

Introducción

- ▶ El 35% de las compras en Amazon en 2019 fueron generadas por un sistema de recomendación de acuerdo con MacKenzie y ese mismo año de reportó una ganancia de \$87.4 billones de dólares en su último cuarto (Palmer, 2020).
- ▶ La primer conferencia AMC Recsys fue llevada a cabo en 2007 donde presentó un problema por Netflix que buscaba mejorar su sistema de recomendación en un 10% y dando un premio de 1 millón de dólares. En 2009 concluyó el reto y se presentó la solución en la conferencia. A partir del siguiente año se creó RecSys Challenge donde una empresa presenta un reto de sistemas de recomendación donde se premia y presenta la mejor solución con el fin de generar nuevas maneras de implementar recomendaciones (Said, 2016).
- ▶ El trabajo toma los datos del RecSys Challenge 2022 que consiste las vistas y compras de ropa en una tienda en línea; se deben generar 100 recomendaciones de artículos por sesión. A partir del modelo ganador del reto se toma la base para generar las recomendaciones usando diferentes métodos de filtrado colaborativo y LightGBM (Zzh, Zhang, & Wentao, 2022).

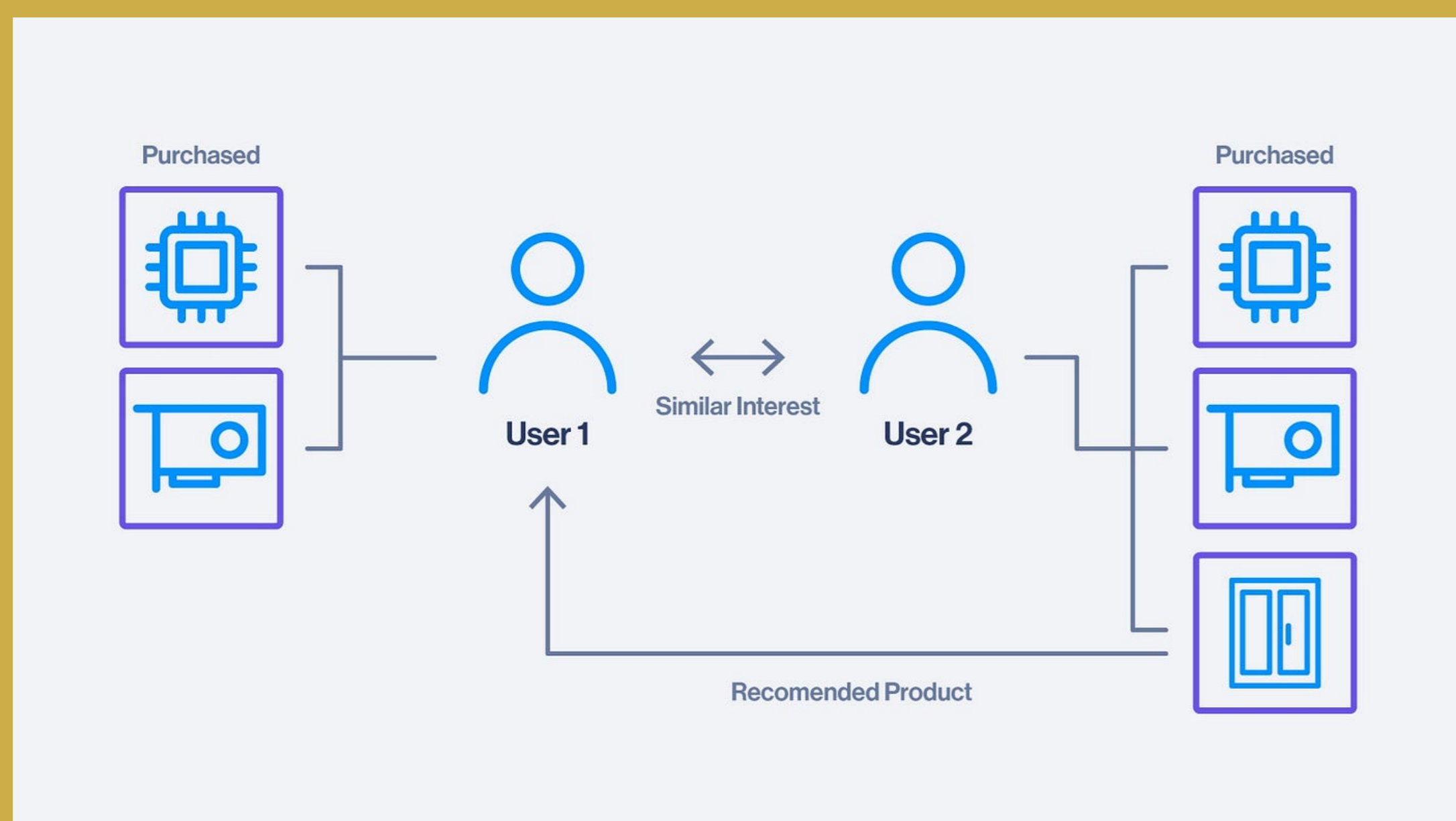


Figure 1: Sistema de Recomendación

Metodología

Modelo de Datos

- ▶ Los datos son dados por la competencia que consta de 18 meses de compras, siendo el último mes el solicitado a predecir. Consta de 1.1 millones de sesiones con posibilidad de usar 1 millón de sesiones para entrenar el modelo y la evaluación se hace a 100,000 sesiones.
- ▶ La figura 2 muestra el modelo de datos de las sesiones. Se cuenta con las vistas de productos antes de comprar el producto final. Se tiene un catálogo de características por producto.

