

Reporte de hallazgos

Informe de Hallazgos

Efrén Jiménez

18/03/2020

Contenido

¿Cómo se asegura de que los clientes puedan pagar sus préstamos?	3
¿Podemos aprobar clientes con alta certeza?	3
¿Qué atributos en los datos podemos considerar estadísticamente significativos para el problema en cuestión?	3
¿Qué información concreta podemos derivar de los datos que tenemos?	4
¿Qué métodos probados podemos usar?	4

¿Cómo se asegura de que los clientes puedan pagar sus préstamos?

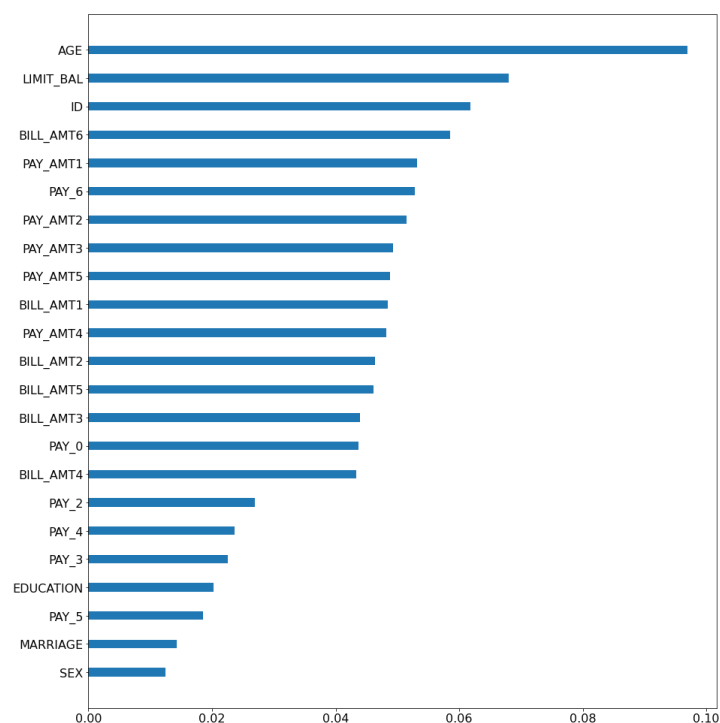
Asegurar que un cliente puede hacer el pago de sus préstamos es bastante difícil, debido a que predecir el comportamiento financiero de una persona es muy complicado debido a una cantidad n de variables.

¿Podemos aprobar clientes con alta certeza?

Existe a nivel estadístico una posibilidad de aprobar tarjetas de clientes que, si van a pagar en los siguientes meses, aunque clasificar el comportamiento financiero de un cliente es bastante difícil, pero utilizando los datos históricos de pago de créditos anteriores se podría brindar una idea a la gerencia de cómo va a hacer el comportamiento financiero de clientes específicos en el tiempo.

¿Qué atributos en los datos podemos considerar estadísticamente significativos para el problema en cuestión?

Las variables AGE, LIMIT_BAL, BILL_AMT6, PAY_AMT1 son las más representativas en modelos de clasificación como Naive Bayes, lo cual podrían usarse para realizar algunos análisis sobre estas variables.



¿Qué información concreta podemos derivar de los datos que tenemos?

La exploración de datos del conjunto de información brindado de las tarjetas de crédito muestra las siguientes posibilidades de negocios:

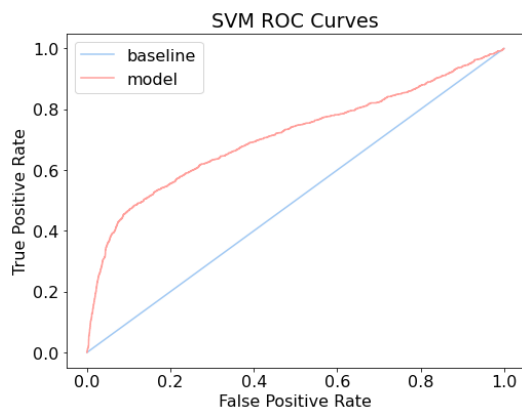
- El set de datos esta desbalanceado en entre el pago y no pago de la cuota del préstamo.
- El número de mujeres es mayor que el de los hombres para las categorías de pago predeterminadas y no predeterminadas, pero no podemos ver sus respectivas proporciones
- El conjunto de datos consta de aproximadamente el 78% (23364) personas con estado predeterminado y el 22% (6634) con estado no predeterminado.
- Hay un mayor porcentaje de mujeres que hombres en la categoría de pago predeterminado.
- El porcentaje de clientes que tienen un título de posgrado es mayor en la categoría de pago predeterminado.
- Las personas que tienen un solo estado tienen un mayor porcentaje de incumplimiento que los casados.
- La edad alcanza un máximo de alrededor de 28-29 años en la categoría de pago predeterminado y tiene un pico más bajo alrededor de alrededor de 27-28 años en la categoría de pago predeterminado según lo visualizado en los gráficos
- La tasa de incumplimiento aumenta durante la recopilación de datos de abril del 2005 a septiembre del 2005.

¿Qué métodos probados podemos usar?

Se verifico el desequilibrio de datos, la visualización de las características y se entendió la relación entre diferentes variables.

Utilizamos tanto la división de datos y la validación cruzada de entrenamiento para evaluar la efectividad del modelo para clasificar el objetivo.

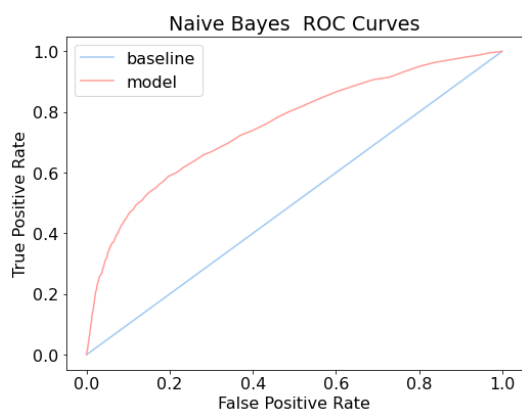
Los modelos utilizados fueron: SVM, Logistic Regression, Naive Bayes, K-NEIGHBOURS



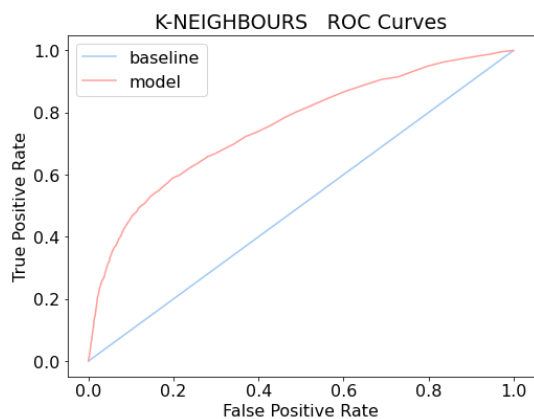
SVM	
Recall	34%
Precision	68%
ROC	71%



Logistic Regression	
Recall	36%
Precision	63%
ROC	76%



Naive Bayes	
Recall	36%
Precision	63%
ROC	76%



K-NEIGHBOURS	
Recall	36%
Precision	63%
ROC	76%

Tenemos 3 modelos que brindan la misma medida en la clasificación de un cliente si va a bajar su cuota o no, por lo cual no se recomienda utilizarlos porque los porcentajes de clasificar correctamente es muy baja.