

Aplicaciones de Machine Learning: Evaluación Final

Gonzalo Lezma

October 2020

AutoBank Technologies Inc. es un neobanco fintech que ofrece créditos vehiculares a sus clientes. Estos clientes entran a la web de Autobank y pueden pedir un préstamo vehicular. Cuando el cliente quiere pedir un crédito vehicular entra a la web e ingresa su DNI y datos relativos a su solicitud; por ejemplo: qué tipo de auto se quiere comprar, cuál es el plazo (número de años) en el que quiere pagar el crédito, y el monto que se quiere prestar. Inmediatamente, en cuestión de milisegundos, la web de Autobank consulta internamente el Score de riesgo del cliente en una base de datos interna, la tasa que esta cobrando la competencia por creditos similares, y el costo de fondos del banco. En base a toda esa información, el servidor web calcula una tasa ofertada hacia el cliente y se la muestra en pantalla. Esta tasa es la tasa de interés que pagaría el cliente de aceptar tomar el crédito vehicular.

En el archivo [RdProData.csv](#) se encuentran las siguientes variables.

| Lista de variables | |
|--------------------|---|
| Variable | Descripción |
| FICO | Score o puntuación de riesgo. Mayor score significa un riesgo mas saludable por lo tanto la probabilidad de que es un cliente buen pagador del crédito es alta. Menor score significa un riesgo menos saludable, es decir la probabilidad de que el cliente pague el crédito al banco es baja |
| Tier | Variable FICO discretizada en 4 segmentos |
| Approve | Fecha en que el usuario esta ingresando a la web que es igual a la fecha en que se aprueba el credito |
| Date | Plazo del crédito vehicular |
| Term | Monto que se el cliente solicita en su solicitud de préstamo |
| Amount | Tipo de auto para el que se solicita el préstamos |
| Car Type | Tasa de interés que esta cobrando la competencia |
| Competition | Variable binaria que es igual a 1 si el cliente decide aceptar la oferta del banco y pagar la tasa ofrecida. Si la variable es 0 significa que el cliente no acepta el credito a la tasa ofertada |
| Rate | Tasa de interes que paga el banco cuando levanta fondos del mercado profesional |
| Accept? | Partner automotriz que internamente tiene el banco para financiar autos nuevos |
| Cost of funds | Tasa ofertada por el banco |
| Partner | |
| Rate | |

1 Analítica descriptiva (5 pts)

Se pide realizar un análisis descriptivo de las variables (Se debe mostrar como mínimo las distribuciones y correlaciones de las variables). Sugerencia: Utilizar el comando `pairs` visto en clase. Explique que variables tienen correlación positiva con la aceptación de créditos y proponga una interpretación para una potencial causalidad. Por ejemplo: ¿Cuál es la correlación entre la tasa de interés ofertada por el banco (propuesta) y la aceptación de los clientes? ¿A que se puede deber este resultado?

2 Segmentación de clientes (5 pts)

Actualmente el personal de la empresa ha agrupado la variable FICO en cuatro Tiers y esta segmentando únicamente en base a esta variable. ¿Podría Ud. sugerir una mejor segmentación de los clientes en base a las demás variables disponibles? (Sugerencia: Utilice k-means para clusterizar a los clientes con diferentes variables y proponga la clusterización que mas se ajuste con su criterio de negocio)

3 Estimaciones de disposición de pago del cliente (5 pts)

Utilice la información dada en el archivo csv para realizar una estimación de la disposición de pago de los clientes. Sugerencia: Desarrolle un modelo en donde la variable dependiente sea `Accept`). ¿Cuál de los métodos vistos en clase (regresión, logístico, CART y Random Forest) predice mejor la disposición a tomar el crédito? Realice un boxplot para poder comparar los modelos de manera visual (Usar los comandos `"resample"` y `"bwplot"` vistos en clase)

4 AB Testing (5pts)

Los ingenieros de Autobank han desarrollado un chatbot como alternativa que aparentemente podría ser mejor para solicitar los créditos de los clientes por su facilidad y naturalidad de uso. El equipo de marketing analytics ha decidido probar si esta hipótesis es cierta con datos. Bajo los supuestos usuales ($\alpha = 0.05$, Poder de la prueba 0.8 y $B=0.5$) ¿cuál debe ser la cantidad de clientes que se deben repartir entre el chatbot y la página web para poder detectar una mejora de 0.1 (10%) en la probabilidad de tomar el crédito?

5 Opcional (para vuestra diversión)

El departamento de marketing no esta seguro del lift que quiere detectar con el AB testing, además, le han dicho que este tipo de pruebas se van a realizar con frecuencia. Tomando en cuenta esto, su jefe le pide sugerencias para poder hacer que estos cálculos se realicen sin necesidad de conocer R y desde una página web para que todo el equipo de marketing pueda saber fácilmente cuántos clientes se necesitarían para diferentes niveles de detección de lift. Sugerencia: Despliegue en la web, una shiny app con un slider que represente el lift (por ejemplo desde 0.01 hasta 0.2) y un texto que muestre cuantos clientes se necesitan en cada experiencia (utilice el comando `renderText` para desplegar el resultado en vez del comando `renderPlot` visto en clase). Para facilidad mantenga $p_1 = 0.1$.