FALCON EYE

BIOMETRIA POR VOZ

PROYECTO FINAL INTELIGENCIA ARTIFICIAL 1

Diego Alejandro López Camacho – 2200162
Gabriel Fernando Reyes Guevara -2200141
Silvia Valerie Guarín Manrique - 2202690

Biometría de voz

- La biometría por voz es una tecnología de reconocimiento y autenticación que se basa en los rasgos naturales, únicos e intransferibles de la voz humana.

- Parámetros físicos de la voz (tamaño y forma de la laringe y de la cavidad nasal y craneal), pero también de su comportamiento (frecuencia, entonación y el acento).

El fraude de identidad es un delito en el que un individuo utiliza los datos personales de otro, sin autorización, para engañar o estafar a otras personas por medio de redes sociales o tros sitemas digitales.

OBJETIVOS

Identificar al interlocutor cuando accese al perfil mediante la recopilación de datos sonoros obtenidos a través de la voz del usuario.

Analizar la captura de los rasgos únicos de cada usuario, su aparato fonador para demostrar que los datos ingresados son validos.

autenticar en cuestión de segundos y con una simple prueba de voz que el usuario que accede a la cuenta corresponde a los datos optenidos.

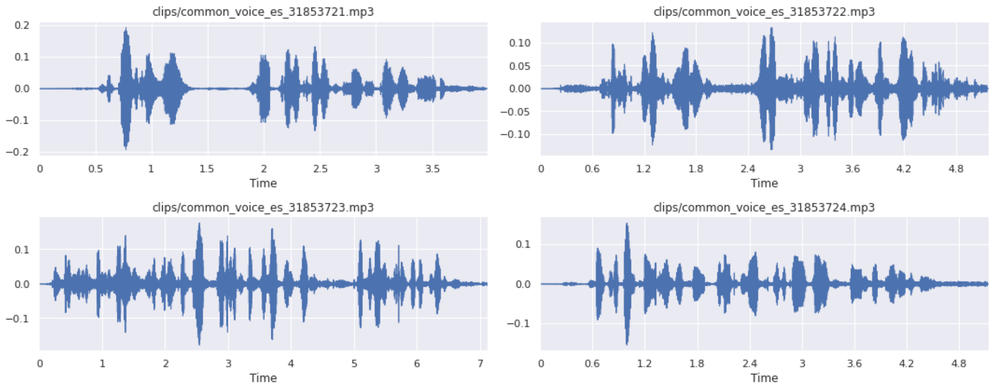
Librerías a utilizar

Si bien existen múltiples bibliotecas de Python que le permiten trabajar con datos de audio, para este ejemplo, usaremos librosa, noisereduce y soundfile, donde se carga un archivo MP3 y trazamos su contenido.

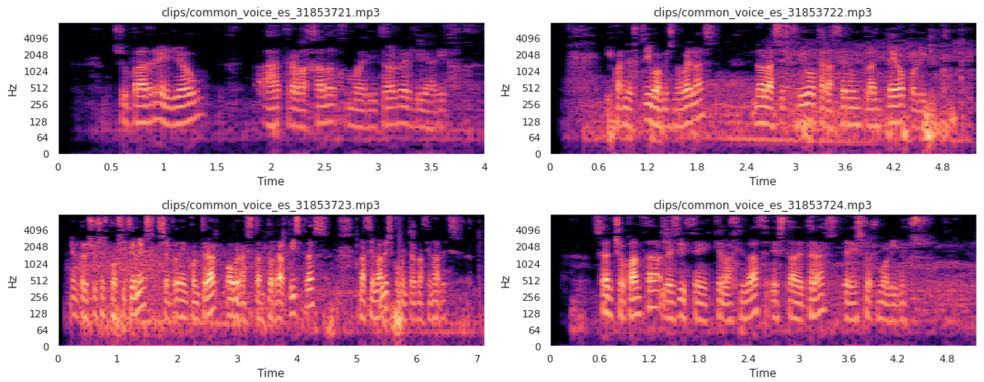
```
# { display-mode: "form" }
colab_requirements = [
    "pip install librosa",
    "pip install noisereduce",
    "pip install soundfile",


]

import sys, subprocess
```

Visualización de datos

- Se evidencia en las graficas de onda que el 3 tiene un ruido de fondo variable que cubre múltiples frecuencias, mientras que el ruido de fondo en la muestra 4 es bastante constante. Esto es también lo que vemos en las figuras de arriba. La muestra 3 es muy ruidosa en todo momento, mientras que la muestra 4 es ruidosa solo en unas pocas frecuencias.



