



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

1. Introducción

Como se expuso en la entrega anterior, 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 125 pacientes y finalmente un archivo de texto que contiene información demográfica de cada paciente.

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 da cuenta de los falsos positivos.

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.

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. 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.

2. Exploración descriptiva del dataset

La exploración del dataset se realizó haciendo uso de un notebook de colab, donde se encontró que los datos de los pacientes que se incluyeron en el dataset, están clasificados tanto por edades como por la posición del estetoscopio en el momento de realizar la medición. Tomando como base la clasificación por edades de la figura 1, se encontró que de los 125 pacientes hay:

- 36 bebés
- 7 infantes
- 7 adolescentes
- 1 adultos jóvenes
- 12 adultos
- 62 adultos mayores

Bebés	0-5 años
Infantes	6-12 años
Adolescentes	13- 19 años
Adultos jóvenes	20-34 años
Adultos	35 -60 años
Adultos mayores	< 60 años

Figura 1. Clasificación poblacional por edades.

Respecto a la clasificación según la posición del estetoscopio, se describe en la competencia que hay 7 posibles casos, a continuación se muestra cada una de las posiciones con el respectivo número de mediciones para cada posición. Cabe aclarar que se realizó más de una medición por paciente, con diferentes posiciones de estetoscopio, debido a esto hay más datos de mediciones que de pacientes.

- 36 bebés
- 7 infantes
- 7 adolescentes
- 1 adultos jóvenes
- 12 adultos
- 62 adultos mayores

3. Iteraciones de desarrollo

El preprocesado de datos que se realizó fue el mismo para todas las iteraciones, se decidió clasificarlo según la posición del estetoscopio, debido a esto se realizó una tabla con la siguiente estructura:

Tabla 1. Descripción del procesamiento del dataset

Edad	Sexo	PI	Pr	AI	Ar	Tc	LI	Lr	Diagnóstico	Audio
74	1	0	1	0	0	0	0	0	1	

Donde la primera columna se refiere a la edad correspondiente a cada paciente, siguiente está la columna del sexo, que se clasifica como 1(Masculino) o 0 (Femenino), las columnas referentes a cada ubicación del estetoscopio se clasifican con el valor de 1 si esa es la posición y 0 al resto de columnas, posteriormente se encuentra la columna de diagnostico que corresponde a los casos de sano y enfermo; esta clasificación se realizó tomando como base el archivo de información de cada medición donde se informaba si habían sibilancias o crepitancias presentes en los ciclos respiratorios, según esta información se clasifica al paciente como sano (sin sibilancias o crepitancias) con el número 0 y como enfermo (con sibilancias o crepitancias) con el número 1, y finalmente encontramos la columna de audio que cuyo contenido es el resultado del calculo de la densidad espectral del ciclo respiratorio al que se hace referencia en cada fila del dataframe, con estos datos se realiza el entrenamiento, prueba y validación del algoritmo de machine learning.

3.1. Iteración 1:

3.1.1. Modelo supervisado: Para este modelo se utilizó una red neuronal “MLPClassifier” con los siguientes parámetros, MLPClassifier(hidden_layer_sizes=(8,8,8), activation='relu', solver='adam', max_iter=500). Se destinó un 70% de los datos para train y un 30% para test.

3.1.2. Resultados, métricas y curvas de aprendizaje: Luego de realizar las pruebas al modelo con los datos de test, se obtuvieron los siguientes resultados.

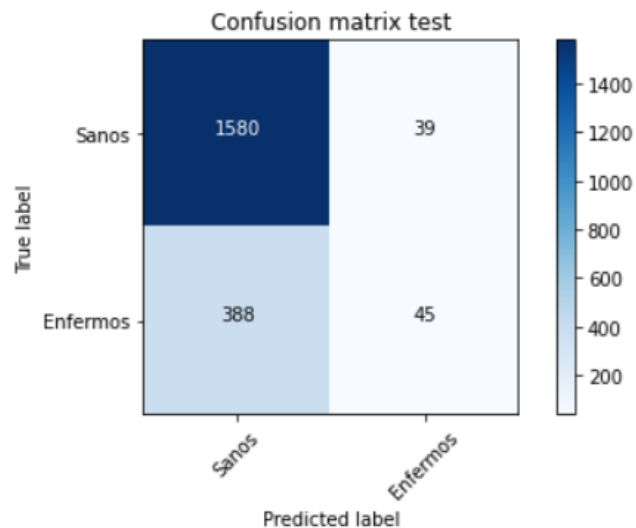


Figura 2. Matriz de confusión iteración 1.

	precision	recall	f1-score
Sanos	0.80	0.98	0.88
Enfermos	0.54	0.10	0.17
accuracy			0.79
macro avg	0.67	0.54	0.53
weighted avg	0.75	0.79	0.73

Figura 3. Métricas de desempeño iteración 1.

