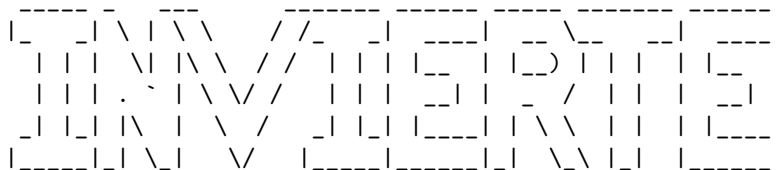


Caso Advanced Analytics - Banco BICE



Introducción

INVIERTE es el nuevo proyecto de una conocida institución financiera llamada YO INVIERTO. INVIERTE posee una cartera de clientes, la cual puede invertir en distintos productos de inversión, algunos de estos pueden ser:

- Depósitos a plazo.
- Fondos de inversión.
- Fondos mutuos.
- Acciones.

El objetivo de toda institución financiera es mantener los saldos ya sea en cuentas corrientes, cuentas vistas o, en este caso particular, en inversiones, puesto que de esta manera pueden administrar los fondos generando rentabilidades por comisión de administración, comisión por compra o venta, movimientos, retiros tempranos o parte de la rentabilidad asociada al instrumento financiero propiamente tal.

Problema

Producto de fenómenos macroeconómicos y una fuerte competencia en la oferta de productos de inversión, INVIERTE ha sufrido fugas de capital invertido por sus clientes. Dado lo anterior, se le pide poder entrenar un modelo que asocie una probabilidad o asigne una etiqueta binaria de 1 o 0 para cada cliente sobre si retirarán sus fondos en el siguiente período. **Como es de esperar usted debe encontrar un modelo que prediga los 1's (etiqueta positiva), es decir, los clientes que retirarán sus patrimonios invertidos en INVIERTE. Dicho de otra manera, usted no debe obviar los 0's, pero sí el foco está en el rendimiento en los clientes que se fugarán.**

Diccionario de Bases y Variables

Para esta tarea se le entregarán 6 fuentes distintas de datos, el detalle se encuentra a continuación:

1. **id_label.csv**: Base con id's (columna *id*) de clientes junto con su etiqueta (columna *label*) de 1 o 0 según corresponda. La columna *label* contiene **1's** cuando un cliente retira parte de sus fondos (no importa cuanto, basta que retire 1 peso para considerarse un 1) durante el mes siguiente (marzo). Por

otro lado, la etiqueta de **0's** quiere decir que el cliente mantiene o aumenta sus fondos en INVIERTE. (no importa cuanto, basta que agregue 1 peso para considerarse un 0)

2. **demograficos.csv**: Datos demográficos de los clientes de INVIERTE.
 - a. *fec_nacim*: Fecha de nacimiento.
 - b. *flg_empleado*: S: El cliente es colaborador de YO INVIERTE. N: No es colaborador.
 - c. *gls_nacioncorta*: País de residencia del cliente.
 - d. *gls_residencialarga*: Categoría de la residencia donde vive el cliente.
 - e. *gls_profesionlarga*: Profesión del cliente.
 - f. *gls_sep_bieneslarga*: Indica si el cliente está casado o no. En caso de ser vacío, el cliente está casado, si el valor es “*sin separación*” está separado sin separación de bienes y “*con separación*” está separado con separación de bienes.
3. **cuenta_banco.csv**: Cuenta banco es un producto de YO INVIERTE que te permite tener una tarjeta llamada **inversionista_corriente**. Si tienes tarjeta en cuenta banco, los ejecutivos de cuenta banco te asignan una clasificación segun tu perfil (variable o columna *clasificacion_cliente*).
4. **deudas.csv**: Resumen de las deudas por periodo que mantienen los clientes de INVIERTE con entidades crediticias.
 - a. *periodo*: Periodo en año y mes (YYYYMM) de la deuda.
 - b. *entidades*: Cantidad de entidades crediticias con las que el cliente mantiene créditos consumo.
 - c. *disponible*: Monto que el cliente tiene disponible para utilizar como deuda/prestamo.
 - d. *deuda_1 a deuda_4*: Monto adeudado tipo 1 a 4 (representan deudas que son independientes entre si).
 - e. *mora_1 a mora_4*: Monto adeudado impago en etapa 1 a 4 (a mayor etapa, mayor tiempo impago).
 - f. *mora_5 a mora_6*: Monto adeudado impago indirecto en etapa 1 y 2 (representan deudas que no fueron tomadas directamente por el cliente).
 - g. *mora_7*: Otros tipos de montos adeudado impagos.
5. **Base carpeta cierre_patrimonio**: Esta carpeta contiene archivos *.csv* que tienen los cierres de patrimonios de los clientes al final del mes. Digamos, en el archivo 1, el cliente 1, en la fecha indicada por el archivo o la columna *fecha_cierre* indica que el patrimonio al final de ese mes es el número indicado en la columna *patrimonio*.
6. **aportes_rescates_12m.csv**: Este archivo contiene un consolidado de la suma de todos los aportes y rescates de los clientes durante 12 meses (suma). Digamos, $n_{aportes} = 14$ quiere decir que el cliente x ha hecho 14 aportes en los últimos 12 meses. Por otro lado, $n_{rescates} = 14$ quiere decir que el cliente ha rescatado 14 veces en los últimos 12 meses.

