

Memoria

**Trabajo Final de la Asignatura: Reconocimiento Notas Musicales**

**Grado:**

Ingeniería Informática

**Asignatura:**

| Inteligencia Artificial |
| --- |

**Alumnos:**

Víctor Ernesto Martínez Leal

Wilfredo Oswaldo Hernández Argueta

**Enero, 2022**

## 

[**INTRODUCCIÓN**](#_isbctlb0c8) **3**

[Notas musicales](#_rbq59pc52djq) 4

[Armónicos:](#_rg9f9tlauxwd) 4

[De dónde salen las notas musicales](#_1zddhcch2pgd) 5

[Nombrando las notas](#_gshxvjvgpiv8) 7

[Calculando nuevas notas](#_23q2u7351gnd) 7

[Cómo leer los sonidos](#_607uc12rmnnv) 8

[Volumen de los sonidos](#_o0dbs38kldnp) 9

[Inteligencia Artificial](#_mpmlsxg33xws) 10

[Red Neuronal](#_uks3t6f9t0ph) 11

[Deep Learning](#_jfli3q1d4ejl) 11

[Reconocimiento de Notas musicales](#_4camckxlwbzk) **13**

[Planteamiento](#_i7k3och43bhx) 14

[Paso 1 populado de datos:](#_9ska10vn62g6) 14

[Paso 2 extracción de datos:](#_orr5vp1e3pv4) 14

[Paso 3 tratado de datos:](#_40nm03nbo5ov) 15

[Paso 4 segregar data:](#_4t0pbhrwx8iv) 16

[Entrenamiento](#_gokqu5j71gpg) 16

[Pruebas](#_ssaoii3mth8j) 17

[Limitaciones](#_t4j0p880afhz) 19

[**Conclusiones**](#_ewk87xjb0tn2) **20**

[**Bibliografía**](#_ievm6wfyvxss) **21**

## 

## 

## 

# 

# INTRODUCCIÓN

Sonidos y notas musicales o melodías, aunque toda nota musical es un sonido no todos los sonidos son notas musicales, para entender esta diferenciación y saber que hace que un conjunto de vibraciones propagadas en algún medio terminen siendo una nota musical o cualquier otro sonido anti melódico, o en los instrumentos musicales mismos ¿qué es la **afinacion** y por qué uno mal afinado suena tan “mal”?.

La respuesta a todo esto se encuentra en las propiedades de estas oscilaciones específicamente en su frecuencia, la variación y o continuidad de la onda resultante por el sonido, el resultado que produce y percibimos está implícito en su frecuencia.

Una frecuencia puede tener teóricamente cualquier valor numérico positivo medido en Hertz (Hz).

Un hertz es una unidad de medida del sistema internacional en el cual 1Hz representa un ciclo completo por cada segundo, al observar en un gráfica del tiempo como variable independiente se dibujara una onda, se considera un ciclo completo al volver a su posición original.

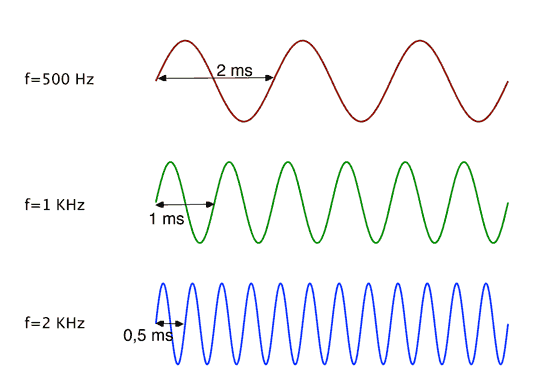


Figura 1: Onda de 500 Hz

Para representar matemáticamente un onda de sonido pure se utiliza una función sinusoidal en la que A: Amplitud y f: Frecuencia pueden ser modificadas:

Figura 2: Fórmula Sinusoidal

Al aumentar la frecuencia la agudeza percibida del sonido aumenta.

## Notas musicales

Las notas musicales son al final estas ondas propagándose por el aire aunque no cualquier onda se encuentran acotadas dentro de un grupo finito y discreto entre el teórico infinidad posible de ondas sonoras.

Son un grupo específico de frecuencias que agradan al oído y tienen melodía o más bien “armonia”.

### Armónicos:

Una onda estacionaria con sus nodos fijos en los extremos, estas vibraciones pueden ir duplicando su frecuencia en cada paso para crear armónicos nombrados por niveles n1, n2, n3 una onda “fija” en sus extremos hace que solo sea posible este tipo de vibraciones que se conocen como los armónicos.

La longitud puede ser n en distancia la regla matemática es que si el primer armónico vibra a una frecuencia determinada el segundo armónico para esa distancia L vibrará al doble, tercero al triple y así sucesivamente.

## 

Figura: Primero, Segundo Y Tercer Armónico

Las notas musicales no son solo onda simples e individuales es una combinaciones específicas de armónicos que cada una a cierto volumen o amplitud de onda general el sonido combinado que puede obtener una nota musical.

Cada instrumento musical genera para cada nota la misma distribución de armónicos simples combinados pero con diferente volumen dándole a cada uno su sonido característico de que una guitarra suene como una guitarra y una flauta como una flauta.

