

Clasificación No Supervisada de Pistas por Género Musical

*Trabajo Práctico I- Data Mining en Ciencia y Tecnología- 2do Cuatrimestre de 2020

TESTORI, Franco FCEN + FIUBA <i>Universidad de Buenos Aires</i> CABA, Argentina. <franco.testori@hotmail.com>	HRUSZECKI, Darío FCEN + FIUBA <i>Universidad de Buenos Aires</i> CABA, Argentina. <dario.hruszecki@gmail.com>	CHAMORRO, Luciano L. FCEN + FIUBA <i>Universidad de Buenos Aires</i> CABA, Argentina. <luc.chamo@gmail.com>	PEZZOT, Nicolás FCEN + FIUBA <i>Universidad de Buenos Aires</i> CABA, Argentina. <nicolaspezzot@hotmail.com>
---	---	---	--

Resumen—La clasificación automática de música es un tema importante en el campo de *Music Information Retrieval* con muchas aplicaciones interesantes. En este campo, se produce el análisis computacional de señales musicales para aplicaciones tales como el reconocimiento de género o sistemas de recomendación. Muchas aplicaciones se basan en medir una similaridad entre piezas musicales para encontrar patrones. En este trabajo se busca aplicar técnicas de aprendizaje no supervisado mediante *clustering* para rescatar tendencias en los comportamientos de las pistas.

El código del desarrollo del presente trabajo se encuentra en el repositorio https://github.com/francotestori/dmctuba_01.

I. INTRODUCCIÓN

Music Information Retrieval es una ciencia interdisciplinaria que abarca múltiples campos de investigación y aplicación.

Este campo de estudio, tiene usos muy variados desde su aplicación en el desarrollo de sistemas de recomendación como también en la categorización y clasificación de señales musicales automática según su género.

En la actualidad hay múltiples aplicaciones de MIR que se están desarrollando y que permiten que esta ciencia vaya tomando relevancia en el campo de **Machine Learning**.

Uno de los factores que más motivan el desarrollo de este campo de estudio es encontrar aplicaciones que permitan comprender y mejorar la experiencia de los oyentes ante la nueva oferta masiva que se encuentra actualmente online.

Es importante resaltar que la música, más allá de componente meramente objetivos como lo son las señales sonoras subyacentes; tiene una riqueza subjetiva asociada a la valoración emocional de quien la oye. Y

esta información, que compete a campos de estudio sociales y psicológicos, brinda mucha riqueza en los diversos análisis investigativos de MIR.

Dada a la amplitud de este campo de estudio, en el último tiempo han habido varios estudios de aplicación de MIR. Algunos de los usos recientes, utilizados en papers investigativos, son:

1. Algoritmos de clasificación para la categorización en géneros musicales.
2. Algoritmos de etiquetado para detección de emociones.
3. Algoritmos de clustering para detección de estilos musicales.
4. Algoritmos de recomendación semi-supervisados.

Teniendo en mente estos antecedentes, el objetivo general de nuestro trabajo consiste en determinar si las pistas musicales (**tracks**) se agrupan naturalmente por géneros a partir de los distintos atributos que hacen a un track con distintas perspectivas y niveles de observación.

Para este trabajo, utilizaremos tracks provenientes de **Spotify** y aplicaremos distintos algoritmos de clustering, junto con otras técnicas de análisis para intentar comprender la relación entre los distintos atributos de nuestra muestra y sus etiquetas de género.

Por ende, nos proponemos a:

- explorar los distintos atributos que conforman a un track musical.
- buscar comprender y analizar el etiquetado de géneros de Spotify y analizar la bondad del mismo.
- aplicar distintos algoritmos de clustering para comprender si nuestros tracks musicales se agrupan naturalmente en los géneros preestablecidos.