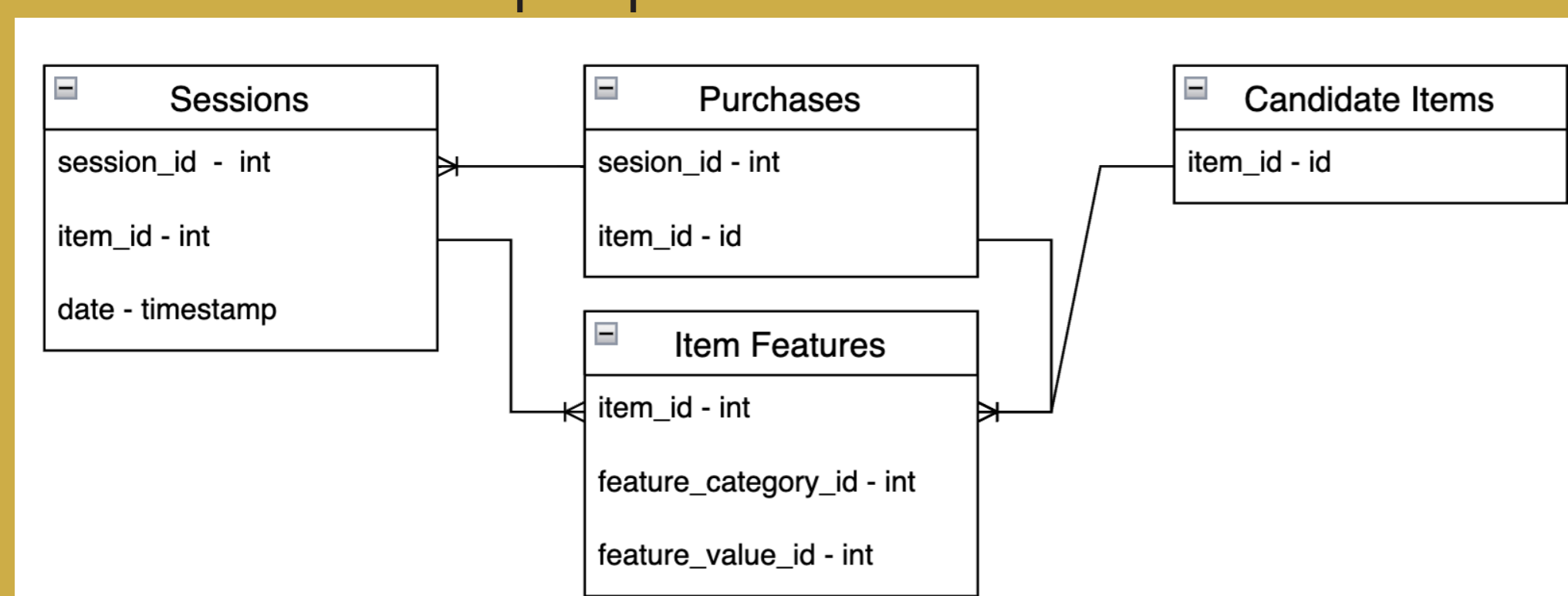


Figure 2: Modelo de datos de compras en línea.

Pipeline del Modelo

- ▶ La figura 3 muestra el proceso para obtener la recomendación:
 1. Feature Prepare: se realizan algoritmos de filtración colaborativa que se realizan por sesión y por productos en general. Además se usan otros métodos para caracterizar artículos como PCA o Word2Vec.
 2. Retrieval: se busca reducir el número de productos a sugerir por sesión para no evaluar la cantidad de productos disponibles y mejorar el procesamiento.
 3. Feature Transform: una vez reducido los posibles artículos a recomendar se le añaden características hechas en la primera etapa.
