

The background features a stylized collage of musical elements. On the left, a red acoustic guitar and a yellow trumpet are partially visible against a black and teal geometric pattern. In the center, a black vinyl record with a yellow label is shown next to a white speaker icon. A grey circle containing a white musical note is positioned at the bottom left. The right side of the slide has a light beige background with a large, thin-lined grey circle. The title text is overlaid on this circle.

Uso de algoritmos de Machine Learning en la clasificación de géneros musicales de pistas de audio

Saúl Contreras - Defensa de Proyecto para
adquirir la Licenciatura en Ingeniería en
Ciencias de la Computación y TI



Tabla de Contenidos

01

Justificación

03

**Marco
Teórico**

02

Objetivos

04

Metodología

Tabla de Contenidos

05

Desarrollo

07

Recomendaciones

06

Conclusiones

08

Preguntas

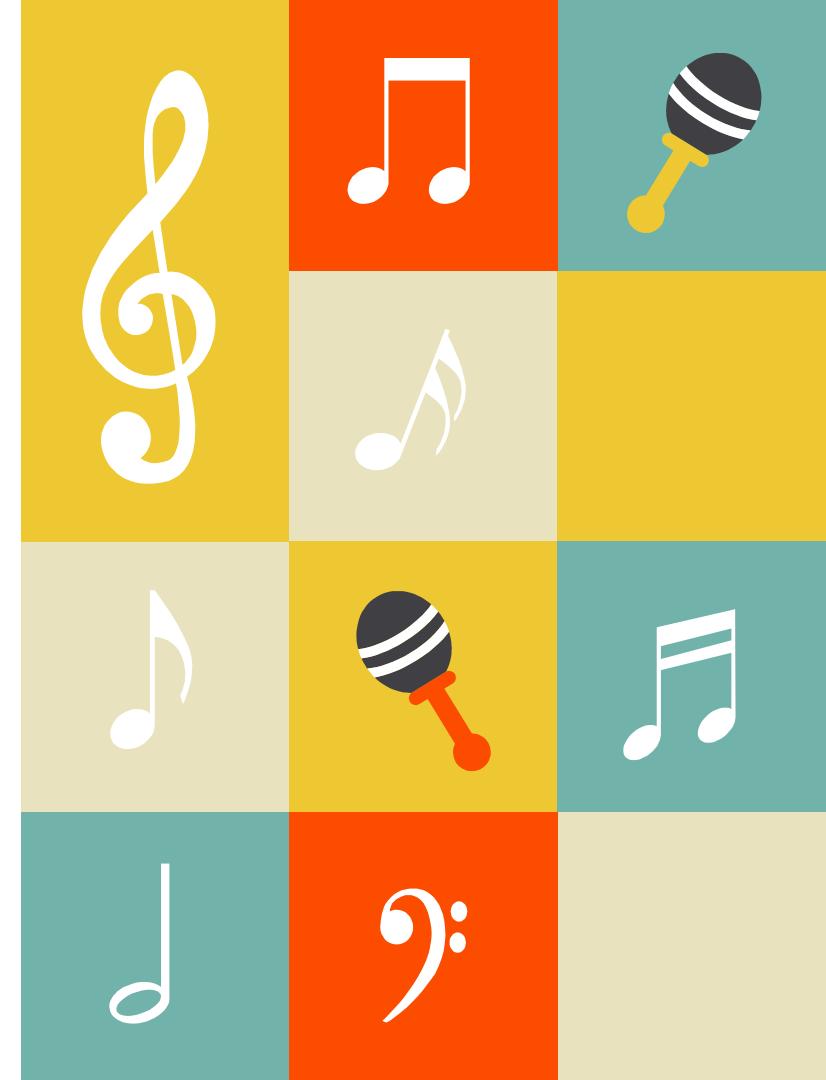
Introduction

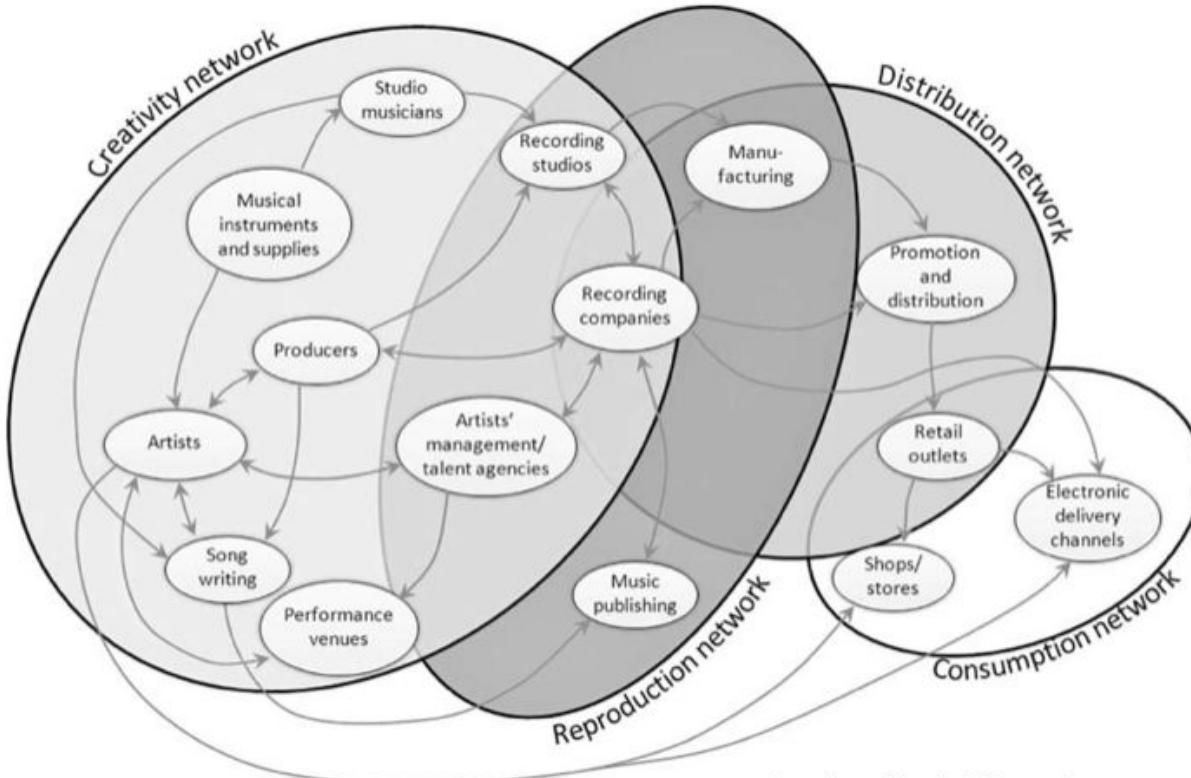
En el proyecto de graduación se podrá observar tres fases específicas. La extracción y análisis de datos proveídos por distintas APIs de servicios de Streaming. La construcción de algoritmos de Machine Learning para clasificar cada uno de los géneros musicales. La elaboración de un sitio web donde se pueda obtener estos modelos.



01

Justificación





Leyshon: Musical Networks
Como aparece en "Music Industry", Patrick Wikström.





“Es necesario un sistema de clasificación para poder enlazar a los productores con los artistas y viceversa ...

... la clasificación por género es uno de los mejores sistemas para emparejar dichos roles”.

—Richard James Burgess





A&R



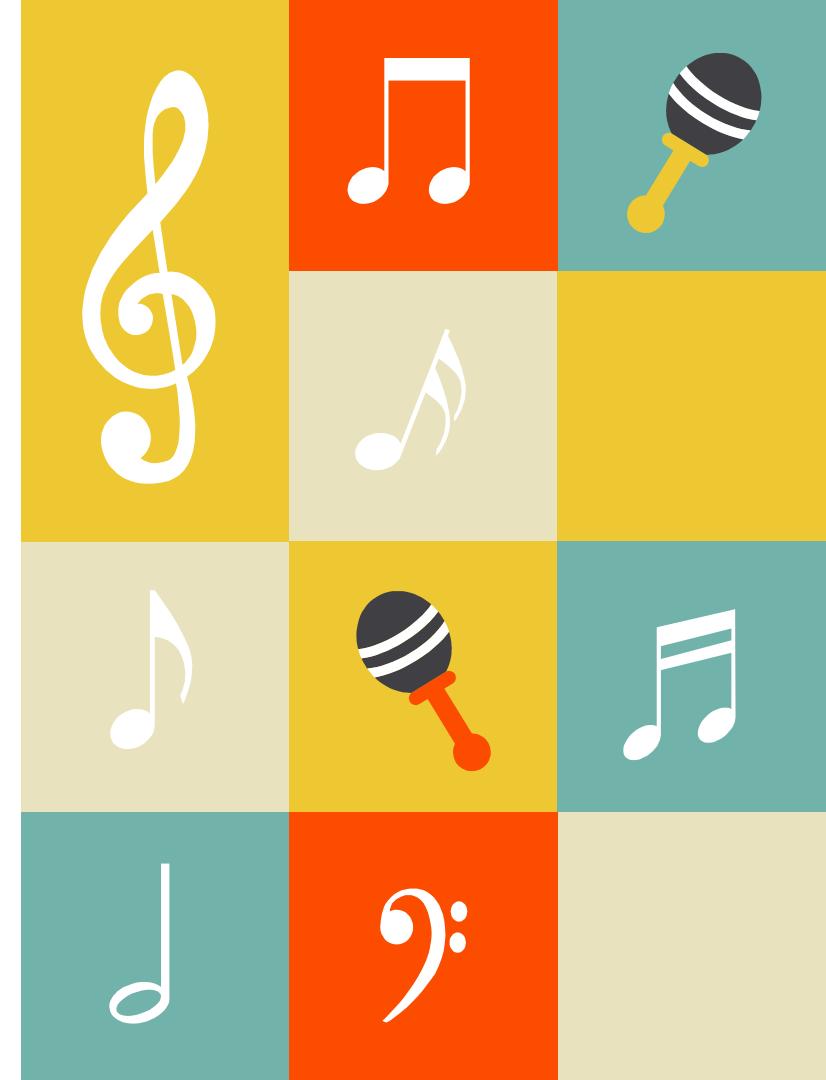


60k

Canciones son subidas a Spotify diariamente

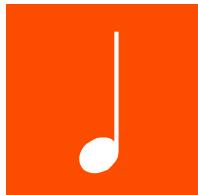
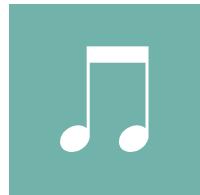
02

Objetivos





**Analizar métodos de clasificación de
machine learning para clasificar géneros
musicales de pistas de audio con una
efectividad mayor al 75 %.**





Objetivos Específicos



Construir un sitio web de acceso público que permita determinar el género musical de una nueva pista de audio.



Construir modelos de machine learning que permitan clasificar los géneros musicales de nuevas pistas de audio.

