Perreología

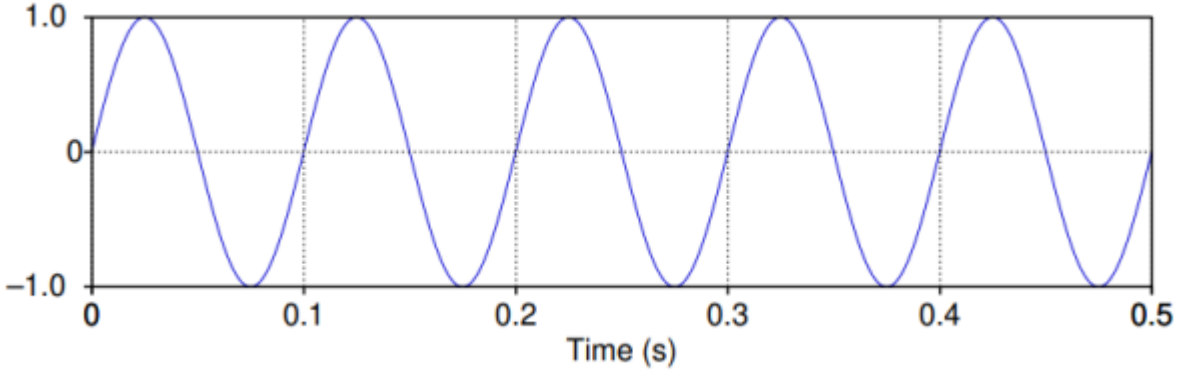
¿Cómo podemos usar la música en análisis estadísticos?

Metodología. ¿Qué es el sonido?

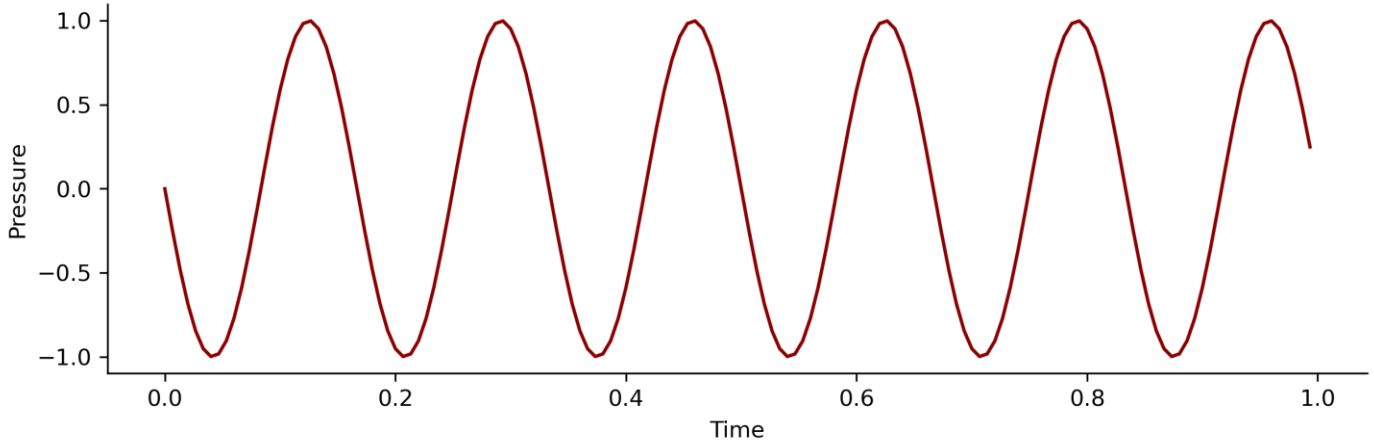
El sonido es una onda producida por variaciones en la presión del aire.

Thomas Edison inventa el fonógrafo en 1887

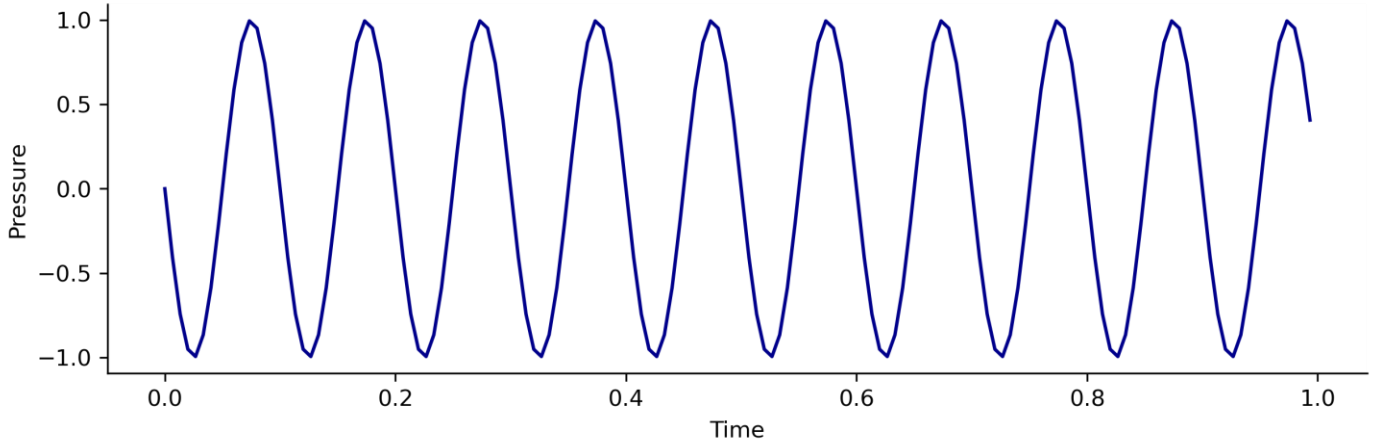
Onda con frecuencia 10 hz y amplitud de 1.



Metodología. ¿Qué es el sonido?

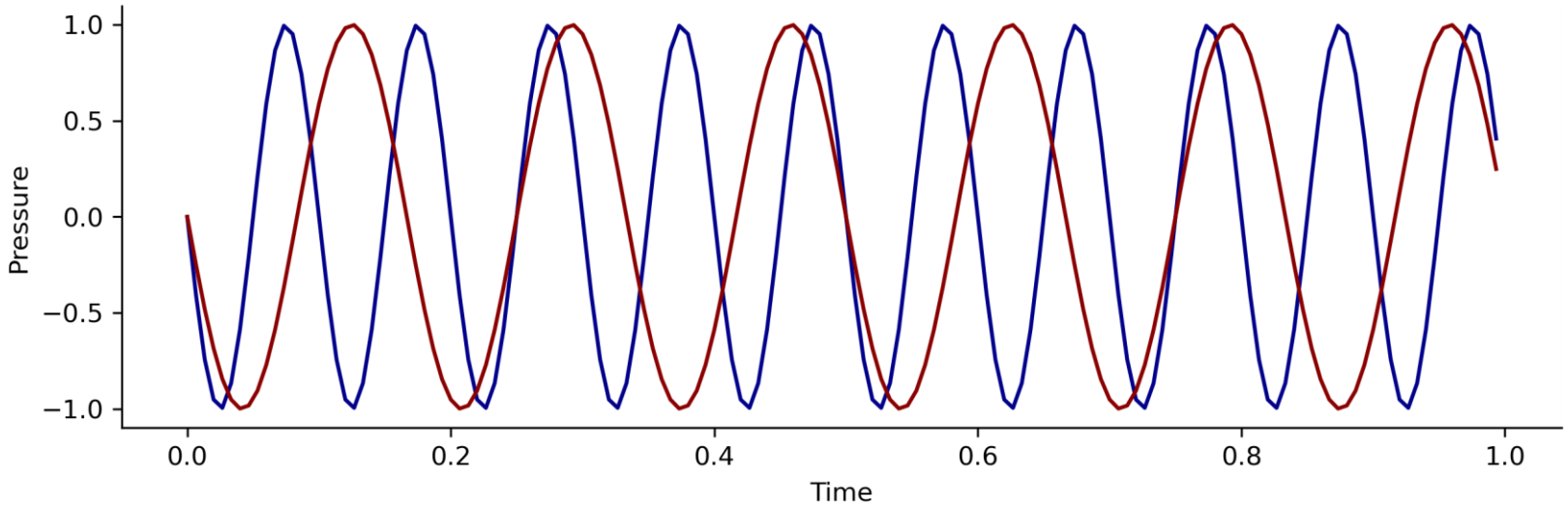


Re – 294 Hz



La – 440 Hz

Metodología. ¿Qué es el sonido?

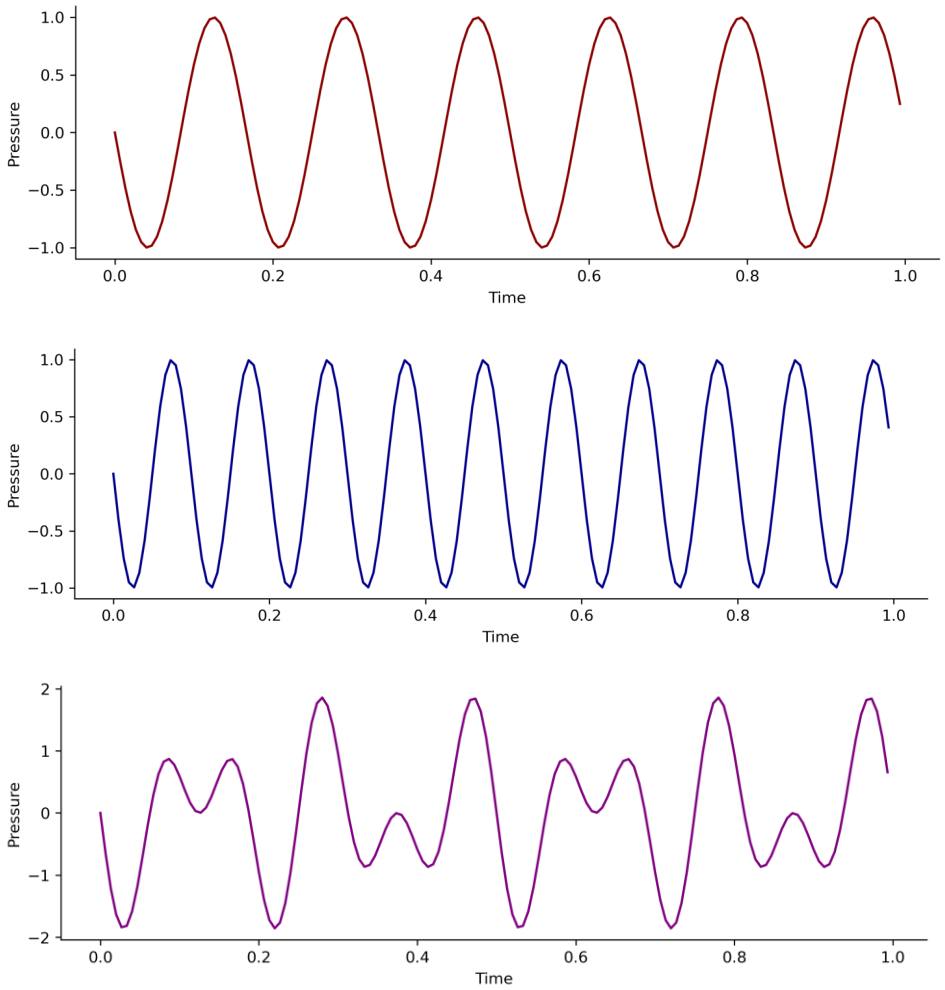


Re – 294 Hz

+

La – 440 Hz

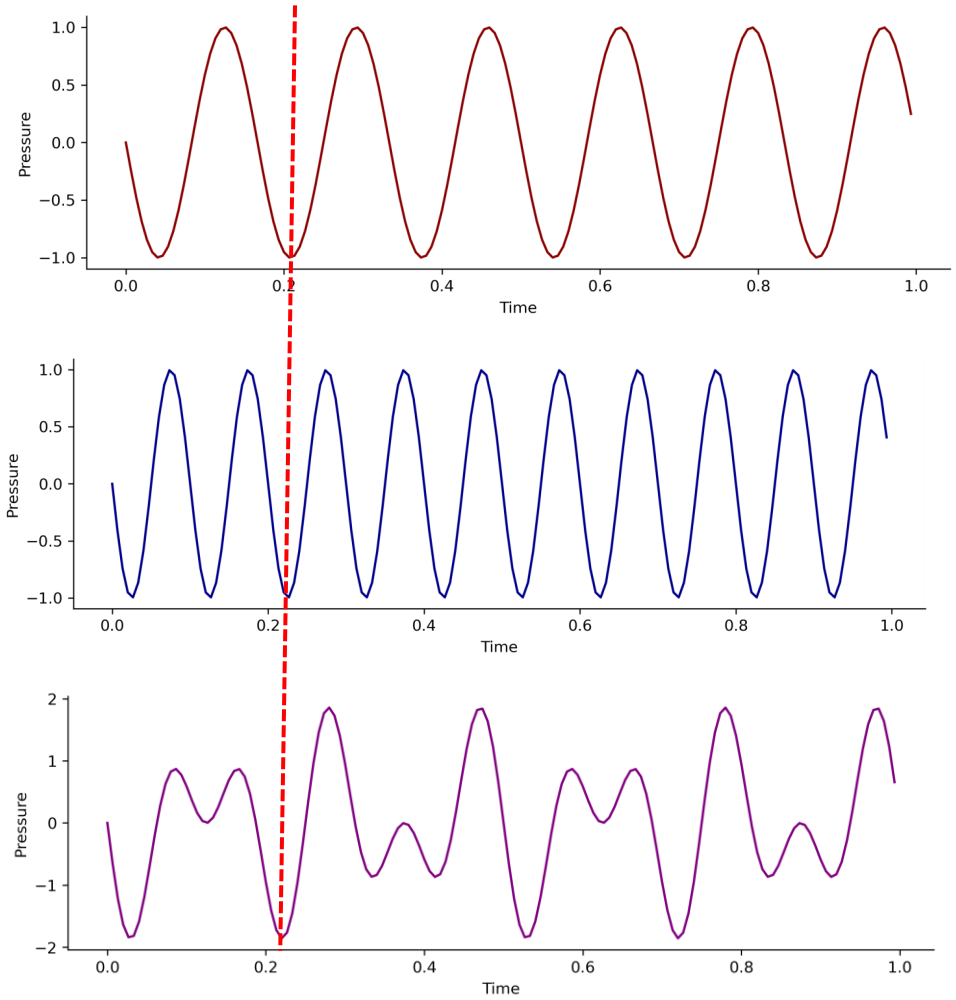
Metodología. ¿Qué es el sonido?



Re – 294 Hz

La – 440 Hz

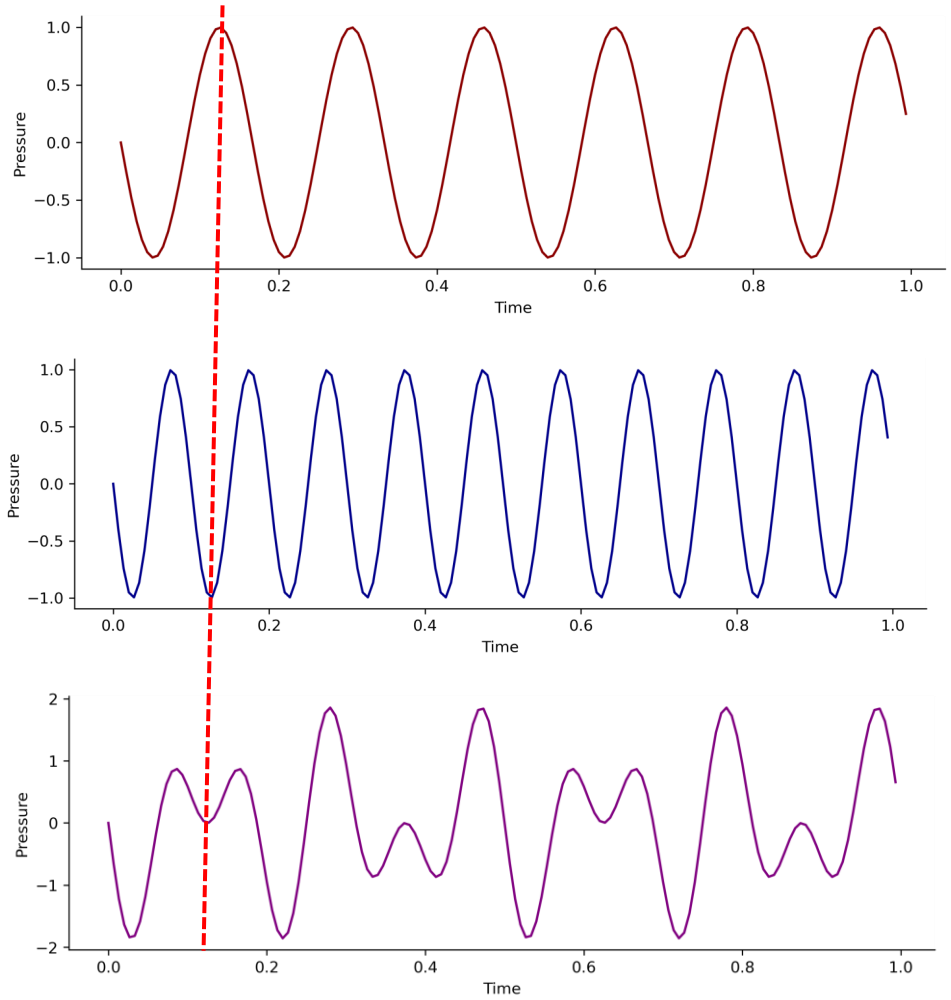
Metodología. ¿Qué es el sonido?



