Randall Ramírez Solano

randallramirezs@gmail.com

Credit One report

# Introducción

Credit One, como empresa analista de riesgos para empresas que otorgan crédito, es afectada si los clientes de éstas dejan de cancelar sus créditos por encima de la probabilidad inicialmente estimada, pues es un indicador que el análisis de riesgos realizado por Credit One está fallando, y porque, si los clientes no pagan, las empresas pierden dinero y por tanto Credit One también pierde.

En el siguiente reporte se describe más ampliamente el problema, se hace un resumen del trabajo de análisis realizado y se responden las preguntas planteadas en el problema. Al final se añaden conclusiones y recomendaciones.

# El problema

Un aumento en las tasas de incumplimiento de los clientes es malo para Credit One ya que su negocio está aprobando préstamos para clientes en primer lugar. Es probable que esto resulte en la pérdida de los clientes comerciales de Credit One.

Por tanto, es necesario investigar las siguientes preguntas:

* ¿Cómo se asegura de que los clientes puedan pagar sus préstamos?
* ¿Podemos aprobar clientes con alta certeza?

Para resolver estas preguntas, se debe centrar en los problemas que se pueden resolver:

* ¿Qué atributos en los datos podemos considerar estadísticamente significativos para el problema en cuestión?
* ¿Qué información concreta podemos derivar de los datos que tenemos?
* ¿Qué métodos probados podemos usar para descubrir más información y por qué?

No se debe perder de vista las lecciones aprendidas por Credit One en ocasiones anteriores:

* No podemos controlar los hábitos de gasto de los clientes.
* No siempre podemos pasar de lo que encontramos en nuestro análisis al "por qué" subyacente

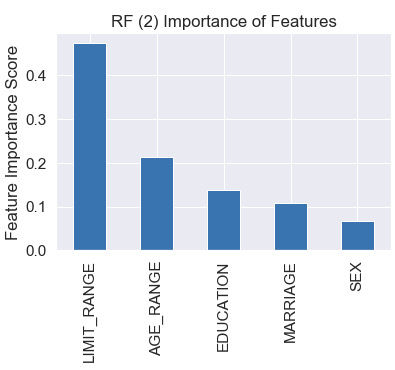
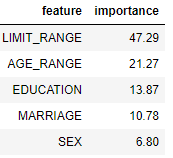
# Resultados del análisis

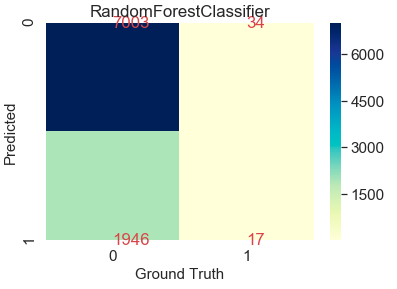
El análisis realizado reveló que existe un modelo de predicción que se ajusta con 82.7% de certeza (*accuracy*) a identificar los clientes que caen (caerán) en default. Existen 3 variables cuya importancia combinada es de un 78.8%: PAY\_0, PAY\_2 y PAY\_3, las cuales representan el estado de pago de los últimos 3 meses.

Lo anterior, aunque parece prometedor, debe analizarse en el contexto de otras consideraciones. En primer lugar, las variables más importantes a la hora de predecir el default, se obtienen hasta que el crédito ha sido otorgado y se necesita al menos 3 meses de pagos para tener el input necesario para tomar una decisión. Por tanto, **no es posible bajo ese modelo predecir con anticipación si el cliente estará o no en default ante de otorgar el crédito**. En segundo lugar, aunque la certeza (*accuracy*) del modelo es alta, no lo es el indicador Kappa 37.83%. Según algunos autores, este valor para considerarse aceptable debería ser superior al 75%. Esto significa que, aunque la certeza sea relativamente alta, **habrá una serie de malas predicciones que podría llevar a tomar decisiones equivocadas con posibles buenos clientes, o, por el contrario, a no hacer nada cuando es necesario evitar un default**.