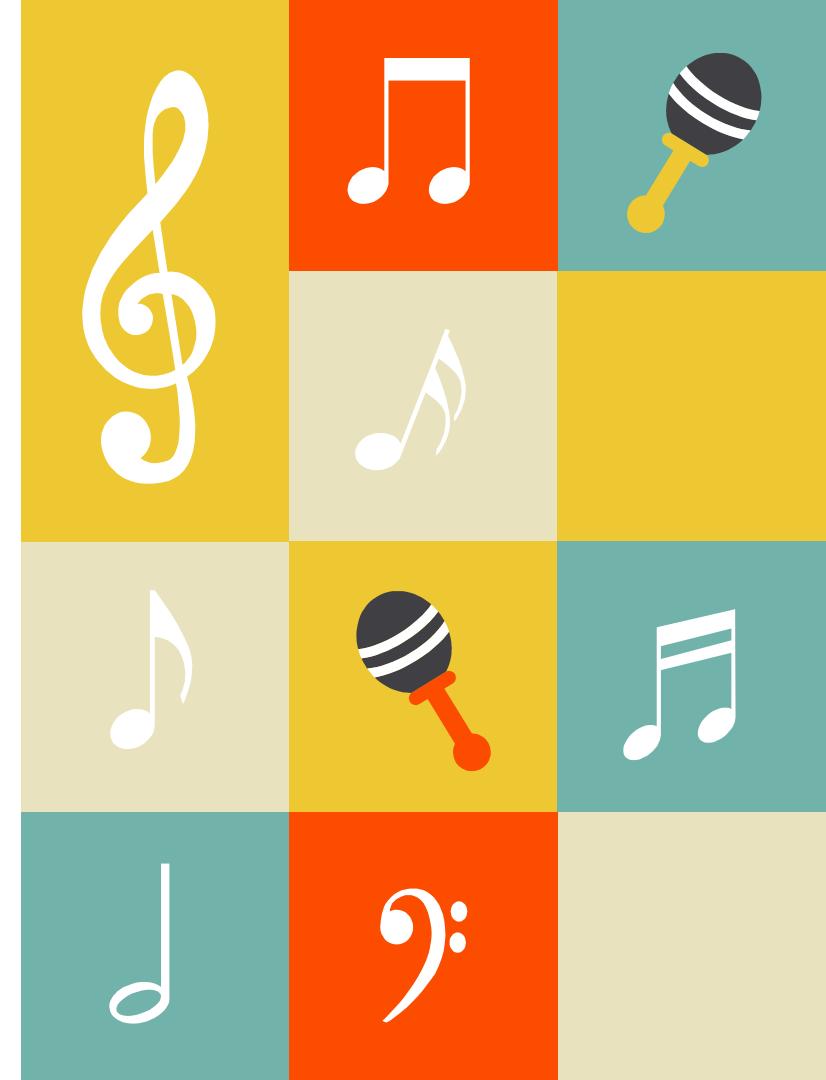


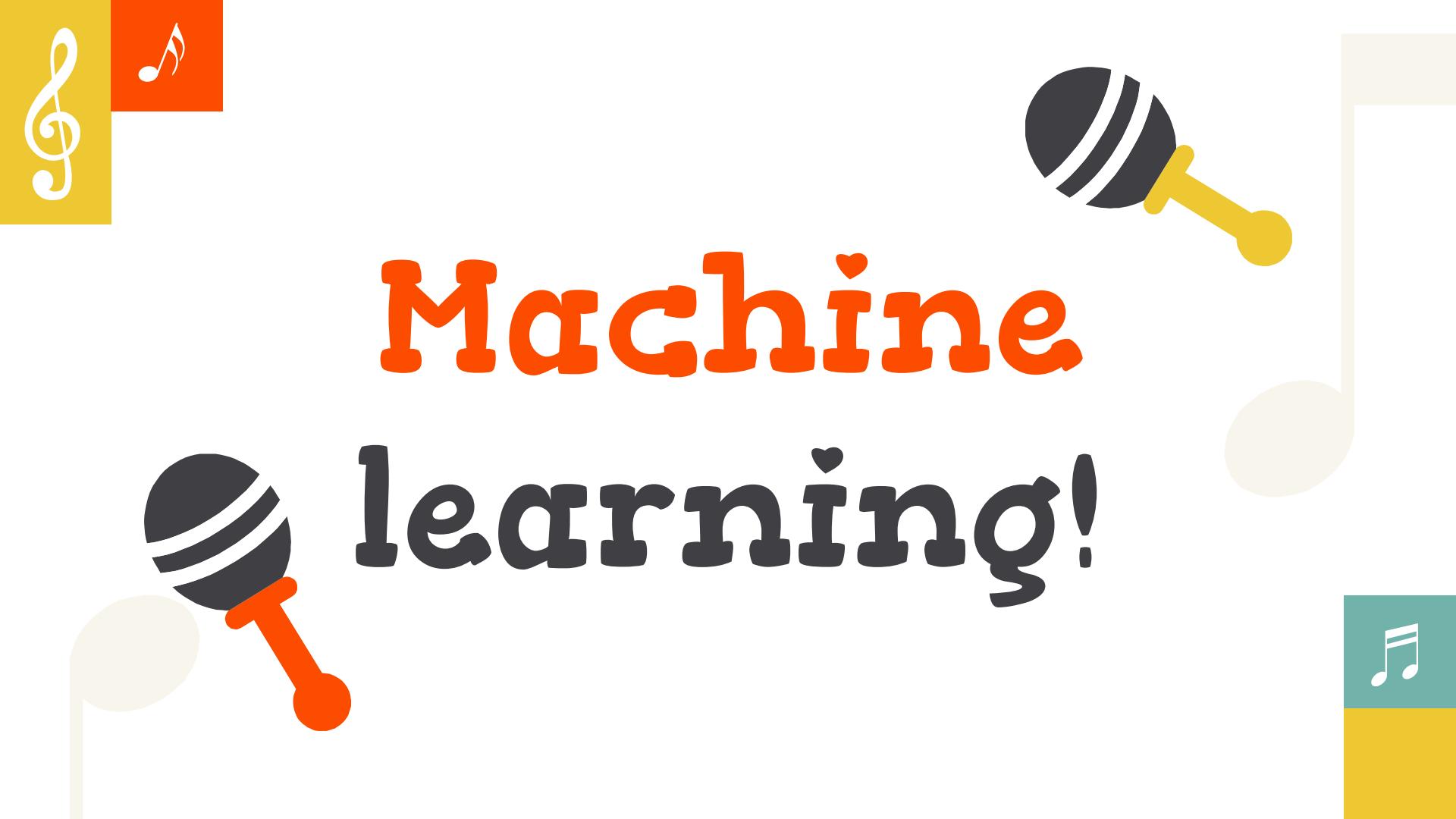
Explorar los datos en las APIs públicas de servicios de Streaming.



03

Marco Teórico

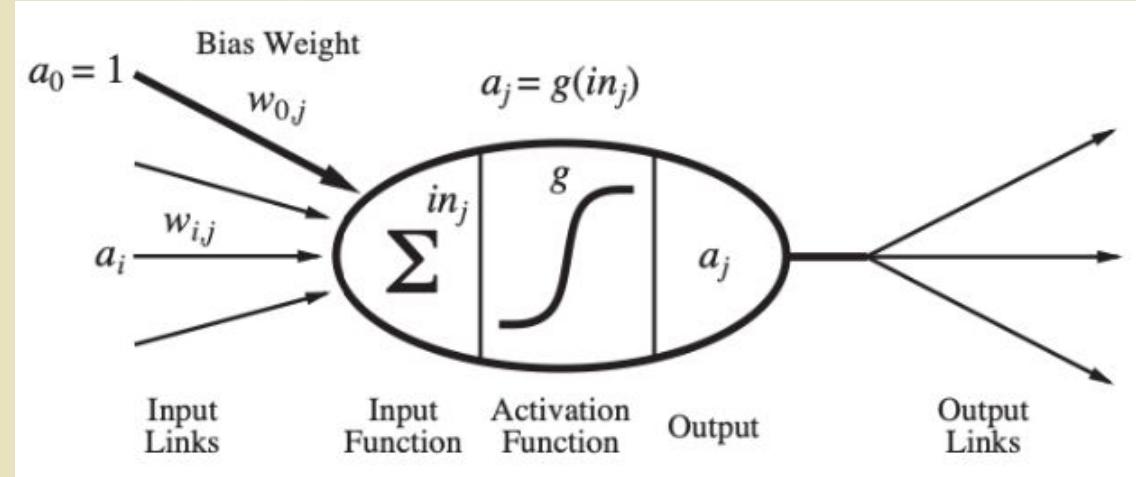




Machine
learning!



Redes Neuronales





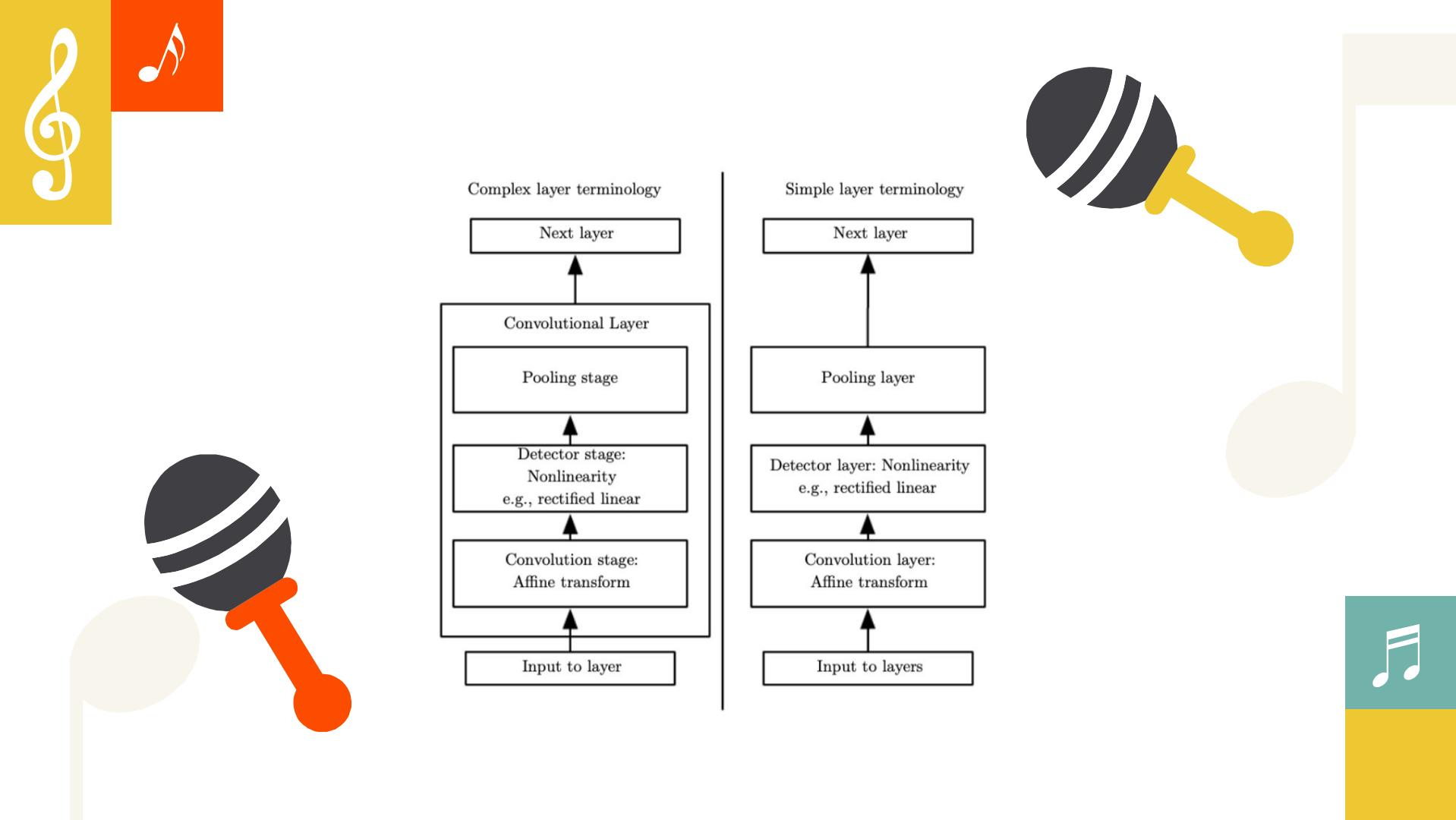
Redes Convolucionales



Una red neuronal que utiliza la operación matemática de Convolución.

$$s(t) = \int_{-\infty}^{\infty} x(a)\omega(t-a) da$$





Redes Convolucionales



Convolución

Una operación lineal



Pooling

Una capa típica en este tipo de redes



Parameter Sharing

Compartir Parámetros

Convolucion

$$s(t) = \int_{-\infty}^{\infty} x(a)\omega(t - a) da$$

Convolucion

$$s(t) = (x * \omega)(t) = \sum_{-\infty}^{\infty} x(a)\omega(t - a)$$

Redes Convolucionales



Convolución

Una operación lineal



Pooling

Una capa típica en este tipo de redes



Parameter Sharing

Compartir Parámetros

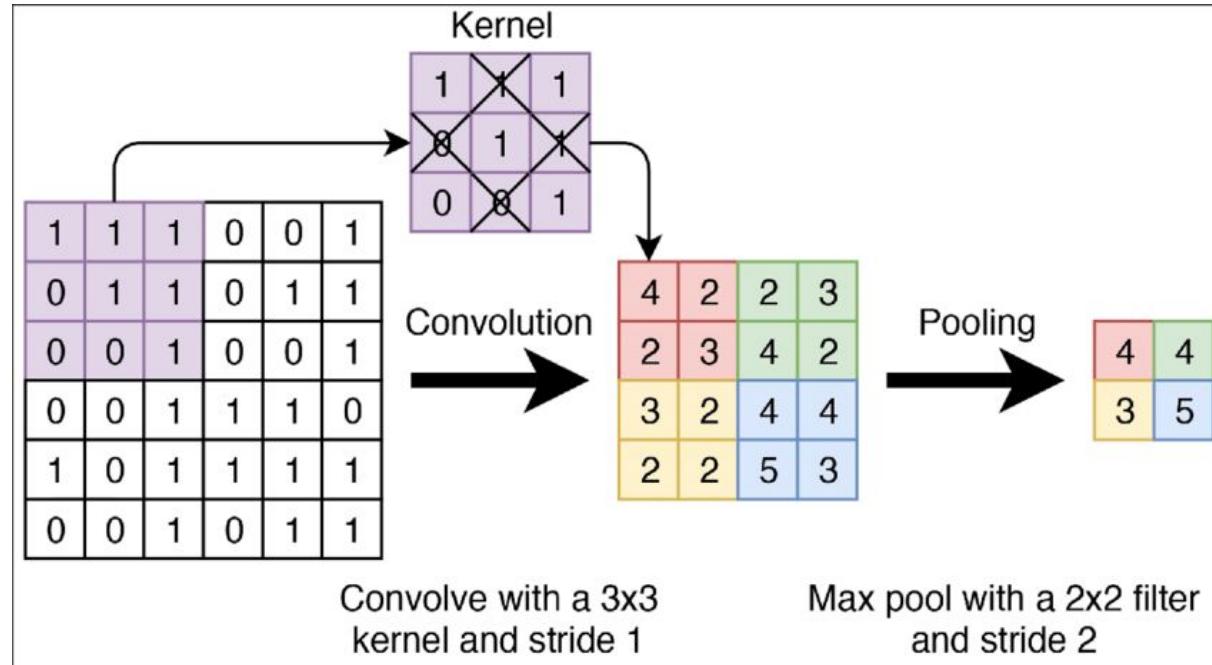
Pooling



Ejecuta varias convoluciones en paralelo para producir un set de activaciones lineales

Detección: cada activacion linea es ejecutada a travez de una función de activación no lineal

Pooling: para modificar la salida de la siguiente capa.



Redes Convolucionales



Convolución

Una operación lineal



Pooling

Una capa típica en este tipo de redes

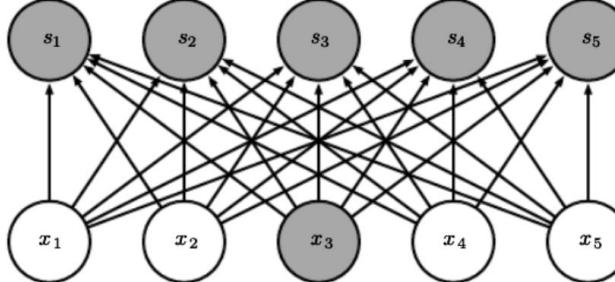
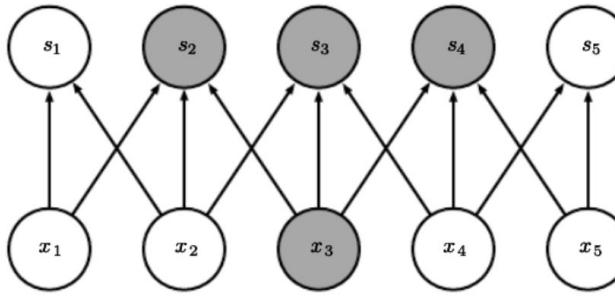