Re – 294 Hz

La – 440 Hz

Metodología. ¿Qué es el sonido?

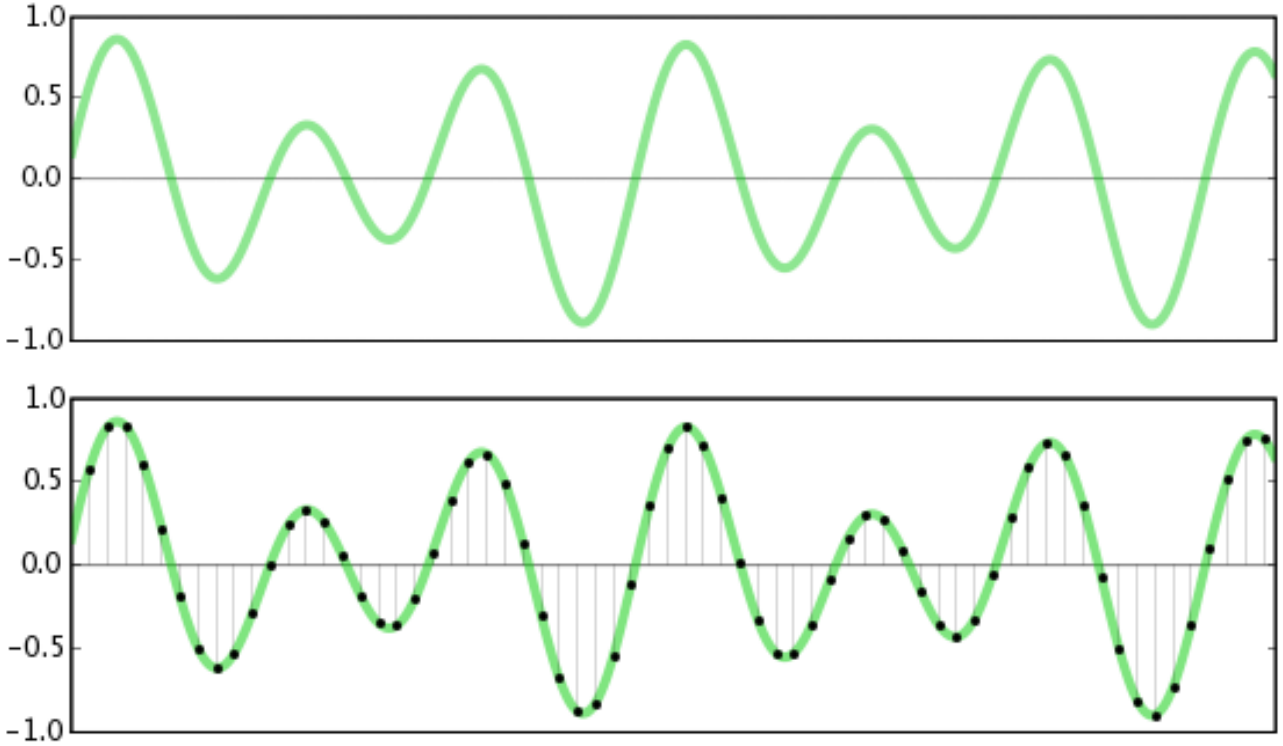


Re – 294 Hz

La – 440 Hz

Metodología. ¿Cómo se almacena el sonido?

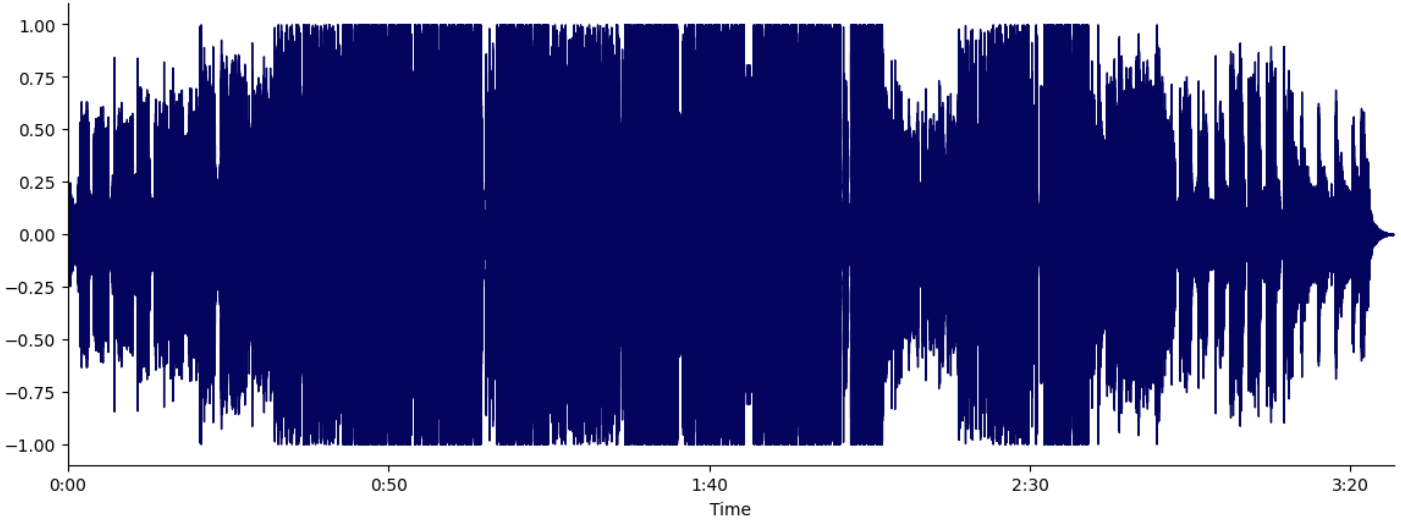
La grabación digital mide en intervalos periódicos el valor de la onda acústica para capturar su forma.



Metodología. ¿Cómo entienden los computadores el sonido?

La canción *Yonaguni* del artista Bad Bunny dura 3:27. Con un sr de 48,000 Hz, esta se representa como una cadena de:

$$207 \times 48,000 = 9,936,000 \text{ dígitos}$$

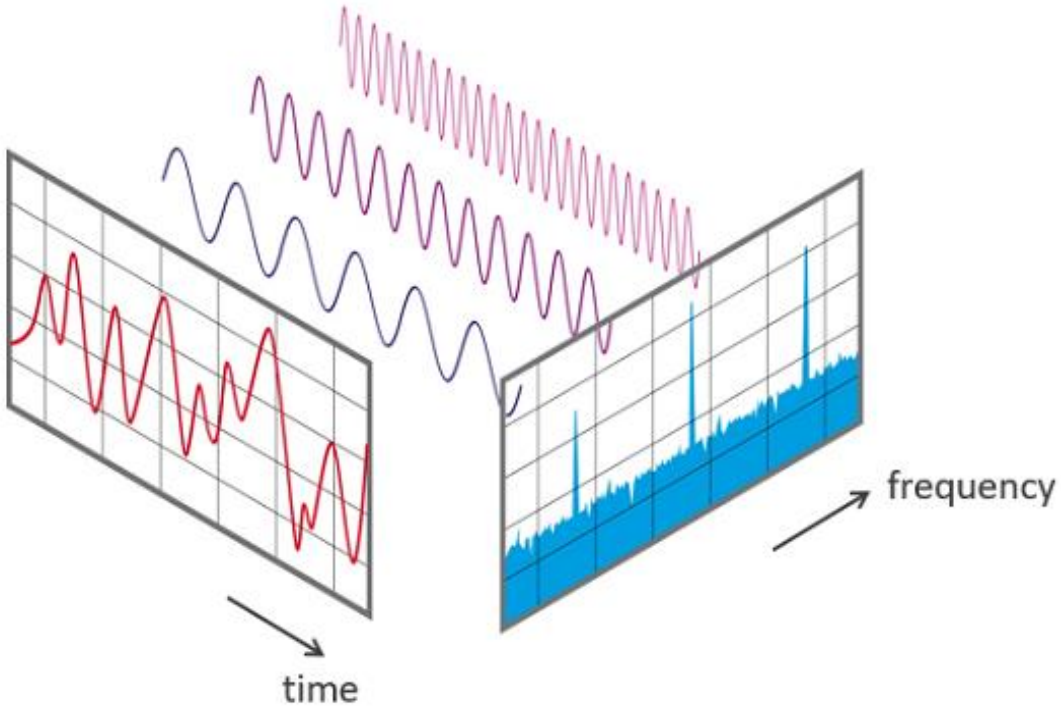


Metodología. Transformación de Fourier Discreta (DFT)

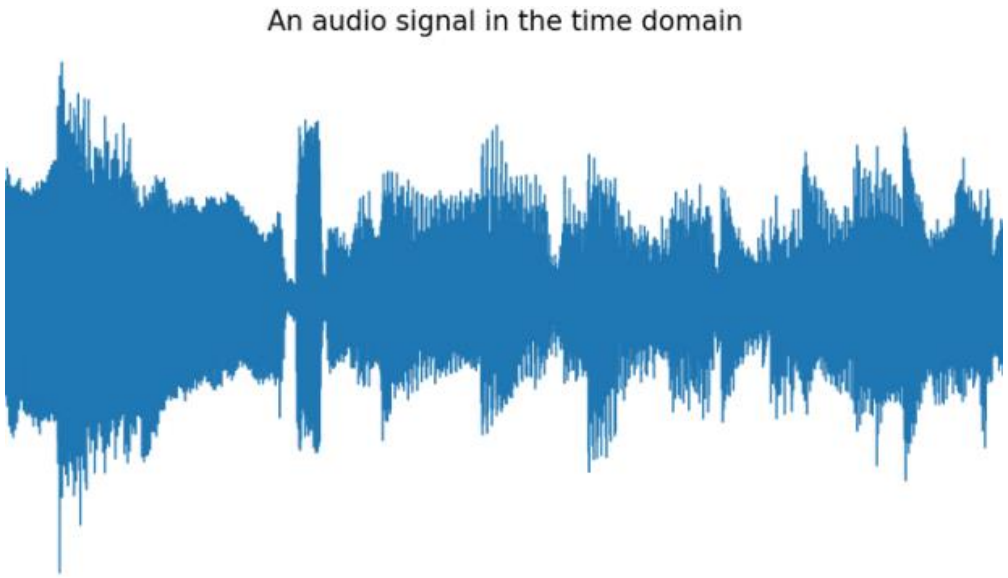
Vamos a descomponer el sonido en sus frecuencias constituyentes

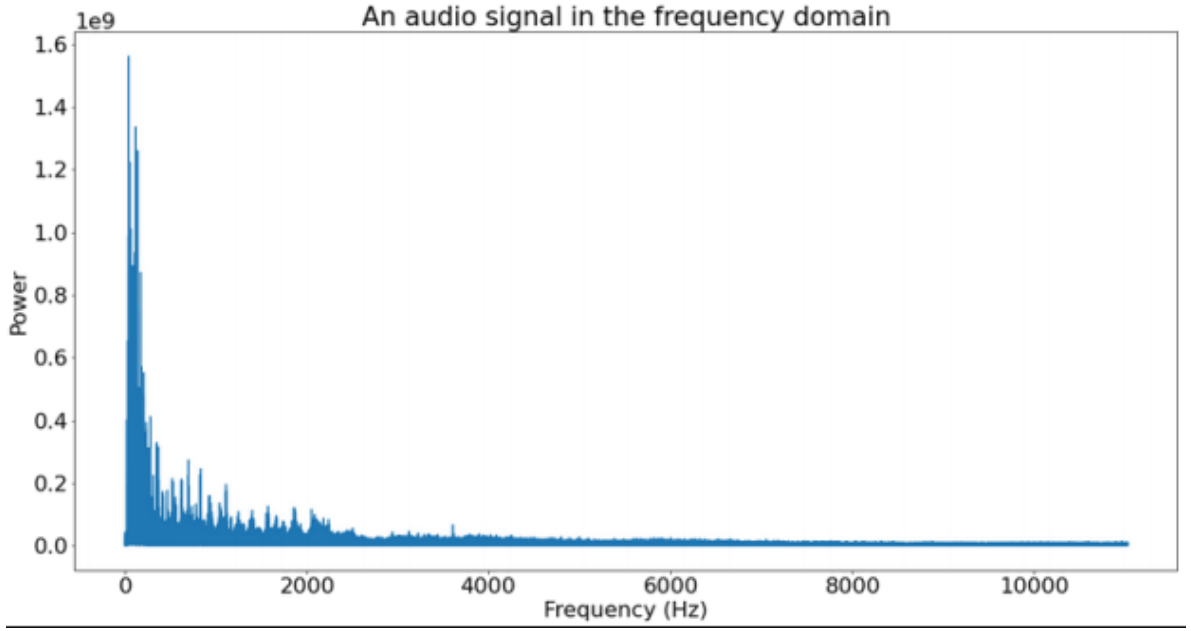
$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{2\pi i}{N} kn}$$

$$X_k = \sum_{n=0}^{N-1} x_n \cdot \left[\cos\left(\frac{2\pi}{N} kn\right) - i \cdot \sin\left(\frac{2\pi}{N} kn\right)\right]$$

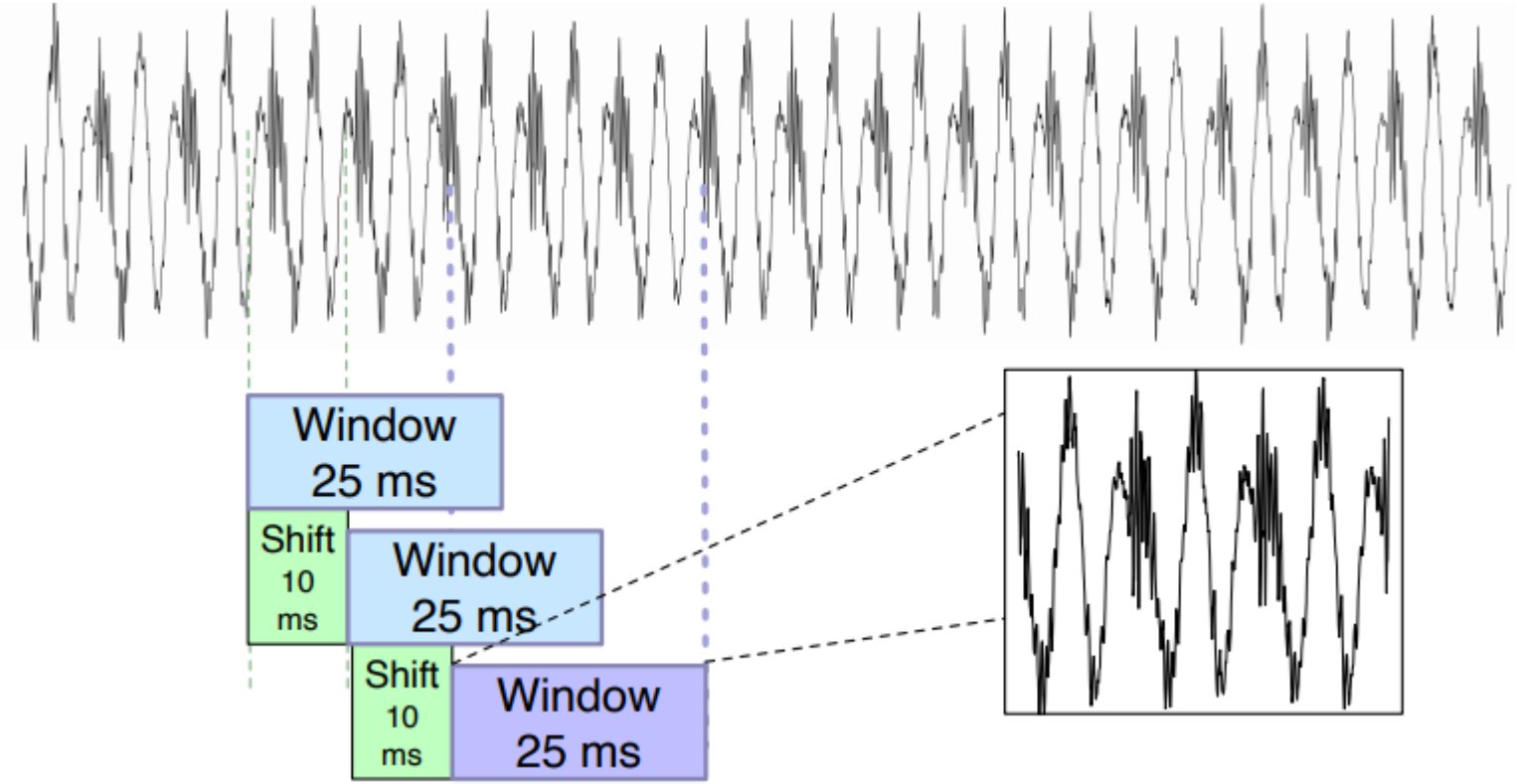


Metodología. Transformación de Fourier Discreta (DFT)





Metodología. Transformación de Fourier de Tiempo Corto (STFT)



Usaremos ventanas de 1,024 dígitos.

