

Laboratorio 1: Métricas de Desempeño

Inteligencia Artificial - CC3085

José Antonio Mérida Castejón

22 de enero de 2026

Task 1: El Dilema del Negocio

Imagine que usted ha sido contratado como Lead Data Scientist para tres startups diferentes. Para cada caso, responda las preguntas justificando su respuesta.

Caso A: “MediScan AI”

Están desarrollando un modelo para detectar un tipo de cáncer raro en etapas tempranas a partir de radiografías.

1. *En este contexto, ¿qué es peor: un Falso Positivo o Falso Negativo?*

En este caso, un **falso negativo** es significativamente más peligroso (o peor). En el caso de un **falso negativo**, el paciente abandona completamente la posibilidad de tener la enfermedad y se elimina cualquier posibilidad de recibir tratamiento oportuno. Mientras que en un **falso positivo**, el paciente continúa con el monitoreo y exámenes necesarios a pesar de no tener la enfermedad. Una opción genera gastos monetarios en exámenes (o tratamientos), mientras que el otro resulta siendo potencialmente mortal.

En mi opinión, el modelo debería de tener un *threshold* sumamente bajo para indicar una potencial presencia de la enfermedad. La información proveída por el modelo podría servir como un tipo de *screening* temprano, seguido de un régimen establecido de exámenes para continuar con el monitoreo del paciente. En caso que la tecnología se vea limitada las únicas opciones fueran *tratamiento o no tratamiento*, seguiría considerando peor un **falso negativo** aunque se deberían de examinar los efectos secundarios o demás consecuencias del tratamiento.

2. *Basado en lo anterior, si tuviera que optimizar el modelo priorizando una sola métrica entre Precisión (Precision) y Sensibilidad (Recall / Sensitivity), ¿cuál escogería y por qué?*

Definitivamente sería **recall**. Utilizando precisión, el modelo se estaría enfocando en minimizar los **falsos positivos**. Por otro lado, el **recall** busca minimizar los **falsos negativos**. En otras palabras, utilizando **recall** estamos entrenando al modelo para identificar la mayor cantidad de casos de cáncer correctamente sin importar que tan seguro esté, mientras que utilizando **precision** lo estamos entrenando a priorizar la certeza absoluta antes de indicar cáncer. Esto sería como un doctor que observa ciertas indicaciones de la enfermedad, dónde priorizar **precision** incentiva al doctor a ignorar los indicadores y no informar al paciente (negligencia, si me preguntan mi opinión).

3. ¿Por qué el Accuracy sería una métrica peligrosa para presentar a los inversionistas en este caso específico?

Primero que nada, el **accuracy** realmente no cuenta la historia completa. Para cualquier modelo que estemos diseñando, es clave tomar múltiples métricas en cuenta a la hora de analizar su rendimiento y utilidad en el mundo real. Esto cubre las métricas de **recall** y **precision** discutidas anteriormente, dónde buscamos analizar más allá y descubrir un poco más sobre que tipos de errores comete el modelo. Luego, podemos considerar que el **accuracy** de por si es sumamente inapropiado debido a la rareza de la enfermedad que buscamos predecir. Si tuviésemos 1 radiografía (o paciente) de cada 20 presentando esta enfermedad, podríamos obtener un **accuracy** del 95 % simplemente prediciendo que ninguno de los pacientes está enfermo. Por último, como discutimos anteriormente, tenemos una situación de la vida real dónde las consecuencias de cada tipo de error son completamente diferentes. El presentar esta métrica a los accionistas sería un error total, ya que no representa adecuadamente el rendimiento del modelo *respecto al problema que se busca resolver*.

Caso B: “SpamGuard“ (Filtro de Correos)

Están creando un filtro de spam para una corporación grande.

1. ¿Qué error causaría más molestia y pérdida de productividad a los empleados: que un correo de spam llegue al inbox (FP o FN dependiendo de su definición) o que un correo importante de un cliente se vaya a la carpeta de spam?

En este caso, un correo de spam llegando al inbox sería una inconveniencia ligera para el trabajador que lo reciba. Mientras tanto, si un correo importante va a la carpeta de spam tenemos una situación catastrófica. El no ver el correo puede tener consecuencias dentro de la empresa, económicas o hasta legales para la persona que lo reciba. Adicionalmente, esta incertidumbre llevaría a que el resto de empleados estuviesen revisando la carpeta de spam constantemente. Como resultante, existiría una pérdida de productividad significativa, dónde el modelo contribuye al problema en vez de resolverlo.

2. ¿Qué métrica priorizaría aquí: **Precision** o **Recall**? (*Defina cual es su clase positiva*)