# Proyecto final de machine learning

# Planteamiento del Problema

El planteamiento del proyecto final consta de utilizar, analizar e interpretar un grupo de 100 audios, los cuales tienen las siguientes características:

1. La mitad de los audios son narrados por un hombre y la otra mitad es narrada por una mujer.
2. Por cada narración la misma persona realiza 3 cambios de voz, por lo cual para cada narración hay 5 audios diferentes con diferentes tonalidades (normal, grave, suave, aguda, nasal).

Con estos audios se quiere implementar un feature el cual nos permita entrenar un sistema machine learning y así poderlo evaluar.

Una solución que se planteó para este proyecto fue: análisis de frecuencias bajas (sonidos graves) de los audios, los cuales nos permitan identificar si el audio corresponde a un hombre o a una mujer.

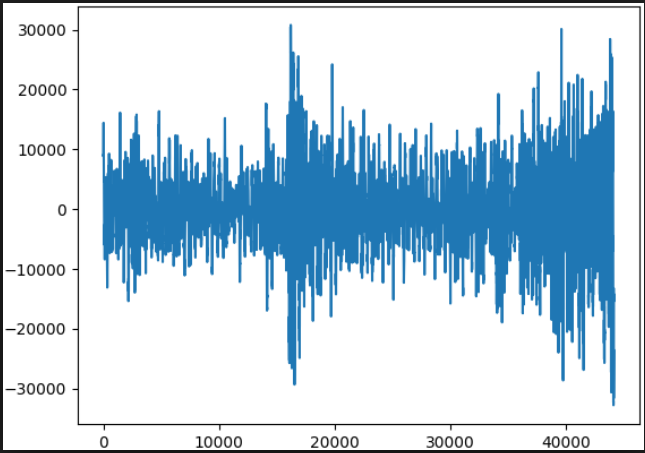
# Teoría de señales audibles

Una **señal de audio** es una señal analógica la cual normalmente está dentro del rango de frecuencias audibles por los seres humanos (frecuencias que podemos escuchar), que están aproximadamente entre los 20 y los 20. KHz.

Estas señales se pueden representar en varias maneras:

1. **Señales en el dominio del tiempo**

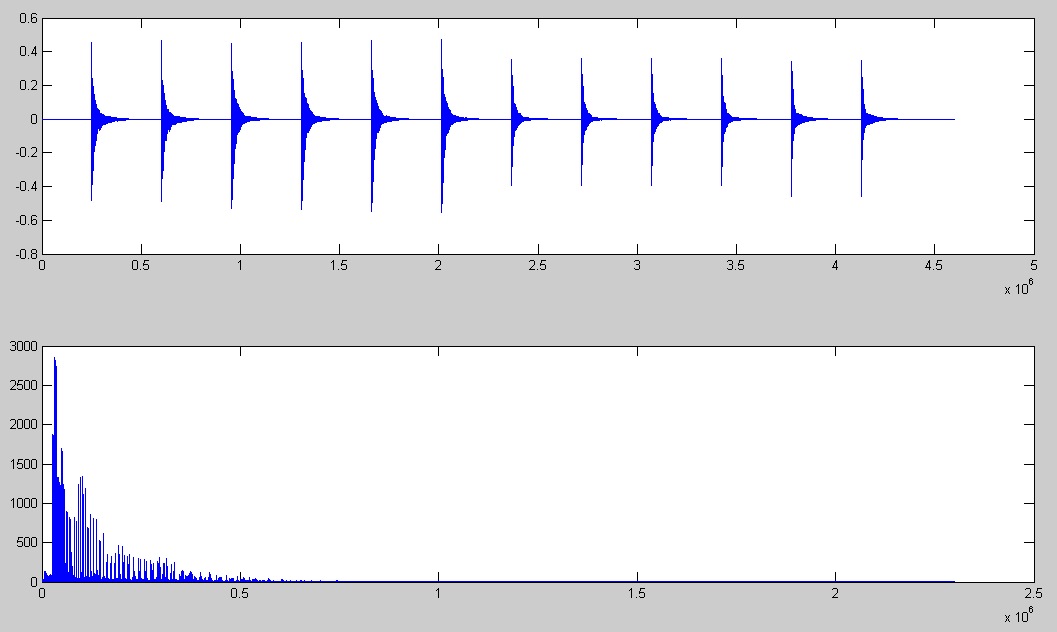
Es la representación gráfica de la potencia de un audio en el dominio del tiempo



***Figura 1 señal de audio en el dominio del tiempo***

1. **Señales en el dominio de la frecuencia**

Es la representación gráfica de la potencia de un audio en el dominio de la frecuencia, esta representación permite analizar más afondo las componentes y características de una señal, para esto previamente se requiere una señal en el dominio del tiempo y luego a esta señal se le realiza una transformada de Fourier para dejarla en el dominio de la frecuencia.



