

Modelo de Rodamiento y Perfilamiento de Clientes para la Gestión de Mora

Objetivos

1. Definir y explicar los conceptos clave relacionados con el rodamiento crediticio: buckets de mora, curación y matriz de transición.
2. Construir una matriz de transición que muestre el comportamiento histórico de los clientes entre un mes y otro.
3. Identificar las variables más influyentes en el rodamiento de clientes, con el fin de aportar información útil para la gestión de riesgo y la toma de decisiones.
4. Clasificar a los clientes en grupos de riesgo (bajo, medio y alto) a partir de las probabilidades generadas por el modelo.

Contenido

Planteamiento del problema.....	3
Data.....	3
Conceptos	4
Modelo de Rodamiento	5
Balance de clases en la variable respuesta.....	7
Entrenamiento y Prueba	8
Perfiles de Clientes	10
Grupos de Riesgo	11
Conclusiones.....	12

Planteamiento del problema

En la gestión crediticia, uno de los mayores retos es controlar la mora, ya que no todos los clientes mantienen un pago estable: algunos se atrasan de forma temporal, otros se recuperan y otros caen en incumplimiento. Por eso, el área de riesgo busca anticipar el deterioro de la cartera para ajustar provisiones, el área de cobros necesita información para priorizar gestiones y recuperar con mayor efectividad.

Data

El dataset Loan Data de Kaggle contiene información de solicitantes de préstamos, con datos personales, financieros y de crédito, y es usado para analizar riesgo crediticio y predecir aprobación o incumplimiento. Una parte clave son las columnas Payment Status (X6 a X11), que registran el historial de pagos de los últimos seis meses y permiten construir los buckets de mora: vigente (0), mora temprana (1–30 días), intermedia (31–60 días) y severa (más de 60 o 90 días), facilitando así el análisis de rodamiento. Y finalmente la variable respuesta Y (incumplimiento al siguiente mes, en este caso, el siguiente mes es octubre)

Variable_original	Nombre_ingles	Nombre_espanol	Descripcion
X1	credit_amount	monto_credito	Monto del crédito otorgado (incluye individual y familiar)
X2	gender	genero	Género (1 = masculino; 2 = femenino)
X3	education	educacion	Nivel educativo (1 = posgrado; 2 = universidad; 3 = secundaria; 4 = otros)
X4	marital_status	estado_civil	Estado civil (1 = casado; 2 = soltero; 3 = otros)
X5	age	edad	Edad en años
X6	payment_status_sep2005	estatus_pago_sep2005	Estado de pago en septiembre 2005
X7	payment_status_aug2005	estatus_pago_ago2005	Estado de pago en agosto 2005
X8	payment_status_jul2005	estatus_pago_jul2005	Estado de pago en julio 2005
X9	payment_status_jun2005	estatus_pago_jun2005	Estado de pago en junio 2005
X10	payment_status_may2005	estatus_pago_may2005	Estado de pago en mayo 2005
X11	payment_status_apr2005	estatus_pago_abr2005	Estado de pago en abril 2005
X12	bill_amount_sep2005	monto_factura_sep2005	Monto facturado en septiembre 2005
X13	bill_amount_aug2005	monto_factura_ago2005	Monto facturado en agosto 2005
X14	bill_amount_jul2005	monto_factura_jul2005	Monto facturado en julio 2005
X15	bill_amount_jun2005	monto_factura_jun2005	Monto facturado en junio 2005
X16	bill_amount_may2005	monto_factura_may2005	Monto facturado en mayo 2005
X17	bill_amount_apr2005	monto_factura_abr2005	Monto facturado en abril 2005
X18	payment_amount_sep2005	monto_pago_sep2005	Monto pagado en septiembre 2005
X19	payment_amount_aug2005	monto_pago_ago2005	Monto pagado en agosto 2005
X20	payment_amount_jul2005	monto_pago_jul2005	Monto pagado en julio 2005
X21	payment_amount_jun2005	monto_pago_jun2005	Monto pagado en junio 2005
X22	payment_amount_may2005	monto_pago_may2005	Monto pagado en mayo 2005
X23	payment_amount_apr2005	monto_pago_abr2005	Monto pagado en abril 2005
Y	default_next_month	incumplimiento_mes_siguiente	Indica si el cliente incumplió el pago en el mes siguiente (1 = sí, 0 = no).

https://www.kaggle.com/datasets/jakeshbohaju/loan-data?utm_source=chatgpt.com

Conceptos

En los bancos es común que algunos clientes no paguen sus préstamos a tiempo, y por eso las instituciones necesitan maneras de medir y vigilar esos atrasos. Para entender cómo lo hacen, primero hay que conocer algunos conceptos clave.

