**PRUEBA PRÁCTICA**

Gracias por su interés en participar en la convocatoria para pertenecer a la Gerencia Analítica Personas.

Se buscan personas comprometidas, analíticas, con buena actitud de servicio y que se destaquen por realizar un trabajo oportuno y de calidad… además, personas con unos sólidos conocimientos en técnicas de modelación, estadística, matemáticas, minería de datos y bases de datos y como si fuera poco, capaces de definir adecuadamente un problema, generar las hipótesis, asumir supuestos, obtener conclusiones poderosas y comunicar los resultados de los modelos de una forma práctica y amigable para el usuario final, de tal forma que se asegure que los resultados de los modelos se utilicen en la estrategia del banco.

El propósito de esta prueba es medir sus capacidades para desarrollar modelos predictivos, realizar supuestos, filtrar y utilizar información relevante, concluir y comunicar adecuadamente los resultados de los modelos.

Puede usar cualquier herramienta (R, Python, SAS Guide/Miner/, SPSS, etc.) y cualquier tipo de modelación o técnica.

El trabajo es personal. **No** se permite intercambiar con otros participantes en el proceso o acudir a ayuda o asesoría externa.

Pueden realizar los supuestos que consideren necesarios. No es necesario utilizar todas las bases o todas las variables. Esto depende de la forma en que usted aborde el problema. No hay una solución única. Inclusive puede darse el caso en que no se tenga un modelo viable.

**CONTEXTO:**

Las Tarjetas de Crédito son un medio de pago en donde al cliente se le asigna un cupo de crédito que se empieza a consumir haciendo compras en establecimientos de comercio o bien retirando dinero (avances). Al momento de la compra o del avance, al cliente se le pregunta a cuantas cuotas desea diferir dicha transacción, pudiendo escoger entre 1 y 36 meses. Cada mes se debe pagar lo correspondiente a intereses y abonos a capital. A medida que va pagando, se empieza a liberar de nuevo el cupo de crédito de tal forma que el cliente pueda seguir utilizando su TDC. Cuando las TDC utilizan todo su cupo de crédito, no es posible seguir transando con ellas por lo que el cliente debe recurrir a solicitar un aumento de cupo o bien abonar capital a la deuda para liberar cupo. Algunas tarjetas de crédito tienen atados cuotas de manejo (generalmente se pagan trimestralmente) y algunos cargos recurrentes (débitos automáticos). Igualmente, hay algunas tarjetas de crédito que el cliente no mueve (no transa) ya que muchos clientes las mantienen no como medio de pago sino como un medio para hacer frente a alguna eventualidad.

Mas información acerca de las Tarjetas de crédito se puede consultar en la página www.grupobancolombia.com

**OBJETIVO DEL TRABAJO:**

Se le han proporcionado las siguientes bases:

* **ESTADOTDCyyyymm.csv:** Esta base ilustra el estado en el que se encuentra la tarjeta en la fecha yyyymm (Fin de mes).
* **TRXyyyymm.csv:** En este archivo se encuentran los movimientos de las TDC en el mes yyyymm.
* **SISTINFyyyymm:** En este archivo se encuentran características socio demográficas, tenencias de productos y de saldos consolidados por cliente en todo el banco en la fecha yyyymm (a fin de mes).

Con estas bases se pide construir un modelo de deserción que calcule la probabilidad que un cliente cancele su TDC en los meses de Noviembre2015, Diciembre2015 o Enero del 2016.

Para esto se le ha proporcionado en el archivo [Resultadosericasanchez@hotmail.com](mailto:Resultadosericasanchez@hotmail.com) en la hoja resultados las TDC vigentes al 31 de Octubre del 2015. La idea es que en la columna Prob se ingrese la probabilidad de que cada tarjeta sea cancelada entre el periodo Noviembre2015, Diciembre2015 o Enero del 2016. Solo se les debe calcular probabilidad a las TDC que hay listadas en la hoja resultados.

