

Prueba Final

Supongamos que usted trabaja en el servicio de salud y recibe muestras que provienen de mujeres con cáncer de mama.

Los médicos han extraído características y las han anotado, su trabajo es crear un modelo que sea capaz de identificar si un paciente tiene o no cáncer.

Recordemos que un falso positivo no es tan preocupante como un falso negativo, ya que en el futuro se le hacen más pruebas a las pacientes y hay oportunidades de descubrir que estábamos en un error.

Sin embargo, un falso negativo puede llevar a que el cáncer se desarrolle sin supervisión durante más tiempo del necesario y podría llevar a daños más graves o incluso la muerte de la paciente.

Teniendo esto en cuenta, desarrolla un modelo que funcione lo mejor posible y explica qué decisiones has tomado en su elaboración y por qué.

A entregar obligatoriamente:

- Link a un repositorio público de Github que contenga al menos:
 - Un archivo Jupyter Notebook con todas las celdas ejecutadas en orden. Es decir, que antes de subir el archivo a github habéis limpiado el notebook y luego lo habéis ejecutado desde el principio).
 - Un archivo Readme en el que se explica el proyecto y el ejercicio. Tened en cuenta que este repositorio puede servir como CV en el futuro y que los recruiters suelen mirar los archivos Readme.md
 - Una carpeta data con el dataset.

En el notebook debe aparecer el proceso de preprocesado de datos desde los archivos originales a ser posible.

En el notebook debéis probar al menos con 3 modelos, evaluarlos y decidir cual es el mejor, justificando la respuesta en base a las matrices de confusión que aparecen al evaluar el error en training y en test.

El dataset y su descripción aparecen aquí:

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>