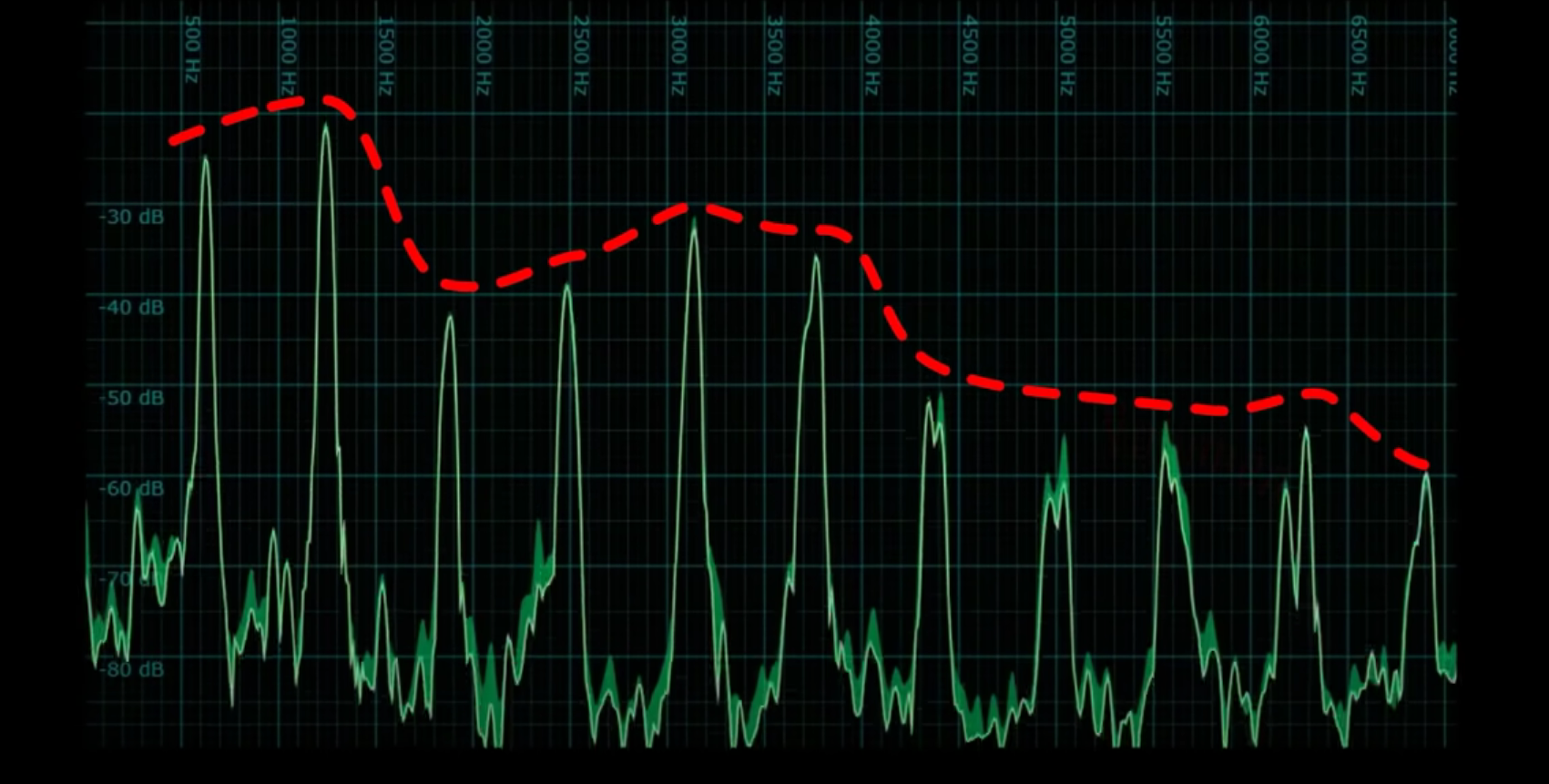


Figura: Sonido de instrumento musical con armónicos vibrando.

### De dónde salen las notas musicales

Al tomar una “cuerda” de longitud L, al reducir su longitud a la mitad seguirá manteniendo los mismos armónicos característicos lo que en esencia hace que sea el mismo “sonido” con una distinta frecuencia o lo que es lo mismo, dividir la longitud de la cuerda por dos es como multiplicar su frecuencia en 2.

Las notas creadas con este método se llaman “**Octavas**”, se pueden crear las octavas que se quiera al dividir o multiplicar su frecuencia por 2.

Usando la relación ⅔ se obtienen nuevas notas musicales o conocidos como “Semitonos” son las denominadas “**Quintas**” obteniendola multiplicando o dividiendo su frecuencia por la relación ⅔.

Las octavas producen la misma nota en distinta frecuencia mientras que su quinta produce nuevas notas diferenciales con el oído manteniendo un ritmo melódico esto porque mantienen muchos armónicos en común pero agrega nuevos mientras que las octavas siempre son los mismos.

Usando sus quintas y octavas se obtiene un patrón que se repite a cada octava dando las distintas escalas musicales.

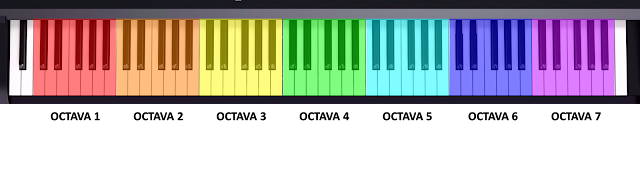


Figura: Octavas de un piano

Siendo estas las notas musicales, teniendo una nota definida para encontrar la misma nota en distinta octava basta con multiplicar o dividir su frecuencia por 2.

Las octavas comparten los armónicos con su frecuencia con la relación 2f , dos cuerdas que no generan armónicos en común tendrán sonidos disonantes es decir no son la misma nota.