Esta probabilidad se debe poner en un archivo que se llame Resultadosxxxxx donde xxxxx es su mail. La información se debe poner en la hoja Resultados conservando el mismo formato y los títulos de las columnas. En una hoja del archivo se anexa un ejemplo de la manera como se debe diligenciar el archivo (ver archivo [Resultadosericasanchez@hotmail.com](mailto:Resultadosericasanchez@hotmail.com) en la hoja Ejemplo). Solo se debe enviar el archivo con los resultados al correo [erisanch@bancolombia.com](mailto:erisanch@bancolombia.com) y al correo [ericasanchez@hotmail.com](mailto:ericasanchez@hotmail.com)

**EVALUACION DEL MODELO:**

Los modelos que ustedes envíen se les calculará el estadístico AUC (Area Bajo la Curva) y se seleccionaran los modelos que tengan mejores AUC.

**Como calculo la AUC?**

La base que ustedes me están enviando con la probabilidad de deserción corresponden a las TDC que estaban vigentes el 31 de Octubre del 2015.

Cada una de estas TDC vigentes al 31 de Octubre del 2015 yo  les consulta el estado que tenían tres meses después (es decir al 31 de Enero del 2016).

Las tarjetas que se cancelaron en este periodo de tiempo (en Noviembre, en Diciembre o en Enero)  son las cancelaciones verdaderas (1) y las que siguen vigentes son TDC no canceladas (0).

Luego con esta información y las probabilidades que ustedes me envían calculo el estadístico AUC. Este estadístico nos mide que tan bien está pronosticando (clasificando) el modelo. Es decir, mide que las TDC que tengan una alta probabilidad sean las que realmente se cancelan en una ventana de tiempo de tres meses y las que tengan menores probabilidades sean las que no se cancelan.

Para que tengan un punto de comparación, lo ideal es que el AUC se mueva en el rango entre 0.5 y 1. Mientras más cercano se encuentre al 1 es mejor. Una AUC cercana a 0.5 o menor significa que el modelo no clasifica bien (es mejor coger al azar una TDC y adivinar si se va a cancelar o no).

**DESCRIPCION FUENTES DE DATOS:**

**ESTADOTDCyyyymm:**

* Cedula1: Nro. de identificación del Tarjeta Habiente.
* Franquicia: Amex, MasterCard, Visa.
* Tipo Tarjeta: Representa el tipo de tarjeta. Se agrupan en normales (clásicas, azul), oro (oro, verde, dorada) y las Platinum y las especiales (Ecard, Ideal).
* Número Tarjeta1: Indica el número de la Tdc.
* Cupo Global: Indica el cupo de crédito que tiene la tarjeta. (Pesos)
* Disponible: Indica del cupo global cuanto tiene a ese corte disponible para realizar compras o avances.(Pesos)
* Fecha Emisión: Fecha en la que se compró la tarjeta.
* Fecha Ult Aumento Cupo: Fecha en la que se realizó el último aumento de cupo. Cuando esta en 0 significa que nunca se ha realizado un aumento de cupo.
* Total Mes en Mora: Número de meses que ha estado en mora la tdc (durante toda la vida de la tarjeta).
* Altura de Mora: Indica la mora actual que tiene la tdc. Viene en rango: 0= No está en mora, 30= está en mora hace menos de 30 días, 60= está en mora hace menos de 60 días. 90= está en mora hace menos de 90 días.
* Código Vendedor: Indica el código del vendedor que emitió la TDC.
* Código Estado Tarjeta: Indica si la tarjeta está Activa o vigente (Codigo=0) y código= 3000000000 (indica que la tarjeta está cancelada por el cliente).

**TRXyyyymm:**

* Cedula1: Nro de identificación del Tarjeta Habiente dueño de la tdc.
* Número Tarjeta1: Indica el número de la Tdc que realizó la trx.
* Fecha de transacción: Indica la fecha en que se realizó la trx.
* Codigo Transacción:

03 AVANCE (El cliente retira dinero en efectivo de un cajero o de una sucursal)

04 ABONO (Pago mensual que hace. Incluye capital + intereses + Cuota Manejo)

06 COMPRA NACIONAL: Compra que se hace en un establecimiento de Colombia. Se encuentra en Pesos.

07 COMPRA INTERNAL: Compra que se hace en un establecimiento internacional. Para Visa se encuentra en pesos y para Amex y Master Card se encuentra en dólares.

69 CUOTA DE MANEJO: Es la cuota de manejo que paga el cliente. Generalmente se liquida trimestralmente.

* Valor Transacción: Indica el valor de la transacción.
* Numero Cuotas Diferidas: Indica el número de cuotas en el que el cliente desea pagar la compra. Generalmente el cliente puede escoger entre 0 (el cliente paga la totalidad de la compra al siguiente mes) y 36 (el cliente va a pagar la compra en 36 meses).

