Modelo de Predicción de Éxito de Canciones Basado en Descriptores de Audio

Ignacio Barrera and Rodrigo Nazar
Pontificia Universidad Católica de Chile
Second Semester 2021

Link a Carpeta Drive: https://drive.google.com/drive/folders/11dnsGEZ5HEydEkV6MI$_Z8kWAKNbkT3P-$

***Abstract*—El presente trabajo busca modelar la predicción del nivel de éxito de una canción según el ranking de canciones más escuchadas en la plataforma Spotify. Se utiliza una metodología de SOCP para minimizar la región factible para los descriptores de audio de canciones, con lo que se generan restricciones para que una nueva canción sea similar a las utilizadas como datos. Se muestra una simulación computacional basada en datos de Spotify y de elaboración propia. Se concluye con una comparativa con años anteriores.**

I. Introducción

A. Motivación del Problema

En la búsqueda de la construcción de la identidad propia, el ser humano ha procurado separar aquellas cosas que son de su agrado de las que no lo son, ya sea que le son indiferentes o le resulten molestas. En este ámbito, la música es un gran ejemplo de estas clasificaciones inconscientes que realizamos a diario. Los factores que determinan las preferencias son de lo más variados, y discutiblemente pueden llegar a ser considerados imposibles de modelar en su totalidad. Sin embargo, a nivel colectivo es posible notar la preferencia sobre ciertas composiciones, las cuales, aún sin poder describirse en su totalidad, son abordables de forma matemática con más facilidad que al hacerlo de forma individual. En particular, se pueden observar ciertas tendencias en cuanto a géneros musicales, ritmos y otros factores, así como también de otras no apreciables sin herramientas matemáticas y computacionales. El presente trabajo propone una metodología para el uso de categorías clásicas y descriptores de audio para la predicción del éxito de una canción, así como también una caracterización de la evolución de los factores que "determinan" el éxito de las mismas dependiendo del año.

B. Revisión Bibliográfica

Cuando se busca describir una canción en términos distintos a lo puramente audible, o bien, descomponerla en partes más simples de analizar, las soluciones suelen ser variadas. Existen en primer lugar las aproximaciones clásicas como su género (rock, reggae, country), sin embargo, la información que proveen estas categorías es insuficiente para sacar algún tipo de conclusión. Lena y Peterson [1] señalan que gran parte de los géneros más escuchados hoy en día pueden ser rastreados hasta su orígen en un número limitado de géneros, los cuales en el caso de la música escuchada en Estados Unidos son tan solo cuatro.

Un abordaje más intuitivo del problema es propuesto por Askin y Mauskapf [2], quienes en su estudio utilizan 8 categorías en base a las que describen a una canción. Según los autores, estas categorías permiten encontrar un perfil de popularidad que se mantiene relativamente consistente dependiendo del año. Estas categorías son su danzabilidad, tempo, energía, acusticidad, vivacidad, discurso, instrumentalidad y valencia. Es importante notar que todos estos corresponden a parámetros "superficiales", en el sentido de que abordan la forma en que se aprecia la canción, y no su estructura.

A la interpretación anterior, se añade una forma más cuantitativa de análisis, basada en descriptores de audio. Se entiende como descriptores o características de audio al resultado de la aplicación de algoritmos computacionales, que toman como entrada una señal de audio y entregan uno o más valores numéricos. El propósito de un descriptor es que este modele de manera correcta un fenómeno que se desea estudiar, por eso son un pilar fundamental en el análisis de señales.

Existen muchas formas de clasificar los descriptores. Un ejemplo es el dominio en el cual se operan los datos: procesar la señal en el tiempo, su espectro (la magnitud de su transformada de *Fourier*) u otra proyección a otro dominio. Otro ejemplo de clasificación es si son completamente autónomos o si necesitan parámetros externos que regulen su comportamiento. En el libro de Weihs [3] se presenta la teoría del análisis de señales basado en descriptores.

En cuanto a la descripción de tendencias la literatura más cercana se encuentra en los sistemas recomendadores, puesto que en esencia, una recomendación acertada de cualquier clase consiste en modelar correctamente las preferencias de un individuo respecto a elementos de algún conjunto y predecir si un nuevo item será de su agrado. Al respecto, las aproximaciones al problema son variadas. Maldonado y López [4] utilizan un SOCP para crear un clasificador para una Support Vector Machine. Otro algoritmo usado se encuentra en la descomposición por valores singulares, como la realizada por Mehta [5], en la cual aprovecha la ventaja de este método para reducir la cantidad de datos usados y para seleccionar solo aquellos que son más relevantes, con lo cual genera una recomendación.