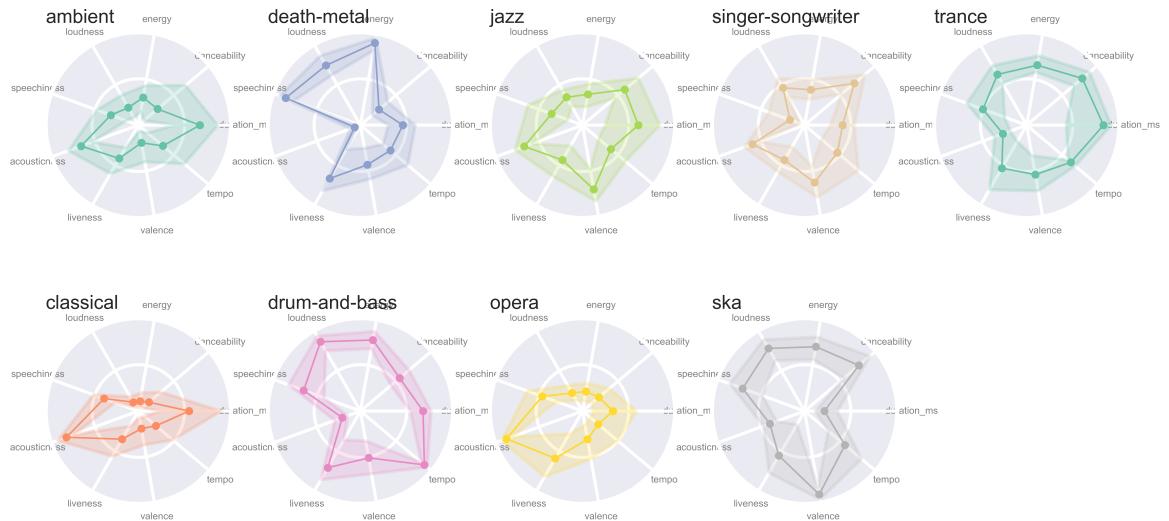


Figura I.1. Gráfico Radar de los *features* de alto nivel por cada género del set de datos. La línea gruesa indica la mediana, y el sombreado es el rango intercuartil. Las características se encuentran escaladas por cuantiles (dando una medición robusta de cada *feature* para hacer los estadísticos comparables).

- juzgar y concluir respecto de las agrupaciones observadas y el etiquetado asignado por Spotify a nuestra muestra musical.

Se toman como punto inicial tres sets de datos para el análisis de la canción:

- Los datos de la pista.
- Los *features* descriptores de la canción.
- Las componentes de bajo nivel del audio que describen su estructura y contenido musical: ritmo, tono y timbre.

I-A. Datos y Atributos

I-A1. Adquisición de Datos: La base de datos está compuesta por canciones de distintos artistas (tanto de renombre como no) a lo largo de un gran espectro de géneros musicales.

Para obtener las características de los audios, se hizo uso de la API de Spotify. A través de su plataforma Echo Nest, Spotify efectúa un *feature engineering* de las canciones en su repositorio y aplica distintos algoritmos para identificación musical, sistemas de análisis de audio y clasificación de géneros.

El estudio se realiza sobre 4092 registros descargados a través de la API de Spotify. Los datos descargados pertenecen a los sets de *Tracks*, *Audio Features* y *Audio Analysis*.

Aquí se trata con música de 9 géneros, los que serán descompuestos y analizados a lo largo de esta sección para encontrar qué caracteriza a cada uno.

I-A2. Preparación de los Datos: Los datos de la API de Spotify requieren de un preprocesamiento para realizar los agrupamientos de piezas musicales.

Después de la unión de los tres sets y la eliminación de registros que tienen campos vacíos, se obtuvo un ensamble de 4082 canciones.

El set de datos cuenta con *features* de alto nivel propuestos por la plataforma Echo Nest como buenos descriptores de las canciones. Algunos de estos índices se encuentran asociados al ritmo, la energía, la intensidad, el contenido verbal, la acústica, etc. Se pueden encontrar patrones generales de estos datos realizando un análisis del gráfico de la Fig. I.1.

- La ópera y la clásica son los que más contenido acústico tienen. Seguidas por la música ambiente, el jazz y las singer-songwriter.
- Las canciones de trance son las que en general suelen tener mayor duración (seguidas por las ambiente, drum-and-bass y jazz).
- Drum-and-bass suele ser el género de mayor índice de canción en vivo (*liveness*), junto con death-metal y ska.
- Death-metal es incuestionablemente el género más energético, seguido por drum-and-bass, ska y trance.
- La ópera y la clásica comparten las características de baja energía, baja valencia, bajo tempo, alta acústica y bajo contenido verbal.
- Los géneros más rítmicos son singer-songwriter, trance, jazz y ska.

Se cuenta también con datos de bajo nivel prove-



Figura I.2. Gráfico Radar de las estadísticas extraídas de *pitch*, *timbre* y *confidence* por cada género del set de datos. La línea gruesa indica la mediana, y el sombreado es el rango intercuartil. Las características se encuentran escaladas por cuantiles (dando una medición robusta de cada *feature* para hacer los estadísticos comparables).

nientes de las series temporales de los audios. Estos describen la estructura de la pista y el contenido musical en formato temporal, junto con el ritmo, el tono y el timbre.

Las variables del set *Audio Analysis* son temporales, por lo que es necesario aplicar una transformación sobre estas series para rescatar las características principales. A los fines prácticos, en este trabajo se plantea extraer la mediana y el rango intercuartil (IQR) de cada serie temporal.

