

MANUAL DE USUARIO VV DATA INSIGHTS

Nota: Para la entrega del modelo se realizó una pseudointerfaz, es decir, se deben correr los dos modelos (el de b2b y el de b2c) y al finalizar se desplegará una pequeña interfaz interactiva donde se pueden cambiar los parámetros para ver las predicciones del modelo.

1. Descripción general

Este sistema de recomendación permite generar sugerencias personalizadas de productos para clientes B2C o empresas B2B utilizando tres enfoques distintos:

- **Modelo XGBoost:** basado en características tabulares y comportamiento histórico.
- **Modelo LightFM:** basado en aprendizaje de representaciones con características cruzadas usuario-producto.
- **Modelo Híbrido:** combinación ponderada de los dos anteriores (usando un parámetro α ajustable).

Los modelos presentan una descripción mucho más detallada en el documento de análisis, sin embargo, han sido entrenados y evaluados con métricas reales (AUC, precisión, recall, F1-score).

2. Requisitos previos

Para correr el modelo es necesario descomprimir el archivo de “Algoritmo de Recomendación” allí dentro se encuentran:

- Dos archivos .ipynb, uno para cada enfoque del problema (B2B y B2C, nombrados adecuadamente)
- Todos los archivos de extensión .pkl y .json para la futura implementación de interfaz
- Una carpeta de data que contiene dos bases de datos modificadas a partir de los datos iniciales entregados para la elaboración del reto.

Nota: Todo el proceso previo, de manejo de datos, entendimiento de los mismos y creación de nuevas variables está dentro del archivo comprimido “Reto Corona” enviado la entrega pasada y adjunto en esta entrega final, a partir de ese proceso se generaron las dos bases de donde se genera el modelo de recomendación.

- Plataforma: **Google Colab**

Se recomienda correr el archivo “recomendaciónB2B.ipynb” y “recomendacionB2C.ipynb” desde el entorno de Google Colab ya que por la magnitud de las bases y la complejidad computacional de los algoritmos, si se corre

desde un entorno local como Visual Studio se consumirá toda la memoria RAM y el programa no se ejecutará.

Para esto se debe tener una cuenta de Gmail activa.

3. Uso paso a paso

0. Descargar localmente las bases de datos “b2b_nuevo” y “b2c_nuevo”, ponerlas en el mismo ambiente de Google Colab donde se encuentran los archivos “recomendaciónB2B.ipynb” y “recomendacionB2C.ipynb” y correr de manera independiente cada uno.

1. **Ingresar el ID del cliente:**

Una vez compilado todo el archivo la última celda desplegará automáticamente una pseudo interfaz donde el usuario podrá escribir el ID del cliente. Ese formato es el mismo para las dos bases.

- Ejemplo B2C: 11
- Ejemplo B2B: B2B_01

Para efectos prácticos les anexamos ID's para que puedan probar el funcionamiento de la herramienta.

B2B: 'B2B_01' 'B2B_02' 'B2B_03' 'B2B_04' 'B2B_05' 'B2B_06'

B2C: Un número desde el 1 al 419226

2. **Elegir el número de recomendaciones**

En una barra se puede seleccionar de 1 a 20 la cantidad de productos recomendados que se desea conocer

3. **Ajustar el parámetro α (entre 0 y 1)**

En otra barra se define cuánto peso darle al modelo LightFM en el enfoque híbrido. Ya que los modelos tienen enfoques distintos se deja a discreción de la empresa que modelo se ajusta más y que modelo genera resultados más coherentes con respecto al modelo de negocio de esta.

4. **Presionar el botón “Comparar Modelos”.**

Para obtener el resultado, al haber puesto los parámetros necesarios solo hace falta presionar el botón y se desplegarán los resultados.

4. Resultados presentados

Una vez generado el análisis, se mostrarán:

a. Recomendaciones personalizadas

- Tabla de productos recomendados con sus *scores* individuales por modelo y un *score* promedio.
- Los productos están ordenados de mayor a menor probabilidad de intención de compra.