Parameter Sharing

Compartir Parámetros



Parameter Sharing



Modelo LeNet

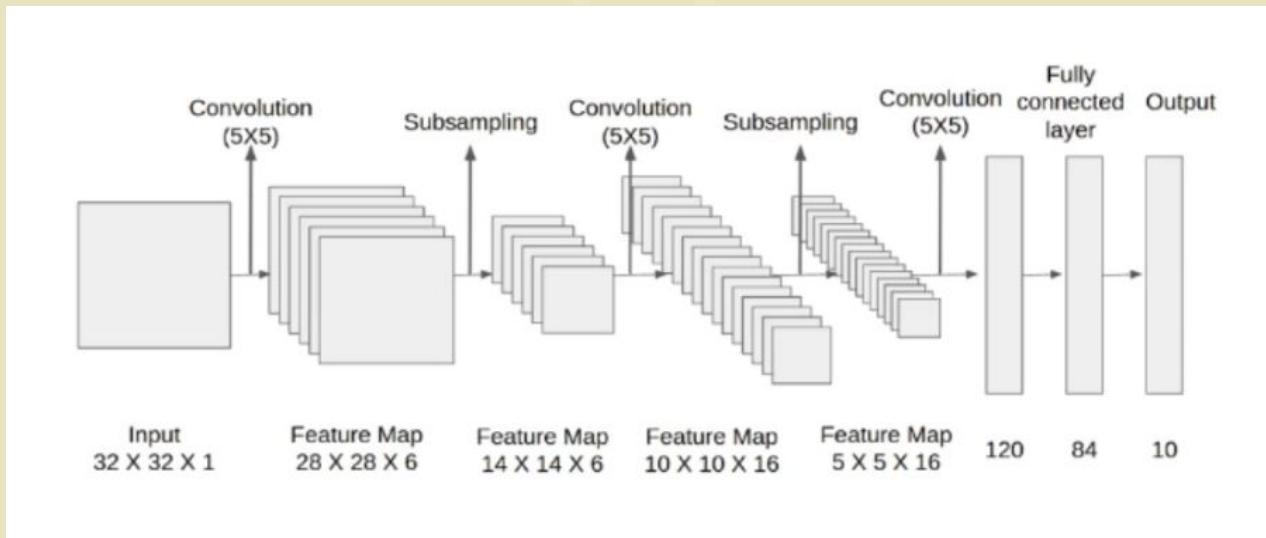
1998

Yann LeCun

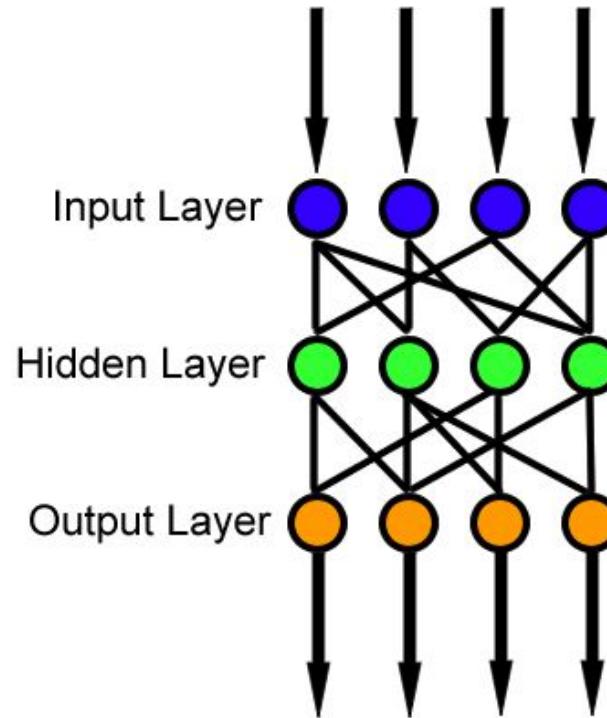




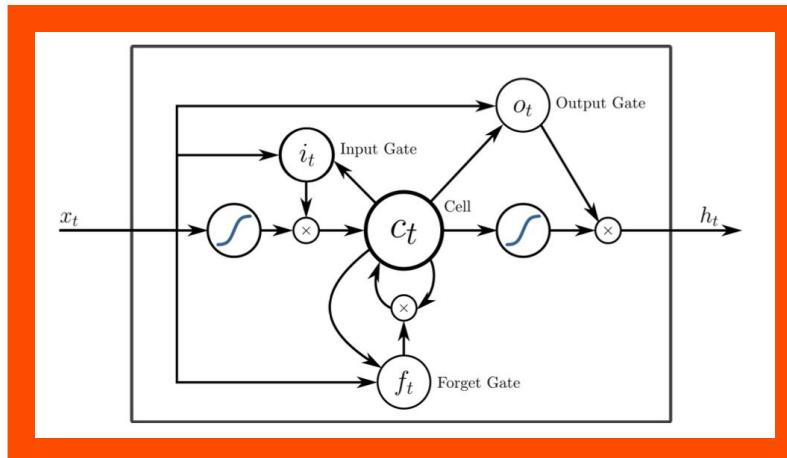
Modelo LeNet



Modelo Feed Forward



LSTM Model



Conexiones de retroalimentación



Long Short Term Memory



Escritura a mano,
reconocimiento de voz,
traducción por
máquina



Teoría Musical!



La música y el ruido son ambas mezclas de ondas de sonidos de diferentes frecuencias. Las frecuencias que componen a la música se pueden identificar como discretas. Mientras el ruido esta compuesto de frecuencias aleatorias en donde es difícil identificar las frecuencia dominante.

–Glenn Elert





Objetivos Específicos



Venus

Venus is the second planet from the Sun



Mars

Despite being red, Mars is a very cold place



Mercury

Mercury is the closest planet to the Sun



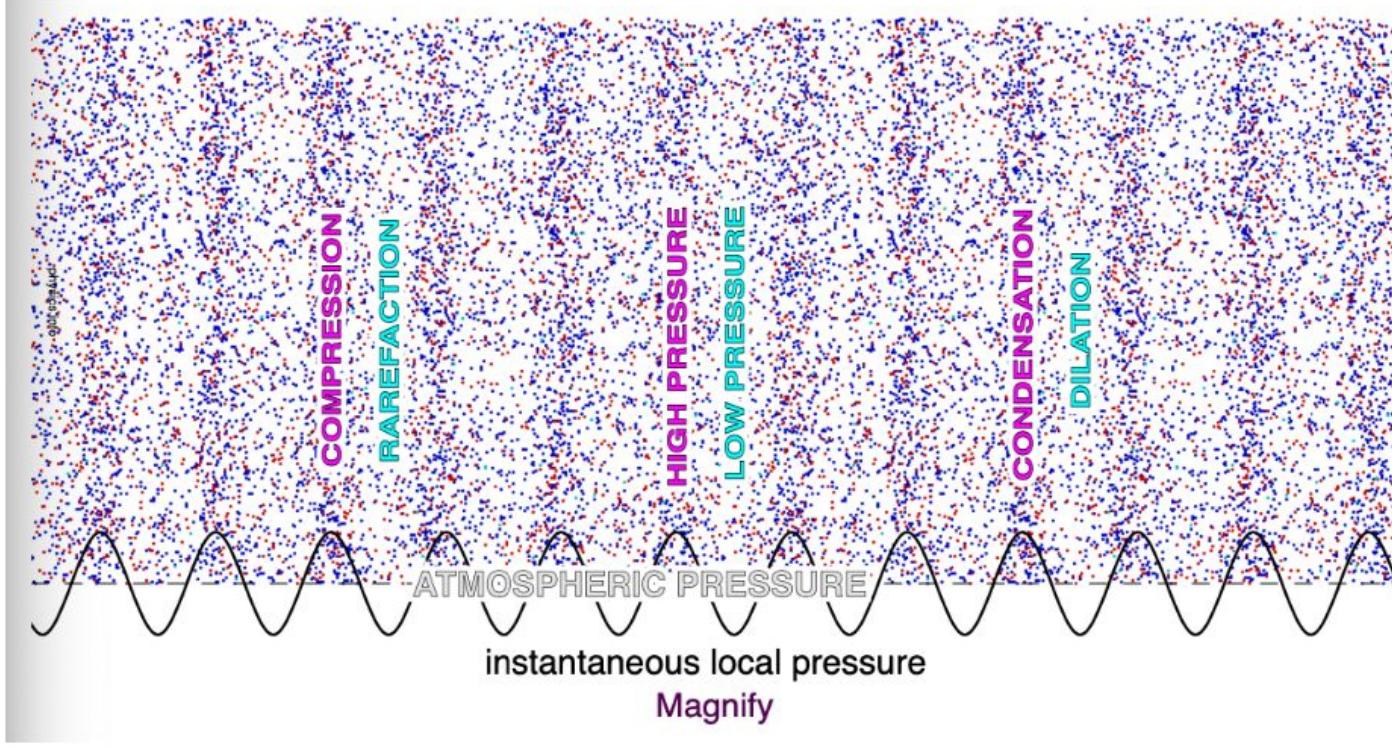
Saturn

Saturn is composed of hydrogen and helium





Sonido



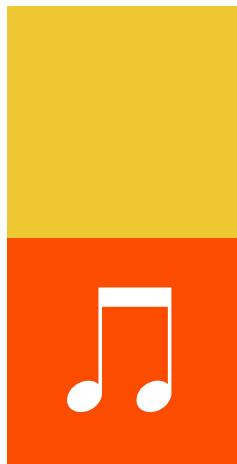
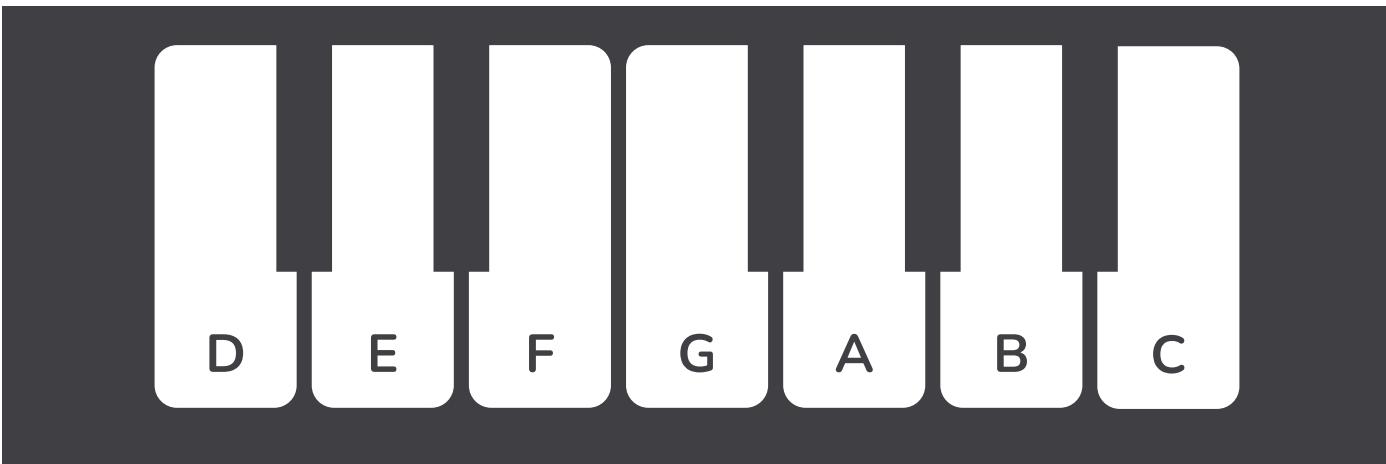


Sonido

$$Y(t) = A \sin(2\pi f t)$$



Nota Musical





Ritmo

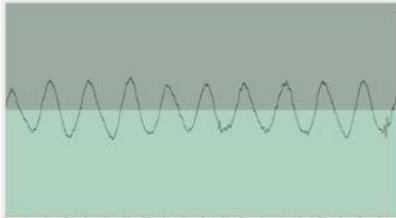
El ruido es la capacidad directa que tiene el ser humano de procesar información temporal. La percepción humana de estos patrones está determinada por la duración de las figuras musicales.

—Krumhansl

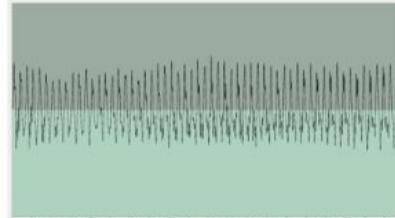




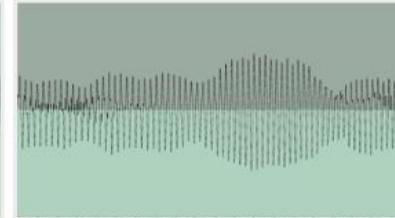
Combinación lineal en la música



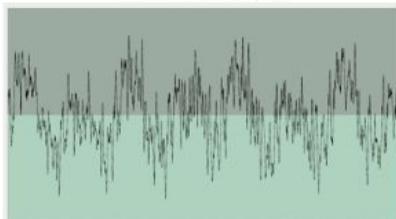
Bass voice singing



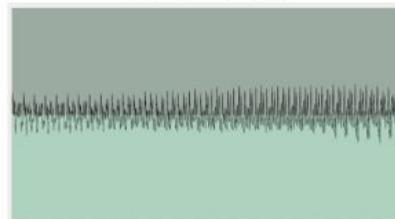
Tenor voice singing



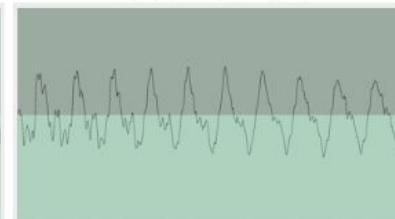
Soprano voice singing



Pipe organ



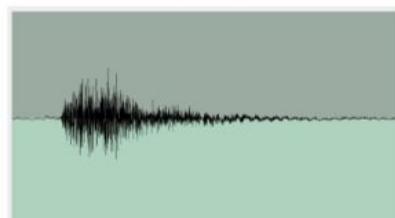
Trumpet



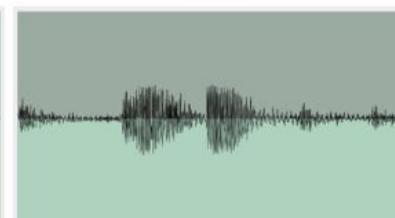
Violin



Conga drum



High hat cymbal

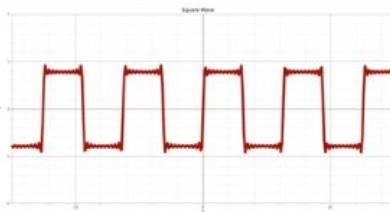


Woodblock

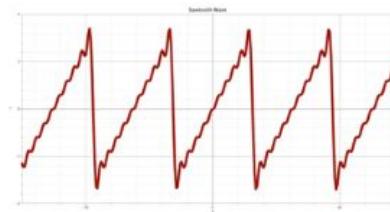


Combinación lineal en la música

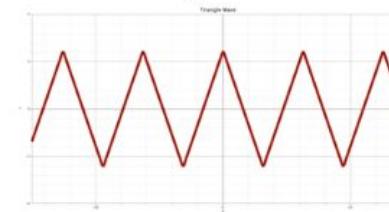
square wave



sawtooth wave



triangle wave



$$y = \sum \frac{1}{2n-1} \sin(2n-1)x$$

$$y = \sum \frac{(-1)^{n+1}}{n} \sin nx$$

$$y = \sum \frac{1}{(2n-1)^2} \cos(2n-1)x$$





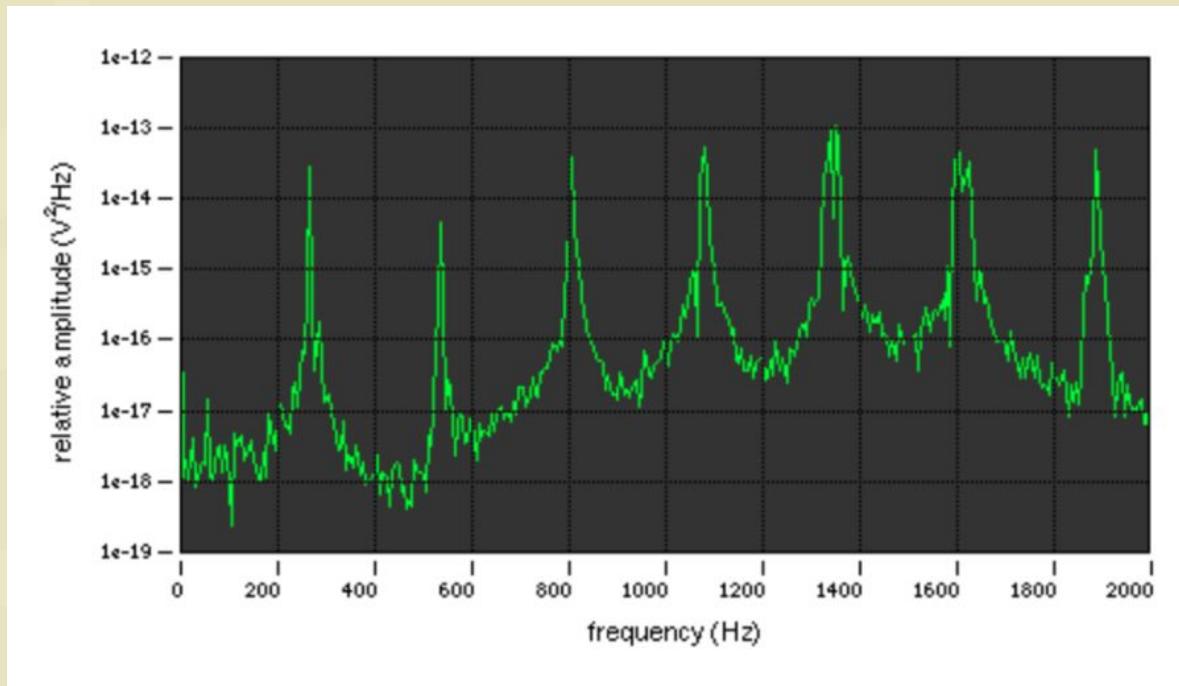
Análisis de Fourier

$$\int_{-\infty}^{\infty} f(x) e^{-i2x\pi\epsilon} dx$$





Análisis de Fourier





Síntesis de sonido

Método	Definición
Síntesis Aditiva	Se crea una onda resultante de las ondas almacenadas en la memoria.
Modulación de Frecuencias	La síntesis FM necesita tan sólo dos osciladores: la señal portadora y la señal moduladora.
Síntesis por tabla de ondas	Con pequeños fragmentos procedentes de sonidos reales, digitalizados y almacenados en ROM
Sampler	El «sampler» a diferencias de los métodos anteriores permite que el sonido sea guardado en RAM y no en ROM