A su vez, también se hace uso del recurso de la primer diferencia de las series, y también se extraen la mediana y el rango intercuartil de ellas. Estos estimadores se plantean con el objetivo de extraer conocimiento de la derivada de las series.

Algunas de estas medidas de resumen pueden ser vistas en el gráfico de la Fig. I.2. Estos datos aportan riqueza al análisis para una posterior separación de clusters. Se pueden distinguir características particulares para cada uno de los gráficos radar.

Una vez decididos los *features* que se le alimentarán a los algoritmos de *clustering*, se procede a transformarlos con la intención de minimizar el efecto de valores atípicos, estandarizar los rangos de datos y simplificar las medidas de distancia en los distintos campos.

Se efectuaron pruebas con las siguientes transformaciones:

- Escalamiento MinMax - Transformación lineal de todos los valores al intervalo[0; 1].

- Distribución uniforme por cuantiles - Toma los centiles y los redistribuye para redistribuirlos en una uniforme.
- Normalización por cuantiles - Toma los centiles y los redistribuye para convertirlos en una normal.
- Transformación por potencia - Escalamiento robusto que asemeja la distribución a una gaussiana mediante el método Yeo-Johnson.

I-B. Géneros

Uno de los análisis preliminares que efectuamos consistió en validar si los géneros asignados por Spotify a cada uno de los tracks categorizaba correctamente a nuestra muestra o si la misma se encontraba aleatoriamente etiquetada.

En primer lugar, analizamos las distintas tendencias al clustering de nuestras muestras (para cada uno de los datasets transformados generados previamente) y utilizamos el **Estadístico de Hopkins** para comprobarla.

Con este análisis comprobamos que todos nuestros datasets transformados, tenían valores cercanos a 0, lo cual indicó que los valores se encontraban uniformemente distribuidos.

Luego buscamos analizar la bondad de la clasificación de nuestra muestra en base al uso de la etiqueta “género”. La métrica que elegimos para validar fue el **Coeficiente de Silhouette**.

Como podemos observar en la siguiente tabla, los valores de Silhouette obtenidos, en su mayoría, son cercanos a 0 con una tendencia positiva (a excepción del dataset sin transformar que tiene una tendencia

negativa). Entonces podemos asumir que nuestros datasets tienden hacia agrupamientos no muy relevantes (aunque su inclinación hacia los valores positivos indicarían un acercamiento a un agrupamiento moderado).

Dataset	Indice de Hopkins	Coeficiente de Silhouette
Reducido	0.045	0.046
Robusto	0.081	0.028
Min-Max	0.077	0.044
Uniforme	0.173	0.039
Normalizado	0.065	0.028
No escalado	0.076	-0.260

Cuadro I.1

INDICE DE HOPKINS Y COEFICIENTES DE SILHOUETTES TOMANDO COMO CLUSTERS LAS ETIQUETAS DE GÉNERO

Para finalizar, analizamos los valores medios para los Coeficientes de Silhouette para cada una de las etiquetas de género como se ve en la siguiente tabla.

Género	Cantidad de Tracks	Silhouette (Media)
ambient	421	-0.066
classical	396	-0.042
death-metal	477	0.282
drum-and-bass	486	0.005
jazz	442	-0.007
opera	477	0.111
singer-songwriter	447	0.044
ska	461	0.047
trance	475	0.011

Cuadro I.2

MEDIA DE SILHOUETTE POR GÉNERO PARA EL DATASET REDUCIDO (MEJOR COEFICIENTE)

De los resultados obtenidos, encontramos que los géneros que peor se agrupan son **ambient**, **classical** y **jazz**. Y en cuanto al resto, tienen valores muy cercanos a 0. Podemos observar este comportamiento en el siguiente gráfico.

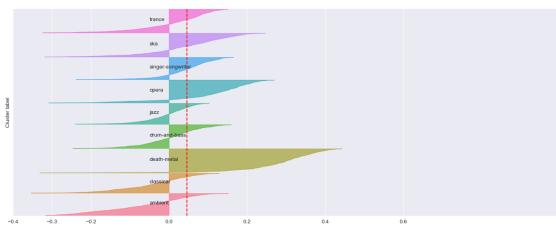


Figura I.3. Silhouettes por cluster en etiquetas de género.

I-B1. Visualización: En la Fig. I.4 se muestra un gráfico t-SNE. Aquí se visualizan las relaciones entre los distintos individuos. Las transformaciones realizadas en la sección anterior ayudaron a homogeneizar la

distribución para reducir el efecto de los outliers y acercar más la función a una forma normal, lo cual se estima será de ayuda para diseñar implementaciones de clasificadores más eficientes.