Los **buckets de mora** o **tramos de mora** son conjuntos predeterminados de clientes según los días que llevan atrasados: vigente (al día o 0 días de mora), 1-30 días, 31-60, 61-90, y así sucesivamente.

La **curación** ocurre cuando un cliente que estaba atrasado paga y logra ponerse al día, ya sea totalmente (curación total) o bajando a un bucket de menor atraso (curación parcial).

El **rodamiento** es cuando los clientes se mueven entre los buckets de mora con el paso del tiempo: si empeoran (caen en más atraso), si se mantienen igual o si mejoran/curan al **cierre de mes o fecha de corte**.

La matriz de transición es una tabla que muestra cómo se movieron los clientes de un mes al siguiente según sus días de mora indicando cuántos clientes se mantuvieron en el mismo estado, cuántos se curaron (pasaron a un bucket con menos mora) y cuántos empeoraron (pasaron a un bucket con más mora).

Buckets de Mora		may								Total de Clientes
abr		0 días	31 a 60 días	61 a 90 días	91 a 120 días	121 a 150 días	151 a 180 días	181 a 210 días	211 +	
0 días		19812	683							20495
31 a 60 días		770	1368	114						2252
61 a 90 días		37	41	27	48					153
91 a 120 días			8	4	18	5				35
121 a 150 días			2	1	1	3	2			9
151 a 180 días		1	2		1	2	1	9		16
181 a 210 días								37		37
211 +								1	1	2
Total de Clientes		20620	2104	146	68	10	3	47	1	22999

Interpretación: Para los 3 primeros buckets de mora:

Clientes que estaban al día en abril (20,495):

- 19,812 se mantuvieron al día en mayo.
- 683 cayeron en atraso de 31 a 60 días (rodaron).

Clientes con atraso de 31 a 60 días en abril (2,252):

- 770 se recuperaron y pasaron a al día (curación total).
- 1,368 se mantuvieron en 31–60 días.
- 114 empeoraron y pasaron a 61–90 días.

Clientes con atraso de 61 a 90 días en abril (153):

- 37 se recuperaron y pasaron a al día.
- 41 redujeron atraso a 31–60 días (curaron parcial).
- 27 se mantuvieron en 61–90 días.
- 48 empeoraron a 91–120 días.

