

# Clasificación jerárquica de géneros musicales

---

Rodolfo Figueroa Soriano

3 de diciembre de 2021

Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas  
UNAM

- |                                   |                    |
|-----------------------------------|--------------------|
| 1. Motivación                     | Canciones          |
| 2. Fundamentos                    | Artistas           |
| 3. Preliminares                   | 4.2 Modelado       |
| 3.1 Objetivo                      | 5. Resultados      |
| 3.2 Alcance                       | 5.1 Jerarquías     |
| 3.3 Flujo de trabajo              | 5.2 Clasificación  |
| Construcción de jerarquías        | 5.3 Interpretación |
| Clasificación                     | 6. Conclusiones    |
| Interpretación                    | 7. Trabajo futuro  |
| 4. Desarrollo                     | 8. Referencias     |
| 4.1 Integración de la información |                    |

## Motivación

---

# Motivación

- El género de una canción encapsula muchas propiedades en una etiqueta fácil de entender.
  - Instrumentos
  - Ritmo
  - Estilo
- Es un excelente punto de partida para algoritmos de recomendación.

# Motivación

- Al día, Spotify publica 60 000 canciones nuevas.[2]
- ¡Casi 2 canciones por segundo!
- Asignarle un género a cada una manualmente requeriría de mucho tiempo y trabajo.
- Es posible que dos personas asignen a una misma canción un género diferente.

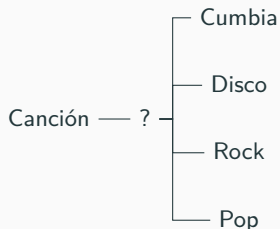
**Necesitamos una solución automatizada**

## Fundamentos

---

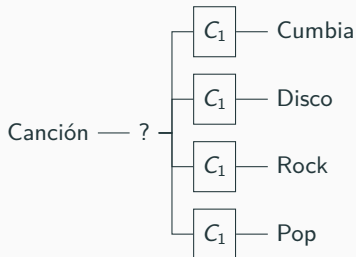
# Fundamentos

Dada una instancia de datos, se desea determinar a qué clases (cero o más) pertenece.



# Fundamentos

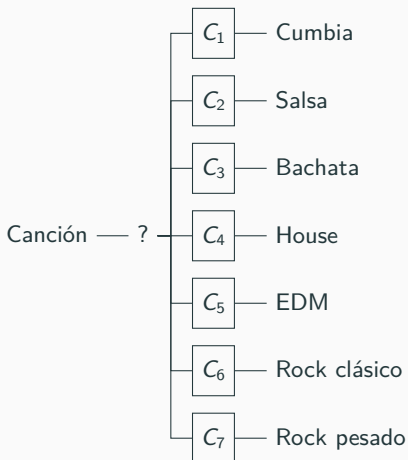
Una solución es entrenar un *clasificador binario* para cada clase:





# Fundamentos

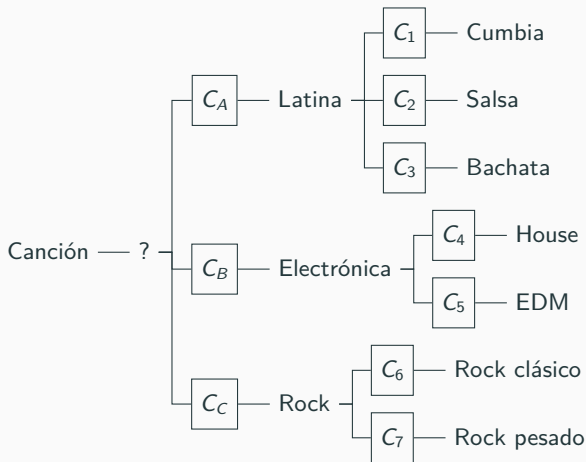
Si las clases son muchas o muy similares, el problema se vuelve más difícil.