Los saltos serán de 512 dígitos.

Metodología. Espectrogramas

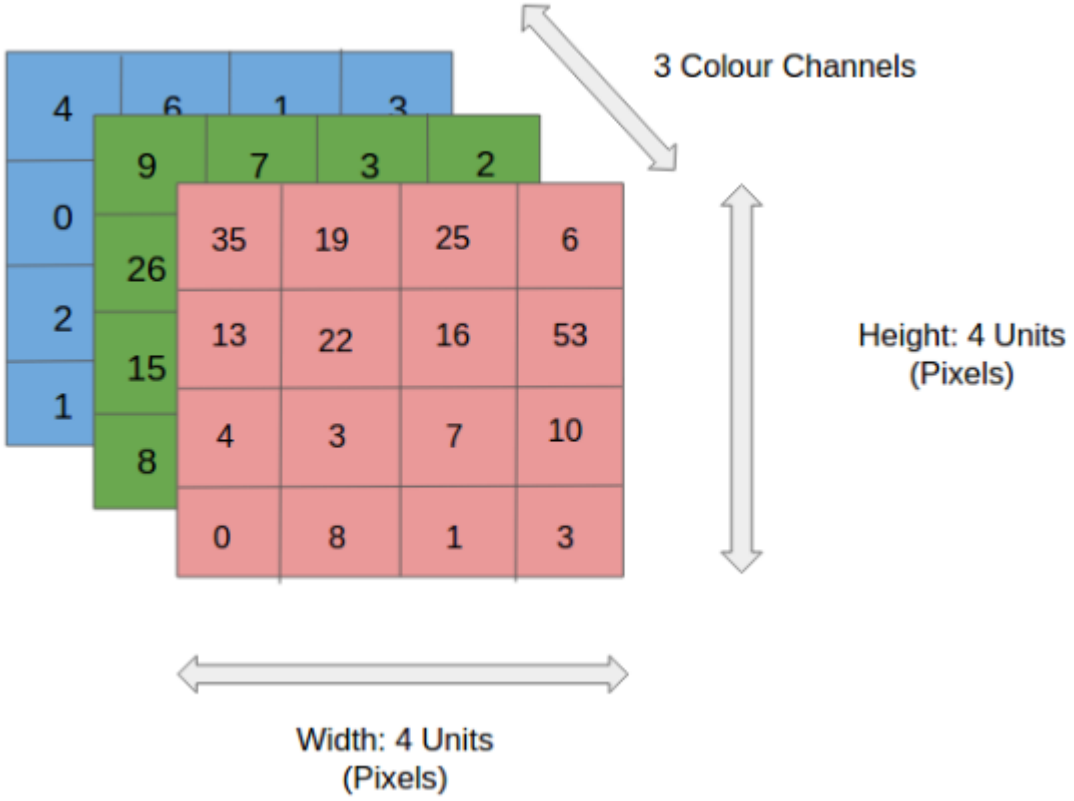
Frecuencia 1000 Hz	Intensidad: 10 dB	Intensidad: 0 dB	Intensidad: 0 dB
Frecuencia 800 Hz	Intensidad: 0 dB	Intensidad: 10 dB	Intensidad: 0 dB
Frecuencia 600 Hz	Intensidad: 5 dB	Intensidad: 5 dB	Intensidad: 5 dB
Frecuencia 400 Hz	Intensidad: 1 dB	Intensidad: 0 dB	Intensidad: 1 dB
	Segundo 0 - 9	**Segundo 10 - 19**	**Segundo 20 - 29**

Metodología. Espectrogramas

Espectrograma

10 dB	0 dB	0 dB
0 dB	10 dB	0 dB
5 dB	5 dB	5 dB
1 dB	0 dB	1 dB

Imagen



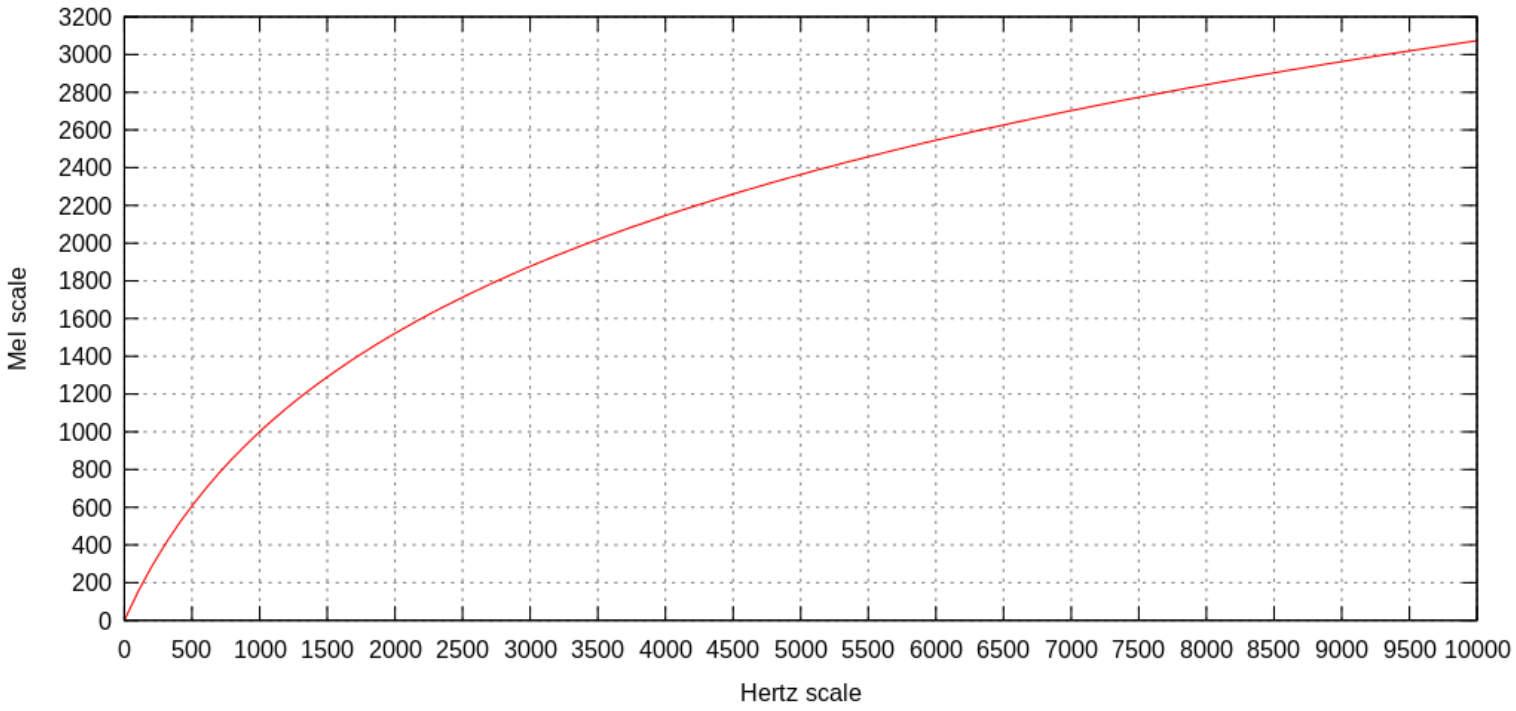
Metodología. Escala de Mel

100 Hz

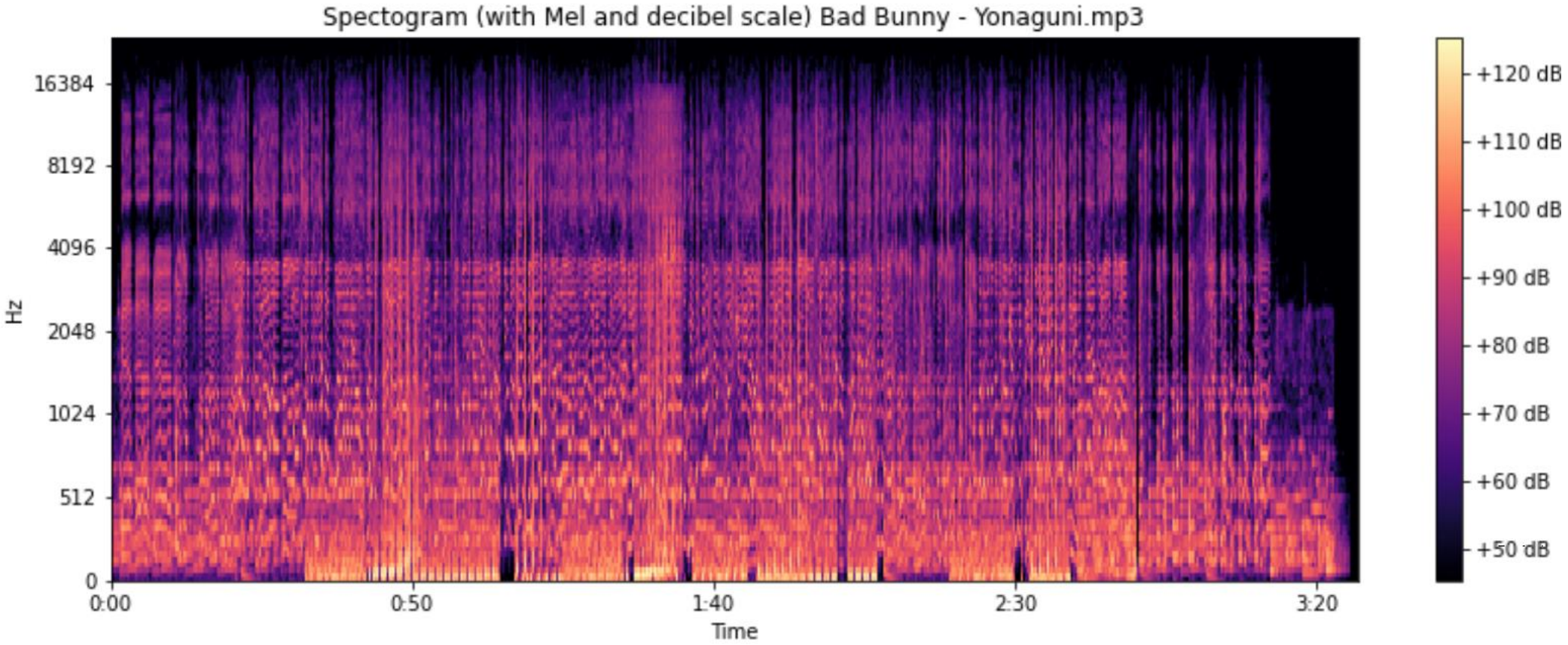
200 Hz

10,000 Hz

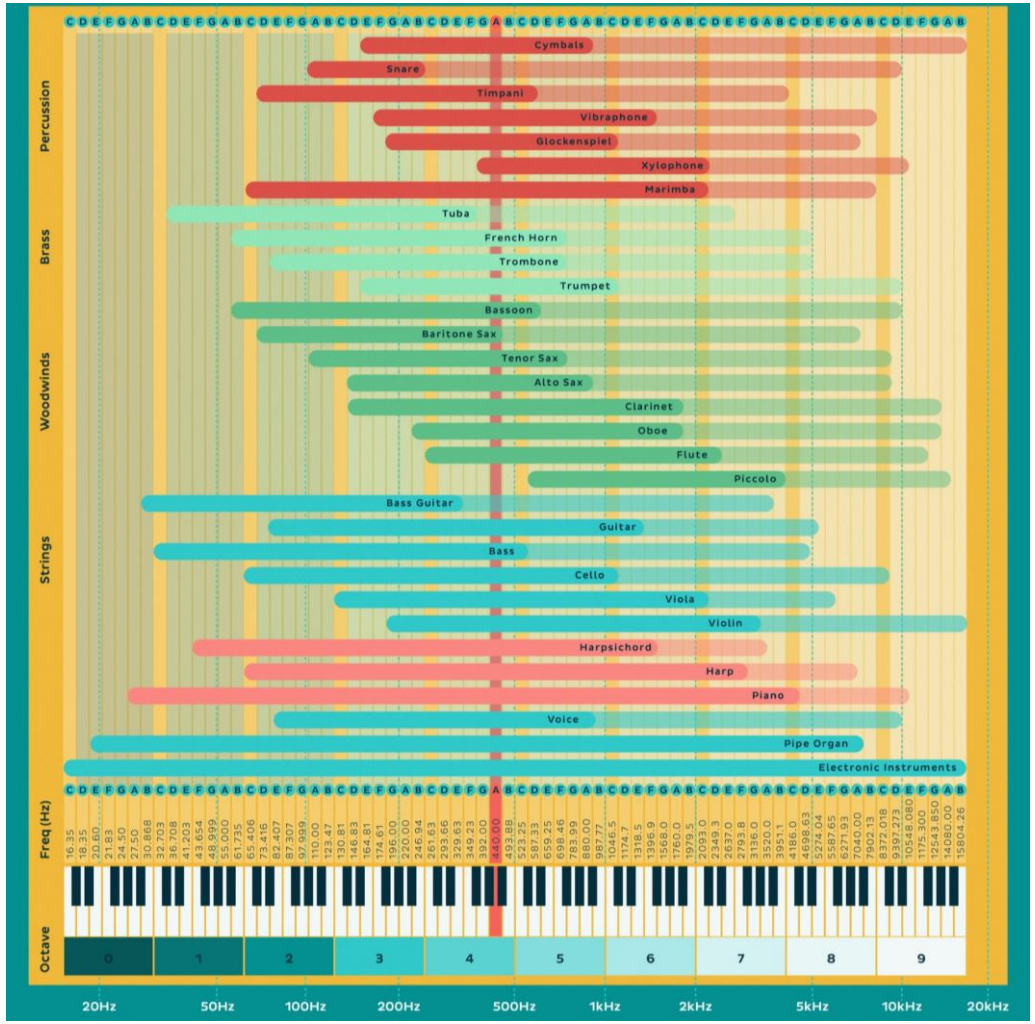
10,100 Hz



Metodología. Espectrogramas



Metodología. Espectrogramas



Metodología. Construcción de los *embbedings* musicales

- Hershey et al. (2016). CNN architectures for large-scale audio classification
- Redes Neuronales Convolusionales (CNN) para clasificar audio.
- Entrenaron arquitecturas con 5.24 millones de horas de audio extraído de vídeos de YouTube con sus respectivas etiquetas.

Arquitecturas	Tiempo de entrenamiento	AUC	mAP
Fully connected	35h	0.851	0.058
AlexNet	82h	0.894	0.115
VGG	184h	0.911	0.161
Inception V3	137h	0.918	0.181
ResNet-50	356h	0.926	0.212

Metodología. Construcción de los *embbedings* musicales

- Benavides et al. (2021): Visual representations to evaluate the heterogeneous effects of urban Parks restoration on crime.
- Representación vectorial de imágenes con CNN VGG19 pre entrenada.
- Modelo de tópicos LDA para darle interpretabilidad a los vectores.

