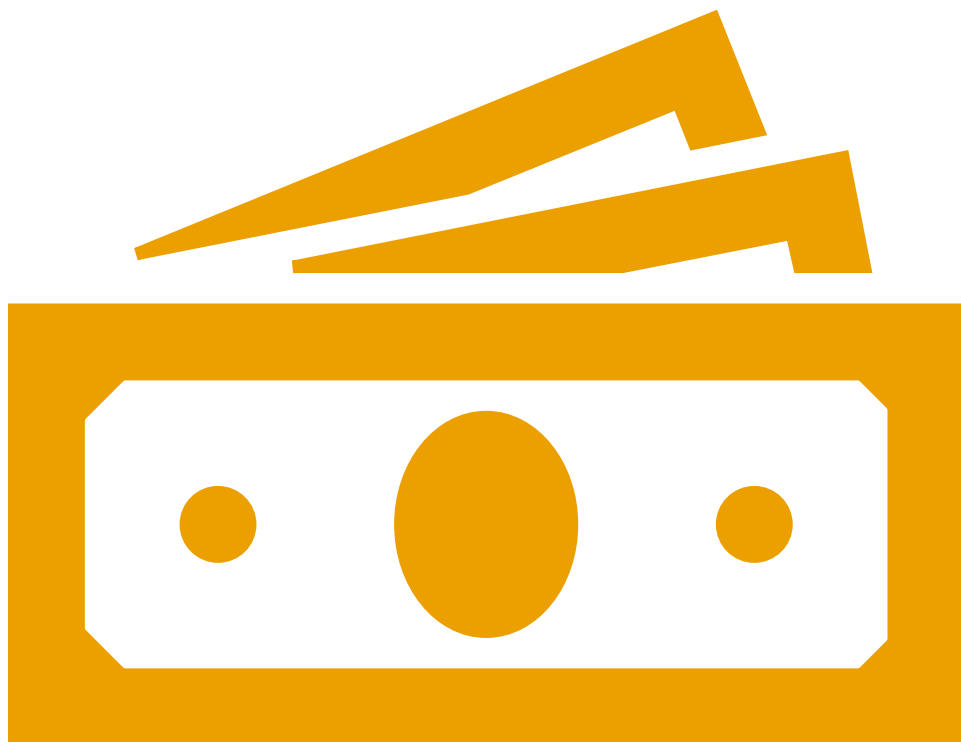

Modelo de Riesgo de Impago para Financiera Horizonte S.A.

Proyecto MLOps 3

Edwin de Leon





Contexto del Proyecto

- Empresa: **Financiera Horizonte S.A. (FHSA).**
- Sector: **Préstamos de consumo y microcréditos.**
- Problema: aumento de morosidad **5% → 9%** tras recesión.
- Objetivo: **predecir riesgo de impago a 12 meses** para originación digital.
- Impacto económico:
 - Pérdida actual: **USD 450,000 / año.**
 - Ahorro esperado con modelo: **USD 135,000 / año.**
 - ROI: **< 1 año.**

2. Stakeholders Clave

Stakeholder	Interés	Requisito
CFO	Financiero y ROI	Ahorros cuantificables, tiempo de implementación
Gerente de Riesgo	Precisión y trazabilidad	Recall alto, explicabilidad, auditoría
Producto/Marketing	Escalar originación	Mantener tasa de aprobación >70%
Equipo ML/Datos	Implementación	Pipeline reproducible, validación cruzada temporal
Operaciones de crédito	Reglas de negocio	Integración con flujo manual, casos especiales
Legal/Compliance	Regulaciones	Trazabilidad, logs, explicabilidad para auditores

Necesidades del Negocio



Modelo con alta sensibilidad a defaults ($\text{recall} \geq 0.80$) pero manteniendo precisión razonable para no cortar crecimiento



Interpretabilidad: Importancia de features para explicar rechazos a clientes y auditores



Latencia: Inferencia $< 200\text{ms}$ en promedio para integración en flujo de originación



Métricas definidas: AUC-ROC, curva Precision-Recall, tasa de falsos positivos aceptable según política



Integración: Pipeline de scoring en producción (APIs, logs, monitoreo de drift)