C. Preguntas de Investigación

En base a lo anteriormente expuesto, la pregunta que se buscará responder es: ¿Qué requisitos debe cumplir una canción para estar incluida en la lista de "éxitos"?

La pregunta será estudiada desde el punto de vista de los descriptores, tanto los propuestos en [2] como los mencionados en [3]. Debido a que se tendrá una idea de qué requisitos cumplen, se probará el modelo con una nueva canción para evaluar su efectividad. También se planteará la pregunta:

¿Una canción popular en un determinado año, ¿será popular en los años siguientes?, ¿lo hubiese sido en años anteriores?

por lo que se estudiará la modelación para más de un periodo de tiempo.

II. Descripción del Problema

Se utilizará la base de datos de Spotify para obtener el top 100 de las canciones más escuchadas a nivel mundial en el año 2020. Para esto, se accedió a la base de datos mediante la API en modo desarrollador, donde es posible obtener una muestra de 30 segundos de cada canción del top100, su posición en el ranking y una serie de descriptores clásicos. El problema a resolver fue, tomando como base la canción, obtener en primer lugar los 12 descriptores que proponemos, los que se basan en características matemáticas de la forma de onda de cada canción. Estos descriptores fueron unidos a una lista de 13 descriptores que entrega Spotify.

El objetivo será obtener la proporción de aquellos descriptores que permitan generar restricciones que encierren las características mínimas (apreciables o no) para determinar el "éxito" de una canción. Cada canción será categorizada según su lugar en el ranking, siendo la cercanía al primer puesto el nivel de "éxito" de la canción. Como se puede intuir, cuanto menor sea el valor, más exitosa será, por lo que tratará de un problema de minimización. Para abordar el problema se escoge usar Programación Cónica de Segundo Orden (SOCP), debido a la forma "descendente" que tendrán los descriptores respecto a su valor en el ranking (cuanto menor sea su lugar, su relevancia en el modelo será mayor). La idea detrás de la elección del SOCP es la generación de un cono de segundo orden en n dimensiones (siendo n el número de descriptores), que contenga los valores descriptivos de cada canción. Cuanto más restrictivo sea el espacio generado por el cono en una característica específica, más relevante será cumplir los requisitos de esta para poder estar en una lista de éxitos, lo que se traduce en una condición razonable para tener "éxito" (aunque como se señaló al comienzo, esto dependerá de muchos más factores, y solo intentaremos reducir la complejidad del problema).

III. Modelo matemático

Se propone la implementación de un modelo SOCP para generar restricciones que determinen la elegibilidad de una canción. El modelo presentado es el siguiente

A. Notación

1) Indices
   - i = índice del descriptor i.
   - j = índice de la canción j.
2) Conjuntos
   - $\mathcal{I}$: Conjunto de descriptores de audio extraidas de una canción.
   - $\mathcal{J}$: Conjunto de canciones ordenadas según el *ranking top*.
3) Parámetros
   - $\alpha_{i,j}$ = Descriptor i-ésimo de la canción j-ésima.
   - z_j = Posición en el *ranking* de la canción j.
4) Variables
   - $x_{i,j}$ = Coeficiente mínimo del descriptor i-ésimo de la j-ésima canción.
   - $y_{i,j}$ = Coeficiente máximo del descriptor i-ésimo de la j-ésima canción.

En este punto es importante aclarar el parámetro z_j. Este valor va asociado a una canción específica, y representa la posición en el ranking que tiene la canción. Este valor

B. Modelo

$$\min \quad \sum_{i\in\mathcal{I}} y_i - x_i \tag{1a}$$

$$\text{subject to} \quad z_j \geq \sqrt{\sum_{i\in\mathcal{I}} \left(\alpha_{i,j} \cdot x_i\right)^2} \quad \forall\, j \in \mathcal{J} \tag{1b}$$

$$x_i \geq 0. \tag{1c}$$

La función objetivo (1a) será la suma de los pesos de los descriptores que encierra a los puntos. Se busca minimizar dicho valor para que la región contenida sea lo más precisa posible. La restricción (1b) representa justamente esta situación. En primer lugar pone un límite inferior a la región al intentar maximizar los coeficientes (esto es, encontrar el cono más pequeño que contenga todos los datos) Por último, la restricción (1d) asegura la positividad de los valores, pues solo trabajamos con descriptores positivos.