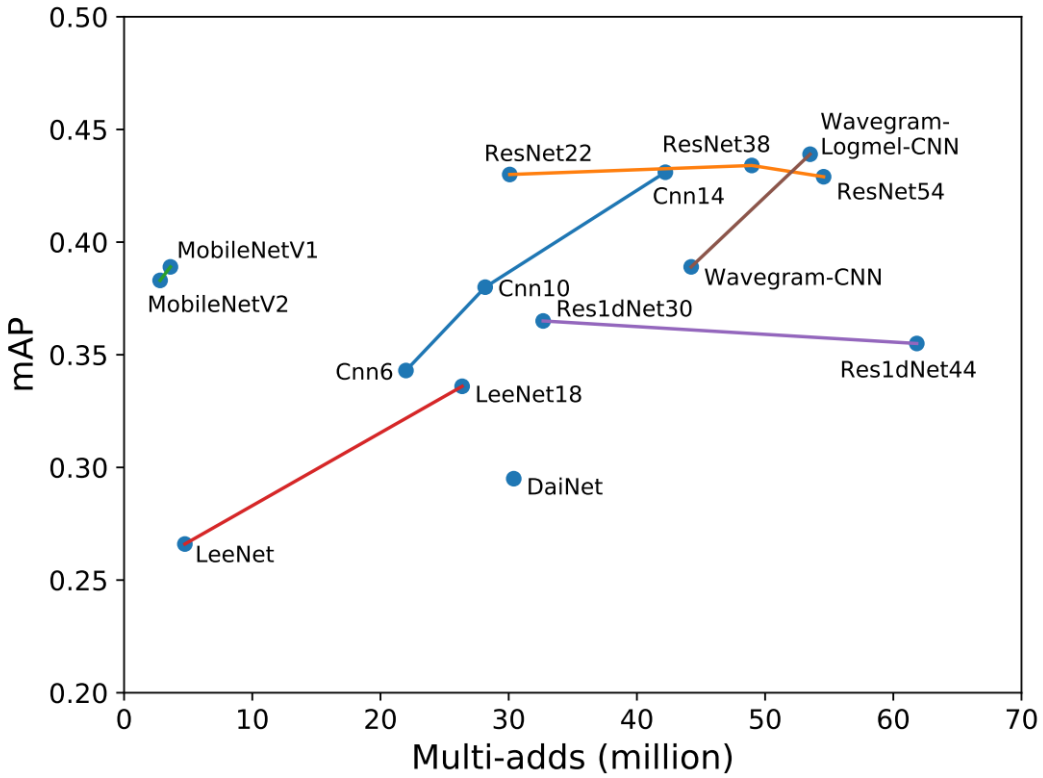
Metodología. Construcción de los *embbedings* musicales

- Acosta y Camargo (2019): Predicting city safety perception based on visual image content.
- Representación vectorial de imágenes con CNN VGG19 pre entrenada.

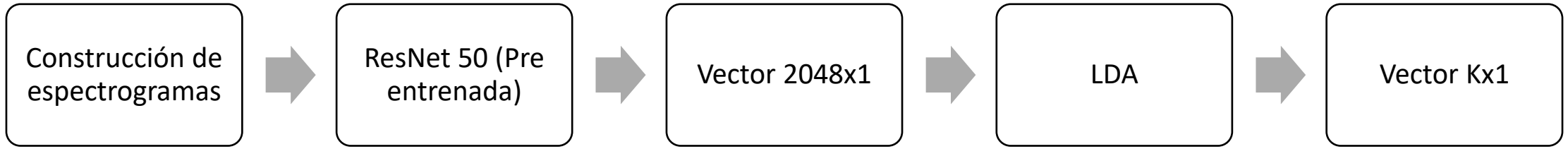
- Benavides et al. (2021): Visual representations to evaluate the heterogeneous effects of urban Parks restoration on crime.
- Modelo de tópicos LDA para darle interpretabilidad a los vectores. Estimación causal mediante Diff-in-Diff.

Metodología. Construcción de los *embbedings* musicales

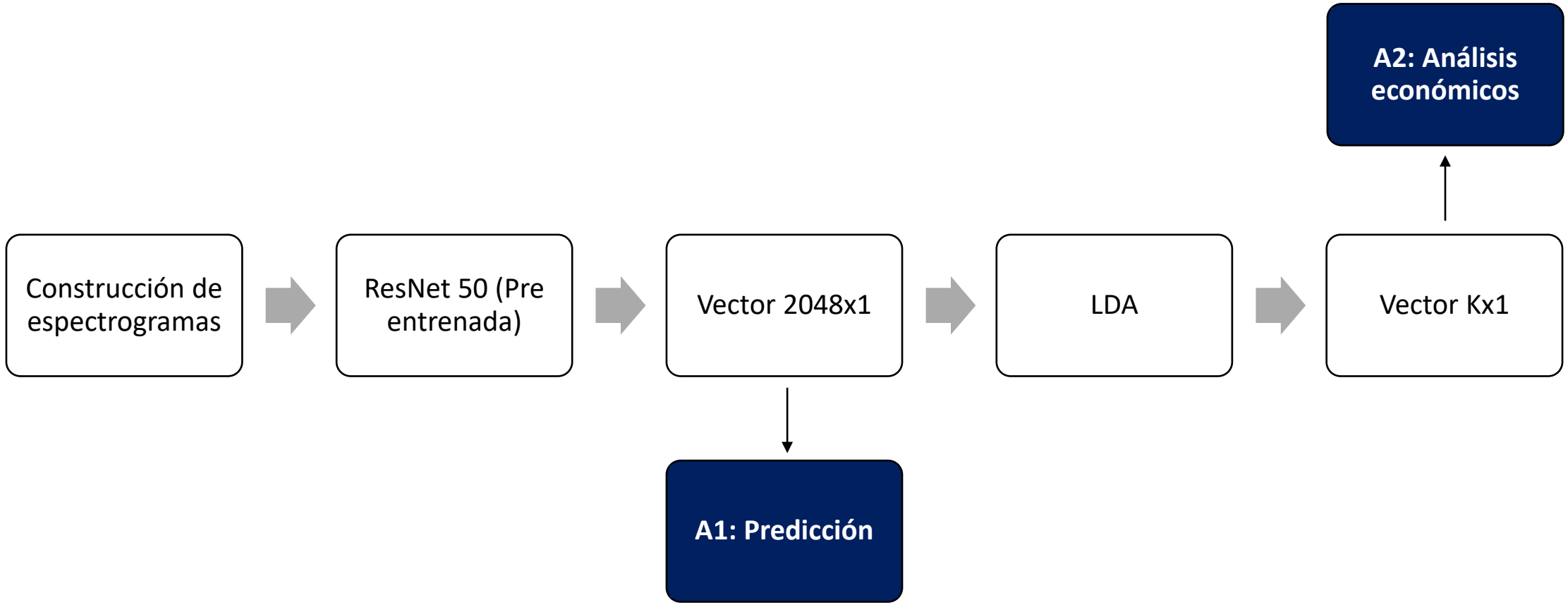
- Kong et al. (2020): Panns: Large-scale pretrained audio neural networks for audio pattern recognition.



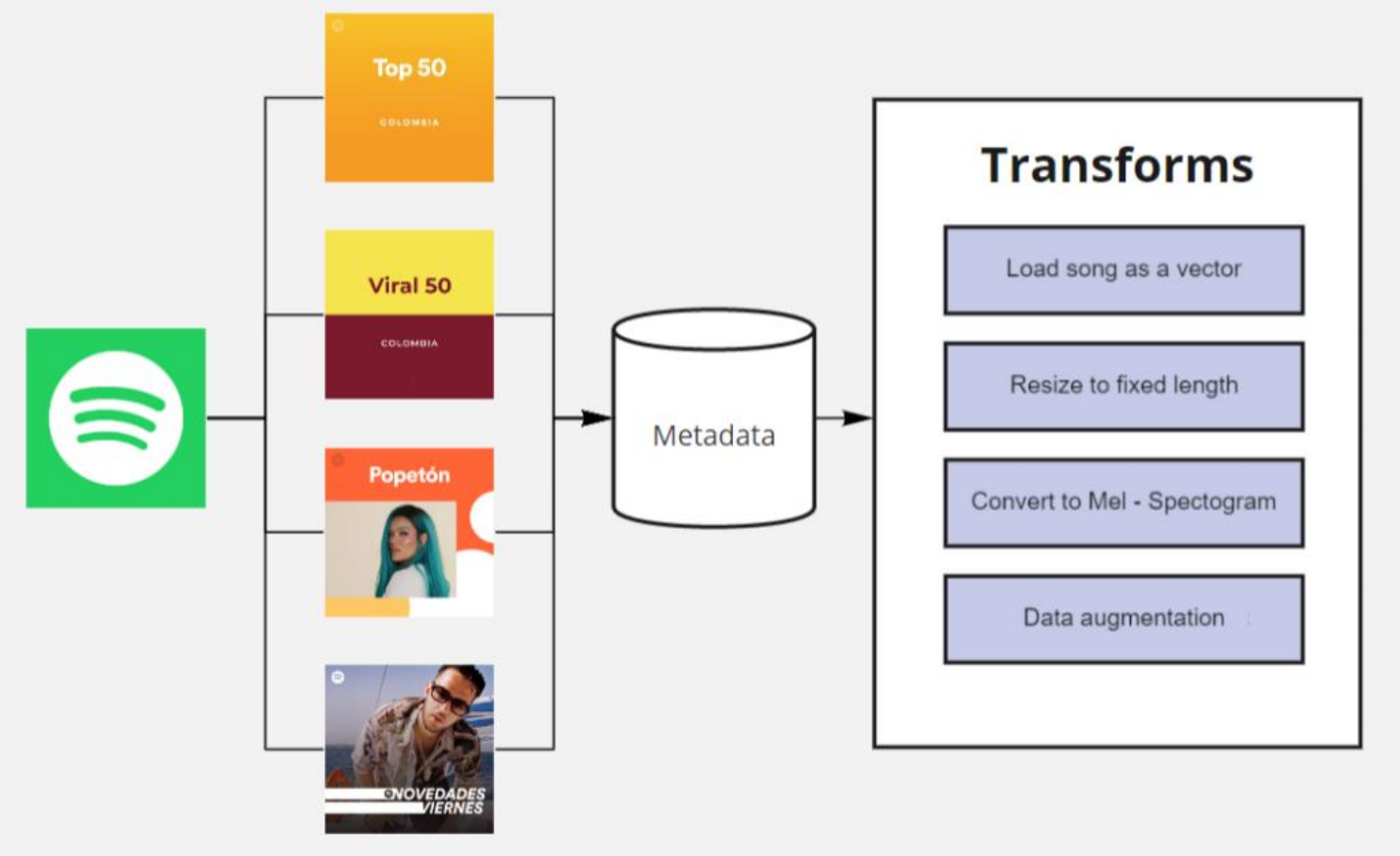
Metodología. Construcción de los *embbedings* musicales



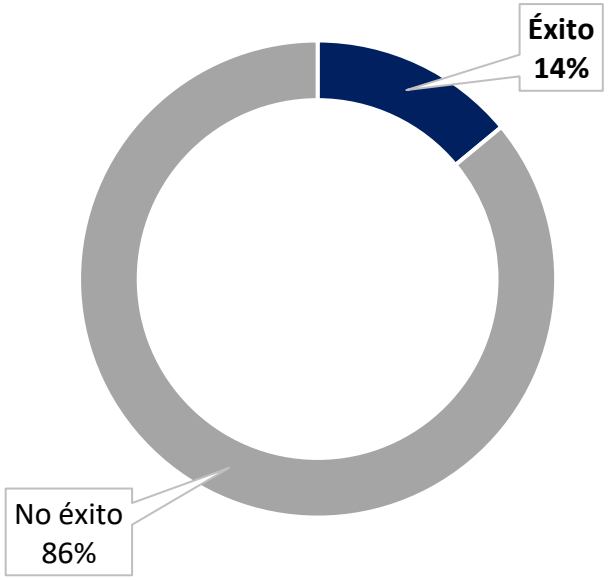
Metodología. Construcción de los *embbedings* musicales



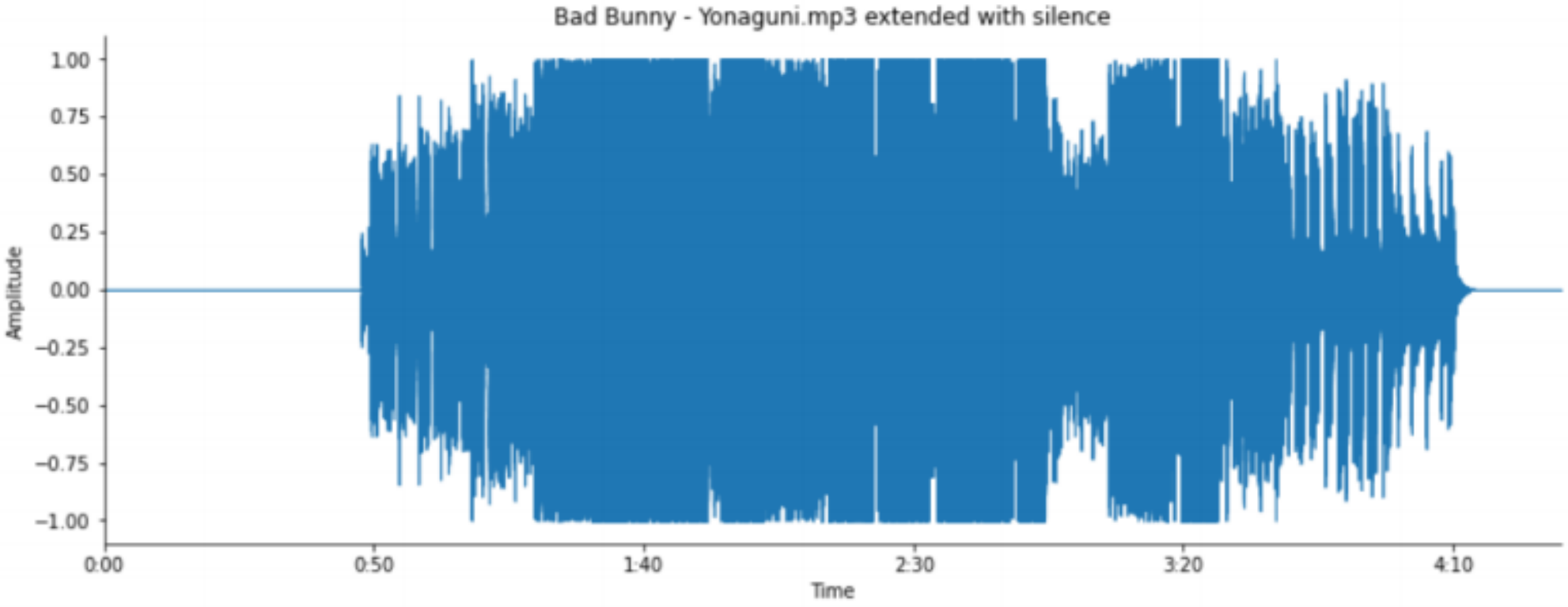
Datos



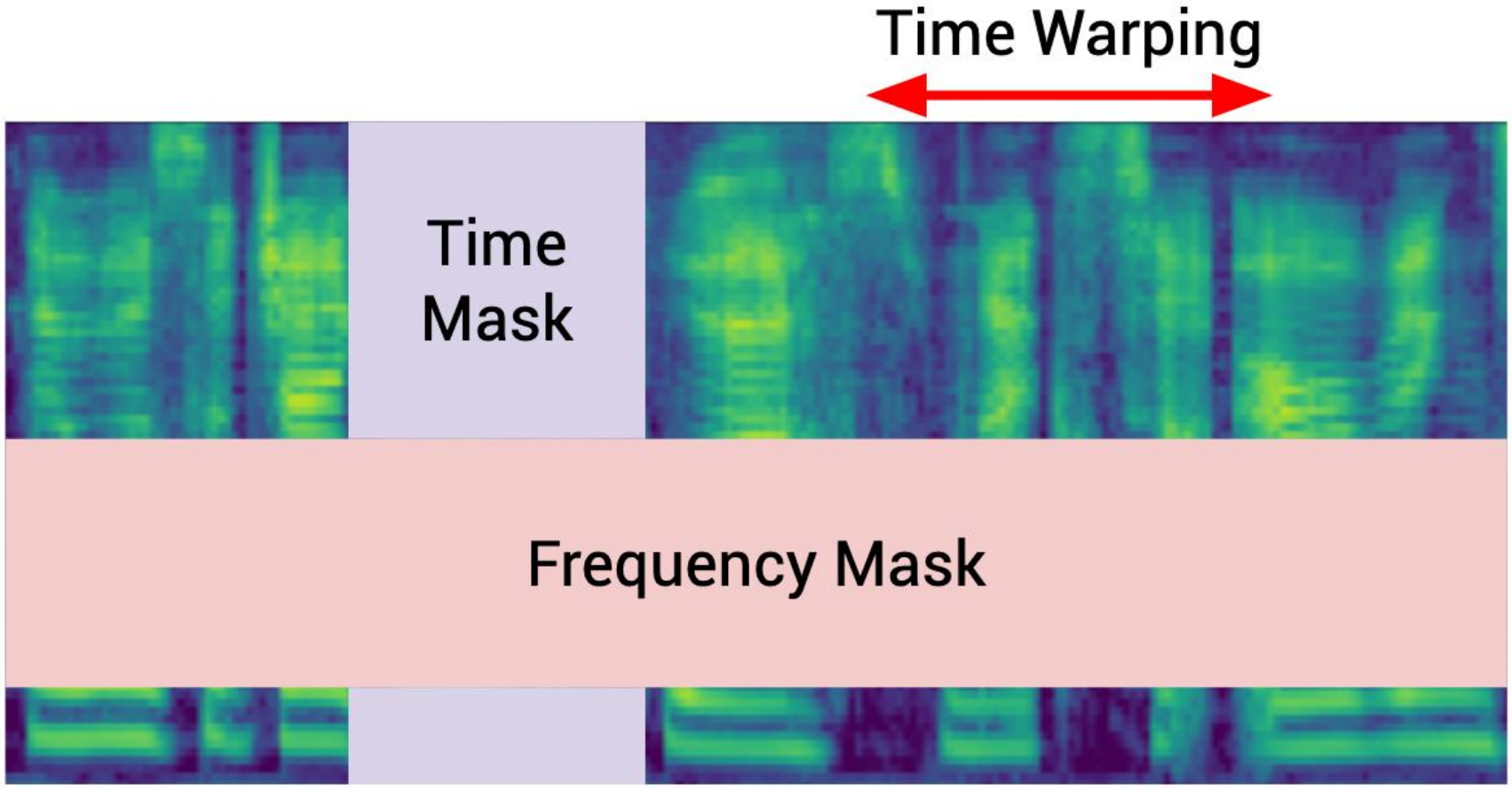
Porcentaje de éxitos %



Datos. Normalización en la duración



Datos. Crecimiento artificial de la muestra



Aplicación 1. Buscar una aguja en un pajar

- Cada 10 segundos se suben 10 canciones nuevas a Spotify (2021).
- Cerca del 16% de los ingresos de las disqueras se destinas A&R.
- Al año se gastan alrededor de $5.800 millones en creación de nuevas estrellas.
- Solo 1 de cada 10 artistas firmados es económicamente viable.

