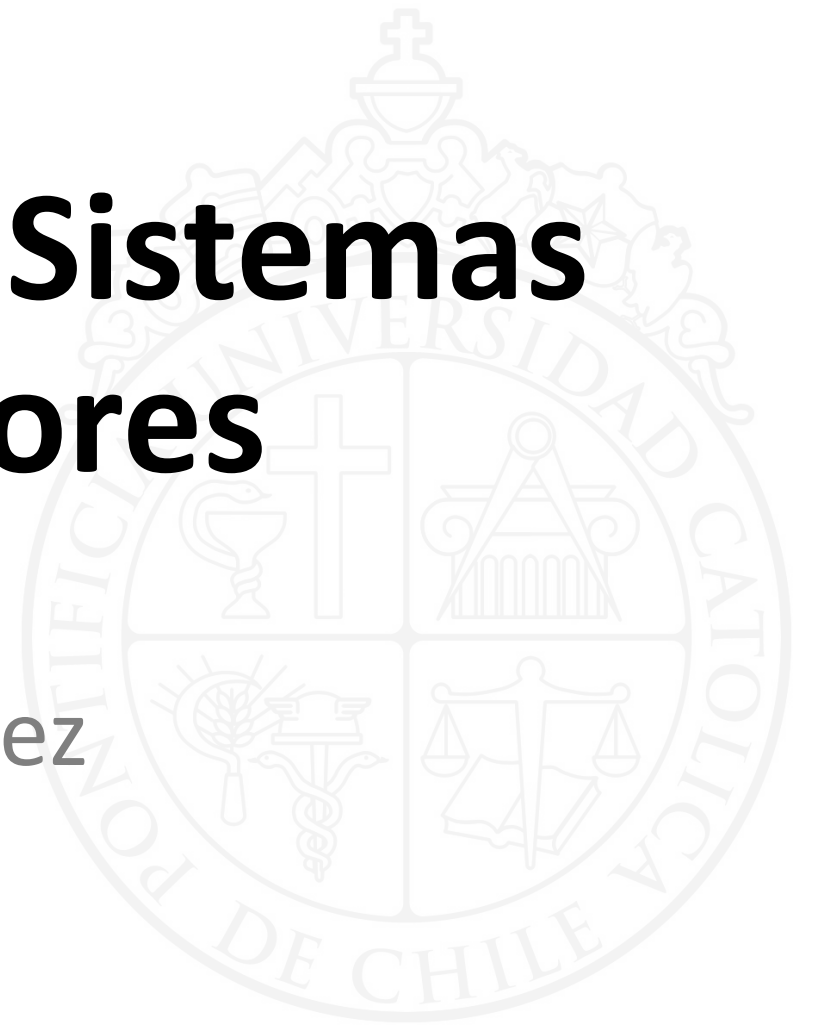


Introducción a los Sistemas Recomendadores

Vicente Domínguez

Magister de Inteligencia Artificial UC



Breve presentación



Equipo Docente

Vicente Domínguez (Profesor)

Ingeniero Civil de Computación PUC,

Magister en Ciencias de la Ingeniería PUC,

Instructor Adjunto del Departamento de Ciencia de la Computación,

Machine Learning Operations Engineer, Zerofox

[Perfil Google Scholar](#)

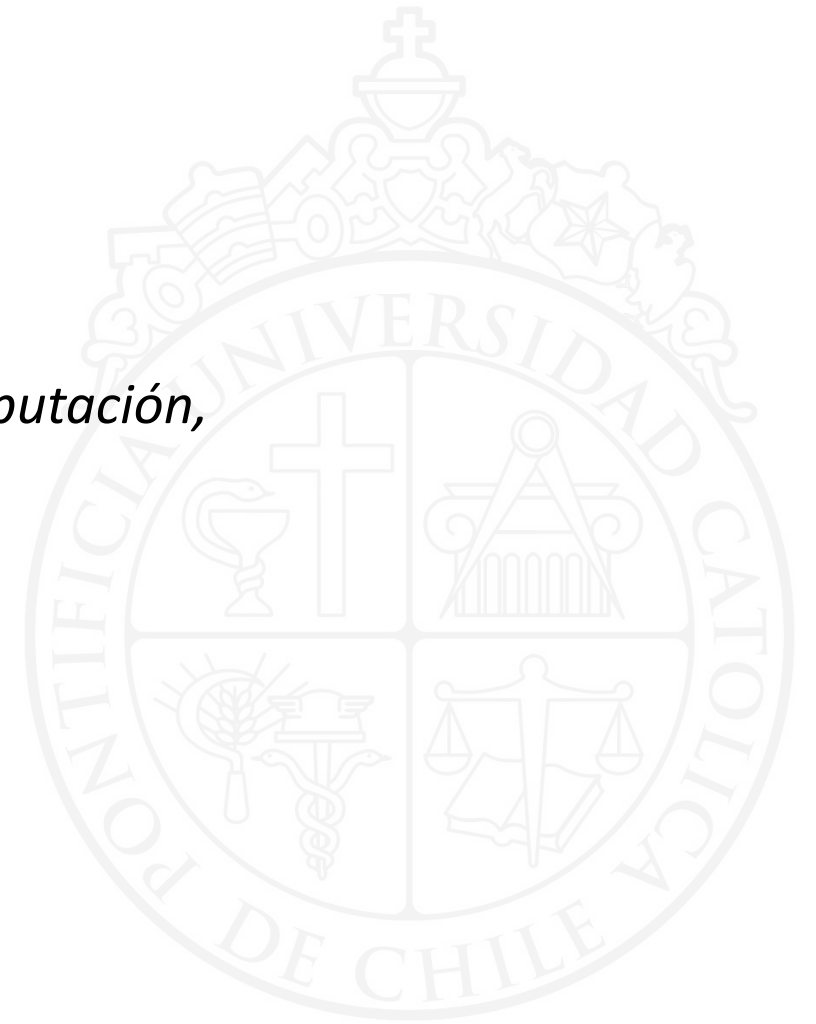
Michelle Madrid (Ayudante)

Estudiante de Ingeniería, Major Computación PUC

Desarrolladora de Software en Okane Capital

E-mails:

- vidominguez@uc.cl
- msmadrid@uc.cl

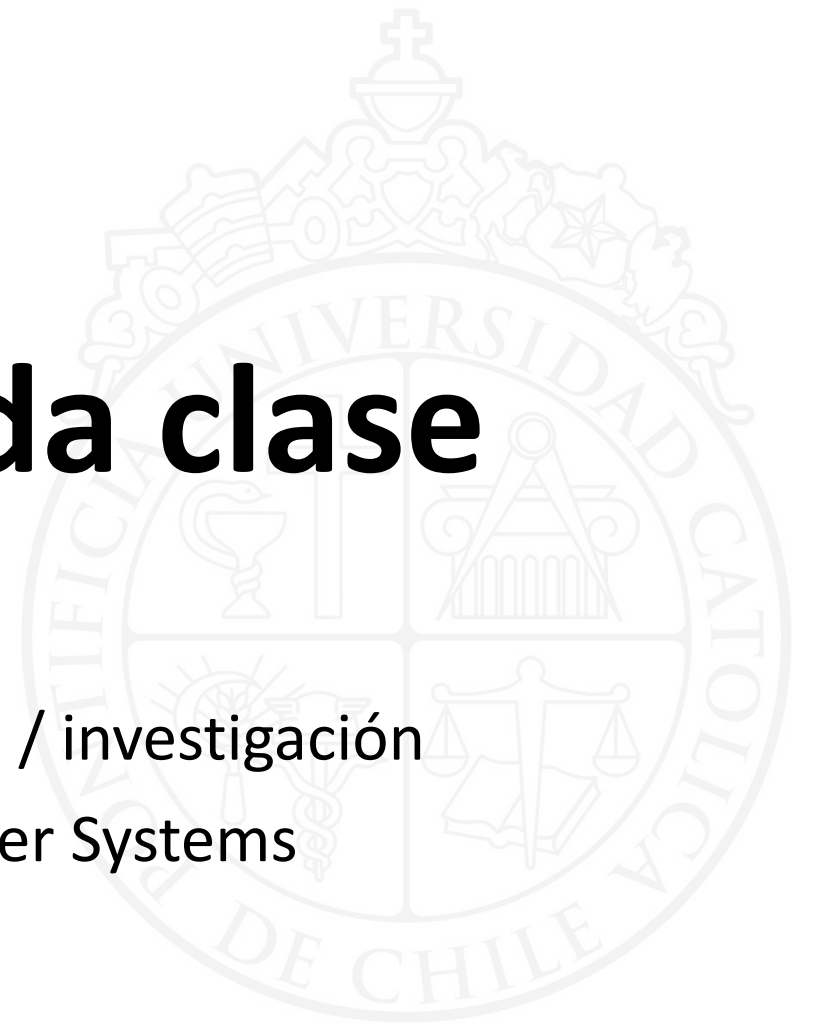


Estructura de cada clase

Materia - Teoría

Casos prácticos de uso en la industria / investigación

Práctico - Hands on Recommender Systems



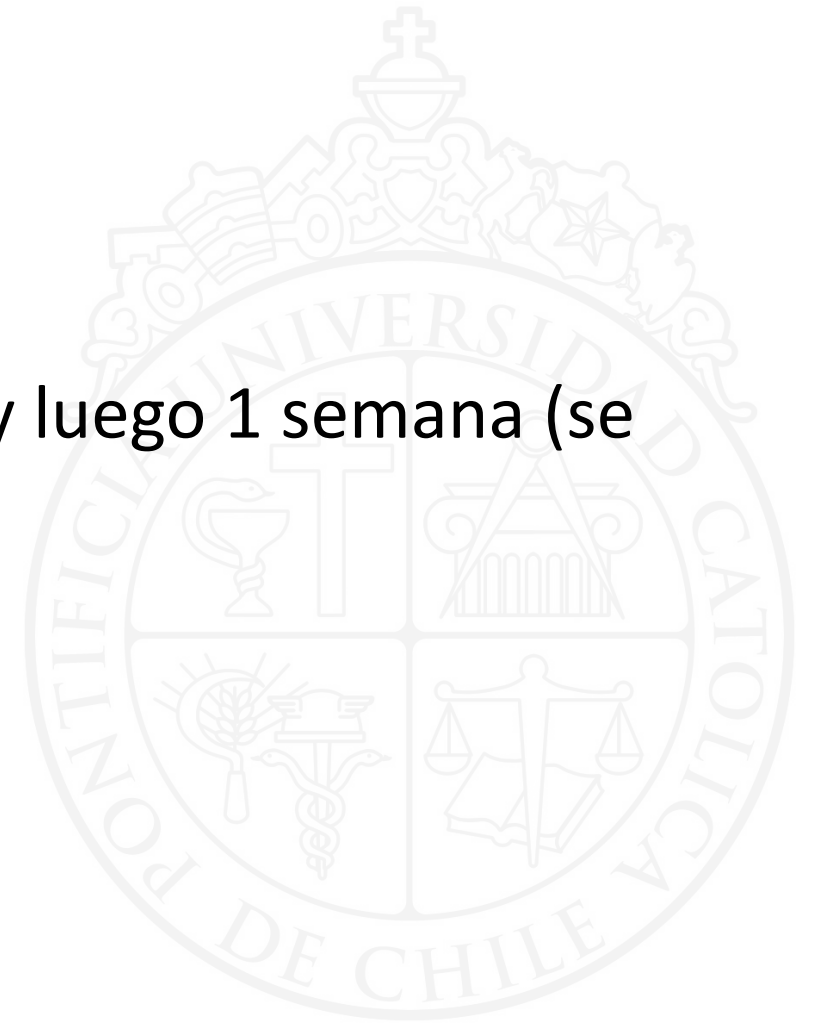
Programa del Curso

1. **Introducción , Recomendación No personalizada , Filtrado Colaborativo (User y Item), Evaluación basada en métricas de error (RMSE, MAE) + Práctico.**
2. Slope One , Recomendación basada en métodos latentes (Factorización Matricial) + **Práctico.**
3. Recomendación basada en feedback implícito , Optimización basada en ALS y BPR, Evaluación basada en métricas de ranking , novedad, serendipia + **Práctico.**
4. Recomendación basada en contenido (reglas de asociación, metadata y texto) + **Práctico.**
5. Recomendación basada en contexto , ensembles y recomendación híbrida. + **Práctico.**
6. Deep Learning para recomendación (texto , imágenes, multimodal, secuencial) + **Práctico.**
7. Reinforcement Learning (Bandits) + Conversational Recsys + **Práctico**
8. Repaso general , dominios de aplicación e investigación reciente , recomendación basada en grafos.
9. Cierre calificaciones, aplicaciones

Evaluación

Promedio de ejercicios prácticos.

Tiempo para realizarlos. Al final de la clase y luego 1 semana (se entrega en la siguiente clase en general)



En esta clase

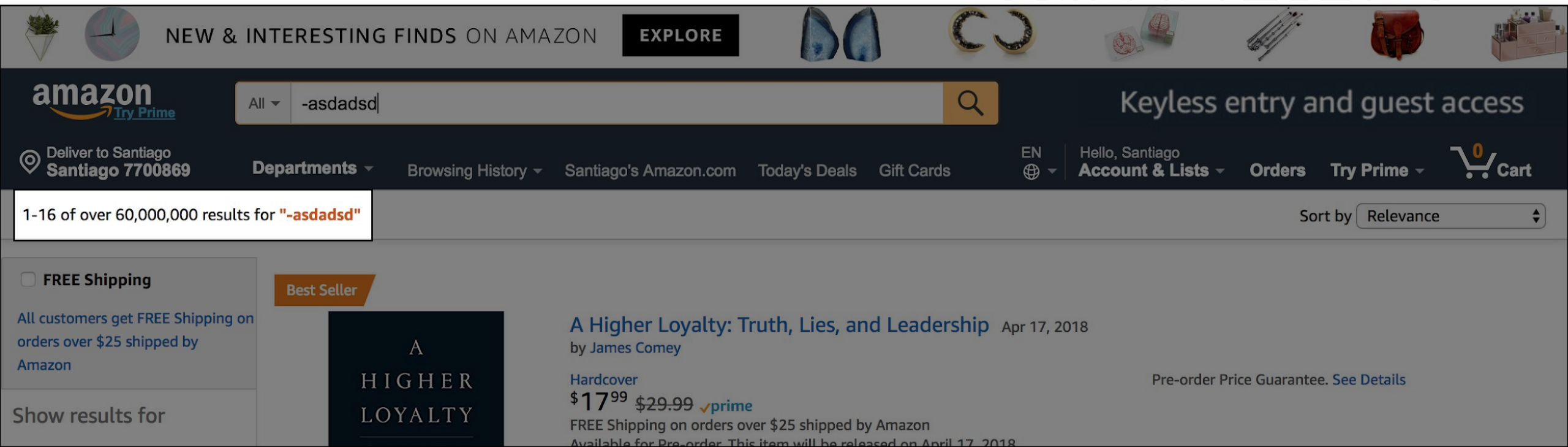
1. Introducción a los sistemas recomendadores.
2. Recomendaciones no personalizadas.
3. Filtrado colaborativo basado en usuarios.
4. Filtrado colaborativo basado en ítems.
5. Evaluación basada en métricas de error (RMSE, MAE).
6. Práctico.



Objetivos de sistemas recomendadores

- Ayudar al usuario con la sobrecarga de la información y que se ajuste a sus preferencias.
- Muchas alternativas , tiempo limitado para escoger.

Amazon: + de 60.000.000 de productos para escoger.



The screenshot displays the Amazon homepage with a search bar containing the text "-asdadsd". The search results show 1-16 of over 60,000,000 results. A product titled "A Higher Loyalty: Truth, Lies, and Leadership" by James Comey is featured, with a price of \$17.99 (reduced from \$29.99) and a Prime badge. The product is marked as a "Best Seller" and has a "Pre-order Price Guarantee" label. The Amazon logo and navigation links are visible at the top.

NEW & INTERESTING FINDS ON AMAZON EXPLORE

amazon Try Prime

All ▾ -asdadsd 🔍

Keyless entry and guest access

Deliver to Santiago Santiago 7700869

Departments ▾ Browsing History ▾ Santiago's Amazon.com Today's Deals Gift Cards

EN ▾ Hello, Santiago Account & Lists ▾ Orders Try Prime ▾ Cart

1-16 of over 60,000,000 results for "-asdadsd" Sort by Relevance ▾

☐ FREE Shipping

All customers get FREE Shipping on orders over \$25 shipped by Amazon

Show results for

Best Seller

A HIGHER LOYALTY

A Higher Loyalty: Truth, Lies, and Leadership Apr 17, 2018 by James Comey

Hardcover

\$17⁹⁹ \$29.99 ✓prime

FREE Shipping on orders over \$25 shipped by Amazon

Available for Pre-order. This item will be released on April 17, 2018

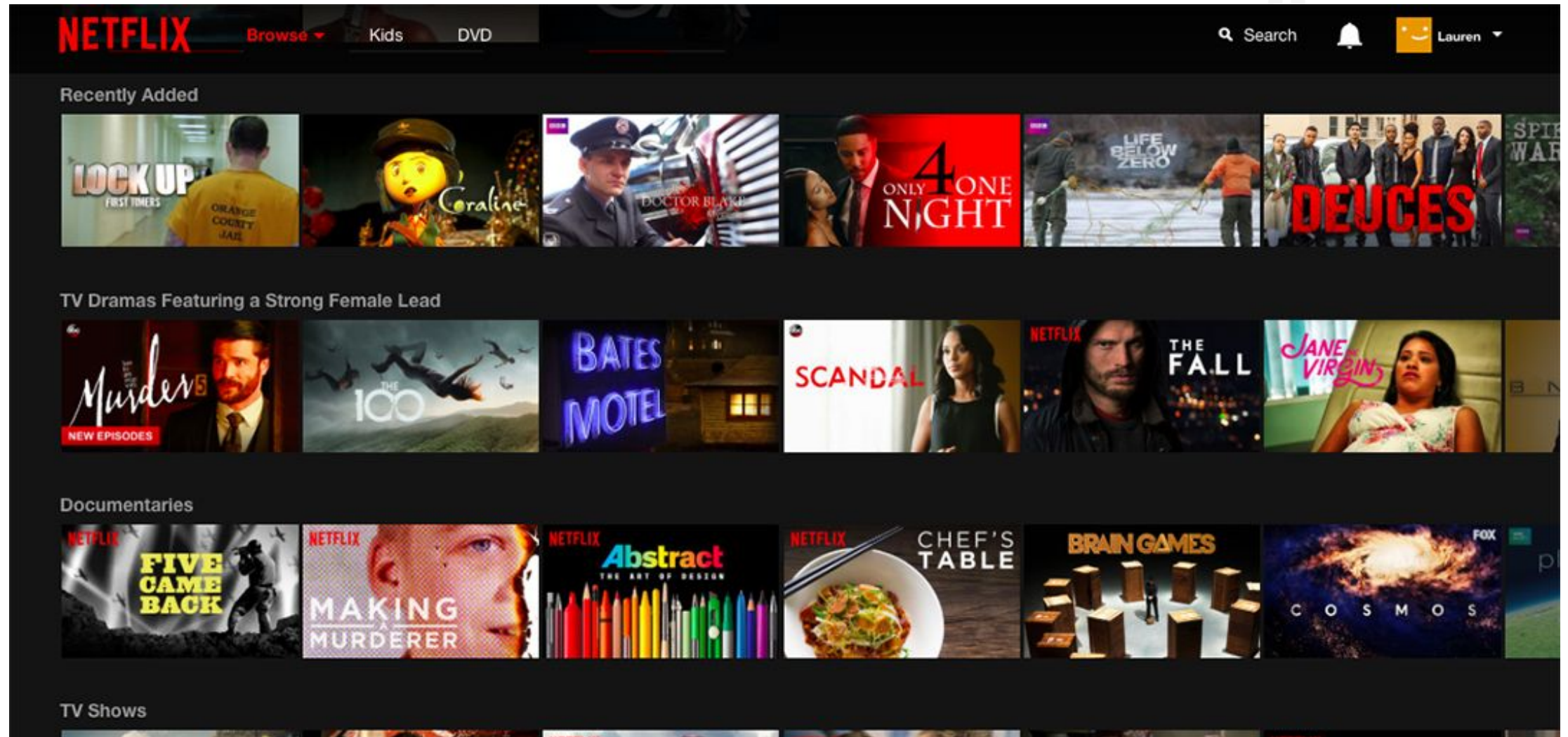
Pre-order Price Guarantee. See Details

Objetivos de sistemas recomendadores

- Aumentar KPI de la empresa o sistema
 - Mejorar el *engagement* del usuario.
 - Aumentar las ventas de la empresa.
 - Aumentar el ticket promedio de la orden.
 - Aumentar *click rate*.
 - *Posicionamiento orgánico de la empresa.*