Podemos simplificar el problema construyendo una **jerarquía de clases** [1]:

- Hacemos a cada clase hija de una **superclase**.
- Calculamos la probabilidad de pertenecer a cada clase y superclase.
- Combinamos estas probabilidades para obtener la probabilidad final de pertenecer a la clase.

# Fundamentos



## Ventajas

- Evita el desbalance entre ejemplos positivos y negativos.
- Simplifica el entrenamiento y la inferencia.
- Es fácil de interpretar.
- Puede ayudar a descubrir nuevas relaciones entre clases.

## Desventajas

- La complejidad es mayor, ya que hay más clasificadores.
- Los datos deben de tener una jerarquía explícita o latente.
- Muchas veces no es fácil encontrar dicha jerarquía.

## Preliminares

---

## Objetivo

Diseñar un sistema de reconocimiento de géneros musicales. Este debe de:

- dar buenos resultados.
- ser extendible a un número arbitrario de géneros.
- ser interpretable.

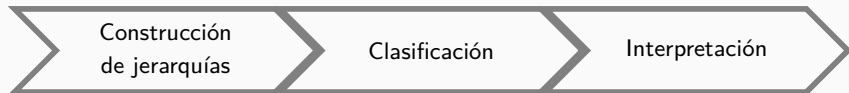
## Alcance

- Hasta donde sabemos, el reconocimiento de géneros en las plataformas mayoritarias (Spotify, Apple Music, etc.) se hace de manera manual.
- Muchas veces se hacen aproximaciones para ahorrar trabajo, por ejemplo, asignárselo al álbum o al artista en vez de a canciones individuales.
- Un sistema completamente automatizado podría producir mejores resultados, de manera más rápida y sin necesidad de intervención humana.

# Preliminares

## Flujo de trabajo

El proceso de análisis puede dividirse en tres pasos principales:





## Flujo de trabajo

- Podemos encontrar superclases agrupando los datos con algoritmos de *clustering*.
- Si hay demasiadas dimensiones, los puntos están muy separados como para encontrar *clusters* significativos.

**Solución:** Reducir el número de dimensiones.

## Flujo de trabajo

Para reducir las dimensiones, utilizamos:

### **UMAP: Uniform Manifold Approximation and Projection**

- Reducción de dimensiones **no-lineal**.
- Utiliza análisis topológico para proyectar los puntos de una variedad a otra de menores dimensiones.

## Flujo de trabajo

Posteriormente, agrupamos los puntos proyectados.

### **HDBSCAN: Hierarchical Density-Based Spatial Clustering of Applications with Noise**

- Aglomeramiento basado en densidad (DBSCAN), combinado con aglomeramiento jerárquico (H).
- Asigna a cada punto un *cluster* o lo clasifica como ruido.
- Permite controlar el tamaño, forma y número de *clusters* con hiperparámetros.

# Preliminares

## Flujo de trabajo

- Una vez agrupados los datos, tendremos una taxonomía con **géneros** y **supergéneros**.
- Entrenamos un clasificador binario para cada uno de los nodos.

### **XGBoost: eXtreme Gradient Boosting**

- Utiliza árboles de decisión con potenciación de gradiente.
- Robusto ante datos desbalanceados o mixtos.
- Fácil de interpretar.

# Preliminares

## Flujo de trabajo

- Los clasificadores aprenderán a distinguir entre cada género y supergénero, utilizando las características musicales.
- Queremos entender el proceso de decisión para saber qué hace a cada género único.

### **SHAP: SHapley Additive exPlanations**

- Entre mayor sea la magnitud del valor SHAP, mayor será la contribución.
- Características con valores positivos “convencen” al clasificador de dar una respuesta positiva (sí pertenece al género).
- Características con valores negativos lo “convencen” de dar una negativa.

## Desarrollo

---

## Integración de la información

- 22 millones de canciones.
- Cada canción tiene uno o más artistas asociados.