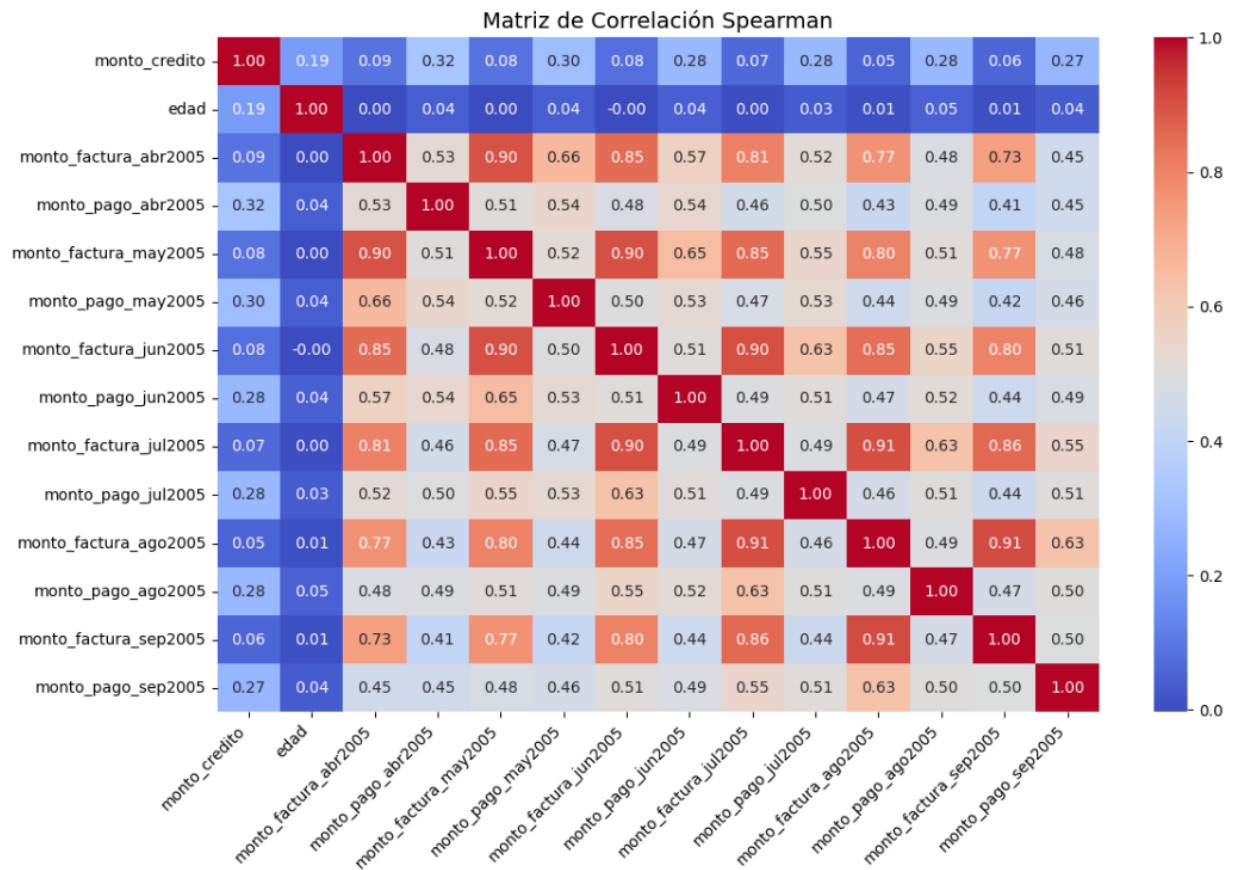
Modelo de Rodamiento

Es un modelo que ayuda a anticipar clientes tienen más probabilidad de moverse de un bucket de mora a otro en el siguiente mes.

Para este proyecto se usó el árbol de decisión debido a que es sencillo y se construye como un esquema de reglas condicionales fáciles de interpretar.

Para este ejemplo, se descartaron las variables de estatus de pago, educación, género y estado civil porque son poco explicativas del riesgo real en el negocio y pueden causar sesgos.

Correlaciones.



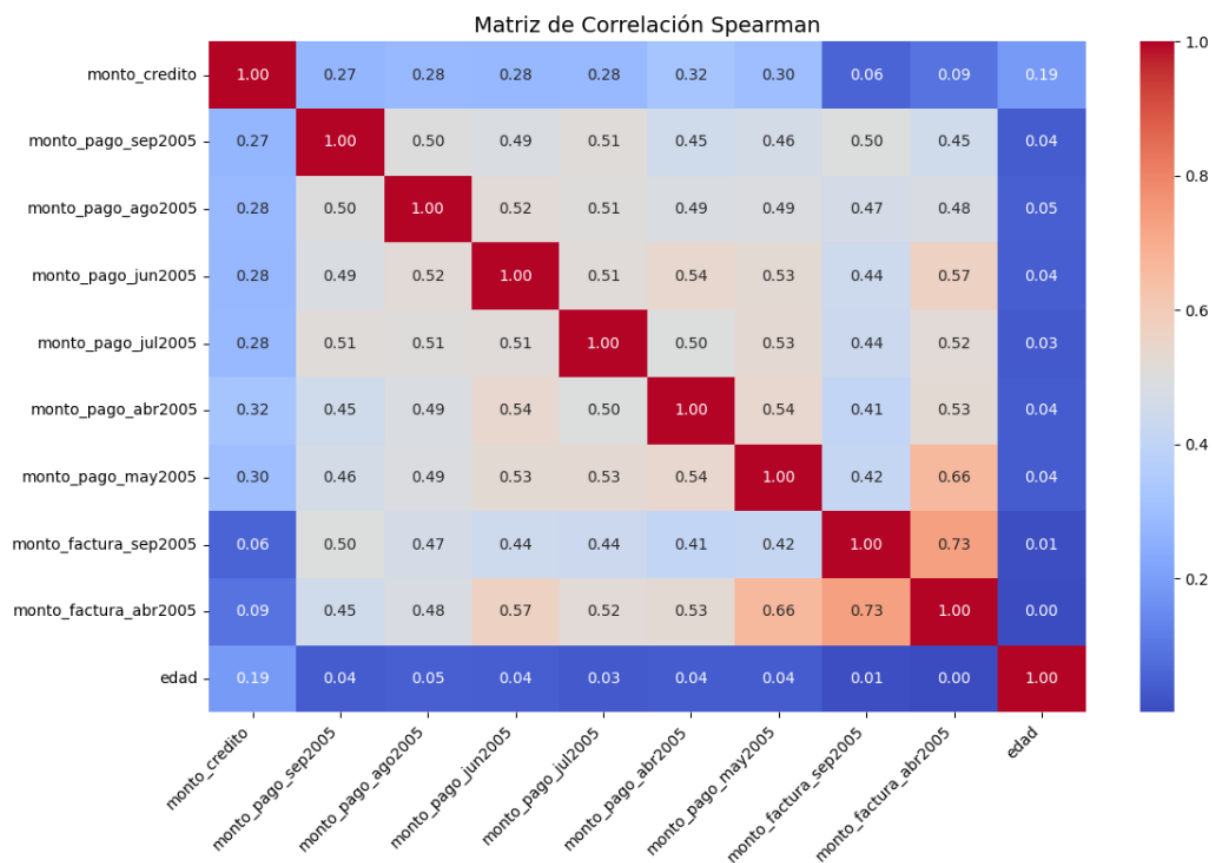
Se sacan las variables que tiene mayor correlación (> 0.75) con las demás:

```
[ 'monto_factura_ago2005',
  'monto_factura_jul2005',
  'monto_factura_jun2005',
  'monto_factura_may2005' ]
```

