
Proyecto Riesgo Relativo: Análisis de datos “ Bancarios”

Minerva Bobadilla
Antofagasta, Chile
22-08-2024

Caso

De análisis

En este proyecto se centró en un análisis de datos para una entidad Bancaria. En el contexto reciente, la disminución de las tasas de interés en el mercado ha desencadenado un aumento notable en la demanda de solicitudes de crédito. Los clientes ven una oportunidad favorable para financiar compras importantes o consolidar deudas existentes, lo que ha llevado a una afluencia de solicitudes de préstamo en el banco “Super Caja”.

Se desarrolló el análisis inmerso en el mundo financiero, explorando los datos del banco con el objetivo de segmentar y evaluar la situación crediticia de sus clientes.

El contexto nos otorga el problema que se enfrenta: la necesidad de automatizar y optimizar el proceso de análisis crediticio para gestionar eficazmente el riesgo de incumplimiento.

El objetivo principal es mejorar la eficiencia y la precisión en la evaluación del riesgo crediticio, permitiendo al banco tomar decisiones informadas sobre la concesión de crédito y reducir el riesgo de préstamos no reembolsables.

Por lo anterior es necesario Validar (refutar o confirmar) estas hipótesis mediante el análisis de los datos, y proporcionar recomendaciones estratégicas basadas en los hallazgos .

Riesgo relativo: “Análisis Bancario”

Para el estudio se cuenta con una base de datos de 34790 usuarios, la cual se proceso con técnica de Análisis de Datos, cálculo de riesgo relativo y desarrollo la metodología de segmentación (Cuartiles) para validación de hipótesis y clasificar por grupos con mayor riesgo relativo.



Riesgo relativo: “Análisis Bancario”

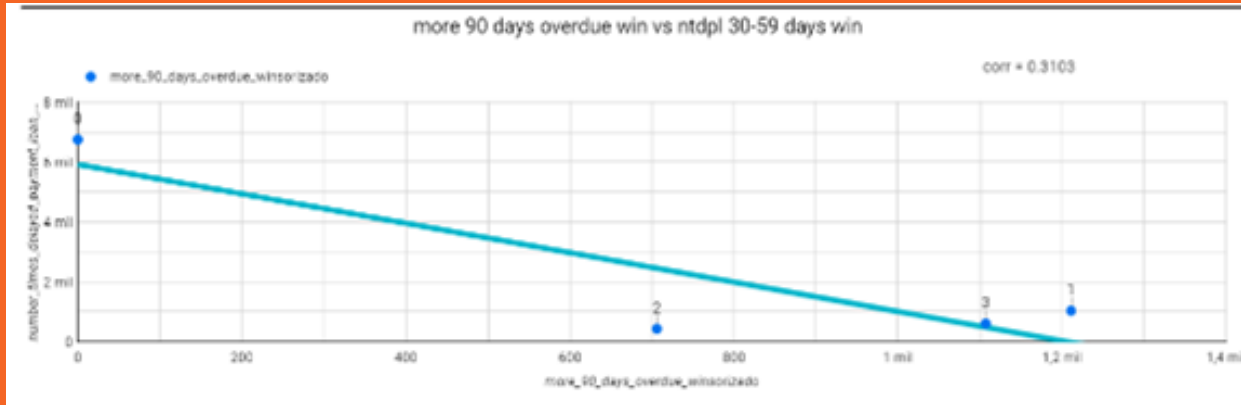


Figura N° 1: Correlación entre variables more 90 days overdue vs number times delayed payment loan 30 - 59 days.

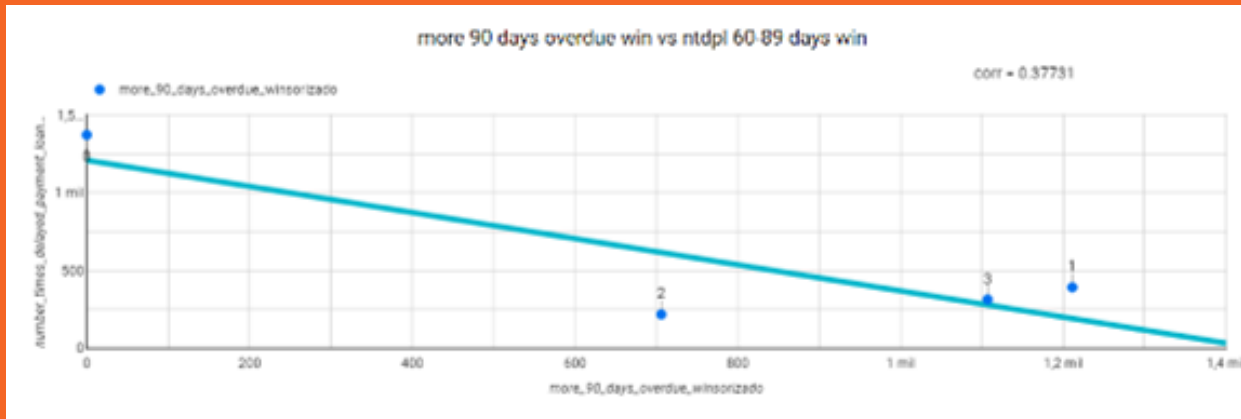


Figura N° 2: Correlación entre variables more 90 days overdue vs number times delayed payment loan 60 - 89 days.

