Clasificación de Audio para predecir Sentimientos

Arnoldo Oliva

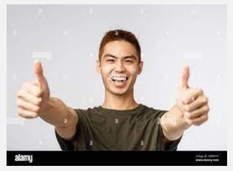
UANL FCFM MCD

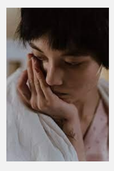
Motivación y resumen

El análisis de sentimiento puede tomar muchas formas, por ejemplo el análisis de texto donde se tokenizan palabras; y, de igual manera, el procesamiento de audio puede entrar en el mismo análisis al permitir identificar que palabras esta diciendo una persona.

En el presente proyecto se aborda un análisis de sentimiento fundamentado en el procesamiento de audio aunque por el lado de las características del sonido y no tanto del contenido (las palabras en sí), ya que en las muestras las personas repiten frases sin mucha profundidad en su significado, por lo tanto el proyecto se enfoca en **cómo** se dice la palabra, para clasificar el sentimiento imperante de esa persona en dicha muestra.

Lo anterior mencionado se realizará usando técnicas de procesamiento de audios y análisis de los mismos, como por ejemplo, análisis de frecuencias, espectogramas, MFCC (extracción de características), entre otros. Las muestras se component de frases de no más de 20 segundos segmentados en dos listas de más de 2100 records de archivos .wav/.mp3 con las etiquetas de “Feliz” o “Triste” (Happy o Sad).





Importancia del proyecto

El por qué del Proyecto se centra en la capacidad de poder detectar con que emoción una persona dice algo centrandonos únicamente en la manera en que lo dijo (no el significado de las oraciones en sí).

Dicho proceso podría servir para Predicción del estado emocional de un cliente recién atendido, si quisiéramos aterrizarlo a un caso de uso real. Esto es demasiado útil para la empresa porque le serviría para conocer si el cliente quedó satisfecho o todo lo contrario, y a su vez, esto le permitiría saber a la empresa como proceder con dicho cliente respecto a su atención al mismo.

Objectives

• **Objetivo general:**
Mediante el uso de técnicas de análisis de audio y Machine Learning, predecir la emoción preponderante de dicho audio.

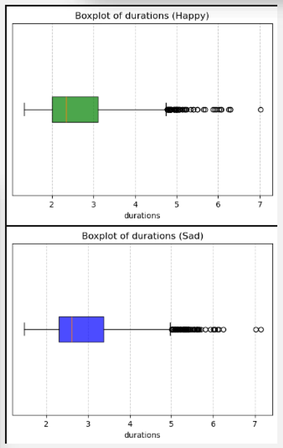
•Objetivo específico 1: Procesar los datos de tal manera que se lleven a cabo los análisis descriptivos y el algoritmo de clasificación (por ejemplo, eliminación de ruido extra, entre otros).

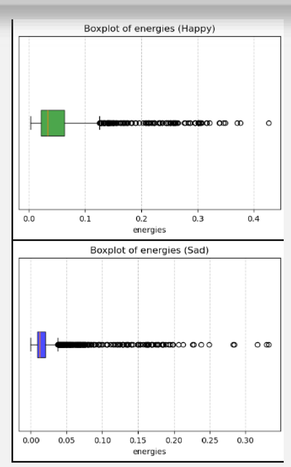
•Objetivo específico 2: Obtener características relevantes de los audios que nos permitan tener cierto aprendizaje, ya sea a través de espectogramas, análisis de frecuencias, extracción de características, entre otros.

Methods

Ya que los datos estaban contenidos en carpetas conteniendo audios .wav, se parsearon de tal manera que fueran leídos en formato de arrays en python y a su vez se obtuvo su “rate” (tasa de muestreo se refiere al número de muestras de audio tomadas por segundo, medida en Hz).

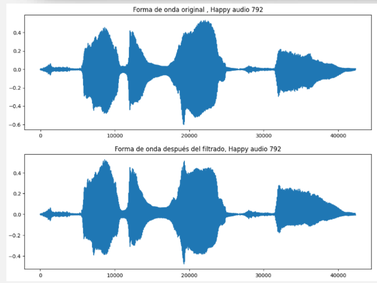
Como segundo paso se procedió a realizar análisis descriptivos (y comparativos) de las diversas ondas, para la representación visual solo se trabajaron con pequeñas muestras y para características generales, como por ejemplo, duraciones de los audios, amplitud de los sonidos, y “energies” (cuantifican la "fuerza" o "amplitud" general de una señal).

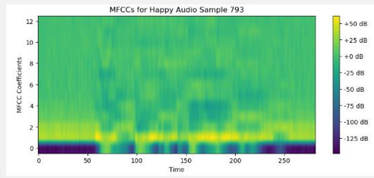


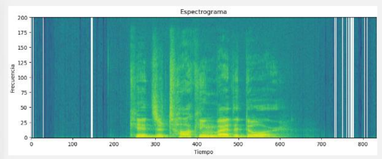


Una comparativa general de los audios de Happy vs Sad (verde y azul respectivamente), es que las duraciones de Sad duran más, pero las amplitudes (energies) de Happy son mucho más grandes (puede ser que lo que se pronuncia se hace con más “enjundia”, y más preciso por ser más corto.

Debido a que no existe manera visual sencilla para representar todas las muestras de manera visual se trabajo esto con submuestras de ambas etiquetas. Se visualizaron las Ondas en manera visual con gráficos de frecuencias y espectogramas (y de paso se les retiro el ruido adicional ajustando más las muestras con transformaciones de Fourier/Frequency Spectrum). Así mismo se realizaron extracciones de características con MFCC.







Una comparativa general de los audios.

Con un cutoff de los frequency spectrum se acotan los audios, removiendo el ruido adicional (gráficos de frecuencia de onda superiores). El gráfico de en medio es un MFCC, o extracción de características y el de abajo es un espeectograma que es una representación visual de la intensidad de las diferentes frecuencias en una señal de audio a lo largo del tiempo.

Results/Discussion

Future Directions

.