Existen variables que son conocidas antes de otorgar crédito: SEX, MARRIAGE, EDUCATION, y AGE. Incluso se puede incluir LIMIT\_RANGE, el cual es definido al momento de otorgar el crédito. **¿es posible con estas variables predecir con certeza si un cliente estará en default y así no aprobar el crédito?**

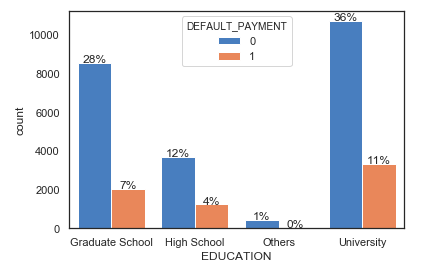
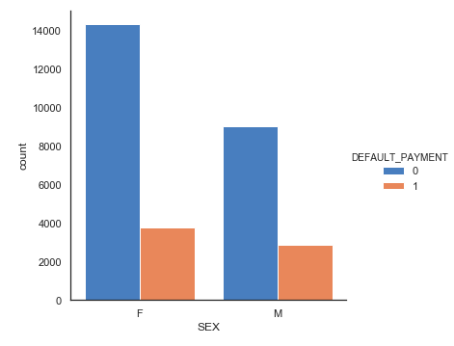
Primero debe observarse en el gráfico anterior que estas variables tienen un peso muy bajo a la hora de predecir el default. Por tanto, preliminarmente, se puede deducir que no pueden utilizarse como predictoras de default.

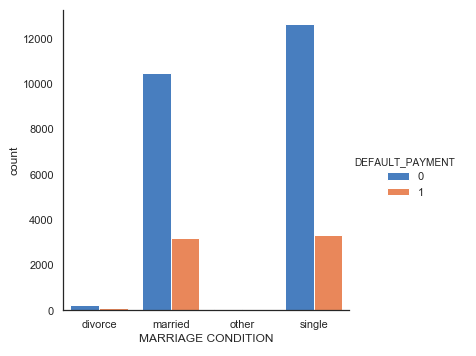
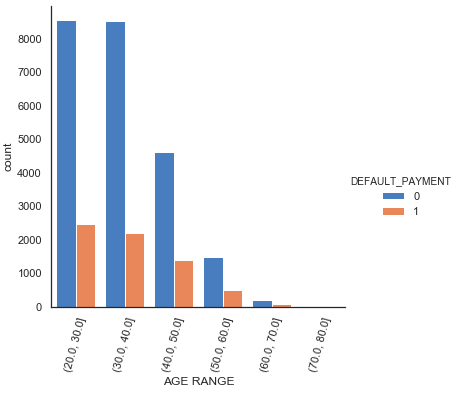
Para comprobarlo, se creó un modelo utilizando solo estas variables como predictoras. El resultado es que Límite de Crédito sea la variable más importante. Luego continúan la Edad, Educación, Condición Marital y Sexo.

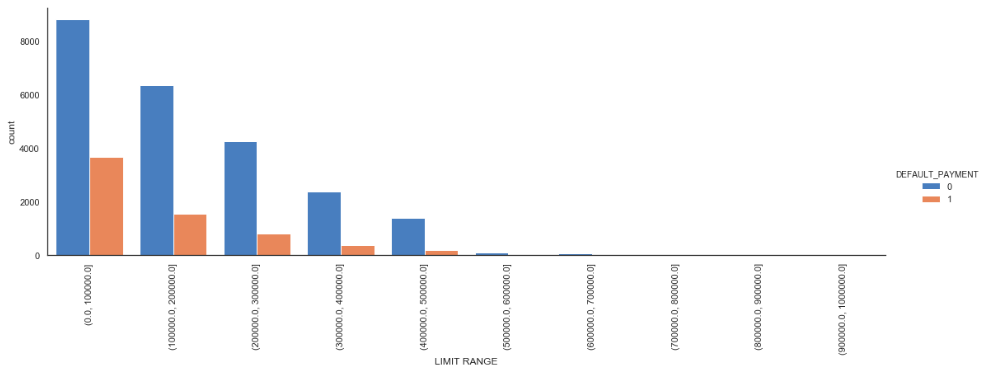
El modelo tiene una certeza (*accuracy*) de 78% lo cual es aceptable. Pero el indicador de confianza Kappa está por debajo del 1% y por tanto, **no hace confiable el modelo**.