IV. Experimentos computacionales

En esta sección se expone el método de resolución problema propuesto anteriormente. También se presentan los datos utilizados y los experimentos realizados.

A. Metodología

Se extrajeron datos de 30 segundos de canciones del top100 del año 2020, en primera instancia pues es la cantidad disponible en la API, y en segundo lugar, esta cantidad resulta representativa para cada canción. Como señalan Huihui et al [7], no es necesario extraer los descriptores de toda la canción, sino que basta con escoger 30 segundos representativos, debido a que por lo general, las canciones tienen muchos segmentos repetidos o similares en términos de su estructura. Los datos que no calzan en esta categoría son limitados, por lo que serán considerados como outlayers.

1) Datos Utilizados: Los descriptores, *rankings* y parámetros que se utilizaron en los experimentos provienen de dos fuentes:

- Bases de datos externas:
  Se utilizó el conjunto de datos de *Top Spotify Tracks of 2020* [8]. Como se mencionó antes, esta contiene una

tabla con las 100 canciones más escuchadas de *Spotify* del año 2020, en conjunto a 13 descriptores de cada una, entre los que se encuentran algunos propuestos por [2]. Las características incluidas son: danzabilidad, energía, tono, volumen, modalidad (menor o menor, en el sentido armónico), presencia de palabras habladas, acusticalidad, instrumentalidad, vivacidad, valencia, tempo, duración y estructura de tiempo de los compases. Las interpretaciones de estos descriptores están descritas por McIntire [6]. Con esto, se tiene una matriz de 100 filas que representan cada canción y 13 columnas de descriptores asociadas a estas. Cada fila ordenada con su posición en el *ranking* de canciones más escuchadas.

- Descriptores calculados: Se creó una base de datos de 11 descriptores calculados para cada una de las 100 canciones mencionadas anteriormente, la cual está compuesta por: el centroide, media, varianza, oblicuidad, kurtosis, valor RMS, amplitud máxima, razón de cruces por cero, centroide espectral, pendiente espectral y planicie espectral. Las interpretaciones de estos descriptores están descritas en [3], sin embargo, capturan puntos claves que describen analíticamente una parte de la canción.

2) Procedimiento: Los distintos pasos y ambientes que se utilizaron se muestran en el diagrama de bloques de la figura 1. Este muestra el flujo de pasos para el proceso de extracción de descriptores propios. El método que se utilizó con la base de datos de *Spotify* es análogo, sólo que no cuenta con los dos primeros pasos de lectura de la señal y la extracción de sus características, ya que estas venían precalculadas. En la figura también se especifican los datos de entrada y de salida de cada paso.

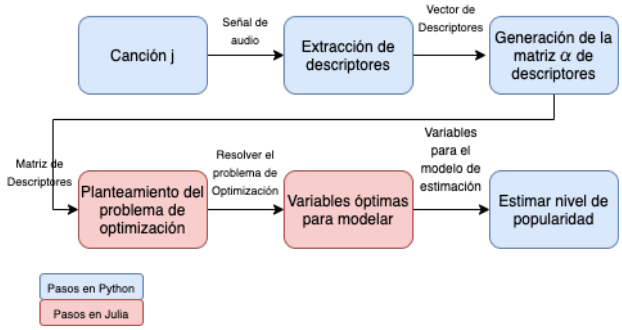

Fig. 1: Procedimiento del método.

Los descriptores fueron ingresados a Julia y mediante Ipopt se resuelve el modelo planteado en la sección anterior.

3) Ambiente Computacional: La obtención de las características y manejo de los datos fueron programados en rutinas de *Python 3.7*. Por otra parte, el modelo de optimización y su solución fueron calculados con la librería *Ipopt* de *Julia 1.6*.

