**¿Por qué importa A/B Testing?**

A/B Testing es una herramienta fundamental para tomar decisiones basadas en evidencia. En lugar de asumir que un cambio en un modelo, producto o estrategia será mejor, lo probamos directamente con datos. Esto evita implementar cambios que podrían no aportar mejoras o incluso empeorar el rendimiento.

En el contexto de modelos predictivos, el A/B testing permite comparar modelos y verificar si realmente hay una mejora estadísticamente significativa en métricas como accuracy, F1, RMSE o AUC. Sin A/B testing, podríamos estar implementando modelos que solo parecen mejores por casualidad o por overfitting.

**2. ¿Qué tipos de A/B Testing hay?**

Existen diferentes tipos de A/B testing, dependiendo de qué se quiere probar y cómo se diseña el experimento.

* El **A/B testing clásico** compara dos variantes (modelo A contra modelo B). Se mide una métrica objetivo en ambos grupos y se prueba si la diferencia es significativa.
* El **A/B/n** extiende esto a más de dos variantes. Por ejemplo, podrías comparar tres modelos distintos para ver cuál es mejor.
* El **multivariante** no compara variantes completas, sino combinaciones de elementos. Por ejemplo, se prueban distintas combinaciones de parámetros o técnicas para entender no solo cuál funciona mejor, sino **por qué**.
* El **testing secuencial** permite analizar los resultados a medida que llegan los datos. Esto puede acelerar decisiones si una variante se vuelve claramente superior desde el inicio. Aquí entra el enfoque SPRT (Sequential Probability Ratio Test), que permite detener un experimento si hay suficiente evidencia.
* La **inferencia causal** se usa cuando no puedes hacer una asignación completamente aleatoria (como en A/B clásico), pero quieres estimar el efecto de un modelo, tratamiento o política, controlando variables que podrían sesgar el análisis.
* El **multi-armed bandit** reemplaza el A/B tradicional con un enfoque adaptativo. En lugar de asignar tráfico de manera fija, el sistema va ajustando las probabilidades de asignación según qué opción va funcionando mejor, maximizando el beneficio mientras aprende.

**3. ¿Qué enfoques existen para analizar A/B testing?**

Hay dos enfoques principales: el **frecuentista** y el **bayesiano**.

* El enfoque **frecuentista** formula una hipótesis nula ("no hay diferencia entre A y B") y calcula un p-valor: la probabilidad de observar una diferencia como la que se ha medido si la hipótesis nula fuera cierta. Si el p-valor es menor a un umbral (por ejemplo, 0.05), se rechaza la hipótesis nula. Este enfoque es el más tradicional y el más usado.
* El enfoque **bayesiano** parte de una creencia previa y actualiza esa creencia con los datos observados. En lugar de hablar de p-valores, calcula directamente la probabilidad de que una variante sea mejor que otra. Por ejemplo, podrías decir que "hay un 85% de probabilidad de que el modelo A sea mejor que el modelo B".

Ambos enfoques son válidos, pero tienen ventajas distintas. El bayesiano es más intuitivo para tomar decisiones prácticas, mientras que el frecuentista es más conservador y tradicionalmente aceptado.

**4. ¿Cómo se implementan?**

Para aplicar A/B testing, debes:

1. Definir una **métrica de éxito** clara (por ejemplo, accuracy, AUC, tiempo de respuesta).
2. Calcular el **tamaño mínimo de muestra** necesario para detectar diferencias significativas (esto se puede hacer con herramientas online).
3. Asegurar una **asignación aleatoria** entre grupos.
4. Ejecutar el experimento sin detenerlo antes de tiempo, a menos que uses un enfoque secuencial.
5. Analizar los resultados con un enfoque estadístico claro (frecuentista o bayesiano).

Para modelos de machine learning, el A/B testing puede aplicarse en un entorno simulado o en producción. En simulación, se puede usar un conjunto de validación o test para dividir entre modelos A y B, y comparar métricas. En producción, puedes redirigir tráfico de usuarios reales o peticiones de predicción a los distintos modelos.