



IA 2 - UIS

Nov 16, 2023  
4:00 p.m.

# FALCON EYE - Voice Recognition

“Elevating Security through AI-Powered Voice Recognition Excellence”

# Integrantes



**DIEGO LOPEZ**

2200162



**GABRIEL REYES**

2200141

# Integrantes



**DIEGO LOPEZ**

2200162



**GABRIEL REYES**

2200141

# OBJETIVOS

FALCON EYE propone una solución avanzada de seguridad en la era digital

- Recopilar un dataset exhaustivo de archivos de audio proporcionado por Mozilla. Realizar tareas de reprocesamiento para garantizar la calidad y uniformidad de los datos, como la normalización del audio y la extracción de características relevantes.

- Diseñar y entrenar una CNN utilizando técnicas de aprendizaje profundo y convolución para reconocer patrones complejos en los archivos de audio. Implementar algoritmos de optimización para mejorar la precisión y eficiencia del modelo.

- Desarrollar algoritmos de clasificación que permitan identificar el género (hombre/mujer) y estimar la edad en base a los patrones de audio identificados por la CNN

# Dataset a Utilizar

Common Voice

moz://a

Hay 6002 filas y 9 columnas

Column Names

0 client\_id

1 path

2 sentence

3 up\_votes

4 down\_votes

5 age

6 gender

7 accent

8 set

# Common Voice

moz://a

Hay 6002 filas y 9 columnas

Column Names

0 client\_id

1 path

2 sentence

3 up\_votes

4 down\_votes

5 age

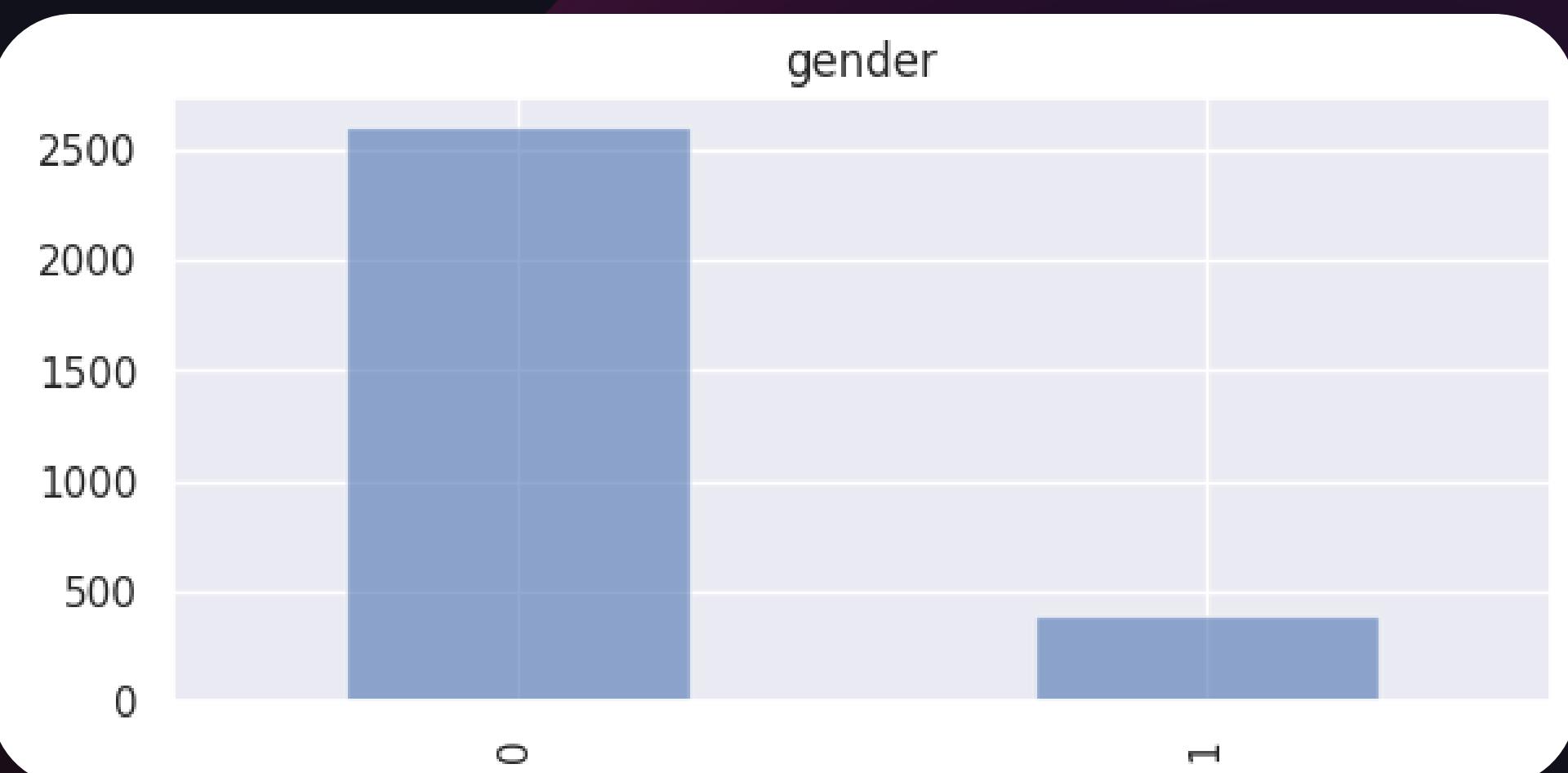
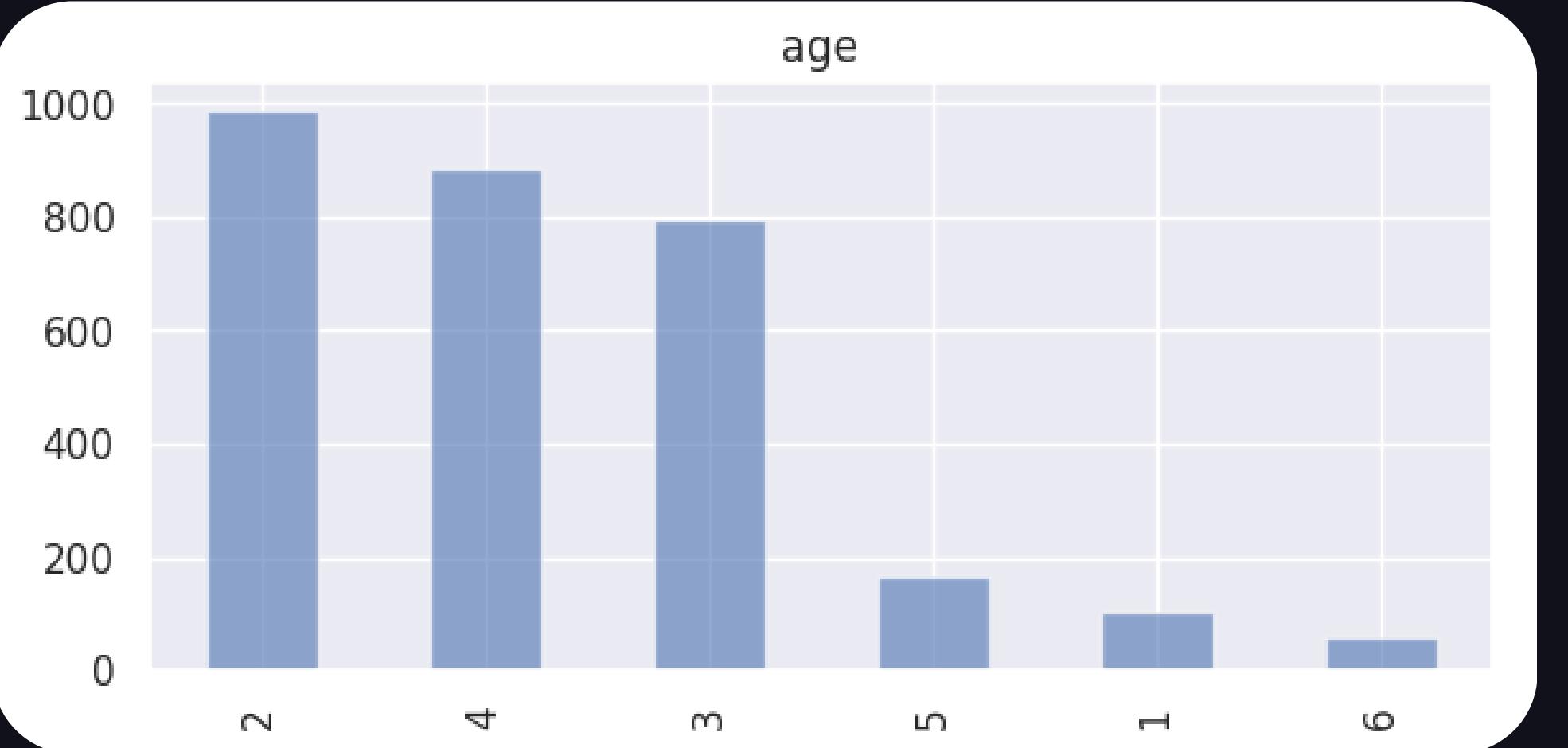
6 gender