V. Resultados

A. Valores obtenidos

Para el conjunto de datos de *Spotify* se obtuvieron los siguientes vectores de coeficientes óptimos para el cono inferior:

$$x_{Spotify} = \begin{bmatrix} 0.2819 \\ 0.1626 \\ 0.0184 \\ 0.0040 \\ 0.1131 \\ 0.7841 \\ 2.5237 \\ 4.7999 \\ 8.4720 \\ 0.2382 \\ 1.1124 \cdot 10^{-5} \\ 1.5555 \cdot 10^{-12} \\ 0.0066 \end{bmatrix}$$

Mediante las ideas propuestas por Mehta et al[5], una descomposición *SVD* sería práctica para casos como los de los descriptores 11 y 12, cuyos valores son de órdenes casi despreciables. Estos dos descriptores corresponden al tempo y a la estructura de los compases, los cuales (al menos para 2020) tienen muy poca relevancia en términos del éxito.

Cada vector representa los coeficientes que genera el cono más restrictivo que contiene los descriptores de las canciones. Con esto se busca delimitar la región en la cual se espera tener una canción exitosa, en base a sus descriptores.

A continuación, se muestra el cono para una de las dimensiones de los descriptores. Específicamente, en la figura 2 se muestra el de la danzabilidad del top 100 de *Spotify*. Se puede apreciar que todos los valores de este descriptor son positivos (motivo de la restricción de positividad). Además, se observa que la pendiente es muy baja (cercana a 0.2819), esto debido a que la danzabilidad resulta ser una característica común de los grandes éxitos.

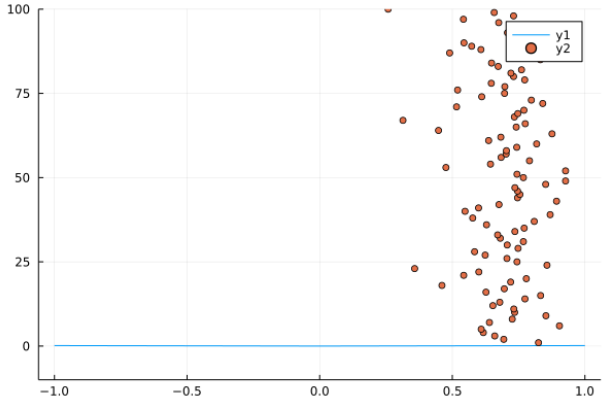

Fig. 2: Cono de la danzabilidad del top 100 de *Spotify*

Para otro descriptor, tal como la modalidad, se tiene la situación de la figura 3, en la cual se muestran 50 datos de canciónes con el objetivo de apreciar mejor las magnitudes, es posible observar una pendiente muy baja (específicamente, cercana a 0.1131). Esto quiere decir que los datos están aglutinados abajo.

En contraposición se tiene el caso de la vivacidad, cuya pendiente debe tener un valor de 8.4720, lo que se muestra

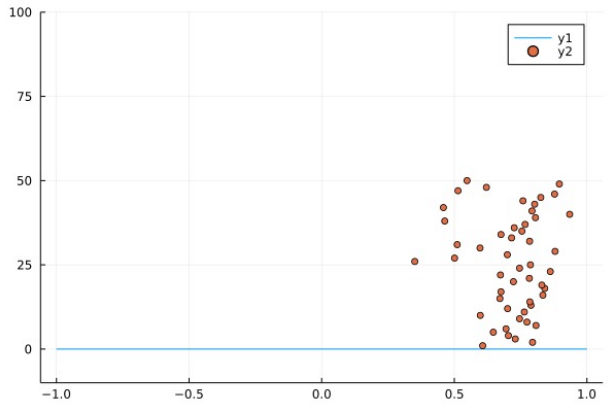

Fig. 3: Cono de la modalidad del top 100 de *Spotify*

en la figura 4, nuevamente con 50 datos. Esta pendiente es perceptiblemente mayor a la anterior, lo que implicaría que la vivacidad de una canción es un factor relevante al momento de determinar su éxito.

