo de la pista en beats per minute (BPM), que es una medida del ritmo de la música.

V. PROCEDIMIENTO

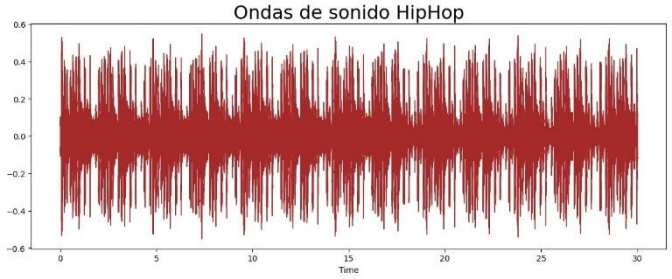
A. Graficación de Formas de Onda

Iniciando el proyecto, decidimos entender a profundidad las características esenciales de los datos con los que estábamos trabajando a través de las formas de onda. Al observar estas representaciones gráficas, esperamos descubrir similitudes y diferencias que nos ayuden a comprender mejor las características únicas de cada género y, quizás, a encontrar patrones subyacentes que definan su identidad sonora.

La forma de onda de música clásica revela un patrón sinusoidal distintivo, típicamente asociado con la pureza y la claridad del sonido. Con una frecuencia fija y una amplitud estable, esta forma de onda refleja la precisión y la belleza que caracterizan a la música clásica. Su duración corta sugiere una ejecución meticulosa y una atención cuidadosa a cada nota. Es un reflejo de la tradición y la sofisticación de la música clásica, donde la excelencia técnica se valora profundamente.



Por otro lado, la forma de onda del hip hop es un estudio de contrastes y dinamismo. Con frecuencia y amplitud fluctuantes, esta forma de onda es un reflejo de la complejidad y la vitalidad de este género musical. Su estructura no sinusoidal y la presencia de armónicos revelan la riqueza y la profundidad de los sonidos utilizados en el hip hop, que van desde samples hasta ritmos percusivos. La duración variable de la onda sugiere una narrativa en constante evolución y una exploración continua de nuevos ritmos y texturas.

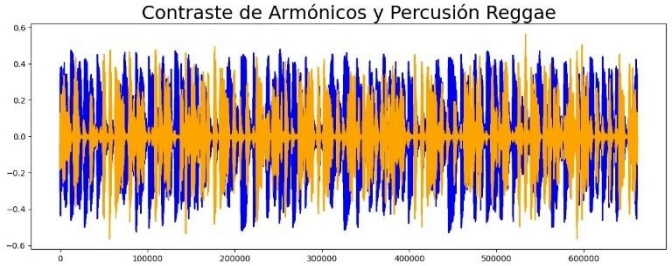


Desde un punto inicial del proyecto, nos encontramos con claros patrones o características que se logran computarizar y diferenciar entre sí.

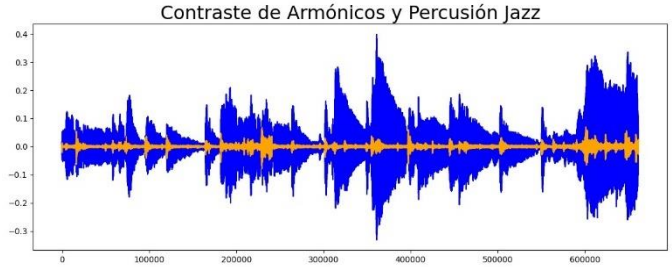
B. Graficación de Armonías y Percusiones

Lo que buscamos ahora, es extraer otros componentes propios de la música: las armonías y las percusiones. Comparando ahora los géneros de Reggae y Jazz, podemos llegar a ciertas conclusiones.