***Figura 2 señal de audio en el dominio de la frecuencia***

Para el análisis de señales de audio tendremos que manejar casi siempre los audios en el dominio de la frecuencia, ya **que es la representación toda la energía de la señal, distribuida en cada valor de la frecuencia, es decir que, al sumar toda la energía, frecuencia por frecuencia, obtendríamos el valor total de la frecuencia.**

Ya comprendido los dominós de representación de una señal de audio, como esta se representa y por qué es necesaria su interpretación, ahora procedemos a comprender lo que son las frecuencias audibles y los tonos.

-El **espectro audible** o **campo tonal**, se encuentra conformado por toda la gama de frecuencias que pueden ser percibidas por el oído humano su rango consta de los 20 Hz a los 20KHz.

El espectro audible podemos subdividirlo en tonos:

1. Tonos graves (frecuencias bajas, desde los 16 Hz a los 256 Hz).
2. Tonos medios (frecuencias medias, desde los 256 Hz a 2 kHz).
3. Tonos agudos (frecuencias altas, desde 2 kHz a los 16 kHz).

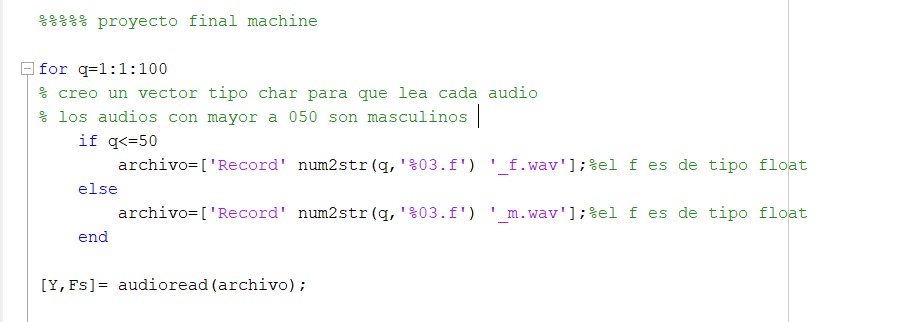
Teniendo en cuenta el rango de frecuencias graves, ***el objetivo del proyecto es analizar cada audio en el rango de los tonos graves y entrenarlo así para que identifique la voz de un hombre o la voz de una mujer, donde todo este análisis se planteara y se dará solución con los audios en el dominio de la frecuencia***

Ya con el conocimiento necesario de análisis de señales, procedemos a realizar la generación de nuestro dataset utilizando Matlab

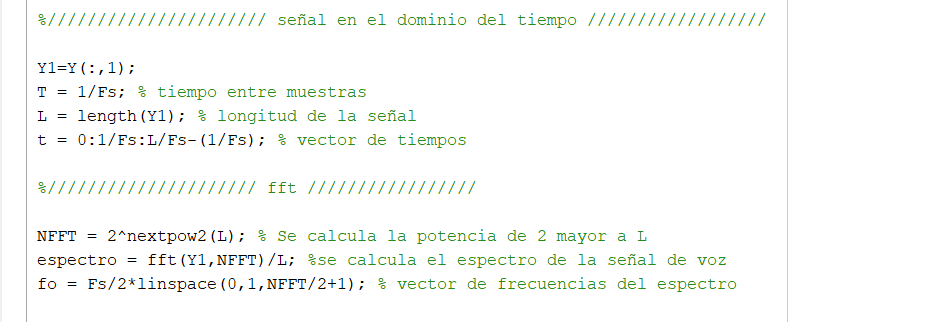
# Programa Matlab

Lo primero que hacemos es la lectura de los audios, si tenemos en cuenta de que hay un total de 100 audios debemos hacer el tratamiento, para cada uno de los audios, donde el objetivo es que se analice audio por audio y cada valor se almacene en una fila diferente de una matriz, de aquí la necesidad del ciclo “for” el cual nos va a permitir leer cada audio uno por uno hasta completar los 100.

También se utiliza el comando num2str el cual nos permite transformar un numero en un carácter permitiéndonos así ir variando audio por audio al igual que va avanzando el contador.

