

Reporte Modelos de Recomendación: Tienda Mayorista11

1.Objetivo

El objetivo de este proyecto es proporcionar un sistema de recomendación para la distribuidora de productos **Mayoristas11**, que sugiera productos relevantes a comercios locales que aún no hayan comprado productos del portafolio. La intención es además de la colocación de más productos del portafolio , aumentar el ticket promedio.

2. Hipótesis Planteadas

Las opciones de enfoque fueron dos: un sistema de recomendación basado en un filtro colaborativo con similaridad coseno y vecinos cercanos, y otro basado en el algoritmo PageRank aplicado a las relaciones entre comercios y productos.

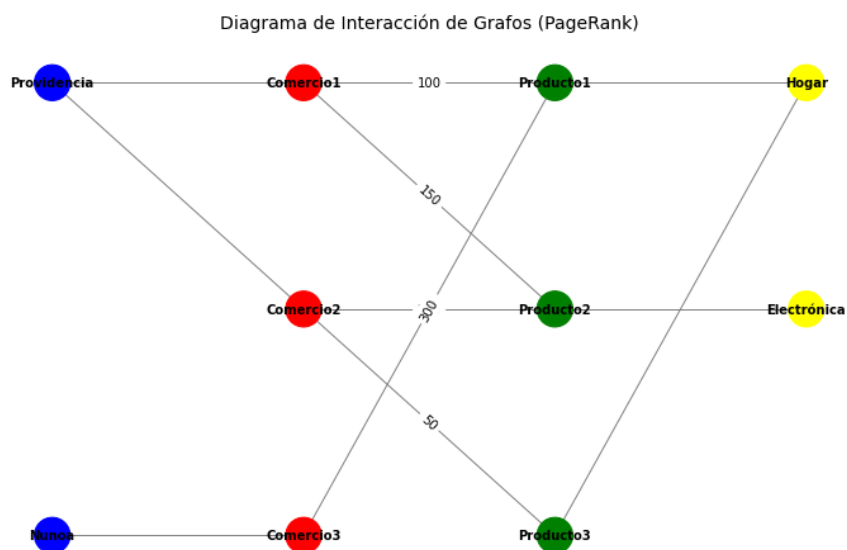
- **Hipótesis 1:** El modelo de filtro colaborativo será más efectivo en generar recomendaciones precisas para comercios con un historial de compras significativo, aprovechando la similaridad entre ellos. Se basa en buscar lo que compren los comercios más parecidos en cuanto a total de ventas por producto, para luego recomendar lo que estos comercios compraron. El modelo entregará los productos rankeados por orden de relevancia.
- **Hipótesis 2:** El modelo PageRank ofrecerá una cobertura más amplia, generando recomendaciones para todos los comercios, incluso para aquellos con menos historial de compras, al considerar las relaciones globales en la red de productos y comercios. El modelo entregará los productos rankeados por orden de relevancia.

3. Modelamiento de datos.

Dataset : El primer desafío para crear un modelo de recomendación es separar en conjunto de train y test. En este caso el dataset tiene la particularidad que no tiene un identificador de transacción, o timestamp. Además si se observa el dataset se tiene que todos los clientes registran compras de todos los productos del portafolio de productos, entonces considerando esto lo que se hará es por cada comercio quitar un 20% de los productos para simular la "no compra" de ciertos productos de parte de un comercio.

Modelo Colaborativo de Similaridad Coseno : Se modela mediante una matriz de similaridad de comercios y clientes. En este punto tomaremos como referencia total_price, que indica el total del valor monetario de los productos vendidos.

Modelo Pagerank de grafos: El modelo de grafos tiene 4 tipos de nodos : commerce, **product**, **district**, **category** . Además entre comercio y producto la arista tiene un peso que corresponde al precio total(total_price de los productos comprados por producto. Además a cada producto le corresponde una categoría y a cada comercio un district. Ejemplo en la figura siguiente



4. Evaluaciones

Como lo que se pretende es aumentar la colocación de productos en tiendas por parte de Mayoristas11, los modelos se miden en términos de cobertura. Además como se quiere también aumentar el ticket promedio.

	Colaborativo	PageRank
Comercios con cobertura	75%	100%
Promedio cantidad productos recomendados	7.36	10
Número de productos recomendados	50	50

Tabla 1. Cobertura modelos de recomendación.

rank_colab	precio promedio colab	rank_pagerank	precio promedio pagerank
1	49.76	1	92.52
2	46.68	2	85.43
3	53.44	3	77.66
4	61.72	4	69.48
5	54.178082	5	61.97
6	52.085714	6	54.44
7	62.384615	7	46.32
8	52.068182	8	37.85
-	-	9	27.06
-	-	10	17.50

Tabla 2. Ticket promedio de acuerdo al ranking de recomendación.

5. Conclusiones

El Filtro colaborativo en algunos casos podría no entregar una recomendación, esto es debido a que el comercio podría tener poco historial de datos.

En el filtro colaborativo cuando recomienda productos de un cliente similar del mismo district, todos sus productos tienen la misma ponderación. Lo que puede introducir una distorsión.

PageRank logró ofrecer recomendaciones a todos los comercios en el conjunto de datos, mientras que el filtro colaborativo no pudo generar recomendaciones para aquellos con poca interacción.

Pagerank entrega siempre una recomendación de los productos disponibles aún sin comprar a todos los comercios.

PageRank recomienda productos con un ticket más alto que el filtro colaborativo. Esto se observa al ver la Tabla 2. Los productos más relevantes recomendados por el PageRank son más caros en promedio que los recomendados por el modelo colaborativo.

Los resultados de PageRank quedan condicionados por el peso de la relación entre comercios y productos, que el caso del modelo fue el precio total por comercio y producto. Esto podría influir en que los productos recomendados tiendan a ser los más populares.