Nombre	Tipo	Rango
Duración	Real	$[0, \infty)$
<i>acousticness</i>	Real	$[0, 1]$
<i>danceability</i>	Real	$[0, 1]$
<i>energy</i>	Real	$[0, 1]$
<i>explicit</i>	Entero	$\{0, 1\}$
<i>instrumentalness</i>	Real	$[0, 1]$
<i>key</i>	Entero	$[0, 11]$
<i>liveness</i>	Real	$[0, 1]$
<i>loudness</i>	Real	$[0, 1]$
<i>mode</i>	Entero	$\{0, 1\}$
<i>speechiness</i>	Real	$[0, 1]$
<i>tempo</i>	Real	$[0, \infty)$
<i>time signature</i>	Entero	$[3, 7]$
<i>valence</i>	Real	$[0, 1]$

**Tabla 1:** Características numéricas de canciones.

## Integración de la información

- 228 000 artistas.
- Cada artista tiene cero o más géneros asociados.

Nombre	Géneros
The Beatles	psychedelic rock, beatlesque, rock, classic rock, british invasion, merseybeat
Skrillex	electro, edm, brostep, complextro
Vicente Fernández	ranchera, mariachi

**Tabla 2:** Géneros de algunos artistas de ejemplo.



## Integración de la información

Para la limpieza:

### Canciones

- Eliminamos todas las canciones con valores faltantes.
- Estandarizamos **solo para la proyección**.
- Para la clasificación tomamos las características en sus escalas originales.

### Artistas

- Quitamos todos los que tuviesen menos de 1000 seguidores.
- Quitamos todos los que no tuviesen géneros registrados.

## Modelado

1. Agrupamos las canciones por artista y calculamos el promedio de sus características. Esto nos da un punto por artista.
2. Proyectamos estos puntos a dos dimensiones utilizando UMAP.
3. Generamos los *clusters* con HDBSCAN.

Realizamos *grid search* exhaustivo, tomando en cuenta:

- Índice de validez de *cluster* basada en densidad (DBCV). Un valor más alto indica que el *cluster* representa de mejor manera a los puntos.
- Tamaño mínimo, máximo y promedio de los *clusters*.
- Cuántos puntos son clasificados como ruido.

## Modelado

- Para determinar a qué supergénero pertenece cada género, contamos cuántas veces aparece en cada *cluster*.
- Lo asignamos al supergénero en el que aparezca más veces.

Por ejemplo, un género con la distribución de la derecha se asignaría al *cluster* 0.

# de cluster	Frecuencia
0	100
1	20
2	10
3	2
4	3
5	50
6	1

**Tabla 3:** Distribución de géneros de ejemplo.

## Modelado

La arquitectura de los clasificadores es:

- *Modelo*: XGBoost
- *Función objetivo*: Entropía cruzada
- *Método de optimización*: Optimización bayesiana con 50 iteraciones y 5-fold cross-validation.

Evaluamos tres clasificadores:

- Clasificador sin información de la jerarquía (vainilla).
- Clasificador jerárquico
- Clasificador jerárquico con superclasificadores perfectos.

## Modelado

Utilizamos dos métricas para evaluar la clasificación:

- Precisión promedio: Área debajo de la curva precisión-sensibilidad. Una mayor puntuación indica que el clasificador comete menos errores.
- Error de cobertura

## Modelado

### Error de cobertura

- Métrica especial para clasificación multi-etiqueta.
- Al final de la clasificación, tenemos la *probabilidad* de que la canción pertenezca a cada uno de los géneros.
- Ordenando de mayor a menor, tenemos un *ranking* de más probable a menos probable.
- El error de cobertura nos dice cuántos de los géneros más probables tenemos que reportar para estar seguros de que incluimos todos los verdaderos.