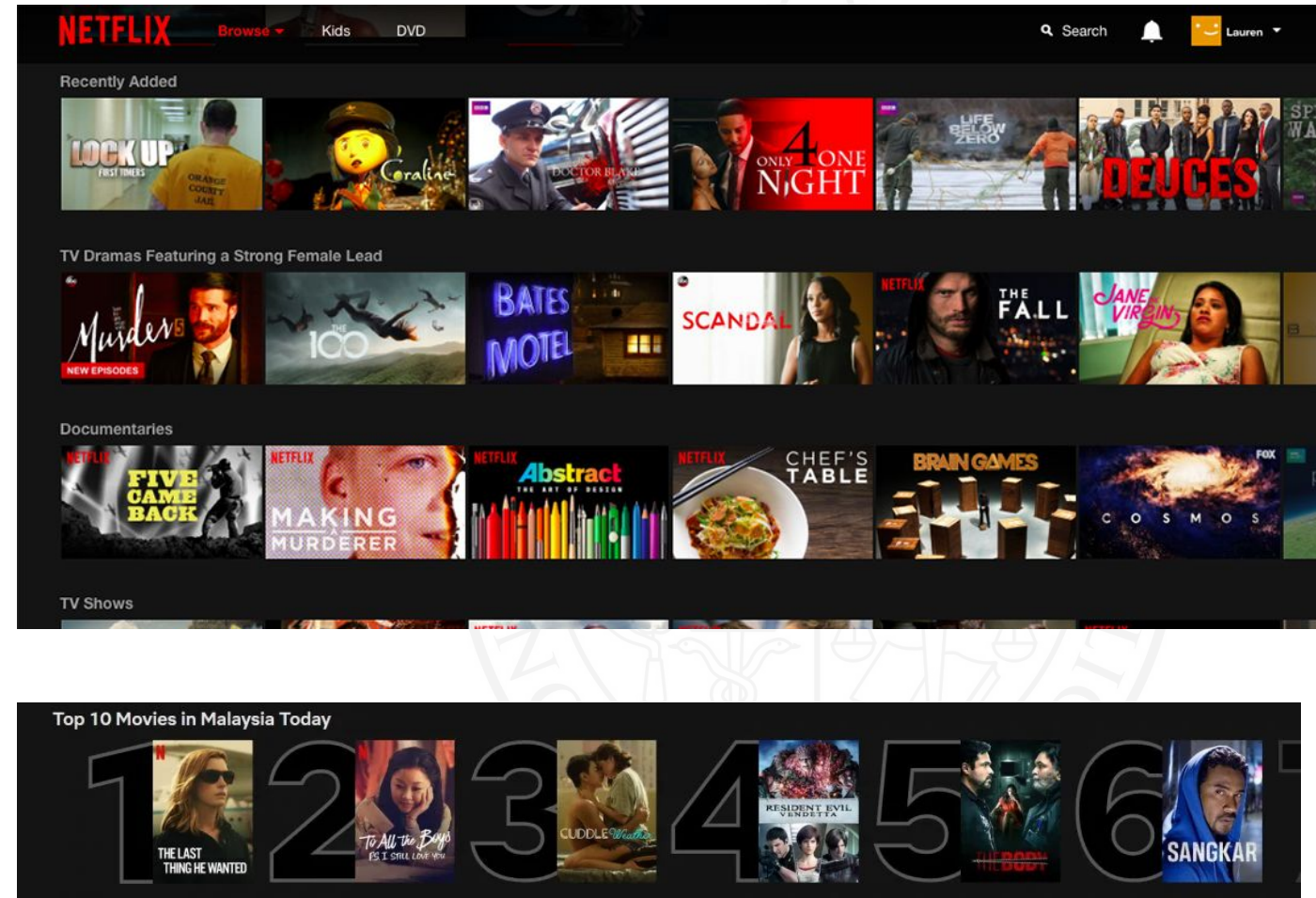


¿Dónde entran los algoritmos de recomendación en este caso?





¿Dónde entran los algoritmos de recomendación en este caso?

1. Ordenar los géneros de las películas dependiendo de mis intereses.
2. Dentro de cada género ordenar de ese género cuál es la que más me puede interesar.
3. Cambiar la carátula de la imagen dependiendo de mis intereses.
4. Más populares , más vistas en chile.



Si ya adquirí el producto ...

¿Por qué la recomendación en este ejemplo no es la adecuada?



Nuevo | +500 vendidos

Impresora a color multifunción HP LaserJet Pro M182nw con wifi blanca 220V - 240V

★★★★★ (34)

MÁS VENDIDO 12° en Impresoras HP

\$ 398.990
en 6x \$ 66.498 sin interés

[Ver los medios de pago](#)

Voltaje: 220V - 240V


220V - 240V

Lo que tienes que saber de este producto


- Imprime, escanea y hace copias.
- Impresión doble faz manual.
- Tecnología de impresión: láser.
- Posee pantalla LCD para visualizar el proceso.
- Tiene entrada USB.
- Capacidad máxima de 150 hojas.
- Soporta papel tamaño Carta, Legal, Ejecutivo, A4, A5, A6, B5 JIS, B6 JIS, Oficio, Tarjeta postal JIS, Tarjeta postal doble JIS, Sobre N10, Sobre monarca, Sobre B5, Sobre C5, Sobre DL, A5R, 16K, Personalizado (76 x 127 mm - 216 x 356 mm).
- Incluye accesorios.
- Práctica y funcional tanto para uso personal como profesional.

[Ver características](#)


Quienes vieron este producto también compraron



\$ 189.500
6x \$ 31.583 sin interés
Envío gratis
Impresora a color multifunción Epson EcoTank L3250 con w...



\$ 60.555
6x \$ 10.093 sin interés
Envío gratis
Impresora simple función Brother HL-1202 negra y...



\$ 130.990
6x \$ 21.832 sin interés
Envío gratis
Impresora a color multifunción HP Smart Tank 500 negra...



Jac Rayner
@GirlFromBlupo



Dear Amazon, I bought a toilet seat because I needed one. Necessity, not desire. I do not collect them. I am not a toilet seat addict. No matter how temptingly you email me, I'm not going to think, oh go on then, just one more toilet seat, I'll treat myself.



¿Dónde entran los algoritmos de recomendación para hacer un carro de compras en un sitio de e-commerce?

Categorías

¿Qué estás buscando?