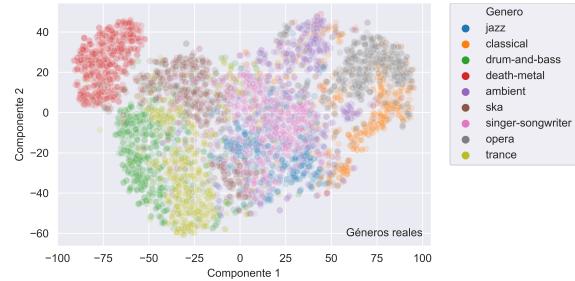


Figura I.4. Visualización resultante de la aplicación del algoritmo de reducción de dimensionalidad t-SNE (perplejidad: 20).

Una vez más, con este gráfico es posible identificar los distintos géneros y las asociaciones con sus semejantes.

No perfectamente segregados, pero sí se notan buenos delineamientos y tendencias que ayudan a un buen entendimiento del dataset. El t-SNE da un soporte para la visualización de esta información.

Un grupo que está altamente segregado del resto es death-metal, cuya mayor relación parece encontrarse con los géneros drum-and-bass y ska (y no coincidentemente, son los géneros más pesados del set).

Después, un agrupamiento particular es drum-and-bass y trance, que se encuentran en un extremo del set, siendo que también se condice con los géneros con alto *liveness*.

En el otro extremo del dataset se encuentran la música clásica y la ópera, con un cierto solapamiento el uno con el otro. Y no coincidentemente está la música ambiente cerca de ellos. Estos tres géneros son los menos intensos.

Las canciones de jazz y singer-songwriter se encuentran solapadas en la visualización, en una posición céntrica y son las más espaciadas. Esto puede ser indicador de que son las más balanceadas del set, siendo los géneros más versátiles.

II. EXPERIENCIA 1: K-MEANS

Se propone comenzar con el método de K-means para probar la medida de clusterización.

En este algoritmo es necesario especificar la cantidad de clusters a usar. Para ello, en este trabajo se usa el método del codo. Este análisis consiste en definir la cantidad de manera que se minimiza la variación total entre clusters (*WSS - within-cluster sum of squares*). Esto puede verse en el gráfico II.1.

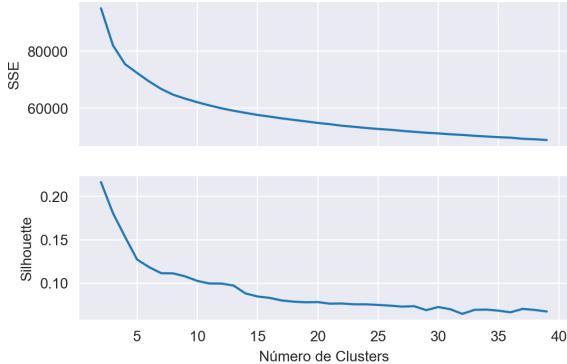


Figura II.1. Gráficos del SSE y Silhouette en función de la cantidad de clusters para el caso de K-means.

También, se aplica el criterio de Silhouette a modo de comparación, y se puede identificar una alta similaridad entre ambos gráficos, por lo que se sugiere que se aplique el valor de 9 clusters como óptimo.

II-A. Métodos

La clasificación por k-means consiste en definir los grupos para minimizar la variación total intra-cluster. En este trabajo se usa el algoritmo de Elkan, que define el total de variación entre clusters como la suma de distancias cuadradas euclídeas entre los ítems y sus correspondientes centroides.

Este experimento resulta en la clusterización graficada en la Fig. II.2.

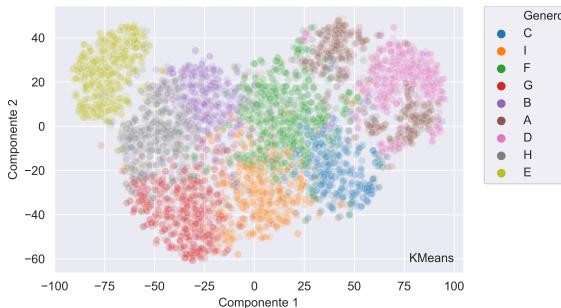


Figura II.2. Clustering por K-Means.

II-B. Resultados y Discusión

Para efectuar un análisis más detallado de la solución, se propone usar como medidas de validación a los índices Van Dongen y Rand ajustado.

Un índice de Van Dongen pequeño y un Rand ajustado alto son indicadores de una mayor afinidad entre la categorización oficial y la generada por el algoritmo.