Aplicación 1. Buscar una aguja en un pajar

- Objetivo: Predecir la probabilidad de que una canción sea un hit en Colombia.
- Esta aplicación permite:
  - Ampliar la cantidad de información que se procesa por unidad de tiempo.
  - Realizar recomendaciones algorítmicas: rankings de probabilidades, asignación de recursos.

Aplicación 1. Buscar una aguja en un pajar

Especificaciones del modelo:

- CNN: ResNet-50
- Pesos: ImageNet
- Optimizador: ADAM
- Tamaño del batch: 8
- Tasa de aprendizaje: 0.1, 0.01. 0.001, **0.0001**, 0.00001.
- Función de pérdida: Weighted Binary Cross Entropy

Aplicación 1. Buscar una aguja en un pajar

Especificaciones del modelo:

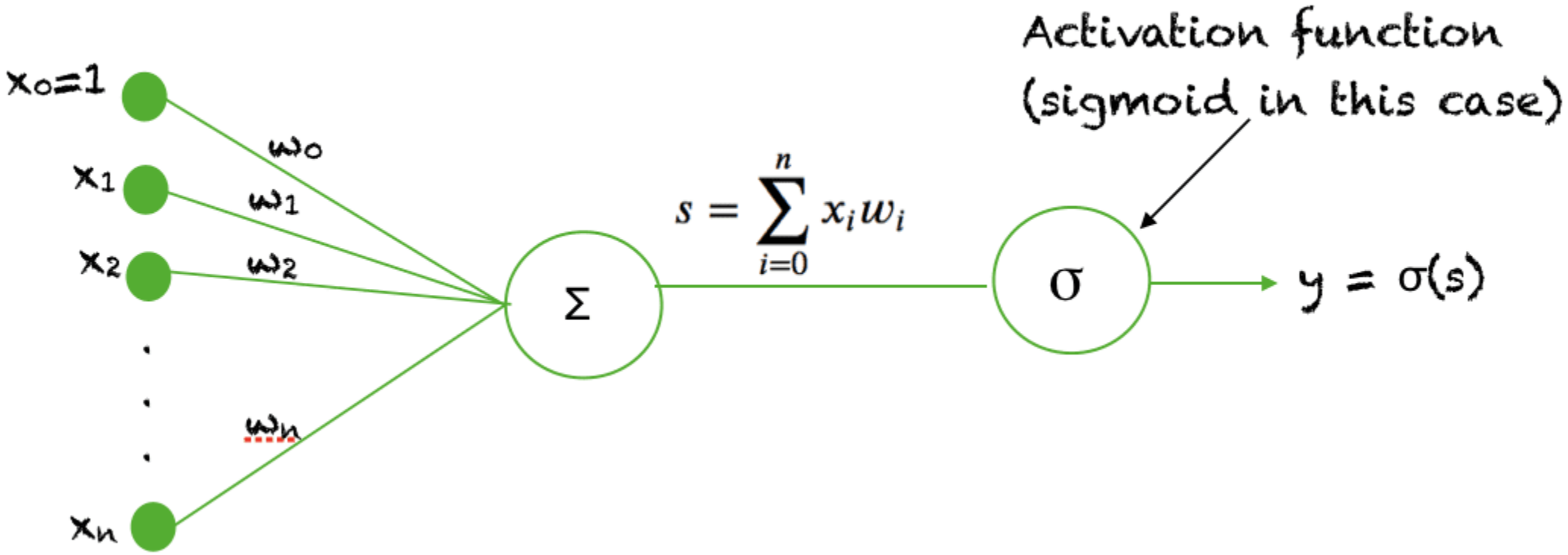
- Última capa: Sigmoidea/Logit:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

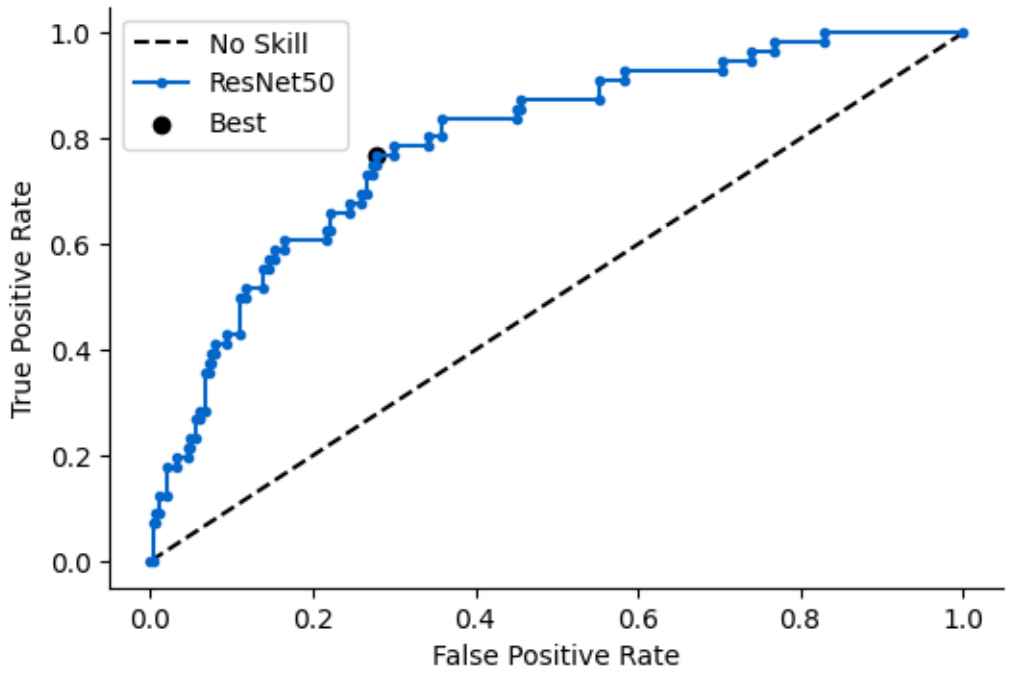
- Función de pérdida:

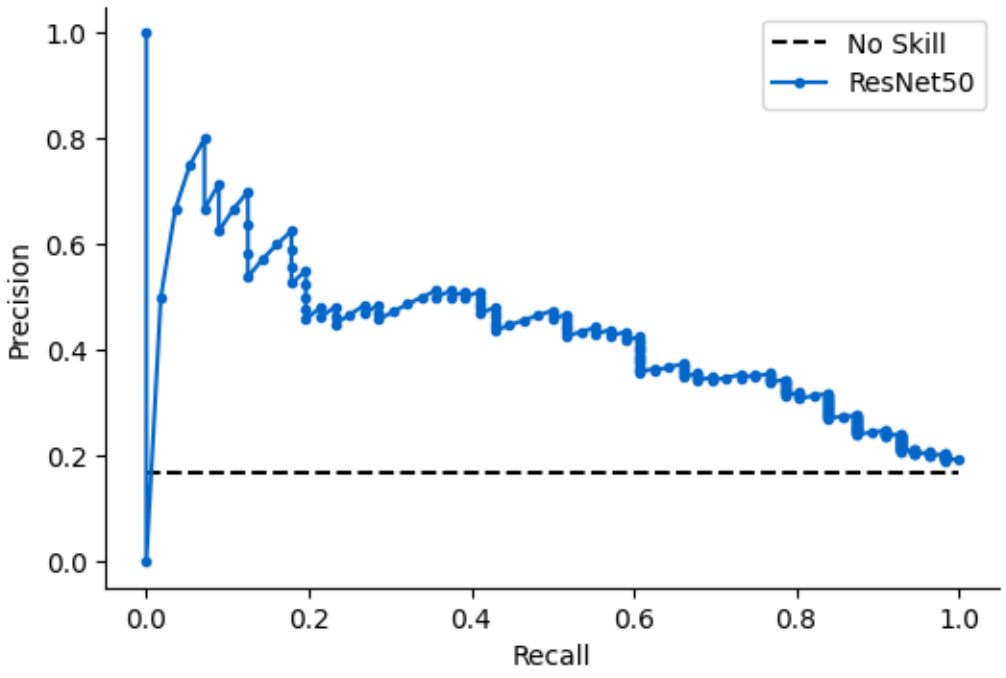
$$l_i = \omega_i [y_i \cdot \log \sigma(x_i) + (1 - y_i) \cdot \log(1 - \sigma(x_i))]$$

Aplicación 1. Buscar una aguja en un pajar



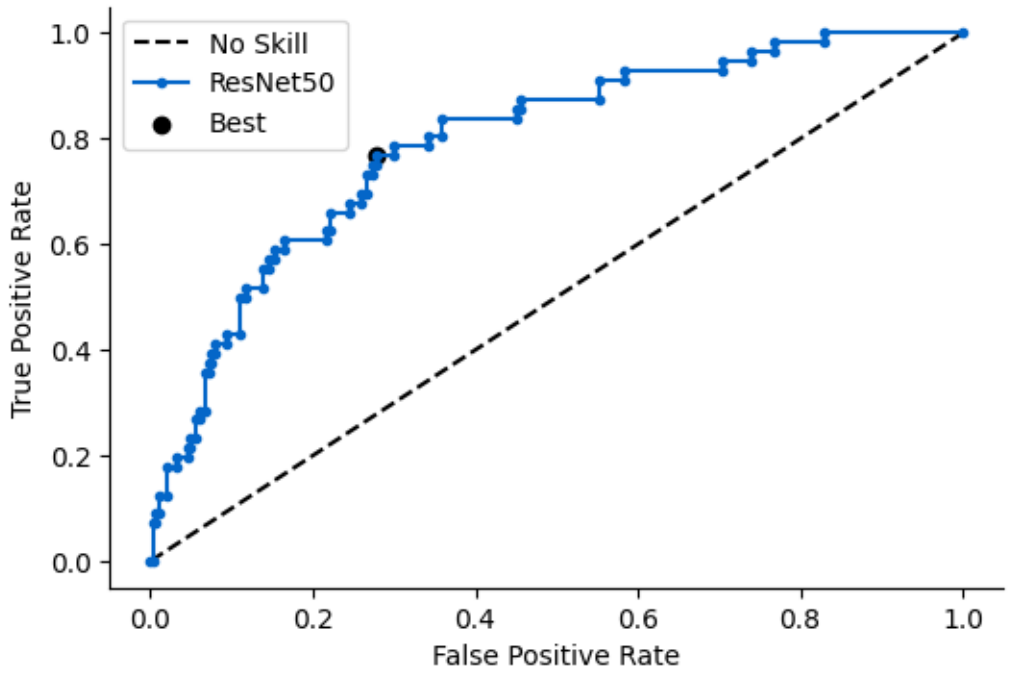
Aplicación 1. Resultados





$$\hat{y}_i = \begin{cases} 1, & \sigma(x_i) \geq 0.54 \\ 0, & \sigma(x_i) < 0.54 \end{cases}$$

Aplicación 1. Resultados



Métrica	Puntaje
AUC	0.7652
F1	0.4970
Accuracy	0.7478
Precision	0.3889
Recall	0.6885

$$\hat{y}_i = \begin{cases} 1, & \sigma(x_i) \geq 0.54 \\ 0, & \sigma(x_i) < 0.54 \end{cases}$$

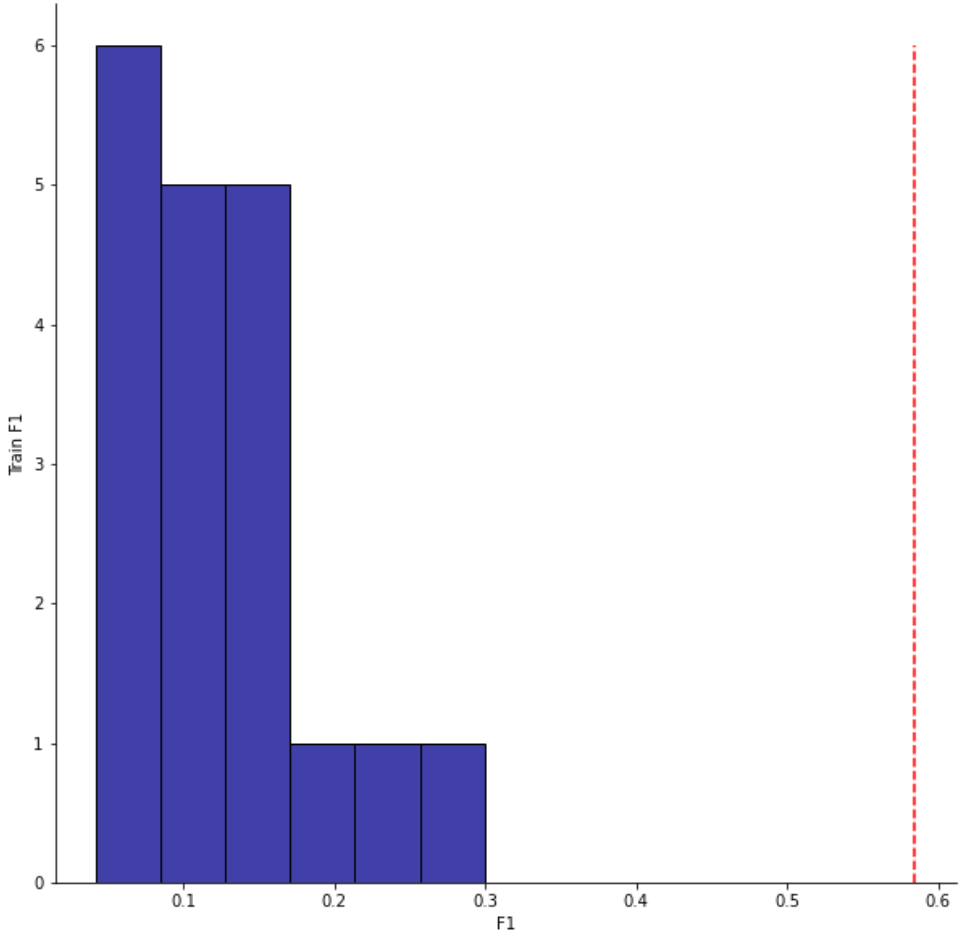
Aplicación 1. Pruebas de robustez

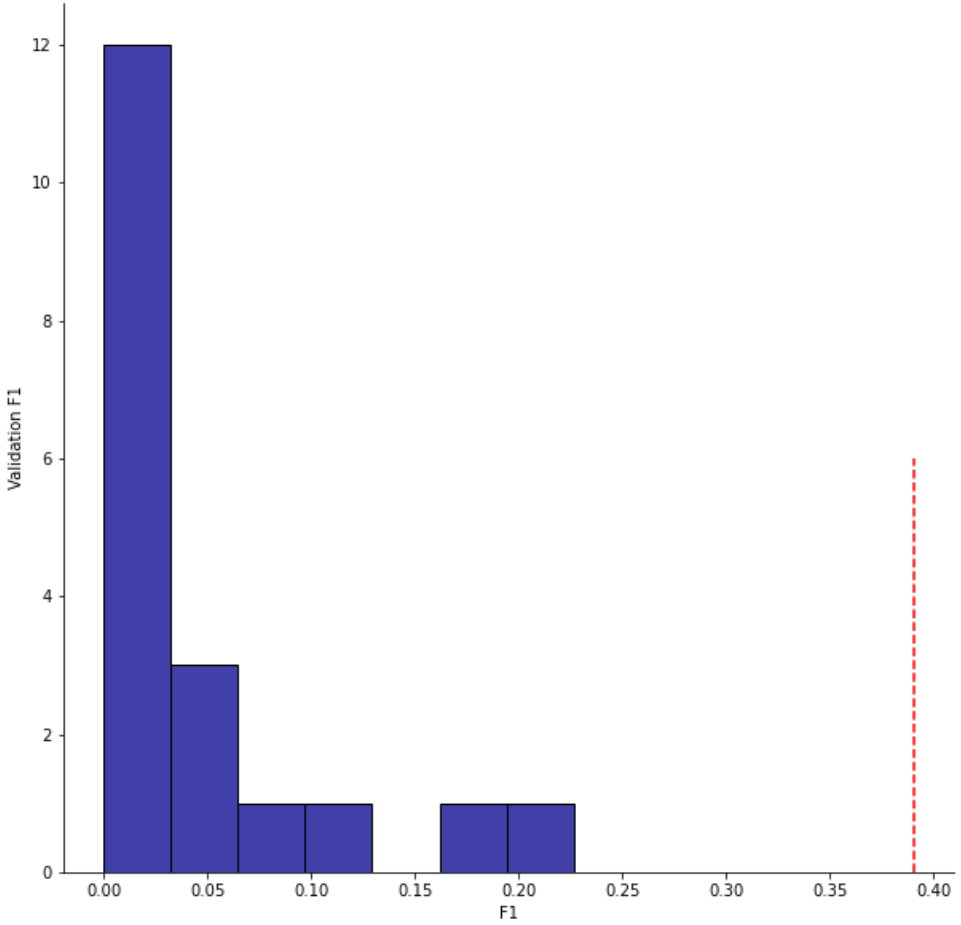
- Ojala y Garriaga (2010). Permutation test for studying classifier performance.
- Sea $\widehat{D}$ un conjunto de *k* versiones permutadas (D') de la base original D. El valor empírico del *p-valor* para el clasificador f se calcula como:

$$p = \frac{\left|\left\{D' \in \widehat{D} : e(f, D') \leq e(f, D)\right\}\right| + 1}{k + 1}$$

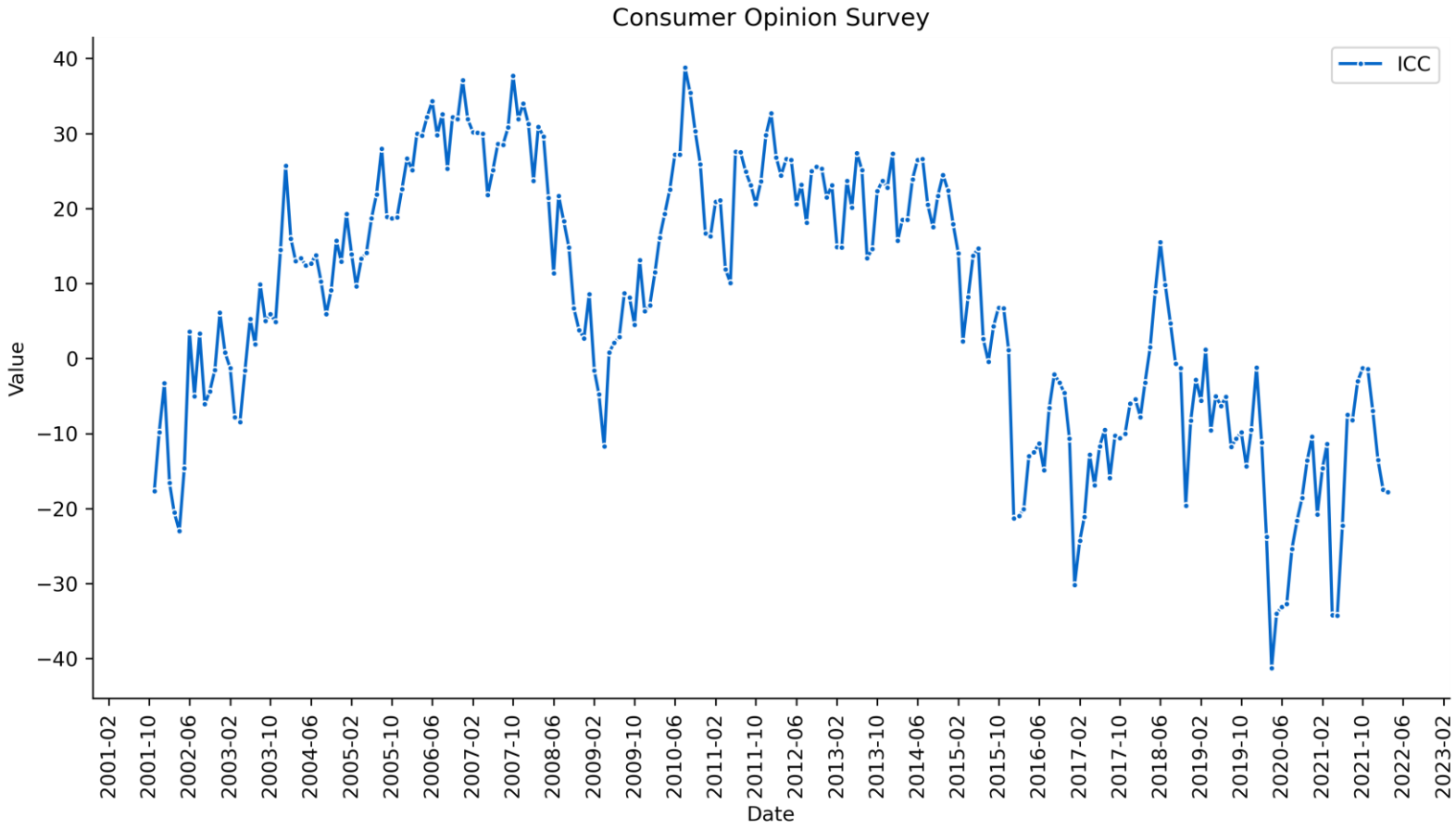
- Es decir, el porcentaje de veces que el clasificador se desempeñó mejor en la muestra permutada que en la original.

Resultados preliminares



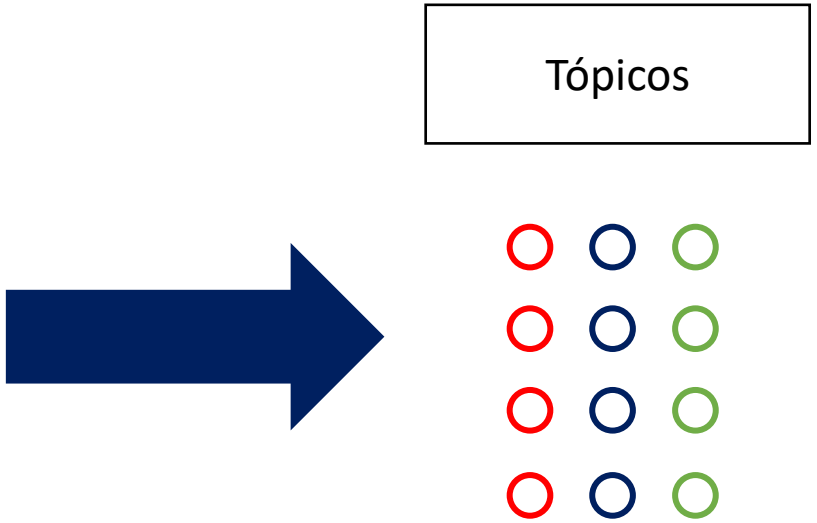


Aplicación 2. Índice de Confianza del Consumidor



Aplicación 2. Relación entre la música y el ICC

ID	Característica 1	Característica 2	…	Característica 2048
1	$w_{1,1}$	$w_{1,2}$	…	$w_{1,2048}$
2	$w_{2,1}$	$w_{2,2}$	…	$w_{2,2048}$
3	$w_{3,1}$	$w_{3,2}$	…	$w_{3,2048}$
4	$w_{4,1}$	$w_{4,2}$	…	$w_{4,2048}$



Aplicación 2. Relación entre la música y el ICC

Objetivo:

- Representar los embbeding musicales a partir de una combinación de *k* tópicos.
- Construir una serie de tiempo de la intensidad de los k tópicos de la canción promedio entre las 200 más escuchadas en Colombia cada mes.

Aplicación 2. Latent Dirichlet Allocation (LDA)

Utilizamos un modelo probabilístico generativo para un corpus de canciones.

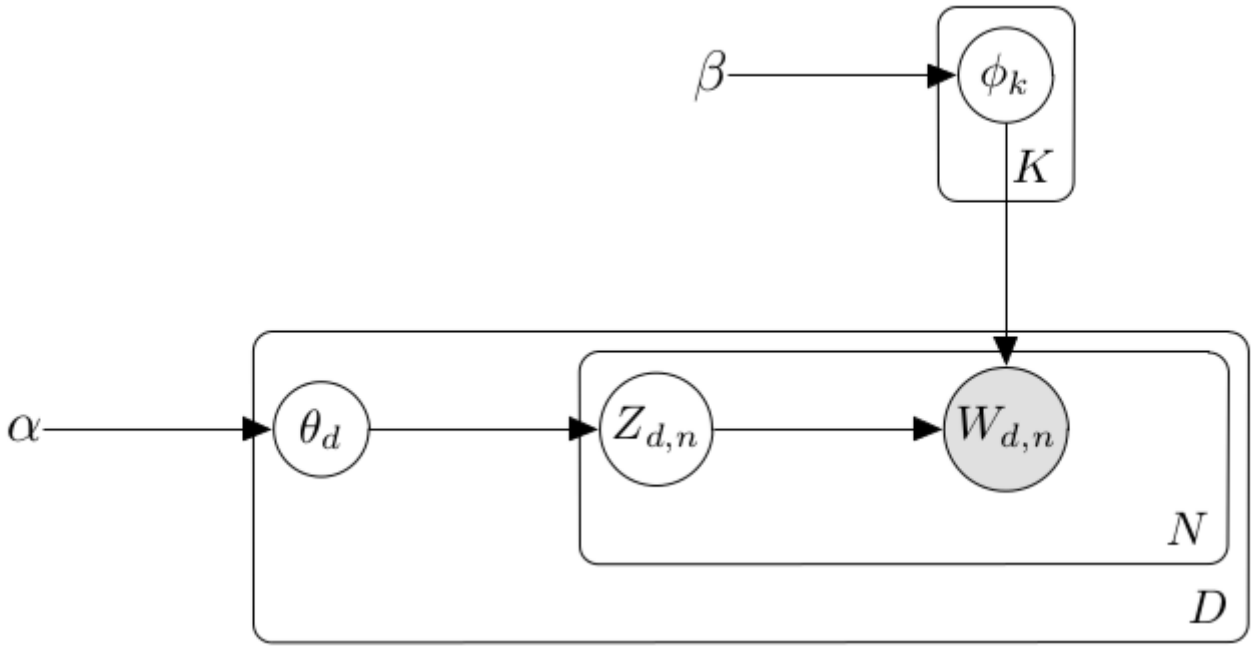
Modelar las **canciones** a partir de una combinación aleatoria de **tópicos latentes**, en donde cada tópico está descrito por una **distribución de características**.

Aplicación 2. Latent Dirichlet Allocation (LDA)

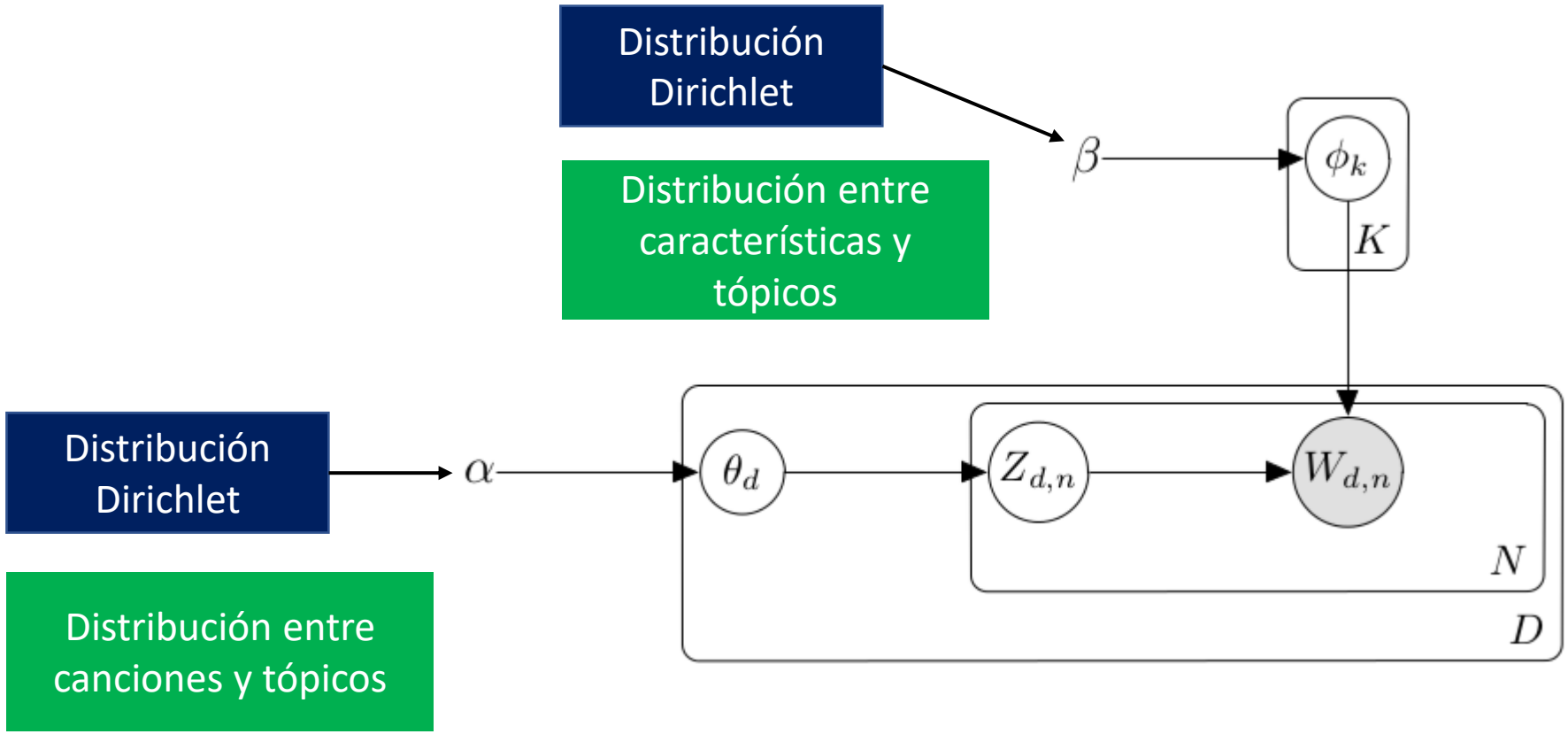
Notación:

- Una **característica** es la unidad básica de los datos discretos. Definimos una característica como un ítem de un diccionario indexado por $\{1, \cdots, V\}$. Se representan las características mediante vectores de base uno, es decir, un vector donde solo un elemento es 1 y el resto son 0. Así, usando superíndices para denotar componentes, la v-ésima palabra en el diccionario se representa mediante un V-vector w tal que $w^v = 1$ y $w^u = 0$ para $u \neq v$.
- Una **canción** es una secuencia de N características denotadas por $\boldsymbol{w} = (w_1, \mathrm{w}_2, \cdots, w_N)$, en donde w_N es la eneava característica de la secuencia.
- Un **corpus** es una colección de M canciones denotadas por $D = (\boldsymbol{w_1}, \boldsymbol{w_2}, \cdots, \boldsymbol{w_m})$