Envío a Av. De La Pl. 1324

Desde hoy - 11:00 hrs. - \$2.990

Ingresar

1 - 16 de 310 productos

ver 16 productos por página

Ordenar por Destacados

Marca

Acuenta

4

Argos

1

Athletic Works

1

Autodrive

2


Black Bull

1

Precio

\$500 - \$489.990

\$500\$+245mil\$+489mil



Papas


Papa Malla, 2 Kg

\$2.850

-

1

+




Lay's

Papas Fritas Corte Americano, 380 g

\$2.290

12%


~~\$2.590~~



Marco Polo

Papas Fritas Rústicas Sal de Mar, 185 g

\$2.450





Lay's

Papas Fritas Corte Americano, 230 g

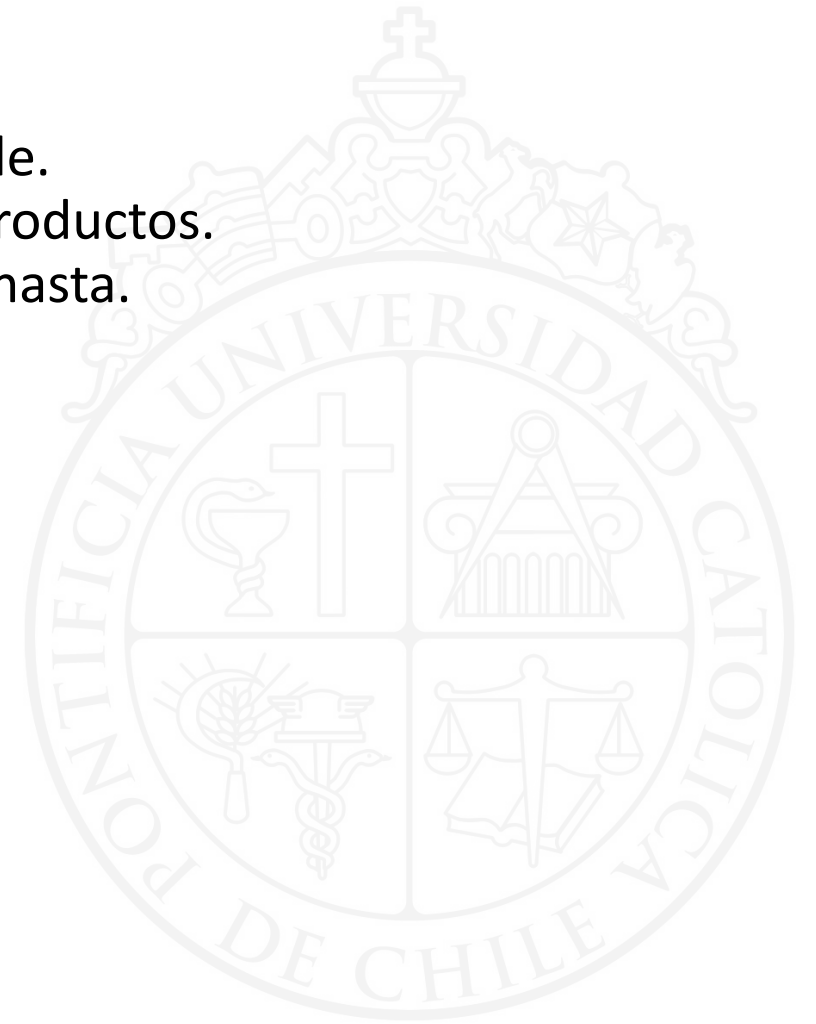
2 x \$3.160

Normal: \$2.050 c/u

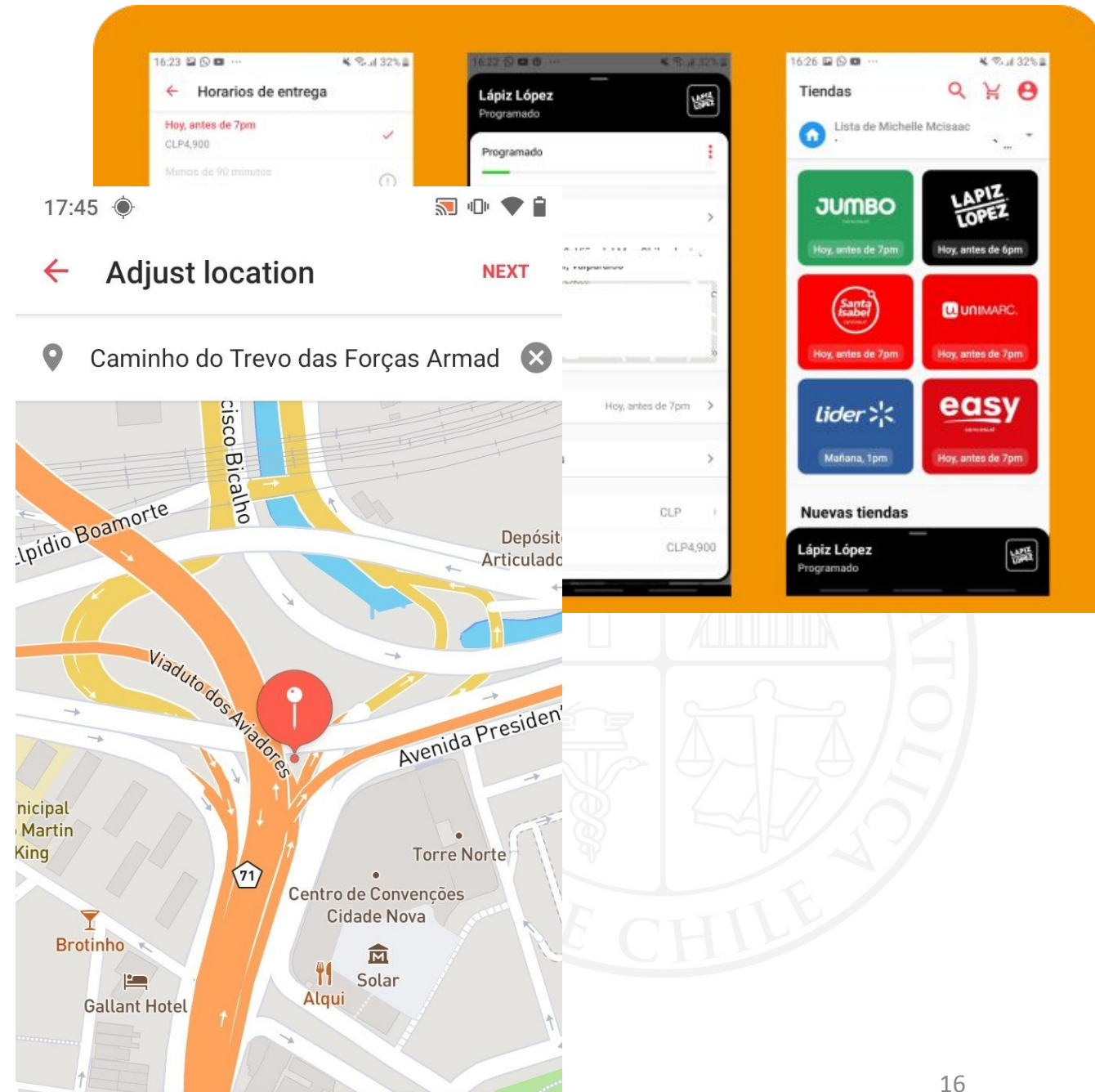
Carro (2)		✕
1	 <div> Papás, Papa Malla, 2 Kg \$2.850 </div>	\$2.850
Eliminar		<div> <div>−</div> <div>1</div> <div>+</div> </div>
1	 <div> Lider, Leche Semidescremada, 1 L \$895 </div>	\$895
Eliminar		<div> <div>−</div> <div>1</div> <div>+</div> </div>
<div> <div>🗑️</div> Vaciar Carro </div>		
Subtotal		\$3.745
Pagando con Tarjeta Lider		\$3.745
<div>Ir a pagar</div>		

Recomendación secuencial

1. Dada una canasta recomendar productos para agregarle.
2. Dadas canastas anteriores recomendar conjuntos de productos.
3. Recomendar el primer producto para agregar a una canasta.



¿Dónde entran los algoritmos de recomendación a la hora de hacer un pedido?



Recomendación contextual

Dado el contexto:

- ubicación
- tamaño del pedido
- productos del pedido
- medio de transporte del shopper
- pedidos del shopper
- clima
- distancia
-

Recomendar (asignar) un shopper/repartidor para que se encargue de mi pedido.



Problema de recomendación

- Seleccionar un conjunto de ítems para un usuario C que maximicen su utilidad.
- La función de utilidad genera una predicción de **ranking**

$$\forall c \in C, s'_c = \arg \max_{s \in S} u(c, s)$$

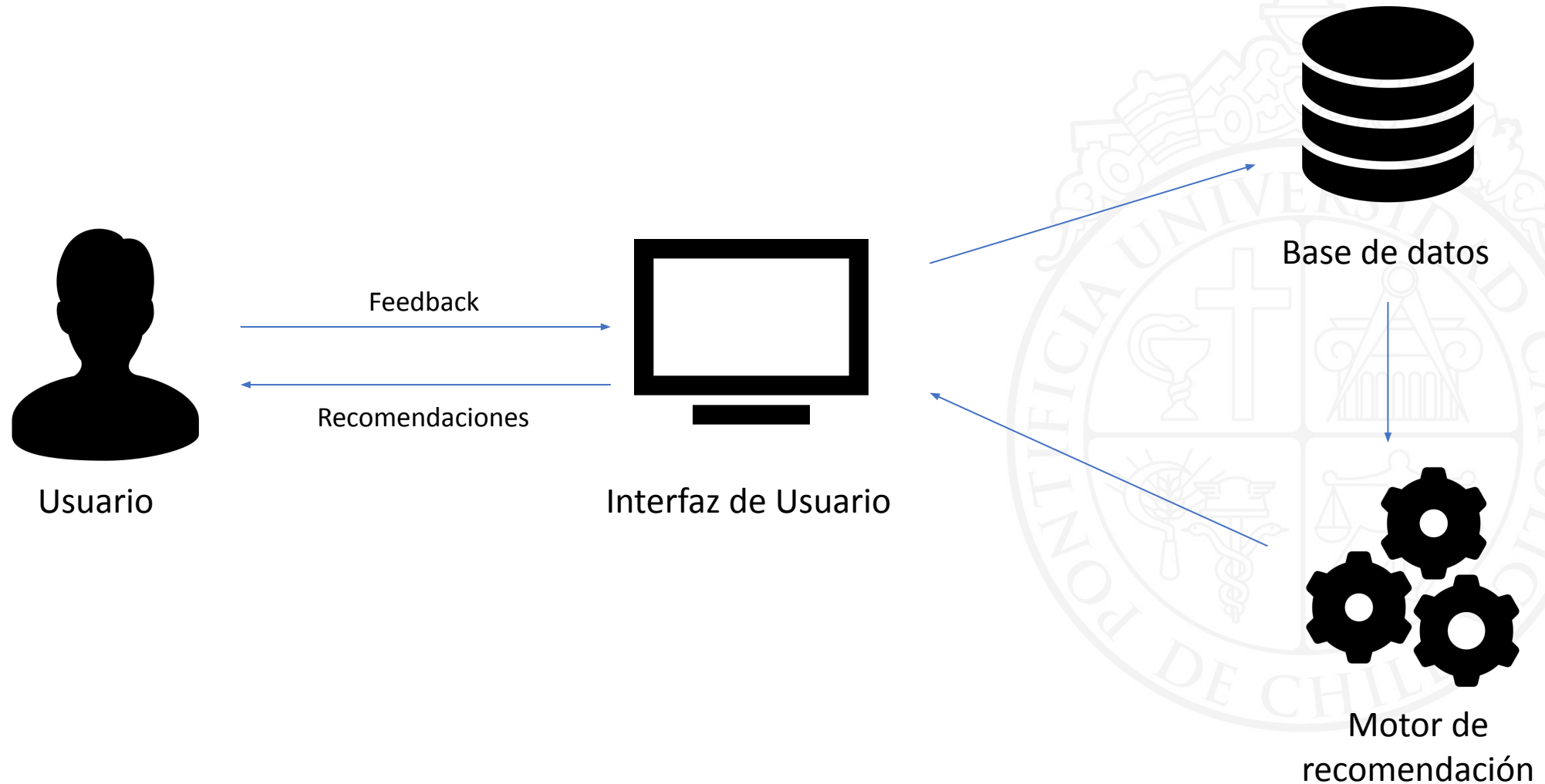
$u : C \times S \rightarrow R$, función de utilidad