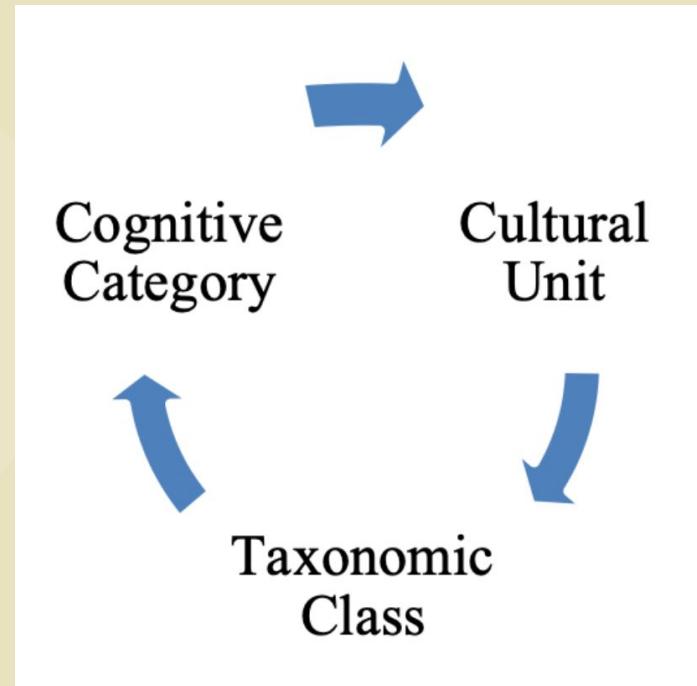




¿Qué es un Género Musical?



Ciclo de tres fases de Mendoza

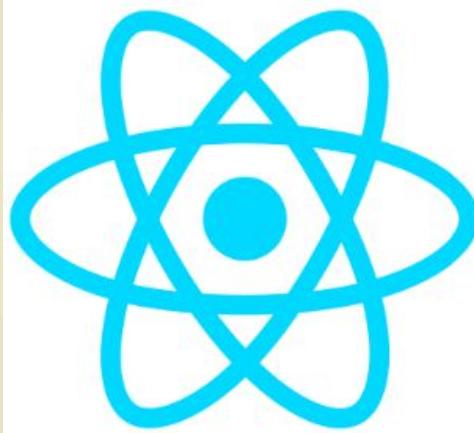




Herramientas y modelos!



Herramientas



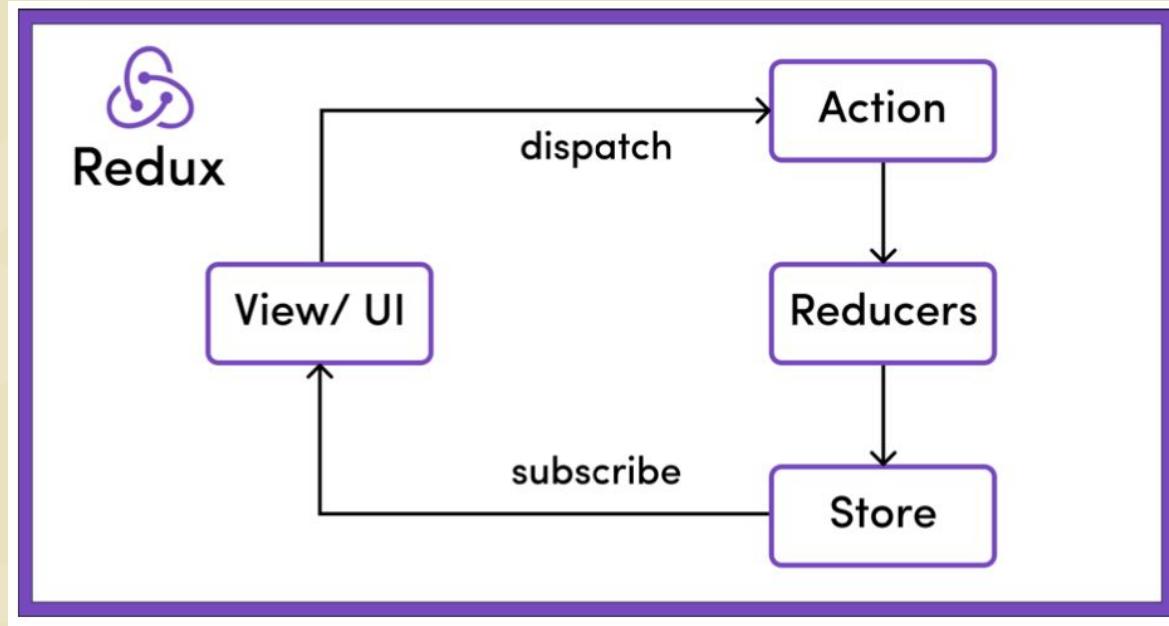
React





Redux







Flask

web development,
one drop at a time





TensorFlow





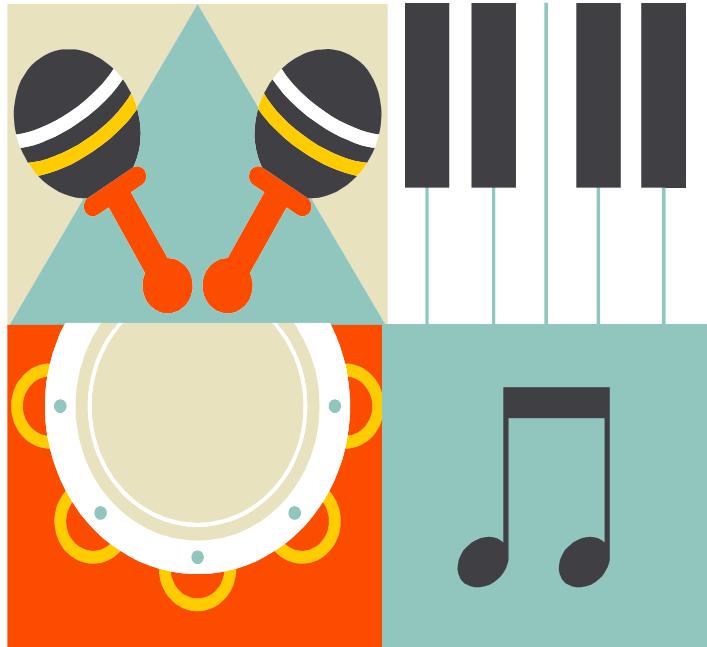
K





Modelos!





O3

Metodología

Metodología



Obtención de Datos

Obtención de datos de los servidores de servicios de Streaming.



Unificación de datos

Obtener un único set de datos

Metodología



Análisis Exploratorio

Realizar un análisis de los datos obtenidos



Elección de algoritmos

Escoger 3 algoritmos.

Metodología



Elaboración de los algoritmos

Realizar un análisis de los datos obtenidos



Elaboración del sitio web

Escoger 3 algoritmos.

Metodología



Publicación del sitio web

Realizar un análisis de los datos obtenidos

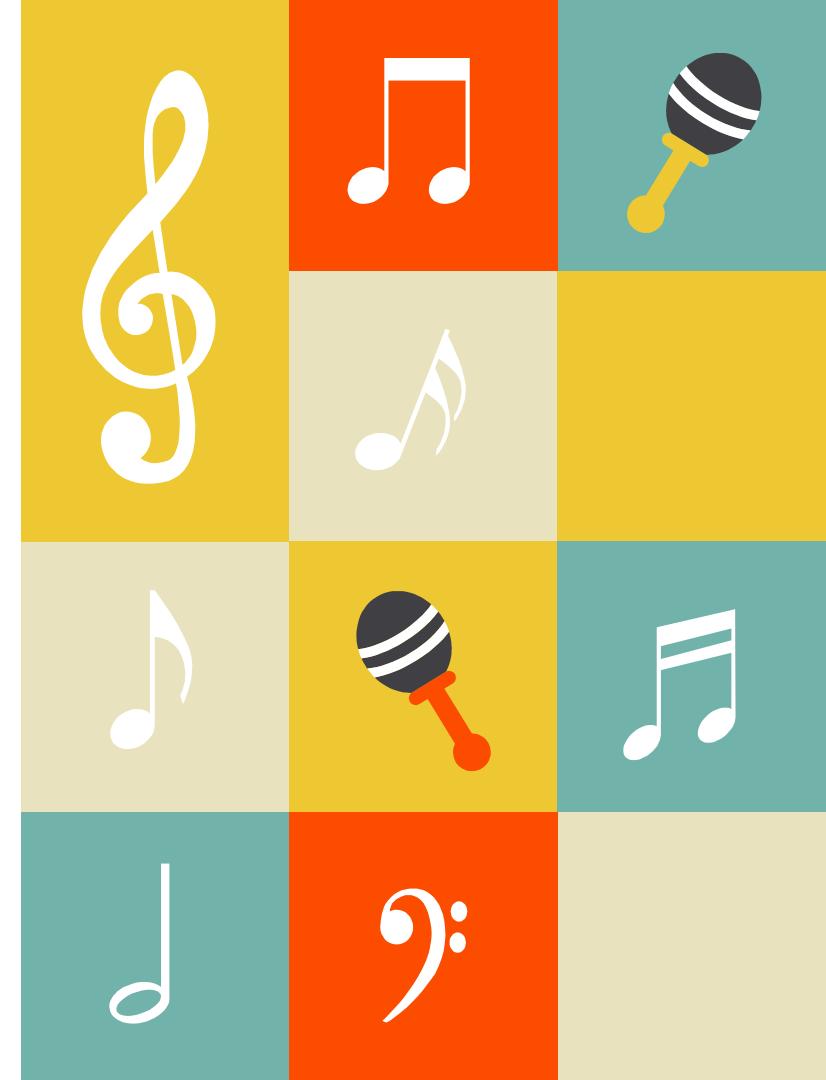


Verificación de los datos

Escoger 3 algoritmos.

05

Desarrollo



Obtención de datos





Ventajas de Soundcloud API



Género

Estructura de DB

El género de las canciones está estrictamente relacionado con las mismas canciones



El género es asignado por el artista.



Ventajas de Soundcloud API



Género

Las canciones son descargables o se puede encontrar canciones descargables entre los datos.



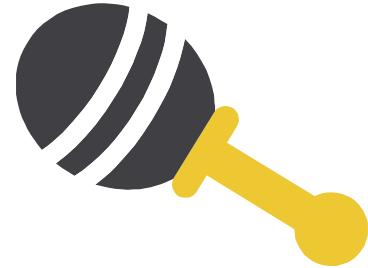
El acceso a los datos es gratuito



Cumbia	Indian	Rap/Hiphop/hip-hop/hip hop
Bachata	Arabic	Funk
Salsa	Soul	Afrobeat
Reggae	Dance/EDM/Electronic/Electro	Dembow
Gospel	Alternative/Indie	Bossa Nova
Punk	Future/Futuristic	Rock
Latin	Folk/Folkloric	Pop
Trap	Norteña/Nortena/Banda/Corridos	Jazz
Country	R&B/R & B/rnb	House
Disco	Reggaeton/Reguetón	Oriental
Acoustic/Singer/Songwriter	Classical/Classic	Metal
Hardstyle	Samba/Brazilian	Acapella
World	Ambient	HardCore