Riesgo relativo: “Análisis Bancario”

Las Hipótesis entregadas por la entidad Bancaria y que se deben validar, para saber qué hace que un cliente sea riesgoso son las siguientes:

- I) Los más jóvenes tienen un mayor riesgo de impago.
 - II) Las personas con más cantidad de préstamos activos tienen mayor riesgo de ser malos pagadores.
 - III) Las personas que han retrasado sus pagos por más de 90 días tienen mayor riesgo de ser malos pagadores.
-

Riesgo relativo: “Análisis Bancario”

I) Los más jóvenes tienen un mayor riesgo de impago.

Tabla N° 1: Cálculo del riesgo relativo por cuartiles en Variable “Age”

non_exposed_quartile ▼	risk_relative ▼	rango_min_age ▼	rango_max_age ▼	cantidad_default_flag_1 ▼	cantidad_default_flag_0 ▼
1	1.0	21	41	296	8402
2	0.685810810810...	41	51	203	8495
3	0.435860921286...	51	62	129	8568
4	0.172317108415...	62	80	51	8646

De acuerdo a los resultados de la tabla N° 1 podemos indicar que el cuartil 1 con más riesgo de impago son el segmento de 21 hasta 41 años de edad, por lo anterior se valida la hipótesis que los más jóvenes tienen un mayor riesgo de impago.

Riesgo relativo: “Análisis Bancario”

II) Las personas con más cantidad de préstamos activos tienen mayor riesgo de ser malos pagadores.

Tabla N° 2: Cálculo del riesgo relativo por cuartiles en Variable “total loan”

non_exposed_quartile ▼	risk_relative ▼	rango_min_total_loan ▼	rango_max_total_loan ▼	cantidad_default_flag_1 ▼	cantidad_default_flag_0 ▼
1	1.0	1	5	259	8439
2	0.664092664092...	5	8	172	8526
3	0.544464140876...	8	11	141	8556
4	0.413174915417...	11	57	107	8590

Con los resultados de la tabla N° 2 podemos indicar que el cuartil 1 de la variable total de crédito, son quienes tienen mayor riesgo relativo y son el segmento de 1 hasta 5 créditos, por lo anterior se debe refutar la hipótesis dado que no tenemos especificación de cuáles son los créditos activos.

Riesgo relativo: “Análisis Bancario”

III) Las personas que han retrasado sus pagos por más de 90 días tienen mayor riesgo de ser malos pagadores.

Tabla N° 3: Cálculo del riesgo relativo por cuartiles en Variable “number times delayed payment loan 30 - 59 days, number times delayed payment loan 60 - 89 days y more 90 days overdue”

non_exposed_quartile ^	risk_relative_ntdl_30_59_dw	risk_relative_ntdl_60_89_days_wins	risk_relative_m90dayswins
1	1	1	1
2	0,84	1,1	0,78
3	1,42	0,67	0,83
4	32,48	14,64	35,12

Riesgo relativo: “Análisis Bancario”