7 accent

8 set

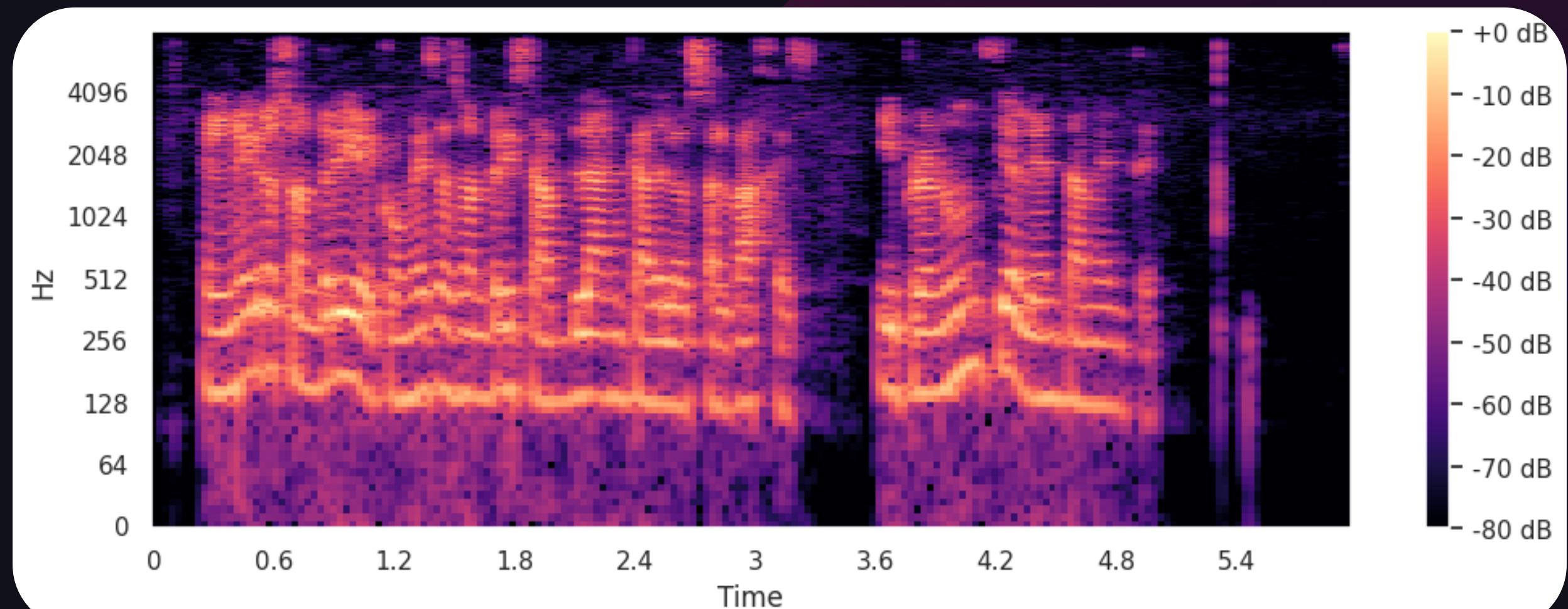
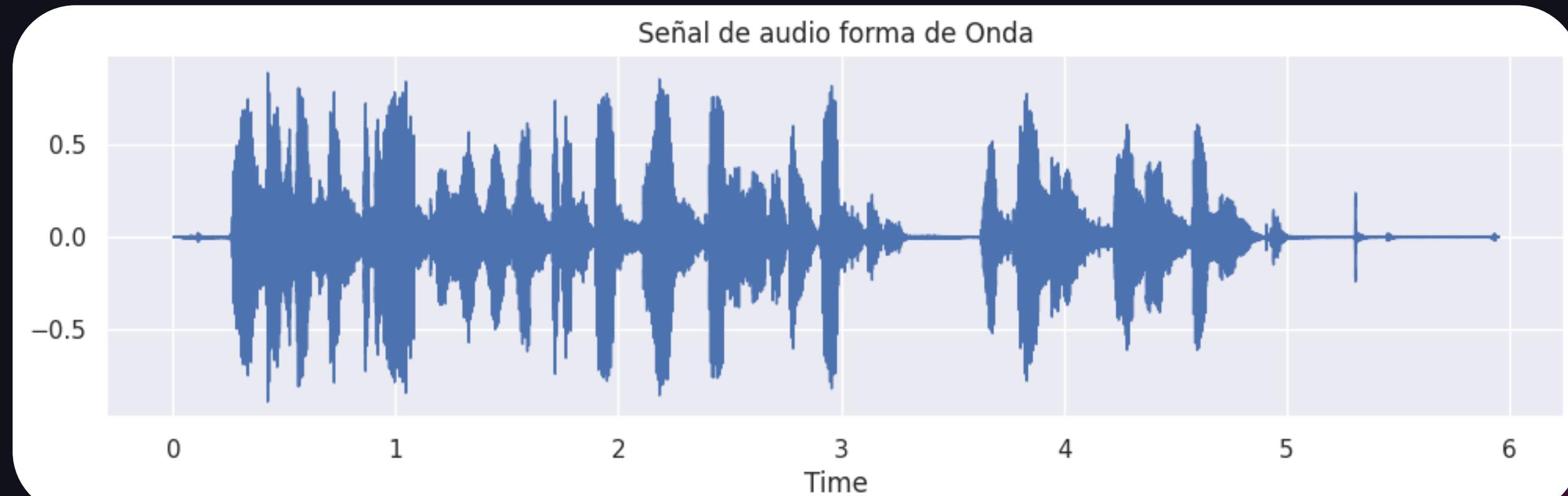
(2994, 4)

