#### 

**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.

1. 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.

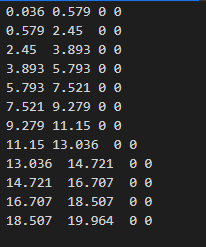


Figura 1. Forma de archivos de texto con anotaciones.

1. 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 partir de la cual salen las siguientes fórmulas.

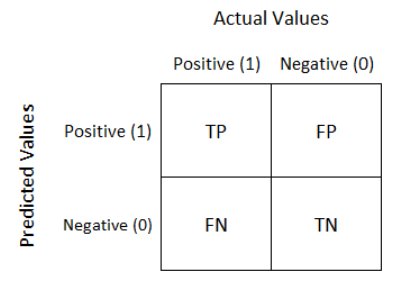


Figura 2. Matriz de confusión.

(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.

(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:

(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 sobrecalifique 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 al hora de evaluar un modelo.

1. DESEMPEÑO DESEABLE.

#### ENTREGA 1[¶](https://rramosp.github.io/ai4eng.v1/content/M00_intro_udea.html#entrega-1)

Esta primera entrega sólo consta de un archivo en el que (1) describas el problema predictivo a resolver, (2) el dataset que vas a utilizar, (3) las métricas de desempeño requeridas (de machine learning y de negocio); y (4) un primer criterio sobre cual sería el desempeño deseable en producción.

Por ejemplo:

1. dadas las caracetrísticas de una casa (superficie, localización, etc.) vamos predecir el precio de venta de una casa en el mercado.
2. vamos a usar el dataset de kaggle esta competición (poner el enlace), que tiene X número de muestras (casas) y tales columnas (poner la lista de columnas, o si es muy grande, poner las que se consideren más representativas para dar una idea de cómo es el dataset).
3. como métrica de machine learning vamos a usar el MAE (Mean Absolute Error) que es el define la propia competencia (podría ser otro si el de la competencia es complicado). Por ejemplo, si el precio de una casa es 100K y un modelo predice 120K, entonces el error es del 120%. Pero si el precio es de 40K y el modelo predice 20K el error es del 50%. Como métrica de negocio se podría usar el incremento en ventas gracias a la utilización del modelo.
4. si las ventas no aumentan más de un 10% no merece la pena poner el modelo en producción ya que el coste de desarrollo y mantenimiento no cubriría las ganacias adicionales de ese aumento.

tened en cuenta que el proyecto va hasta la métrica de machine learning (tenéis que reportar el rendimiento de los modelos). La métrica de negocio y el punto (4) es más una reflexión de cómo se usaría el modelo es un caso hipotético de que se integrase en la operación de una empresa u organización

Otros ejemplos de los puntos 3 y 4:

* Ejemplo 1: nuestro modelo de predicción de la patología X en pacientes debería de tener un porcentaje de acierto >80%, pero también un false negative rate <5%, ya que es una patología grave y es preferible no fallar una detección de un paciente que verdaderamente tiene la patología, aunque eso implique que aumente el número de falsos positivos.
* Ejemplo 2: según el departamento de marketing de cierta empresa, un modelo de predicción del siguiente producto que compre un cliente debería de tener un porcentaje de acierto de al menos 50%, ya que se usará el modelo para sugerir recomendaciones a los usuarios. Si el porcentaje de acierto es menor sería contraproducente porque perderíamos ventas.