 4. Model Train & Inference: se entrena y se obtienen resultados a través de un Ranking LightGBM que permite evaluar los productos por grupos.

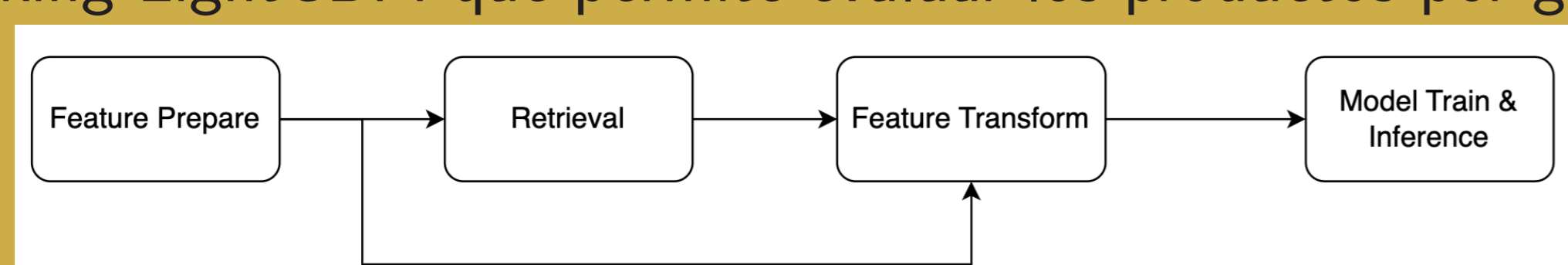


Figure 3: Pipeline para obtener recomendaciones

Metodología

Collaborative Filtering Similarity

- ▶ Collaborative Filtering Similarity toma en cuenta la interacción de los artículos dentro de cada sesión. Toma en cuenta el tiempo entre dos artículos, la cantidad de artículos entre ambos artículos y las sesiones que tienen interacción los artículos.