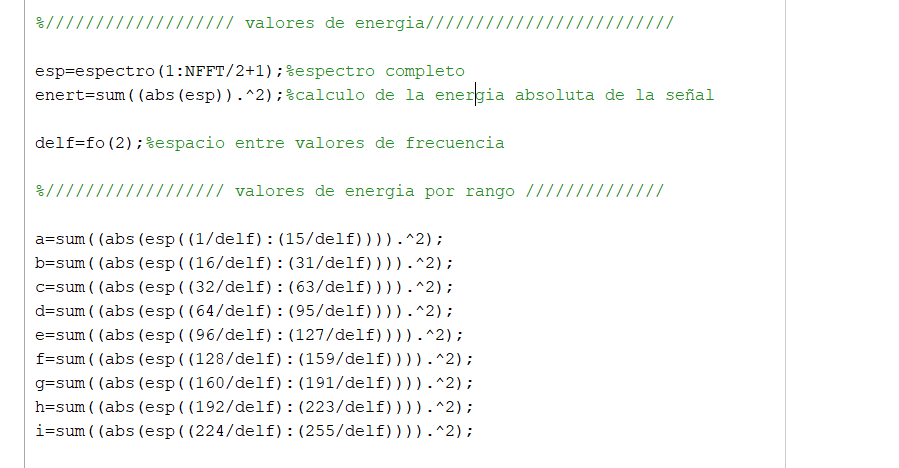


Ya con la señal de audio dentro del sistema procedemos a convertirlas en el dominio del tiempo, para luego hacerle la debida transformada de Fourier y dejarla planteada en el dominio de la frecuencia.



Ya con la señal en el dominio de la frecuencia procedemos a calcular el valor de la energía total de la señal teniendo en cuenta que se deben tener en cuenta los valores positivos de la señal de allí el valor absoluto de la función, y todo eso elevado a la “dos” debido a que estamos hablando de energía.

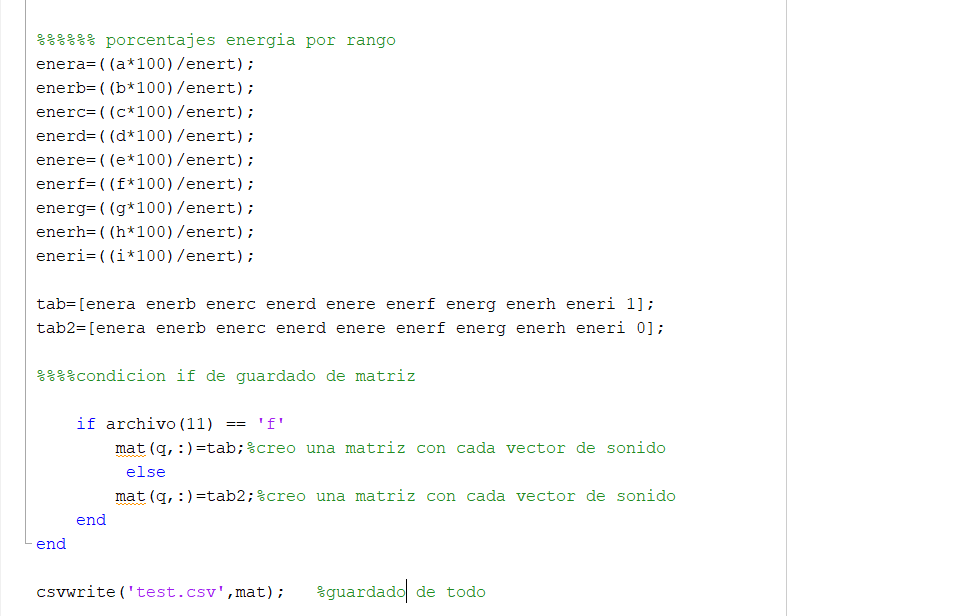
También teniendo en cuenta el rango de tonos graves (0 Hz a 256Hz) lo dividimos en sub tonos, cada uno con un ancho de banda de 16 Hz , donde para cada banda se le calcula el valor de energía, pero además para asignar el rango de cada banda en Hz, es necesario pasar el valor número del eje del espectro , ya que este no maneja los valores de frecuencias de allí la declaración de la variable delf la cual mide el valor de diferencia entre los puntos.