Sabiendo que una nota es una frecuencia y su relación con otra con al constante 2 en frecuencia, que frecuencias son las específicas que nombran a una nota las definió en 1955 la Organización Internacional de Estandarización para tener una nota de referencia y todos los instrumentos utilizarla para acordas que nota deberia tener cierta frecuencia y afinarse acorde, la nota de referencia definida es LA con 440 Hz es decir la cuerda o aire debe hacer vibrar todos sus armónicos 440 veces por segundo a partir de esta frecuencia base se puede calcular la frecuencia de el resto de notas con relaciones matemáticas usando sus **quintas** y **octavas**.

### Nombrando las notas

Hay 2 notaciones reconocidas para nombrar en esencia las misma notas la Organización Internacional de Estandarización definió los 440 Hz para la nota “La” que sería el sistema latino pero actualmente se usa otro equivalente anglosajón comprendidos por las siguientes notas y frecuencias.

| **Latino** | **Anglosajón** | **Hz** |
| --- | --- | --- |
| Do | C | 261.63 |
| Reb | C# | 277.18 |
| Re | D | 293.66 |
| Mib | D# | 311.13 |
| Mi | E | 329.63 |
| Fa | F | 349.23 |
| Solb | F# | 369.99 |
| Sol | G | 392.00 |
| Lab | G# | 415.30 |
| La | A | 440.00 |
| Sib | A# | 466.16 |
| Si | B | 493.88 |

Figura: Nombre de notas musicales y sus frecuencias.

### Calculando nuevas notas

Teniendo como referencia La a 440 Hz ya que con las octavas y sus semitonos tenemos una escala de 12 notas con una relación constante que las separa con la fórmula:

Figura: Formula para calcular notas musicales

Siendo “n” el semitono ya sea superior (positivo) o inferior(negativo) se puede calcular cualquier de los otros 11 semitonos en cualquier escala.

= 880

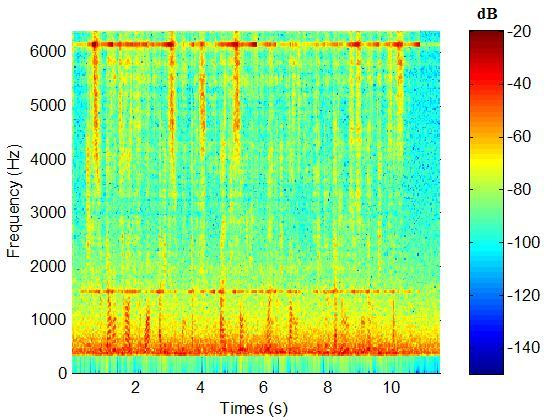
Figura: Calculando “La” en la siguiente octava.

= 493,88

Figura: Calculando “Si” en la misma octava.

## Cómo leer los sonidos

Para poder leer o visualizar un sonido se debe convertir una señal análoga las ondas a un medio digital reproducible y cuantificable es lo que hace un **espectrograma** una representación visual del espectro de frecuencias que varía con el tiempo son representaciones gráficas de 3 dimensiones (Tiempo, frecuencia y volumen)

****

**Figura: Espectrograma de ejemplo**

### Volumen de los sonidos

Para cuantificar la intensidad o volumen de un sonido se emplea una unidad denominada decibelio (dB) es una escala logarítmica un aumento del doble en la presión de sonido en el medio corresponde a un aumento de solo 6 dB.

## 

## Inteligencia Artificial

Qué es inteligencia puede ser abordado desde distintos puntos, la definición de inteligencia no se tiene completamente definido pero implica la capacidad de razonamiento utilizando información para resolver problemas o alcanzar un objetivo de manera autónoma.

En término de inteligencia artificial es la teoría de desarrollo de un sistema computacional que sea capaz de realizar tareas que se considere que requiere inteligencia o intervención humana. Esto puede ser cumplido en diversos campos, desde toma de decisiones, percepción visual, reconocimiento del lenguaje, predicciones o reconocimiento de patrones entre otros.

Para el caso de uso de una clasificación de elementos ya existentes para un posterior reconocimiento de estos usando nuevos inputs que antes no se habían visto se considera aprendizaje el **Machine Learning** es un subconjunto de la IA (Inteligencia Artificial), que “aprende” de datos, identificando patrones para predecir o tomas decisiones con mínima intervención externa.

### 

### 

### Red Neuronal

También conocidas como redes neuronales artificiales son una serie de algoritmos del machine learning y son parte del núcleo de los algoritmos del deep learning trata de imitar las neuronas biológicas se señalan entre sí.

Estas neuronas se separan por capas donde cada nodo es una neurona, teniendo en general la capa de entrada las intermedias que pueden ser múltiples capas conocidas como capas ocultas y la capa final de salida.

Las redes neuronales se entrenan para aprender y mejorar su precisión con las iteraciones las que permiten una vez finalizado ese entrenamiento clasificar y agrupar datos.

Cada nodo funciona como su propio modelo de regresión lineal formado por datos de entrada una ponderación que se auto corrige con el tiempo y determinan la importancia de cada variable así como un sesgo y una salida.

### Deep Learning

Subconjunto del machine learning son métodos basados en redes neuronales con aprendizaje supervisado o no supervisado usando muchas capas de extracto progrsivo de información de las capas anteriores a partir del input inicial.

En cada capa aprende por separado a identificar características básicas a partir de una representación abstracta del input inicial que se va descomponiendo más y más en cada capa donde conceptos de alto nivel son redefinidos en término de bajo nivel en una jerarquía denominada **deep architecture.**

El sonido puede ser representado como los espectrogramas este conjunto de datos en 2 dimensiones para aprender los patrones específicos esto puede ser tratado como una imagen. El “problema” con las imágenes es que al ser información necesariamente representada en al menos 2 dimensiones para que tengan un sentido una red neuronal convencional solo podría recibir vectores de una dimensión haciendo difícil esta clasificación para los espectrogramas.