Estructura del dataset

1. Datos del cliente

customer_id, age, gender, country

2. Actividad y uso de la plataforma

signup_date, last_login_days, total_orders, last_purchase_date, days_since_last_purchase

3. Comportamiento de compra

avg_order_value, promotion_usage, loyalty_points

4. Interacción y soporte

support_tickets, payment_issues

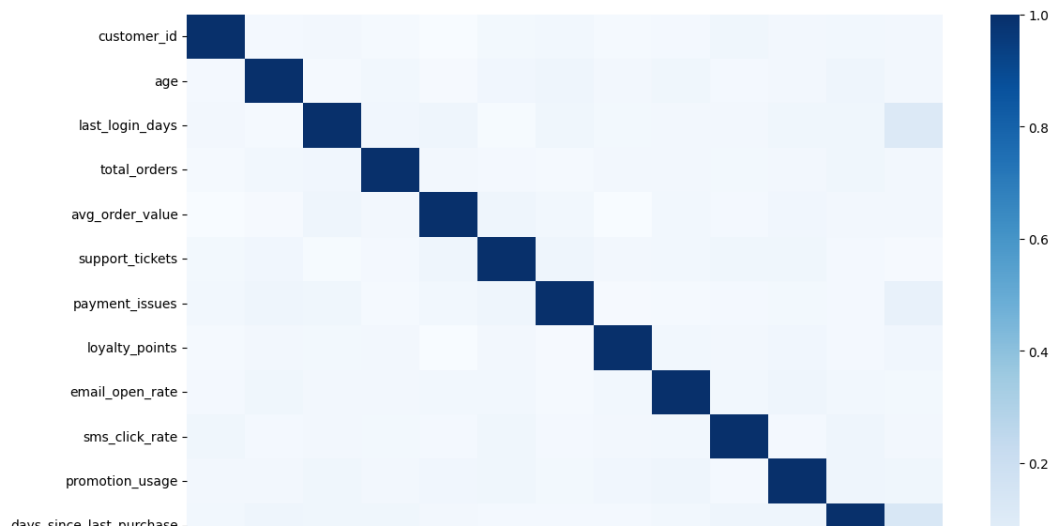
5. Engagement con marketing

email_open_rate, sms_click_rate

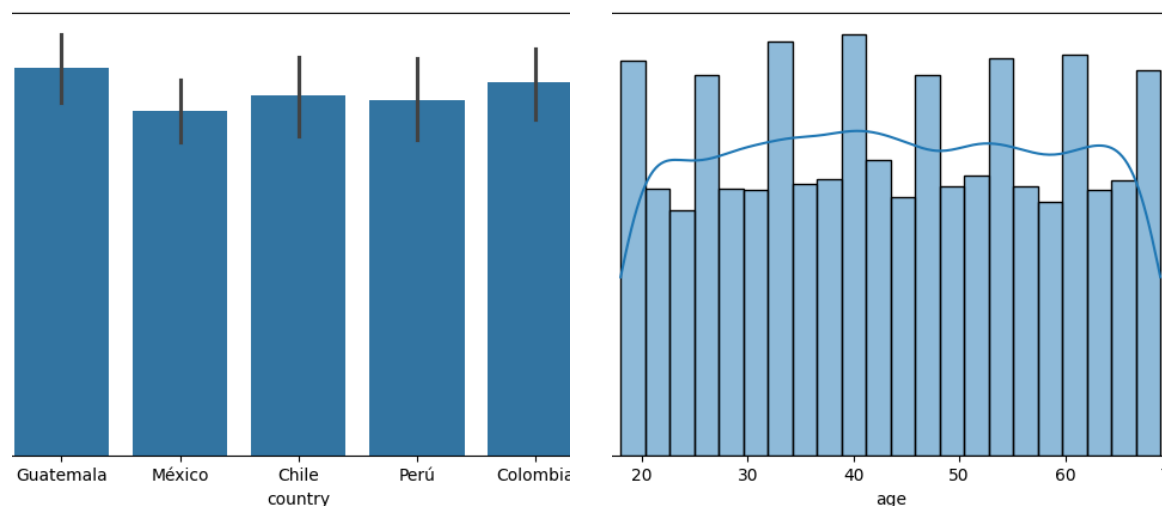
Variable objetivo

churn (0 = activo, 1 = abandono)

Comprensión de Datos



- **Churn por país:** Guatemala reporta el nivel más alto; México el más bajo. Existen diferencias geográficas relevantes.
- **Correlación:** No se observa multicolinealidad; la mayoría de variables son independientes.
- **Edad:** Distribución uniforme entre 18 y 70 años; no hay concentraciones marcadas.
- **Conclusión:** El dataset es variado, equilibrado y adecuado para modelos de clasificación que capturen relaciones no lineales.