En la figura 2 se ve la matriz de confusión referente al primer modelo, con valores de TP = 1580, FP = 39, FN = 388 y TN = 45. A partir de los valores anteriores se hizo el cálculo de las métricas de desempeño definidas en el primer informe (accuracy, recall, F1 score), teniendo como resultado un accuracy = 75%, un recall = 79% y un F1 = 73%. Los resultados anteriores nos dicen que este modelo tiene un comportamiento regular ya que no alcanza el desempeño mínimo esperado (80% accuracy, bajo% recall), además nos indican que tiene un buen manejo de los falsos positivos ya que la métrica accuracy es alta, pero también que presenta muchos falsos negativos porque el recall es de casi el 80%, esto lo confirma la matriz de confusión. Finalmente la media armónica F1 coincide con el resultado ya que un porcentaje del 73% para un modelo indica que de manera general no funciona del todo bien. Al analizar más a profundidad, en la figura 3 se observa que el recall para los sanos es de 98%, lo que indica que hubo muy pocos pacientes sanos que se clasificaron como enfermos, pero el

recall asociado a los enfermos es del 10% lo que indica que muchos enfermos se clasificaron como sanos, ambos resultados se confirman en la matriz de confusión.

3.2. Iteración 2:

- 3.2.1. Modelo supervisado:** Para este modelo se utilizó un SVM con los siguientes parámetros SVM($\gamma = .1$), cabe resaltar que de todos los casos realizados con este modelo, este fue el que presentó mejores resultados, a pesar de que de manera general no lo son. También se destinó un 70% de datos para train y un 30% para test.
- 3.2.2. Resultados, métricas y curvas de aprendizaje:** Luego de realizar las pruebas al modelo con los datos de test, se obtuvieron los siguientes resultados.

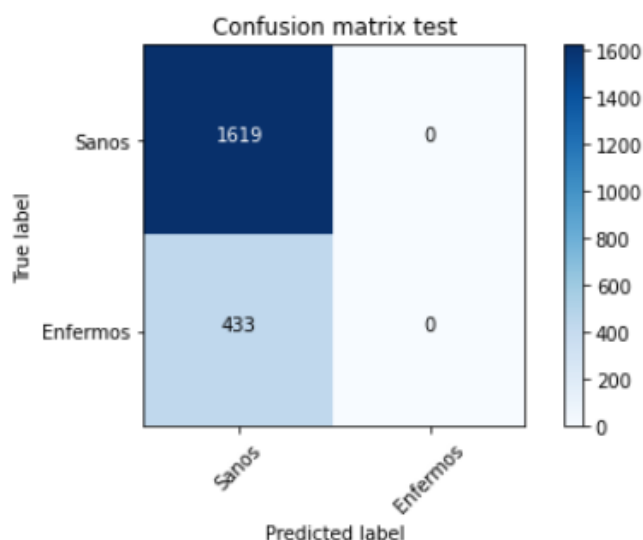


Figura 4. Matriz de confusión iteración 2.

	precision	recall	f1-score
Sanos	0.79	1.00	0.88
Enfermos	0.00	0.00	0.00
accuracy			0.79
macro avg	0.39	0.50	0.44
weighted avg	0.62	0.79	0.70

Figura 5. Métricas de desempeño iteración 2.

En la figura 4 se ve la matriz de confusión referente al segundo modelo, con valores de TP = 1619, FP = 0, FN = 433 y TN = 0.

A partir de los valores anteriores se hizo el cálculo de las métricas de desempeño definidas en el primer informe (accuracy, recall, F1 score), teniendo como resultado un accuracy = 62%, un recall = 79% y un F1 = 70%. Los resultados anteriores nos dicen que este modelo tiene un comportamiento muy malo ya que lo único que está haciendo es clasificar a todos los pacientes como sanos, esto lo podemos ver de la matriz de confusión y además en la figura 5 se ve que el recall para sanos es del 100% lo que termina de confirmar este comportamiento. También se puede ver que los valores de accuracy, recall y F1 están en cero para la fila de enfermos, lo que hace que tenga sentido que no haya clasificados enfermos en la matriz de confusión y que nos dice que el modelo no está clasificando bien los pacientes. Dos de los tres SVM implementados con los diferentes parámetros, entregaron este resultado de las métricas. Finalmente este modelo no cumple tampoco con las métricas de desempeño esperadas.