para finalmente quedar con:

```
[ 'monto_credito',
  'monto_pago_sep2005',
  'monto_pago_ago2005',
  'monto_pago_jun2005',
  'monto_pago_jul2005',
  'monto_pago_abr2005',
  'monto_pago_may2005',
  'monto_factura_sep2005',
  'monto_factura_abr2005',
  'edad' ]
```

Correlaciones



Balance de clases en la variable respuesta

Se observa un desbalance a favor de la clase 0

count	
incumplimiento_mes_siguiente	
0	17826
1	5173

Se aplica un submuestreo para reducir la clase mayoritaria, para este ejemplo se dejó en un 50% y 50%

count	
incumplimiento_mes_siguiente	
0	5173
1	5173

Entrenamiento y Prueba

Se aplica un modelo de árbol de decisión con los siguientes hiper - parámetros:

```
DecisionTreeClassifier
DecisionTreeClassifier(max_depth=5, min_samples_leaf=50, random_state=42)
```

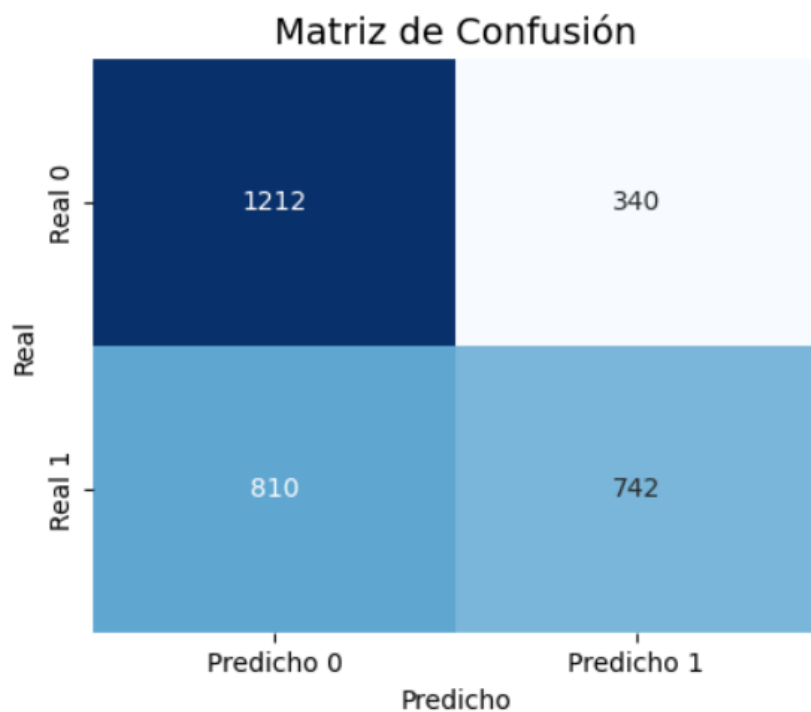
La prueba tuvo el siguiente resultado:

Reporte de clasificación:

	precision	recall	f1-score	support
0	0.60	0.78	0.68	1552
1	0.69	0.48	0.56	1552
accuracy			0.63	3104
macro avg	0.64	0.63	0.62	3104
weighted avg	0.64	0.63	0.62	3104

AUC: 0.67

Interpretación: El modelo acierta en un 63% de los casos (accuracy), también reconoce mejor a los buenos pagadores (recall clase 0 = 0.78) que a los malos (recall clase 1 = 0.48). Su capacidad global para diferenciar entre buenos y malos pagadores se muestra en un AUC de 0.67, lo que indica que el desempeño es aceptable, pero todavía necesita mejorar para identificar con mayor seguridad a los clientes de más riesgo.



Interpretación: La matriz muestra que el modelo acierta con la mayoría de los clientes que cumplen sus pagos, pero se le escapan muchos de los que incumplen. En otras palabras, reconoce bien a los buenos pagadores, aunque todavía falla en identificar a una parte importante de los que representan mayor riesgo.

Perfiles de Clientes

Algo importante, es que se pueden encontrar características de los clientes (como ingresos, monto de deuda, tipo de producto o forma de pago) para identificar cuáles influyen más en que avancen al siguiente bucket. Como ejemplo, las primeras condiciones:

```

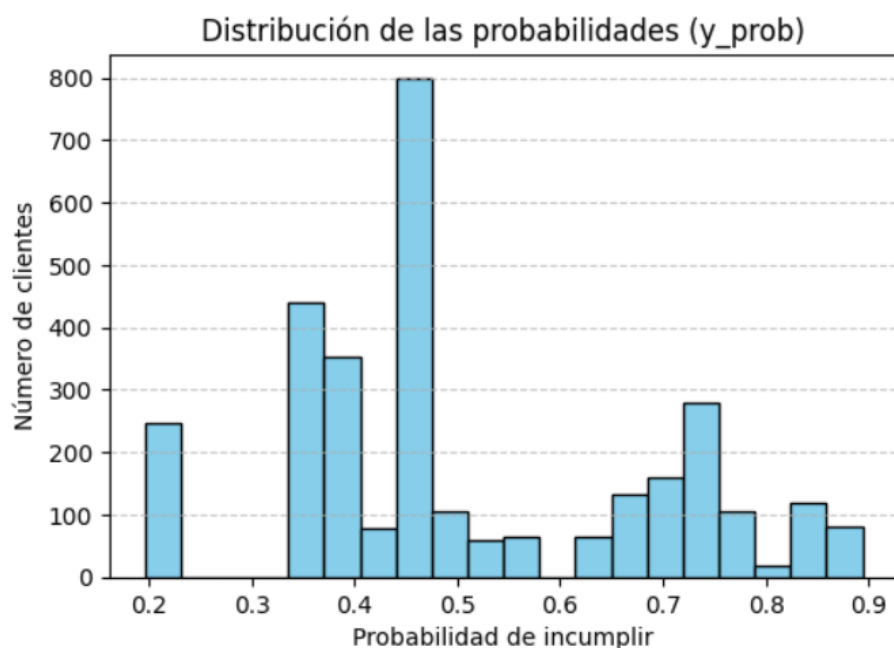
|--- monto_credito <= 115000.00
|   |--- monto_pago_sep2005 <= 52.50
|       |--- monto_factura_sep2005 <= 45.50
|           |--- edad <= 30.50
|               |--- class: 0
|           |--- edad > 30.50
|               |--- class: 1
|       |--- monto_factura_sep2005 > 45.50
|           |--- monto_factura_abr2005 <= 30.00
|               |--- class: 1
|           |--- monto_factura_abr2005 > 30.00
|               |--- monto_pago_jun2005 <= 671.50
|                   |--- class: 1
|               |--- monto_pago_jun2005 > 671.50
|                   |--- class: 1
|       |--- monto_pago_sep2005 > 52.50
|           |--- monto_pago_ago2005 <= 1.50
|               |--- monto_factura_sep2005 <= 3449.50
|                   |--- class: 1
|               |--- monto_factura_sep2005 > 3449.50
|                   |--- monto_pago_jul2005 <= 2285.00
|                       |--- class: 1
|                   |--- monto_pago_jul2005 > 2285.00
|                       |--- class: 1

