

Credit Scoring con Ciencia de Datos

Exploración de técnicas avanzadas para evaluar riesgos crediticios utilizando ciencia de datos y aprendizaje automático.

Andrea Liberato Moore

Elvis Donayre

Ronal P.



Credit Scoring

"Imagine que cada decisión financiera que tomamos, desde otorgar un préstamo hasta fijar una tasa de interés, pudiera ser respaldada por una herramienta que no solo reduce riesgos, sino que también abre puertas a oportunidades para aquellos buenos clientes por descubrir y para aquellas empresas con ansías de crecer.

Imagine facilitar con total seguridad la toma de decisiones que permitan que las oportunidades de negocio se maximicen y cada cliente reciba una oferta adaptada a su perfil.

Hoy, vamos a explorar cómo el Credit scoring basado en machine learning está transformando la manera en que evaluamos el riesgo crediticio, haciéndolo más justo, eficiente y predictivo. En este viaje, no solo veremos datos; veremos cómo estos datos pueden contar historias que impulsan decisiones estratégicas y generan impacto."

Objetivo Principal

El objetivo principal es predecir la probabilidad de que un cliente cumpla o incumpla con el pago de un préstamo, sin embargo tradicionalmente los datos en que se basa el sistema para definir si se otorga o no son muy limitados.



- ¿Cómo evitamos otorgar créditos a clientes que eventualmente no pagarán, generando pérdidas?
- ¿Rechazar créditos a clientes que sí hubieran pagado, perdiendo oportunidades de negocio?

Importancia y objetivos del Credit Scoring

- 1 Evaluación de riesgos
- 2 Decisiones de préstamos más informadas
- 3 Mejora de la eficiencia operativa
- 4 Maximización de la rentabilidad





