



Tarjetas de Crédito:

Un caso de Clasificación

Escrito por Eduardo Aguilar, UCA.

Este caso está basado en el [repositorio](#) de la Universidad de California Irvine UCI, [kaggle](#) y la solución encontrada en el sitio oficial de [INSEADAnalytics](#).

Crédito

Nuestro mundo moderno depende del crédito. La economía entera está basada en la disponibilidad de las personas de “comprar ahora – pagar después”. En efecto, hace doscientos años prestar dinero era un privilegio, pero en la sociedad industrializada de ahora es casi considerado un derecho. Pero proveer de dinero es un negocio riesgoso, ya que los prestamistas difieren en su capacidad de pago. Por ello es importante desarrollar una metodología que nos ayude a diferenciar entre personas que son potencialmente sujetos de crédito de las que no lo son.

El caso de Taiwán

Hace poco más de una década, los emisores de tarjetas de crédito en Taiwán experimentaron una crisis de morosidad en las tarjetas de crédito. En el año 2016, con el propósito de incrementar la cuota de mercado, los bancos establecidos en Taiwán emitieron un exceso de tarjetas de crédito a aplicantes que eran muy riesgosos. Al mismo tiempo, muchos tarjetahabientes utilizaron de forma irresponsable su crédito y acumularon una deuda demasiado grande para poder cancelarlas. La crisis causó un derrumbe en la confianza del consumidor Taiwanés.

Para evitar este problema, las compañías de crédito intentan predecir potenciales defaults, o evaluar la probabilidad de default de un cliente. Por esta razón, es necesario predecir de forma efectiva potenciales defaults y minimizar las pérdidas.

La alta gerencia decidió suspender la emisión de tarjetas de crédito y préstamos para **consumo personal temporalmente**, y encargó a José, el jefe de riesgos financieros, la tarea de definir una estrategia que pueda reducir la tasa de defaults en los créditos otorgados lo más pronto posible, sin reducir significativamente la emisión de tarjetas.

José recientemente tuvo una capacitación acerca de cómo los modelos predictivos pueden ayudar a alcanzar este objetivo. La técnica que él aprendió se llama regresión logística, que era en esa época el modelo más popular y ampliamente aceptado en el sector bancario por su transparencia y facilidad de interpretabilidad. Por lo cual decidió construir un modelo basado en ello.

Los Datos

Usted puede encontrar los datos originales en este el repositorio de [UCI](#).

Un emisor de tarjetas de crédito reunió información de 30,000 consumidores, recopilando datos con información de 24 variables, incluyendo factores demográficos, datos de crédito, historial de pagos y estados de cuenta de tarjetas de crédito desde abril 2005 a septiembre 2005, así como información del resultado: ¿El cliente pagó o no? Las variables se detallan a continuación:

1. ID: Identificador del cliente
2. LIMIT_VAL: Monto del crédito en dólares
3. SEXO
4. EDUCACION: 1=postgrado, 2=universidad, 3=bachillerato, 4=otros, 5=desconocido, 6=desconocido
5. MARRIAGE: Estatus Marital (1=casado, 2=soltero, 3=otros)

6. EDAD
7. PAY_0: Estatus de pago en septiembre 2005 (-2=no consumo, -1=paga debidamente, 0=usa crédito rotativo, 1=atraso de un mes, 2=atraso de dos meses, ... 8=atraso de ocho meses, 9=atraso de nueve meses y más.
8. PAY_2: Estatus de pago en agosto 2005
9. PAY_3: Estatus de pago en julio 2005
10. PAY_4: Estatus de pago en junio 2005
11. PAY_5: Estatus de pago en mayo 2005
12. PAY_6: Estatus de pago en abril 2005
13. BILL_AMT1: Monto a pagar septiembre 2005
14. BILL_AMT2: Monto a pagar agosto 2005
15. BILL_AMT3: Monto a pagar julio 2005
16. BILL_AMT4: Monto a pagar junio 2005
17. BILL_AMT5: Monto a pagar mayo 2005
18. BILL_AMT6: Monto a pagar abril 2005
19. PAY_AMT1: Monto de pago anterior septiembre 2005
20. PAY_AMT2: Monto de pago anterior agosto 2005
21. PAY_AMT3: Monto de pago anterior julio 2005
22. PAY_AMT4: Monto de pago anterior junio 2005
23. PAY_AMT5: Monto de pago anterior mayo 2005
24. PAY_AMT6: Monto de pago anterior abril 2005
25. DEFAULT: 1=sí, 0=no

El análisis

José decidió tomar todas las características disponibles para intentar construir un modelo predictivo que pudiera mejorar la gestión crediticia del banco. Su objetivo era encontrar las respuestas a las siguientes preguntas:

1. ¿Cuáles son los factores clave que pueden ayudar a determinar que clientes entrarán en default?
2. ¿Cuál sería el desempeño estimado de un modelo que aprobara el 60%, 70% y 80% de las solicitudes?
3. Dado que el costo de no predecir correctamente un default es -100,000 y la ganancia por predecir correctamente un cliente que no entra en default es +20,000, ¿cuál sería el porcentaje de aprobación óptimo?