El reggae es un género musical que se caracteriza por su distintivo ritmo de percusión, marcado por el uso prominente del "skank", un patrón rítmico de guitarra rítmica que enfatiza los tiempos 2 y 4 del compás. Esta percusión distintiva, combinada con líneas de bajo profundas y pulsantes, establece el fundamento rítmico del reggae. Además, los acordes simples y repetitivos de la guitarra eléctrica o teclado proporcionan la armonía básica que sirve como lienzo para la expresión vocal y la instrumentación melódica.

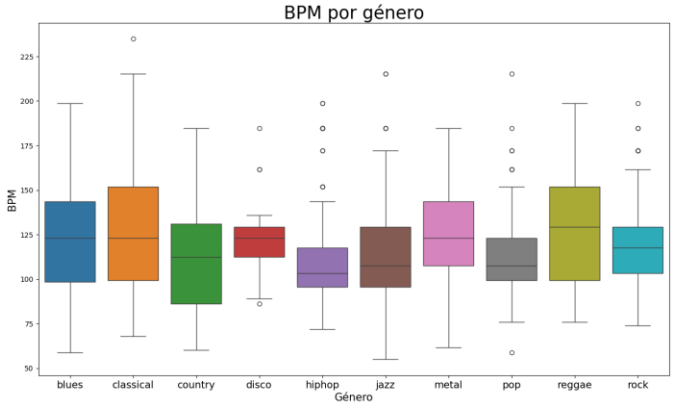


Si bien la percusión sigue siendo un componente vital en el jazz, a menudo toma un segundo plano en comparación con otros géneros, como el reggae. En lugar de enfocarse en ritmos marcados y repetitivos, la percusión en el jazz suele adoptar un papel más sutil y de apoyo, proporcionando una estructura rítmica flexible sobre la cual los músicos pueden construir sus improvisaciones.



C. BPM

Siguiendo en la profundización de los datos, decidimos ver cómo llegan a variar los beats por minuto (BPM) de cada género del dataset. A través de la mediana, la caja (rango intercuartílico), los bigotes y los outliers, se revelan tendencias significativas.



Por ejemplo, géneros como "blues" y "jazz" muestran una amplia variabilidad en los BPM, mientras que "disco" y "country" presentan una dispersión más estrecha. Además, se observan numerosos outliers en géneros como "classical", "hiphop" y "pop", indicando canciones con BPM atípicamente altos o bajos. Podemos encontrar nuevamente patrones, además de la tendencia de ciertos géneros a valores específicos de BPM (Como es el caso del disco, que su rango de en esta característica se encuentra muy estrecha y rondando los 120, aproximadamente).

D. Entrenamiento de Modelos

El código realiza varias etapas para preparar los datos y evaluar diferentes modelos de predicción. A continuación, se describe el proceso paso a paso:

Lectura de datos:

- Se carga el archivo CSV ubicado en la ruta especificada en la variable `path` utilizando la función `pd.read_csv`.

Eliminación de columnas innecesarias:

- Se elimina la columna 'length' que representa la duración de las grabaciones, ya que no es relevante para la predicción.
- Se eliminan las primeras columnas innecesarias utilizando `iloc`.

Separación de características y etiquetas:

- `y` se define como la columna 'label', que contiene las etiquetas (clases) de las muestras.
- `X` se define como todas las columnas excepto 'label', que contienen las características (features) de las muestras.

Normalización de datos:

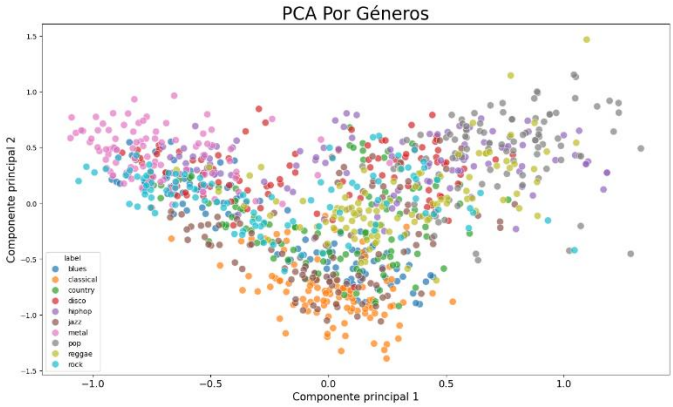
- Se crea un objeto `MinMaxScaler` para escalar las características entre 0 y 1.
- Se ajustan y transforman las características con el escalador.
- Los datos escalados se convierten nuevamente en un DataFrame para facilitar su uso posterior.