Fuentes de Datos



Historial crediticio



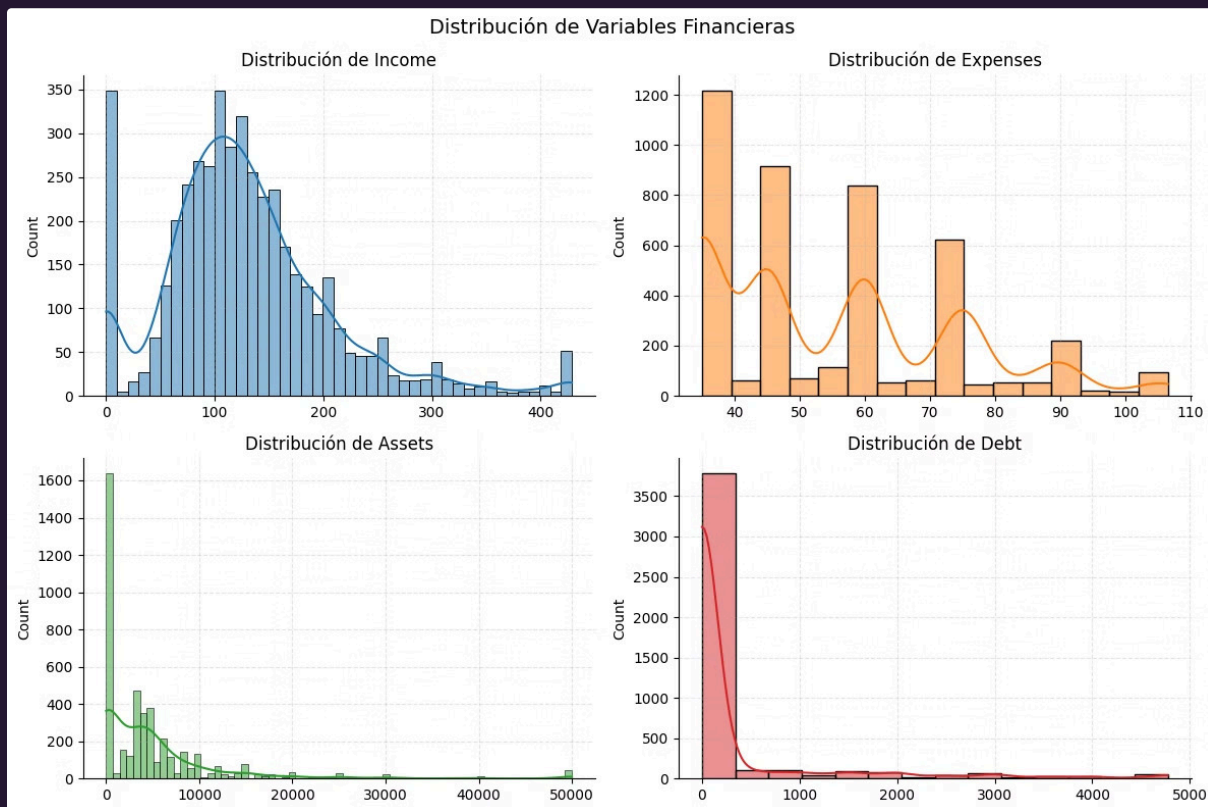
Información financiera



Datos demográficos

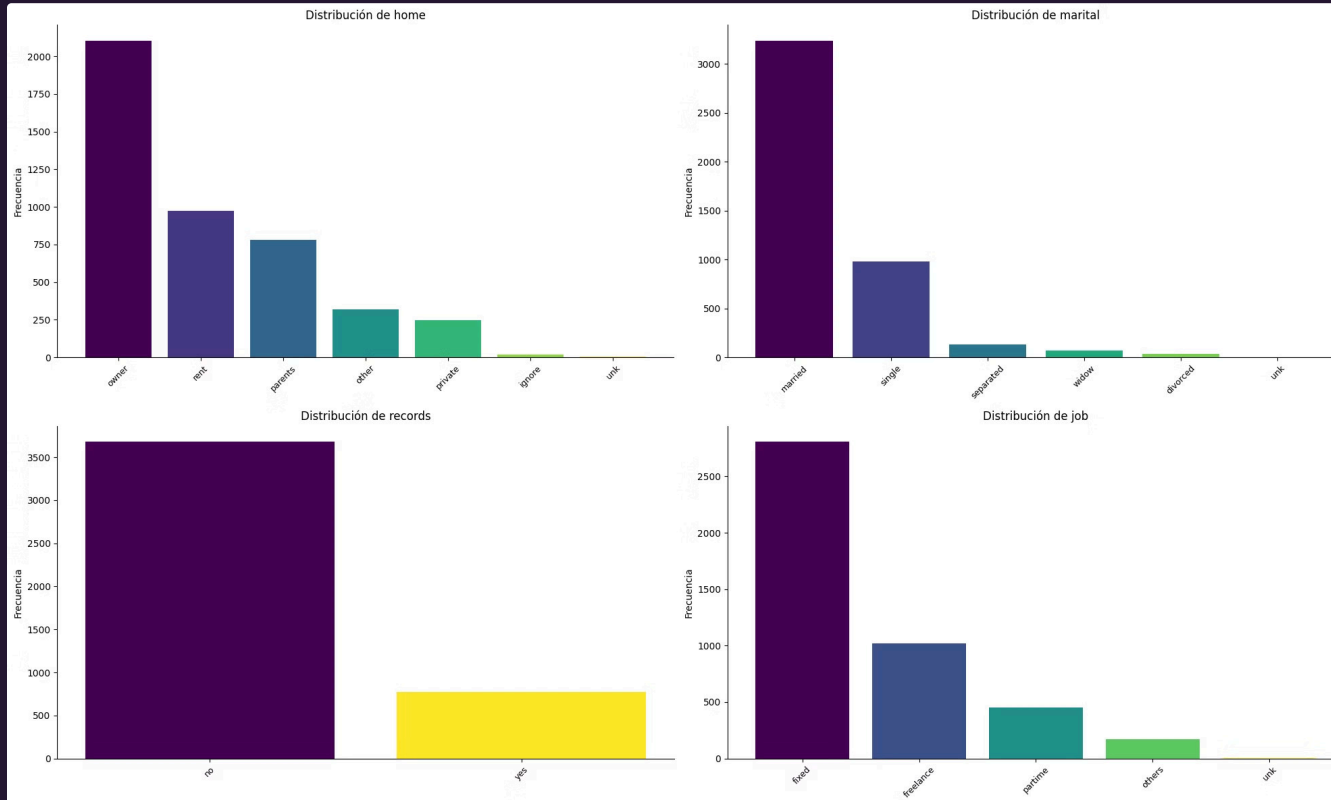
Análisis de datos financieros

"Las variables financieras nos permiten evaluar la capacidad de pago y el nivel de endeudamiento de los solicitantes."



Aquí podemos destacar que la mayoría de nuestros clientes tienen un valor de ingreso al rededor de 100 pero también hay un número muy alto de