Por lo que para poder procesar alimentos de más de una dimensión pueden ser utilizadas redes neuronales convolucionales especializadas en encontrar y clasificar patrones usando enormes datasets en su fase de aprendizaje, son especializadas en poder comprender la estructura espacial siendo el mejor caso de uso para las imágenes. Identifican elementos básicos y generales para que posteriormente las siguientes capas pueden discernir los elementos más complejos iniciales.

La red neuronal aprende a definir los valores del filtro que usa para extraer estas características simples, este filtro también es conocido como “kernel” no es más que una matriz en base a la cual ejecutara ciertas operaciones secuencialmente sobre cada grupo de pixeles de las imágenes devolviendo una versión reducida y simplificada de la misma, cada capa configura su propio kernel hasta llegar a la última capa que toma estos mapas de características. Que al final son detecciones de las detecciones de las capas de la red una vez obtenidos las capas de características estos pueden ser introducidos finalmente en una red neuronal multicapa para analizar lo que se encontraba en la imagen inicial.

El aprendizaje de la red convolucional es aprender a encontrar los valores adecuados para los filtros mientras que la red neuronal a entender que significa el resultado de estos.

Para poder clasificar se usan métodos de clasificación que consisten en decir a que pertenece el resultado dentro de un número finito de posibles respuestas, un número limitado de clases o categorías que tienen sentido en el contexto en que se clasifica como “si” o “no”, “líquido”, “sólido” , “gaseoso” o las notas musicales y estas pueden ser recibidas como probabilidades de acierto las que acompañan la respuesta como un “89% de probabilidad de que sea un correo spam”, en cuanto a la regresión es dar la respuesta dentro de un valor numero en un conjunto finito de resultados.

# Reconocimiento de Notas musicales

Como se mencionó anteriormente una nota musical siempre por convenio consta de los mismos grupos de n armónicos combinados independientemente de su fuente aunque distintos instrumentos suenen de manera ligeramente distinta la esencia del sonido de la nota siempre es reconocible es porque la única diferencia es el volumen que otorga a cada armónico dándole esto su sonido característico a cada uno.

Por lo que para reconocer una nota basta con encontrar sus armónicos y compararlos con los de la nota en cuestión o con todos los armónicos en resonancia hacer los cálculos necesarios para encontrar el resto de notas a partir de una de referencia usando sus octavas y cuartas.

Pero cómo puede ser reconocido un sonido, dado que las notas musicales a pesar de repetirse continuamente en el espectro a distintas frecuencias estas son finitas y concretas y reciben un “label” característico que las nombra visto anteriormente.

Las notas se pueden clasificar en esos y todos los sonidos con frecuencias fuera de estos “labes” no son notas musicales.

Usando una aproximación usando Reconocimiento de los patrones de los espectrogramas generados llevaría a la capacidad de poder etiquetar las notas a partir de estos.

## 

## Planteamiento

Con un dataset de audios de notas musicales individuales con las clases definidas (“do”, “re”, “mi” etc.) utilizando el espectrograma para el reconocimiento de patrones en base a sus frecuencias.

En lugar de usar el espectrograma directamente como una imagen los valores de la variable amplitud se leerán como números indicando su valor y la variable tiempo representada como posición en un array.

### Paso 1 populado de datos:

Leer el fichero database.sqlite que funciona como una mapper al dataset completo por lo que la información principal y metadatos como tipo de nota, instrumento etc son almacenadas en este archivo haciendo referencia a cada fichero .wav del dataset.

*#read and format dataset*

*import* sqlite3

con = sqlite3.connect('database.sqlite')

data = con.execute('''

select

sa.id,

sa.instrument,

sa.note as class,

sa.filename

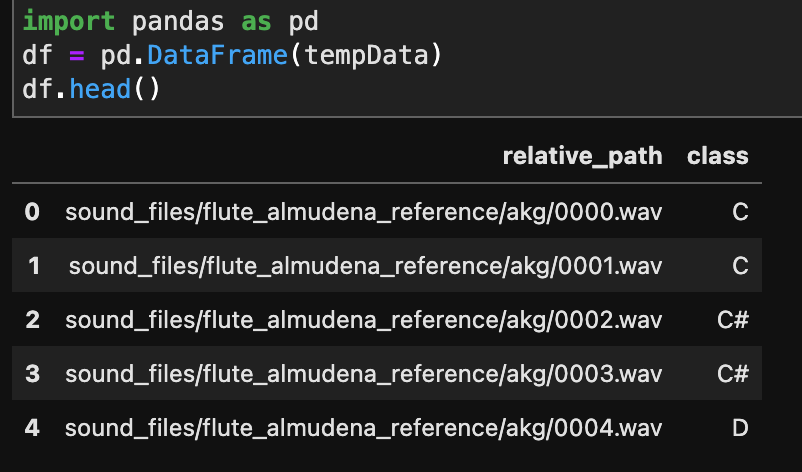
from "sounds+akg" as sa where not(sa.klass like 'scale%')

''').fetchall()