3.3. Iteración 3:

3.3.1. Modelo supervisado: Para esta iteración se utilizó un random forest classifier con un max_depth=20, se iba probando con diferentes valores de max_depth, pero este último resultó siendo el que mejores resultados arrojaba.

3.3.2. Resultados, métricas y curva de aprendizaje: luego de realizar las pruebas, se obtuvieron los siguientes resultados.

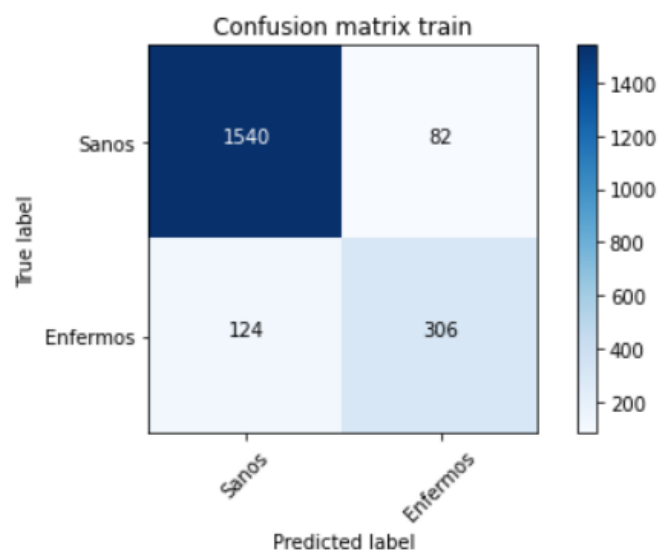


Figura 6. Matriz de confusión iteración 3.

	precision	recall	f1-score
Sanos	0.93	0.95	0.94
Enfermos	0.79	0.71	0.75
accuracy			0.90
macro avg	0.86	0.83	0.84
weighted avg	0.90	0.90	0.90

Figura 5. Métricas de desempeño iteración 2.

al comparar con las anteriores iteraciones, podemos darnos cuenta que es la que mejores resultados nos arroja, donde tenemos un tasa de FP=28% , FN=5%, TP=95% y TN=72%, en comparación con las iteraciones anteriores, este modelo nos muestra resultados muy buenos, ya que nos permite clasificar con una precisión del 93% sujetos sanos y una del 79% sujetos enfermos, al final por presentar las mejores métricas, este es el modelo elegido para solucionar el problema de clasificación.

4. Retos y consideraciones de despliegue

En general todo el proyecto representó un reto que consistió de varias etapas, desde aprender las herramientas necesarias para el machine learning de manera autodidacta, hasta finalmente desarrollar un modelo predictivo que cumpliera con los requerimientos mínimos de desempeño. En el pregrado de Bioingeniería se ven unas nociones básicas de machine learning, donde se trabaja con datos extraídos de Kaggle; lo anterior, hizo que fuera más sencillo el poder escoger un set de datos que se acomodara a un proyecto que fuese relacionado con nuestra área.

5. Conclusiones

- 5.1. El machine learning y la inteligencia artificial son las herramientas del futuro, a través de estas se pueden realizar infinidad de proyectos que van desde predecir subidas o bajadas en la bolsa, hasta predecir si una persona padece o no una enfermedad. La versatilidad que brindan estos programas se debe a la gran cantidad de formas de realizarlos y también a la infinidad de parámetros que se pueden variar para obtener el modelo deseado según sea cada caso. Además, cada vez

estas herramientas son más accesibles a todas las personas por medio del internet, así como también cursos de acceso libre para aprender a utilizarlas.

- 5.2.** Si bien tenemos una gran herramienta, de nada nos sirve esta si no tenemos datasets con características útiles para poder entrenarlo, es de vital importancia tener los datos organizados, tener conocimiento del tema o estar apoyado de alguien con conocimiento del proyecto para poder organizar los datos de una mejor manera, muchos parámetros no es siempre lo más importante, muchas veces la mayoría resultan ser basura y nos llevan a realizar cambios al modelo sin esto requerirse.