clientes que no presentan ingresos. Y aunque la mayoría no tienen obligaciones financieras importantes tampoco cuentan con ahorros o activos.

Análisis de datos Demográficos



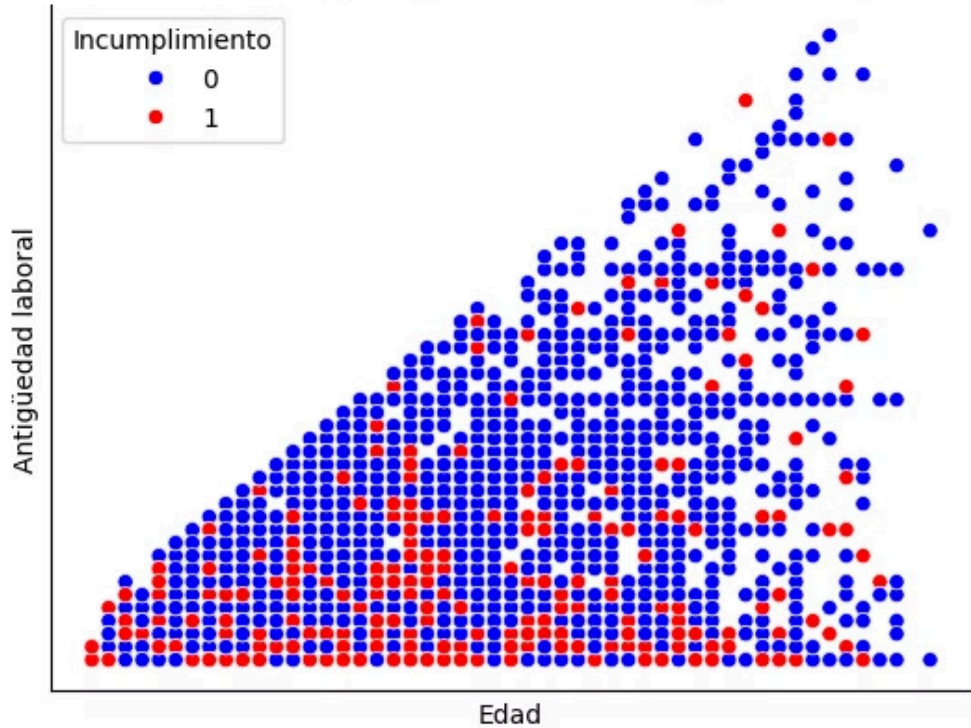
Aquí vemos la distribución de los datos demográficos.

Vemos que la mayoría de clientes son propietarios.

Hay un mayor número de clientes casados.

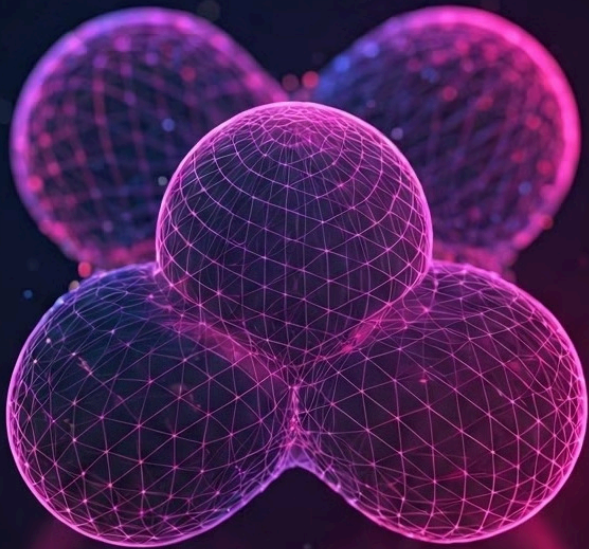
Tienen trabajo fijo y no tienen historial crediticio previo.

