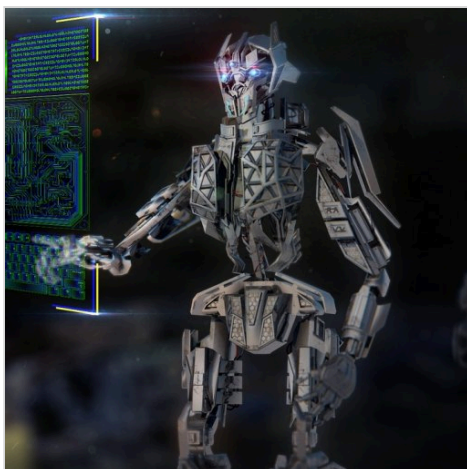


L U M B R E



Evaluación de recomendadores

Grado: 7

Materia: Sistemas de Recomendación · Inteligencia Artificial

11/10/2026

Antes de empezar

Juzgar un recomendador no es solo acertar la nota: es colocar los buenos ítems arriba, abarcar catálogo y no vivir de un puñado de éxitos. Las métricas de ranking, cobertura y diversidad capturan facetas distintas.

QUÉ APRENDERÁS

- Medir un sistema de recomendación con precisión en top-k, cobertura y diversidad, más allá del error de valoración.

Evaluación de recomendadores

Acierto no basta

Un recomendador que siempre sugiere las diez películas más famosas falla: acierta a veces pero aburre y oculta el catálogo. Evaluar bien exige mirar ranking, cobertura y variedad a la vez.

VIDEO

Métricas para evaluar sistemas de recomendación

Vídeo en español que mide un recomendador por ranking, cobertura y más allá del error de nota.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=ZgBDZ8c1r2Q>

Del error al ranking

Predecir notas con RMSE es útil, pero al usuario le importa el orden: qué aparece arriba. Las métricas de lista premian colocar lo relevante en las primeras posiciones.

Precisión y exhaustividad en top-k

- Precision@k: de lo recomendado, cuántos interesan de verdad.

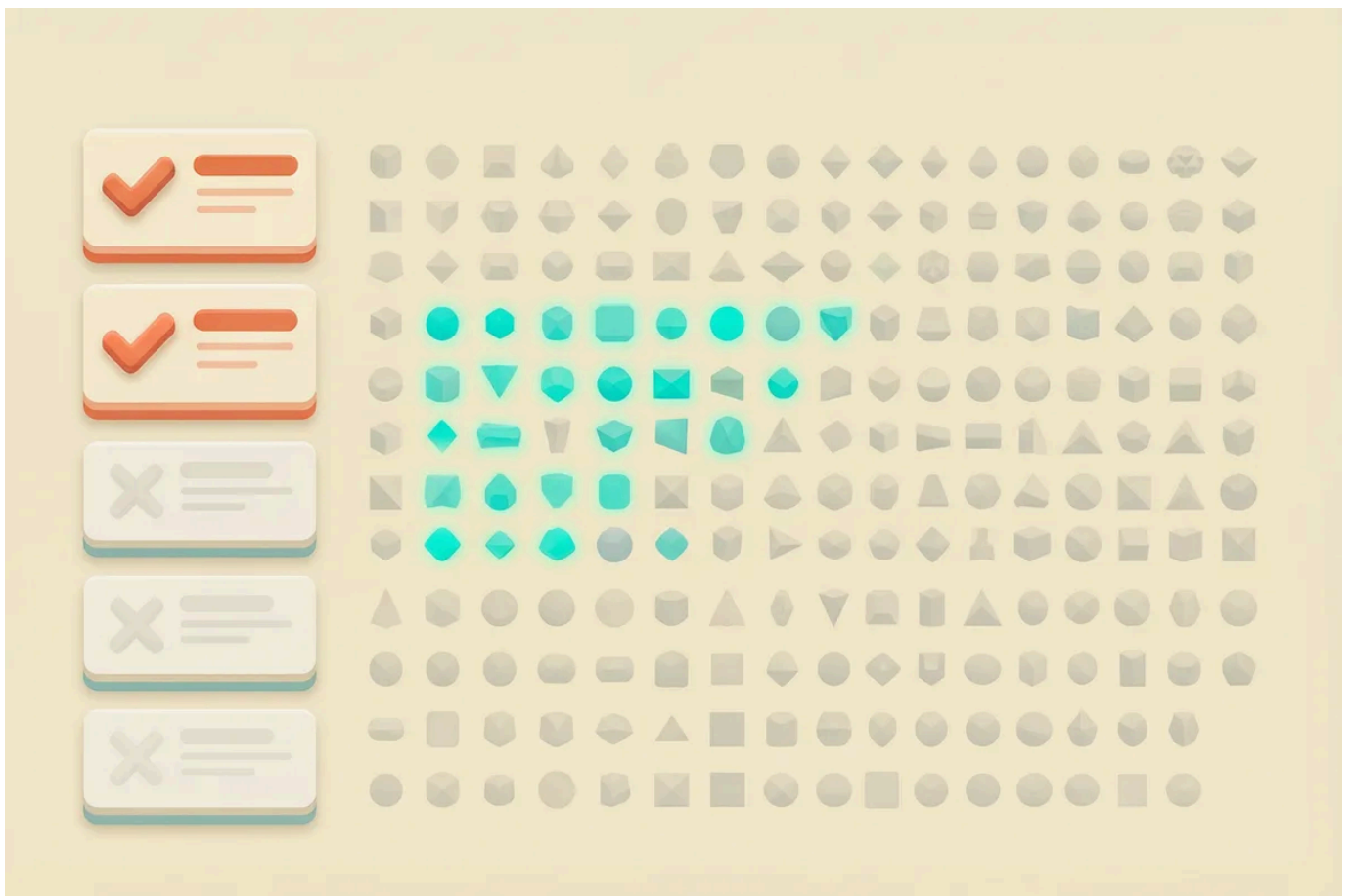
- Recall@k: de lo que interesaba, cuántos recuperaste.
- Suelen tirarse de la mano: más lista sube exhaustividad y baja precisión.