**SISTINFyyyymm:**

* Cedula1: Nro. de identificación del Tarjeta Habiente.
* Sexo: M=Masculino, F=Femenino
* Estado del Cliente: Indica el estado del cliente en el banco. Activo (El cliente ha usado al menos uno de sus productos en los últimos 3 meses). Inactivo (El cliente No ha usado ninguno de sus productos en los últimos 3 meses).
* Fecha de Nacimiento. Fecha de nacimiento del cliente. Edad del cliente.
* Fecha de Vinculación. Indica la fecha en que el cliente se vinculó a Bancolombia. Hace cuánto es cliente del Banco.
* Valor Ingresos: Los ingresos reportados por el cliente como salarios. (Esta en Pesos).
* Estado Civil: Casado, Viudo, Divorciado, Soltero, etc..
* Nivel Académico: Los más comunes son Primario, Bachillerato, Universidad, Tecnología, Especialización.
* Profesión: Código de la profesión del cliente.
* Ocupación: Las más comunes son empleado, jubilado, estudiante, ama de casa, etc..
* Es\_Nómina\_Bancolombia: Indica si el cliente recibe la nómina por medio de la cuenta Bancolombia.
* Ciudad: Ciudad de residencia del cliente.
* Ivc Cantidad Portafolio: Indica el número total de productos que tiene el cliente con el banco.
* Ivc Tiene Grupo: Indica el número de grupos de productos que tiene el cliente con el Banco.

Por ejemplo, el cliente puede tener 3 cuentas de ahorro y 2 tarjetas de crédito. En IVC Portafolio serían 5 productos y en IVC Grupo tendría 2.

* Tamaño Comercial: Es la suma de todos los saldos de los productos de captación (CDT, Cuenta de ahorro, cuenta corriente, fondos de inversión) y de Cartera (Cartera de consumo, Cartera Hipotecaria, Cartera TDC, etc..). Se encuentra en miles de pesos.
* Total Cartera: Es la suma de todos los saldos que el cliente tiene en cartera con Bancolombia (Cartera de consumo, Cartera Hipotecaria, Cartera TDC, etc..)
* Total Captaciones: Es la suma de todos los saldos que el cliente tiene en ahorros (CDT, Cuenta de ahorro, cuenta corriente, fondos de inversión)
* Max Mora: Es la mora máxima que el cliente ha tenido en los últimos doce meses con todos sus productos de riesgo con Bancolombia.
* Resultado Consulta BIBIC: Indica si el cliente pasa el score de riesgo de Bancolombia. N= No pasa el score de riesgo y por lo tanto al cliente no se le puede prestar y vacío o S indica que al cliente si se le puede prestar dinero. Indica el comportamiento financiero del cliente con sus productos de riesgo con Bancolombia.
* Score: Es el score de riesgo que tiene el cliente con el sistema financiero. A mayor score, menor probabilidad de pérdida. Un score superior a 550 es un score aceptable. Este score refleja el comportamiento de crédito del cliente con todos los productos financieros que tenga en el sistema financiero.
* Saldo No Rotativo Entidad: El saldo de los productos de crédito (sin incluir tdc) que el cliente tiene con Bancolombia. (Miles de pesos)
* Saldo No Rotativo Mcdo: El saldo de los productos de crédito (sin incluir tdc) que el cliente tiene con entidades diferentes a Bancolombia. (Miles de pesos)
* Saldo Prom Tdc Entidad (Trim): El saldo de las TDC que el cliente tiene con Bancolombia. (Miles de pesos)
* Saldo Prom Tdc Mcdo (Trim): El saldo de las TDC que el cliente tiene con entidades diferentes a Bancolombia. (Miles de pesos)
* Cupo Tdc Entidad: Es el cupo de TDC que el cliente tiene con Bancolombia. (Miles de pesos)
* Cupo Tdc Mcdo: Es el cupo de TDC que el cliente tiene con entidades diferentes a Bancolombia. (Miles de pesos)

**ARCHIVOS**

En el siguiente enlace se encuentran todos los archivos que se requieren para la realización del ejercicio. Copiarlo y pegarlo en el navegador para acceder a las fuentes de información.

<https://drive.google.com/folderview?id=0BzgTckwEAkdVSnhCWm41MjZfS3c&usp=sharing>