



REPORTE

Credit One

Bryan Fallas L.



RESUMEN

DESCRIPCIÓN GENERAL DEL PROBLEMA

Se ha dado un aumento en la tasa de clientes que han incumplido con el pago de los préstamos que han obtenido de varios socios durante el último año. Credit One, como servicio de calificación crediticia, podría arriesgarse a perder clientes comerciales dada la falta de credibilidad en sus servicios.

OBJETIVO DEL PROYECTO

El objetivo principal del presente proyecto es otorgarle a la empresa Credit One una solución de mejora en el proceso de predicción que permita minimizar el riesgo de otorgación de préstamos a clientes susceptibles al incumplimiento de los pagos mediante un proceso de ciencia de datos. Con esto se es posible mejorar la predicción realizada que determina si el cliente efectuará o no el pago de un crédito.

Para dar solución al presente problema, Credit One plantea dos interrogantes las cuales desea buscar una pronta respuesta con base a la información proporcionada: ¿cómo podemos asegurar que los clientes pueden pagar sus préstamos? Y, ¿se pueden aprobar el crédito a clientes con certeza que este efectuará el pago de este?

RECURSOS DISPONIBLES

La empresa Credit One ha facilitado algunos datos históricos para la realización de esta tarea. Esta información proporcionada cuenta con un total de 6 meses de datos crediticios de 3 000 clientes; la cual, además contiene una serie de datos personales de los mismos que incluyen la edad, el género, su nivel de educación, el estado civil y el límite de balance de crédito otorgado al cliente.

ANÁLISIS Y VISUALIZACIÓN DE DATOS

Dado que se comprende que no es posible controlar los hábitos de gasto de los clientes, se le da un enfoque al proyecto en el que se trata de determinar que atributos se pueden considerar estadísticamente significativos para el problema en cuestión; además, se busca obtener información concreta derivada de los datos proporcionados por Credit One.

A continuación, se resaltan aquellos hallazgos que se consideran pueden dar un potencial valor comercial a la empresa; así como recomendaciones que se le brindan a Credit One con respecto a la información proporcionada para el análisis. Importante resaltar que el atributo que especifica el sexo del cliente fue descartado del análisis, dado que éste puede resultar discriminatorio al considerarse un factor que incida en el otorgamiento o no del crédito.

ANÁLISIS CONSIDERANDO LA EDAD DEL CLIENTE

La mayor cantidad de clientes se ubican dentro del rango de edad que va desde los 30 hasta los 40 años aproximadamente, tal y como se puede observar en la figura 1. Por ello, la mayor concentración de observaciones del conjunto de datos proporcionado se ubica dentro de este rango de edad.

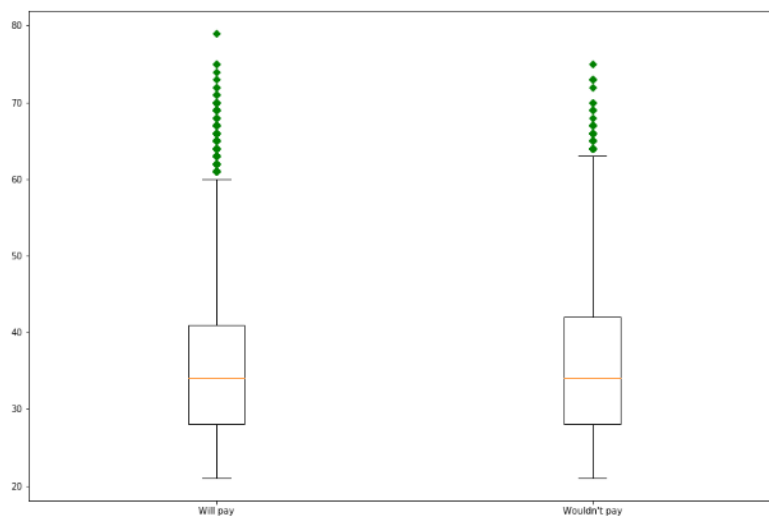


Figura 1. Diagrama de caja de la edad del cliente respecto a su comportamiento del pago del crédito el siguiente mes.