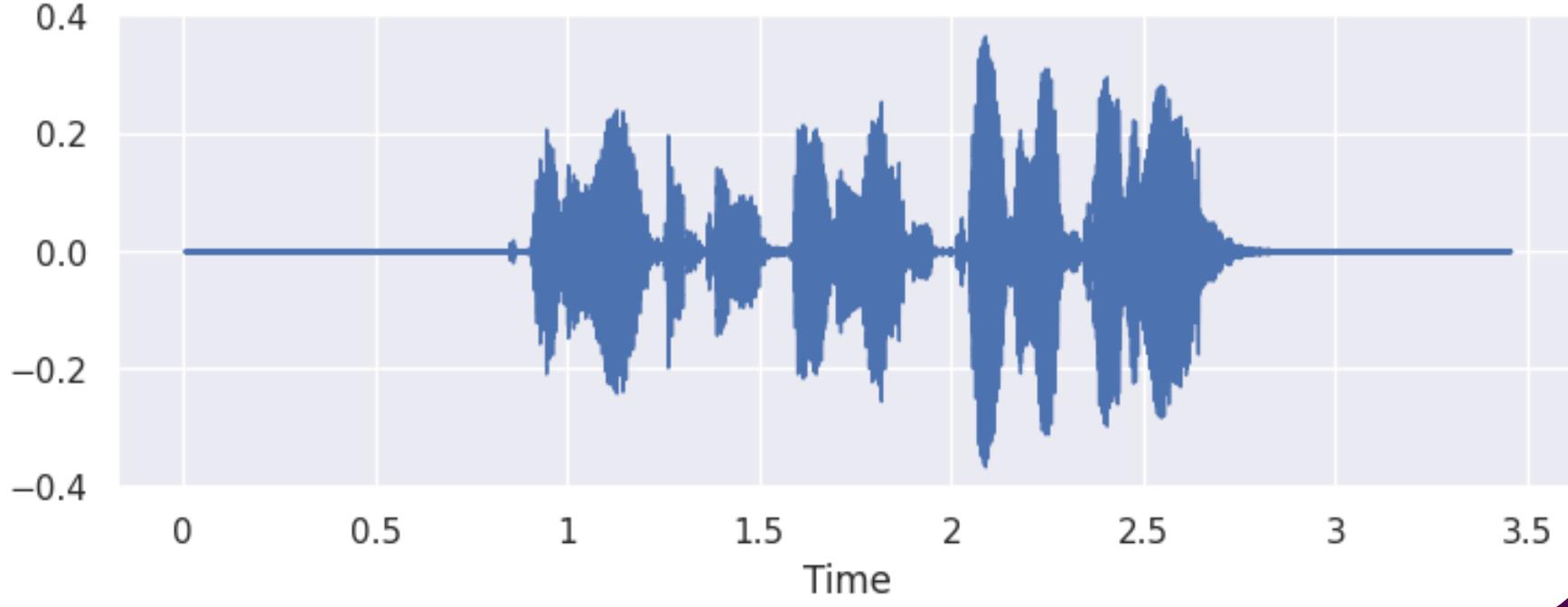
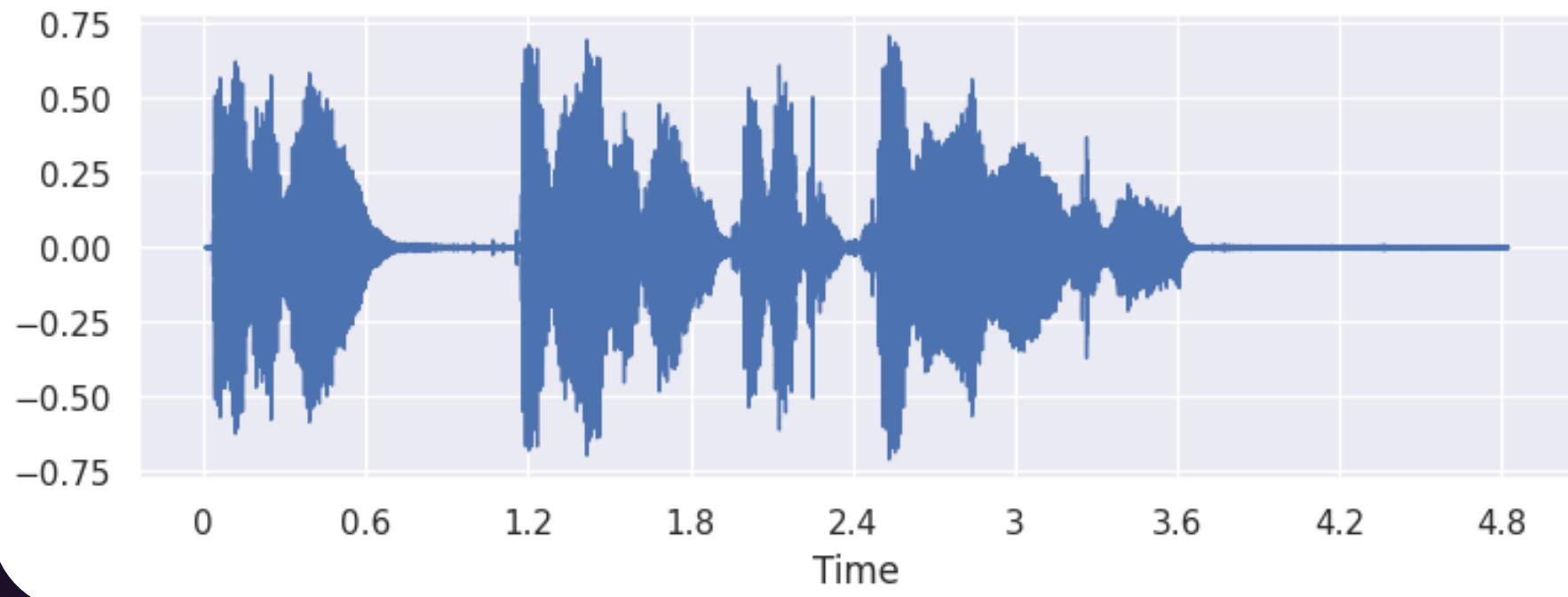
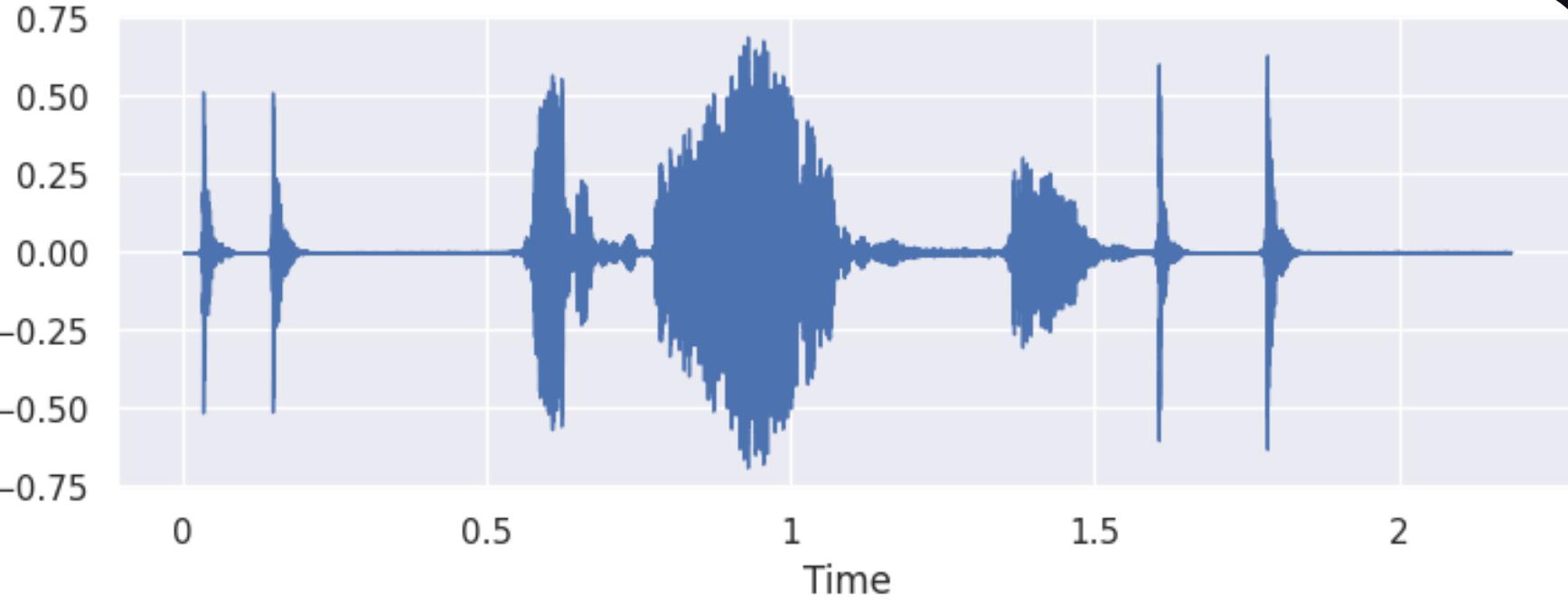
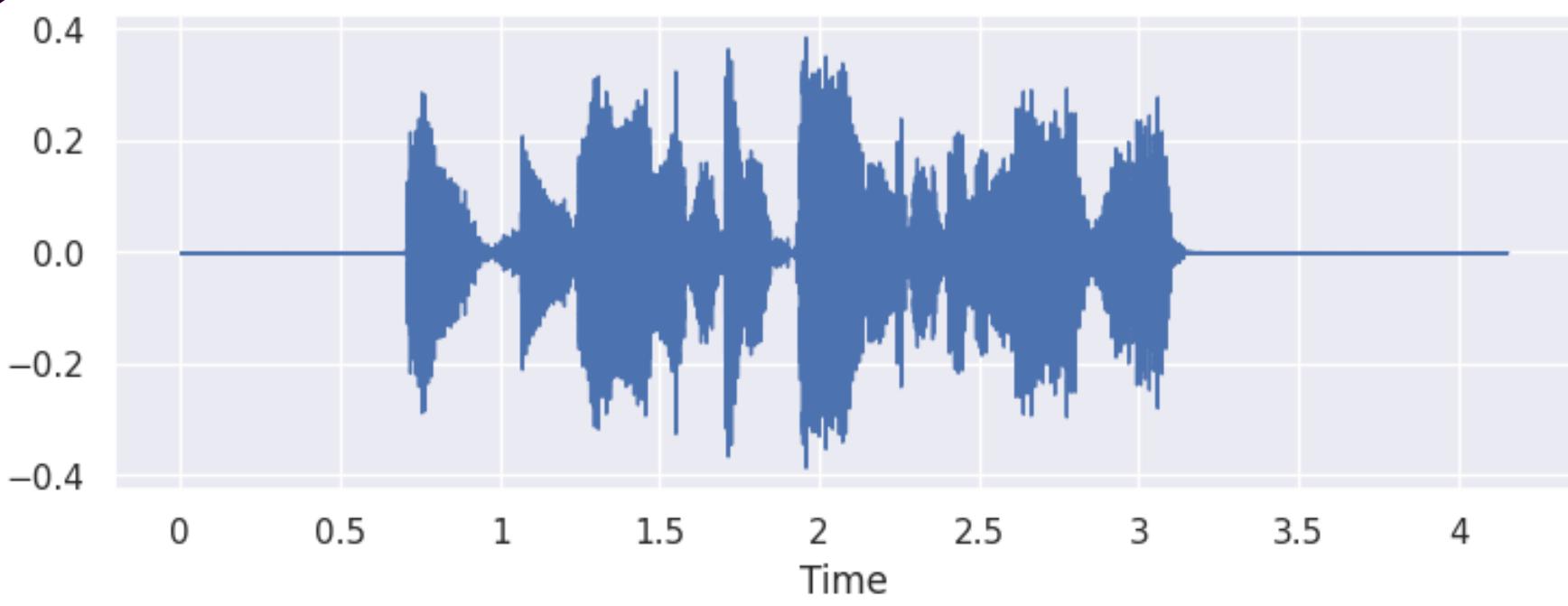
	path	age	gender	set
0	f8dcacb4795e...	2	0	train
1	07a3f0a61f65...	2	0	train
2	9284e0d655fd...	2	0	train
3	79729598b44...	2	0	train
4	ed7a3cfb5fd7f...	2	0	train