División de los datos en conjuntos de entrenamiento y prueba:

- Se divide el conjunto de datos en entrenamiento (70%) y prueba (30%) utilizando la función `train_test_split`, asegurando reproducibilidad con un `random_state` fijo.

E. PCA

Para el análisis por PCA genera una gráfica de dispersión en la que cada punto representa una instancia de datos en un espacio bidimensional definido por los dos componentes principales obtenidos. Esta visualización permite observar cómo las instancias se distribuyen en función de la variabilidad capturada por estas dos dimensiones principales. Además, la gráfica utiliza colores diferentes para distinguir entre las diferentes clases o categorías de los datos, lo que facilita la identificación de patrones o agrupaciones en el conjunto de datos.
La utilidad principal de esta gráfica radica en proporcionar una representación visual de alta dimensión de los datos, lo que facilita la interpretación y comprensión de su estructura subyacente. Al observar la distribución de las clases en el espacio bidimensional, se pueden identificar relaciones, agrupaciones o separaciones entre ellas, lo que puede ser útil para análisis exploratorios, detección de patrones y toma de decisiones en tareas de clasificación o agrupación. Además, al utilizar PCA para reducir la dimensionalidad, se preserva la mayor parte de la variabilidad de los datos, lo que garantiza que la visualización refleje de manera significativa la estructura original del conjunto de datos.



Acá podemos encontrar ciertas relaciones, como puede ser el clúster de rosa (que le pertenece a la etiqueta de 'Metal') y el clúster azul celeste (que le pertenece a la etiqueta de 'Rock'): presentan una mayor concentración en el costado izquierdo y se mezclan entre sí, lo cual se puede deber a su similitud de ritmos e instrumentos.
Además, también se logra presenciar una clara concentración y mezcla de dos clústeres, naranja y marrón en la zona inferior (música clásica y jazz, respectivamente). Esto se puede deber a la complejidad musical que ambos géneros presentan.

F. Prueba de Modelos

Evaluación de diferentes modelos:

Definición de una función para probar modelos:

- La función `probar` toma un modelo y un título como argumentos.
- Ajusta el modelo con los datos de entrenamiento.
- Realiza predicciones sobre los datos de prueba.
- Calcula y muestra la precisión del modelo.
- Devuelve el modelo entrenado, las predicciones y la precisión.

KNN (K-Nearest Neighbors):

- Se crea un clasificador KNN con 19 vecinos.
- Se entrena y evalúa el modelo usando la función `probar`.
- Se grafica la matriz de confusión de las predicciones.

Random Forest:

- Se crea un clasificador Random Forest con 1000 árboles y una profundidad máxima de 10.
- Se entrena y evalúa el modelo usando la función `probar`.
- Se grafica la matriz de confusión de las predicciones.

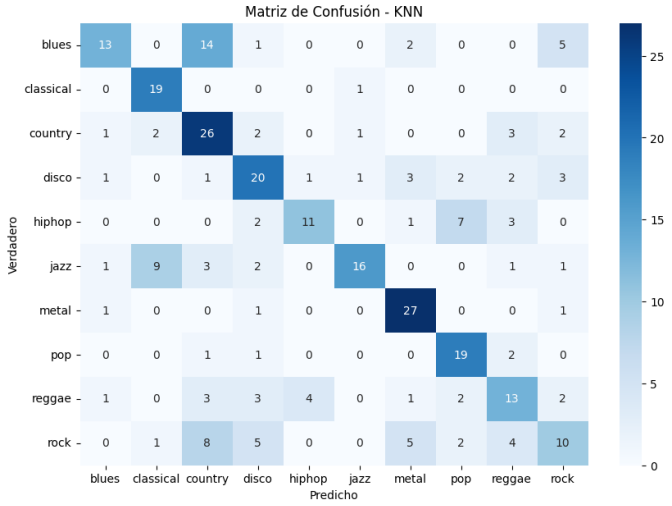
Red Neuronal (Neural Network):

