

# 1 Hoja de Instrucciones

## 1.1 Alcance del modelo

El sistema predictivo desarrollado estima la probabilidad de que un cliente caiga en mora ( $\geq 30$  días) antes de otorgar un crédito. Para ello se integraron tres componentes analíticos que cumplen funciones distintas dentro del proceso:

- LDA: Se usó solo para interpretar el riesgo. Muestra qué variables separan mejor a morosos y no morosos. No es el modelo final.
- XGBoost (modelo principal): Es el clasificador final por su buen desempeño y estabilidad. Se mejoró con SMOTE y ajuste de umbral para detectar más clientes en riesgo.
- KMeans + PCA (segmentación): Se emplearon para agrupar clientes en perfiles de riesgo y apoyar decisiones diferenciadas.

## 1.2 Datos utilizados

El modelo se entrenó con la base anonimizada proporcionada por la institución y únicamente con variables disponibles antes de aprobar el crédito, es decir, información que el analista realmente conoce al evaluar una solicitud. Las variables utilizadas fueron: SCORE, MONTO DESEMBOLSADO, EDAD, NIVEL DE ESTUDIOS, INGRESOS FIJOS, GASTOS DE SOSTENIMIENTO, ACTIVOS, TIEMPO EN ACTIVIDAD, TIEMPO EN RESIDENCIA, CÓDIGO CIU, ESTADO CIVIL y ESTRATO. Estas características describen la situación económica, laboral y demográfica del solicitante, por lo que son adecuadas para estimar riesgo previo al desembolso. No se incluyó ninguna variable que dependa del comportamiento posterior del crédito.

## 1.3 Supuesto del modelo

El modelo asume que la información registrada por Dentix es completa, consistente y capturada sin errores importantes. También se requiere que los datos usados en producción mantengan la misma estructura que los utilizados durante el entrenamiento. Finalmente, el umbral operativo recomendado (0.15) debe aplicarse de manera uniforme para garantizar coherencia en las decisiones de crédito.

## 1.4 Métricas clave del modelo

Las métricas principales para evaluar el desempeño del modelo en Dentix son el recall y la precisión, porque determinan qué tan bien se identifican los clientes morosos y qué tan confiables son las alertas emitidas. El recall de morosos es la métrica más crítica, ya que mide cuántos clientes de alto riesgo se detectan correctamente antes de financiar un tratamiento. La precisión complementa este

criterio indicando qué proporción de las alertas realmente corresponde a casos de mora.

El F1-score resume el equilibrio entre ambas métricas en un contexto de desbalance de clases, mientras que la matriz de confusión permite visualizar de forma clara los aciertos, falsos positivos y falso negativos del modelo. En conjunto, estas métricas permiten validar si el sistema mantiene un desempeño estable y adecuado para la operación crediticia de Dentix.

## 1.5 Riesgos del modelo

El modelo depende totalmente de la calidad y consistencia de los datos capturados por Dentix. Errores en variables clave como ingresos, score o antigüedad pueden generar decisiones incorrectas. Un punto crítico es la variable TOTALINGRESOS, que actualmente no puede usarse para el análisis debido a su alto nivel de valores faltantes, sin embargo, es una variable relevante y se recomienda mejorar su captura, ya que podría incrementar la capacidad predictiva del modelo en futuras versiones. Otro riesgo es el sobre-rechazo: si el umbral se baja demasiado, se podrían negar tratamientos a clientes que sí pagarían. Por ello, se recomienda monitorear el modelo de manera periódica y recalibrarlo cuando la composición de la cartera o las condiciones del negocio cambien.

## 1.6 Uso adecuado del modelo

El modelo debe utilizarse siempre con los mismos campos y estructura con los que fue entrenado. La probabilidad estimada debe interpretarse comparándola con el umbral operativo de 0.15: valores por encima indican mayor riesgo de mora y requieren una revisión más estricta, mientras que valores cercanos al umbral (por ejemplo 0.12–0.18) deben analizarse manualmente antes de aprobar un crédito. Si Dentix desea reducir el riesgo de incumplimiento, debe bajar el umbral, ya que esto aumenta la detección de morosos (menos falsos negativos), aunque reduce la tasa de aprobación. Si Dentix desea aprobar más volumen, puede subir el umbral, entendiendo que esto disminuye la detección de clientes riesgosos (más falsos negativos) y aumenta la exposición a mora.

Los cluster obtenidos por KMeans sirven para complementar la decisión. Dentix puede identificar el cluster de un cliente usando las mismas variables estandarizadas y luego visualizar su posición en el gráfico PCA, lo que permite comparar su perfil con los tres segmentos detectados: bajo riesgo, intermedio y alto riesgo. La primera componente (PC1) resume estabilidad laboral, residencia y score, la segunda (PC2) resume condiciones financieras. Si un nuevo cliente aparece en la zona cluster estable, puede aprobarse con mayor confianza, si cae en el cluster vulnerable, debe revisarse con mayor detalle, sin depender únicamente de la probabilidad. Esta combinación de umbral + cluster + juicio humano asegura un uso adecuado del modelo dentro de Dentix.