Estos valores serán usados como *benchmark* frente al resto de los experimentos.

s Clustering	Silhouette	Van Dongen	Adjusted Rand
K-Means	0.108	0.487	0.386

Cuadro II.1

VALIDACIÓN DE LOS CLUSTERS POR EL MÉTODO DE KMEANS CON 9 GRUPOS.

Como punto final de análisis, se cuenta también con la matriz de confusión graficada en la Fig. II.3.

Real	K-Means								
	ambient	3	65	5	1	121	17	3	35
classical	163	0	96	103	0	30	0	0	4
death-metal	1	0	2	0	466	8	0	9	0
drum-and-bass	0	19	0	0	7	10	178	248	15
jazz	7	12	132	4	0	108	13	2	143
opera	27	1	10	375	0	41	0	1	6
singer-songwriter	0	30	71	3	1	247	2	12	81
ska	0	262	0	0	22	9	1	78	105
trance	2	36	0	0	2	25	276	91	43

Figura II.3. Matriz de confusión con la comparación del clustering k-means con 9 grupos contra las categorías de Spotify.

Aquí se visualiza lo esperado en la sección anterior. Hubo géneros que no pudieron ser apropiadamente separados (como lo son la música ambiente y la clásica), pero sí hubo otros géneros con una alta tasa de aciertos (como death-metal y gran parte del ska). En una comparación de los gráficos t-SNE también se pueden apreciar estas características.

III. EXPERIENCIA 2: CLUSTERING JERÁRQUICO

III-A. Métodos

Dentro de los métodos empleados, nos encontramos con el uso de Clustering Jerárquico aglomerativo. Se implementó para distintos datasets con distintas transformaciones y distintas combinaciones de datos. Se consideró la distancia euclídea, dado que la misma es la de mayor uso para trabajar con variables continuas. También se probaron distintas funciones de enlace como average, centroid, ward, single, complete, weighted y mediana. Se empleó el coeficiente de correlación cofenético, el cual indica la correlación de las distancias definidas por la métrica de árbol binario con las distancias originales entre objetos. Para la determinación del número de clusters se utilizó el dendrograma que nos provee información acerca de cómo se van dando los agrupamientos y también el coeficiente de Silhouette, que nos brinda una métrica que compara la cohesión dentro de los grupos contra la separación respecto a los demás grupos. Se calcularon los índices de Rand y de Van Dongen para comparar los agrupamientos realizados para los distintos conjuntos de datos y también con el

género (etiqueta). Se empleó la técnica de reducción de dimensionalidad TSNE, para visualizar los agrupamientos obtenidos.

III-B. Resultados

III-B1. Selección del Método de enlace: Para efectuar la selección del método de enlace se calculó primeiramente el coeficiente de correlación cofenético para cada uno de los métodos como se muestra en la Tabla III.1

Método	Coeficiente de correlación cofenético
average	0.611
centroid	0.581
ward	0.503
single	0.356
complete	0.441
weighted	0.473
median	0.243

Cuadro III.1
COEFICIENTE DE CORRELACIÓN COFENÉTICO PARA DISTINTOS MÉTODOS DE ENLACE

Si bien los métodos average y centroid son los de mayor coeficiente, los mismos generan un cluster extremadamente poblado y el resto de los clusters con muy poca cantidad de observaciones, con lo cual se decidió utilizar el método de ward, dado que genera clusters más homogéneos en cantidad de canciones que tienen más sentido para el caso de estudio.

III-B2. Determinación de la Cantidad de Clusters: Respecto a la cantidad de clusters se calculó el coeficiente de Silhouette para distinto número de clusters, como se puede ver en la Fig. III.1

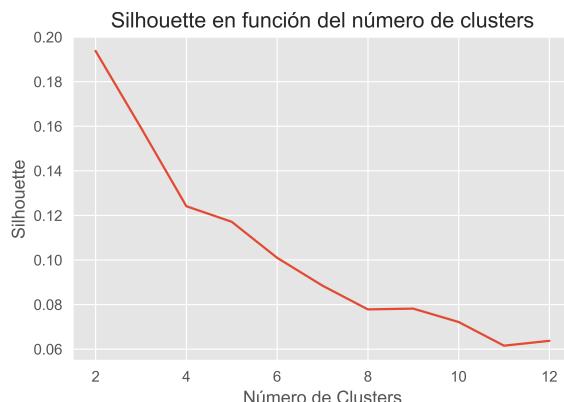


Figura III.1. Gráfico del valor de Silhouette en función a la cantidad de clusters para clustering jerárquico.

En base a esto en 8 clusters podemos ver un amesetamiento y luego una caída bastante pronunciada con lo cual se definió una cantidad igual a 8.