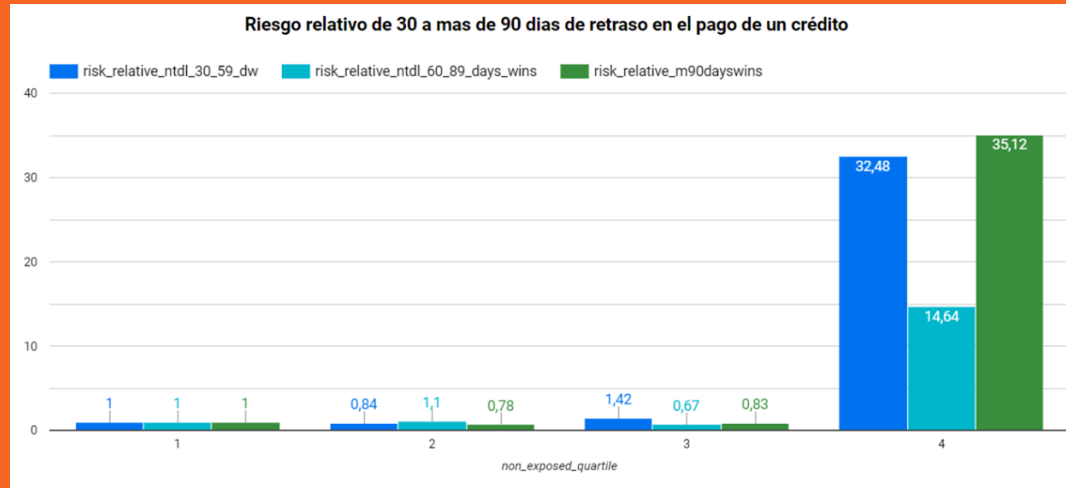


Figura N° 3: Riesgo relativo entre variables more 90 days overdure, number times delayed payment loan 30 - 59 days y number times delayed payment loan 60 - 89 days vs cuartiles.

Con los resultados obtenidos en la figura N° 3 y , se puede indicar que la variable retraso en más de 90 días en el pago, son quienes tienen mayor valor en el cálculo del riesgo relativo. Cabe hacer notar que el cuartil 4 es el segmento de 0 hasta 3 cantidad de veces, por lo anterior se valida la hipótesis “Las personas que han retrasado sus pagos por más de 90 días tienen mayor riesgo de ser malos pagadores”.

Riesgo relativo: "Análisis Bancario"

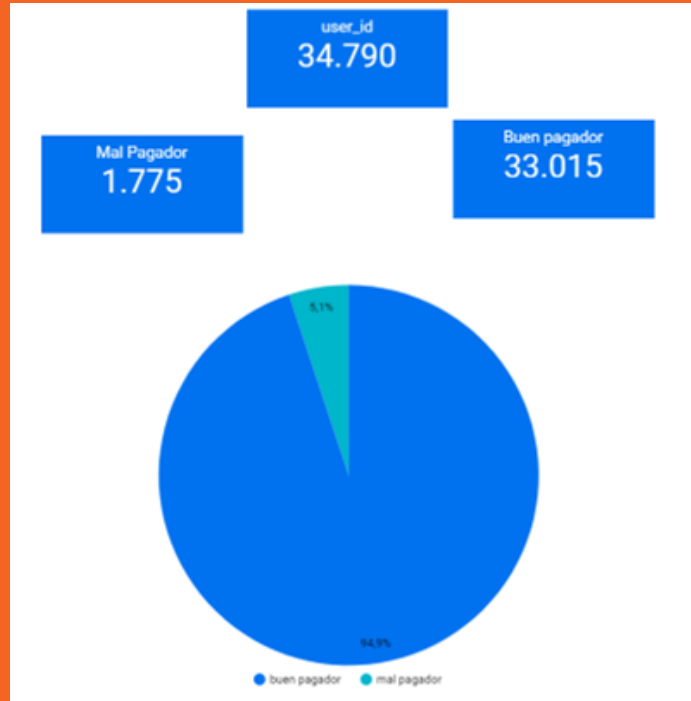


Figura N° 4: Clasificación de clientes de acuerdo a análisis de Riesgo relativo, cumpliendo con más de tres condiciones de riesgo (number times delayed payment loan 30 - 59 days, number times delayed payment loan 60 - 89 days y more 90 days overdue, total loan y using_lines_not_secured_personal_assets_wins)

Riesgo relativo: “Análisis Bancario”