Ya con los valores de energía de cada ancho de banda, procedemos a calcularle el valor porcentaje de cada banda con respecto a la energía total del espectro, siendo estos valores de porcentaje los features de nuestro sistema, para ello solo se procede a realizar la división y eso almacenarlo en diferentes variables.

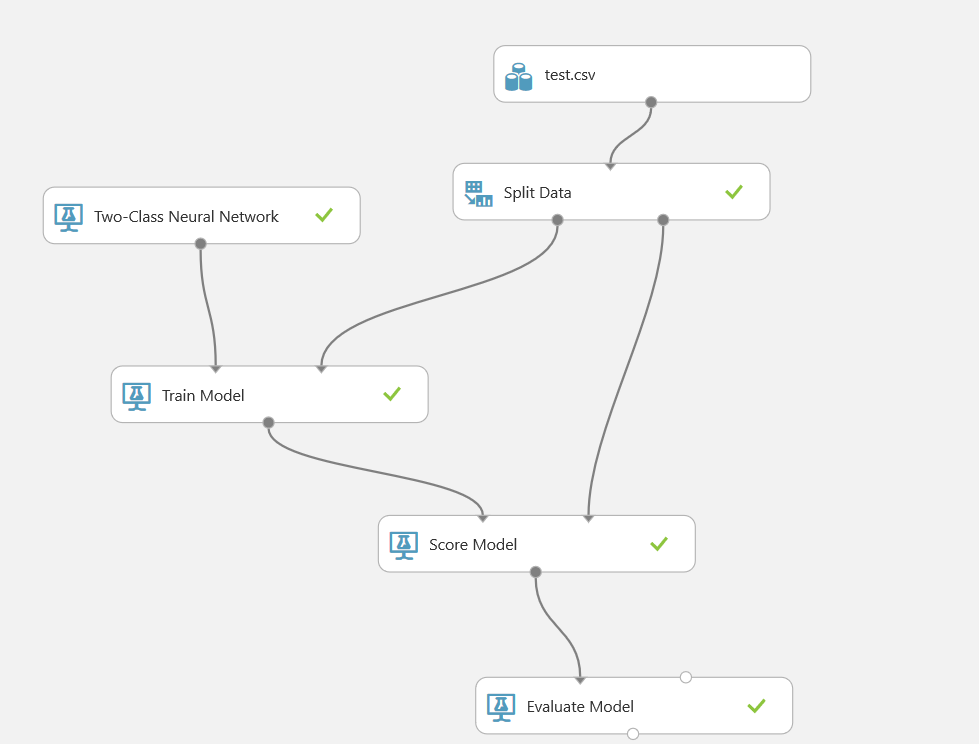
Con la relación de la energía ya calculada, creamos un dos vectores uno con salida “1” y otro con salida “0” vectores que se asignaran a la matriz final del dataset según si es hombre o mujer el que hable en el audio, para esto el sistema lee el nombre del audio lo compara con una condicional, la cual dice, que si el audio es de una mujer la salida es 1 y si es de un hombre la salida es 0 todo esto gracias a uso del comando “if”.

Al final de generar la matriz con los 100 audios analizados procedemos a almacenar los datos en un archivo “.cvs” la cual nos permitirá llevar lo audios a nuestro sistema de aprendizaje.

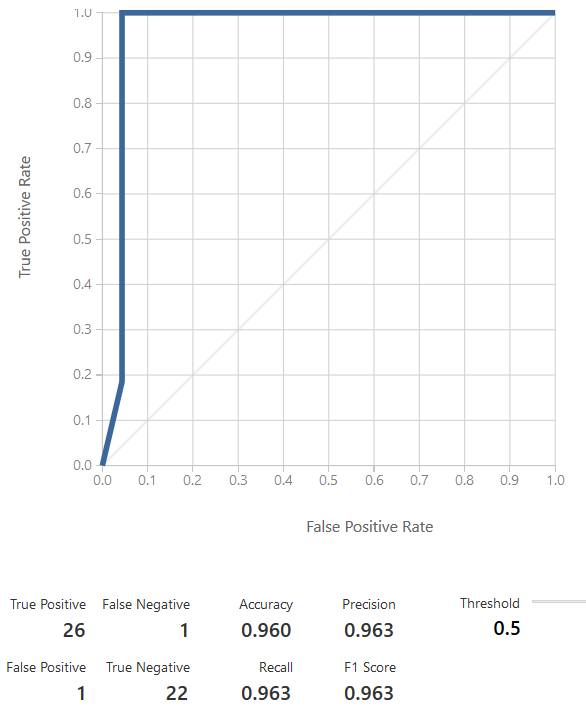
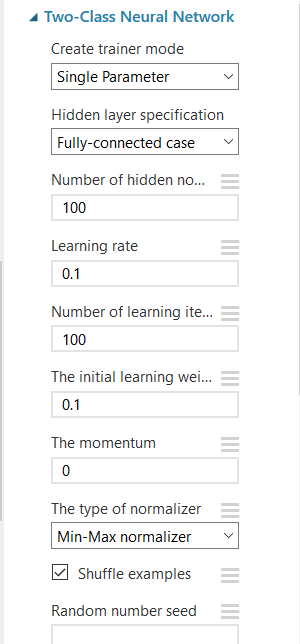


