

Informe del Análisis Exploratorio de Datos para la Identificación Predeterminada del Cliente

Análisis realizado para OneCredit

**Preparado por
Luis Rubén Artavia**

Descripción General

La empresa CreditOne ha ofrecido el servicio de calificación crediticia a un grupo de socios de negocio que reciben solicitudes de préstamo o tarjetas de crédito, pero en el último año ha visto un aumento en el número de clientes que han incumplido con el pago de los créditos que se les ha otorgado.

CreditOne como responsable de la calificación, ve ese comportamiento como un riesgo a perder a sus socios de negocio, por lo que ha solicitado la ayuda para implementar una solución que le ayude mejorar los procesos actuales.

Análisis del Caso

Tenemos los siguientes temas a considerar:

- Se tiene un aumento de la tasa de incumplimiento de los clientes en el pago de sus créditos
- CreditOne es la encargada de la aprobación de los créditos de los clientes
- Impacto negativo en la imagen de CreditOne
- Posible pérdida de socios de negocio para CreditOne, por la pérdida de ingresos

Inicialmente se consideró la investigación de las siguientes preguntas:

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

Dentro de las lecciones aprendidas de eventos similares que han ocurrido antes, tenemos lo siguiente:

- No es posible controlar el hábito de gasto de los clientes
- No siempre es posible pasar del resultado del análisis al “Por qué” de las cosas

Exploración de los Datos

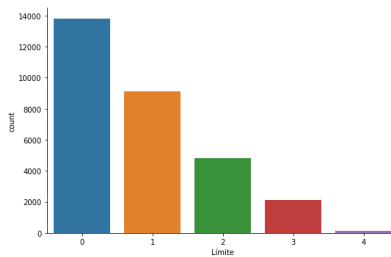
La información que fue suministrada para trabajar en la resolución del problema incluye los siguientes datos:

- Datos del perfil del cliente como género, nivel de estudio o educación, estado civil y edad
- Monto del crédito que le fue otorgado
- Seis meses de información relacionados a los créditos: clasificación del historial de pagos, monto del estado de la cuenta y el pago realizado
- El indicador del comportamiento del cliente, indicando si está debiendo o no

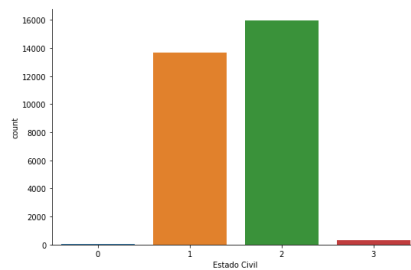
De la exploración, se considera que los atributos que pueden ser considerados como significativos para el problema que se está tratando, tienen que ver con el perfil del cliente, ya que los datos de pagos y compras no son controlables, dependen de factores externos. Por lo tanto, se utilizaron la información del estado civil, el nivel de estudio o educación, género y edad, además, se incluye el monto del crédito otorgado.

Tenemos los siguientes resultados de los datos del perfil del cliente:

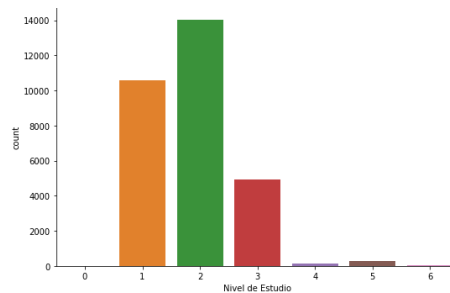
Son pocos los créditos otorgados que tienen un monto alto, aproximadamente el 75% de los créditos son menores a los 500 mil. Los rangos en la gráfica son: 10K a 208K, 208K a 406k, 406k a 604k, 604k a 802k y 802k a 1M:



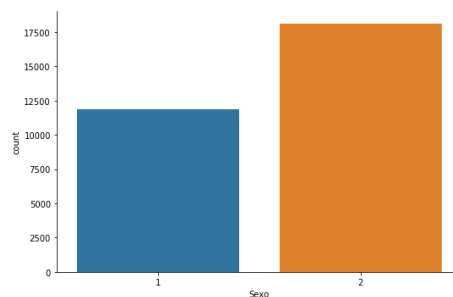
El 99% de los clientes están casados o solteros (1 – casados, 2 – solteros):



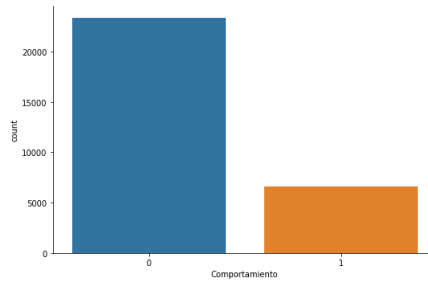
Dentro del nivel de estudio, los graduados universitarios (2) y de la escuela (1) son mayoría, los de colegio (3) son la mitad de los de la escuela:



En cuanto al género, la cantidad de clientes que son mujeres (2) sobrepasa en una tercera parte a los clientes que son hombres (1):



Para completar el análisis del perfil, tenemos que un 22% de los clientes son los que están debiendo o están atrasados con el pago de sus créditos:



Este último resultado del comportamiento general de los clientes será utilizado como base para las recomendaciones finales.

Métodos de Clasificación

Un método que puede ser utilizado para descubrir más información, es la aplicación de algoritmos de clasificación de información. Estos procesos complementarían el análisis manual que realizan los Científicos de Datos de una manera automatizada. En nuestro caso, se probaron los métodos de Support Vector Classification, Random Forest Classifier y Ridge Classifier, obteniendo el que mejor resultados en la etapa de entrenamiento el Random Forest Classifier.

Observaciones

Al realizar comparaciones de las diferentes características de los perfiles con respecto al comportamiento esperado, si está debiendo o no, tenemos las siguientes conclusiones:

- El porcentaje de clientes mujeres que están debiendo es menor que el de los hombres, además, está en un 1% por debajo del porcentaje general.
- El cliente divorciado tiene el porcentaje más alto para aquellos clientes que están debiendo, además, los casados son el segundo grupo más alto. En ambos casos están sobre el porcentaje general. Los soleros están por debajo en un 1% que el general y los otros apenas llegan al 9%.
- Para el caso del nivel de educación, tanto los universitarios y los que tienen un título de colegio, el porcentaje de clientes que están debiendo está sobre el porcentaje general, los que tiene nivel escolar está 3% por debajo y el resto alrededor del 6% están debiendo.
- Con respecto a las edades, para los clientes de 50 años o menos, el porcentaje de personas que están debiendo son similares al general o tiene un 1% menos. Para los de edades superiores el porcentaje aumenta levemente, entre el 23% y el 25%.

Se podría considerar que las mujeres usan más el servicio y son mejores paga que los hombres, que los divorciados son más propensos a dejar de pagar los créditos y los que tienen un nivel de estudio escolar tratan de no estar debiendo.