Relación entre Edad y Antigüedad laboral según Incumplimiento



"Las variables demográficas nos ayudan a entender el perfil del solicitante y su estabilidad laboral en relación incluso a su edad."

Aquí podemos apreciar que entre mayor edad hay mayor antigüedad laboral y menor incumplimiento.



Técnicas de Modelado

1

Regresión logística

2

Árboles de decisión

3

Random Forest

Evaluación del Modelo



Métricas Modelo Regresión Logística

Métrica	Clase 0	Clase 1	Macro Avg	Weighted Avg	AUC
Precisión (Precision)	0.83	0.86	0.84	0.84	-
Sensibilidad (Recall)	0.87	0.81	0.84	0.84	-
F1-Score	0.85	0.84	0.84	0.84	-
Exactitud (Accuracy)	-	-	-	0.84	-
AUC	-	-	-	-	0.92

Métricas Modelo Árboles de decisión

Métrica	Clase 0	Clase 1	Macro Avg	Weighted Avg	AUC
Precisión (Precision)	0.80	0.78	0.79	0.79	-
Sensibilidad (Recall)	0.79	0.79	0.79	0.79	-
F1-Score	0.80	0.79	0.79	0.79	-
Exactitud (Accuracy)	-	-	-	0.79	-
AUC	-	-	-	-	0.92

Métricas Modelo Random Forest

Métrica	Clase 0	Clase 1	Macro Avg	Weighted Avg	AUC
Precisión (Precision)	0.85	0.87	0.86	0.86	-
Sensibilidad (Recall)	0.88	0.84	0.86	0.86	-
F1-Score	0.86	0.85	0.86	0.86	-
Exactitud (Accuracy)	-	-	-	0.86	-
AUC	-	-	-	-	0.93

Análisis de las métricas

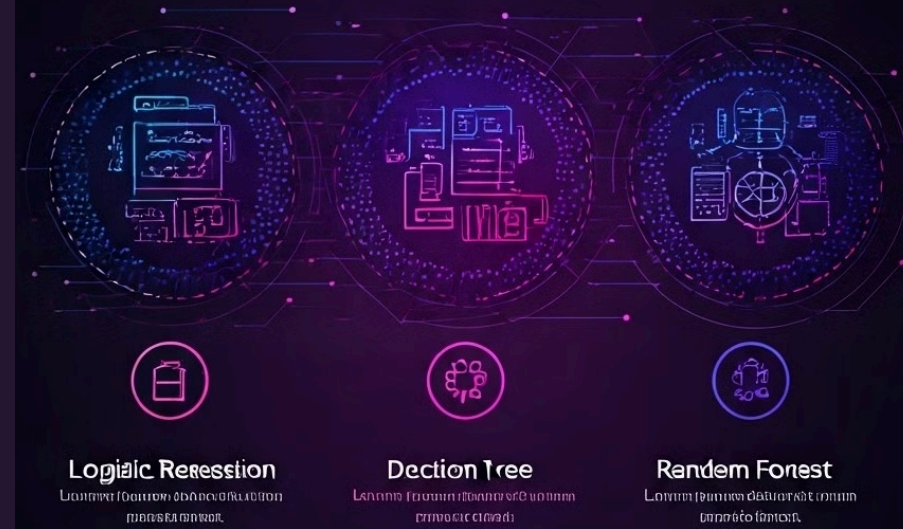
- **Mejor modelo: Random Forest** (mayor exactitud y equilibrio entre clases).
- **Alternativa: Regresión Logística** (buen desempeño y mayor interpretabilidad).
- **No recomendado: Árbol de decisión** (métricas más bajas y propenso a sesgos).