# Proyecto final de Microsoft azure

Con el dataset ya generado procedemos a crear un sistema de aprendizaje utilizando redes neuronales, con un modelo simple de análisis donde el objetivo será ir variando las características del sistema para obtener una mejor grafica de fp/fn y mejores valores de “f1 score” y “accuracy”.



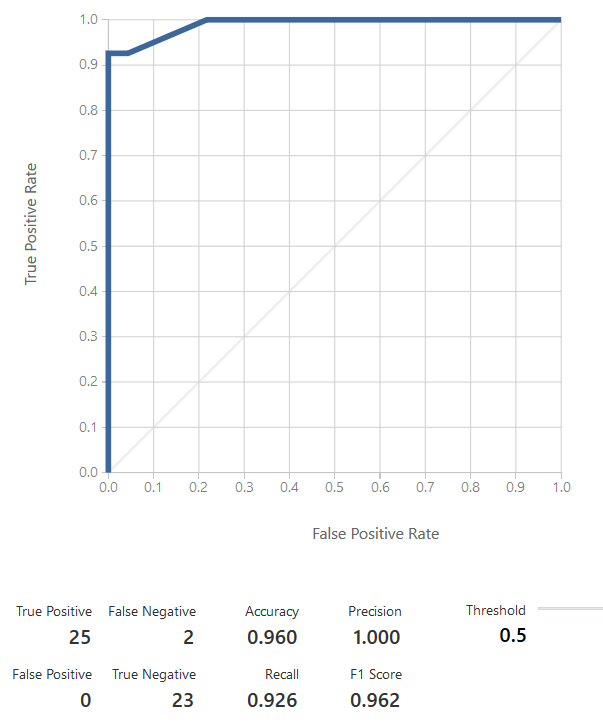
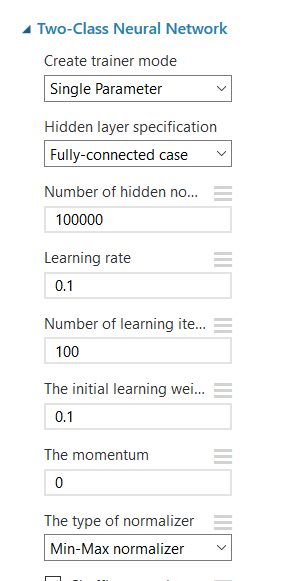
Lo primeros valores a tener en cuenta son los valores por defecto que nos brinda Azure para redes neuronales, a la par de que el Split corresponde a un valor de 50/50 y su división es aleatoria.



***Figura 3. valores iniciales por defecto***

## Variación # de nodos

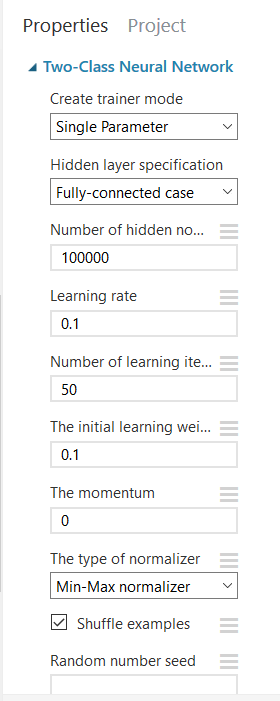
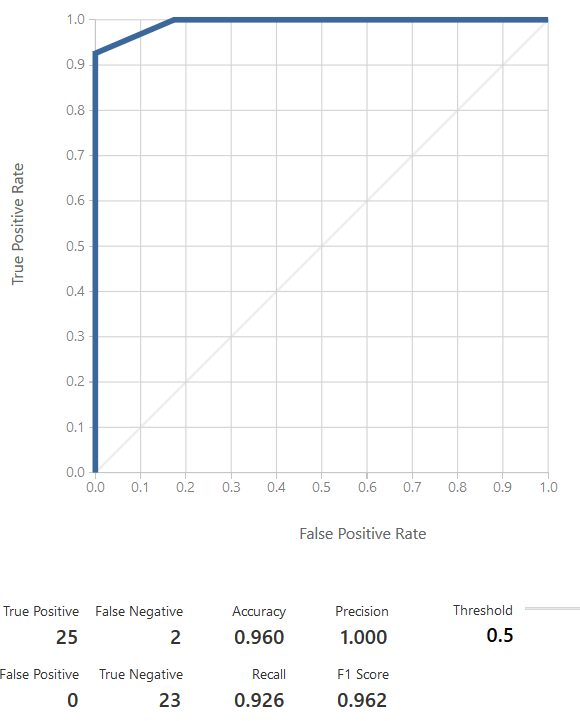
Se aumentaron el número de nodos los cuales mejoraron el valor de precisión



