



## CURSO DE APRENDIZAJE AUTOMÁTICO

### SEGUNDO SEMESTRE

#### TAREA 2

Los departamentos comerciales y de marketing de un banco local han lanzado dos nuevos productos de crédito: Tarjeta Platinum e Internet Global.

Tarjeta Platinum fue diseñada para clientes de alto nivel adquisitivo, que ya poseen otro producto de crédito en el banco y hacen un uso intensivo de su tarjeta. Internet Global es una tarjeta de crédito, de marca Visa, que permite realizar compras internacionales, por internet, efectuando el pago de las mismas hasta en 12 cuotas sin recargo. Este producto también fue diseñado para ser ofrecido, en una primera instancia, a clientes del banco.

El departamento comercial ha encargado al equipo de Ciencias de Datos del banco que efectúe un análisis de la cartera de clientes, de forma de obtener listados de aquellos clientes con mayor propensión a la adquisición de estos nuevos productos. Con estos listados, el área de Contact Center hará los llamados necesarios para la venta de los productos. Por indicación del área Comercial, no se podrán realizar contactos a aquellos clientes que hayan recibido al menos dos llamadas en los últimos 60 días.

#### Sobre los dataset disponibles

##### DatasetTrxs

Este dataset contiene datos de actividad transaccional de un conjunto de clientes. Los siguientes son los atributos del archivo:

CUSTID	Identificador del cliente
BALANCE	Monto del saldo disponible para realizar compras
BALANCEFREQUENCY	Frecuencia de actualización del saldo, puntuación entre 0 y 1 (1 = actualizado con frecuencia, 0 = no actualizado con frecuencia)
PURCHASES	Importe de las compras realizadas
ONEOFFPURCHASES	Monto máximo de una compra realizada de una sola vez
INSTALLMENTSPURCHASES	Cantidad de compras hechas en cuotas
CASHADVANCE	Cantidad de retiros en efectivo

PURCHASESFREQUENCY	Frecuencia de las compras (1 = frecuentemente, 0 = no frecuentemente)
ONEOFFPURCHASESFREQUENCY	Compras frecuentes son realizadas de una sola vez (1= frecuentemente, 0= no frecuentemente)
PURCHASESINSTALLMENTSFREQUENCY	Frecuencia de compras en cuotas (1=frecuentemente, 0=no frecuentemente)
CASHADVANCEFREQUENCY	Frecuencia de retiros en efectivo (1=frecuentemente, 0=no frecuentemente)
CASHADVANCETRX	Cantidad de transacciones de retiro en efectivo
PURCHASESTRX	Cantidad de transacciones
CREDITLIMIT	Limite de crédito
PAYMENTS	Cantidad de pagos efectuados por el usuario
MINIMUM_PAYMENTS	Cantidad de pagos mínimos hechos por el usuario
PRCFULLPAYMENT	Porcentaje de pagos totales hechos por el usuario
TENURE	Tenencia de la tarjeta de crédito (meses)

#### DatasetContactos

CUSTID	Identificador del cliente
PHONE	Teléfono de contacto
CONTACTS	Cantidad de llamados que recibió en los últimos 60 días

**Nota:** Los archivos son generados por diferentes sectores del banco, por lo que pueden tener diferentes formatos.

#### Parte 1

Efectúe un profundo EDA de los datasets entregados por el área de Operaciones Crediticias. Documente todas las observaciones que haga de los datos.

#### Parte 2

Haga la limpieza necesaria de los datasets, considerando datos nulos y outliers. Justifique cada una de las decisiones tomadas.

#### Parte 3

Utilizando el algoritmo k-means efectúe una clusterización del dataset entregado. Detecte aquellos clusters relevantes para el problema, describiéndolos en términos de

negocio. Debe de graficar el resultado de la clusterización, se recomienda el uso de PCA para bajar la dimensionalidad de los datos.

Utilice las técnicas que considere necesarias para determinar la cantidad óptima de clusters, justificando claramente las decisiones tomadas.

#### Parte 4

En sus propias palabras, describa las técnicas de clustering basadas en densidad, en particular el algoritmo DBSCAN. Aplique DBSCAN al dataset y compare los resultados con los obtenidos en la parte 3.

#### Parte 5

Estudie y explique el método de Ward utilizado en clustering aglomerativo, en particular, la implementación de sklearn.

#### Parte 6

Planificando próximas etapas, el banco requiere armar un cuestionario para nuevos clientes, de forma de poder ofrecer el producto a usuarios de los cuales no cuenta con histórico transaccional. Para esto, el Departamento Comercial solicita que se construya, a partir del análisis de datos, el listado de preguntas y un modelo predictivo que en función de las respuestas retorne la propensión de compra de los productos. En caso de que entienda que no es posible cumplir con alguno de estos requerimientos, documente los motivos que lo llevaron a tomar esa conclusión.

#### Parte 7

Ensaye una discusión general del trabajo realizado, haciendo los comentarios y recomendaciones que considere necesarias.

#### Consideraciones generales

- Como ambiente de trabajo puede utilizar Jupyter Notebooks o Google Colab
- Si usa Jupyter Notebooks, el lenguaje de programación seleccionado debe ser Python
- Se recomienda fuertemente el uso del paquete scikit-learn
- Cada grupo debe entregar
  - Notebook
  - En caso de que documente fuera del notebook, entregar el documento anexo en formato PDF
  - Siempre incluya número de grupo, junto con el nombre de cada uno de los estudiante que lo integran
- **El plazo de entrega vence el sábado 23 de octubre a las 23:59.** La entrega se efectuará vía Moodle. **La defensa se llevará a cabo el martes 26 de octubre.**