- Se crea un clasificador MLP (Perceptrón Multicapa) con una arquitectura específica.
- Se entrena y evalúa el modelo usando la función `probar`.

- Se grafica la matriz de confusión de las predicciones.

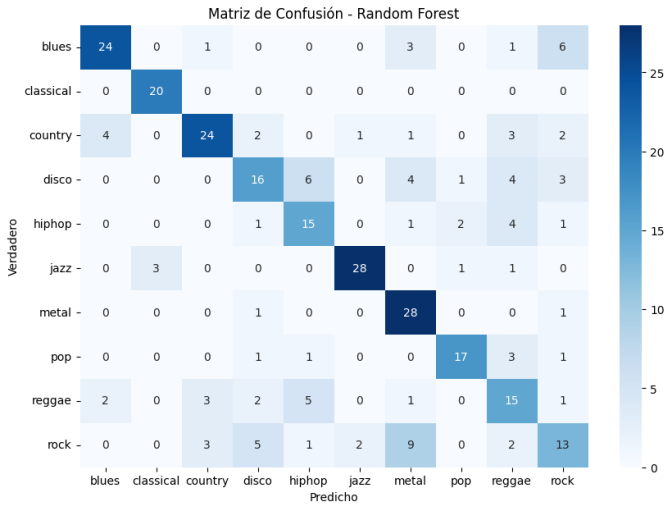
G. Resultados

1) KNN



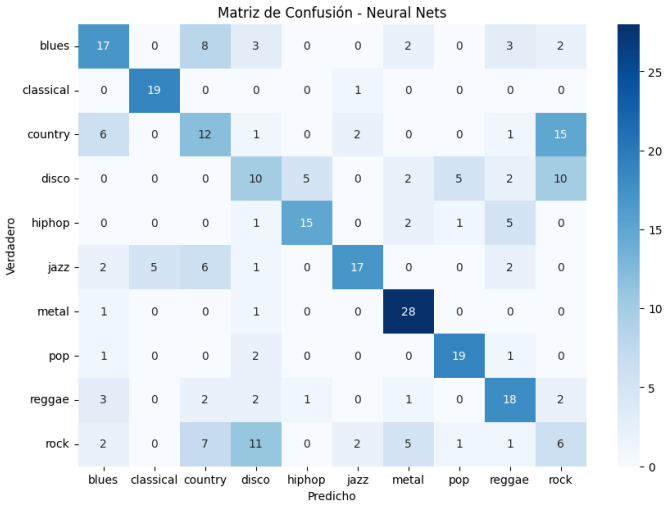
La matriz de confusión proporciona una visión detallada del rendimiento del modelo KNN en la clasificación de géneros musicales, donde la precisión general fue de 0.58. Se destaca la eficacia en la clasificación de géneros como "Country" y "Metal", con valores notablemente altos, mientras que "Disco" también obtuvo resultados satisfactorios. Sin embargo, se observa una confusión significativa entre "Blues" y "Country", lo que sugiere una dificultad para distinguir entre estos dos géneros. Los valores en la diagonal principal indican las instancias correctamente clasificadas, mientras que los errores fuera de esta diagonal señalan las clasificaciones incorrectas. Por ejemplo, se identifican correctamente 13 instancias de "Blues", pero 14 de ellas fueron erróneamente clasificadas como "Country".

