










PREDICTION TASK  <p>Este proyecto desarrolla un sistema de Machine Learning para automatizar la evaluación de riesgo crediticio. Al analizar los datos históricos de los solicitantes, el modelo genera un puntaje de riesgo objetivo que permite a la institución financiera minimizar las pérdidas por incumplimiento de pagos, a la vez que acelera y estandariza el proceso de aprobación de créditos para clientes fiables.</p>	DECISIONS  <p>El puntaje generado se convierte en una decisión binaria: aprobar o rechazar el crédito.</p> <p>También puede dar lugar a reglas intermedias (ej. "aprobar con condiciones", "revisar manualmente").</p> <p>Parámetros clave:</p> <ul style="list-style-type: none"> • Umbral de probabilidad de incumplimiento. • Monto máximo de crédito recomendado. • Condiciones adicionales (ej. aval, seguro) 	VALUE PROPOSITION  <p>Un sistema automatizado que:</p> <ul style="list-style-type: none"> - Reduce pérdidas al identificar mejor a los clientes de alto riesgo - Acelera las decisiones de crédito. - Estandariza la evaluación de solicitudes de forma objetiva. 	DATA COLLECTION  <p>Fuente inicial: dataset histórico de clientes con atributos demográficos, financieros y comportamiento crediticio.</p> <p>Estrategias de actualización:</p> <ul style="list-style-type: none"> • Integración con bases de datos bancarias en tiempo real. • Automatización de cargas diarias/semanales. Validaciones de calidad de datos (valores faltantes, duplicados, inconsistentes). 	DATA SOURCES  <p>El proyecto utiliza el German Credit Dataset de la UCI Machine Learning Repository, preparado originalmente por el Prof. Hofmann. Este dataset contiene 1000 registros de solicitantes de crédito con 20 atributos categóricos y numéricos.</p>
IMPACT SIMULATION  <p>Costo de falsos negativos: aprobar a un cliente de alto riesgo → pérdida financiera.</p> <p>Costo de falsos positivos: rechazar a un cliente confiable → pérdida de oportunidad.</p> <p>Antes del despliegue: simular con datos históricos (cross-validation).</p> <p>Restricciones de equidad: evitar sesgos contra grupos por edad, género, estado civil.</p>	MAKING PREDICTIONS  <p>Predicciones en tiempo casi real (al recibir solicitud de crédito).</p> <p>Latencia aceptable: < 1 segundo.</p> <p>Batch posible para monitoreo y reporting semanal/mensual.</p> <p>Recursos: servidor en la nube con capacidad de ejecutar modelos ML (ej. sklearn, XGBoost).</p>		BUILDING MODELS  <p>Modelos candidatos:</p> <ul style="list-style-type: none"> • Regresión logística (baseline, interpretable). • Árboles de decisión / Random Forest / Gradient Boosting (mayor performance). <p>Actualización:</p> <ul style="list-style-type: none"> • Reentrenar cada 6-12 meses con datos más recientes. <p>Recursos:</p> <ul style="list-style-type: none"> • GPU no es indispensable (dataset pequeño). • CPU/memoria suficiente en entorno productivo. 	FEATURES  <p>El data set se compone de 1000 registros de solicitantes de crédito y 20 atributos categóricos y numéricos. Las variables se agrupan en tres bloques principales:</p> <ul style="list-style-type: none"> • Datos demográficos • Información financiera • Condiciones del crédito y garantías <p>La variable objetivo clasifica a cada cliente como "buen riesgo" o "mal riesgo" crediticio, lo que convierte el problema en una tarea de clasificación supervisada.</p>
	MONITORING <p>Métricas técnicas:</p> <ul style="list-style-type: none"> • Accuracy, Precision, Recall, F1, ROC-AUC. • Drift en distribución de datos (para detectar cambios en clientes) <p>Métricas de negocio:</p> <ul style="list-style-type: none"> • % de créditos incobrables después de aplicar el modelo. • Tasa de aprobación. 			