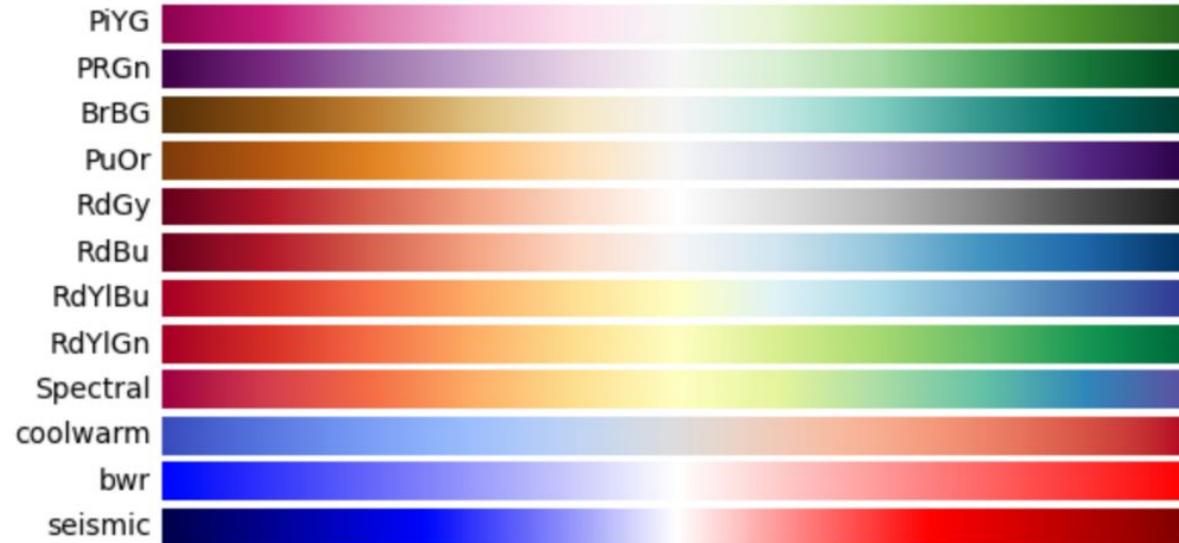


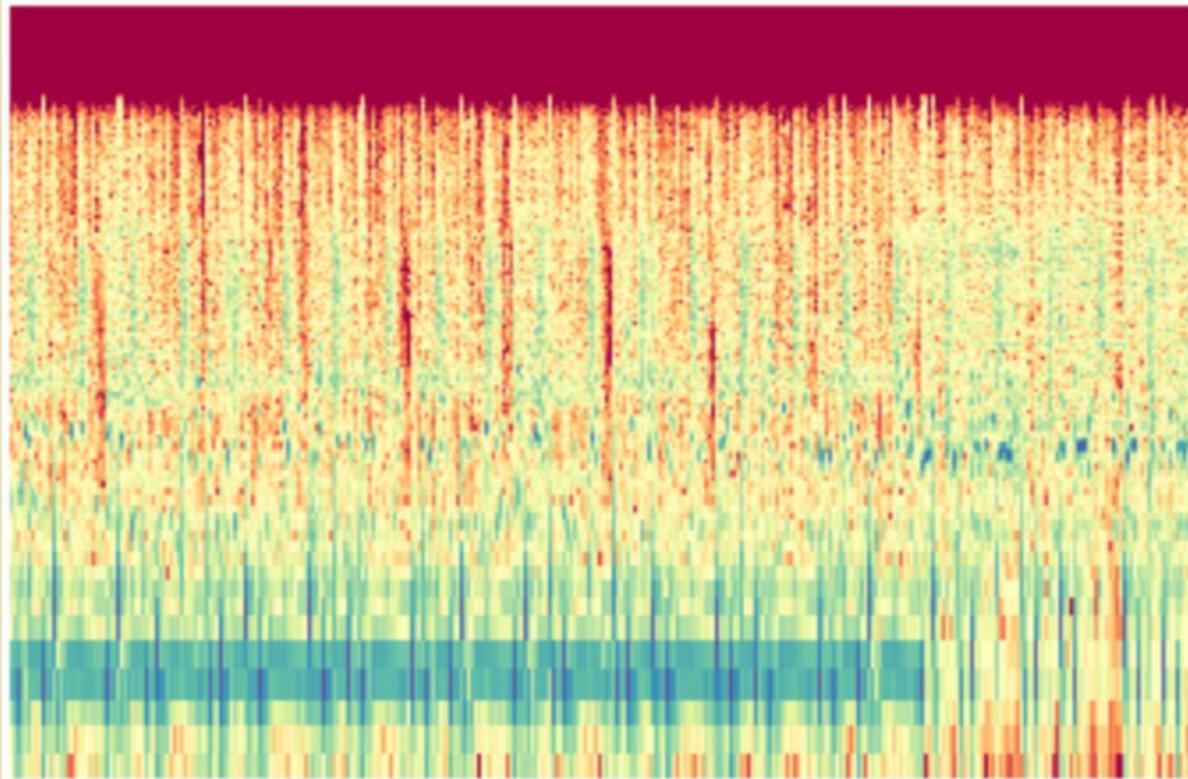
Procesamiento de datos

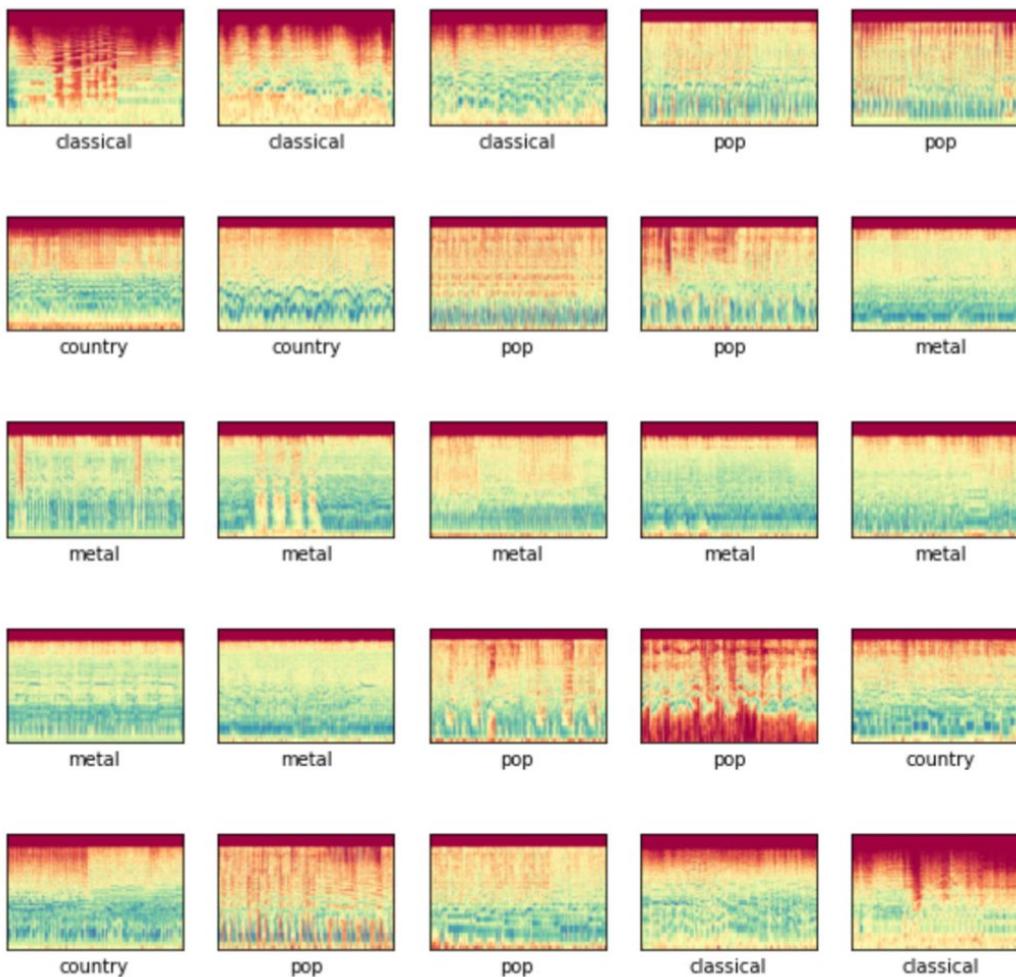


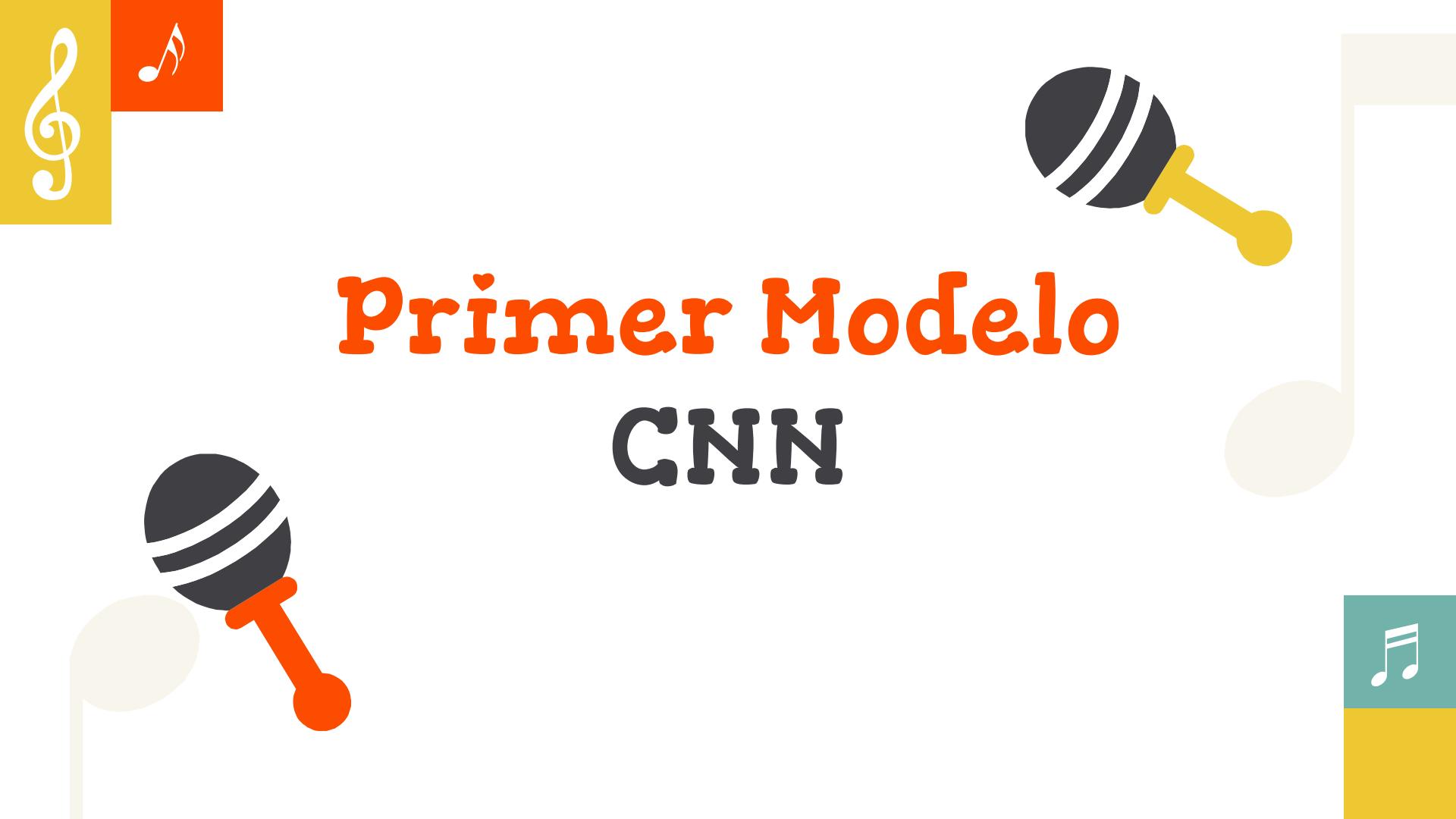


Diverging colormaps









Primer Modelo CNN



Género/Clase	Cantidad de canciones
Jazz	201
Metal	102
Country	104
Blues	105
Hip Hop	317
Disco	193
Reggae	151
Pop	133
Rock	156
Classical	145





Model: "sequential"

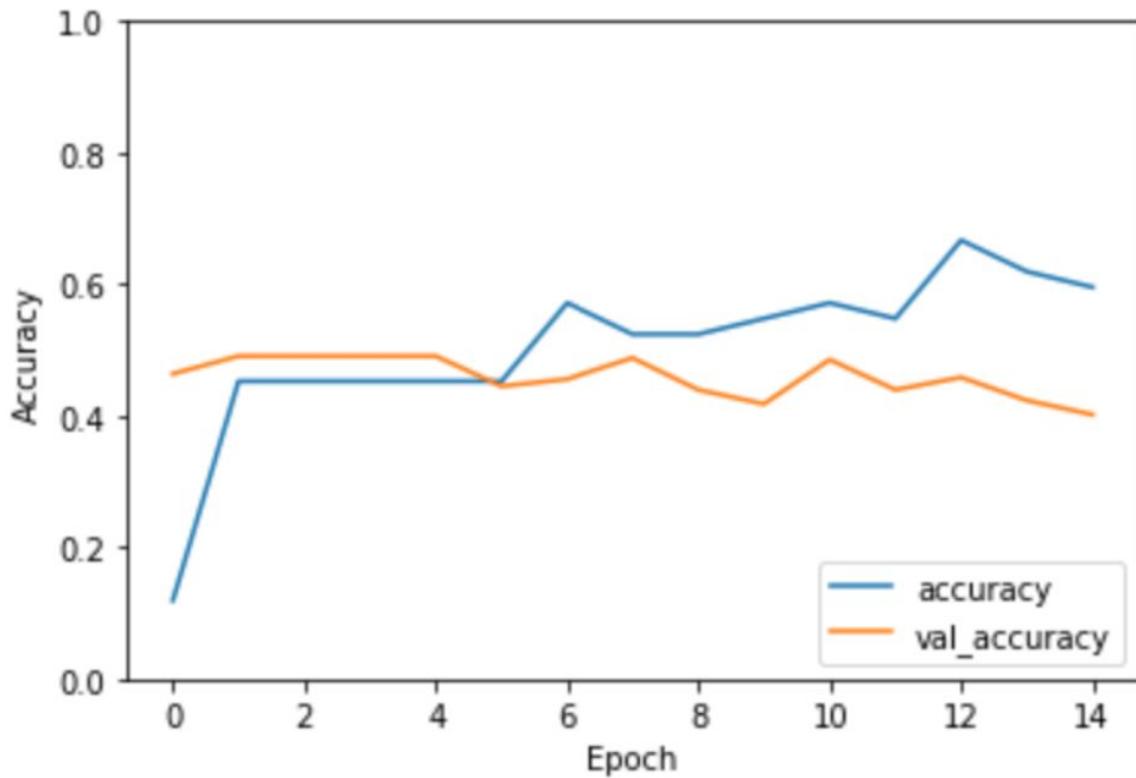
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 200, 200, 32)	896
max_pooling2d (MaxPooling2D)	(None, 100, 100, 32)	0
conv2d_1 (Conv2D)	(None, 98, 98, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 49, 49, 64)	0
conv2d_2 (Conv2D)	(None, 47, 47, 64)	36928
flatten (Flatten)	(None, 141376)	0
dense (Dense)	(None, 64)	9048128
dense_1 (Dense)	(None, 4)	260

Total params: 9,104,708

Trainable params: 9,104,708

Non-trainable params: 0





Elección de Géneros



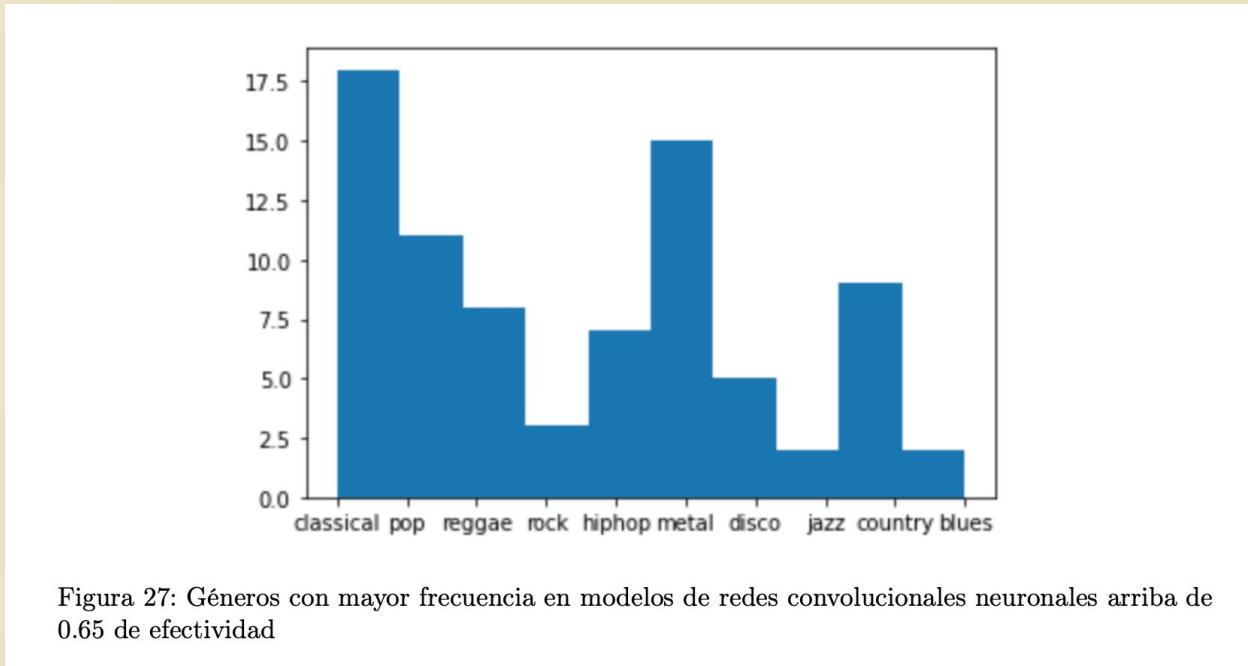


Figura 27: Géneros con mayor frecuencia en modelos de redes convolucionales neuronales arriba de 0.65 de efectividad