Elección de Modelo

Basado en el análisis anterior, se **recomienda la implementación del Modelo de *Random Forest*** por las siguientes razones:

1. **Mejor desempeño general:** Tiene las métricas más altas en precisión, sensibilidad, F1-Score y exactitud.
2. **Equilibrio entre clases:** Aunque todos los modelos tienen un buen equilibrio entre las clases, el **Random Forest** es ligeramente mejor.
3. **Mayor confiabilidad:** La mayor exactitud y F1-Score indican que el **Random Forest** es más confiable para predecir ambas clases.





Desafíos

Datos desbalanceados

Sesgos algorítmicos

Interpretabilidad del modelo

Elección del modelo correcto

Implementación

1

Limpieza y tratamiento
de los datos

2

Desarrollo del modelo

3

Pruebas y validación

4

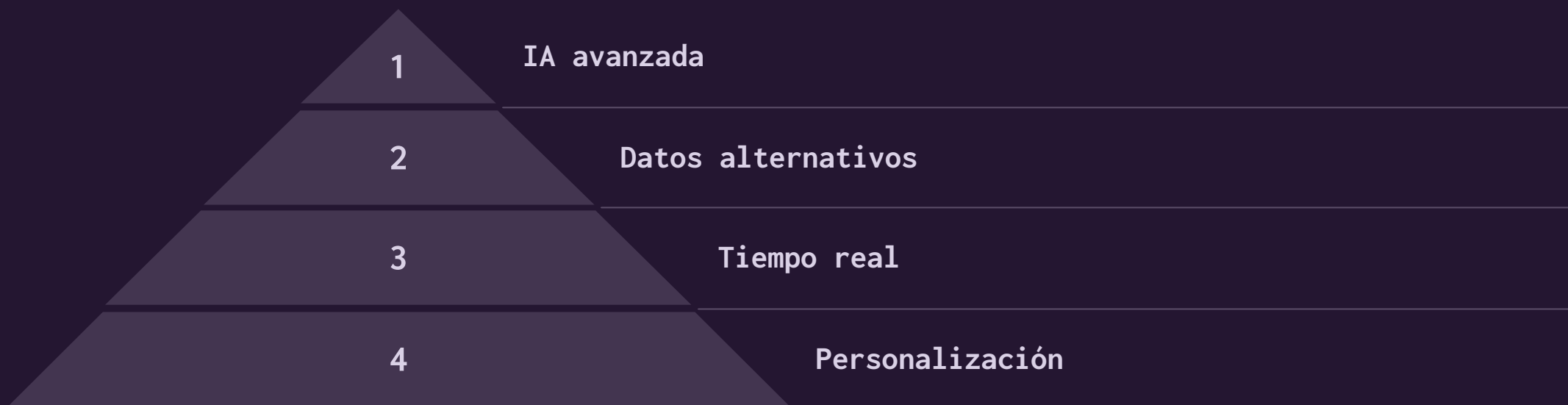
Integración en sistemas

5

Monitoreo continuo



Futuro del Credit Scoring



Conclusión

"En resumen, el **Credit Scoring con Machine Learning** no es solo una herramienta; es una revolución. Con modelos como el **Random Forest**, hemos logrado una exactitud del **86%**, reduciendo riesgos y abriendo nuevas oportunidades de negocio. Pero esto no es el final; es solo el comienzo.

Cada dato que analizamos, cada predicción que hacemos, nos acerca a un futuro donde las decisiones crediticias son más justas, más rápidas y más inteligentes. Y ustedes tienen el poder de ser parte de este cambio.

¿Están listos para transformar la manera en que evaluamos el riesgo crediticio? ¿Listos para tomar decisiones que no solo protejan nuestro negocio, sino que también impulsen su crecimiento? El futuro del crédito está aquí, y comienza hoy. ¡Actuemos juntos!"