No se obtiene alguna diferencia significativa en cuanto a edad entre los clientes que se predicen efectuará o no el pago del préstamo, por lo que se puede concluir que el pago o no del crédito adeudado por el cliente no está determinado por la edad de este.

ANÁLISIS CONSIDERANDO EL LÍMITE DE CRÉDITO DEL CLIENTE

Del conjunto de datos de Credit One se pudo determinar que no hay clientes cuyo límite de crédito se encuentre dentro del rango de los \$800 000 y los \$900 000. Por otro lado, se puede observar en la figura 2 que el porcentaje de clientes que incumplen con el pago del préstamo tiende a disminuir en aquellos casos en el que el cliente dispone de un crédito superior;

comenzando con un porcentaje de incumplimiento en los pagos de un 29.5% en aquellos que tienen un límite de crédito igual o inferior a \$100 000 hasta alcanzar un aproximado de un 13% de casos de incumplimiento en aquellos clientes que disponen de un límite superior a los \$500 000.

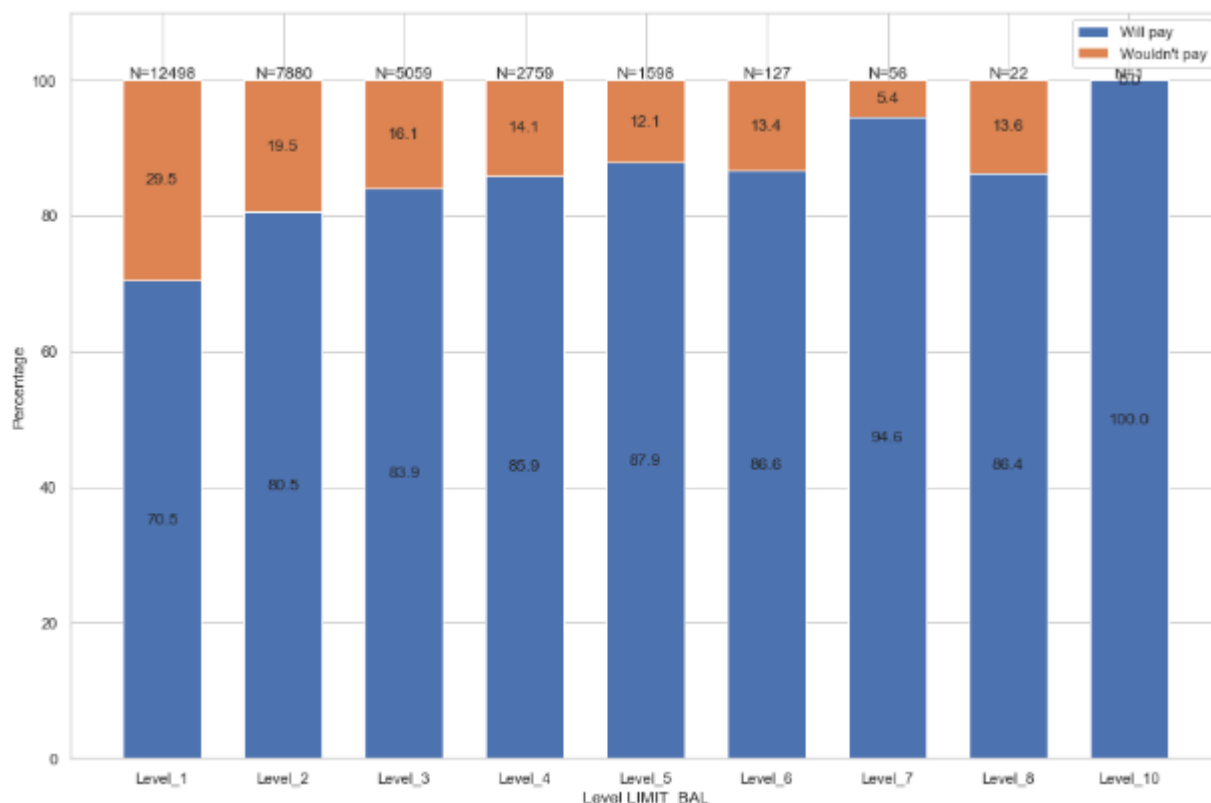


Figura 2. Gráfico de barras del comportamiento del cliente respecto al pago o no del crédito.