2) Random Forest



El modelo de Random Forest empleado ha demostrado un rendimiento notable en la clasificación de diversos géneros musicales. Entre las clases, "Jazz" y "Metal" son las que mejor clasifica, con una alta cantidad de predicciones correctas (28) y relativamente pocas confusiones con otras clases. Por otro lado, el modelo tiende a confundir "Metal" y "Rock", como se observa en la matriz de confusión, donde algunas instancias de estos géneros son incorrectamente clasificadas como el otro. A pesar de estas confusiones, la precisión general del modelo sigue siendo significativa, alcanzando el 66.67%.

3) Redes Neuronales



Por último, el modelo de redes neuronales se queda atrás con un menor rendimiento de clasificación. La clase con mejor puntaje resulta siendo el de “Metal” con 28 predicciones correctas. Las otras clases, generalizando un poco, tienden a tener múltiples confusiones entre sí (Como puede ser el caso de “Country”, que no logra diferenciarse exitosamente de “Rock”)

H. Modelo Final

Después de analizar detenidamente el rendimiento de diferentes modelos de clasificación, se ha llegado a la conclusión de que el modelo final y seleccionado es el de Random Forest. Este modelo ha demostrado siempre los mejores resultados en cuanto a precisión y capacidad para clasificar varios géneros musicales. Destacan los excelentes resultados obtenidos en géneros como Jazz, Metal, Blues, Clásico y Country.

Podemos encontrar diferentes razones por las que este modelo llegó a ser el ganador. Primero, este modelo muestra una notable capacidad para manejar complejas características de la música, como el tempo, la tonalidad y la instrumentación, gracias a su habilidad para capturar relaciones no lineales entre ellas. Además, su robustez frente a datos desbalanceados o ruidosos facilita la generalización y evita el sobreajuste, común en conjuntos de datos musicales. La estrategia de utilizar múltiples árboles de decisión entrenados en subconjuntos de datos diferentes reduce aún más el riesgo de sobreajuste, lo que mejora la capacidad del modelo para generalizar a nuevos datos.

I. Peso de los atributos

En el proceso de evaluación de atributos para determinar su importancia en el modelo de Random Forest, se identificaron cuatro variables como las más influyentes en la predicción del resultado deseado. La selección de estas variables se basó en su contribución significativa a la precisión del modelo y su capacidad para discriminar entre las clases objetivo.

Weight	Feature
0.0247 ± 0.0205	chroma_stft_mean
0.0193 ± 0.0098	perceptr_var
0.0187 ± 0.0090	chroma_stft_var
0.0107 ± 0.0098	mfcc8_var
0.0093 ± 0.0122	spectral_centroid_var
0.0093 ± 0.0115	mfcc1_mean
0.0073 ± 0.0050	zero_crossing_rate_var
0.0073 ± 0.0171	mfcc6_var
0.0067 ± 0.0073	rms_var
0.0060 ± 0.0176	mfcc6_mean
0.0060 ± 0.0098	mfcc4_var
0.0060 ± 0.0088	mfcc20_var
0.0053 ± 0.0161	mfcc7_var
0.0047 ± 0.0090	mfcc4_mean
0.0047 ± 0.0080	mfcc3_mean
0.0040 ± 0.0115	zero_crossing_rate_mean
0.0033 ± 0.0169	mfcc5_var
0.0033 ± 0.0042	mfcc7_mean
0.0033 ± 0.0060	mfcc5_mean
0.0033 ± 0.0042	mfcc16_var
... 37 more ...	

"Chroma_stft_mean" se destacó por su habilidad para capturar la intensidad de las diferentes clases de tonos presentes en la pista de audio, siendo crucial para discernir la composición tonal. "Perceptr_var" emergió debido a su capacidad para reflejar la variabilidad en la información tonal, ofreciendo una visión de la diversidad tonal en la pista. "Chroma_stft_var" fue seleccionada por su capacidad para representar la variabilidad en la intensidad de los tonos, mientras que "spectral_centroid_var" fue considerada relevante por mostrar la variabilidad en el brillo del sonido.