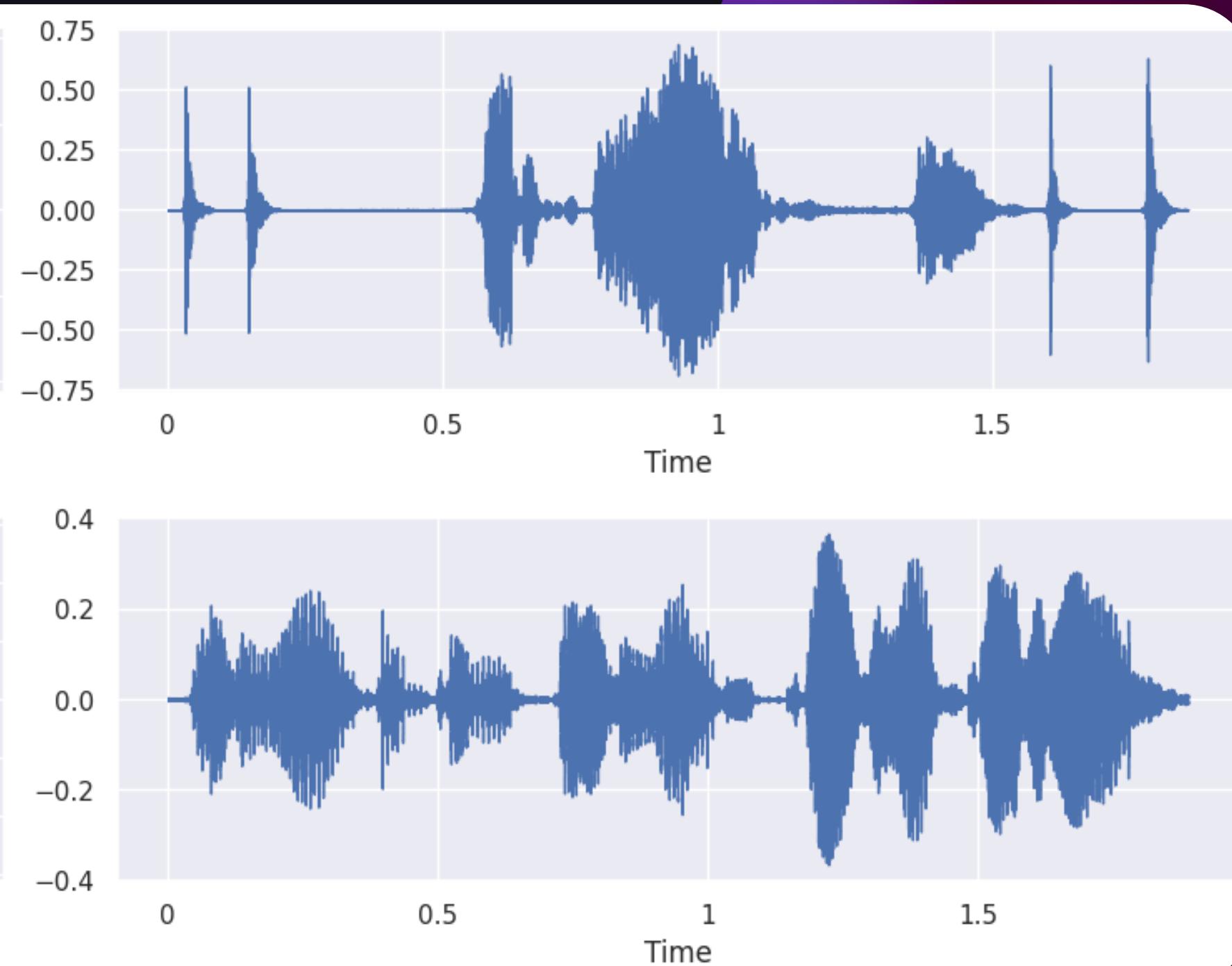
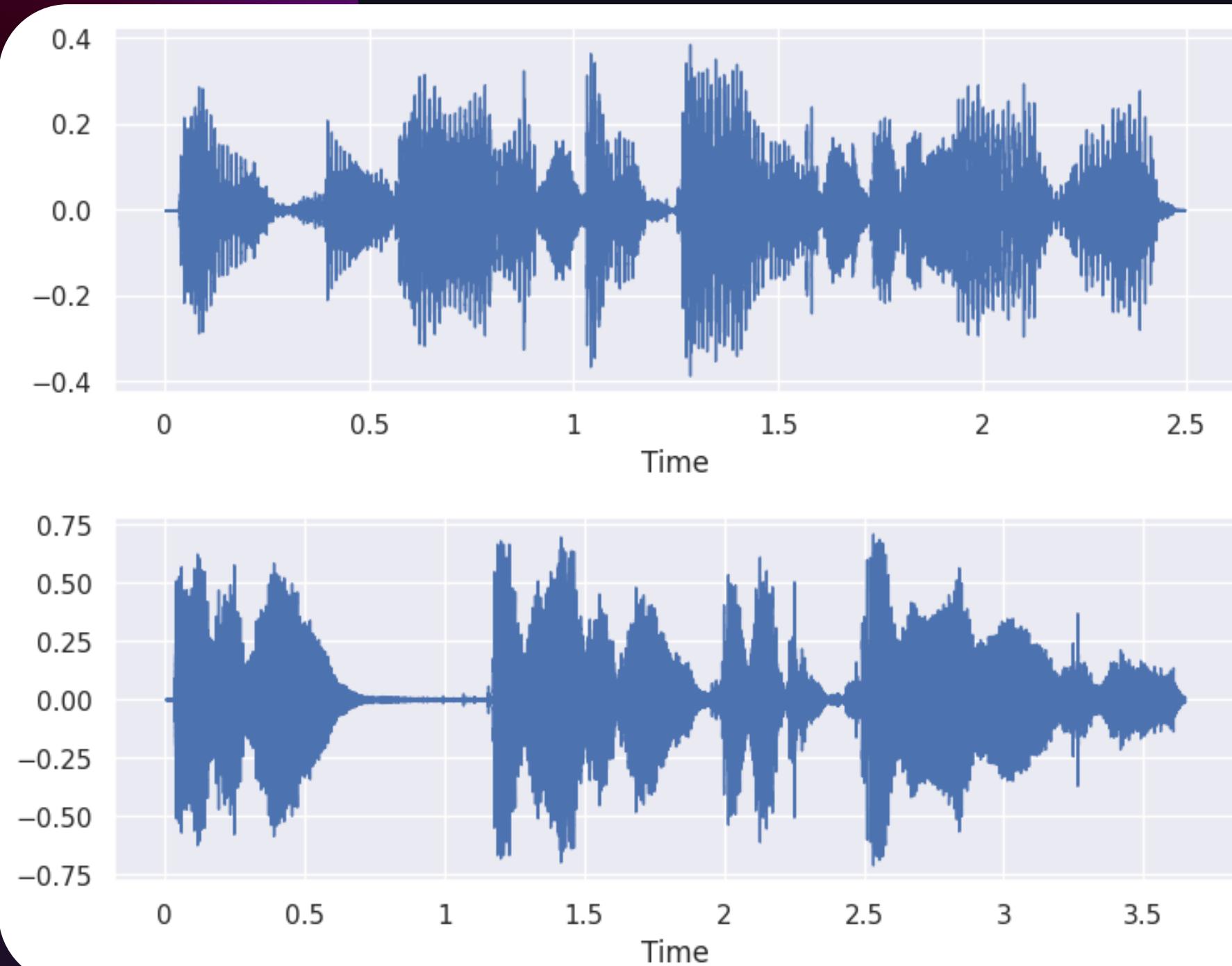
# Common Voice