Género	Probabilidad ( %)
<b>Cumbia</b>	90
<b>Salsa</b>	80
Rap	65
<b>Bachata</b>	60
Metal	30
Rock	20
Pop	10

**Tabla 4:** Ejemplo de resultados de clasificación. Los géneros verdaderos se muestran en negritas.

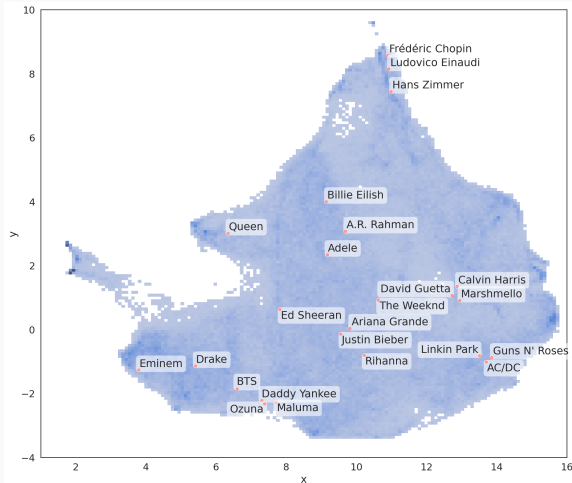
## Resultados

---

# Resultados

## Jerarquías

Después de reducir con UMAP, obtenemos la siguiente proyección de artistas:



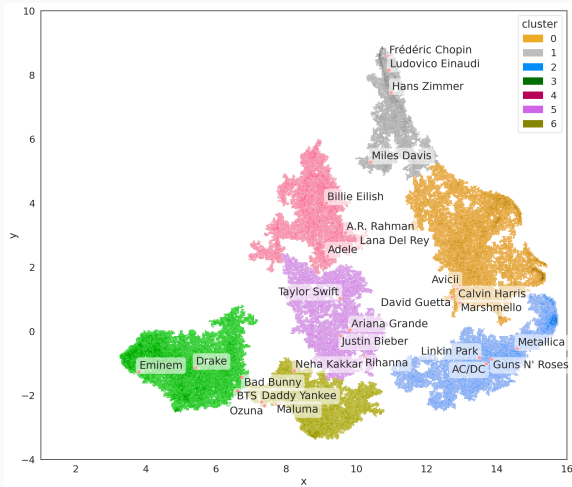
**Figura 1:** Puntos proyectados con UMAP. Se muestran algunos artistas populares.



# Resultados

## Jerarquías

Agrupando con HDBSCAN, los *clusters* resultantes son:



**Figura 2:** *Clusters* obtenidos con HDBSCAN. No se muestran los puntos etiquetados como ruido.

# Resultados

## Jerarquías

Un análisis de los géneros presentes en cada *cluster* sugiere la siguiente nomenclatura para los supergéneros:

Número	Nombre	Géneros comunes
0	Electrónica	house, EDM, electrónica
1	Instrumental	ambient, soundtrack, classical
2	Rock y metal	metal, rock, screamo
3	Rap	rap, hip hop
4	Pop instrumental	folk, indie, pop
5	Pop bailable	pop, dance pop, electropop
6	Ritmos latinos	corrido, tropical, regional mexicano

**Tabla 5:** Nombres y constitución de cada supergénero.

# Resultados

## Clasificación

Evaluando cada modelo sobre el conjunto de validación, obtenemos los siguientes resultados:

Modelos	Precisión promedio	Error de cobertura
Vanilla	0.6060	3.396
Jerárquico	0.6229	3.305
Jerárquico c/hint	<b>0.7501</b>	<b>2.035</b>

**Tabla 6:** Resultados de los distintos clasificadores. El error de cobertura mínimo es 1.504.

# Resultados

## Clasificación

Desglosando la precisión para cada género y supergénero:

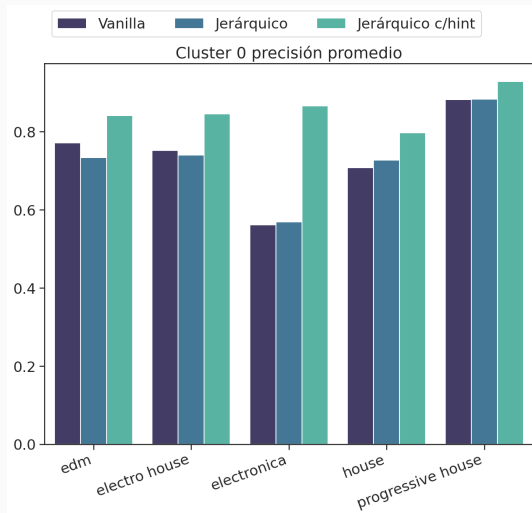


Figura 3: Electrónica.