También se efectuó un dendrograma, como se puede ver en la Fig. Fig. III.2

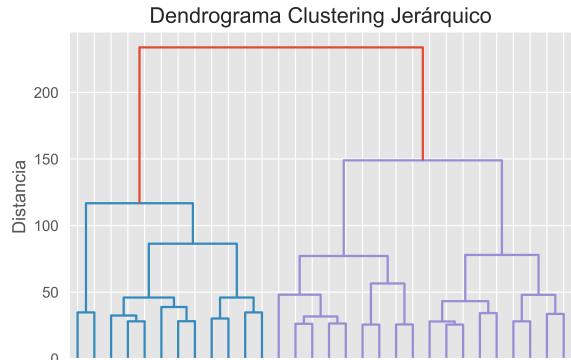


Figura III.2. Dendrograma para clustering jerárquico.

Aquí analizamos el punto en el que se produce un salto grande en la distancia y en base a esto también se definen 8 clusters, donde obtenemos un buen punto de corte.

III-B3. Comparación entre distintos conjuntos: Comparamos los agrupamientos obtenidos con distintos datasets mediante índices como el de Rand y Van Dongen, contra nuestro dataset base que es el dataset Reducido, como se puede ver en la Tabla III.2

Dataset	Indice de Rand	Indice de Van Dongen
Robusto	0.429	0.427
Min-Max	0.325	0.539
Normalizado	0.277	0.587
No escalado	0.039	0.896

Cuadro III.2
INDICE DE RAND Y VAN DONGEN PARA LOS DISTINTOS MODELOS POR CLUSTERING JERÁRQUICO

En la Fig III.3 se visualiza la matriz de confusión donde obtenemos los mejores índices, es decir se compara contra el escalado robusto:

Clusters	0	1	2	3	4	5	6	7	
-Clusters	0	309	19	72	1	4	87	33	136
1	110	0	26	1	0	399	0	6	
2	5	32	1	1	404	0	84	10	
3	0	301	0	23	196	3	1	12	
4	8	0	349	0	0	18	0	1	
5	0	17	0	452	2	0	0	0	
6	121	34	21	0	20	16	144	194	
7	0	293	3	1	41	0	39	32	

Figura III.3. Matriz de confusión comparando dos conjuntos de datos.

Vemos que en general hay un buen grado de similitud en la asignación de los clusters entre estos dos conjuntos de datos.

III-B4. Comparación entre distintos conjuntos: Se efectuó una matriz de confusión, Fig. III.4, entre el número de clusters asignado por el algoritmo para cada conjunto y el género (labels).

	Clusters	0	1	2	3	4	5	6	7
	Labels								
ambient	ambient	79	8	4	0	25	238	19	69
classical	classical	119	0	73	0	0	0	196	1
death-metal	death-metal	3	16	2	460	2	3	0	0
drum-and-bass	drum-and-bass	2	188	0	3	270	2	11	1
jazz	jazz	179	14	14	1	18	13	90	92
opera	opera	25	0	374	0	0	52	2	8
singer-songwriter	singer-songwriter	142	40	2	1	5	14	66	177
ska	ska	3	328	2	12	10	0	103	19
trance	trance	1	102	1	2	337	5	9	18

Figura III.4. Matriz de confusión comparando agrupamientos con etiquetas de género.

Como podemos observar, más de un 90 % de las canciones de Death Metal quedan en un mismo grupo. A su vez la mayoría de las canciones de Opera se encuentran en un único cluster, lo mismo ocurre pero en menor medida con Trance, seguido por Ska y Drum and Bass.

Dataset	Indice de Rand	Indice de Van Dongen
Reducido	0.392	0.488
Robusto	0.415	0.435
Min-Max	0.265	0.569
Normalizado	0.264	0.604
No escalado	0.059	0.838

Cuadro III.3

INDICE DE RAND Y VAN DONGEN PARA LOS DISTINTOS CONJUNTOS DE DATOS POR CLUSTERING JERÁRQUICO

Como podemos observar en la tabla III.3 el set de datos con escalado Robusto es el que nos brinda mejores valores de los índices de Rand (más cercano a 1) y Van Dongen (más cercano a 0). Estos índices acompañan el vínculo entre las etiquetas y los clusters conformados, a mejores métricas, mejor se agrupan de acuerdo al género. No obstante también vemos que el dataset Reducido nos da valores muy similares.