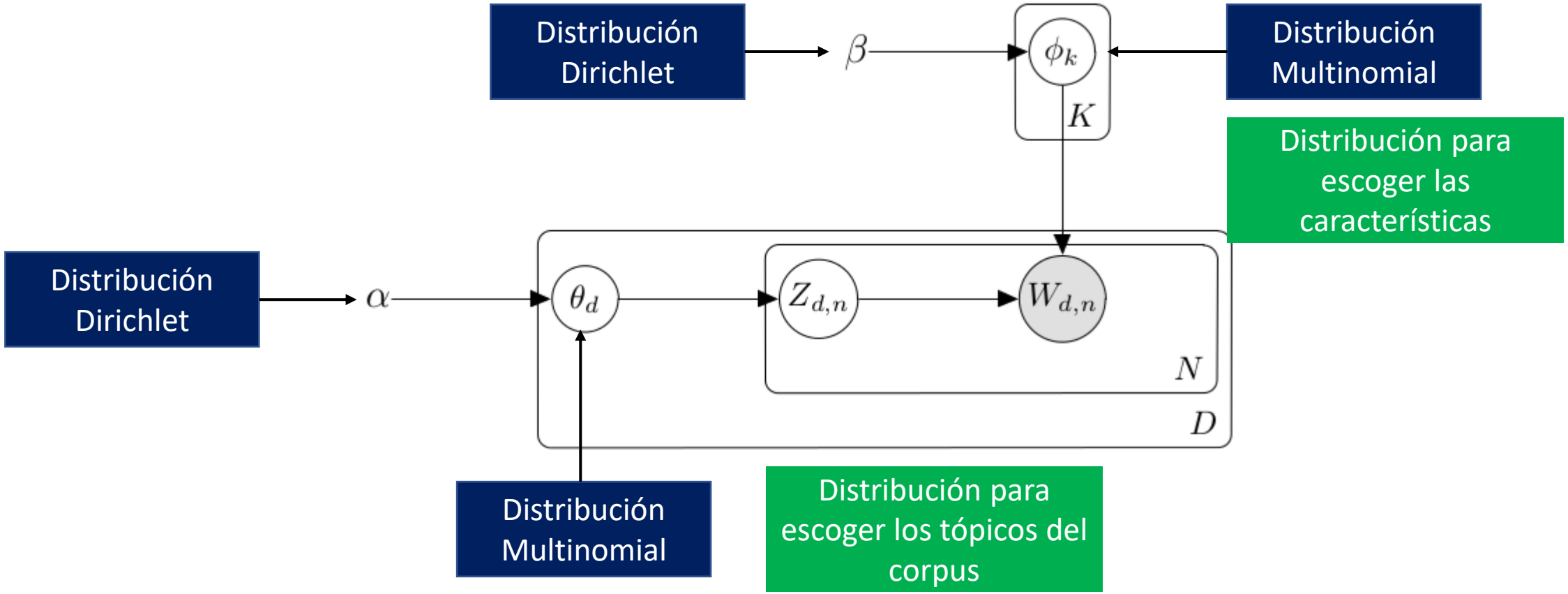
Aplicación 2. Latent Dirichlet Allocation (LDA)



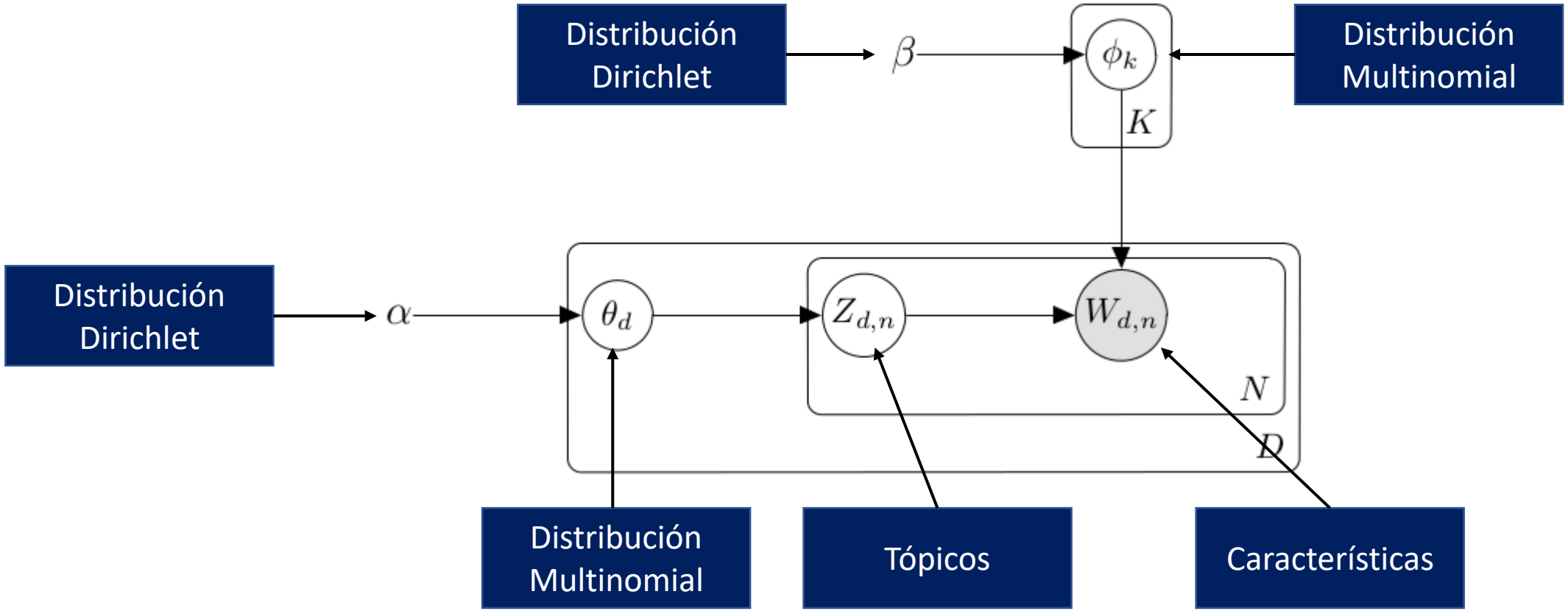
Aplicación 2. Latent Dirichlet Allocation (LDA)



Aplicación 2. Latent Dirichlet Allocation (LDA)



Aplicación 2. Latent Dirichlet Allocation (LDA)



Aplicación 2. Latent Dirichlet Allocation (LDA)

$$P(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\phi} | \alpha, \beta) = \prod_{j=1}^{M} P(\theta_j | \alpha) \prod_{i=1}^{K} P(\phi_i | \beta) \prod_{t=1}^{N} P\left(Z_{j,t} \middle| \theta_j\right) \cdot P(W_{j,t} | \phi z_{j,t})$$

Aplicación 2. Latent Dirichlet Allocation (LDA)

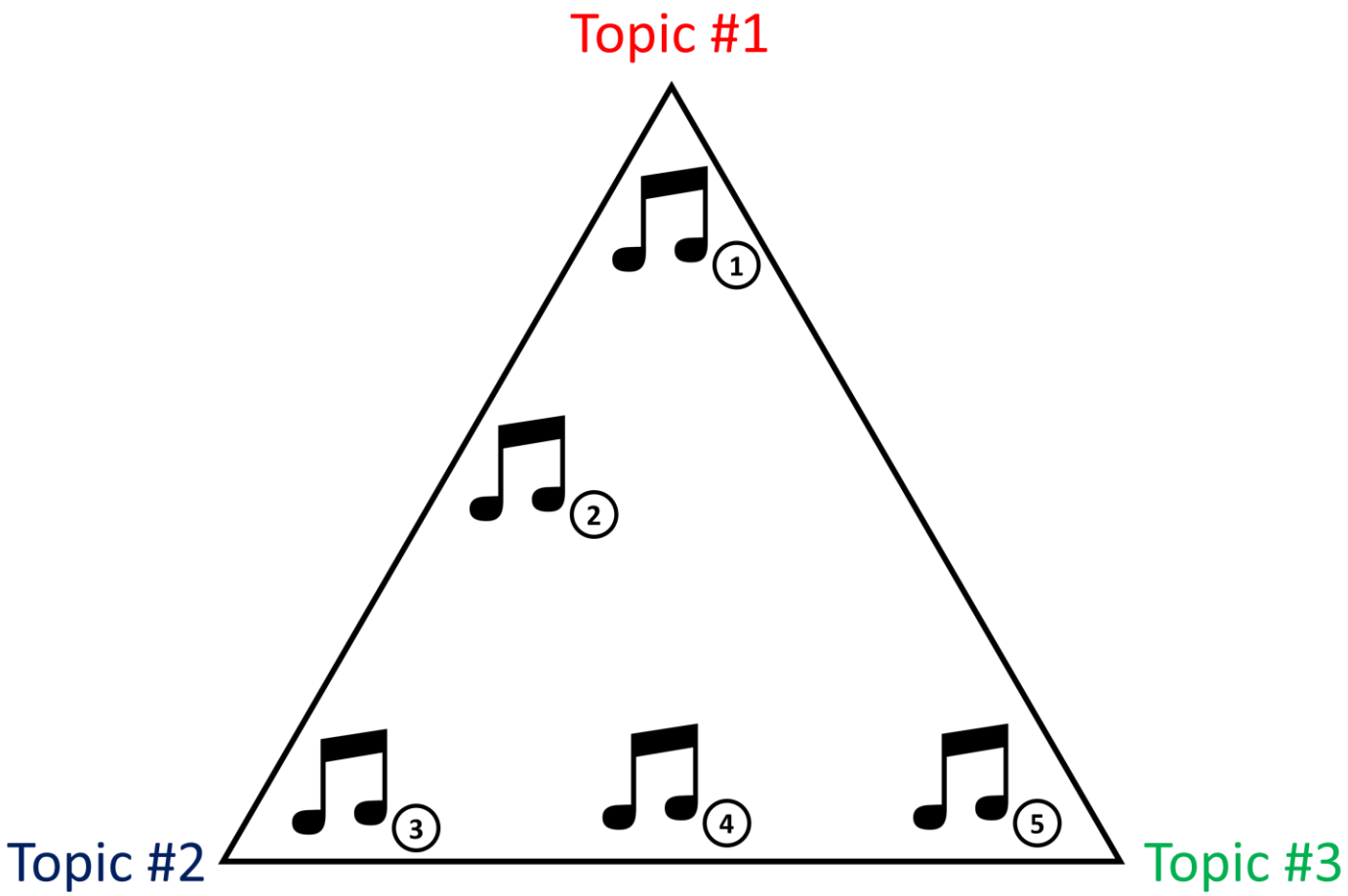
$$P(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\phi} | \alpha, \beta) = \underbrace{\prod_{j=1}^{M} P(\theta_j | \alpha) \prod_{i=1}^{K} P(\phi_i | \beta)}_{\text{Distribuciones Dirichlet}} \underbrace{\prod_{t=1}^{N} P\left(Z_{j,t} \middle| \theta_j\right) \cdot P(W_{j,t} | \phi z_{j,t})}_{\text{Distribuciones Multinomiales}}$$

Aplicación 2. Latent Dirichlet Allocation (LDA)

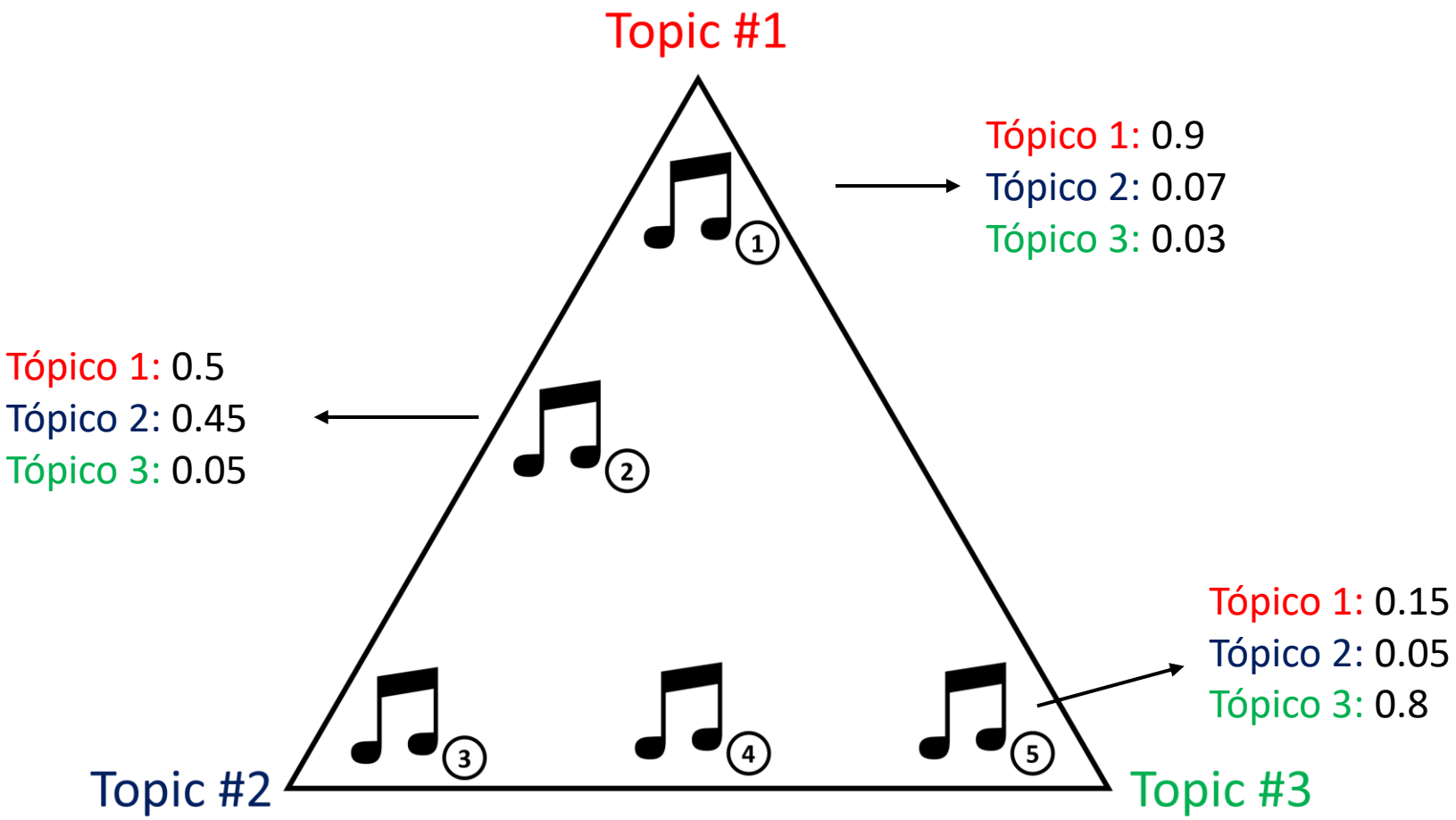
Tópicos | Características | Tópicos | Características

$$P(\boldsymbol{W},\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{\phi}|\alpha,\beta) = \underbrace{\prod_{j=1}^{M} P(\theta_j|\alpha) \prod_{i=1}^{K} P(\phi_i|\beta)}_{\text{Distribuciones Dirichlet}} \underbrace{\prod_{t=1}^{N} P\left(Z_{j,t}\middle|\theta_j\right) \cdot P(W_{j,t}|\phi z_{j,t})}_{\text{Distribuciones Multinomiales}}$$

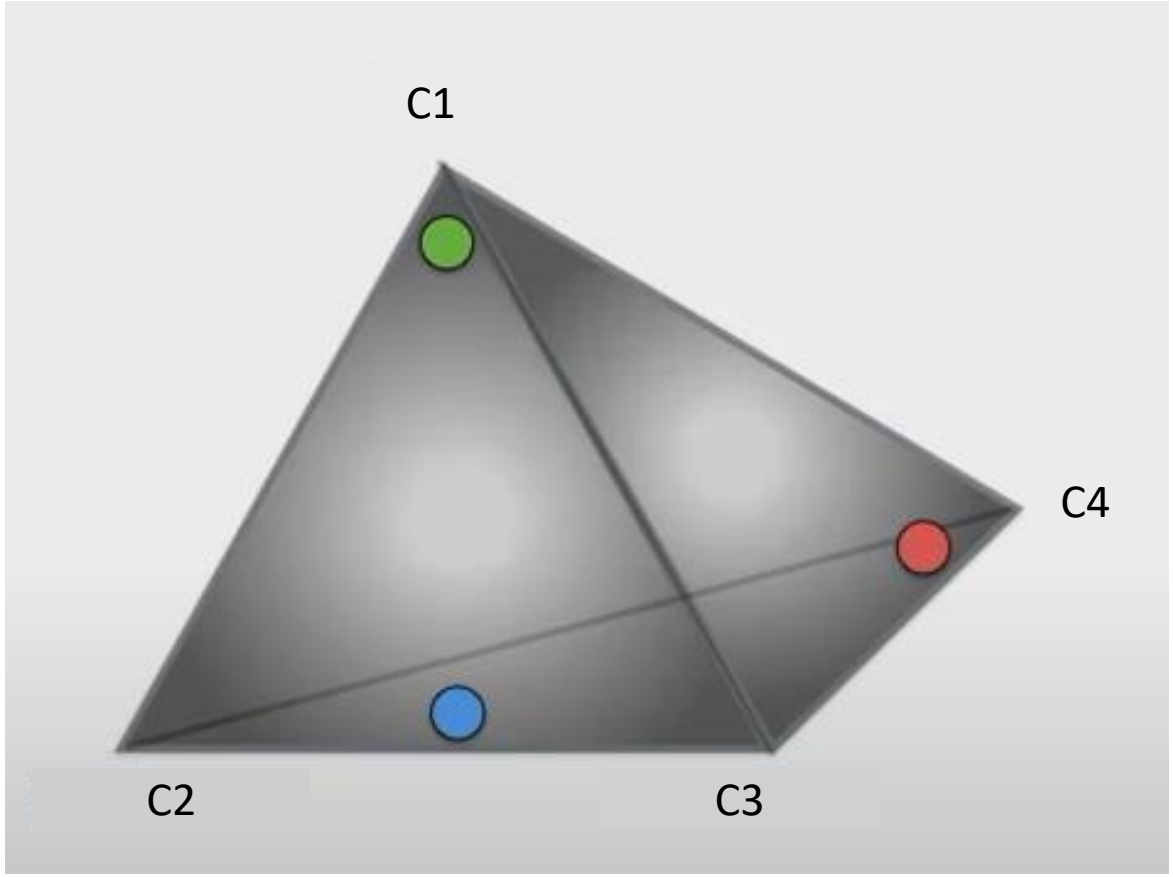
Aplicación 2. Distribución Dirichlet (Canciones-Tópicos)