# Resultados

## Clasificación

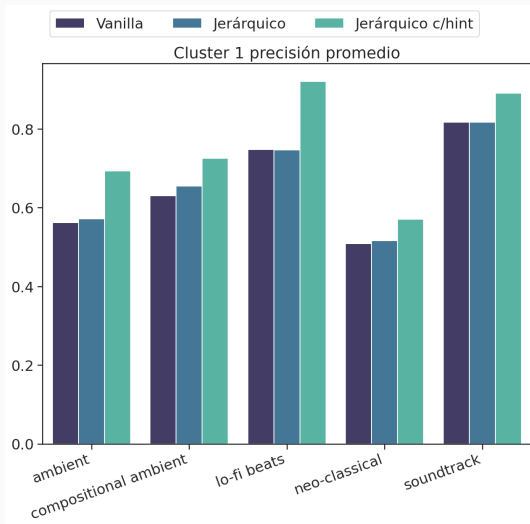
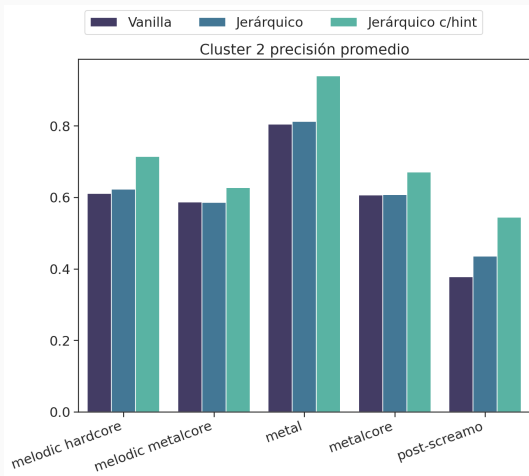


Figura 4: Rap.

# Resultados

## Clasificación



**Figura 5:** Rock.

# Resultados

## Clasificación

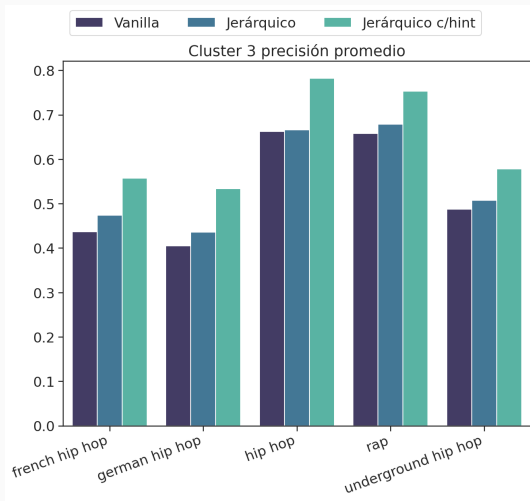


Figura 6: Rap.

# Resultados

## Clasificación

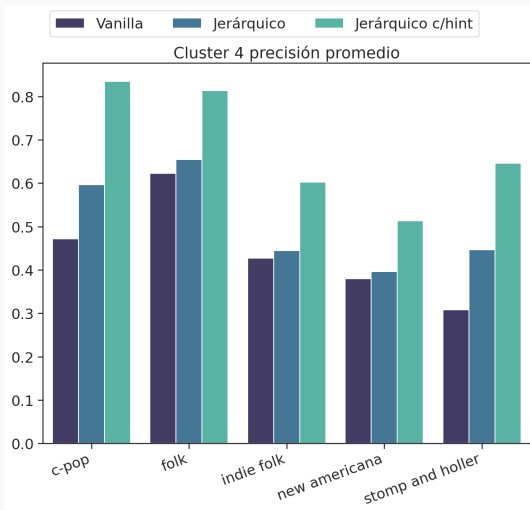


Figura 7: Folk.