R : conjunto de ítems recomendados

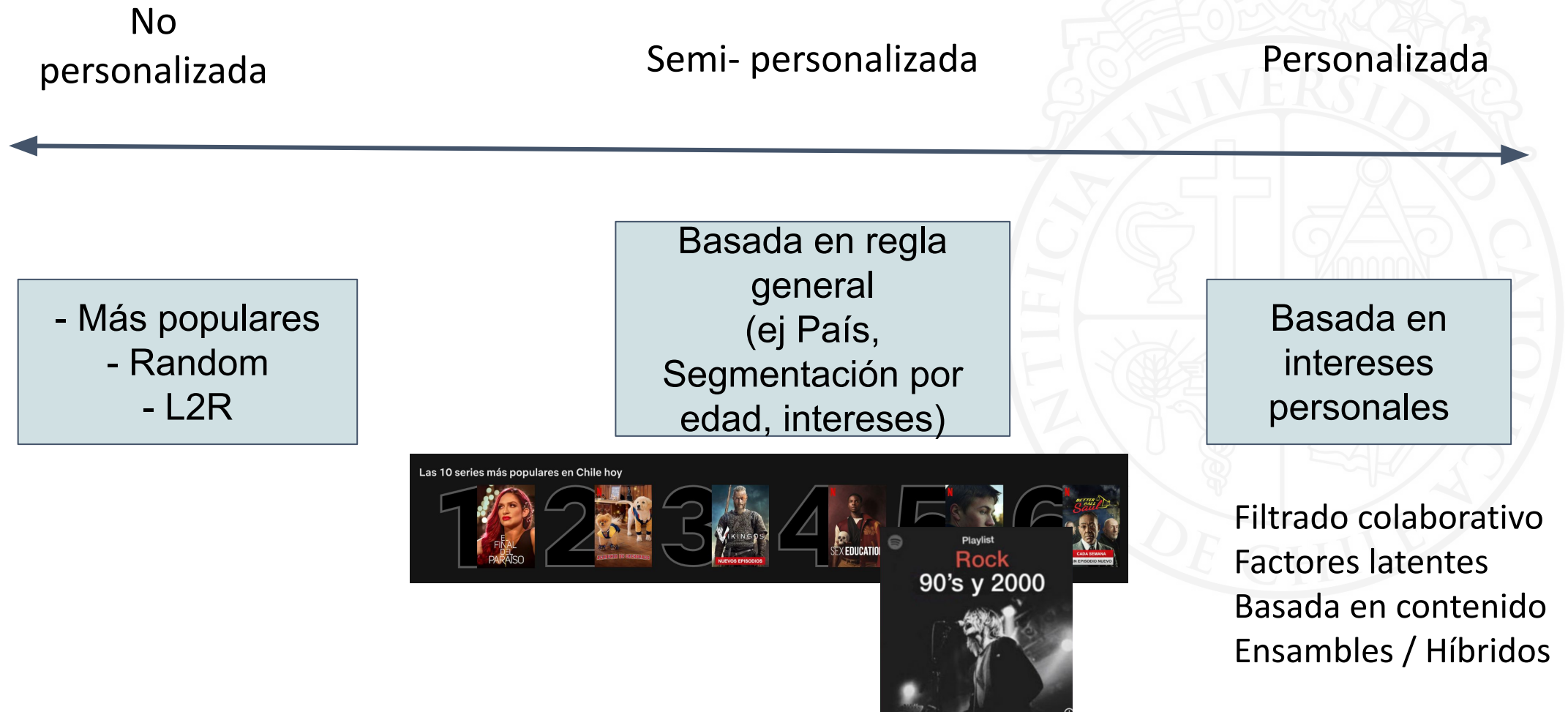
C : conjunto de usuarios

S : conjunto de ítems

Esquema de recomendación



Recomendación no personalizada, semi-personalizada y personalizada.



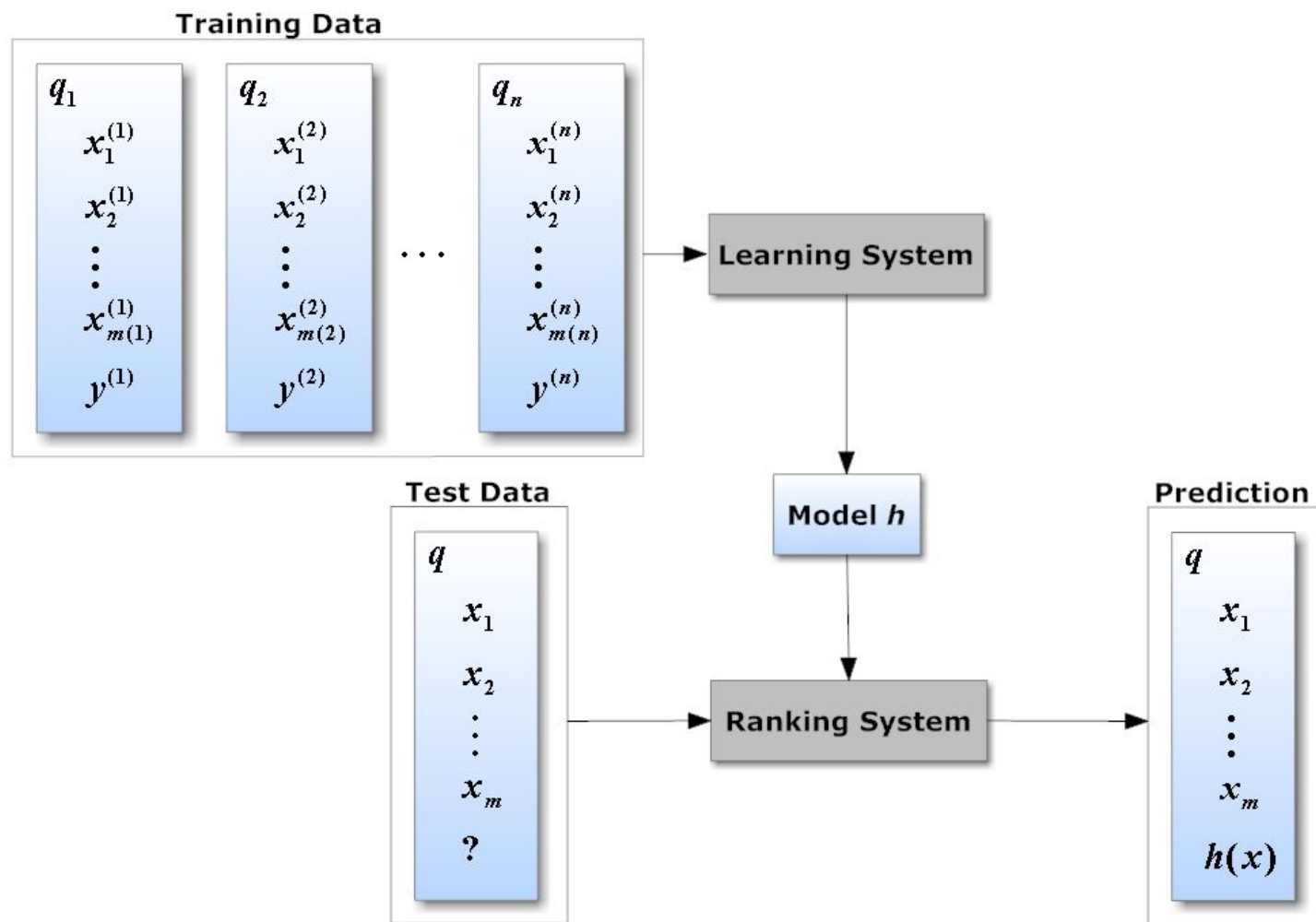
Recomendaciones no personalizadas

Recomendación que no toma información del usuario.

Learning to rank (no personalizado)

Learning to rank es una tarea dentro de Machine Learning que:

Dada una **query** y un conjunto de **items candidatos**, el objetivo es aprender a **ordenar** los que son más relevantes para la query.



Learning to rank

POINTWISE: Obtiene un score de relevancia de cada resultado dada una QUERY y los ordena de mayor a menor.

PAIRWISE: Compara relevancia en pares de items y finalmente los ordena por transitividad.
Ej. $A > B$ y $A < C$, entonces resultaría: C, A, B

LISTWISE: Compara relevancia de listas de un tamaño N de items.
Ej. $A > C > B$ y $C > D > F \rightarrow A, C, B, D, F$

Recomendación personalizada

- Considera información del usuario:
 - Ítems con los que ha interactuado históricamente (rating, compra/reproducción ,like, etc)
 - Información de su perfil
 - Información contextual (ubicación , día , hora, etc..)

Predicción de Ratings

- Los algoritmos que veremos el resto de esta clase simplifican el problema de recomendación considerándolo como un problema de predicción de ratings basado en coocurrencia :

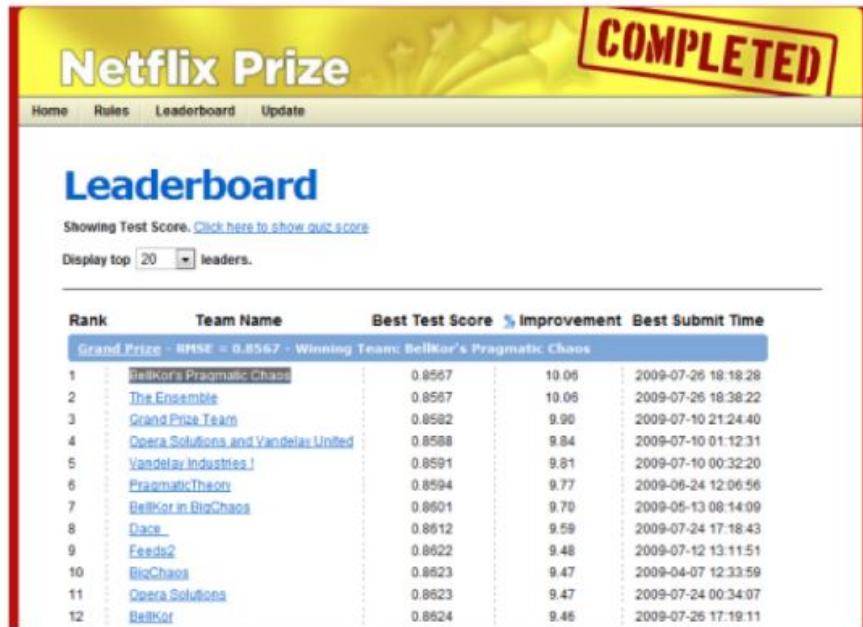
Predict!

	Item 1	Item 2	...	Item m
User 1	1	5		4
User 2	5	1		?
...				
User n	2	5		?

- ... y qué tan buena es mi predicción?

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{r}_{ui} - r_{ui})^2}{n}}$$

Ejemplo: Netflix Prize (2007 - 2009)



NetfliX Prize

Home Rules Leaderboard Update

Leaderboard

Showing Test Score. [Click here to show quiz score](#)

Display top 20 leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries I	0.8591	9.81	2009-07-10 00:32:20
6	Pragmatic Theory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11

¿En qué consistía?