Algunas notas respecto a estas bases de datos:

- a. Si usted hace el conteo de 1's y 0's de la base de *id_label.csv*, debiese llegar a que la base contiene un 85% de 0's y un 15% de 1's aproximadamente.
- b. Si existen valores nulos en la tabla de demográficos quiere decir que son datos que no fue posible recopilar.
- c. No todos los clientes de INVIERTE tienen tarjeta *inversionista_corriente* o dicho de otra forma, no todos los clientes de INVIERTE son parte de cuenta banco, por lo tanto, puede que algunos clientes de INVIERTE no tengan clasificación de acuerdo a su perfil, puesto que no son clientes de cuenta banco.
- d. No todos los clientes de INVIERTE tienen deuda con entidades crediticias.

Desarrollo y Preguntas a Responder

1. Roadmap

- a. ¿Qué esquema, mapa conceptual o carta Gantt propondría para completar este desafío?.
- b. ¿Qué herramientas utilizaría? Explique sus razones.

2. Desarrollo

- a. ¿Qué insights puede obtener de los datos que puedan ser relevantes para la resolución del problema? Apoyese en gráficos, tablas o lo que estime conveniente.
- b. ¿Qué algoritmo/modelo estima adecuado para resolver esta problemática? Justifique su respuesta.
- c. ¿Cómo evaluaría el rendimiento de este modelo? Justifique su respuesta.

3. Resultados

- a. ¿Cuáles son los principales resultados de su modelo? Apoyese en gráficos, tablas o lo que estime conveniente.
- b. ¿Son estos resultados lo suficientemente satisfactorios para proponer el modelo como solución a la problemática?.
- c. ¿Cómo cree que podría mejorar el rendimiento de este modelo? ¿Qué otras variables propondría?.
- d. ¿Cuál sería su metodología para llevar este modelo a producción?.

4. Deploy

Asuma que el modelo fue puesto en producción y ha obtenido los primeros feedbacks.

- a. El stakeholder del proyecto le sugiere que valorice en términos monetarios su modelo, ¿Cómo lo haría? ¿Cree que es importante esta sugerencia para el ciclo de los proyectos de Advanced Analytics?.

- b. El stakeholder del proyecto indica que el modelo no está resolviendo la problemática en cuestión y pide reevaluar la solución propuesta utilizando un enfoque distinto, ¿Qué haría ante esta situación?.

Finalmente, suponga que el modelo ha tenido éxito y buena recepción por las áreas de negocios y se le pide realizar una presentación para distintas entidades del proyecto, ¿Cómo enfocaría sus presentaciones para las siguientes entidades?

- a. El equipo de Advanced Analytics.
- b. Stakeholder del proyecto.
- c. Ejecutivos comerciales que harán uso de las predicciones del modelo.

Presentación

El equipo de Advanced Analytics no solo se debe enfocar en la parte técnica de sus proyectos, sino que también debe ser capaz de promoverlos y mantener una continua interacción con las áreas que harán usos de estos. Dado lo anterior, se le pide presentar sus resultados en una presentación que abarque las preguntas del inciso de **Desarrollo y Preguntas a Responder** o puede ser en formato storytelling, pero que el foco esté centrado en las preguntas del caso.

Algunas consideraciones y orientaciones que pueden ayudar:

- a. No todas las variables les podrían resultar útiles en el modelo, es trabajo del candidato identificar las mejores variables predictoras.
- b. No se limite a utilizar la información de estas variables solo de manera directa, ya que interacciones entre variables y relaciones temporales podrían ayudar a mejorar el rendimiento del modelo.
- c. Las preguntas no tienen respuestas correctas o incorrectas, sientase libre de utilizar los supuestos que estime conveniente para el desarrollo del modelo y la presentación de sus resultados.

¡Te deseamos mucho éxito en esta etapa!

Equipo de Advanced Analytics

Banco BICE