- Al escuchar estas grabaciones podemos observar que esto se debe a mucho ruido de fondo.
- Para comprender mejor cómo se representa esto en el dominio de la frecuencia, veamos los espectrogramas STFT correspondientes.

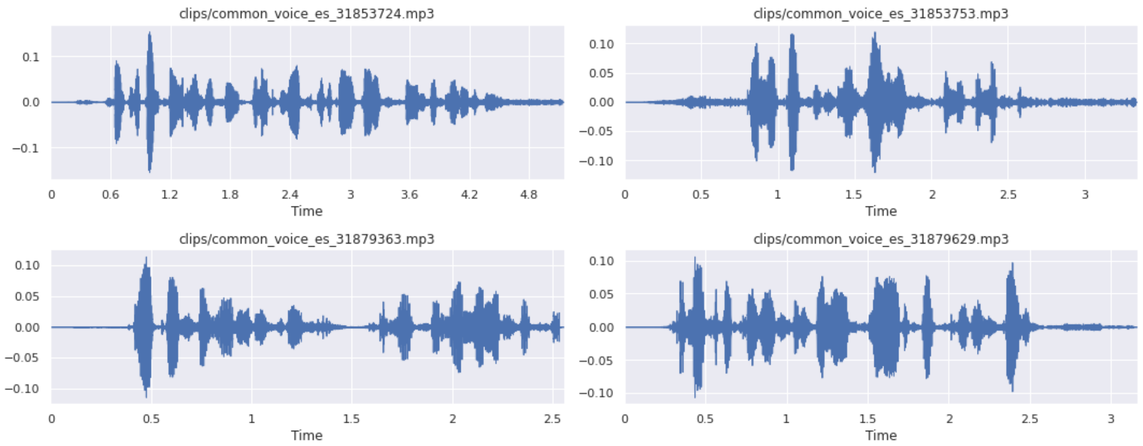


Preparación y limpieza de datos

Se realizo una preparación de los datos porque se necesitaba extraer unas características importantes como lo son: la duración , el tempo , las palabras por minutos , la cantidad de palabras. también se muestra una representación en forma de onda de la oración hablada. Se procede entonces a eliminar el ruido de las muestras.

Resultados de los datos

- Para el paso de recorte podemos usar .effects.trim() la función de librosa. Tenga en cuenta que cada conjunto de datos puede necesitar un top_db parámetro diferente para el recorte, por lo que lo mejor es probar algunas versiones y ver qué funciona bien. En nuestro caso lo es top_db=20.



Ageregando columnas necesarias para el entrenamiento, eliminando las columnas innecesarias.

	path	age	gender	nwords	duration	wps	tempo	f0_mean	f0_median	f0_std	f0_5perc	f0_95perc
0	common_voice_es_31853724.mp3	fourties	male	0	0	0	0	0	0	0	0	0
1	common_voice_es_31853753.mp3	fourties	male	0	0	0	0	0	0	0	0	0
2	common_voice_es_31879363.mp3	fourties	male	0	0	0	0	0	0	0	0	0
3	common_voice_es_31879629.mp3	fourties	male	0	0	0	0	0	0	0	0	0
4	common_voice_es_31879738.mp3	fourties	male	0	0	0	0	0	0	0	0	0