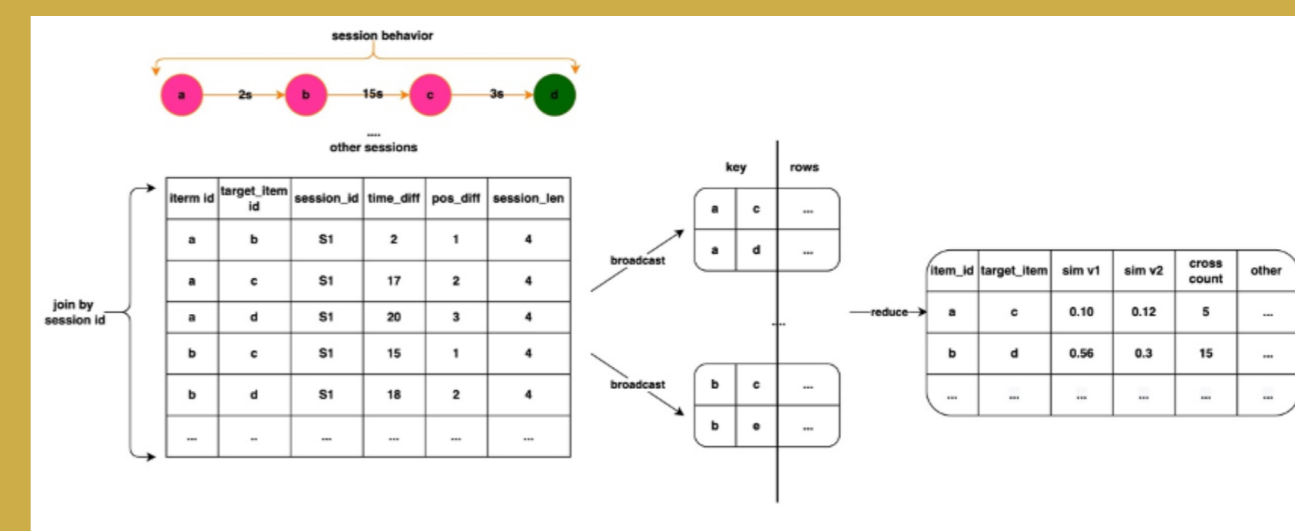


Figure 4: Proceso de Collaborative Filtering Similarity

Swing Algorithm

- ▶ El algoritmo consiste en ver la interacción de artículos pero a través de las sesiones. Toma en cuenta las sesiones totales de cada artículo, los artículos que ve el usuario en cada sesión y las sesiones donde ambos artículos interactúan.
- ▶ El resultado consiste en tomar la similitud de artículos pero basado en las sesiones y no la interacción entre ambos artículos. Sirve para obtener otra característica entre artículos.