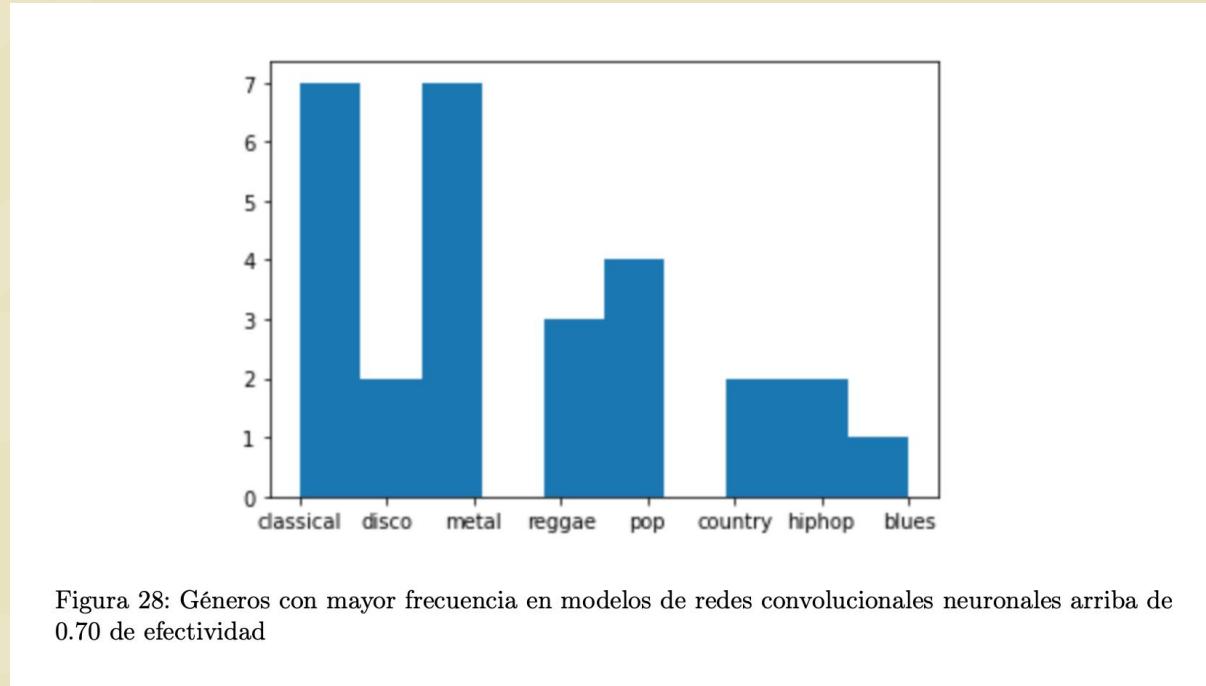


Figura 28: Géneros con mayor frecuencia en modelos de redes convolucionales neuronales arriba de 0.70 de efectividad



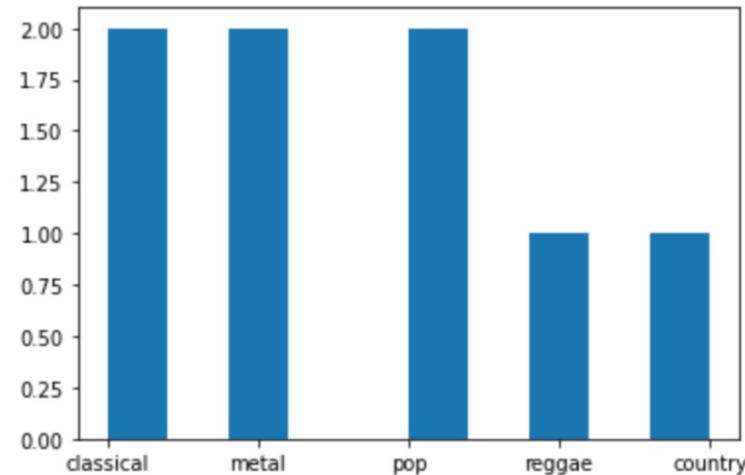


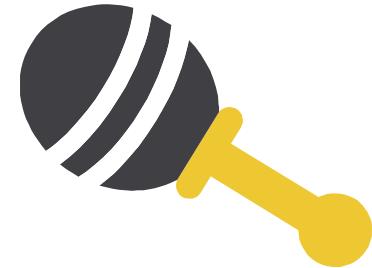
Figura 29: Géneros con mayor frecuencia en modelos de redes convolucionales neuronales arriba de 0.75 de efectividad



Género	Cantidad de canciones
Metal	102
Country	104
Pop	133
Classical	145



Modelo Feed-Forward



Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 198, 198, 64)	1792
flatten (Flatten)	(None, 2509056)	0
dense (Dense)	(None, 16)	40144912
dense_1 (Dense)	(None, 4)	68

Total params: 40,146,772

Trainable params: 40,146,772

Non-trainable params: 0

Figura 30: Estructura del modelo Feed Forward

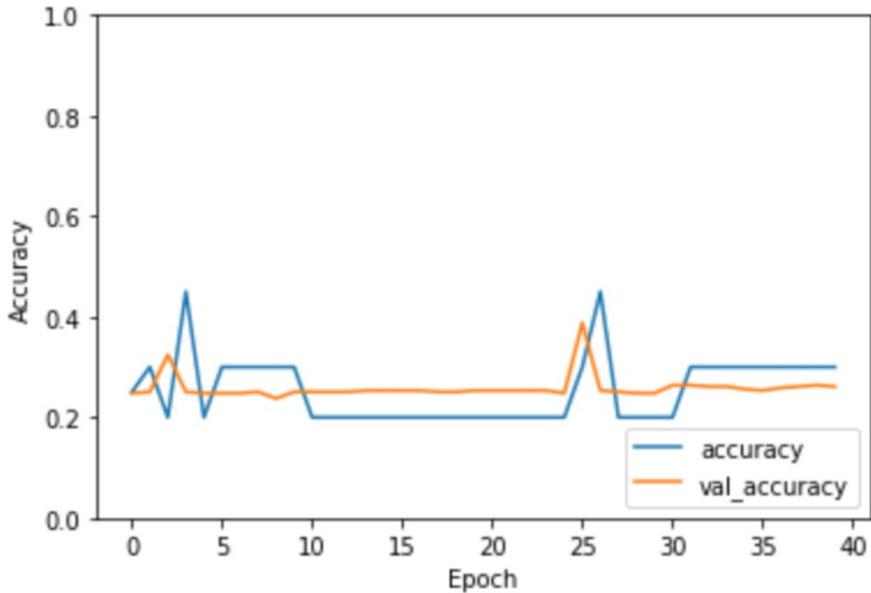


Figura 31: Efectividad del modelo Feed Forward

Modelo Le-Net





Model: "sequential"

Layer (type)	Output Shape	Param #
=====		
conv2d (Conv2D)	(None, 198, 198, 6)	168
average_pooling2d (AveragePo	(None, 99, 99, 6)	0
conv2d_1 (Conv2D)	(None, 97, 97, 16)	880
average_pooling2d_1 (Average	(None, 48, 48, 16)	0
flatten (Flatten)	(None, 36864)	0
dense (Dense)	(None, 128)	4718720
dense_1 (Dense)	(None, 4)	516
=====		
Total params: 4,720,284		
Trainable params: 4,720,284		
Non-trainable params: 0		



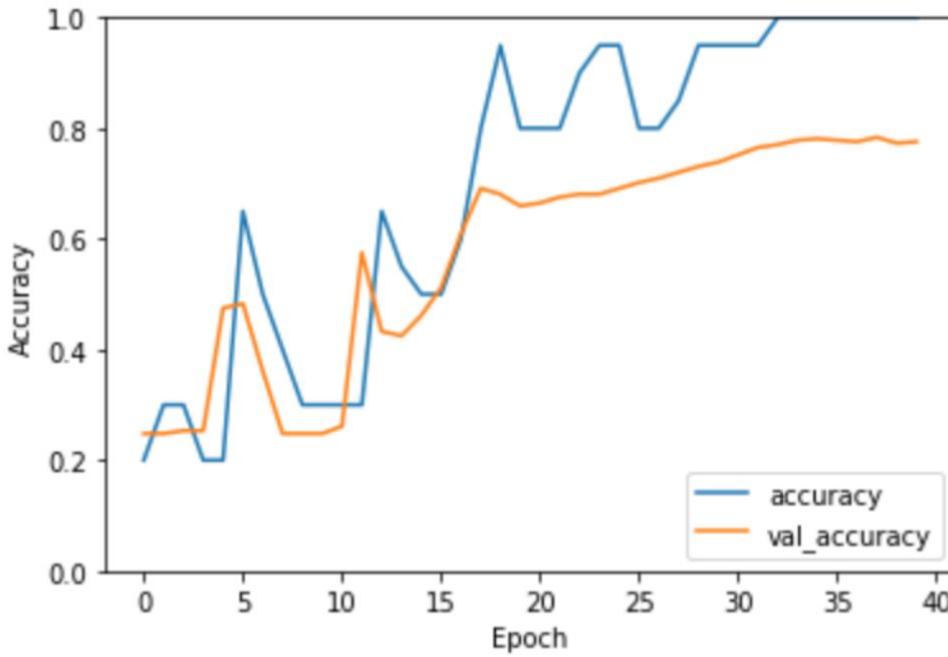
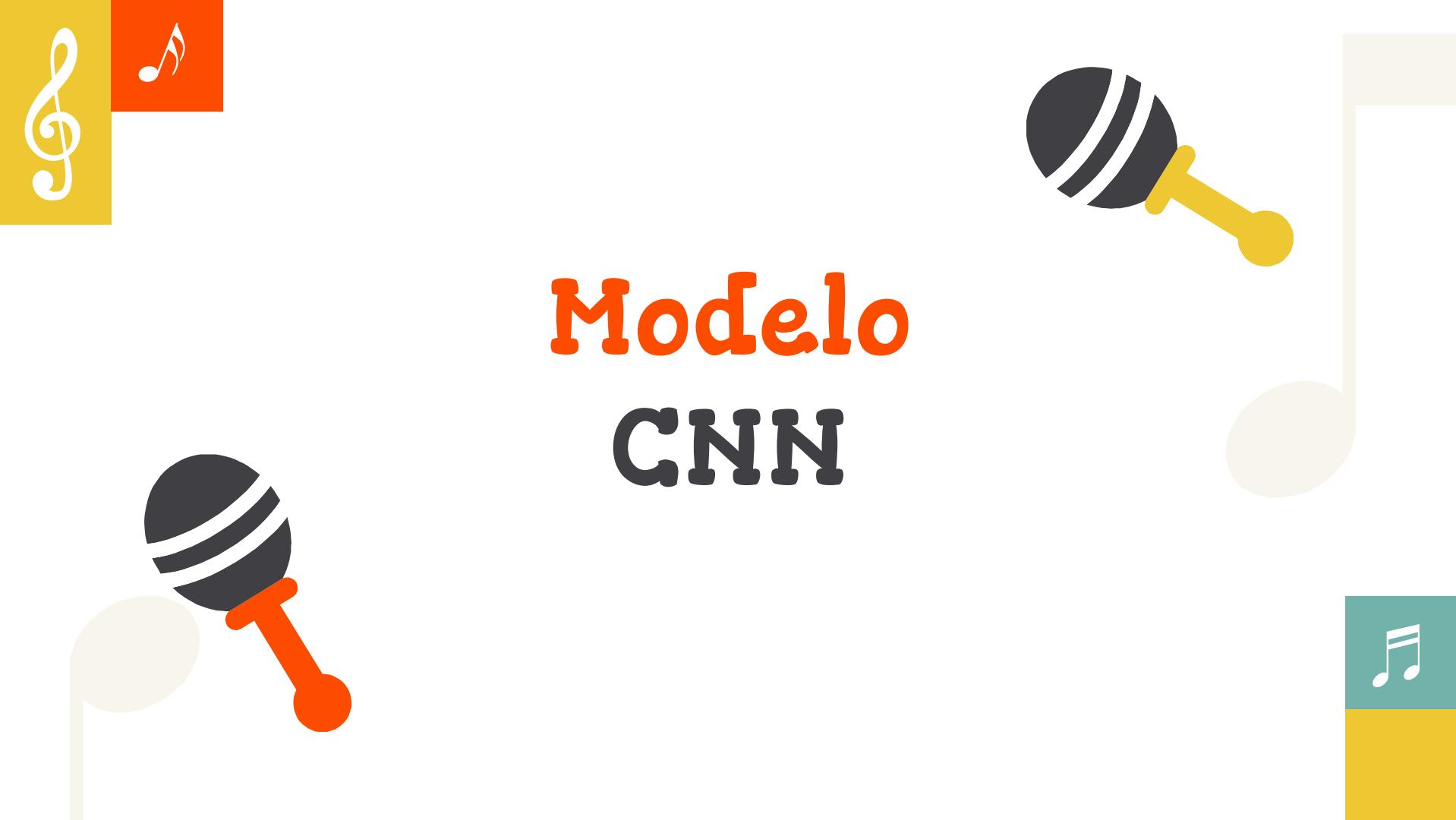


Figura 33: Efectividad del modelo Le-Net

Modelo CNN





Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 200, 200, 32)	896
max_pooling2d (MaxPooling2D)	(None, 100, 100, 32)	0
conv2d_1 (Conv2D)	(None, 98, 98, 64)	18496
max_pooling2d_1 (MaxPooling2 ((None, 49, 49, 64)	0
conv2d_2 (Conv2D)	(None, 47, 47, 64)	36928
flatten (Flatten)	(None, 141376)	0
dense (Dense)	(None, 64)	9048128
dense_1 (Dense)	(None, 4)	260
<hr/>		
Total params: 9,104,708		
Trainable params: 9,104,708		
Non-trainable params: 0		