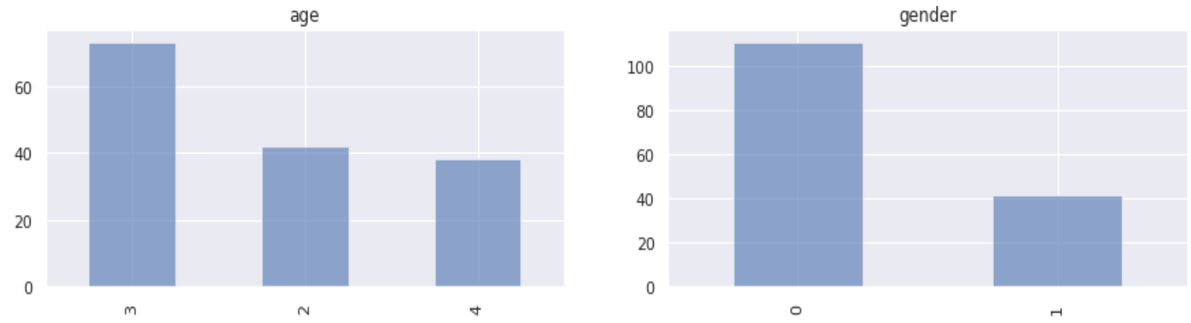
Detección de inicio, duración de audio, tempo

- Aquí se detecta cada una de las palabras habladas para poderlas contar, se calcula la velocidad y duración de cada muestra, extraemos el tempo; posteriormente hacemos la extracción de la frecuencia fundamental, la cual nos permite extraer aun mas características.

	path	age	gender	nWords	duration	wps	tempo	f0_mean	f0_median	f0_std	f0_5perc	f0_95perc
0	common_voice_es_31853724.mp3	fourties	male	30	5.148	5.827506	25.68	0	0	0	0	0
1	common_voice_es_31853753.mp3	fourties	male	17	3.348	5.077658	24.67	0	0	0	0	0
2	common_voice_es_31879363.mp3	fourties	male	14	2.556	5.477308	32.33	0	0	0	0	0
3	common_voice_es_31879629.mp3	fourties	male	18	3.168	5.681818	29.30	0	0	0	0	0
4	common_voice_es_31879738.mp3	fourties	male	21	3.456	6.076389	18.20	0	0	0	0	0