Predecir ratings de Netflix (puerta a puerta) y disminuir en al menos un 10% el RMSE del modelo ya existente (MovieMatch)

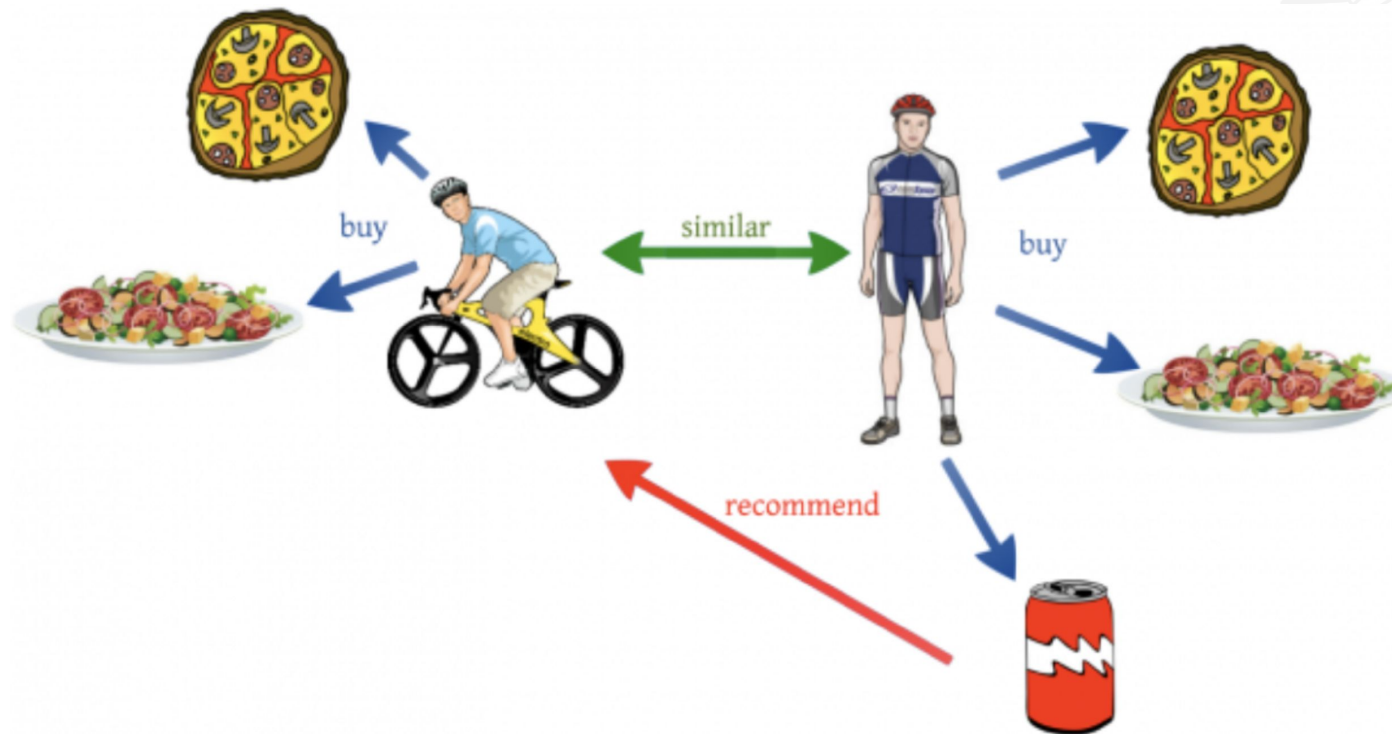
Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40



1 millón de
dólares
20 minutos
antes!!

Filtrado Colaborativo basado en Usuarios

- **Objetivo:** buscar a usuarios similares y recomendar usando una suma ponderada con una métrica de similitud



<https://www.slideshare.net/tantrieuf31/introduction-to-recommendation-systems>

Filtrado Colaborativo basado en Usuarios

- **Objetivo:** buscar a usuarios similares y recomendar usando una suma ponderada con una métrica de similaridad



Usuario A




Usuario B



Usuario Objetivo



Usuario C



Usuario D



Usuario E

Utilizando 5 vecinos cercanos

Similaridad entre usuarios

- **Similaridad entre usuarios:** se basa en los patrones de calificación normalizando por el promedio de cada usuario.
- Fórmula de similaridad coseno:

$$\text{sim}(u, v) = \frac{\sum_{i \in I} [(r(u, i) - \text{avg}(u)) \cdot (r(v, i) - \text{avg}(v))]}{\sqrt{\sum_{i \in I} (r(u, i) - \text{avg}(u))^2} \cdot \sqrt{\sum_{i \in I} (r(v, i) - \text{avg}(v))^2}}$$

- **$\text{sim}(u, v)$** : Similaridad entre usuarios u y v .
- **I** : Conjunto de ítems que ambos usuarios han calificado.
- **$r(u, i)$, $r(v, i)$** : Calificaciones del usuario u y v para el ítem i .
- **$\text{avg}(u)$, $\text{avg}(v)$** : Calificación promedio del usuario u y v .

Generación de recomendaciones

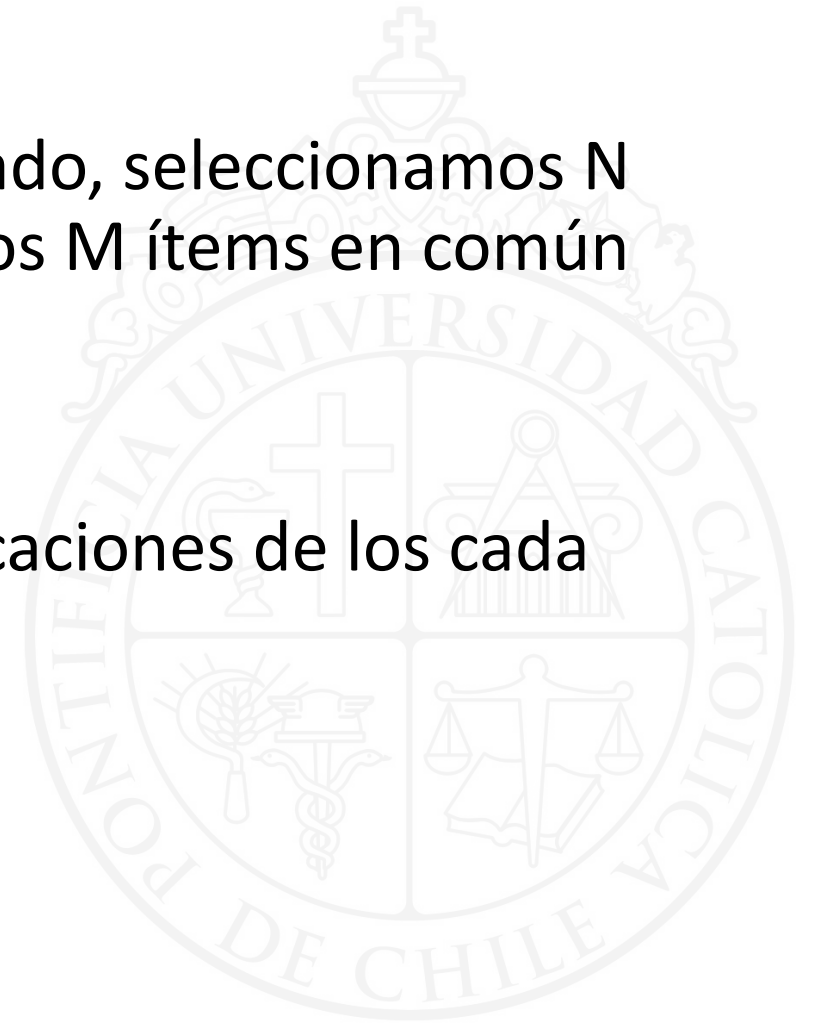
Objetivo: recomendar ítems que aún no han sido calificados por el usuario objetivo, pero que tienen calificaciones altas de usuarios similares que han calificado los mismos ítems.

$$P(u, i) = \text{avg}(u) + \frac{\sum_{v \in V} [\text{sim}(u, v) \cdot (r(v, i) - \text{avg}(v))]}{\sum_{v \in V} |\text{sim}(u, v)|}$$

- $P(u, i)$: rating predicho del ítem i para el usuario u .
- V : Conjunto de los N usuarios más similares a u que han calificado el ítem i
- i : Los ítems considerados son solo aquellos que han sido calificados por los usuarios en V y no por el usuario u .
- $\text{sim}(u, v)$: Similaridad entre usuarios u y v .
- $r(v, i)$: Calificación del usuario v para el ítem i .
- $\text{avg}(v)$, $\text{avg}(u)$: Calificación promedio del usuario v y del usuario u .

Proceso Completo para Todos los Usuarios

- Para cada par de usuario e ítem no calificado, seleccionamos N usuarios similares que comparten al menos M ítems en común con el usuario objetivo.
- Usamos la similaridad para predecir calificaciones de los cada usuario con los ítems.



Generando la Lista de Recomendaciones

- Ordenamos los ítems que el usuario aún no ha calificado según los ratings predichos de mayor a menor.
- Seleccionamos los primeros K ítems para recomendar al usuario.



¿Qué deficiencias puede tener el filtrado colaborativo basado en usuarios?



Debilidades

- Alto costo computacional de recorrer todas las combinaciones de usuarios e ítem.
- Cold start problem: nuevos usuarios y nuevos ítems me obligan a tener que recalcular todo.

Soluciones: limitar el número de usuarios similares considerados, técnicas de factorización de matrices para reducir la dimensionalidad.

Reflexión

¿Qué información captura el modelo de filtrado colaborativo basado en usuarios?

¿Cómo realiza las recomendaciones?

¿Cómo realiza las recomendaciones?

PSEUDO CÓDIGO.

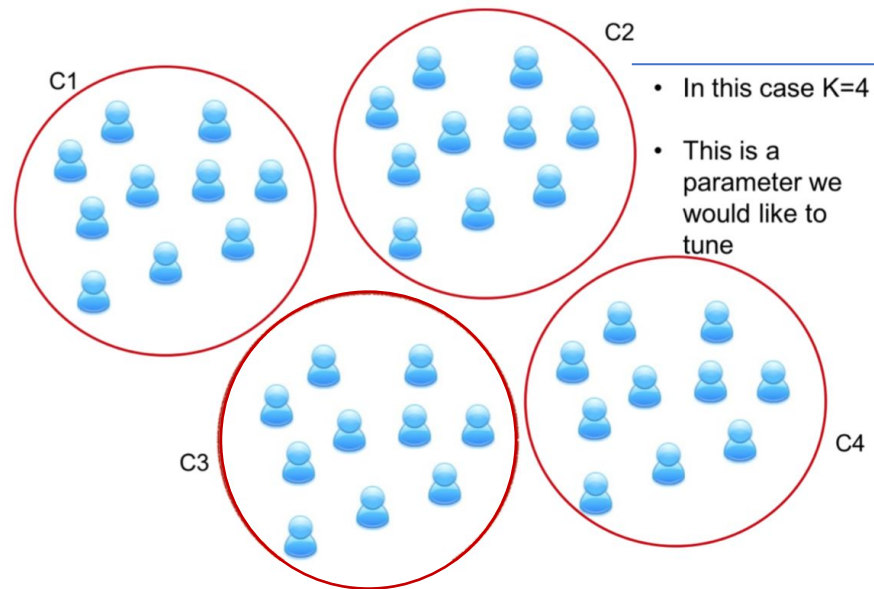
1. Busca usuarios más similares y obtiene un score de similaridad.
2. Recorre ítems de estos usuarios que no hayan sido consumidos por el usuario objetivo.
3. Multiplica el rating de estos ítems por el score de similaridad con el usuario y guardarlos en una lista.
4. Ordena de mayor a menor puntaje ponderado.
5. Recomendando los top N.