Además, al analizar la *matriz de confusión* de las predicciones se puede observar cómo el modelo tiende a predecir clientes buenos como clientes en default (1946 casos, equivalente al 21.6%). Esto implicaría **una reducción de la cartera de clientes**, lo cual, a su vez, se traduce en **reducción de ingresos**.

Cuando se analiza la distribución de clientes en default en las variables conocidas antes de otorgar el crédito, no se identifican patrones marcados que impliquen correlación entre éstas y el default. A continuación se detalla cada caso.





En ninguno de los casos anteriores se observa una correlación fuerte entre default payment y un grupo poblacional específico ni por el límite de crédito otorgado.

Está claro que el default pareciera afectar proporcionalmente más a varones que a mujeres. La mayor parte de los afectados son con estudios universitarios o de posgrado.

En cuanto el estado marital, está claro que afecta por igual a casados y solteros. Nótese que el set de datos no tiene personas divorciadas. Esto puede ser por aspectos culturales del país de donde se obtiene la información (Taiwan) donde la tasa de divorcios no supera el 2.5%.

Finalmente, la mayor cantidad de default se encuentra en la población entre los 20 y 50 años, y proporcionalmente es mayor en la población entre 40 y 50 años. La gran concentración de default se observa en los créditos con límite de hasta $NT 100,000.

Sin embargo, a pesar de esta información, no existe un cruce de variables concreto que prediga con anticipación si la persona estará en default o no.

# Conclusiones y Recomendaciones

1. ¿Podemos aprobar clientes con alta certeza? Esta pregunta no es posible responderla antes de otorgar los créditos porque las variables conocidas en ese momento no permiten crear un modelo con la suficiente confianza. El modelo creado tiene una certeza apenas aceptable, pero las predicciones podrían negar créditos a clientes buenos provocando una disminución de la cartera de crédito y generando reducción de ingresos para los clientes de Credit One.
2. ¿Cómo se asegura de que los clientes puedan pagar sus préstamos? El modelo tiene un poco más de confianza una vez otorgado el crédito pues utiliza las variables de estado de pago, pago mensual y saldos para anticipar con base en los últimos 3 meses si el cliente entrará en default o no. Esta información podría utilizarse para crear estrategias que motiven a los clientes a pagar y así reducir el riesgo de no pago. Se recomienda evaluar alternativas en esta vía. Con los clientes
3. Basado en los resultados del análisis de los datos y de los modelos creados, se concluye que las dos principales preguntas planteadas en el problema no tienen una solución aceptable para Credit One. Por tanto, es necesario buscar alternativas para mejorar las aprobaciones de crédito y asegurar que los clientes puedan pagar sus créditos.
4. Se recomienda incorporar en el set de datos más información del análisis de aprobación de crédito realizado a los clientes para evaluar si esta información puede aportar mayor certeza en la predicción. Algunas sugerencias son: cantidad de hijos, ingreso bruto, ingreso neto, monto de otras deudas, cantidad de empleos en los últimos X años, situación de vivienda (casa propia, alquiler, otro), profesión, ocupación actual, entre otros. El modelo ha demostrado que la información actual es insuficiente y por tanto es necesario incorporar más datos para valorar relevancia.
5. Se recomienda revisar la integridad de los datos brindados en el set de datos y el significado real de la variable *default\_payment*. ¿es una predicción de default o es un default real? Esto pues se encontraron casos donde un cliente no tiene actividad o realiza sus pagos a tiempo y a pesar de esto, el conjunto de datos indica que está en default. Además, se han encontrado incongruencias entre los pagos realizados en el mes y el status de pago del mes. La falta de confianza de los modelos podría tener su fundamento en información inconsistente o datos erróneos.