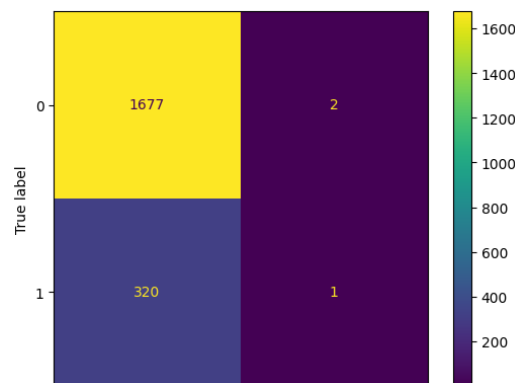
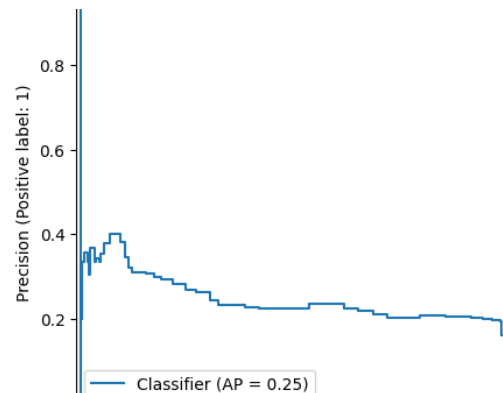
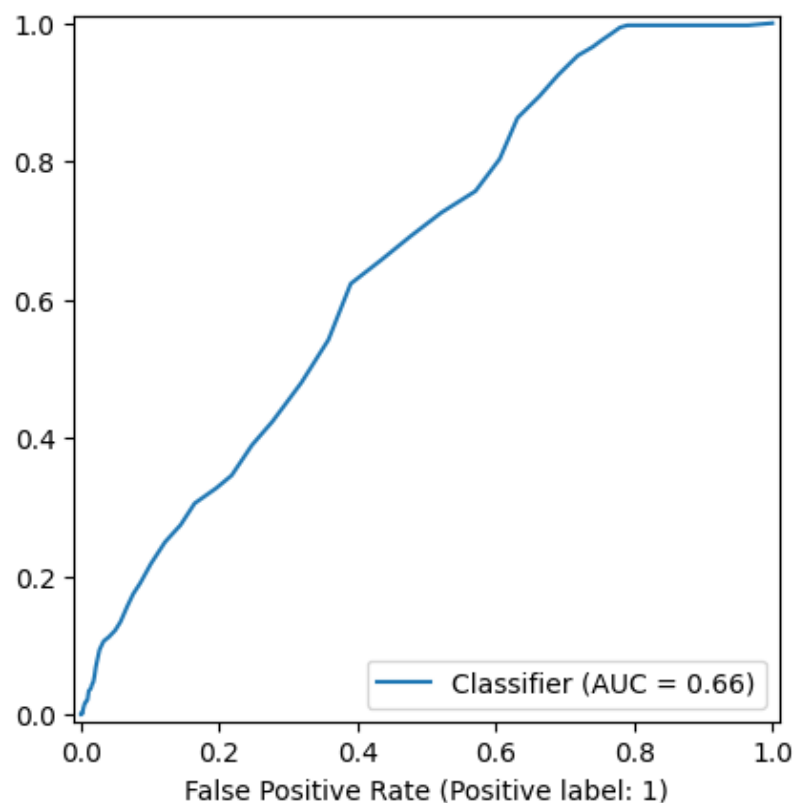