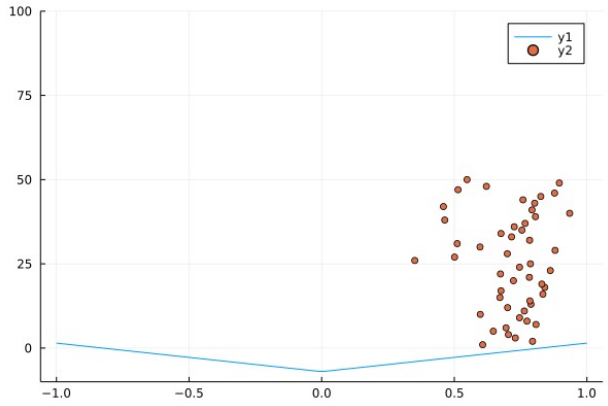

Fig. 4: Cono de la vivacidad del top 100 de *Spotify*

Las figuras 5, 6 y 7 representan respectivamente la energía, el modo y el volumen de las características propuestas. Se aprecia una gran variedad de pendientes, lo que representa las restricciones de algunas dimensiones del problema.

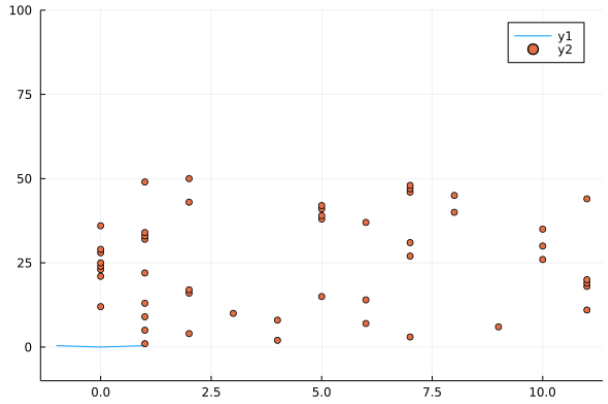

Fig. 5: Cono de la energía del top 100 de *Spotify*

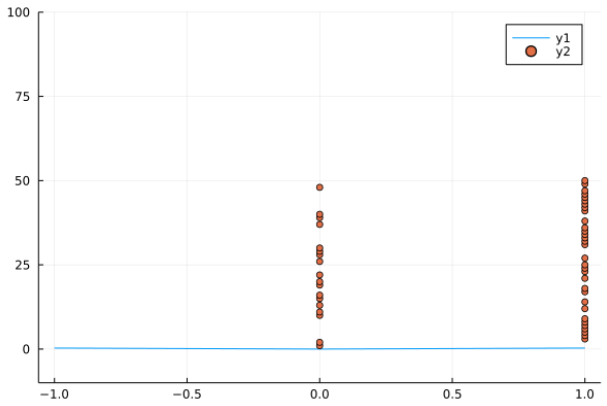

Fig. 6: Cono del modo (mayor o menor) del top 100 de *Spotify*

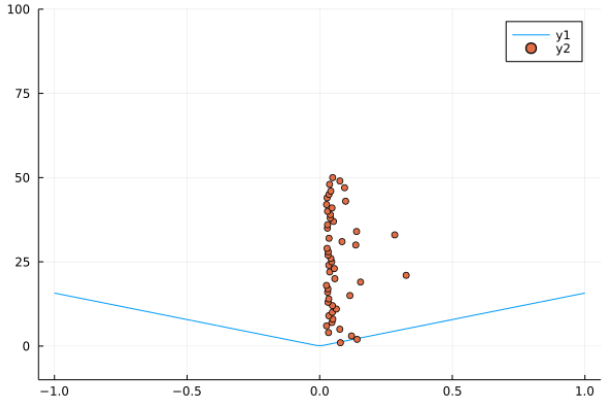

Fig. 7: Cono del volumen del top 100 de *Spotify*

Para evaluar el modelo, se realiza el producto punto entre las features de una canción específica y la concatenación de los coeficientes. Esto es probado con la canción la canción ”Safaera”, del artista Bad Bunny, cuyos datos normalizados (dividiendo por la norma del vector producto) lo ubica en la posición $40, 9$, o 41. La posición de esta canción en dicho año fue la 43, lo que resulta una estimación bastante adecuada.

