

Ficha Técnica: Proyecto 3 Riesgo Relativo

Título del Proyecto

Proyecto 3 Riesgo Relativo

Objetivo

Armar un score crediticio a partir de un análisis de datos y la evaluación del riesgo relativo que pueda clasificar a los solicitantes en diferentes categorías de riesgo basadas en su probabilidad de incumplimiento. Esta clasificación permitirá al banco tomar decisiones informadas sobre a quién otorgar crédito, reduciendo así el riesgo de préstamos no reembolsables. Además, la integración de la métrica existente de pagos atrasados fortalecerá la capacidad del modelo para identificar riesgos, lo que en última instancia contribuirá a la solidez financiera y la eficiencia operativa del banco.

Equipo

Herramientas y Tecnologías

Listado de herramientas y tecnologías utilizadas en el proyecto:

- 1) Big Query (SQL).
- 2) Google Colab (Python).
- 3) Looker Studio.
- 4) Workspace Google (Presentaciones, Gemini y Documentos).
- 5) Videos Looms.
- 6) Git Hub.

Procesamiento y análisis

1) Proceso de Análisis de Datos:

1.1 Procesar y preparar base de datos

- a) Identificar y manejar valores nulos
- b) Identificar y manejar valores duplicados
- c) Identificar y manejar datos fuera del alcance del análisis
- d) Identificar y manejar datos discrepantes en variables categóricas
- e) Identificar y manejar datos discrepantes en variables numéricas
- f) Comprobar y cambiar tipo de dato
- g) Unir tablas
- h) Crear nuevas variables
- i) Construir tablas auxiliares

1.2 Análisis exploratorio

- a) Agrupar datos según variables categóricas.
- b) Visualizar las variables categóricas.
- c) Aplicar medidas de tendencia central.
- d) Visualizar distribución.
- e) Aplicar medidas de dispersión.
- f) Calcular cuartiles, deciles o percentiles.

- g) Calcular correlación entre variables.

1.3 Aplicar técnica de análisis

- h) Aplicar segmentación.
- i) Calcular riesgo relativo
- j) Validar hipótesis.

Resultados y 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 y clasificar por grupos con mayor riesgo relativo. 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.

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.

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).

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.

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).

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.

Limitaciones/Próximos Pasos

Sin observaciones.

Enlaces de interés

[Proyecto 3 Riesgo Relativo.pdf](#)

Dashboard

[Proyecto 3 Riesgo Relativo Matriz de confusion.ipynb](#)

[Proyecto 3 Riesgo Relativo Análisis Bancario \(M. Bobadilla\)](#)

[Presentación video](#)