```

1. Si $\text{monto_credito} \leq 115000.00$ y $\text{monto_pago_sep2005} \leq 52.50$ y $\text{monto_factura_sep2005} \leq 45.50$ y $\text{edad} \leq 30.50 \rightarrow$ clase 0
2. Si $\text{monto_credito} \leq 115000.00$ y $\text{monto_pago_sep2005} \leq 52.50$ y $\text{monto_factura_sep2005} \leq 45.50$ y $\text{edad} > 30.50 \rightarrow$ clase 1
3. Si $\text{monto_credito} \leq 115000.00$ y $\text{monto_pago_sep2005} \leq 52.50$ y $\text{monto_factura_sep2005} > 45.50$ y $\text{monto_factura_abr2005} \leq 30.00 \rightarrow$ clase 1
4. Si $\text{monto_credito} \leq 115000.00$ y $\text{monto_pago_sep2005} \leq 52.50$ y $\text{monto_factura_sep2005} > 45.50$ y $\text{monto_factura_abr2005} > 30.00$ y $\text{monto_pago_jun2005} \leq 671.50 \rightarrow$ clase 1
5. Si $\text{monto_credito} \leq 115000.00$ y $\text{monto_pago_sep2005} \leq 52.50$ y $\text{monto_factura_sep2005} > 45.50$ y $\text{monto_factura_abr2005} > 30.00$ y $\text{monto_pago_jun2005} > 671.50 \rightarrow$ clase 1
6. Si $\text{monto_credito} \leq 115000.00$ y $\text{monto_pago_sep2005} > 52.50$ y $\text{monto_pago_ago2005} \leq 1.50$ y $\text{monto_factura_sep2005} \leq 3449.50 \rightarrow$ clase 1
7. Si $\text{monto_credito} \leq 115000.00$ y $\text{monto_pago_sep2005} > 52.50$ y $\text{monto_pago_ago2005} \leq 1.50$ y $\text{monto_factura_sep2005} > 3449.50$ y $\text{monto_pago_jul2005} \leq 2285.00 \rightarrow$ clase 1
8. Si $\text{monto_credito} \leq 115000.00$ y $\text{monto_pago_sep2005} > 52.50$ y $\text{monto_pago_ago2005} \leq 1.50$ y $\text{monto_factura_sep2005} > 3449.50$ y $\text{monto_pago_jul2005} > 2285.00 \rightarrow$ clase 1

Grupos de Riesgo

Junto a esas reglas, el modelo también genera una probabilidad que indica la posibilidad de que cada cliente en la fase de prueba caiga en la clase 1 (riesgo) o en la clase 0 (no riesgo). Con esas probabilidades se pueden crear grupos de riesgo (bajo, medio, alto), pero para eso hay que decidir puntos de corte. Esos cortes se pueden definir mediante criterios y necesidades del negocio combinadas con técnicas matemáticas. En este caso, se hizo de una manera sencilla y visual, usando un histograma.



- Los clientes con probabilidad menor a 0.35 tienen menor riesgo de rodar (low risk),
- Entre 0.35 y 0.60 se clasifican como riesgo medio (medium risk),
- Finalmente, los mayores a 0.60 se ubican en alto riesgo (high risk).

Estos cortes se propusieron ya que la separación de esos grupos de probabilidades es marcada.

Conclusiones

- ✓ El modelo de rodamiento desarrollado permitió visualizar y predecir el movimiento de los clientes entre los diferentes buckets de mora, mostrando cómo algunos se mantienen al día, otros se atrasan y otros logran curarse.
- ✓ El uso del árbol de decisión facilitó la interpretación de las variables más influyentes en el riesgo crediticio, mientras que la clasificación en grupos de riesgo (bajo, medio y alto) ofreció una forma práctica de segmentar la cartera.
- ✓ Aunque los resultados obtenidos son aceptables, el modelo evidencia la necesidad de seguir afinando las técnicas de predicción para mejorar la detección de clientes con mayor probabilidad de incumplimiento, aportando así valor a la gestión de riesgo de las instituciones financieras.