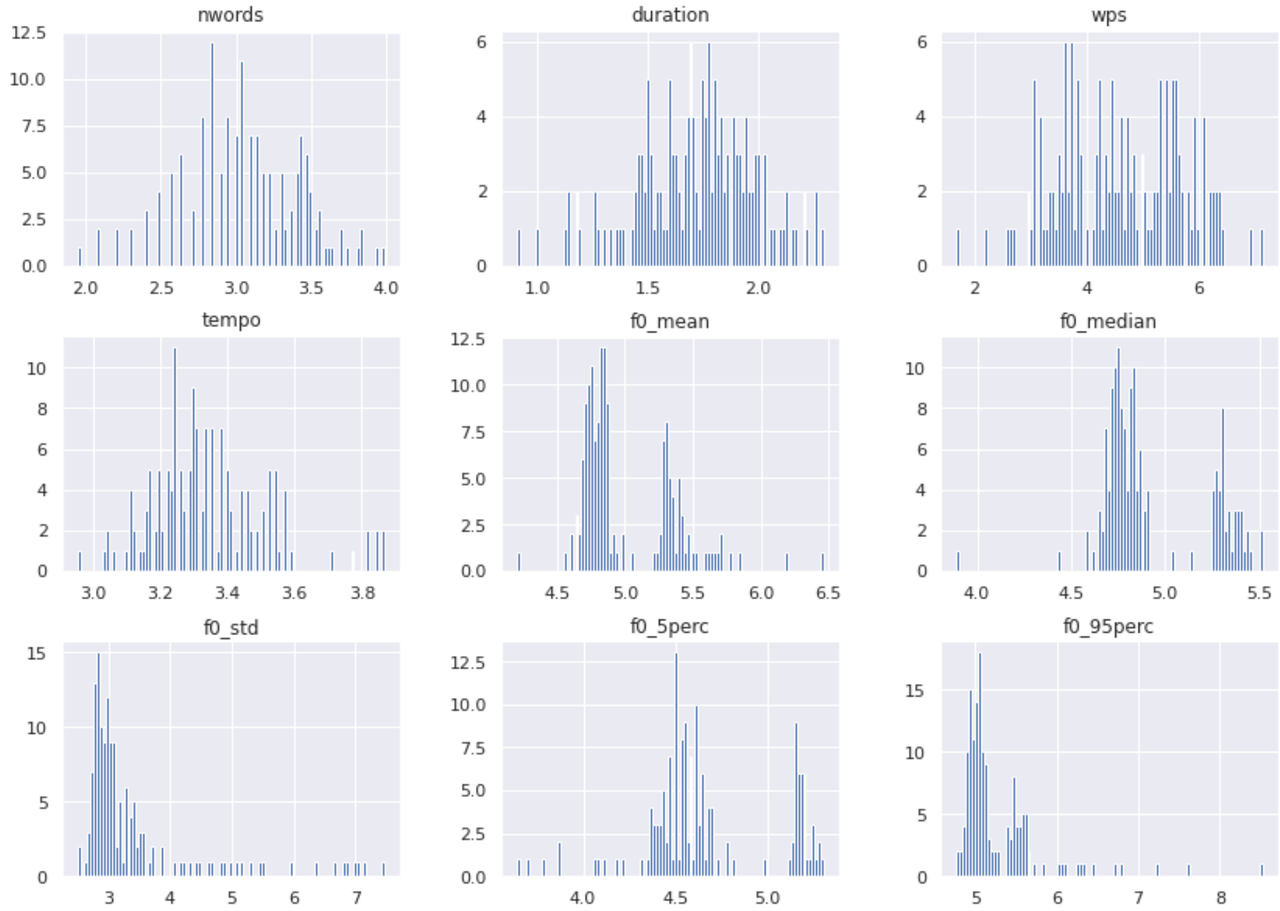
Características se desean clasificar

- Primero, veamos las distribuciones de clase de nuestras clases objetivo potenciales Age y gender.

