

Challenge Técnico DS – Esteban López

A. Objetivo

Desarrollar un sistema de recomendación que sugiera productos relevantes a cada comercio.

B. ¿Qué se hizo?

B.1 Análisis exploratorio de datos

- a) Se cargaron los 3 csv y se unieron mediante identificador de commerce y producto dando un total de 500.000 observaciones.
- b) Para recomendar productos populares (los más vendidos) se deben sumar las cantidades vendidas y agruparlas por comercio y producto para tener un consolidado. Dicha consolidación entregó un total de 5.000 observaciones.
- c) Con estas 5.000 observaciones se realiza un diagrama de cajas y se eliminan valores atípicos/sesgados mediante rango intercuartílico, quedan 4.965 registros. Estos valores sesgados pueden ocasionar eventualmene predicciones sobreajustadas en los modelos por lo que es recomendable omitirlas.

B.2 Modelos de recomendación

Se crearon 6 modelos de recomendación. 1) Primero se evaluará el RMSE de cada modelo, la recomendación es escoger el modelo con menor RMSE, en caso que entregue valores similares se decidirá por 2) el mejor ajuste entre los valores reales y predichos el cuál se revisará mediante un gráfico de densidad.

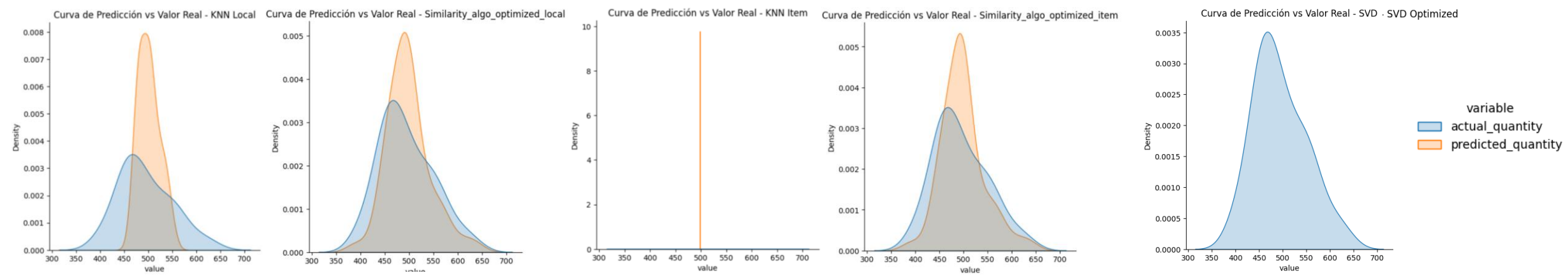
Necesitamos datos de interacción de productos de los locales para construir un sistema de recomendación colaborativo. En este caso se tomarán las cantidades como dato de interacción. Se basarán recomendaciones con similitud del coseno y se usará KNN para encontrar locales similares.

C. Resultados

Estos fueron los modelos usados junto con su RMSE (códigos y resultados generales adjuntos en html)

Modelo	RMSE
Filtro colaborativo por local (KNN Local)	58,56
Filtro colaborativo por local ajustado a hiperparámetro (Similarity_algo_optimized_local)	56,05
Filtro colaborativo por elementos (KNN Item)	55,51
Filtro colaborativo por elementos ajustado a hiperparámetro (Similarity_algo_optimized_item)	55,66
SVD (SVD)	161,22
SVD ajustado a hiperparámetro (SVD Optimized)	161,22

Se concluye que los valores de RSME en los primeros 4 modelos son bastante similares, el filtrado colaborativo por elementos posee el menor valor. Revisemos 6 ejemplos de valor real y predicho para cada modelo utilizando una predicción de las cantidades para un id_commerce=6.



El modelo de filtro colaborativo por elementos ajustados a hiperparámetro (**Similarity_algo_optimized_item**) entrega valores predichos similares a los valores reales, con este modelo se harán las recomendaciones. Otros modelos como el KNN Item o el SVD entregan predicciones alejadas a los valores reales, el primer caso se concentra en entregar predicciones alrededor de la media de la cantidad vendida y el Segundo no entrega predicciones.

D. Conclusiones

```
# Df_grouped es el DataFrame con los datos de transacciones
top_recommendations = get_recommendations_for_district(df_grouped_filtered, 'Macul', new_id_commerce=12, top_n=10, algo=similarity_algo_optimized_item) ;

# Imprimir las recomendaciones
for product_id, predicted_quantity in top_recommendations:
    print(f"Producto {product_id}: Cantidad predicha {predicted_quantity:.2f}")
```

```
Producto 43: Cantidad predicha 601.55
Producto 9: Cantidad predicha 595.53
Producto 41: Cantidad predicha 592.57
Producto 2: Cantidad predicha 581.68
Producto 37: Cantidad predicha 579.61
Producto 32: Cantidad predicha 563.64
Producto 28: Cantidad predicha 561.71
Producto 36: Cantidad predicha 553.78
Producto 27: Cantidad predicha 542.81
Producto 22: Cantidad predicha 538.84
```

Finalmente se expone un extracto de las 10 mejores recomendaciones de productos (según popularidad) para la comuna de Macul con un `id_commerce=12` utilizando el modelo de filtrado colaborativo por elementos ajustado a hiperparámetro.

Nota: Los valores pueden verse modificados en cantidades pequeñas porque cada vez que se entrena el modelo se genera un nuevo grupo de prueba y entrenamiento.

En este estudio de caso, se revisaron 6 formas diferentes de construir sistemas de recomendación:

filtrado colaborativo basado en la similitud (4 modelos), filtrado colaborativo basado en modelos (factorización de matrices) (2 modelos SVD). Una vez que se construyeron estos sistemas de recomendación, podemos usar Pruebas A/B para medir la efectividad de estos sistemas.