J. Encontrar Similitudes (Similitud de coseno)

Métrica utilizada para medir cuán similares son dos vectores en un espacio multidimensional. Es útil en el procesamiento del lenguaje natural, la recuperación de información y en sistemas de recomendación. La similitud del coseno se basa en el cálculo del coseno del ángulo entre dos vectores.

Esta se utilizó para generar una tabla de similitudes que nos permite encontrar canciones parecidas entre sí dado el análisis realizado en el proyecto.

	0	1	2	3	4	5	6	7	8	9	...	990	991
0	1.000000	0.049231	0.589618	0.284862	0.025561	-0.346688	-0.219483	-0.167626	0.641877	-0.097889	...	-0.082829	0.546169
1	0.049231	1.000000	-0.096834	0.520903	0.080749	0.307856	0.318286	0.415258	0.120649	0.404168	...	-0.098111	-0.325126
2	0.589618	-0.096834	1.000000	0.210411	0.400266	-0.082019	-0.028061	0.104446	0.468113	-0.132532	...	-0.032408	0.561074
3	0.284862	0.520903	0.210411	1.000000	0.126437	0.134796	0.300746	0.324566	0.352758	0.295184	...	-0.320107	-0.206516
4	0.025561	0.080749	0.400266	0.126437	1.000000	0.556066	0.482195	0.623455	0.029703	0.471657	...	0.087605	0.017366

K. Implementación

Finalmente, a través de la matriz de similitudes generada anteriormente podemos encontrar, dada una canción a buscar, las diferentes canciones que son más parecidas a esta.

VI. Conclusiones

- **Identificación géneros**: El análisis detallado de características como los espectrogramas, armónicos, percusiones y el tempo ha revelado patrones distintivos asociados con diferentes géneros musicales. Estos patrones no solo proporcionan información sobre la estructura y la instrumentación de la música, sino que también ayudan a definir la identidad sonora única de cada género.
- **Rendimiento y comparación de modelos**: Durante la ejecución del proyecto, llevamos cabo el entrenamiento de 3 modelos de aprendizaje automático: K-Nearest Neighbors (KNN), Random Forest y redes neuronales. Con el dataset trabajado y el diseño elaborado, logramos concluir que Random Forest ha demostrado consistentemente la capacidad más alta para clasificar con precisión una amplia gama de géneros (con una precisión general de 66.67%).
- **Importancia de la analítica:** La importancia de la información y la relación de datos no se limita únicamente a la industria musical (tratando de generar más ventas, segregar público o planear giras), sino que tiene una relevancia inmensa en toda la era contemporánea, donde todo lo que está ocurriendo se encuentra documentado en una base de datos. Estamos en la era de la información, donde los datos y sus parones se pueden traducir en decisiones de valor y poder.

Referencias

[1] «Music Genres Classification using Deep learning techniques». Accedido: 20 de mayo de 2024. [En línea]. Disponible en: https://www.analyticsvidhya.com/blog/2021/06/music-genres-classification-using-deep-learning-techniques/

[2] I. Vatolkin, P. Ginsel, y G. Rudolph, «Advancements in the Music Information Retrieval Framework AMUSE over the Last Decade», en *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Virtual Event Canada: ACM, jul. 2021, pp. 2383-2389. doi: 10.1145/3404835.3463252.

[3] «librosa — librosa 0.10.2 documentation». Accedido: 20 de mayo de 2024. [En línea]. Disponible en: https://librosa.org/doc/latest/index.html

[4] B. L. Sturm, «An analysis of the GTZAN music genre dataset», en *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies*, Nara Japan: ACM, nov. 2012, pp. 7-12. doi: 10.1145/2390848.2390851.