# Resultados

## Clasificación

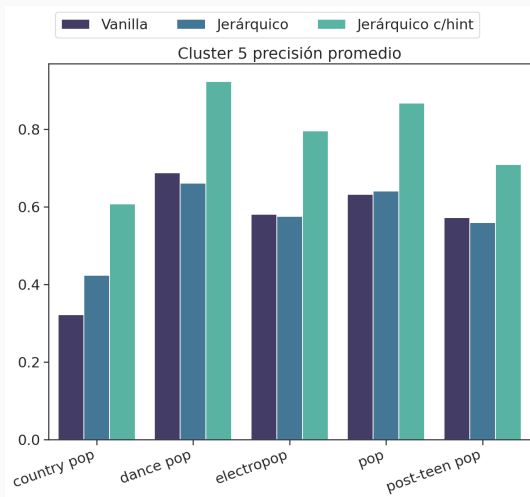


Figura 8: Pop.

# Resultados

## Clasificación

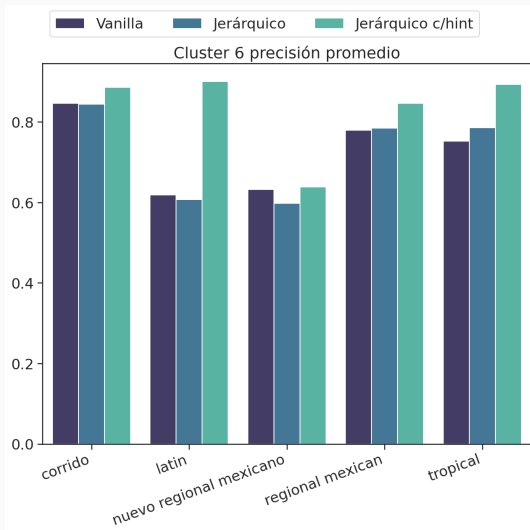


Figura 9: Latina.

# Resultados

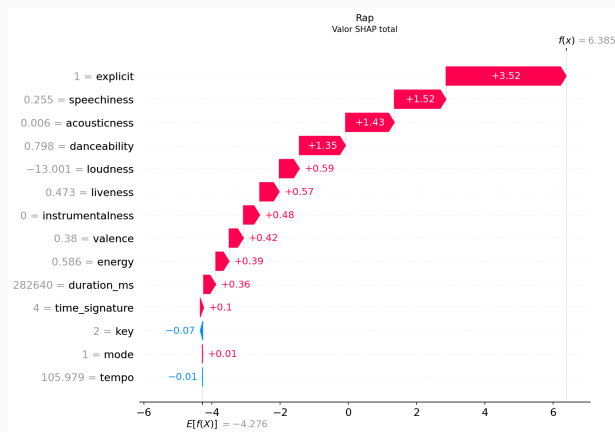
## Clasificación

- En general, los clasificadores jerárquicos tienen un mejor desempeño que el clasificador normal.
- Son especialmente poderosos para desambiguar casos en los que el clasificador normal tiene malos resultados.
- El clasificador con superclasificadores perfectos tiene un desempeño mucho mayor. Esto ilustra la importancia de entrenar estos apropiadamente, ya que sus decisiones afectan a todos sus descendientes.