### Paso 2 extracción de datos:

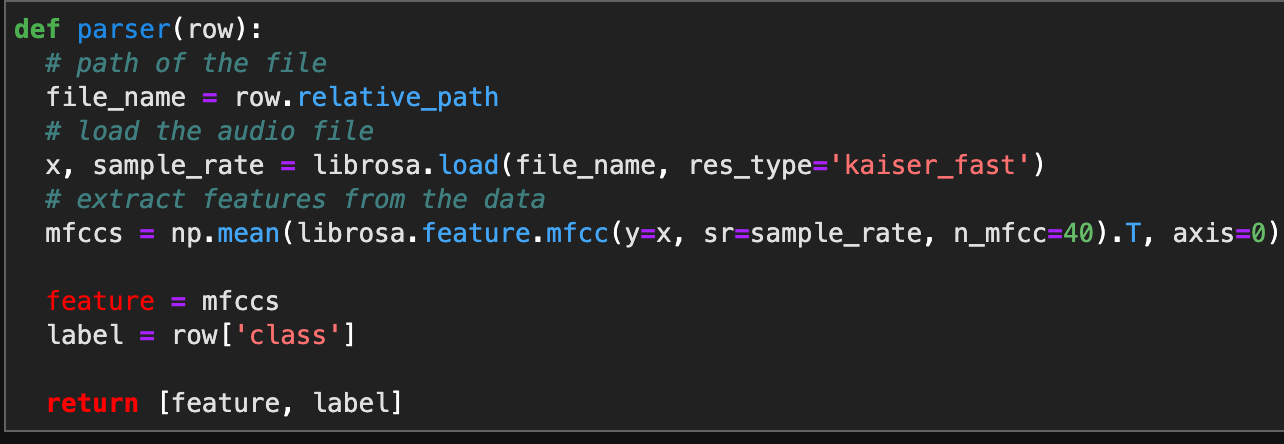
Una vez leídos los archivos del dataset, con ayuda del mapper anterior se procede a crear un diccionario para poder posteriormente ser convertido a un data frame de la librería *pandas* en donde tenemos almacenados los valores del directorio donde se encuentra cada archivo de audio y su nombre de clase.

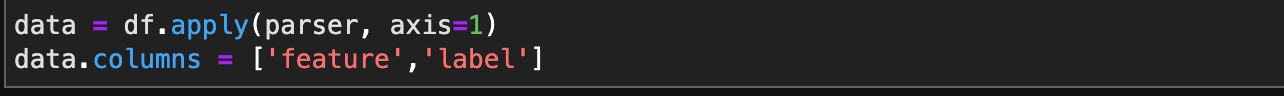
### 



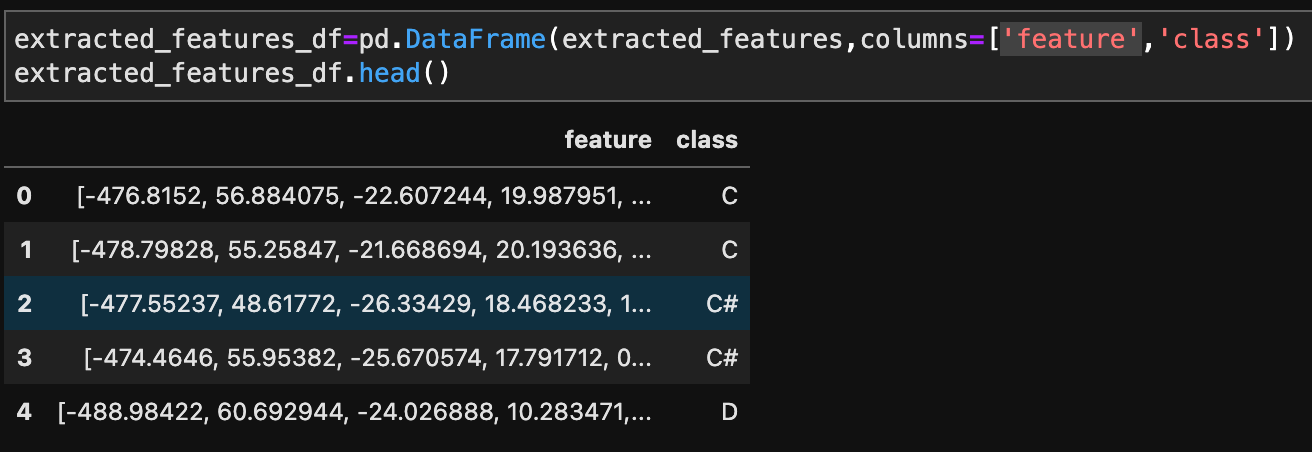
### Paso 3 tratado de datos:

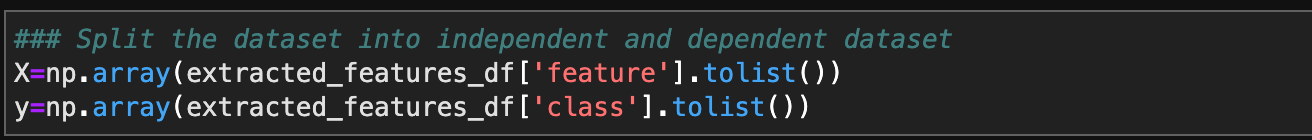
Leer y extraer las características del audio, iterar cada data frame para extraer y se agregan secuencialmente a un array de 2 dimensiones.





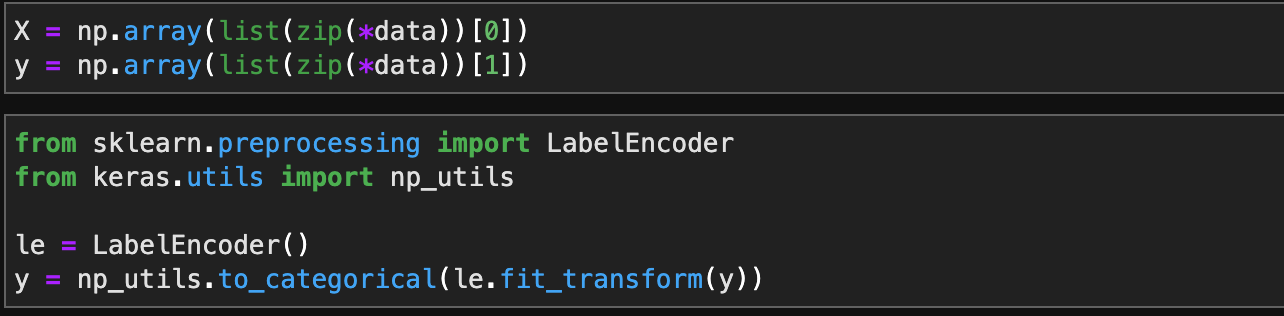
Transformar el array a un dataframe separado en columnas en base a las características de cada audio y su respectiva clase.