F1, media armónica

F1 combina precisión y exhaustividad en un solo número equilibrado. Evita presumir de una de las dos aisladas.

Ranking ponderado por posición

La métrica MRR y nDCG valoran dónde caen los aciertos: cuanto más arriba, mejor. No es lo mismo el primer puesto que el décimo.



Evaluar un recomendador es mirar a la vez el orden de la lista y cuánto catálogo realmente se usa.

Cobertura del catálogo

Qué fracción de ítems aparecen en las recomendaciones. Un sistema de baja cobertura solo recicla éxitos y entierra la cola larga.

Diversidad intra-lista

¿Es variada la lista de cada usuario? Recomendar diez películas idénticas acierta poco y aburre mucho. La diversidad mide el contraste entre los elementos sugeridos.

Serendipia y novedad

Sorprender con algo relevante pero inesperado añade valor. Medir novedad y sorpresa captura esa capacidad de descubrir, no solo de confirmar.

El precio del clic

Optimizar solo el clic a corto plazo empuja al sensacionalismo y la burbuja. Conviene medir satisfacción y bienestar, no solo el clic.

Validación offline frente a online

- Lo offline usa historiales, rápido pero sesgado a lo ya mostrado.
- Lo online prueba con usuarios reales mediante experimentación.
- Combinar ambos da una visión honesta del rendimiento.

El coste de la burbuja

Un recomendador que solo refuerza lo visto encierra al usuario. Diversidad y cobertura también son métricas de salud del sistema, no un lujo.

Errores frecuentes

- Medir solo el error de nota y celebrar listas aburridas que reciclan los éxitos más populares.
- Optimizar únicamente el clic a corto plazo y alimentar el sensacionalismo y la burbuja de filtros.
- Ignorar la cobertura y la diversidad, dejando enterrada toda la cola larga del catálogo.

Ejemplo resuelto: evaluar una lista de recomendaciones

Tienes un recomendador y quieres juzgarlo con honestidad, no solo por acierto:

1. Mides precisión y exhaustividad en top-k sobre un conjunto de prueba.
2. Ponderas la posición de los aciertos con nDCG o MRR.
3. Calculas qué fracción del catálogo aparece: la cobertura.
4. Evalúas el contraste dentro de cada lista: la diversidad.

5. Y confirmas en línea con un experimento real, no solo con el historial offline.

¿Para qué sirve en la realidad?

Un recomendador sano no solo acierta: coloca lo bueno arriba, abarca catálogo y sorprende. Medir ranking, cobertura, diversidad y satisfacción es lo que evita el engagement tóxico y la burbuja.

1. En la evaluación de un recomendador, la cobertura mide principalmente:

- ☐ lo rápido que responde el sistema
- ☐ qué fracción del catálogo llega a recomendarse
- ☐ la nota media asignada
- ☐ el número de usuarios registrados

2. Maximizar solo la exactitud de predicción de notas garantiza listas diversas y con buena cobertura.

- ☐ Verdadero
- ☐ Falso

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Evaluación



4. Ordena las facetas a vigilar al evaluar un recomendador:

Elementos:

- medir acierto básico con precision@k
- valorar si lo bueno sale arriba con nDCG
- comprobar que se abarca el catálogo
- exigir listas variadas, no repetitivas

Acierto	. . .
Posicion	. . .
Cobertura	. . .
Diversidad	. . .

Medidas de evaluación [https://en.wikipedia.org/wiki/Evaluation_measures_\(information_retrieval\)](https://en.wikipedia.org/wiki/Evaluation_measures_(information_retrieval)) Precisión, exhaustividad y métricas de ranking.

5. Une cada faceta de evaluación con la métrica que la mide.

5.	de lo recomendado, cuánto interesa de verdad	-----
5.	de lo relevante, cuánto recuperaste	-----
5.	dónde caen los aciertos en la lista	-----
5.	qué fracción del catálogo llega a recomendarse	-----

Opciones: precision · exhaustividad · ndcg · cobertura

Acierto no basta

Evaluación de recomendadores

Ranking

precision top-k

ndcg y mrr

Acierto

f1 armonica

recall

Salud

cobertura

diversidad

Extra

serendipia

online vs offline

El mapa mental interactivo es interactivo en la versión online.



nDCG https://en.wikipedia.org/wiki/Discounted_cumulative_gain La métrica que premia colocar los buenos ítems en las primeras posiciones.

6. Tu recomendador sube el clic pero muestra siempre lo mismo. ¿Qué dos métricas estarían descuidando y cómo lo arreglarías?

Referencia curricular

Grado:	7
Materia:	Sistemas de Recomendación · Inteligencia Artificial

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-recs-004-evaluacion-de-recomendadores>

Material educativo generado desde Lumbre · lumbre.lat