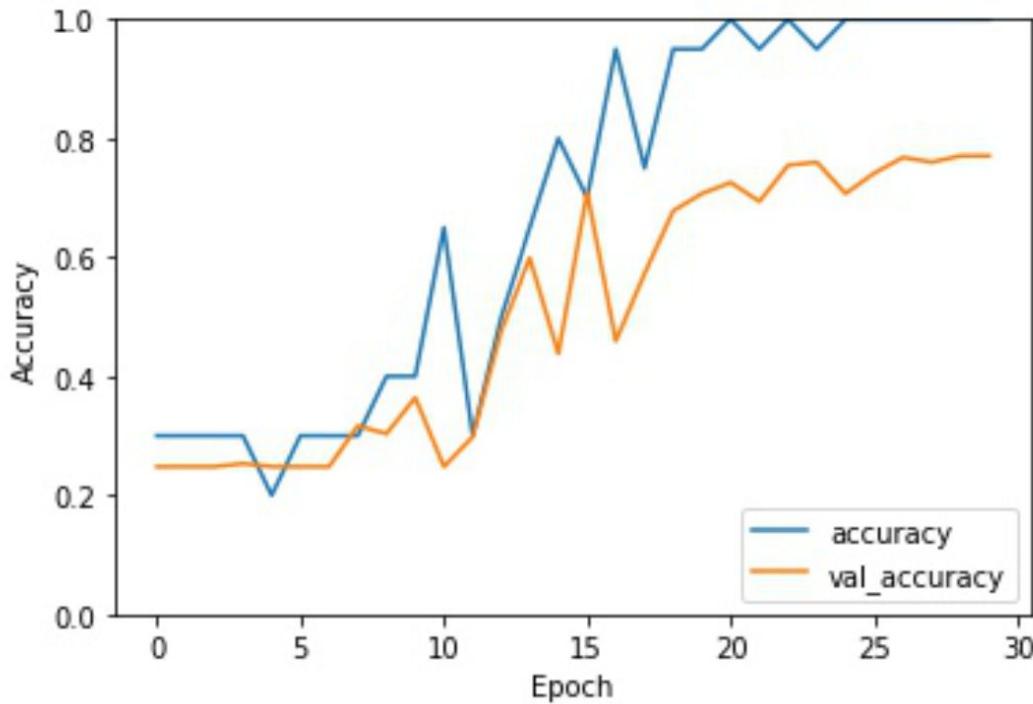


Figura 35: Efectividad del modelo CNN

Modelo LSTM





Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 198, 198, 6)	168
average_pooling2d (AveragePo	(None, 99, 99, 6)	0
conv2d_1 (Conv2D)	(None, 97, 97, 16)	880
average_pooling2d_1 (Average	(None, 48, 48, 16)	0
flatten (Flatten)	(None, 36864)	0
reshape (Reshape)	(None, 1, 36864)	0
lstm (LSTM)	(None, 128)	18940416
dense (Dense)	(None, 4)	516

Total params: 18,941,980

Trainable params: 18,941,980

Non-trainable params: 0



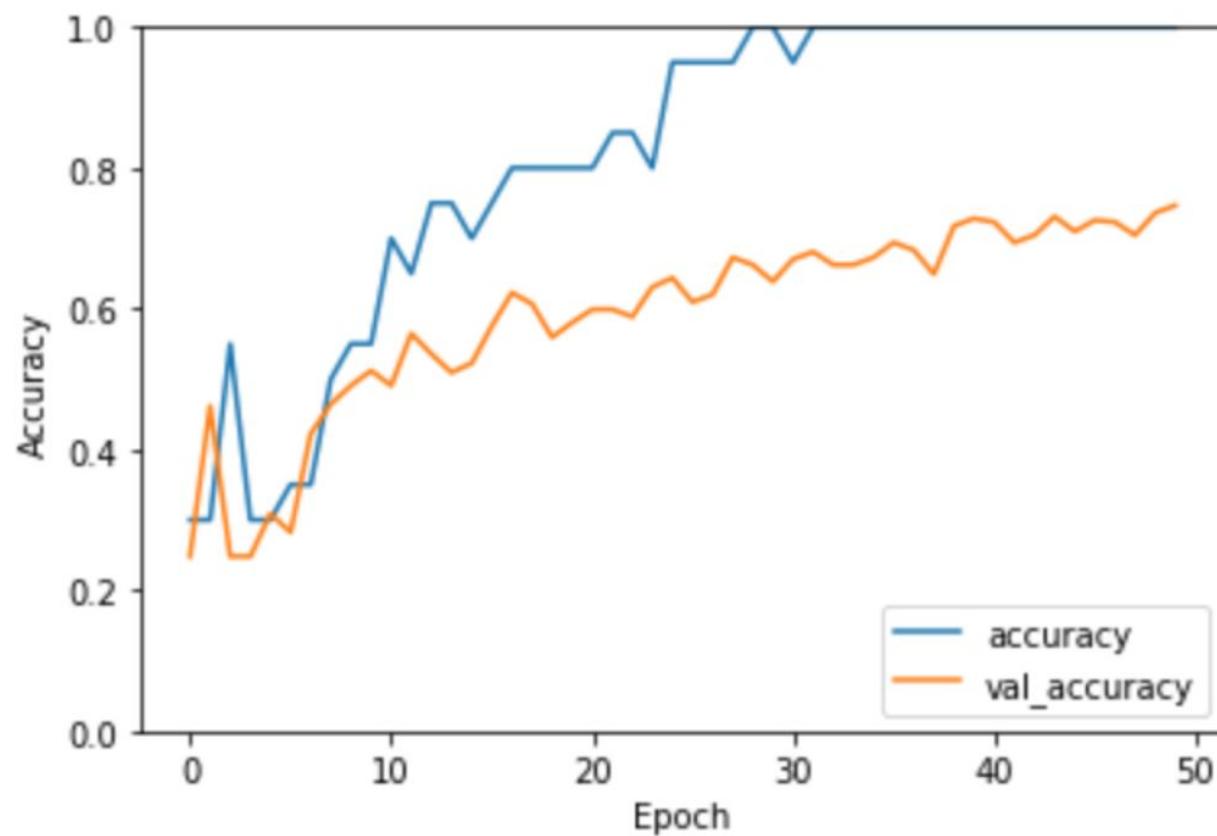


Figura 37: Efectividad del modelo LSTM

Página Web (Backend)



14.1. Estructura de respuesta

```
{  
    "CNN Model": {  
        "data": {  
            "classical": "0.32583416",  
            "country": "0.20497286",  
            "metal": "0.0065581384",  
            "pop": "0.46263486"  
        },  
        "metrics": {  
            "accuracy": "0.7704485654830933",  
            "loss": "0.7446410059928894\n"  
        }  
    },  
    "success": true  
}
```



Página Web (Frontend)



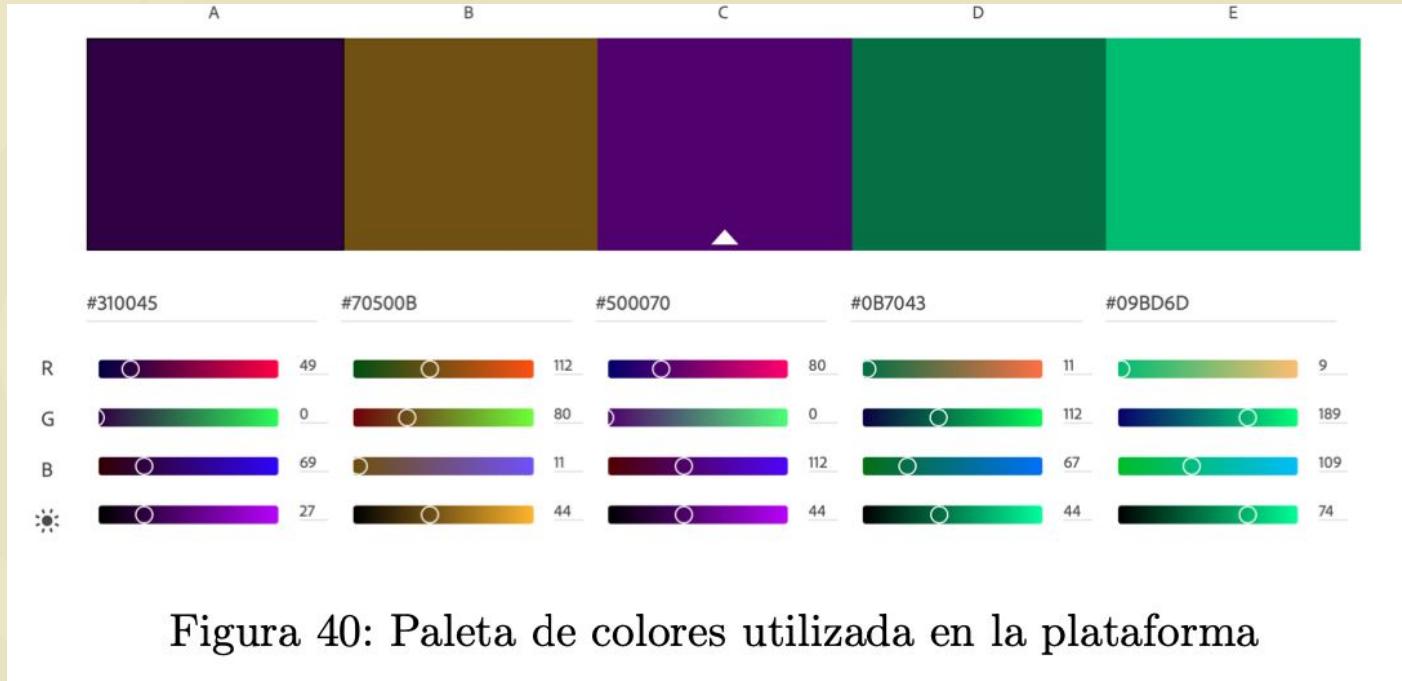


Figura 40: Paleta de colores utilizada en la plataforma





The screenshot shows a dark-themed web application for music genre detection. At the top left is a green logo consisting of a stylized 'V' and 'M'. To its right, the text "MUSIC GENRE DETECTOR" is displayed in white. On the far right of the header are three links: "Home", "How does it work?", and "Documentation". The main title "Find the genre" is centered in large, bold, white font. Below it, the subtitle "of your track" is shown in a smaller, white, sans-serif font. In the center of the page is a dashed blue rectangular input field with a small blue file icon and the text "Upload or drop a file right here". To the right of the input field, the file types ".mp3, .wav" are listed. A green horizontal bar at the bottom contains the text "How it works?" in white. The background of the page is black.

MUSIC GENRE DETECTOR

Home How does it work? Documentation

Find the genre
of your track

Upload or drop a file right here

.mp3, .wav

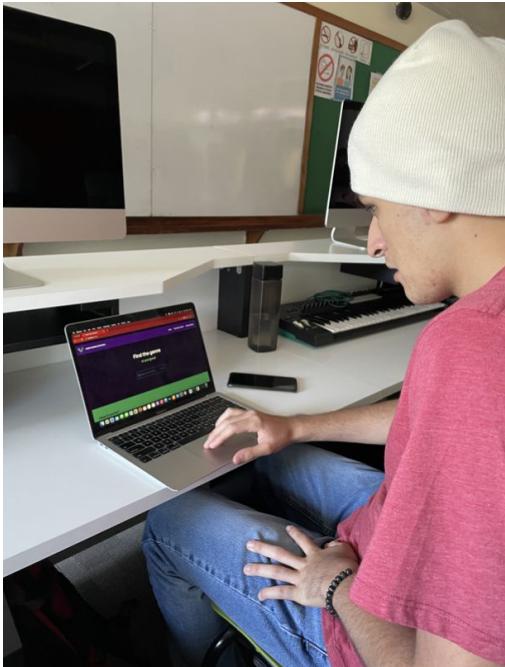
How it works?



Pruebas de usuario



Pruebas de usuario



Usuario 1

Música Experimental y
composición de arreglos
musicales

Problem & solutions



User 2

Manejo y organización de
eventos musicales



Pruebas de usuario



Usuario 3

Compositor y productor
musical

Pruebas de usuario



Usuario 4

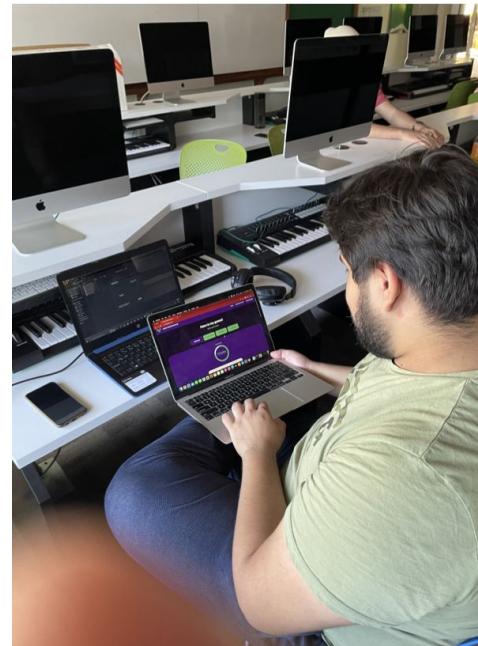
Música Experimental y
composición de arreglos
musicales

Problem & solutions



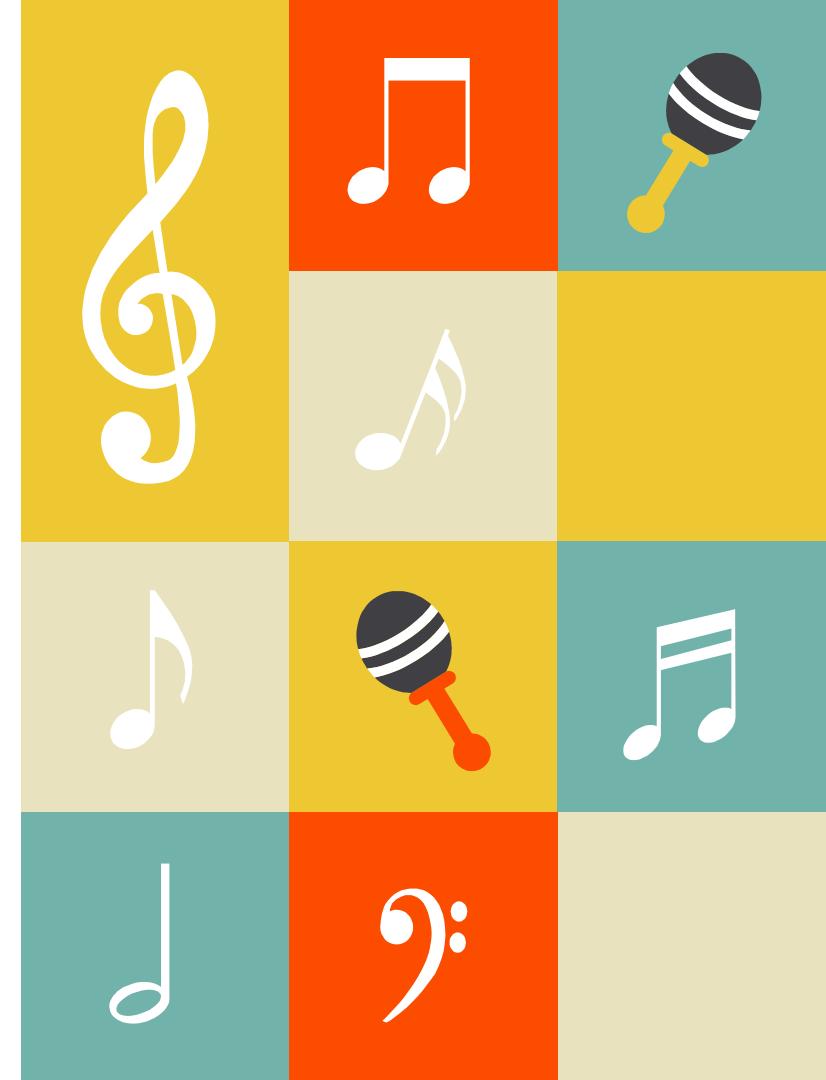
Usuario 5

Composer y Compositor de
música



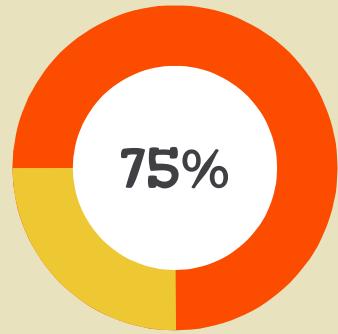
06

Conclusiones





Conclusiones



Los modelos de redes convoluciones representan un alcance positivo para poder identificar géneros con una efectividad mayor al 75 %.