III-B5. Visualización de los clusters y etiquetas de género mediante técnica t-SNE: Se llevó a cabo una reducción de dimensionalidad considerando dos componentes mediante el uso de la técnica TSNE, como se puede observar en la Fig. III.5

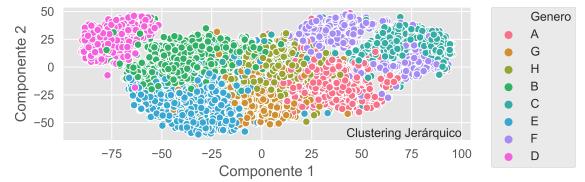


Figura III.5. Visualización de clusters por TSNE para clustering jerárquico

Podemos ver que hay grupos muy bien identificados y separados por esta visualización y algunos otros que se encuentran algo superpuestos.

III-B6. Resultados: A partir de los agrupamientos efectuados vemos que ciertos géneros se agrupan sumamente bien como el death-metal, la ópera y el trance, mientras que otros no tanto, como son la música clásica, el Jazz y Singer-Songwriter.

IV. EXPERIENCIA 3: ALTERNATIVAS AL K-MEANS

Se proponen las alternativas PAM (particionado alrededor de los medoides) y Clustering Espectral.

IV-A. Métodos

Por un lado, el PAM es la alternativa robusta a la implementación K-means. A diferencia del algoritmo K-means, en el que se minimiza la suma total de cuadrados intra-cluster, el algoritmo PAM minimiza la suma de las diferencias de cada observación respecto a su medoide. Las medidas de distancia con las cuales se experimenta son la euclídea, la Manhattan y la coseno.

Y por otro lado, el interés por los métodos espectrales se debe a que su óptimo global en el dominio continuo se obtiene por descomposición de autovalores. Esto significa que los autovectores se tratan como las coordenadas geométricas de un set de puntos. Se puede aplicar un arsenal de heurísticas sobre los nuevos puntos para rescatar particiones óptimas.

En este trabajo se prueban alternativas implementadas en el paquete Scikit-Learn. Se experimenta con las matrices de afinidad basadas en el kernel *rbf* (función de base radial) y en el kernel laplaciano. A cada caso se le ajustan los hiperparámetros de manera de encontrar un máximo en el índice de Silhouette.

Una vez pasadas las variables al dominio transformado, se procede a seleccionar la estrategia para asignar las etiquetas. En este trabajo se prueba con los métodos de K-means y de discretización.

IV-A1. Sensibilidad en la Inicialización: Una inicialización aleatoria de los k puntos centroides resulta en un problema de sensibilidad frente a la inicialización. Este problema afecta la forma final de los clusters, provocando frecuentemente la llegada a óptimos locales muy distintos. La forma de los clusters se

encuentra directamente relacionada con la selección inicial de los centroides.

Para reducir el efecto que provoca una inicialización aleatoria en los algoritmos PAM, se recurre a la técnica K-Meoides++, que es una técnica de inicialización con una evaluación los puntos iniciales en base a la distribución de los datos.

IV-B. Resultados y Discusión

Los resultados de las evaluaciones se encuentran tabulados en la Tabla IV.1.

Clustering	Silhouette	Van Dongen	Adjusted Rand
K-Means	0.108	0.487	0.386
KMedoids (euclidean)	0.081	0.590	0.319
KMedoids (Manhattan)	0.079	0.594	0.294
KMedoids (cosine)	0.072	0.489	0.364
Spectral (laplacian+kmeans)	0.099	0.562	0.336
Spectral (rbf+discretize)	0.118	0.469	0.377
Spectral (laplacian+discretize)	0.091	0.475	0.385

Cuadro IV.1

VALIDACIÓN DE LOS CLUSTERS POR DISTINTOS MÉTODOS (CON 9 GRUPOS).

Aquí se visualiza que el coeficiente de Silhouette resulta similar para todos los casos y, el algoritmo que tuvo la mejor relación en los índices de validación Van Dongen y Rand ajustado es el clustering espectral con el núcleo laplaciano y la estrategia de asignación discreta.

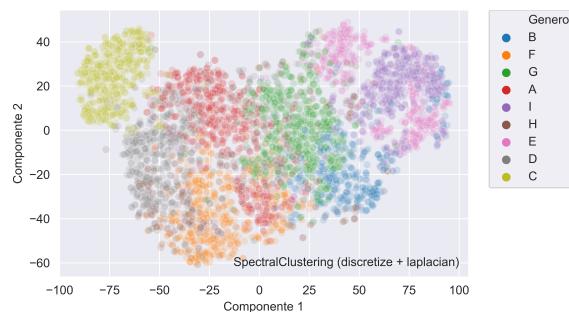


Figura IV.1. Clustering espectral con asignación de categorías discreta y la construcción de la matriz de afinidad es con el método laplaciano ($\gamma = 0,01$).

