
Aceptación de préstamos personales

Universal Bank es un banco relativamente joven que crece rápidamente en términos de adquisición general de clientes. La mayoría de estos clientes son clientes pasivos (depositantes) con diferentes tamaños de relación con el banco. La base de clientes activos (prestatarios) es bastante pequeña y el banco está interesado en expandir esta base rápidamente para atraer más negocios de préstamos. En particular, desea explorar formas de convertir a sus clientes pasivos en clientes de préstamos personales (mientras los retiene como depositantes).

Una campaña que el banco realizó el año pasado para clientes pasivos mostró una tasa de conversión saludable de más del 9% de éxito. Esto ha alentado al departamento de marketing minorista a diseñar campañas más inteligentes con un marketing más dirigido. El objetivo es utilizar k-NN para predecir si un nuevo cliente aceptará una oferta de préstamo. Esto servirá como base para el diseño de una nueva campaña.

El archivo UniversalBank.csv contiene datos sobre 5000 clientes. Los datos incluyen información demográfica del cliente (edad (*age*), ingresos (*income*), etc.), la relación del cliente con el banco (hipoteca (*mortgage*), cuenta de valores (*securities account*), etc.) y la respuesta del cliente a la última campaña de préstamos personales (Préstamo personal, *personal loan*). Entre estos 5000 clientes, solo 480 (= 9.6 %) aceptaron el préstamo personal que se les ofreció en la campaña anterior.

Divide los datos en conjuntos de entrenamiento (60 %) y validación (40 %).

Considera el siguiente cliente:

- Age = 40, Experience = 10, Income = 84, Family = 2, CCAvg = 2, Education_1 = 0, Education_2 = 1, Education_3 = 0, Mortgage = 0, Securities Account = 0, CD Account = 0, Online = 1 y Credit Card = 1. Realiza una clasificación k-NN con todos los predictores excepto el ID y el código postal (ZIP code) utilizando $k = 1$. Recuerda transformar primero los predictores categóricos con más de dos categorías en variables ficticias. Especifica la clase de éxito como 1 (aceptación del préstamo) y utiliza el valor de corte predeterminado de 0.5. ¿Cómo se clasificaría a este cliente?
- ¿Cuál es una opción de k que equilibre entre el sobreajuste y la ignorancia de la información del predictor?
- Muestra la matriz de confusión para los datos de validación que resultan de utilizar el mejor k .

- d. Considera el siguiente cliente: Age = 40, Experience = 10, Income = 84, Family = 2, CCAvg = 2, Education_1 = 0, Education_2 = 1, Education_3 = 0, Mortgage = 0, Securities Account = 0, CD Account = 0, Online = 1 y Credit Card = 1. Clasifica al cliente utilizando el mejor k.
- e. Reparticiona los datos, esta vez en conjuntos de entrenamiento, validación y prueba (50% : 30% : 20%). Aplica el método k-NN con el k elegido anteriormente. Compara la matriz de confusión del conjunto de prueba con la de los conjuntos de entrenamiento y validación. Comenta las diferencias y su razón.