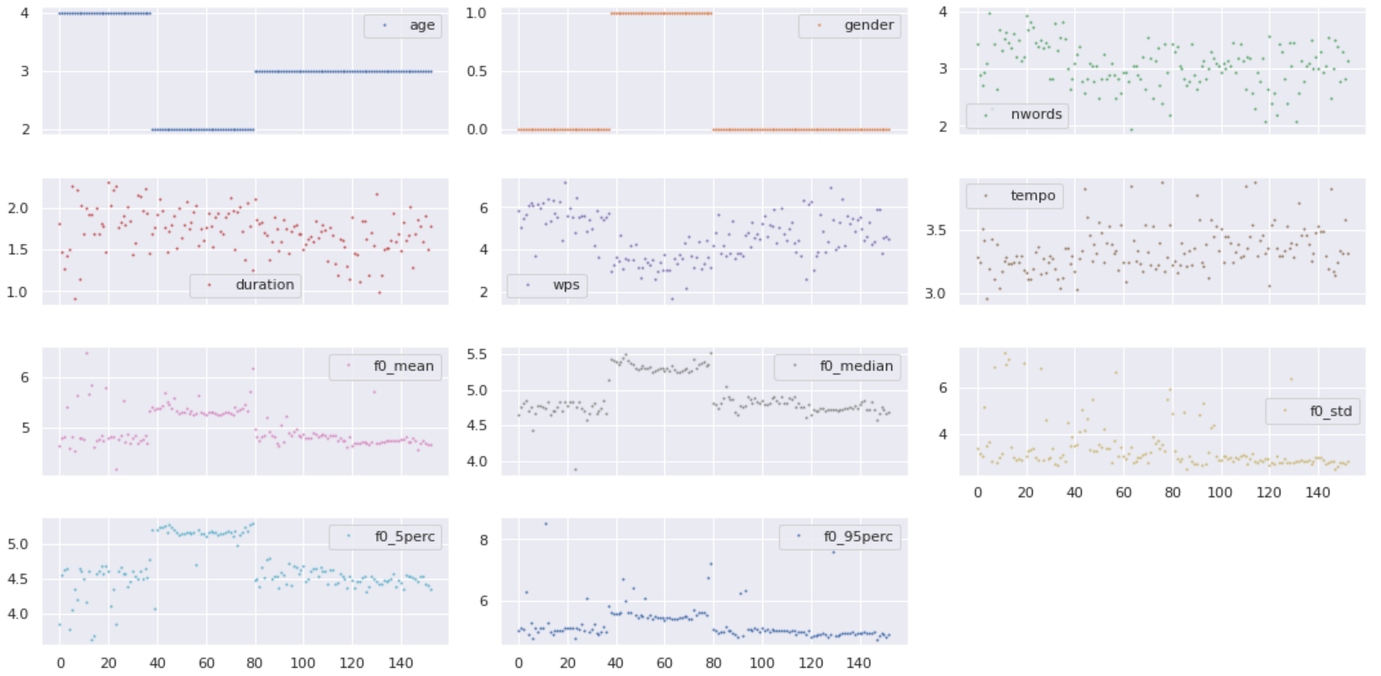


Características extraidas

- A excepcion de words_per_second, la mayoría de estas distribuciones de características están sesgadas hacia la derecha y, por lo tanto, podrían beneficiarse de una transformación logarítmica.



- Como se sospechaba, ¡parece haber un efecto de género aquí! Pero lo que también podemos ver es que algunas f0 puntuaciones (aquí en particular en los varones) son mucho más bajas y más altas de lo que deberían ser. Estos podrían ser valores atípicos, debido a una mala extracción de características. Echemos un vistazo más de cerca a todos los puntos de datos con la siguiente figura.

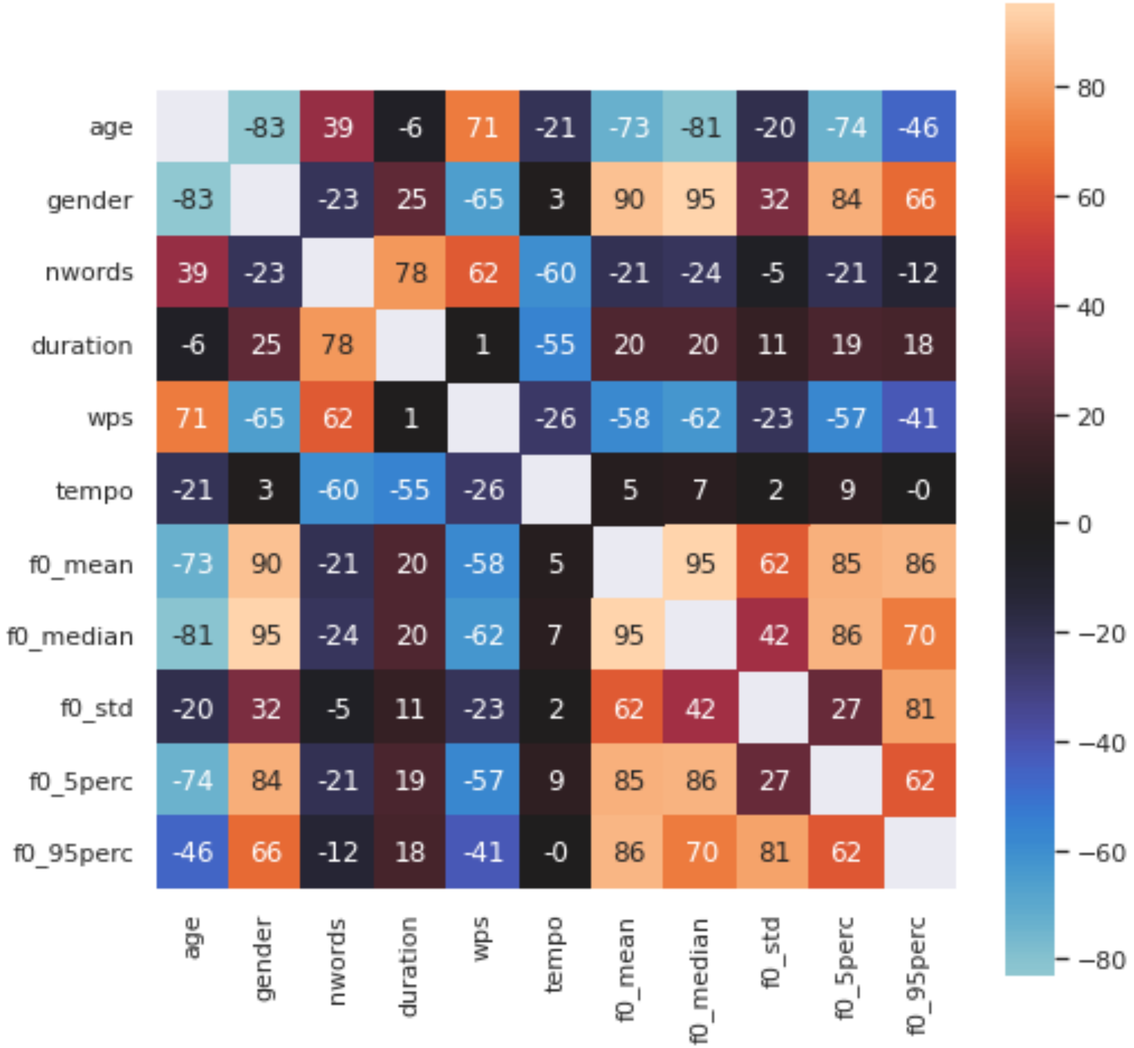


```
[ ]  from scipy.stats import zscore

     # Only select columns with numbers from the dataframe
     d2_num = d2.select_dtypes(np.number)

     # Apply zscore to all numerical features
     d2_num = d2_num.apply(zscore)

     # Identify all samples that are below a specific z-value
     z_thresh = 3
     mask = np.sum(d2_num.abs() > z_thresh, axis=1).eq(0)

     # Only keep the values in the mask
     d2 = d2[mask]
     d2.shape

     (140, 12)
```