Con lo anterior se puede confirmar que entre menor sea el límite de crédito que dispone el cliente, éste tendrá a incumplir con los pagos del préstamo en un mayor porcentaje de los casos. Con esto no se indica como solución aumentarle el límite de balance crediticio a los clientes; sino, prestar mayor atención a aquellos casos en los que el cliente cuenta con un crédito inferior a los \$500 000.

ANÁLISIS CONSIDERANDO EL NIVEL DE EDUCACIÓN DEL CLIENTE

El comportamiento del cliente con respecto al incumplimiento con el pago del préstamo no tiende a disminuir conforme sea más alto su nivel de escolaridad, en donde el porcentaje de clientes que se predice que incumplirán con el pago del préstamo y que tienen un nivel de escolaridad universitaria, alcanza un 23.7%, únicamente superado por aquellos ubicados dentro de la categoría 'High School' (ver figura 3).

Si se considera que la categoría 'Others' hace referencia a grados académicos como técnicos o sin ningún título académico entonces los datos muestran que entre menor sea el nivel de educación del cliente, este tiende a cumplir con el pago del crédito en un mayor porcentaje de los casos. Cabe destacar que el caso de la categoría 'Others' cuenta con una cantidad de

observaciones mucho menor que el resto de los casos; por ello, sería bueno el estudio de más clientes que se encuentren dentro de esta categoría como recomendación para Credit One.

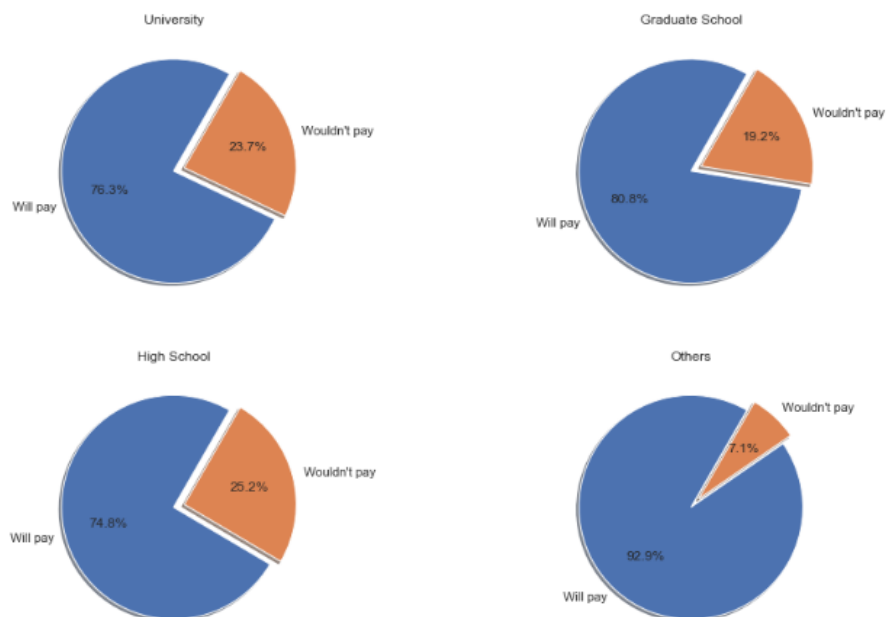


Figura 3. Gráfico de pie considerando el nivel de educación del cliente.

ANÁLISIS CONSIDERANDO EL ESTADO CIVIL DEL CLIENTE

En este aspecto, se puede resaltar que la gran mayoría de los clientes se encuentran dentro de las categorías 'married' y 'single', menos de 400 clientes se encuentran en las dos categorías restantes.