moz://a





# LIMPIEZA



# LIMPIEZA

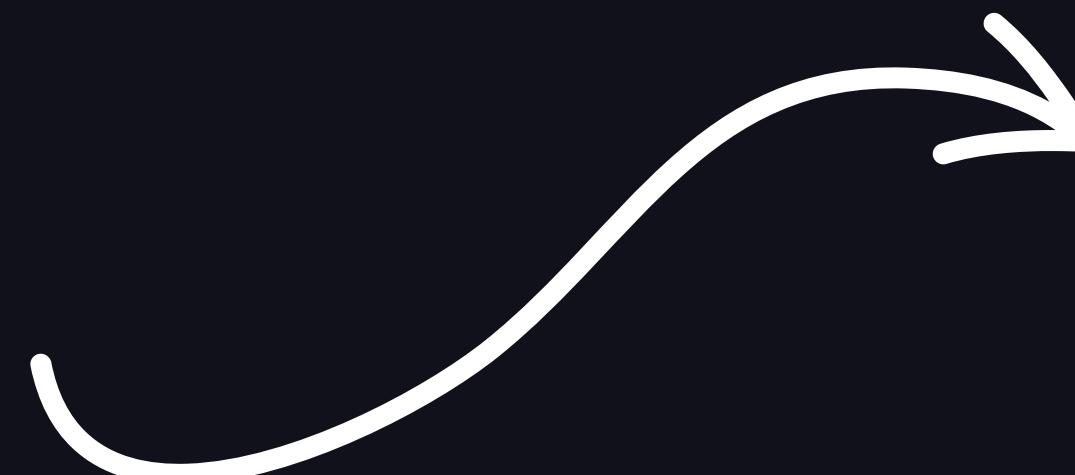
# Extract Features

Column Names
0 client_id
1 path
2 sentence
3 up_votes
4 down_votes
5 age
6 gender
7 accents
8 locale
9 segment

Característica	Descripción
MFCCs (Coeficientes Cepstrales de Mel-Frecuencia)	Representan la envolvente espectral del audio en la escala mel. Los MFCCs son una representación compacta del espectro de frecuencia de un sonido, que se pueden utilizar para caracterizar la voz humana.
Croma	Describe la distribución de energía tonal del audio. El croma es una medida de la presencia de frecuencias tonales en un sonido.
Mel-espectrograma	Representa la energía del audio en el dominio del tiempo-frecuencia. El mel-espectrograma es una representación visual de la energía del audio en función de la frecuencia y el tiempo.