Aplicación 2. Distribución Dirichlet (Canciones-Tópicos)

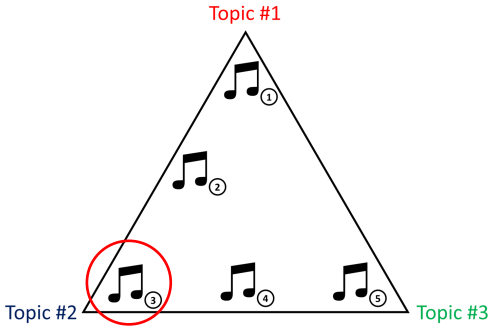


Aplicación 2. Distribución Dirichlet (Tópicos-Características)

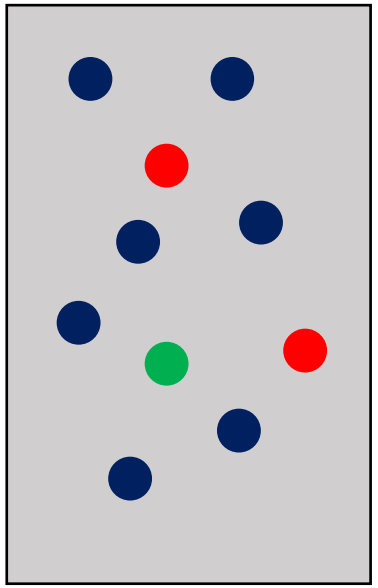


Aplicación 2. Latent Dirichlet Allocation (LDA)

$$P(\boldsymbol{W},\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{\phi}|\alpha,\beta) = \prod_{j=1}^{M} P(\theta_j|\alpha) \prod_{i=1}^{K} P(\phi_i|\beta) \prod_{t=1}^{N} P\left(Z_{j,t}\middle|\theta_j\right) \cdot P(W_{j,t}|\phi z_{j,t})$$

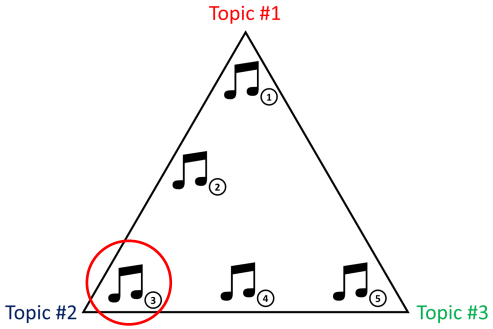


Tópico 1: 0.2
Tópico 2: 0.7
Tópico 3: 0.1

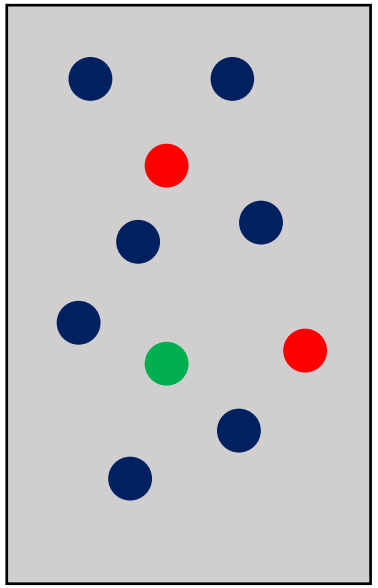


Aplicación 2. Latent Dirichlet Allocation (LDA)

$$P(\boldsymbol{W},\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{\phi}|\alpha,\beta) = \prod_{j=1}^{M} P(\theta_j|\alpha) \prod_{i=1}^{K} P(\phi_i|\beta) \prod_{t=1}^{N} P\left(Z_{j,t}\middle|\theta_j\right) \cdot P(W_{j,t}|\phi z_{j,t})$$



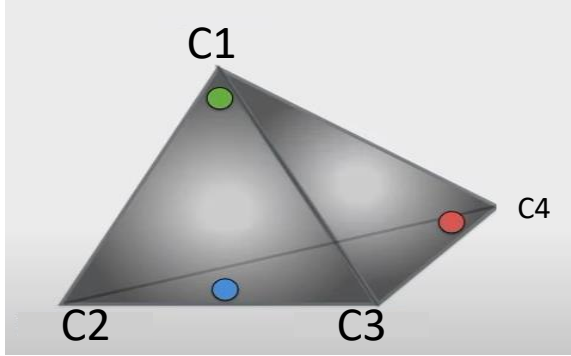
Tópico 1: 0.2
Tópico 2: 0.7
Tópico 3: 0.1



Tópico 2
Tópico 2
Tópico 1
Tópico 2
Tópico 2
Tópico 3
Tópico 1
Tópico 1
Tópico 2

Aplicación 2. Latent Dirichlet Allocation (LDA)

$$P(\boldsymbol{W},\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{\phi}|\alpha,\beta) = \prod_{j=1}^{M} P(\theta_j|\alpha) \prod_{i=1}^{K} P(\phi_i|\beta) \prod_{t=1}^{N} P\left(Z_{j,t}\middle|\theta_j\right) \cdot P(W_{j,t}|\phi z_{j,t})$$



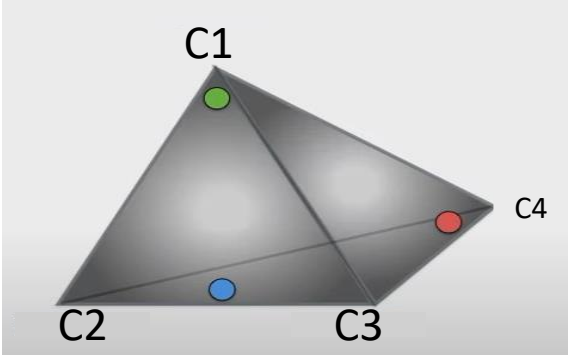
C1	C2	C3	C4
0.1	0.4	0.4	0.1

Tópico 2
Tópico 2
Tópico 1
Tópico 2
Tópico 2
Tópico 3
Tópico 1
Tópico 1
Tópico 2

C1 C3
C3 C2 C4
C2 C2
C3 C2
C3

Aplicación 2. Latent Dirichlet Allocation (LDA)

$$P(\boldsymbol{W},\boldsymbol{Z},\boldsymbol{\theta},\boldsymbol{\phi}|\alpha,\beta) = \prod_{j=1}^{M} P(\theta_j|\alpha) \prod_{i=1}^{K} P(\phi_i|\beta) \prod_{t=1}^{N} P\left(Z_{j,t}\middle|\theta_j\right) \cdot P(W_{j,t}|\phi z_{j,t})$$



C1	C2	C3	C4
0.1	0.4	0.4	0.1

Tópico 2: C3
Tópico 2: C2
Tópico 1
Tópico 2: C3
Tópico 2: C2
Tópico 3
Tópico 1
Tópico 1
Tópico 2: C1

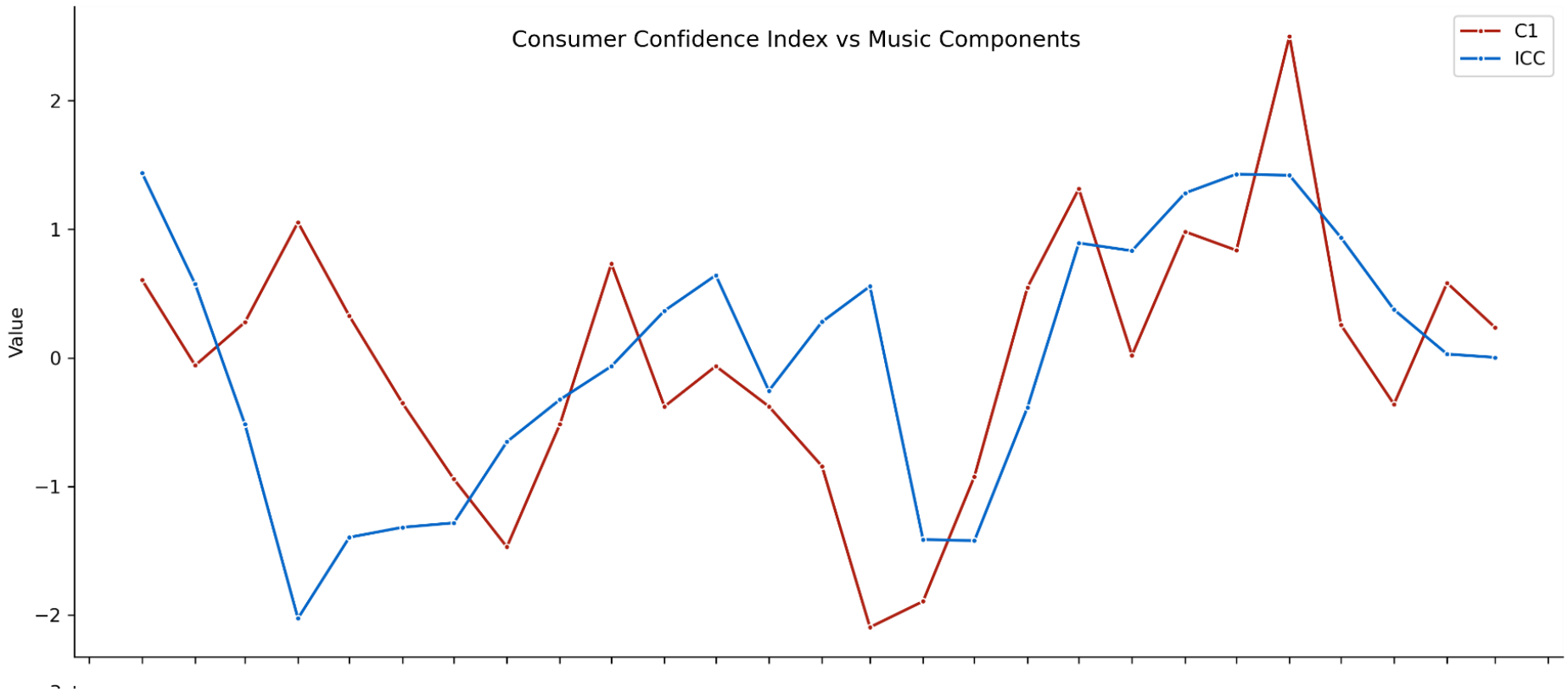
C1 C3
C3 C2 C4
C2 C2
C3 C2
C3

Aplicación 2. Latent Dirichlet Allocation (LDA)

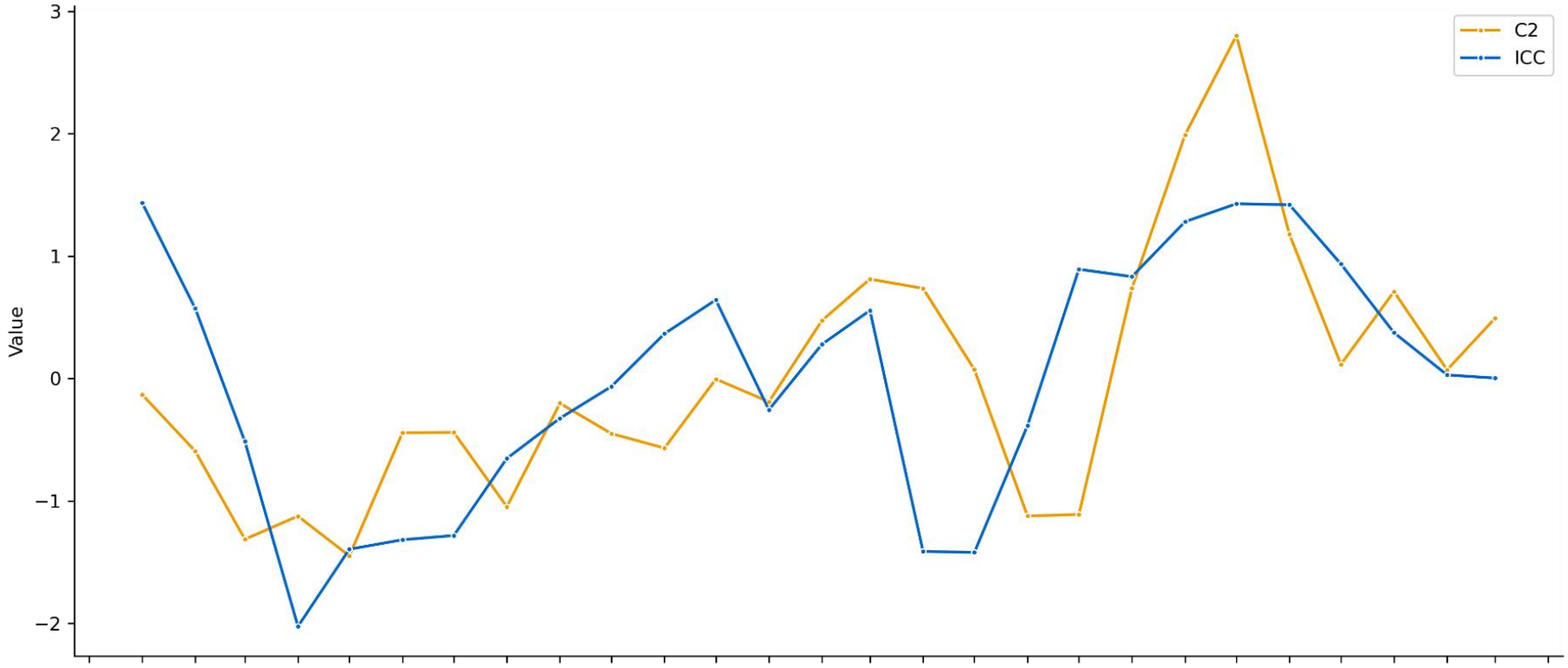
- Se dividió la base en entrenamiento y validación (80/20).
- Se buscaron los mejores hiperparámetros a partir de un grid search utilizando 3-fold CV en train.
  - K: 3 a 20 tópicos.
  - Alpha: 0.01, 0.1, 0.3, 0.5, 0.7, 0.9 y 5.
  - Beta: 0.01, 0.1, 0.3, 0.5, 0.7, 0.9 y 5.
- Los parámetros que maximizaron la log-verosimilitud fueron K=3, Alpha = 0.1 y Beta = 0.5

Aplicación 2. Latent Dirichlet Allocation (LDA)

Aplicación 2. Relación entre Tópicos e ICCC



Aplicación 2. Relación entre Tópicos e ICCC



Aplicación 2. Relación entre Tópicos e ICCC

Aplicación 2. Relación entre Tópicos e ICCC

	Consumer Confidence Index
Component #1 (Reggaeton)	2.453**
	(0.877)
Component #2 (Pop)	−5.631***
	(1.463)
Componente #3 (Rock)	4.355***
	(0.563)
Observations	27
R^2	0.838
Adjusted R^2	0.818
Residual Std. Error	9.030 (df = 24)
F Statistic	41.384*** (df = 3; 24)
Note: Robust standard errors in parentheses	*p<0.1; **p<0.05; ***p<0.01

Comentarios finales

- Una canción de Spotify promedio en calidad estándar pesa 3.96 MB.
- El embedding de la canción promedio pesa 0.0164 MB.
- La metodología propuesta logra representar la música en un elemento **240 veces más pequeño** que el formato original promedio.