

Resumen Ejecutivo: Análisis Discriminante

Mauricio Loera Abundis

Gerardo Javier López García

Emilio Guillen Ramírez

09 de noviembre de 2025

1 Problema y Objetivo Empresarial

LendSmart enfrenta actualmente una tasa de incumplimiento (default) de cartera del 28%, un nivel que la dirección considera insosteniblemente alto. El objetivo de este trabajo fue analizar los datos históricos de solicitudes de préstamo (2022 - 2024) contenidos en el archivo `credit_risk_data.csv` para desarrollar un modelo capaz de predecir la probabilidad de que un nuevo solicitante incurra en *pagado o default*.

El problema central de negocio implica un equilibrio entre dos costos:

- **Aprobar un “mal” Préstamo (Falso Negativo):** pérdida financiera directa si el solicitante incumple.
- **Rechazar un “buen” Préstamo (Falso Positivo):** costo de oportunidad (intereses no percibidos) y cliente potencial insatisfecho.

Nuestro objetivo fue minimizar la aprobación de préstamos de alto riesgo, manteniendo en niveles aceptables el rechazo de buenos clientes. Para ello comparamos dos técnicas: Análisis Discriminante Lineal (LDA) y Análisis Discriminante Cuadrático (QDA).

2 Hallazgos Clave y Análisis

Basándonos en el análisis de los datos, hemos identificado un perfil claro para los solicitantes con mayor riesgo de incumplimiento. De acuerdo con la selección de características con RFE (estimador LDA), los factores más significativos que aumentan el riesgo de *default* son:

`job_stability_score`, `credit_utilization`, `payment_history_score`, `debt_to_income_ratio`.

En términos de negocio, estos factores capturan estabilidad laboral, disciplina de pago histórica, presión de deuda relativa al ingreso y uso del crédito vigente. La **combinación** de estas señales resultó ser totalmente informativa para segmentar clientes que pagan de quienes incumplen.

Perfil de alto riesgo. Sin entrar a coeficientes técnicos, la interpretación bancaria estándar sugiere que:

- *Mayor credit_utilization* y *mayor debt_to_income_ratio* se asocian con mayor probabilidad de *default*.
- *Mejores* puntajes en `payment_history_score` y `job_stability_score` se asocian con menor probabilidad de *default*.

Esta señal es coherente con la intuición del negocio de crédito y respalda reglas de originación más estrictas en solicitantes con alta utilización y alta carga de deuda.

Nota metodológica. La base de datos proporcionada contiene relaciones prácticamente perfectas entre predictores y la variable objetivo. En la práctica, esto implica separabilidad casi perfecta de clases: no hay ruido informativo relevante, no hay mediciones faltantes ni contradictorias, y las correlaciones entre variables son consistentes con el *label*. Bajo este escenario, cualquier clasificador razonable que respete el preprocesamiento estándar converge a métricas perfectas (1.00). Por ello, los resultados reportados muestran un desempeño idéntico (y máximo) tanto para LDA como para QDA.

3 Rendimiento del Modelo

Se ajustaron LDA y QDA con el conjunto de entrenamiento estandarizado y se evaluaron en el conjunto de prueba. Los resultados en *test* fueron:

LDA (test):

- **Accuracy:** 1.00 (500 observaciones)
- **Precision clase 0:** 1.00 **Recall clase 0:** 1.00 **F1:** 1.00 (*support* = 367)
- **Precision clase 1:** 1.00 **Recall clase 1:** 1.00 **F1:** 1.00 (*support* = 133)
- **Promedios** (macro/ponderado): 1.00 / 1.00 / 1.00
- **Validación cruzada:** 1.000 (\pm 0.000)

QDA (test):

- **Accuracy:** 1.00 (500 observaciones)
- **Precision clase 0:** 1.00 **Recall clase 0:** 1.00 **F1:** 1.00 (*support* = 367)
- **Precision clase 1:** 1.00 **Recall clase 1:** 1.00 **F1:** 1.00 (*support* = 133)
- **Promedios** (macro/ponderado): 1.00 / 1.00 / 1.00
- **Validación cruzada:** 1.000 (\pm 0.000)

En el conjunto de datos de prueba, tanto el modelo LDA como el QDA lograron métricas perfectas: Precisión, Exactitud y un AUC (Área bajo la Curva) del 100%. Fundamentalmente para el objetivo empresarial, ambos modelos alcanzaron un Recall (Sensibilidad) del 100%.

Esto significa que el modelo fue capaz de identificar correctamente al 100% de los solicitantes que, de hecho, incumplirían su pago. Esto aborda directamente el objetivo principal de minimizar la aprobación de préstamos incobrables.

4 Recomendación Final

Propuesta. Implementar **LDA** como modelo operativo de primera línea. Aunque QDA y LDA obtuvieron métricas idénticas en esta base, LDA ofrece dos ventajas clave:

- **Interpretabilidad y auditoría:** LDA entrega relaciones lineales claras con cada predictor, lo que facilita explicar decisiones a negocio, cumplimiento y auditoría.
- **Estabilidad y simplicidad:** Con menor complejidad paramétrica, LDA es menos propenso a inestabilidades en *drift* o pequeñas desviaciones distribucionales cuando se escale a producción.

Implicaciones para la gestión del riesgo. Con LDA, las operaciones podrán:

1. **Reducir** la aprobación de créditos de alto riesgo, clave para bajar el 28% de *default*.
2. **Mantener** la fricción comercial controlada, gracias a umbrales de decisión explicables y trazables.

Siguiente paso sugerido. Antes del despliegue definitivo, se sugiere realizar una prueba de robustez con:

- **Validación temporal:** (entrenar en 2022–2023 y probar en 2024).
- **Inyección controlada de ruido/valores faltantes:** para simular condiciones más cercanas a producción.

Si el desempeño permanece alto y estable, se debe avanzar con LDA y establecer un monitoreo de *drift* y *thresholding* por segmento de solicitante.