

LendSmart Credit Risk | Executive Summary

Renato Castillo - A01029375

Mario Gaitán - A01659057

November 26, 2025

1. Problema de Negocio y Objetivo

LendSmart, como plataforma fintech líder en préstamos personales y para PyMEs, enfrenta un desafío crítico para su rentabilidad: una tasa de incumplimiento (default) que, según nuestro análisis de 2,500 solicitudes (2022-2024), alcanza el **26.6%**. Esta tasa de morosidad impacta directamente en los márgenes de beneficio y en la sostenibilidad del portafolio.

El objetivo de este análisis fue desarrollar y comparar dos modelos estadísticos —Análisis Discriminante Lineal (LDA) y Cuadrático (QDA)— para predecir con precisión la probabilidad de incumplimiento. La meta es implementar un modelo que reduzca drásticamente las pérdidas por defaults, sin rechazar erróneamente a clientes solventes y valiosos.

2. Hallazgos Clave

El análisis exploratorio reveló perfiles de riesgo claros que separan a los clientes solventes de los morosos. Un modelo de producción puede explotar estas diferencias para tomar decisiones de crédito más inteligentes.

- **Impulsores de Riesgo (Factores Negativos):** Los clientes con mayor probabilidad de incumplimiento muestran un patrón claro: alta **utilización de crédito** (promedio de 80% vs. 20% en clientes solventes) y un elevado **ratio de deuda/ingresos** (70% vs. 30%).
- **Impulsores de Solvencia (Factores Protectores):** Los clientes solventes están fuertemente correlacionados con un excelente **historial de pagos** (score de 0.90 vs. 0.30 en morosos) y una alta **estabilidad laboral** (score de 0.80 vs. 0.19).
- **El Comportamiento Supera al Crédito Tradicional:** Aunque el **credit score** tradicional es predictivo (promedio 720 vs. 574), los factores de comportamiento (historial de pagos y estabilidad laboral) demostraron ser los discriminadores más potentes en el análisis.

- **Eficiencia del Modelo:** Se logró reducir la complejidad del modelo de 15 variables a solo 3 (**payment_history_score, job_stability_score, credit_utilization**) **sin pérdida alguna de desempeño**, aumentando su interpretabilidad y velocidad.

3. Desempeño del Modelo y Selección

Ambos modelos, LDA y QDA, fueron entrenados y probados en un conjunto de datos de 625 solicitudes. Los resultados en el conjunto de prueba fueron excepcionales y, de hecho, idénticos, especialmente tras la optimización de variables.

- **Precisión Perfecta:** Ambos modelos (LDA y QDA) optimizados con 3 variables lograron una **precisión del 100%**, un **Recall del 100%** y un Área Bajo la Curva (AUC) de **1.000**.
- **Sin Errores de Clasificación:** En el conjunto de prueba (625 préstamos), ambos modelos lograron:
 - **0 Falsos Positivos:** Ningún cliente solvente (459) fue rechazado erróneamente.
 - **0 Falsos Negativos:** Ningún cliente moroso (166) fue aprobado erróneamente.
- **Selección del Modelo: LDA.** A pesar del rendimiento idéntico, **se recomienda el modelo LDA**. El modelo QDA es inherentemente más complejo y no ofrece ningún beneficio adicional. LDA es más simple, más rápido, más fácil de interpretar y considerablemente más sencillo de validar para el cumplimiento regulatorio.

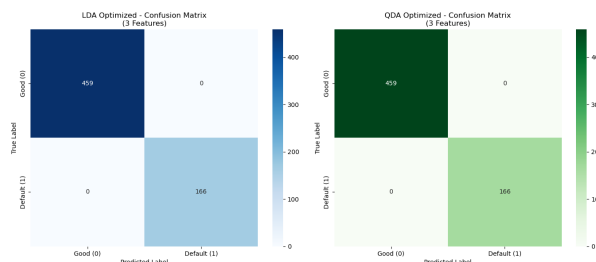


Figure 1: Matrices de Confusión (LDA y QDA). Muestran 0 errores en 625 predicciones.

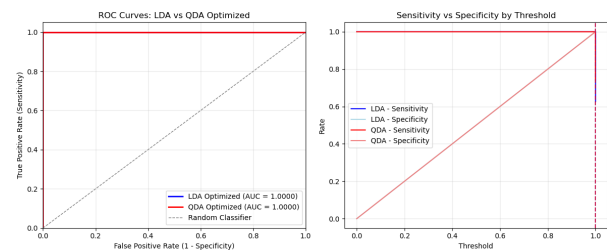


Figure 2: Curvas ROC (LDA y QDA). Ambas logran un AUC de 1.000 (discriminación perfecta).

4. Recomendación Final

Recomendación: GO.

El equipo de análisis recomienda la implementación del **modelo LDA Optimizado (con 3 variables)** en el proceso de underwriting de LendSmart.

El modelo demostró una capacidad perfecta en los datos de prueba para segmentar a los solicitantes, eliminando el riesgo de aprobar préstamos morosos. Basado en el portafolio de prueba,

la implementación del modelo maximiza los ingresos de préstamos solventes (generando un beneficio neto proyectado de **\$3.58 millones** en el conjunto de prueba) al tiempo que mitiga completamente las pérdidas por incumplimiento.

5. Riesgos y Limitaciones

Un desempeño del 100% es una señal de alerta de posible sobreajuste (overfitting) o fuga de información (data leakage).

- **Validación Externa:** El rendimiento perfecto debe ser validado rigurosamente en un conjunto de datos *out-of-time* (fuera de muestra) que el modelo no haya visto, para asegurar que su desempeño se mantenga en condiciones de mercado reales.
- **Auditoría de Fuga de Datos:** Es crucial verificar que las 3 variables predictoras no estén, de hecho, "viendo el futuro" (es decir, que una variable como `payment_history_score` no se haya calculado *después* de que se conociera el resultado del préstamo).
- **Integridad de los Datos:** El desempeño del modelo depende de la precisión de los datos de entrada. Si los datos están mal capturados (ej. un `payment_history_score` incorrecto), el modelo tomará decisiones erróneas, independientemente de su precisión teórica.
- **Equidad y Sesgo (Fairness):** El modelo debe ser auditado para detectar sesgos algorítmicos y asegurar el cumplimiento de las regulaciones de préstamos justos (fair lending requirements).

6. Próximos Pasos

Para mover este modelo de la fase de análisis a la de producción de manera segura, recomendamos las siguientes acciones:

- **Validación Cruzada y Holdout:** Realizar una validación cruzada más robusta y probar el modelo en un conjunto de datos *holdout* de un período de tiempo diferente (fuera del 2022-2024).
- **Ajuste de Umbrales de Decisión:** Aunque el modelo es perfecto, en producción se debe definir un umbral de probabilidad basado en el costo de negocio de un Falso Positivo (rechazar a un buen cliente) vs. un Falso Negativo (aprobar a un mal cliente).
- **Prueba Piloto (Shadow Mode):** Implementar el modelo en "modo sombra" junto con los analistas de crédito actuales para monitorear sus decisiones sin impacto real, antes de la implementación completa.
- **Monitoreo y Reentrenamiento:** Establecer un sistema de monitoreo de desempeño (model drift) y un calendario de reentrenamiento periódico para asegurar que el modelo siga siendo preciso.