Limitaciones del Filtrado Colaborativo

- **Exactitud:** mientras más **K** vecinos considere menor incertidumbre y mayor exactitud.
- **Escalabilidad:** a mayor usuarios, mayor es el costo computacional de encontrar los **K** vecinos más cercanos. Complejidad de K-NN es $O(dnk)$
- **Dispersión:** el método no se ajusta con usuarios o ítems con pocos ratings. Este problema se conoce como “cold start” o “new item problem”.

Alternativas al Filtrado Colaborativo

- Filtrado Colaborativo basado en Usuarios con **Clusters**



- Tomar vecinos dentro del *cluster* de usuarios representándolos con los ratings que han dado.
- Usar algoritmo k-means para calcular obtener *clusters de usuarios*.
- *Generar la recomendación basada en items de usuarios cercanos*

¿De qué otra forma podríamos generar recomendaciones utilizando Filtrado Colaborativo?

Hint: buscar otra entidad para calcular similitudes

Filtrado Colaborativo Basado en Ítems

El filtrado colaborativo basado en ítems se basa en la premisa de que si a una persona le gusta un ítem, es probable que le gusten los ítems similares.

USER-BASED

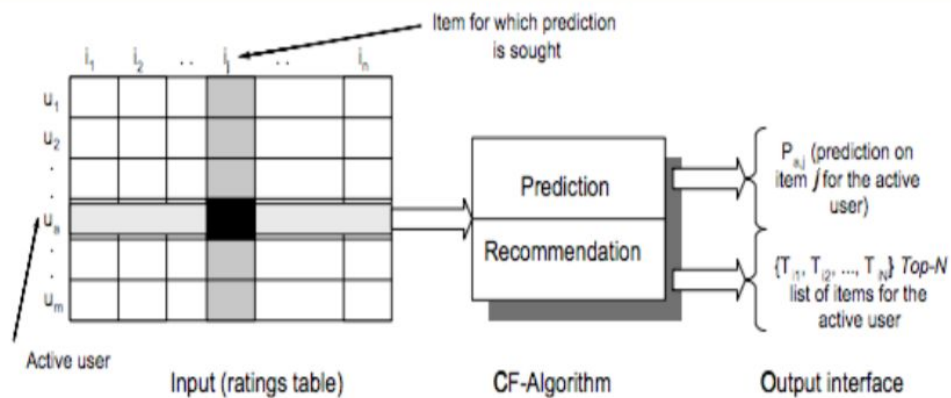


Figure 1: The Collaborative Filtering Process.

ITEM-BASED

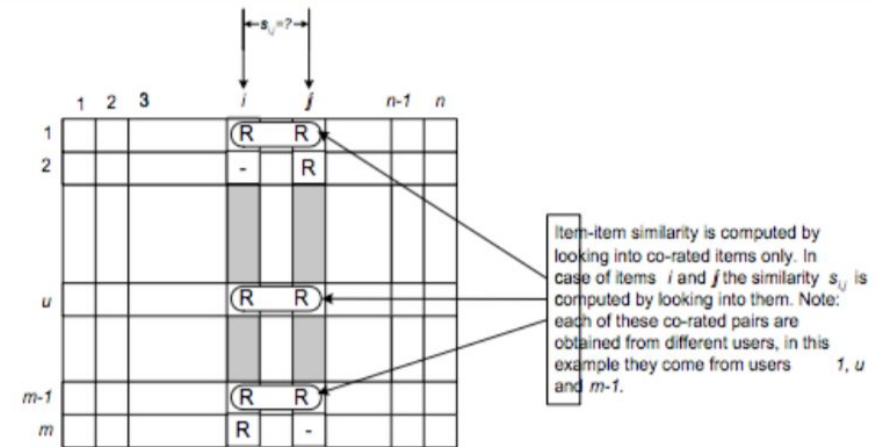


Figure 2: Isolation of the co-rated items and similarity computation

Fórmula de Similitud del Coseno

$$\text{sim}(i, j) = \frac{\sum_u r_{ui} \times r_{uj}}{\sqrt{\sum_u r_{ui}^2} \times \sqrt{\sum_u r_{uj}^2}}$$

- **$\text{sim}(i, j)$** es la similitud entre el ítem **i** y el ítem **j** .
- **r_{ui}, r_{uj}** calificaciones del usuario **u** para el ítem **i** e ítem **j** .
- **Sumatoria del denominador** multiplica ratings para usuarios que han calificado ítems **i** y **j** de manera conjunta.
- **Denominador** normaliza por la norma de las calificaciones de todos los usuarios para los ítems **i** y **j** .

Predicción de Calificación y Exclusión de Ítems Consumidos

$$\hat{r}_{ui} = \frac{\sum_{j \in N(i) \cap I'_u} \text{sim}(i, j) \times r_{uj}}{\sum_{j \in N(i) \cap I'_u} |\text{sim}(i, j)|}$$

$N(i) \cap I'_u$: ítems similares al ítem i que han sido calificados por u .

$|\text{sim}(i, j)|$: valor absoluto de similitud entre ítem i y j

Acotando la Búsqueda de Ítems Similares

Para optimizar el proceso y acotar la búsqueda, puedes usar un umbral de similitud o limitar el número de ítems similares que consideras para las recomendaciones.

Ventajas de Filtrado Basado en Ítems

Escalabilidad: Funciona mejor para conjuntos de datos grandes ya que se precalcula la similitud entre ítems.

Estabilidad: Los ítems suelen ser más estables en sus características que los patrones de calificación de los usuarios.

Calidad de Recomendaciones: En muchos casos, se ha observado que las recomendaciones basadas en ítems son más precisas que las basadas en usuarios.

Análisis de Item Based CF

- [Sarwar et. al. \(2001\) Item-Based Collaborative Filtering Recommendation Algorithms](#)
- Métrica de evaluación

$$MAE = \frac{1}{N} \sum_{i=1}^N |r_{u,i} - \hat{r}_{u,i}|$$

- Resultados relevantes
 - Efecto de la métrica de similaridad
 - Sensitividad de la proporción *Training / Test*
 - Tamaño de vecindario **K**
 - Comparación con otros métodos

Análisis de Item Based CF: comparación de métricas de distancia.

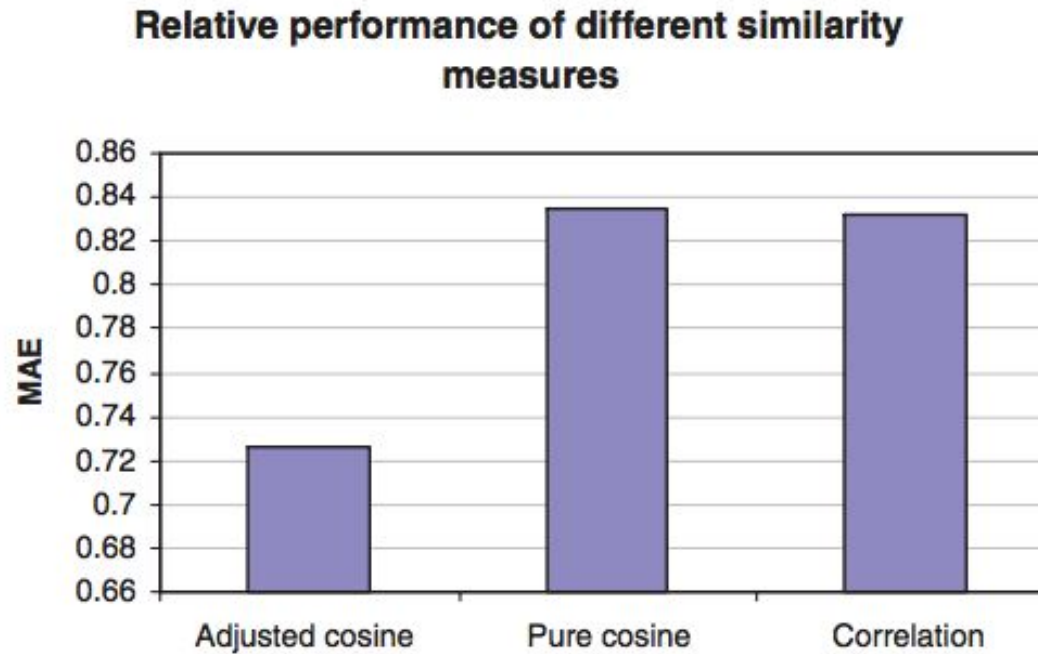


Figure 4: Impact of the similarity computation measure on item-based collaborative filtering algorithm.