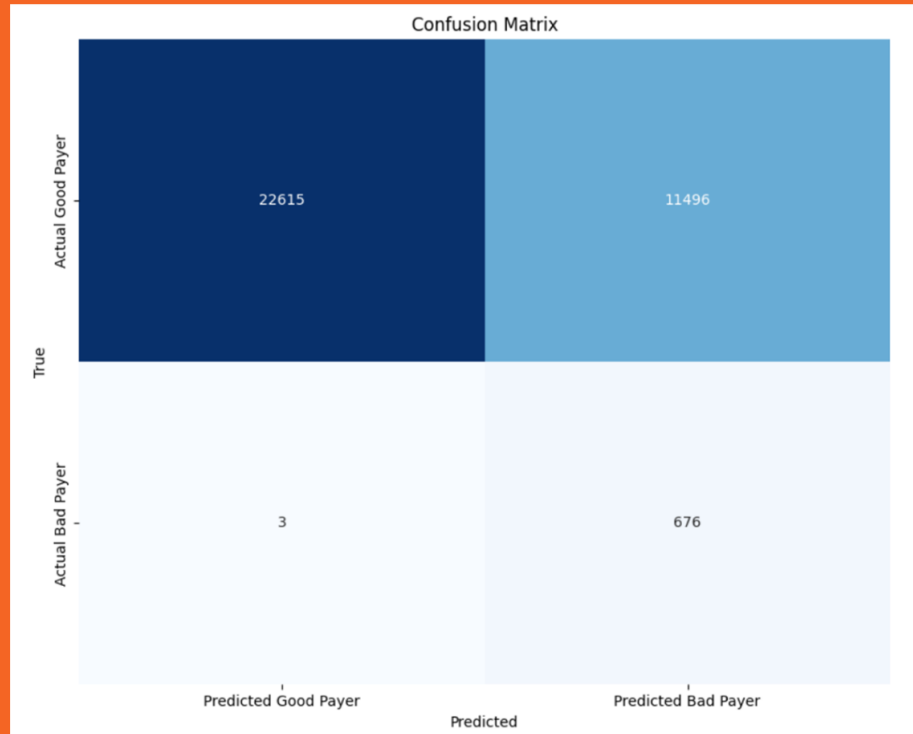


Figura N° 5: Matriz de confusión para validación de clasificación de clientes en buenos y malos pagadores.

Riesgo relativo: "Análisis Bancario"

Tabla N° 4: Parámetros que indican validación de matriz de confusión.

Métricas para la Clase 0 (Buen Pagador):

Precision: 0.9999

Recall: 0.6630

F1 Score: 0.7973

Métricas para la Clase 1 (Mal Pagador):

Precision: 0.0555

Recall: 0.9956

F1 Score: 0.1052

Exactitud (Accuracy): 0.6695

Con los resultados obtenidos en parámetros de matriz de confusión, se puede indicar lo siguiente: el modelo tiene un rendimiento muy bueno en la clasificación de la Clase 0 (Buen Pagador), pero un rendimiento deficiente en la clasificación de la Clase 1 (Mal Pagador).

Conclusiones

Durante el desarrollo del proyecto se utilizó el proceso de Análisis de Datos, cálculo de riesgo relativo y la metodología de segmentación (Cuartiles) para validación de hipótesis. La entidad bancaria planteó tres hipótesis, sobre qué hace que un cliente sea más riesgoso o mal pagador. Estas hipótesis se detallan a continuación y de esto se puede concluir lo siguiente:

I) Los más jóvenes tienen un mayor riesgo de impago:

De acuerdo a los resultados de la tabla N° 1 podemos indicar que el cuartil 1 con más riesgo de impago son el segmento de 21 hasta 41 años de edad, por lo anterior se valida la hipótesis que los más jóvenes tienen un mayor riesgo de impago.

Conclusiones

II) Las personas con más cantidad de préstamos activos tienen mayor riesgo de ser malos pagadores.

Con los resultados de la tabla N° 2 podemos indicar que el cuartil 1 de la variable total de crédito, son quienes tienen mayor riesgo relativo y son el segmento de 1 hasta 5 créditos, por lo anterior se debe refutar la hipótesis dado que no tenemos especificación de cuáles son los créditos activos.

Conclusiones

III) Las personas que han retrasado sus pagos por más de 90 días tienen mayor riesgo de ser malos pagadores.

Con los resultados obtenidos en la figura N° 3 , se puede indicar que la variable retraso en más de 90 días en el pago, son quienes tienen mayor valor en el cálculo del riesgo relativo.

Cabe hacer notar que el cuartil 4 es el segmento de 0 hasta 3 cantidad de veces, por lo anterior se valida la hipótesis "Las personas que han retrasado sus pagos por más de 90 días tienen mayor riesgo de ser malos pagadores".

Además, se informa que del universo de clientes existen 1775 clientes denominados Mal pagador, lo cual corresponde a un 5,1 % del total, cumpliendo con más de tres condiciones de riesgo (number times delayed payment loan 30 - 59 days, number times delayed payment loan 60 - 89 days y more 90 days overdure, total loan y using_lines_not_secured_personal_assets_wins).

Recomendaciones

Dado que la hipótesis II) Las personas con más cantidad de préstamos activos tienen mayor riesgo de ser malos pagadores. no pudo ser validada, entonces se solicita entregar más información: del tipo de crédito, por ejemplo si está activo o se encuentra caducado, datos en el tiempo de clientes, etc, para optimizar análisis encontrando mayor exactitud a los resultados presentados.

En cuanto a los resultados del modelo de acuerdo a la matriz de confusión, éste necesita ser mejorado, especialmente en su capacidad para identificar correctamente a los malos pagadores. Es necesario realizar un análisis más profundo y aplicar técnicas de rebalanceo de datos para obtener un modelo más preciso y confiable.

Muchas gracias