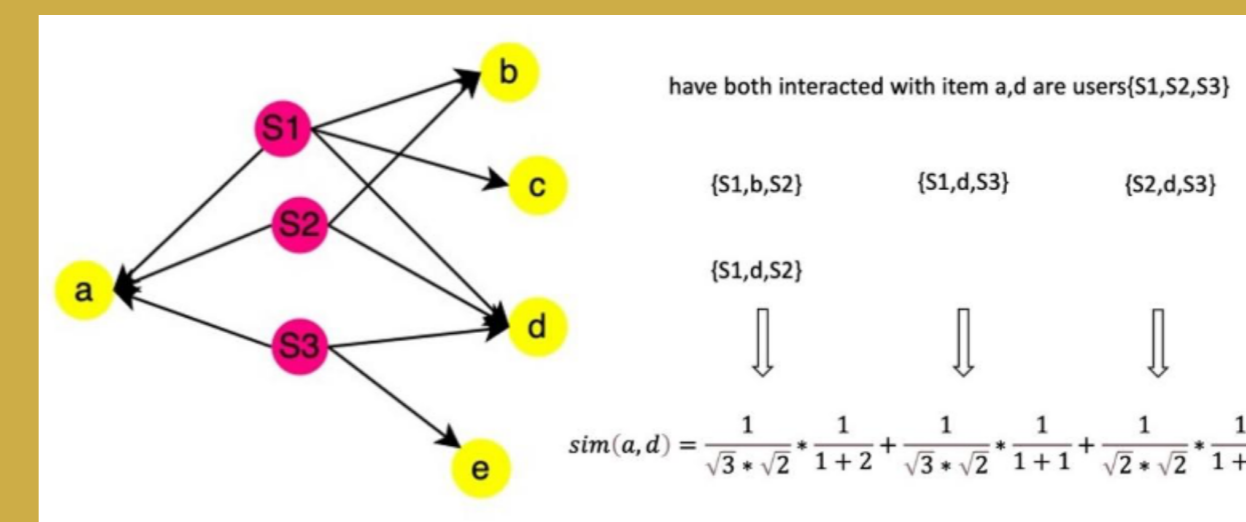


Figure 5: Proceso de Swing Algorithm

LightGBM

- ▶ Para general la recomendación usamos LightGBM que es un modelo basado en árboles. La ventaja sobre otros algoritmos basados en árboles es la escalabilidad, aprendizaje rápido y computación en paralelo.

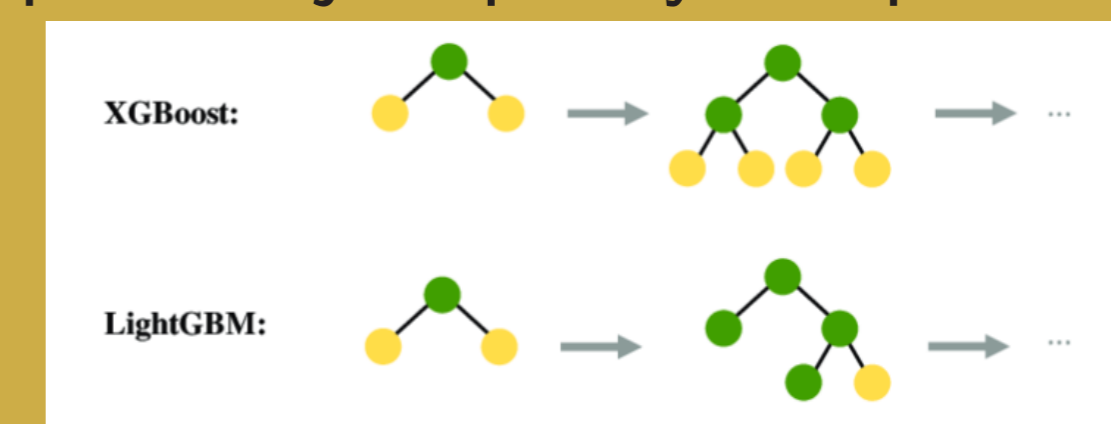


Figure 6: Estructura LightGBM

Resultados

- ▶ Actualmente se reducen los posibles artículos a recomendar, dado que son cerca de 5000 artículos que el usuario puede comprar, con 5 artículos vistos por sesión en promedio y 100 mil sesiones a recomendar; se puede llegar a 2,500 millones de filas. La lista a sugerir se reduce en promedio a 1000 artículos por sesión dando alrededor de 500 millones de filas. La reducción de la lista se hace mediante los métodos de Collaborative Filtering Similarity, Swing Algorithm y generales para asegurar un mínimo de 100 recomendaciones.

Conclusiones

- ▶ Se busca actualmente agilizar el proceso de guardado de las tablas resultantes para los modelos, dado que guardar 500 millones de datos es un proceso lento.
- ▶ Se incorporarán otros algoritmos al modelo buscando mejorar la métrica de evaluación Mean Reciprocal Rank. El ganador obtuvo un resultado de 0.261; se busca replicar ese resultado.

Referencias

- Palmer, A. (2020, January). *Amazon soars after huge earnings beat*. <https://www.cnbc.com/2020/01/30/amazon-amzn-q4-2019-earnings.html>. (Accessed: 2023-8-2)
- Said, A. (2016). A short history of the RecSys challenge. *AI Mag.*, 37(4), 102–104.
- Zzh, Zhang, W., & Wentao. (2022). Industrial solution in fashion-domain recommendation by an efficient pipeline using GNN and lightgbm. In *RecSys challenge 2022*. New York, NY, USA: ACM.