PREPARACIÓN DE LOS DATOS

- **Data Preparation Resumen:**
- No se encontraron duplicados (50k filas).
- Missing imputados con medianas por grupo y KNN.
- Outliers tratados con IQR.
- Se añadieron features clave (ratio, logs, flags, age bands).
- División temporal: 2023-12-31 → Train/Test separados correctamente.
- 36 variables finales listas para modelado.

```
=== LIMPIEZA DE DATOS ===  
Filas iniciales: 50000  
Filas tras eliminar duplicados: 50000  
  
=== TRATAMIENTO DE MISSING VALUES ===  
Missing antes de imputación:  
monthly_income      3460  
credit_score_proxy   2473  
dtype: int64  
Missing después de imputación: 2473  
  
=== TRATAMIENTO DE OUTLIERS ===  
✓ Outliers copiados  
  
=== FEATURE ENGINEERING ===  
Nuevas features creadas: income_to_loan_ratio, log_loan_amount, recent_delinquency_ratio, etc.  
Features finales: 36  
  
=== DIVISIÓN TEMPORAL (Train/Test) ===  
Cutoff: 2023-12-31  
Train: 42879 filas (17.60% defaults)  
Test: 7121 filas (17.81% defaults)  
  
✓ Datasets guardados en processed/
```

MODELADO + MLflow



Curva ROC — AUC = 0.66

- El modelo obtiene un AUC de 0.66, lo cual indica una capacidad moderada para distinguir entre clientes que hacen churn y los que no.

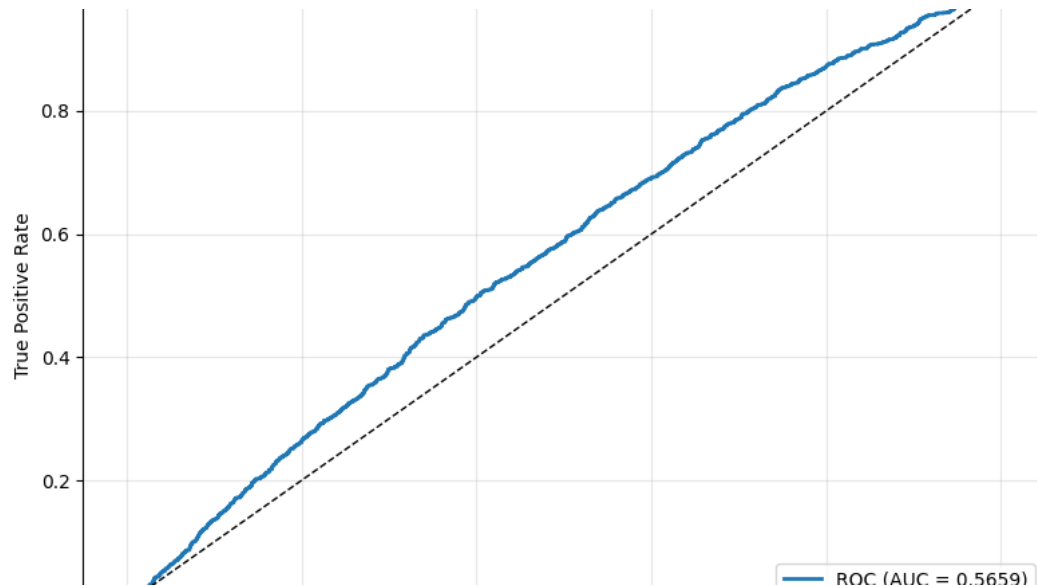
Curva Precision-Recall — AP = 0.25

- El Average Precision (AP) es 0.25, reflejando que la precisión disminuye rápidamente cuando aumenta el recall.

Matriz de Confusión

- El modelo casi no comete falsos positivos, pero falla en capturar la mayoría de los churn reales.

VALUACIÓN Y RESULTADOS



- El modelo no detecta ningún positivo (recall = 0).
- Matriz de confusión: todos los casos predichos como negativos.
- AUC = 0.56 → desempeño apenas superior al azar.
- El modelo debe mejorarse antes de cualquier uso real.