Análisis de Item Based CF: comparación con otros métodos

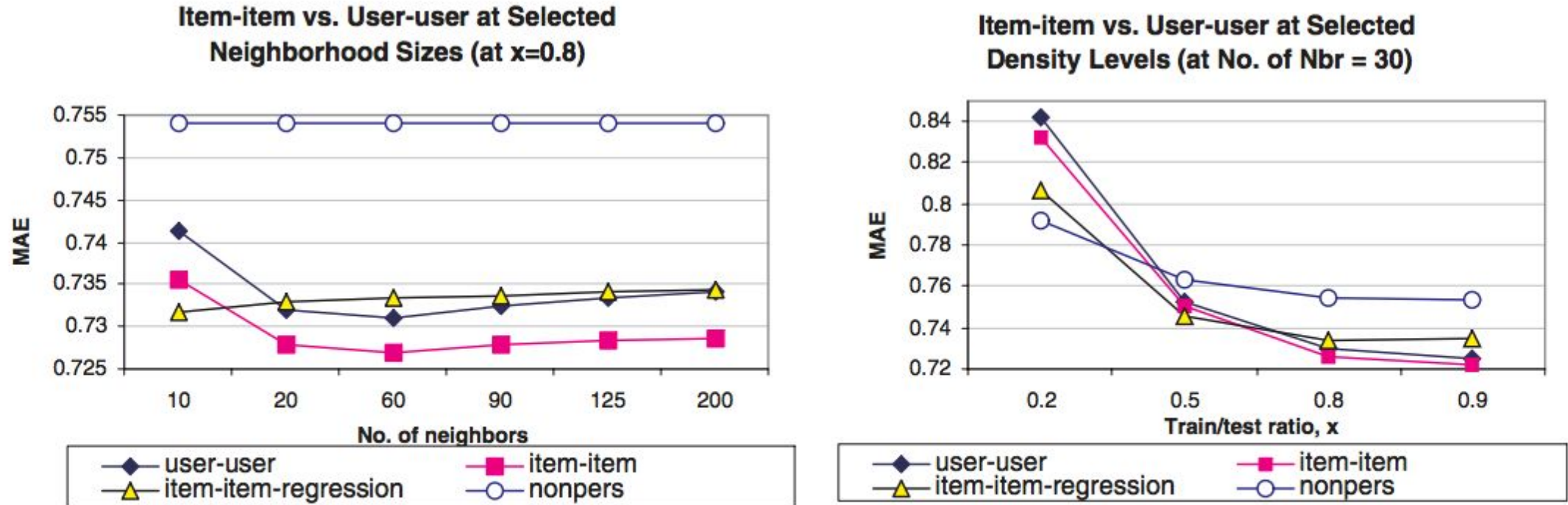


Figure 6: Comparison of prediction quality of *item-item* and *user-user* collaborative filtering algorithms. We compare prediction qualities at $x = 0.2, 0.5, 0.8$ and 0.9 .

Reflexión

¿Qué información captura el modelo de filtrado colaborativo basado en ítems?

¿Cómo realiza las recomendaciones?

¿Cómo realiza las recomendaciones?

PSEUDO CÓDIGO.

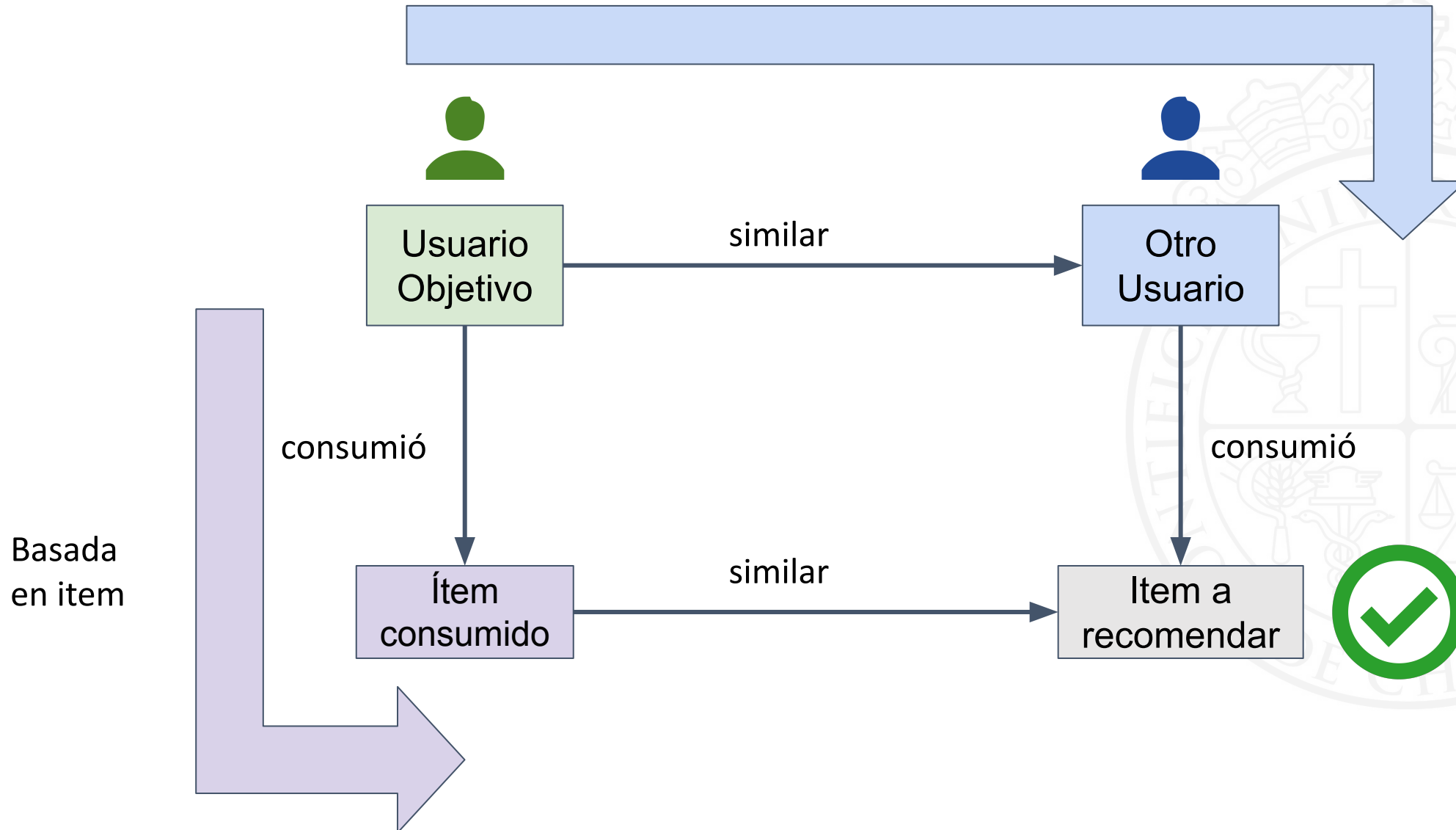
1. Busca items más similares (columnas) a los que consumio el usuario objetivo y obtiene un score de similaridad.
2. Recorre ítems que no hayan sido consumidos por el usuario objetivo.
3. Multiplica el rating de estos ítems por el score de similaridad con los items y guardarlos en una lista.
4. Ordena de mayor a menor puntaje ponderado.
5. Recomendando los top N.

Limitaciones de filtrado colaborativo basado en ítems.

- **Cold start:** Dificultad para recomendar a usuarios o ítems nuevos sin datos previos.
- **Tendencia a la popularidad:** Los ítems populares tienden a ser más recomendados, reduciendo la diversidad.
- **Esparsidad:** Con muchos ítems, solo unos pocos pueden tener suficientes calificaciones para determinar similitudes precisas.

Filtrado colaborativo (resumen)

Basada en usuario



Resumen de la clase

- Introducción a los sistemas recomendadores.
- Recomendación no personalizada.
- Recomendación semi-personalizada.
- Filtrado colaborativo basado en usuario.
- Filtrado colaborativo basado en ítems.



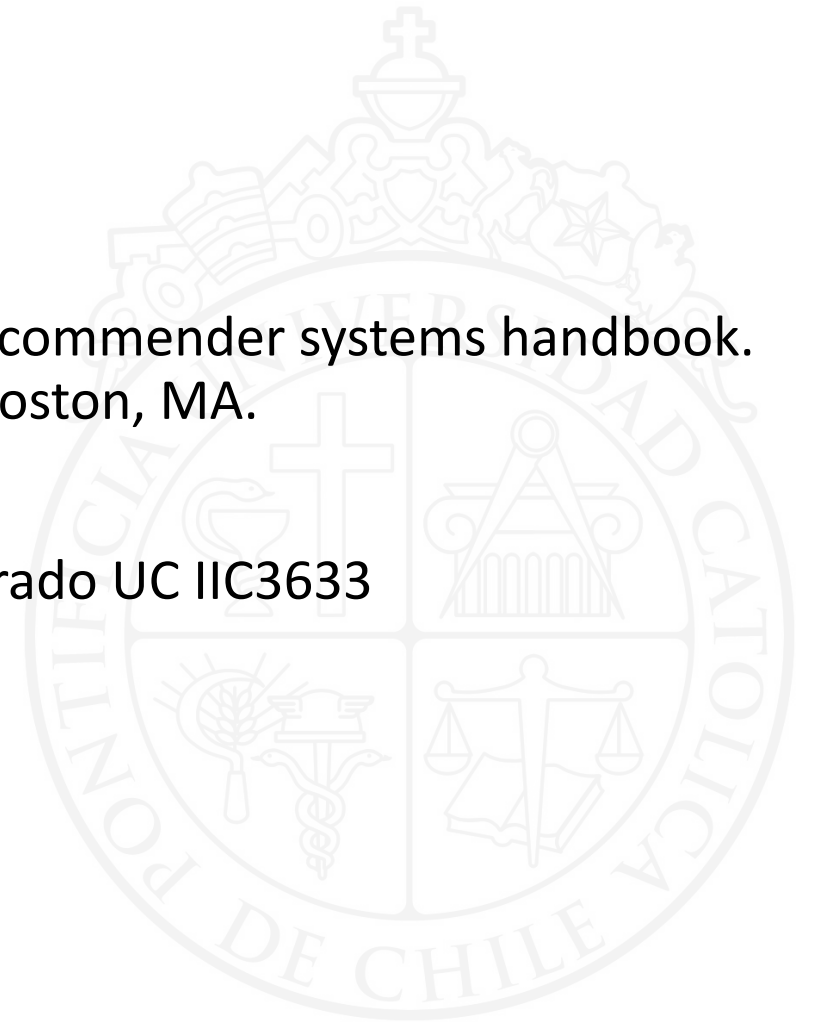
Referencias

- Recommender Systems Handbook

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In Recommender systems handbook (pp. 1-35). Springer, Boston, MA.

- Sitio web y material del curso de recomendadores postgrado UC IIC3633

<https://github.com/PUC-RecSys-Class/RecSysPUC-2022>



Práctico

<https://drive.google.com/file/d/174dBGGfC2BX3zDWGUmpj2Z5CjVr4BL3G/view?usp=sharing>