La segunda parte del vector corresponde a nuestros descriptores, en cuyo caso los valores obtenidos fueron:

$$x_{Propios} = \begin{bmatrix} 0.2860 \\ 0.1307 \\ 0.3047 \\ 0.0046 \\ 0.0871 \\ 15.728 \\ 4.6822 \\ 40.422 \\ 0.7930 \\ 0.2771 \\ 9.4973 \cdot 10^{-6} \\ 3.4142 \cdot 10^{-12} \\ 0.0054 \end{bmatrix}$$

los que corresponden a los descriptores señalados en la sección anterior. Repitiendo el experimento anterior, la canción ”Safaera” queda ubicada en la posición 171, lo que ni siquiera

calza en el top 100. Si se unen los vectores se obtiene una posición relativa de $11,4$. Esto se debe a que las magnitudes son distintas entre sí para los descriptores. Finalmente, un gráfico para estos descriptores se muestra en la figura 8

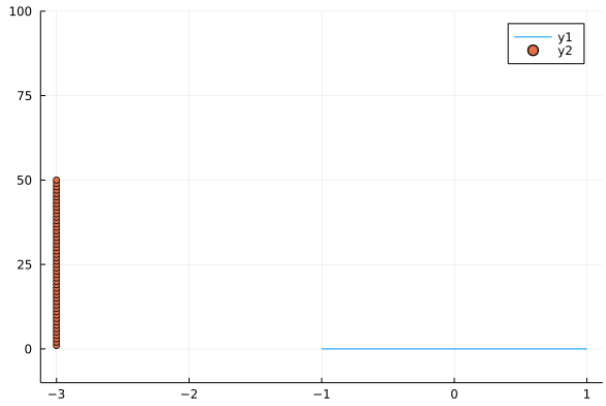

Fig. 8: Cono de Skewness del top 100 de *Spotify* 2010

B. *Comparativa con décadas anteriores*

Se repitió el mismo experimento para periodos de tiempo anteriores. Específicamente, replicando lo propuesto por Bpurreau [9], se escogieron periodos en lugar de años, y específicamente optamos por décadas. La década del 2000 tiene el siguiente vector relativo:

$$(0.216,\ 0.159,\ 0.037,\ 0.002,\ 2.612,\ 11.943,\ 0.312,\ 111.888$$

$$0.902,\ 0.111,\ 1.160 \cdot 10^{-5},\ 1.665 \cdot 10^{-12},\ 0.004)$$

cuya norma es 112.55. Lo respectivo para los 2010, 90 y 80 da una norma de 37.09 39.5 y 42.13 . La misma norma calculada para el año 2020 tiene un valor de 1296.15. Esto muestra una gran variabilidad entre años, y un comportamiento creciente en el último periodo (aunque la falta de datos no permite sacar una buena conclusión). Una gráfica para la acusticidad de la década del 2010 es mostrado en la figura 9.

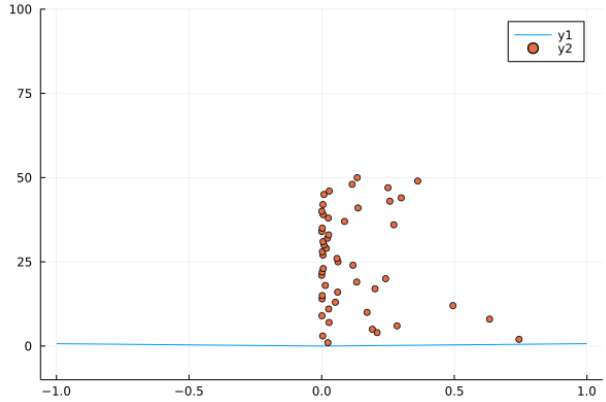

Fig. 9: Cono del acusticidad del top 100 de *Spotify* 2010

C. *¿La música se está homogeneizando?*

Según un estudio realizado por Bourreau et al [9], la música más escuchada a nivel mundial está perdiendo su estabilidad, lo que significaría que las características en su acusticidad están volviendose más variadas cada vez. Nuestros datos señalan que, en décadas anteriores, los valores son significativamente menores, como se muestra en la figura 10