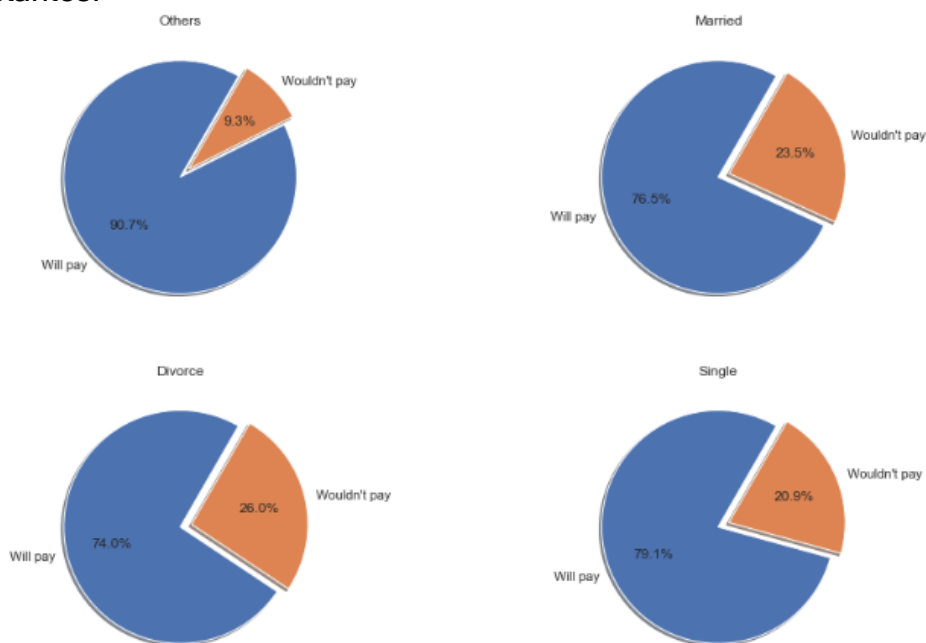


Figura 4. Gráfico de pie considerando el estado civil del cliente.

Analizando los dos grupos más representativos, los porcentajes que reflejan el incumplimiento en los pagos son muy similares, 23.5% para aquellos clientes cuyo estado civil corresponde a casado(a) y un 20.9% para los clientes identificados como soltero(a). Hay un ligero aumento

en el porcentaje que refleja el incumplimiento de los pagos en el caso de aquellos clientes cuyo estado civil corresponde a 'divorce', alcanzando un 26%. Se refleja un bajo porcentaje en la categoría 'others', 9.3%; sin embargo, la cantidad de clientes en esta categoría son pocos, por lo que sería bueno llevar a cabo el análisis con una muestra mayor dentro de esta categoría como recomendación para Credit One. Este comportamiento se puede observar en la figura 4.

ANÁLISIS CONSIDERANDO EL HISTORIAL DE PAGO DEL CLIENTE EN UN SEMESTRE

En este caso se analiza el historial crediticio de los clientes considerando la cantidad de meses en el que se facturó algún monto y el cliente no realizó ningún pago. Esto se determinó sumando la cantidad de meses en el que el pago por parte del cliente fue igual a 0 dólares y restando la cantidad de meses en el que se le factura al cliente un monto de 0 dólares.