***Figura 3. Sistema con aumento de numero de nodos***

# Number of learning iterations

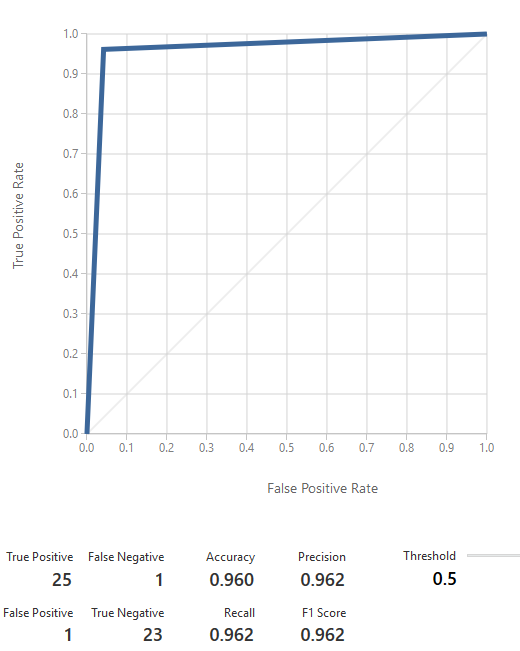
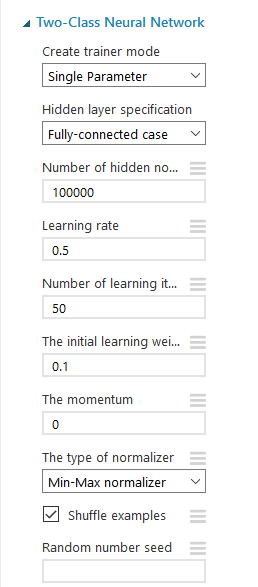
A diferencia del número de nodos, si nosotros variamos los valores de iteración esto empeorara los valores de nuestra gráfica , por esta razón no se evidencia cambio con los valores, pero lo que si se hace es reducir el número de iteraciones , lo suficiente para que no haya cambio y a su vez lo necesario para optimizar un poco el procesamiento del sistema.

***Figura 4. Sistema con learning iterations optimizado***

# Learning rate

El ultimo valor a modificar fue la taza de aprendizaje, la cual al variar entre más baja sea esta taza , los valores son peores, a la par de que el sistema requiere mayor procesamiento al momento de implementar un taza muy baja, por esta razón se buscó un valor alto de taza de aprendizaje la ccual no mejorara los valores de la gráfica de allí el valor final obtenido , el cual aunque baje la precisión , los demás valores aumentan de forma equitativa.



***Figura 5. resultados finales con taza de aprendizaje***

# Análisis y conclusiones

1. Se evidencio por medio de la practica diversos métodos de entrenar un sistema para que detecte dos clases distintas de audios, donde en este caso en particular se aplicó de manera correcta los conocimientos de análisis de señales para entrenar el sistema.
2. Se aplicó el concepto de señal y la teoría de las transformadas de Fourier en una señal real por medio de la herramienta de Matlab, donde se realizó un programa capaz de generar un dataset con base a las necesidades del problema.
3. Se utilizó un sistema machine learning el cual entreno un sistema por medio del método de redes neuronales, cuyos valores de entrenamiento fueron variando, basándose en la equidad entre los valores de Accuracy, precisión, Recall y F1 score. Y para este caso concreto también se tuvo en cuenta la optimización del procesador, evitando asi trabajos innecesarios .