



**UNIVERSIDAD DE ANTIOQUIA**  
**FACULTAD DE INGENIERÍA**

Introducción a la inteligencia artificial para ciencias e  
ingenierías  
Raul Ramos Pollan  
Semestre 2022-1

**Estudiantes:**

Miguel Angel Castaño Cardenas cc 1152225263

Juan Sebastian Pinto Fuentes cc 1007612134

**Entrega 1**

**1. DESCRIPCIÓN DEL PROBLEMA.**

Muchas veces cuando una persona asiste al médico con problemas respiratorios éste nos chequea haciendo uso de un fonendoscopio para detectar o descartar obstrucciones en las vías respiratorias ya que dependiendo el tipo de sonido que escuche puede descartar o predecir ciertas patologías, pero en muchas ocasiones es difícil distinguir con claridad qué tipo de sonido se está presentando, pues en ocasiones es necesario de un oído muy entrenado para poder dar un diagnóstico acertado y aquí es donde nosotros como bioingenieros podemos presentar una solución. Por ejemplo, un sonido sibilante es un signo común de que un paciente tiene una enfermedad obstructiva de las vías respiratorias como asma o enfermedad pulmonar obstructiva crónica (EPOC).

Dadas ciertas características en frecuencia de unas señales respiratorias, se va a predecir si estas contienen ciclos normales y ciclos con sibilancias o ciclos normales y ciclos con crepitantes; aunque claramente lo más importante es el análisis en frecuencia, también podemos complementar el estudio con otros parámetros como la edad, esto debido a que ciertos estudios muestran que según la edad hay mayor o menor posibilidad de encontrar un tipo de afección u otro.

**2. REPOSITORIO.**

Vamos a usar un dataset de Kaggle llamado “Respiratory Sound DataBase” (enlace: <https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database>). Esta base de datos contiene: 920 archivos de sonido .wav, 920 archivos de texto con anotaciones de cada medición donde se evidencia la duración de cada ciclo y si se encuentran crepitancias o sibilancias presentes, se encuentra también un archivo de texto que enumera el diagnóstico de cada paciente, un archivo de texto que explica el formato de nombre del archivo, un archivo de texto que enumera 91 nombres de los

pacientes y finalmente un archivo de texto que contiene información demográfica de cada paciente.

En la figura 1 se puede ver la forma que tiene un archivo de texto con anotaciones, contiene 4 columnas que son: tiempo de inicio del ciclo, tiempo de final del ciclo, crepitancias presentes y sibilancias presentes respectivamente. A partir de las señales de audio filtradas y los archivos de texto con anotaciones es que se evaluarán las métricas de desempeño requeridas.

```
0.036 0.579 0 0
0.579 2.45 0 0
2.45 3.893 0 0
3.893 5.793 0 0
5.793 7.521 0 0
7.521 9.279 0 0
9.279 11.15 0 0
11.15 13.036 0 0
13.036 14.721 0 0
14.721 16.707 0 0
16.707 18.507 0 0
18.507 19.964 0 0
```

Figura 1. Forma de archivos de texto con anotaciones.

### 3. MÉTRICAS DE DESEMPEÑO REQUERIDAS.

Como métricas de desempeño de Machine Learning vamos a usar las métricas accuracy, recall y el F1 score (Harmonic mean). La métrica accuracy, es el número de elementos identificados correctamente como positivo de un total de elementos identificados como positivos, es decir que esta métrica nos da cuenta de los falsos positivos. Un modelo que tenga 0 FP tendrá un accuracy del 100%. En la figura 2 se puede ver la matriz de confusión a partir de la cual salen las siguientes fórmulas.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figura 2. Matriz de confusión.

$$Accuracy = TP + TN / TP + FP + FN + TN \quad (1)$$

La métrica de recall, es el número de elementos identificados correctamente como positivos del total de positivos verdaderos, es decir recall nos da información sobre el rendimiento de un clasificador con respecto a falsos negativos. Un modelo que tenga 0 falsos negativos tendrá un recall del 100%. A continuación encontramos la fórmula para el cálculo del recall.

$$Recall = TP / TP + FN \quad (2)$$

Por último la métrica F1, surge a partir de la necesidad de poder medir en una sola métrica varios valores que evalúan el modelo, para este caso específico con el F1 las métricas que se juntan son el Recall y Accuracy. Para el cálculo de la media armónica se utiliza la siguiente fórmula:

$$F1 = 2 * Accuracy * Recall / (Accuracy + Recall) \quad (3)$$

La media armónica es una especie de promedio cuando accuracy y recall son iguales. Pero cuando accuracy y recall son diferentes, entonces está más cerca del número más pequeño en comparación con el número más grande; esto evita que se sobrevalore un modelo que realmente puede no estar haciendo bien su trabajo. Se sabe que el recall nos da información sobre el rendimiento de un clasificador con respecto a falsos negativos (FN), mientras que la precisión nos proporciona información sobre su rendimiento con respecto a los falsos positivos (FP). Por ejemplo, un algoritmo que fue entrenado trabaja con un dataset de 1000 pacientes, hay 700 sanos y 300 enfermos, sin embargo, el modelo da como resultado una predicción de 800 enfermos y 200 sanos; lo anterior da como resultado un accuracy del 50% y un recall del 100%, a partir de estos resultados podemos concluir que estas métricas podrían dar una falsa idea del buen funcionamiento del modelo, dando una calificación buena a un modelo que prácticamente solo está clasificando a los pacientes como enfermos. El puntaje F1 para el sistema anterior es de 66% lo que nos muestra que en efecto está más cerca del valor menor entre precisión y recall y puede dar una idea más acertada a la hora de evaluar un modelo.

#### 4. DESEMPEÑO DESEABLE.

Al estar tratando con temas relacionados con la salud de las personas y que un diagnóstico dependa de ese resultado, lo ideal es que el resultado arrojado siempre sea el correcto, sin embargo sabemos que esto es prácticamente imposible, por lo tanto estableceremos un rango de tolerancia.

Para establecer este rango nos basamos en la experiencia que tenemos con este tipo de proyectos, en un escenario muy favorable un error del 10% representaría un resultado casi que ideal o excelente para nuestras condiciones, sin embargo

basándose en la complejidad de los datos y la semejanza entre los resultados en frecuencia, podríamos extendernos hasta un 20% en los resultados, es decir que un *Accuracy* y un *Recall* del 80% será considerado como aceptable para el caso.