Comparando la Fig. IV.1 con las etiquetas de géneros de Spotify, se aprecia que hay muchos grupos que fueron correctamente clasificados y que el algoritmo

se acercó a la forma de las nubes de los clusters reales, objetivo que en el resto de los casos no ha sido posible alcanzar.

	0	1	2	3	4	5	6	7	8	
ambient	3	56	2	1	183	40	121	32	4	
classical	0	111	0	0	168	0	10	26	81	
death-metal	2	1	461	4	3	1	5	9	0	
drum-and-bass	9	2	8	315	0	86	4	51	2	
jazz	51	103	0	0	7	49	137	53	21	
opera	4	24	0	0	42	1	29	7	354	
singer-songwriter	69	44	1	12	1	26	252	37	5	
ska	341	0	26	49	0	24	6	27	4	
trance	33	0	3	84	4	235	17	96	3	
	0	1	2	3	4	5	6	7	8	

Figura IV.2. Matriz de confusión con la comparación del clustering espectral con 9 grupos contra las categorías de Spotify.

Esto también puede ser apreciado en la matriz de confusión de la Fig. IV.2, donde los géneros death-metal, drum-and-bass, trance, ska y ópera tuvieron un alto porcentaje de clasificación correcta y aislada.

Un caso que no pudo ser correctamente separado por este algoritmo es la música clásica y la música ambiente, que se encuentran dispersos y combinados a lo largo de dos clusters.

V. ESTANDARIZACIÓN DE LOS MODELOS

Los formatos standard para representar modelos predictivos son herramientas muy útiles que permiten que los mismos sean portables entre distintas tecnologías y ambiente.

Dentro de esos formatos tenemos PMML (Predictive model markup language) y ONNX (Open Neural Network Exchange Format) como los más utilizados.

V-A. Dificultades en la utilización de modelos standard

Una desventaja que tienen, en particular PMML, es el soporte para algunos modelos predictivos solamente.

En el caso de los algoritmos de clusterización en ambos casos solo pudimos generar modelos para KMeans porque es el único caso para el cual se encuentra implementada la lógica.

VI. CONCLUSIONES

En este trabajo se presentó el uso de algoritmos no supervisados para la clasificación de pistas musicales. En términos generales, se pudo evaluar que los algoritmos utilizados para el tratamiento del set de datos crearon panoramas que adicionaron conocimiento sobre el problema a resolver.

Algunos resultados de interés:

1. Las canciones del género death-metal pudieron ser en su mayor parte bien clasificadas con un cluster independiente. Esto es indicador de que los rasgos de este género son muy distintivos del resto.
2. El ska es otro género que fue detectado bien por el algoritmo de Spectral Clustering, pero no pudo serlo por los otros algoritmos. Es un claro ejemplo de los beneficios de la descomposición por autovectores porque permitió una separación del set que no fue posible con un método lineal.
3. La ópera se encuentra bien clasificada en todos los casos también, aunque puede verse que en algunos casos se mimetiza con la música clásica.
4. A pesar de aparentar no solaparse en la visualización t-SNE, dos géneros que no pudieron ser discernidos por la mayoría de los algoritmos son el trance y drum-and-base. Esto puede deberse a la inicialización aleatoria de los algoritmos y podría salvarse con otras condiciones iniciales. Solo el Spectral Clustering logró diferenciarlos.
5. Ninguno de los algoritmos logró una correcta separación de la música ambiente y la clásica. Esto es visible en los gráficos radar del trabajo, donde se perciben distribuciones muy similares de sus variables. Un trabajo a desarrollar en un futuro podría ser la búsqueda de otros *features* que ayuden a la segregación de estos géneros.

REFERENCIAS

- [1] L. I. Kuncheva and D. P. Vetrov. Evaluation of stability of k-means cluster ensembles with respect to random initialization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1798–1808, 2006.
- [2] A. H. Raza and K. Nanath. Predicting a hit song with machine learning: Is there an a priori secret formula? In *2020 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA)*, pages 111–116, 2020.
- [3] Rajib Sarkar and Sanjoy Saha. Music genre classification using emd and pitch based feature. *ICAPR 2015 - 2015 8th International Conference on Advances in Pattern Recognition*, 02 2015.
- [4] Yu and Shi. Multiclass spectral clustering. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 313–319 vol.1, 2003.
- [5] George Tzanetakis, Andrey Ermolinsky, and Perry Cook. Pitch histograms in audio and symbolic music information retrieval. *Journal of New Music Research*, 32, 11 2002.
- [6] Li, Ogihara, Shao, and DingdingWang. *Machine Learning Approaches for Music Information Retrieval*. 01 2009.