Nota: Por la elevada dimensionalidad de variables y productos, la posibilidad de combinaciones es muy elevada y al comparar los resultados de los tres modelos se pueden generar scores negativos ya que cada modelo tiene un enfoque distinto y al buscar generalizarlos ocurre esto.

b. Gráfico de comparación de scores

- Un gráfico de barras comparativo entre modelos: LightFM, XGBoost y Híbrido. Con el número y referencia de productos seleccionado para el cliente seleccionado.

En la foto adjunta se evidencian los 10 productos recomendados para el cliente con ID: 10

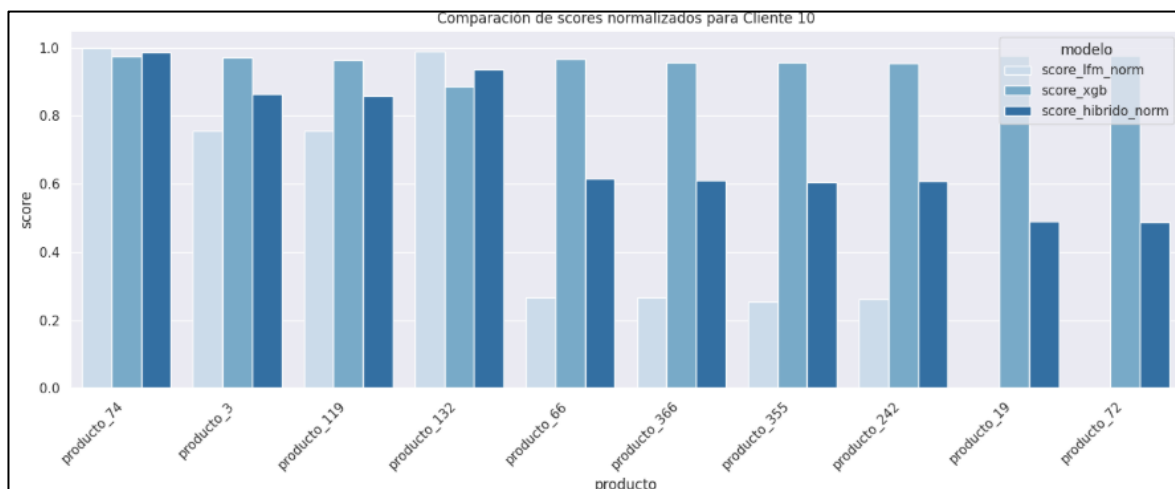


Ilustración 1 Gráfica de barras resultante de la comparación de modelos para el cliente 10

c. Métricas de desempeño global (ya calculadas y cargadas):

Al final de la table se visualizan las métricas específicas de cada algoritmo implementado, para tener un mejor entendimiento de los resultados.

5. Consideraciones adicionales

- El sistema es **reactivo y dinámico**: cada vez que se cambia el ID o el α , se recalculan todas las predicciones.
- Las métricas no se recalculan en tiempo real; fueron calculadas previamente y almacenadas en diccionarios.
- El sistema está preparado para crecer con nuevos datos cargados.
- Los demás archivos que se encuentran dentro del documento comprimido enviado pero que no se mencionan en este manual hacen parte del proceso de elaboración de la siguiente fase que es la Interfaz final con el usuario. Si bien, aún no está finalizado hace parte del trabajo realizado y de las siguientes iteraciones del proyecto.
- Se realizó un pequeño Demo donde se pueden visualizar los resultados de ambos modelos en tiempo real. Este video se encuentra en este [enlace](#).
- Dado que nuestro cliente es el área de Analítica de Corona, optamos por desarrollar una pseudointerfaz funcional, enfocada en la comprensión y validación del modelo por parte de usuarios con conocimientos técnicos en analítica y desarrollo. Esta decisión nos permitió concentrarnos en validar la lógica, precisión y aplicabilidad de los algoritmos propuestos.

Consideramos que el despliegue de una interfaz gráfica o en la nube corresponde a una siguiente etapa del proyecto, que deberá abordarse una vez la solución planteada haya sido aprobada. A partir de esta validación, el objetivo será diseñar e implementar una solución definitiva que sea robusta y escalable, acorde con las necesidades operativas de una organización del tamaño de Corona.