- Dada la poca cantidad de funciones y el hecho de que tenemos distribuciones bastante atractivas con colas pronunciadas, podríamos revisar cada una de ellas y decidir el umbral de corte atípico característica por característica. Pero para mostrarle una forma más automatizada, usemos un enfoque de puntuación z en su lugar.

Mapa de correlación

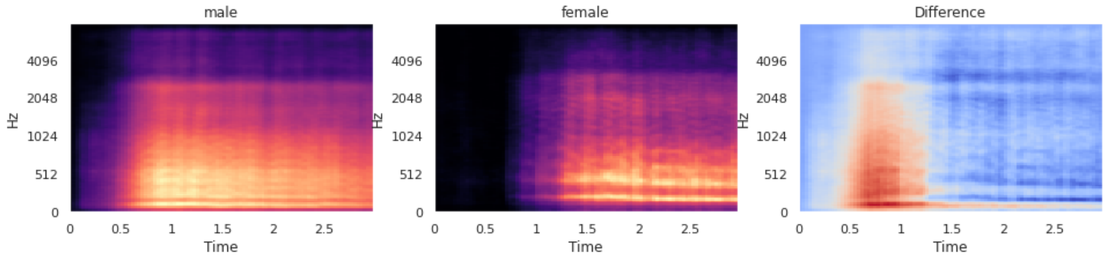
- Como siguiente paso, echemos un vistazo a la correlación entre todas las características. Pero antes de que podamos hacer eso, avancemos y codifiquemos también las características de destino no numéricas haciendo un mapeo manual.
- Ahora estamos listos para usar .corr() la función pandas junto con la de seaborn heatmap() para obtener más información sobre la correlación de características.