SoundCloud API representa una base de datos amplia que permitirá mejorar la efectividad de los modelos para clasificar géneros musicales.





Conclusiones



Las redes convoluciones son funcionales en la clasificación de géneros musicales si se puede eliminar el exceso de ruido que existe entre algunos géneros musicales

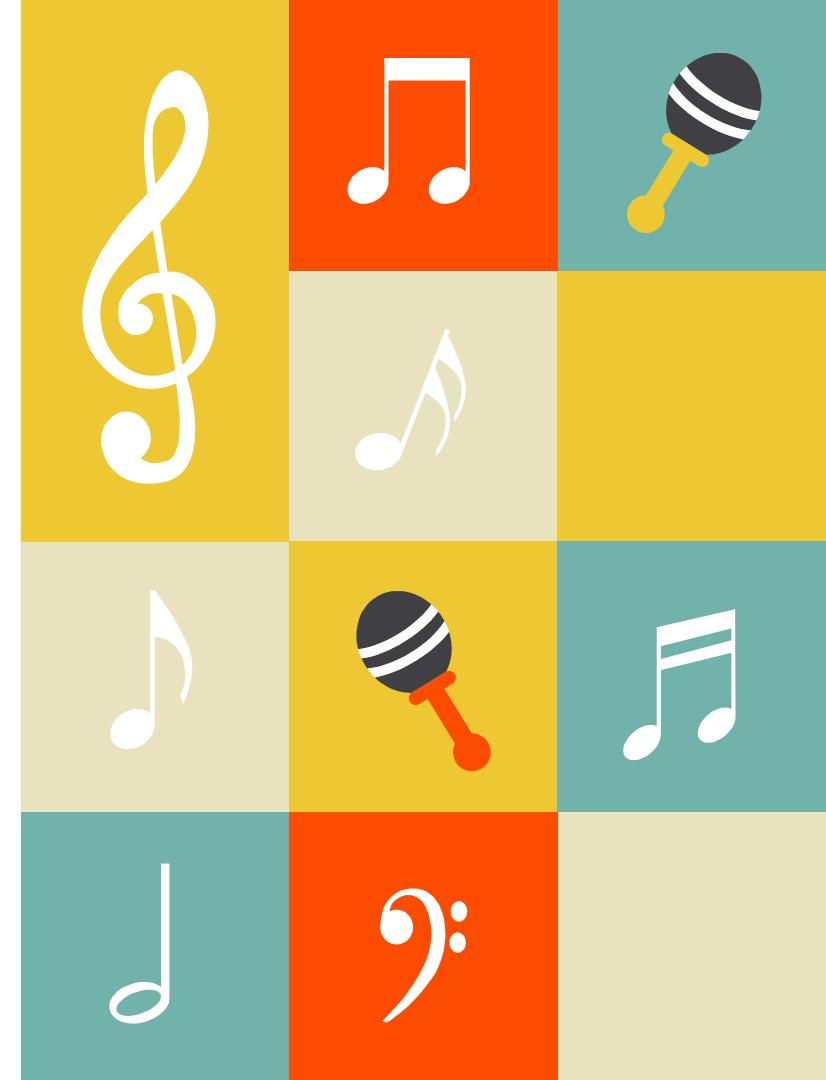


Es posible construir un sitio web público en donde se puede identificar el género musical de una pista de audio.



07

Recomendaciones





Recomendaciones



Mercury

Obtener un conjunto de datos más limpios, con ayuda de musicólogos clasificar cada pista dependiendo de las características teóricas que posee cada género.



Considerar el uso de algoritmos supervisados para la realización de este proyecto.





Recomendaciones



Mercury

Acortar el tamaño de los espectrogramas, es decir hacer un acercamiento a la imagen resultante para poder eliminar las frecuencias más altas y las más bajas.



Realizar un análisis por instrumento.



Recomendaciones



Mercury

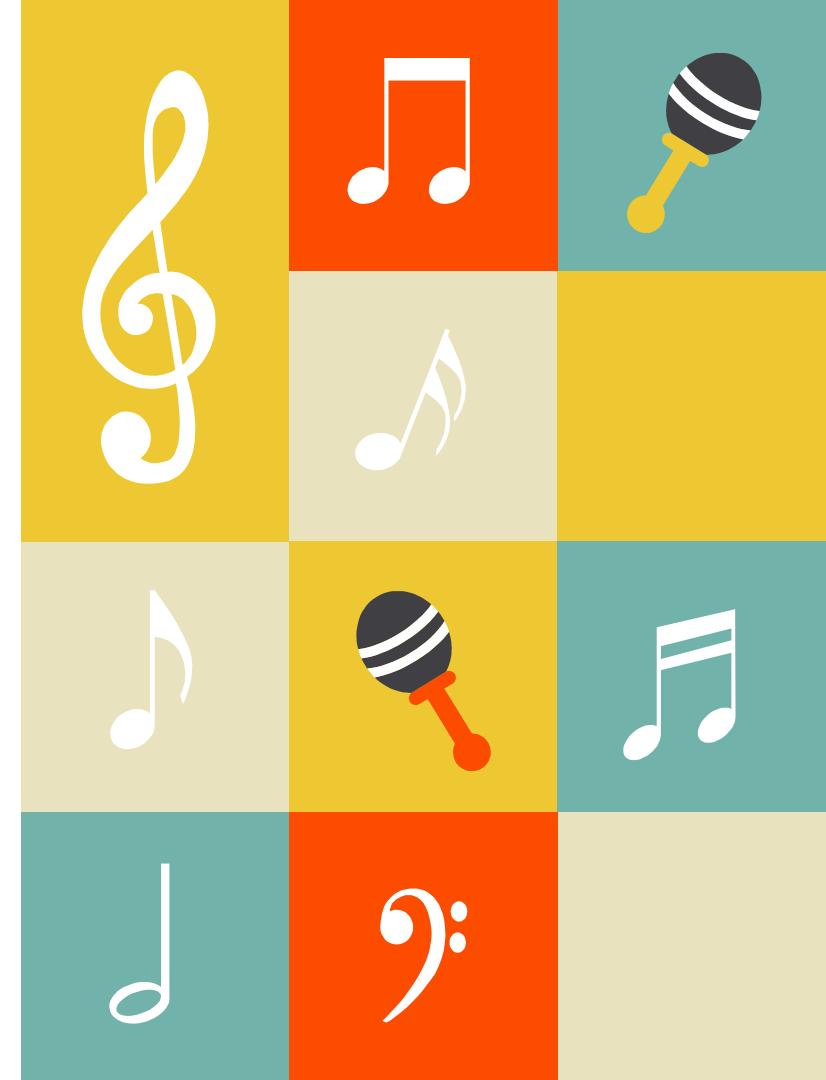
Realizar conexión con labels y miembros de la industria musical para poder enviar notificaciones automáticas a los A&R cuando una canción que se adecúa a su trabajo se sube a la plataforma



Realizar conexiones con productores y artistas

08

Referencias





Resources

- Adobe. (2022). Color Wheel. Adobe Color. Retrieved September 3, 2022, from <https://color.adobe.com/create/color-wheel>
- Aprende machine learning. (2017, November 4). Principales Algoritmos utilizados. Aprende Machine Learning. Retrieved May 10, 2022, from <https://www.aprendemachinelearning.com/principales-algoritmos-usados-en-machine-learning/>
- Burgess, Director of Marketing and Sales Richard James. (2013). The Art of Music Production: The Theory and Practice(4th ed.). Oxford University Press, USA.
- Changsheng Xu, Maddage, N. C., Xi Shao, Fang Cao, Qi Tian. (n.d.). Musical genre classification using support vector machines. 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). doi:10.1109/icassp.2003.1199998
- D. Michie, D.J. Spiegelhalter, C.C. Taylor (eds). Machine Learning, Neural and Statistical Classification, 1994





Resources

- Hochreiter, S., Schmidhuber, J. (1997, November 1). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
<https://doi.org/10.1162/neco.1997.9.8.1735>
- Elert, G. (2022). The Physics Hypertextbook. <https://physics.info>
- Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, 126(1), 159–179. doi:10.1037/0033-2909.126.1.159
- Lind, D., Marchal, W., Wathen, S. (2011). *Statistical Techniques in Business and Economics*. McGraw-Hill Education.
- Matplotlib — Visualization with Python. (n.d.). Retrieved October 22, 2022, from <https://matplotlib.org/>
- Mendoza-Halliday, P. (n.d.). A Theory of the Musical Genre: The Three-Phase Cycle. Programa De Doctorado En Música UNAM, Mexico, Proceedings of the 10th Inter-national Conference of Students of Systematic Musicology (SysMus17), London, UK, September 13-15, 2017. Peter M. C. Harrison (Ed.).





Resources

- Contreras S. Music Genre Detector (2022) Music Genre Detector Platform created in this project. Visitar
<https://suulcoder.github.io/music-genre-detector/>
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D. (1989, December). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4), 541–551.
<https://doi.org/10.1162/neco.1989.1.4.541>
- Lecun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- Loncomilla, P. (2016). Deep learning: Redes convolucionales. Retrieved May 11, 2022, from <https://ccc.inaoep.mx/pgomez/deep/presentations/2016Loncomilla.pdf>





Resources

- ParaBajoelectrico.com. (2022, May 10). Frecuencia de las Notas Frecuencias musicales (Hz) 2022.
<https://parabajoelectrico.com/frecuencia-notas-musicales/>
- Parra, F. (2019, enero 25). Estadística y Machine Learning con R.
<https://bookdown.org/>. Retrieved mayo 11, 2022, from
<https://bookdown.org/content/2274/modelo-lineal-general.htmlintroducción-1>
- Pelchat, N., Gelowitz, C. M. (2020). Neural Network Music Genre Classification. Canadian Journal of Electrical and Computer Engineering, 43(3), 170–173.doi:10.1109/cjece.2020.2970144
- Michels, U. (2022). Atlas de musica / Music Atlas: 1. Alianza.
- Russell, S. J., Russell, S. J., Norvig, P., Davis, E. (2010). Artificial Intelligence: A Modern Approach. Prentice Hall.





Resources

- Salas, R. (2004). Redes neuronales artificiales. academia.edu. Retrieved mayo 11, 2022, from
https://www.academia.edu/download/50358783/Redes_Neuronales_Artificiales.pdf
- Serway, R. A., Jewett, J. W. (2018, January 1). Physics for Scientists and Engineers, Volume 1 (10th ed.). Cengage Learning.
- SoundCloud Developers. (2022). Retrieved October 25, 2022, from
<https://developers.soundcloud.com/>
- Stewart, J. (2015). Calculus. Cengage Learning.
- Valencia-Aguirre, J., Daza-Santacoloma, G., Acosta, C. D., Castellanos-Domínguez, G. (2010, diciembre). Comparación de Métodos de Reducción de Dimensión basados en Análisis por Localidades. *Tecnológicas*, (25), 131-150.
<https://repositorio.itm.edu.co/bitstream/handle/20.500.12622/860/127-Manuscrito- 237-1-10-20170208.pdf?sequence=1isAllowed=y>





Resources

- Swales, J.M. (1990). *Genre Analysis*. Cambridge: Cambridge University Press.
- Zhou, Chellappa. (1988). Computation of optical flow using a neural network. *IEEE International Conference on Neural Networks*.
<https://doi.org/10.1109/icnn.1988.23914>
- Ingham, T. (2021, 1 marzo). Over 60,000 tracks are now uploaded to Spotify every day. That's nearly one per second. *Music Business Worldwide*.
<https://www.musicbusinessworldwide.com/over-60000-tracks-are-now-uploaded-to-spotify-daily-thats-nearly-one-per-second/>
- Redux · An Introduction. (2016, June 28). *Smashing Magazine*.
<https://www.smashingmagazine.com/2016/06/an-introduction-to-redux/>
- Jorda, S. (1997). *Guía Audio digital y MIDI*.
<https://www.ccapitalia.net/reso/articulos/audiodigital/pdf/09-Sintesis.pdf>





Resources

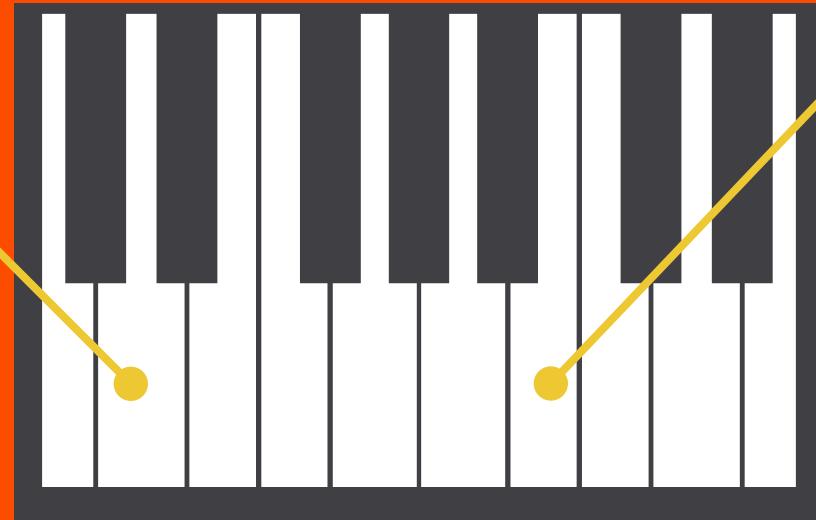
- React – A JavaScript library for building user interfaces. (2022). React. Retrieved Oc- tober 26, 2022, from <https://reactjs.org>
- Team, K. (2022). Keras: the Python deep learning API. Retrieved October 26, 2022, from <https://keras.io/>
- Welcome to Flask — Flask Documentation (2.2.x). (2022). Retrieved October 26, 2022, from <https://flask.palletsprojects.com/en/2.2.x/>
- Google. (2015, November 9). TensorFlow: Open source machine learning [Video]. You- Tube. https://www.youtube.com/watch?v=oZikw5k_2FM
- Fazt. (2017, December 3). Modelo Cliente Servidor, Explicación Simple [Video]. You- Tube. <https://www.youtube.com/watch?v=49zdlyLSwhQ>





¿Preguntas?

Este es el momento para hacer preguntas



Thanks!

CREDITS: This presentation template was created by [Slidesgo](#),
including icons by [Flaticon](#), and infographics & images by [Freepik](#)