# Extract Features

Column Names
0 client_id
1 path
2 sentence
3 up_votes
4 down_votes
5 age
6 gender
7 accents
8 locale
9 segment



Característica	Descripción
Contraste espectral	Captura la distribución de energía en diferentes bandas de frecuencia. El contraste espectral es una medida de la variación de la energía en diferentes bandas de frecuencia.
Tonnetz	Representa la tonalidad del audio en la escala Tonnetz. El Tonnetz es una escala de tonos que se utiliza para representar la tonalidad de un sonido.

# MODELO

```
from keras.models import Sequential
from keras.layers import Dense, Dropout, BatchNormalization
from keras.optimizers import Adam
from keras.callbacks import EarlyStopping

model = Sequential()
model.add(Dense(193, input_shape=(193,), activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.2))

model.add(Dense(128, activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.3))

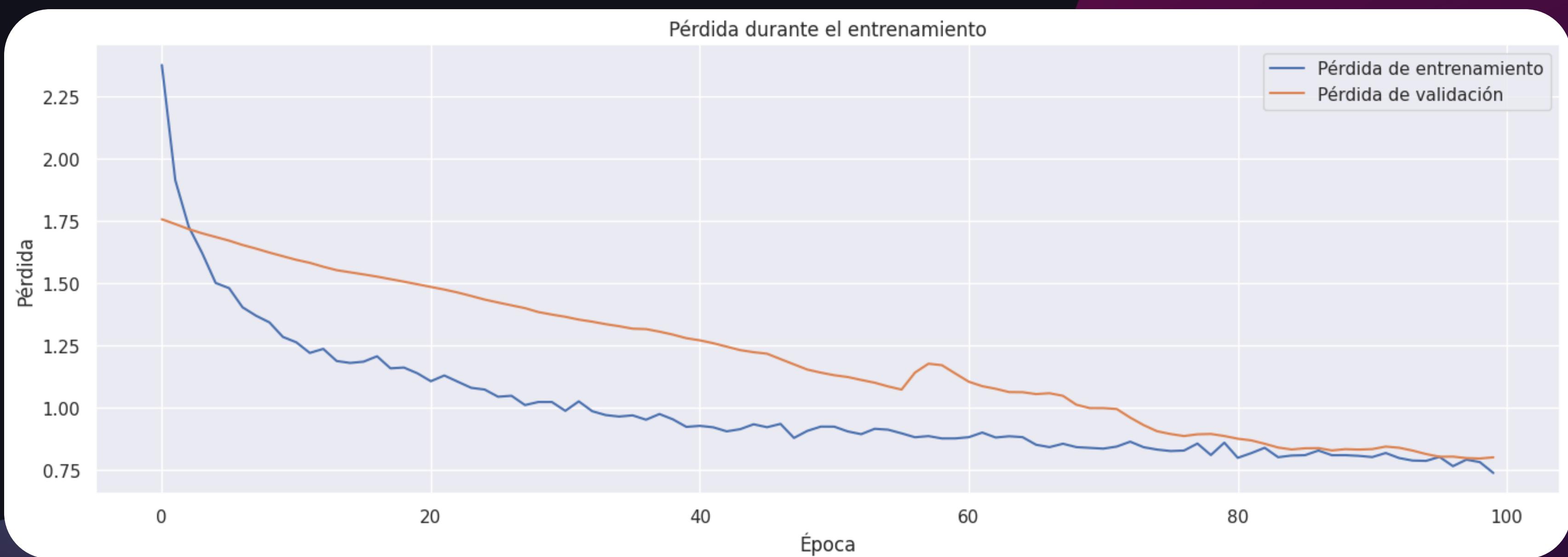
model.add(Dense(64, activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.4))

model.add(Dense(6, activation='softmax'))

# Compiler
opt = Adam(learning_rate=0.001)
model.compile(loss='categorical_crossentropy', metrics=['accuracy'], optimizer=opt)

early_stop = EarlyStopping(monitor='val_loss', min_delta=0, patience=20, verbose=1,
```

# RESULTADOS



# RESULTADOS



# PREDICCIONES

	path	age	gender	set
0	55-man.wav	5	0	nuevo
1	20-man.wav	2	0	nuevo
2	19-woman.wav	2	1	nuevo
3	35-woman.wav	4	1	nuevo
4	65-woman.wav	6	1	nuevo
5	27-woman.wav	3	1	nuevo
6	14-woman.wav	1	1	nuevo
7	15-man.wav	1	0	nuevo

# PREDICCIONES

	path	age	gender	set
0	55-man.wav	5	0	nuevo
1	20-man.wav	2	0	nuevo
2	19-woman.wav	2	1	nuevo
3	35-woman.wav	4	1	nuevo
4	65-woman.wav	6	1	nuevo
5	27-woman.wav	3	1	nuevo
6	14-woman.wav	1	1	nuevo
7	15-man.wav	1	0	nuevo

Porcentaje de predicciones correctas para la edad: 50.00%  
Porcentaje de predicciones correctas para el género: 62.50%

	path	age	gender	set	Edad_pred	Genero_pred
0	55-man.wav	5	0	nuevo	6	0
1	20-man.wav	2	0	nuevo	2	0
2	19-woman.wav	2	1	nuevo	2	0
3	35-woman.wav	4	1	nuevo	4	1
4	65-woman.wav	6	1	nuevo	2	0
5	27-woman.wav	3	1	nuevo	4	1
6	14-woman.wav	1	1	nuevo	2	0



# CONCLUSIONES



# Gracias!