

Desafío 4 - Proyecto integrador

Modelo clasificador de venta de tarjetas de crédito

Grupo 8: Bozzano Santiago, Pérez Curiel Joaquín, Devoto Marcos, Florentín Alejo, Hernández María Paz y Lopera Erazo José Alejandro.

1) Tema de investigación

El objetivo de este proyecto es desarrollar y poner en producción un modelo de clasificación que busca identificar a los posibles clientes (prospectos de acá en adelante) que más probabilidad tienen de aceptar el producto tarjeta de crédito (TC) ante una campaña dirigida de llamadas salientes por parte de la red de sucursales de la entidad bancaria. De esta forma esperamos poder llevar adelante campañas de venta más asertivas y a su vez, en caso de necesidad (como puede ser el cumplimiento del plan comercial), hacer pre-embozados (imprimir tarjeta antes de concretar venta) y enviarlos al domicilio asegurándose una alta tasa de aceptación del plástico (venta).

El dataset con el que trabajaremos contiene información socioeconómica, de bancarización y el resultado de la gestión de los 12.712 prospectos que fueron precalificados y gestionados para la venta de una TC. Los features con los que cuenta el dataset son los siguientes:

- **Venta:** 1 indica venta del producto y 0 el rechazo. Es nuestra variable target.
- **Edad.**
- **Masculino:** 1 masculino, 0 femenino.
- **Capital:** 1 si el prospecto vive en la capital de la provincia, 0 si vive en el interior.
- **Oferta:** margen/límite crediticio ofrecido.
- **NSE:** nivel socioeconómico. Puede asumir los valores D1, C3, C2, C1, B, A (de menor a mayor).
- **Actividad:** categoría de inscripción en AFIP según facturación anual. Las categorías pueden ser: monotributista (A a K), autónomo, informal (no registrado) o pasivo (jubilado).
- **Empleador:** 1 si tiene contratado al menos un empleado, 0 si no tiene.
- **Jubilado:** 1 jubilado, 0 activo (no jubilado).
- **Cant_bcos:** cantidad de entidades financieras con las que tiene deuda vigente en algún producto activo como TC o préstamo.
- **Monto_bcos:** monto adeudado a todas las entidades financieras con las que tiene deuda vigente.
- **Banc_12m:** indica en cuántos de los últimos 12 meses (es decir, en el último año) tuvo deuda con alguna entidad financiera.
- **Banc_13_24:** indica en cuántos de los meses 13 al 24 para atrás (es decir, hace dos años) tuvo deuda con alguna entidad financiera.
- **Banco_1:** indica cual es la principal entidad con la que tiene deuda vigente. Puede asumir el nombre de la entidad en caso de ser un banco, Financiera en caso de ser una entidad no bancaria (excepto Naranja, principal competidor del banco), Naranja, Mutual, Retail y 0 en caso de no tener deuda vigente con ninguna entidad.

2) Antecedentes sobre el tema

Durante la investigación de antecedentes constatamos que la literatura relacionada con machine learning, modelos de clasificación y sistema bancario, se enfocan principalmente en la detección de fraude en TC, utilizando para esto herramientas y análisis estadístico para el reconocimiento de patrones de comportamiento fraudulento. Para ello la mayoría de los sistemas de detección actuales ofrecen dos tipos de alerta: por calificación probabilística y por cumplimiento de reglas, donde en el primer tipo de alerta lo más usual son los modelos predictivos para una calificación "score".

Otra rama de la literatura se enfoca en el análisis y predicción de modelos para estudiar la tasa de "churn" (cancelación/baja) por parte de clientes de tarjetas de crédito (u otros productos activos).

Bibliografía estudiada:

- Fraude: MACHINE LEARNING APLICADO A DATOS FINANCIEROS: http://opac.pucv.cl/pucv_txt/txt-7500/UCC7987_01.pdf
- Desertores: Customer Churn Analysis and Prediction Using Data Mining Models in Banking Industry: <https://ieeexplore.ieee.org/document/8935884>
- Compra de producto / servicio: MODELO DE PROPENSIÓN A LA TOMA DE CRÉDITOS DE CONSUMO EN UNA EMPRESA DEL SECTOR FINANCIERO: <https://repositorio.udd.cl/bitstream/handle/11447/4100/Modelo%20de%20propensi%C3%B3n%20a%20la%20toma%20de%20cr%C3%A9ditos%20de%20consumo.pdf?sequence=4&isAllowed=y>

3) Aporte esperado

En este proyecto buscamos aplicar técnicas de machine learning para desarrollar un modelo que permita predecir la propensión de un prospecto a aceptar una oferta de TC de la entidad bancaria.

Con este modelo buscamos identificar y evaluar cuales son las variables relevantes que inciden sobre nuestro target para así conocer el perfil del prospecto tomador del producto, optimizando la selección de los mismos, aumentando las tasas de aceptación del producto e incrementando la cartera de clientes del banco, así como también disminuirán los costos de adquisición de bases (dado que será más asertiva su compra) y de gestión y pre-embozados (ya que no se trabajaran prospectos con baja propensión a tomar el producto).

4) Disponibilidad de datos e infraestructura

El dataset está conformada por 12.712 prospectos precalificados y gestionados, enriquecido por el bureau de crédito (el cual tiene como soporte las fuentes más confiables de la Argentina como es la AFIP, ANSES, Boletines Oficiales, INPI y el Banco Central de la República Argentina) con el que la entidad trabaja y por la variable target.

La variable target que utilizaremos en nuestro modelo es la venta o rechazo del producto (1 y 0 respectivamente). El resultado "Venta" se obtuvo de la validación del alta del producto en los sistemas internos de la entidad, por lo tanto, es una venta efectivamente consumada, mientras que el resultado

“Rechazo” surge de la gestión de las bases de prospectos enviadas a la red de sucursales, es decir, es la respuesta informada por el ejecutivo de la sucursal que logró el contacto con el prospecto.

5) Plan de trabajo y cronograma tentativo

TAREA / SPRINT	Sprint 1: 16/12/21	Sprint 2: 23/12/21	Sprint 3: 13/01/22	Sprint 4: 20/01/22	Sprint 5: 27/01/22	Sprint 6: 10/02/22
Búsqueda y limpieza del data set						
Identificación de features y relaciones entre variables						
Elección e implementación de modelo de ensamble						
Tuneo de hiperparámetros						
Pipeline y análisis de features						
Utilización de la librería Pickle						
Finalización de notebook y desarrollo de presentación						