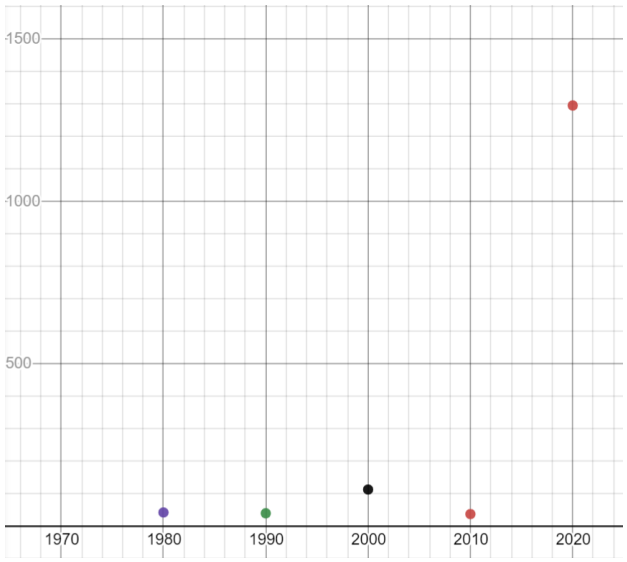

Fig. 10: Normas de *playlist*

La gráfica para 2020 puede tener un valor tan alto debido a tratarse de un periodo más corto, sin embargo, en los otros periodos puede verse una tendencia al alza (salvo en el 2010), por lo que nuestros datos podrían estar correlacionados con lo propuesto por Boureau.

VI. Conclusiones

Las características de la música escuchada por la gente son un asunto sumamente complejo. Al no poder abordarlo completamente, se utilizan distintas formas de abordaje del problema, entre las cuales algunas resultan más convenientes. El presente trabajo experimentó con la inclusión de métricas cuantitativas para intentar describir los gustos populares de mejor manera, y, de ese modo, predecir el éxito de una canción. Pese a lo anterior, las métricas analíticas probaron ser inferiores a las ya utilizadas por grandes empresas como Spotify, debido a que su capacidad de predicción deja mucho que desear. Esto puede deberse a factores como la alta variabilidad en los ritmos de la época en que fue escrito este trabajo, sin embargo, sería interesante repetir el experimento en épocas más homogeneas como las de principio de siglo.

Una conclusión interesante obtenida, es que la música popular sí presenta una mayor dispersión en cuanto a sus características conforme avanza el tiempo, tal como lo propone Bourreau, no obstante, esta no resulta concluyente debido a casos como la década del 2010. En nuestro trabajo probamos que la norma del cono que los contiene no necesariamente sube, por lo que su dispersión (y por lo tanto sus diferencias,

tanto en nuestros predictores como en los clásicos de Spotify) no sería creciente en el siglo pasado. Sin embargo, el año 2020 pareciera ser una excepción notable, por lo que se necesitará información futura para conseguir una respuesta más clara.

VII. Distribución del trabajo

El trabajo fue colaborativo en casi todo el proceso. Rodrigo abarcó mayormente temas de programación, de obtención y tratamiento de datos. Mientras que Ignacio tocó temas de bibliografía y modelación, sin embargo, el trabajo fue realizado con ambos presentes, por lo que fue repartido e informado constantemente.

References

[1] J. Lena and R. Peterson. Classification as Culture: Types and Trajectories of Music Genres. *American Sociological Review*, vol. 73, pp. 697-718, 2008.

[2] N. Askin and M. Mauskapf. What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music. *American Sociological Review*, vol. 82, no. 5, pp. 1-35, 2017.

[3] C. Weihs, D. Jannach, I. Vatolkin, G. Rudolph. Music Data Analysis: Foundations and Applications. *Chapman & Hall/CRC*, ed. 2, pp. 147–162, 2017.

[4] S. Maldonado and J. Lopez. Imbalanced data classification using second-order cone programming support vector machines. *Pattern Recognition*, vol. 47, no. 5, pp. 2070–2079, 2014.

[5] R. Mehta. Session Based Music Recommendation using Singular Value Decomposition (SVD). *San Jose State University Scholar Works*, Master's Projects. 237., 2012.

[6] G. McIntire. A Machine Learning Deep Dive into My Spotify Data, Version 1. Consultado el 20 de Septiembre, 2020 de *https://opendatascience.com/a-machine-learning-deep-dive-into-my-spotify-data/*, 2017.

[7] H. Huihui, X. Luo, T. Yang and S. Youqun. Music Recommendation Based on Feature Similarity. *IEEE International Conference of Safety Produce Informatization*, 2018.

[8] J. Song. Spotify DataSets, Version 1. Consultado el 20 de Septiembre, 2020 de *https://www.kaggle.com/jsongunsw/spotify-datasets*, 2017.

[9] M. Bourreau. Does digitization lead to the homogenization of cultural content?. *Economic Inquiry*, 2021.