# Resultados

## Interpretación

Para entender los valores SHAP, visualicemos el proceso de clasificación del clasificador de **rap** para la canción *Fight the Power*, de Public Enemy (que es de rap):

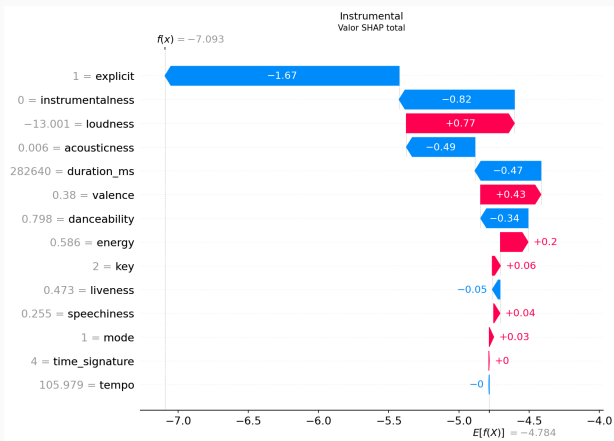


**Figura 10:**  
Proceso de  
decisión para el  
clasificador de  
rap.

# Resultados

## Interpretación

Ahora, veamos qué pasa usando la misma canción pero para el clasificador **instrumental**:



**Figura 11:**  
Proceso de  
decisión para el  
clasificador  
instrumental.

# Resultados

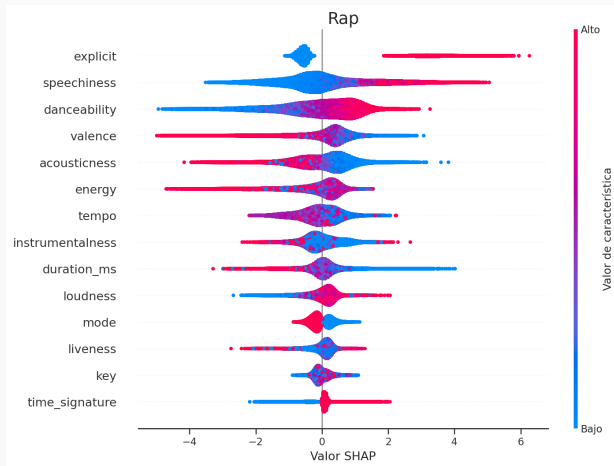
## Interpretación

- Podemos ver que ambos clasificadores le dieron mucha importancia a si la canción era explícita o no.
- En general, el valor SHAP para cada clasificador y característica varía de instancia a instancia.
- Para algunas canciones, una misma característica lo puede convencer de que es positivo, y para otras de que es negativo.
- Si graficamos la distribución de los valores para cada característica, podemos ver hacia qué lado tienden en promedio.

# Resultados

## Interpretación

Haciéndolo para el clasificador de rap:

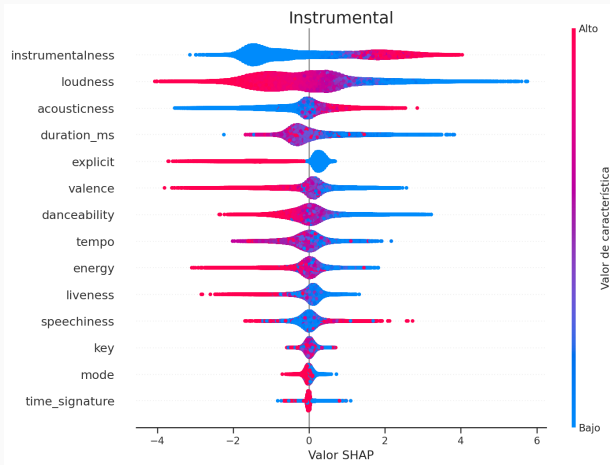


**Figura 12:**  
Distribución de  
valores SHAP del  
clasificador de  
rap.

# Resultados

## Interpretación

Y para el instrumental:



**Figura 13:**  
Distribución de  
valores SHAP del  
clasificador  
instrumental.



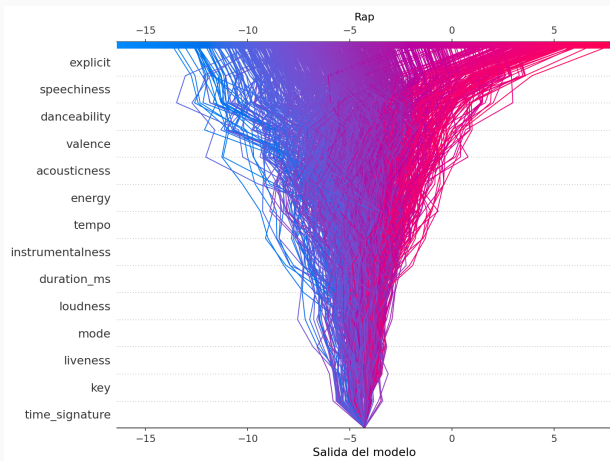
# Resultados

## Interpretación

- Otra manera de visualizar esto es graficando las “trayectorias” que siguen las canciones.
- Al “pasar” por cada característica se desvían a la izquierda si esa característica decrementó su valor SHAP, o a la derecha si lo incrementó.
- Sirve para detectar *outliers*, ya que estos mostrarán una desviación excesiva.

# Resultados

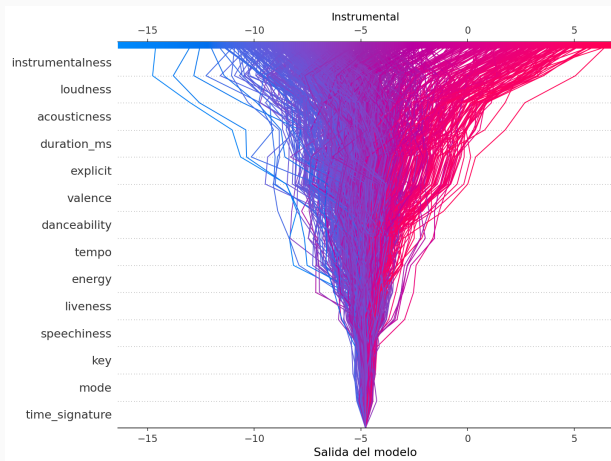
## Interpretación



**Figura 14:**  
Trayectorias de  
1000 canciones de  
rap al pasar por el  
clasificador de  
rap.

# Resultados

## Interpretación



**Figura 15:**  
Trayectorias de  
1000 canciones  
instrumentales al  
pasar por el  
clasificador  
instrumental.

## Conclusiones

---

# Conclusiones

- A pesar de haber hecho una reducción drástica de dimensiones, los *clusters* encontrados representan muy bien diversos aspectos de las canciones.
- Los clasificadores jerárquicos producen mejores resultados que los normales.
- Sin embargo, como las predicciones de los géneros dependen directamente de su supergénero padre, es vital que los clasificadores superiores tengan un muy buen desempeño.
- El análisis con los valores SHAP mostró muchas propiedades interesantes de cada género musical.

## Trabajo futuro

---

# Trabajo futuro

- Los supergéneros obtenidos dependen de la calidad de la proyección y del *clustering*. Distintos hiperparámetros producían *clusters* radicalmente diferentes, por lo cual sería interesante ver qué otros supergéneros pueden derivarse, y cómo afectan estos los resultados de la clasificación.
- La selección y preprocesamiento de características introducidas a los clasificadores fue mínima, ya que se deseaba obtener una caracterización de los datos a través de las importancias que los árboles asignaban a cada una. La clasificación podría repetirse utilizando modelos mucho más refinados.
- Para esta práctica, se seleccionaron los 5 géneros más populares para cada supergénero. Se espera que el desempeño del clasificador vainilla se degrade mucho más rápido que los jerárquicos conforme aumenta este número, aunque no ha sido probado.

## Referencias

---



- [1] SILLA, C. N., AND FREITAS, A. A.  
**A survey of hierarchical classification across different application domains.**  
*Data Mining and Knowledge Discovery* 22, 1-2 (apr 2011), 31–72.
- [2] SPOTIFY.  
**Spotify Stream On - YouTube, 2021.**