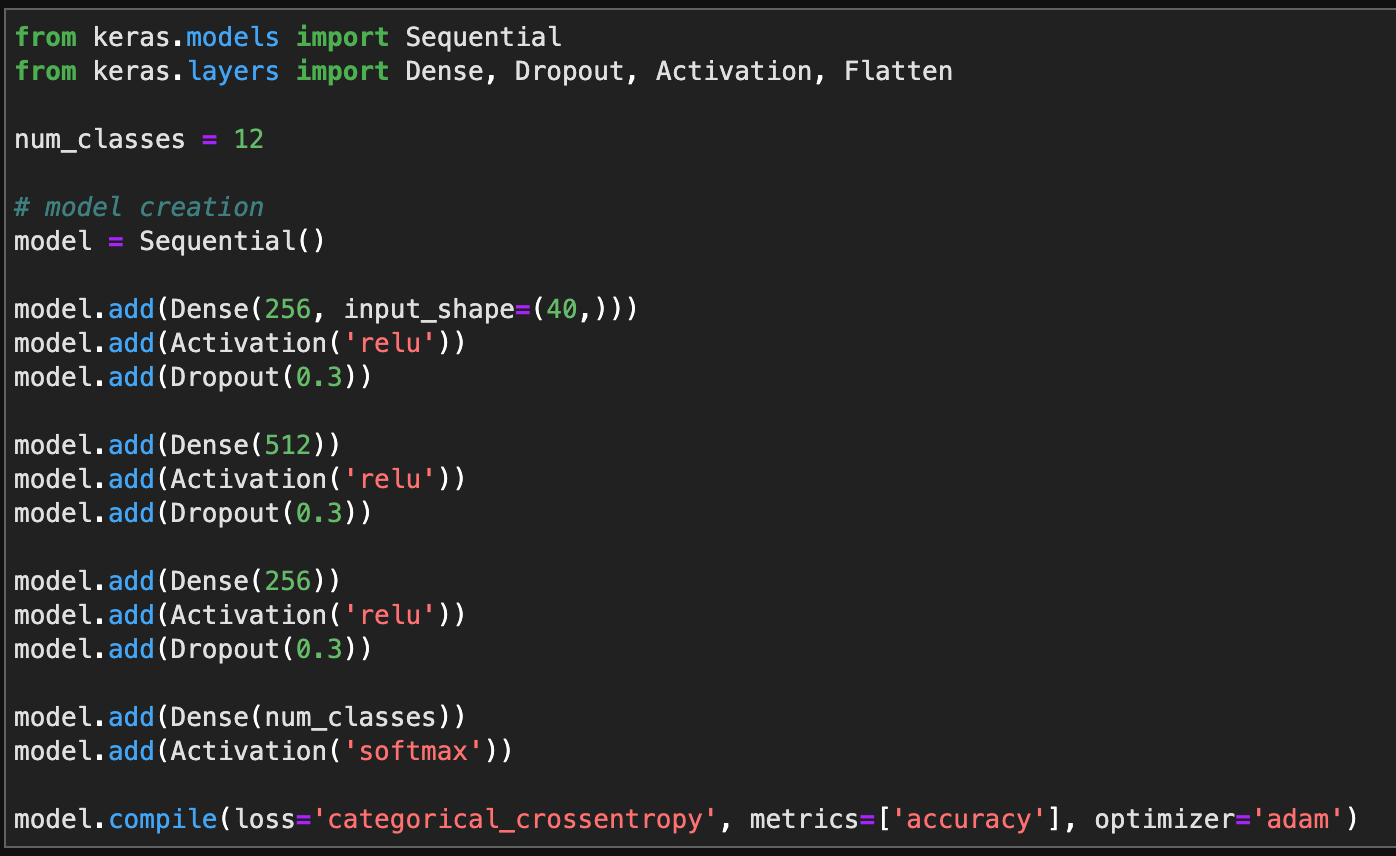
### Paso 4 segregar data:

Separar del dataset una muestra representativa de los datos para el entrenamiento del modelo.



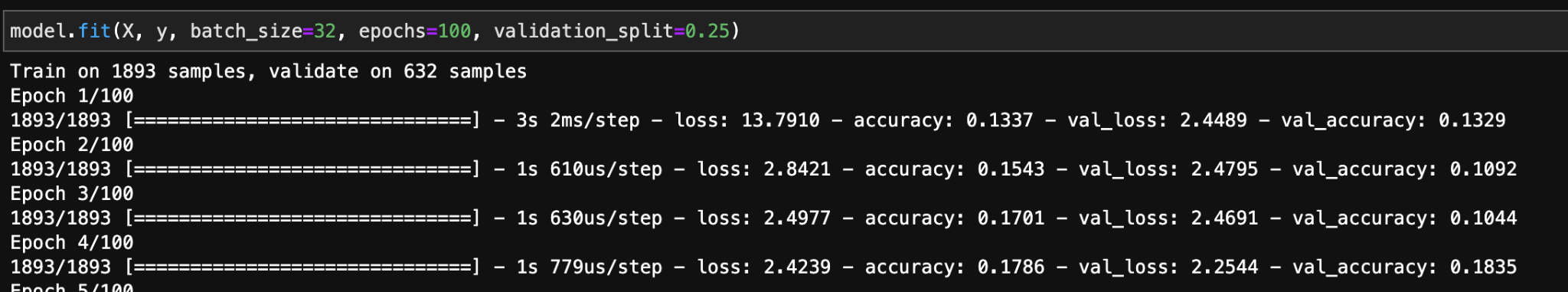
## Entrenamiento

Comenzamos configurando nuestro modelo de datos:



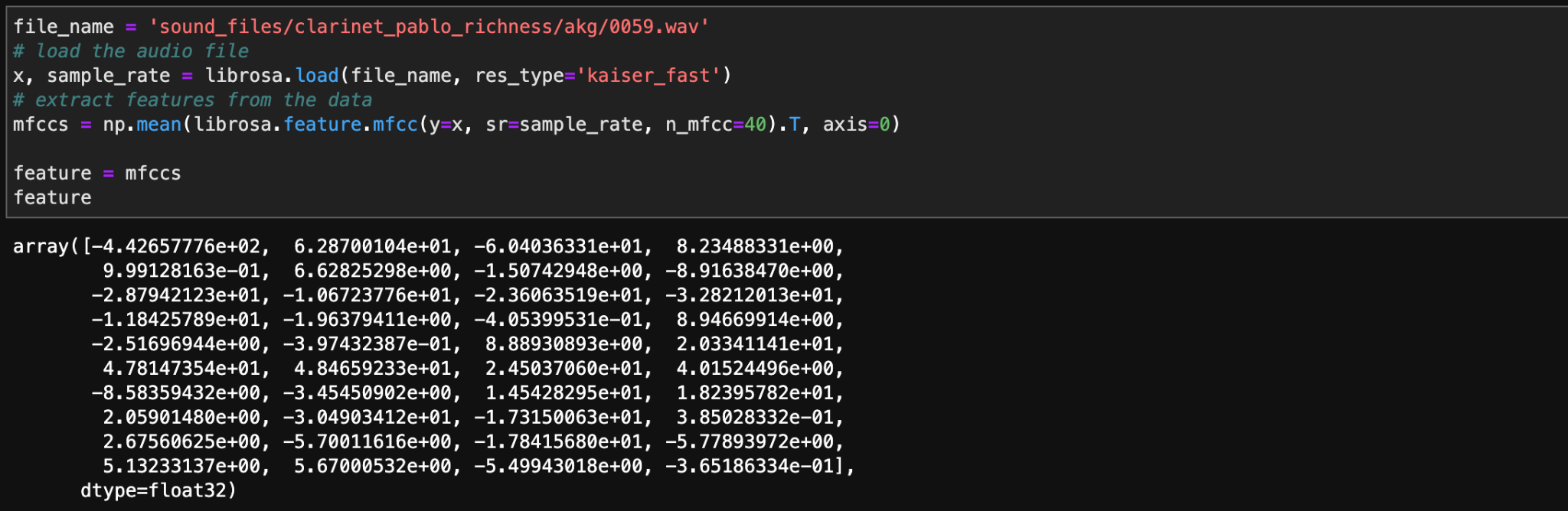
En este modelo de 4 capas, definimos la cantidad de clases que usaremos para las últimas capas.

Luego entrenamos nuestro modelo definiendo la cantidad de interacciones que realiza con la data de muestra, este nos mostrará el desarrollo y mejora dentro de la precisión del modelo en cada iteración del entrenamiento.

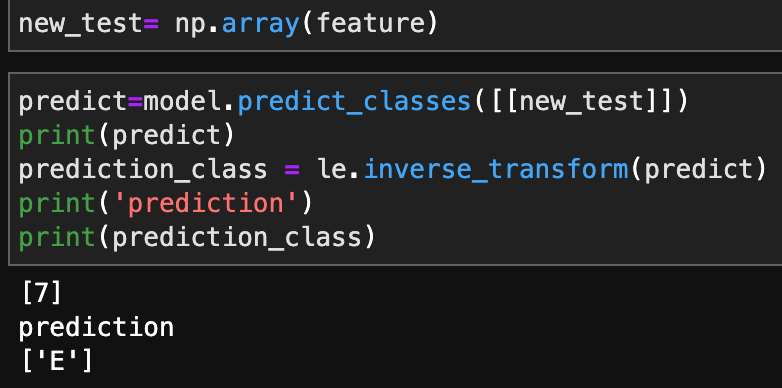


## Pruebas

Para probar nuestro modelo necesitamos aplicar a un audio valido, nuestro extractor de características



Luego damo



## Limitaciones

El sistema es capaz de reconocimiento individual de notas, y sus diferentes armónicos dentro de la escala de frecuencias, los sonidos deben de estar grabados con anterioridad, estos sonidos son presentados sin variantes de ruido en el entorno y con una longitud es decir duración del sonido de al menos 3 segundos, cada uno de los sonidos deben estar en formato .wav, se deben evitar instrumento con sonidos enarmonicos que suben de frecuencia para emular una nota diferente, ya que existen estándares que ciertos instrumentos no respetan, sea el caso del saxofón y el mi sostenido, siendo que el estándar dice que mi sostenido es igual a un fa, pero el saxofón puede mantener una frecuencia intermedia entre estas notas y por lo tanto el sistema no será capaz de reconocer y dará una nota aproximada.