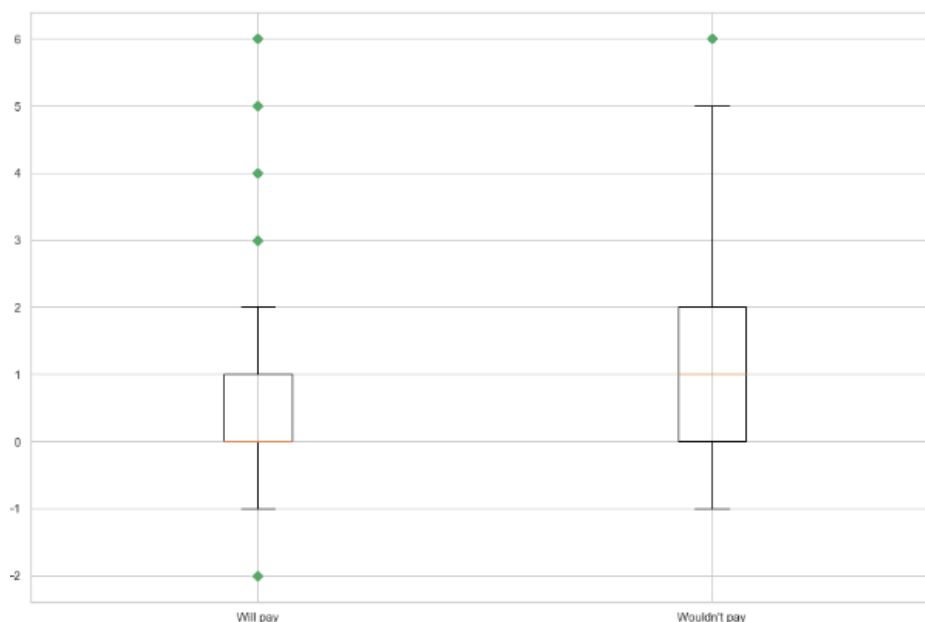


Figura 5. Gráfico de caja del histórico semestral crediticio del cliente.

De los resultados obtenidos (ver figura 5) se puede destacar que para el caso en el que se predice que el cliente efectuará el pago el siguiente mes se observa que se tiene una mediana de 0 en el conjunto de datos; es decir, el historial crediticio de los 6 meses en análisis refleja que no hubo algún mes en el que se haya facturado y no se haya reportado ningún pago por parte del cliente. En este caso, la mayor concentración de datos oscila entre los valores de 0 o 1; además, escasas observaciones se encuentran por encima de 2 meses en el que se haya identificado el comportamiento en análisis.

Por otro lado, para el caso en el que se predice que no se efectuará el pago el siguiente mes, la mediana es de 1 y la mayor concentración de datos corresponden a valores de 0, 1 y 2. Además, se refleja una mayor concentración de datos que van de 2 hasta 5 con respecto a los casos en el que se predice se efectuará el pago. Con esto se puede afirmar que si en el historial del cliente se reportan meses en el que no se hayan realizado pagos del crédito otorgado y las facturaciones corresponden a valores que reflejan un monto superior a \$0 (hizo uso del crédito), se puede predecir que el cliente no efectuará el pago el siguiente mes.

MODELOS DE PREDICCIÓN

Se llevo a cabo el desarrollo de diferentes modelos de clasificación basados en los datos proporcionados por Credit One. Para esto, se utilizaron tres algoritmos de clasificación distintos: Random Forest, Support Vector Machine y K-Nearest Neighbors. Con base a los resultados obtenidos mediante la implementación de los modelos, se procede entonces a plantear una solución al problema partiendo de aquel que presente un mejor rendimiento.

Es importante resaltar que, para el preprocesamiento de los datos, se hace uso de PCA (Principal Component Analysis) para reducir la dimensionalidad de los datos, aplicado únicamente a las variables numéricas del conjunto de datos; es decir, al historial de crédito de cada uno de los clientes. Por otro lado, como se menciona anteriormente, se descarta del conjunto de datos el atributo que indica el sexo del cliente, con el fin de evitar que el modelo realice las predicciones tomando este aspecto como parámetro, dado que puede resultar discriminatorio.

Además, se discretizan algunos datos continuos; en este caso se aplica a dos atributos de la siguiente manera:

- AGE: se discretiza en 4 depósitos Young adults (1) que incluyen los clientes de 15 a 29 años, Adults (2) de 30 a 44 años, Middle-aged adults (3) de 45 a 59 años y Older adults (4) que corresponde a los clientes que tienen entre 60 y 99 años.
- LIMIT_BAL: se discretiza en 10 depósitos con límites cada \$100 000.