Características del espectrograma

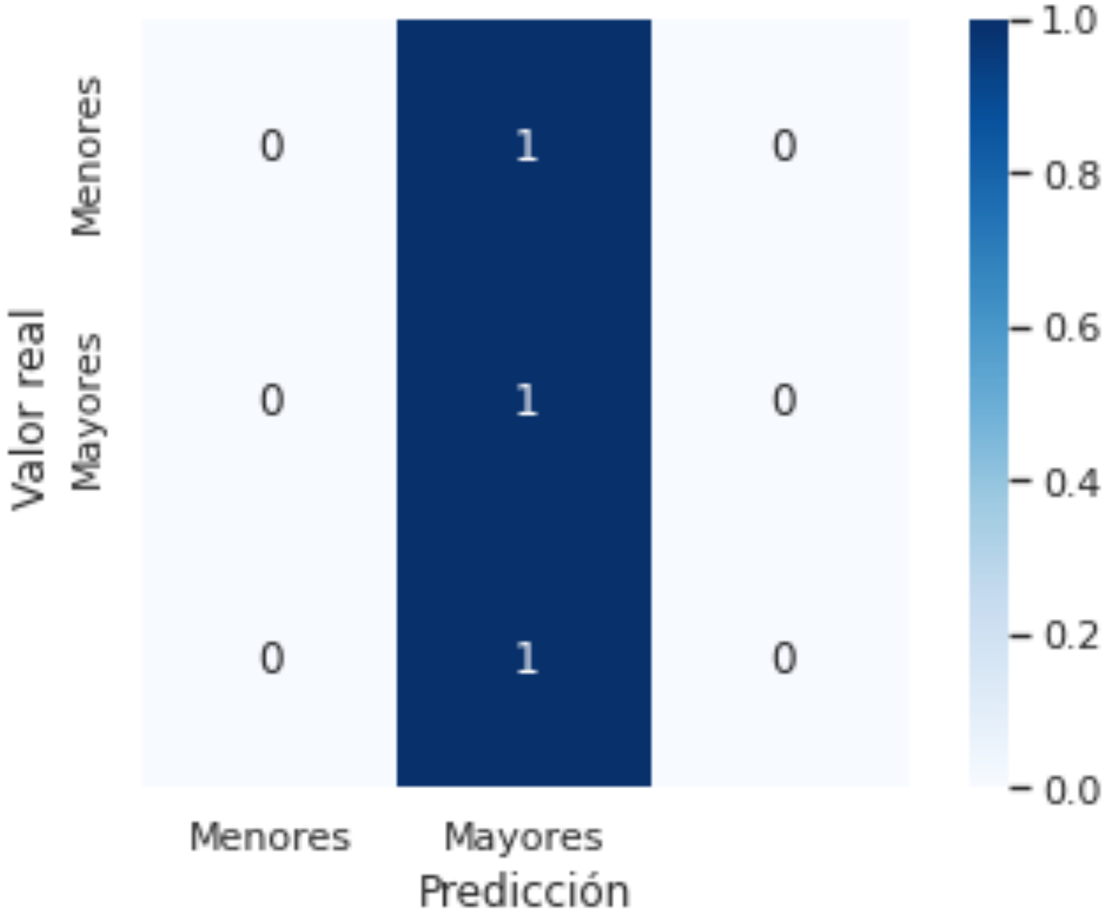
- tenemos varias opciones de representación de audio, como forma de onda o espectrograma STFT, mel o mfccs. Para esta exploración en particular, continuaremos utilizando los espectrogramas de mel.

- No obstante, antes de poder hacerlo, debemos tomar en cuenta un factor importante: las muestras de audio tienen diferentes longitudes, lo que se traduce en espectrogramas de diferentes longitudes. Para normalizar todas las grabaciones, las cortaremos a una duración de exactamente 3 segundos. Es decir, las muestras demasiado cortas se alargarán y las muestras demasiado largas se recortarán.



Evaluación del DataSet con estimadores

- Tomamos los datos del archivo CSV y los combinamos con un sencillo modelo RandomForest para ver hasta qué punto podemos predecir la edad de un hablante. Para empezar, carguemos los datos y dividámoslos en un conjunto de entrenamiento y otro de prueba.
- GaussianNB
- RandomForest



Conclusión

La biometría por medio de voz permite determinar la identidad de una persona para tener acceso inmediato a datos personales (redes sociales) con el fin de generar una barrera de seguridad al momento de ingresar a las cuentas de los usuarios, con esta herramienta se tienen en cuenta aspectos fisiológicos de las personas que los hacen únicos, también es de acceso remoto y es eficaz así halla ruido al fondo de donde se realice la autenticación.

Thank You!