Es importante tener un buen micrófono para la grabación del sonido, evitar también los casos extraordinarios de afinación dentro de percusiones o sonidos que superen las 8 octavas de sonido, es decir frecuencias desde 261.63hz hasta 493.88hz.

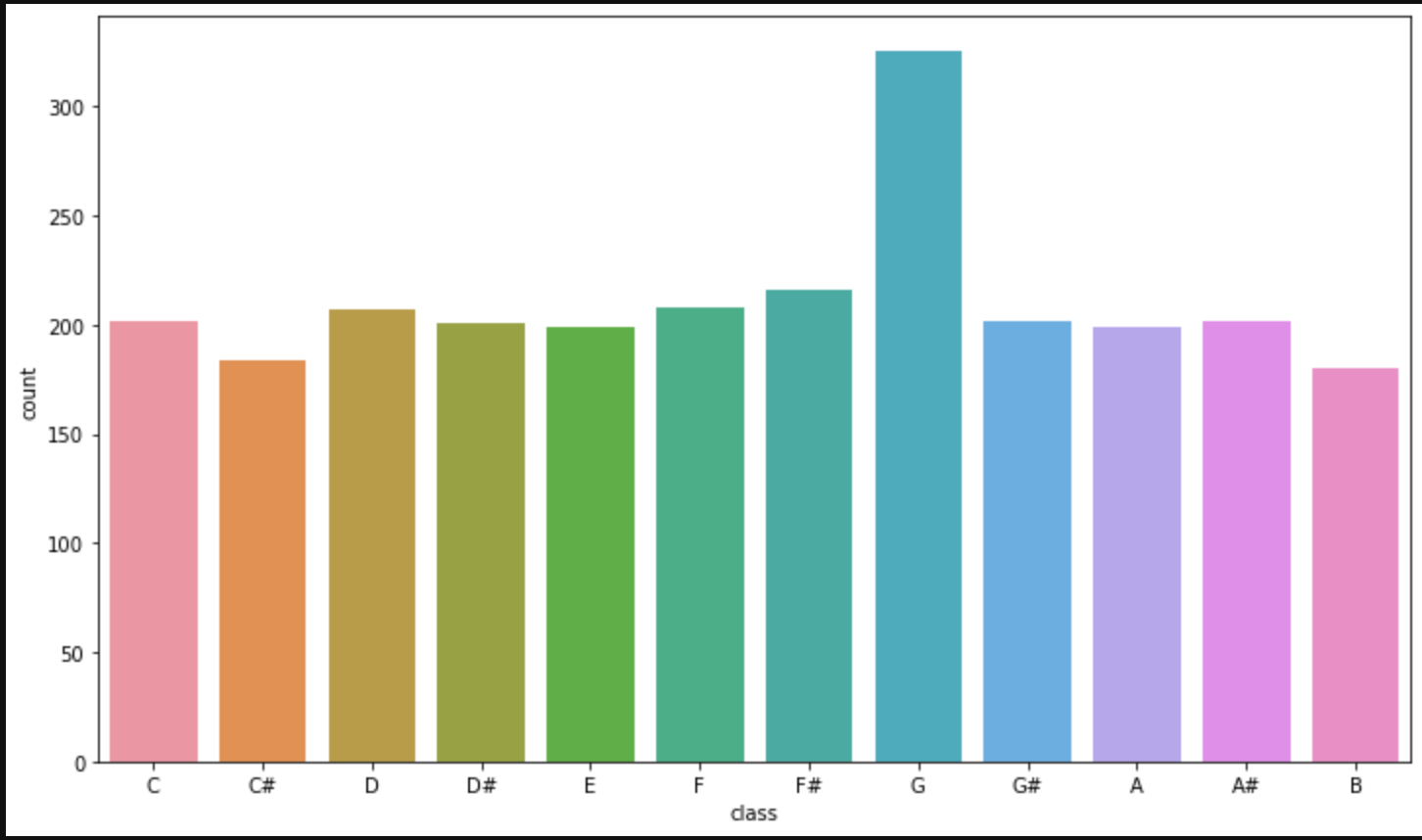
La inteligencia no reconocerá sonidos grabados en tiempo real, solo sonidos pregrabados, verificar que el archivo no esté corrupto y sea perfectamente reconocible al oído humano.

# Conclusiones

Para los modelos de clasificación en deep learning es importante una muestra variada y lo suficientemente grande para una correcta inferencia de los datos.

Una problemática que se ha encontrado en el desarrollo del proyecto es la correcta selección de los datos de entrenamiento, al tener una muestra sesgada de datos, una incorrecta selección de datos, puede provocar una predicción errónea del modelo de datos, en este caso en particular, una clase posee una cantidad de datos mayor a la mayoría de clases, y una incorrecta selección de datos hace que cada predicción de la red neuronal sea la clase con mayor cantidad de datos.

La correcta selección de datos y de clases del modelo ha hecho un incremento en la predicción de datos desde un 13% a un 95% de precisión.



# 

# 

# Bibliografía

Learning, D. (2020). Deep learning. High-Dimensional Fuzzy Clustering.

Yan, L. C., Yoshua, B., & Geoffrey, H. (2015). Deep learning. nature, 521(7553), 436-444.

Jaramillo, A. M. J. (2007). Acústica: la ciencia del sonido. ITM.

Cerdá, J. F. M. (2016). La música y las matemáticas: generando las notas musicales. Aularia: Revista Digital de Comunicación, 5(1), 29-32.

Bernard, M., Poli, M., Karadayi, J., & Dupoux, E. (2021). Shennong: a Python toolbox for audio speech features extraction. arXiv preprint arXiv:2112.05555.