La tabla 1 muestra el rendimiento de los modelos en las predicciones:

Tabla 1. Parámetros de rendimiento de los modelos de clasificación.

Modelo	Precisión	Kappa
Random Forest	81.6%	0.361
Support Vector Machine	78.4%	-
K-Nearest Neighbors	78.5%	0.102

De los resultados anteriores, se puede determinar que aquel modelo que presenta un mejor rendimiento corresponde al que utiliza el algoritmo de Random Forest Classifier con una precisión y un valor de kappa superior a los dos modelos restantes.

Se muestra ahora la importancia que este modelo seleccionado da a cada uno de los atributos:

Feature ranking:

1. feature 13: PC3 (0.123410)
2. feature 10: PC0 (0.123131)
3. feature 11: PC1 (0.117059)
4. feature 12: PC2 (0.116422)
5. feature 14: PC4 (0.116359)
6. feature 4: PAY_0 (0.106774)
7. feature 5: PAY_2 (0.052598)
8. feature 6: PAY_3 (0.037276)

```
9. feature 1: EDUCATION (0.033255)
10. feature 0: LIMIT_BAL (0.032570)
11. feature 3: AGE (0.032558)
12. feature 7: PAY_4 (0.030197)
13. feature 9: PAY_6 (0.030029)
14. feature 8: PAY_5 (0.025503)
15. feature 2: MARRIAGE (0.022858)
```

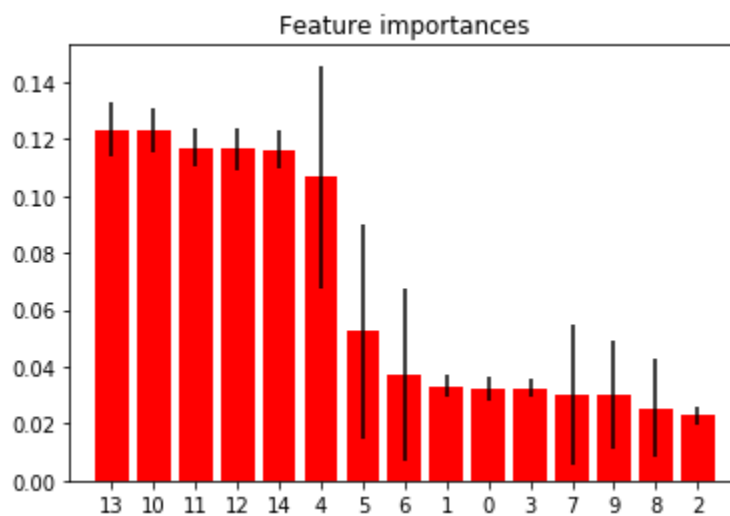


Figura 6. Importancia de los atributos para el modelo de Random Forest.

CONCLUSIONES

No es posible asegurarse si un cliente efectuará o no el pago del crédito, dado que esto lo determinan muchos factores que involucran los hábitos de gasto de los clientes; sin embargo, es posible predecir, con una precisión aproximada de un 81%, si el cliente efectuará o no el pago del crédito mediante un modelo de predicción. Respecto a la pregunta planteada por Credit One, con los datos proporcionados se fue capaz desarrollar un modelo de predicción con un buen porcentaje de confiabilidad, que permite determinar si los clientes pueden pagar el préstamo con un alto porcentaje de aciertos. Sin embargo, si se desea mejorar el rendimiento del modelo, puede ser necesario que se requieran más datos con el fin de analizar el comportamiento de más clientes.

Cabe destacar que las predicciones realizadas en este caso toman en consideración el historial de crédito del cliente, incluso, el modelo les da una mayor importancia a estos atributos para llevar a cabo la predicción, por lo que, al darse el caso que se presente un nuevo cliente sin dicho historial crediticio, su predicción no va a ser precisa y, por consecuencia, su conclusión puede ser errónea.