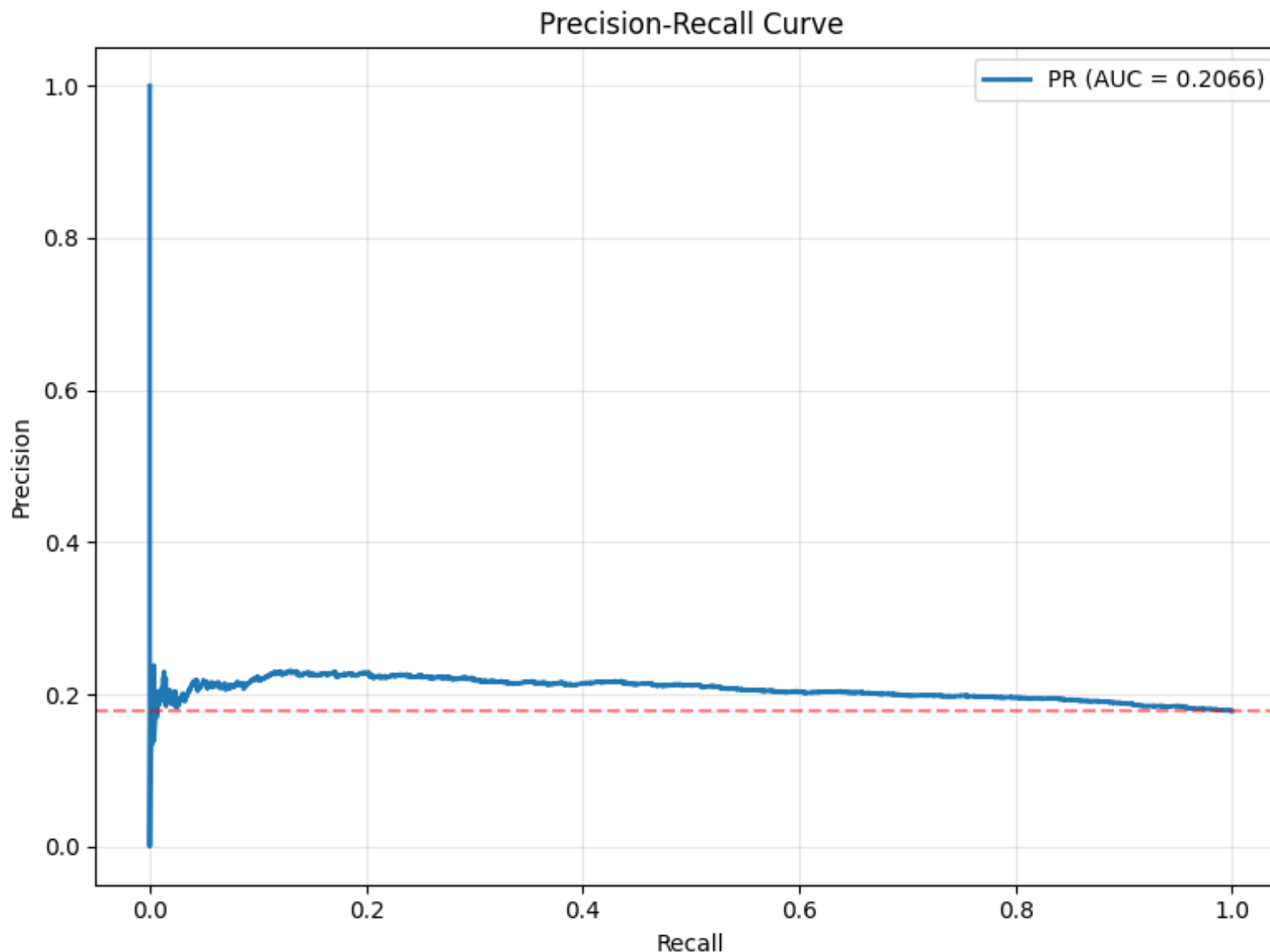
```
X_train shape: (42879, 33), y_train: (42879,)
X_test shape: (7121, 33), y_test: (7121,)

=== ENTRENANDO MODELO ===
✓ Modelo entrenado

=== MATRIZ DE CONFUSIÓN ===
TN=5853, FP=0, FN=1268, TP=0
Recall (sensitivity): 0.0000
Precision: nan
Specificity: 1.0000
F1-Score: 0.0000
```

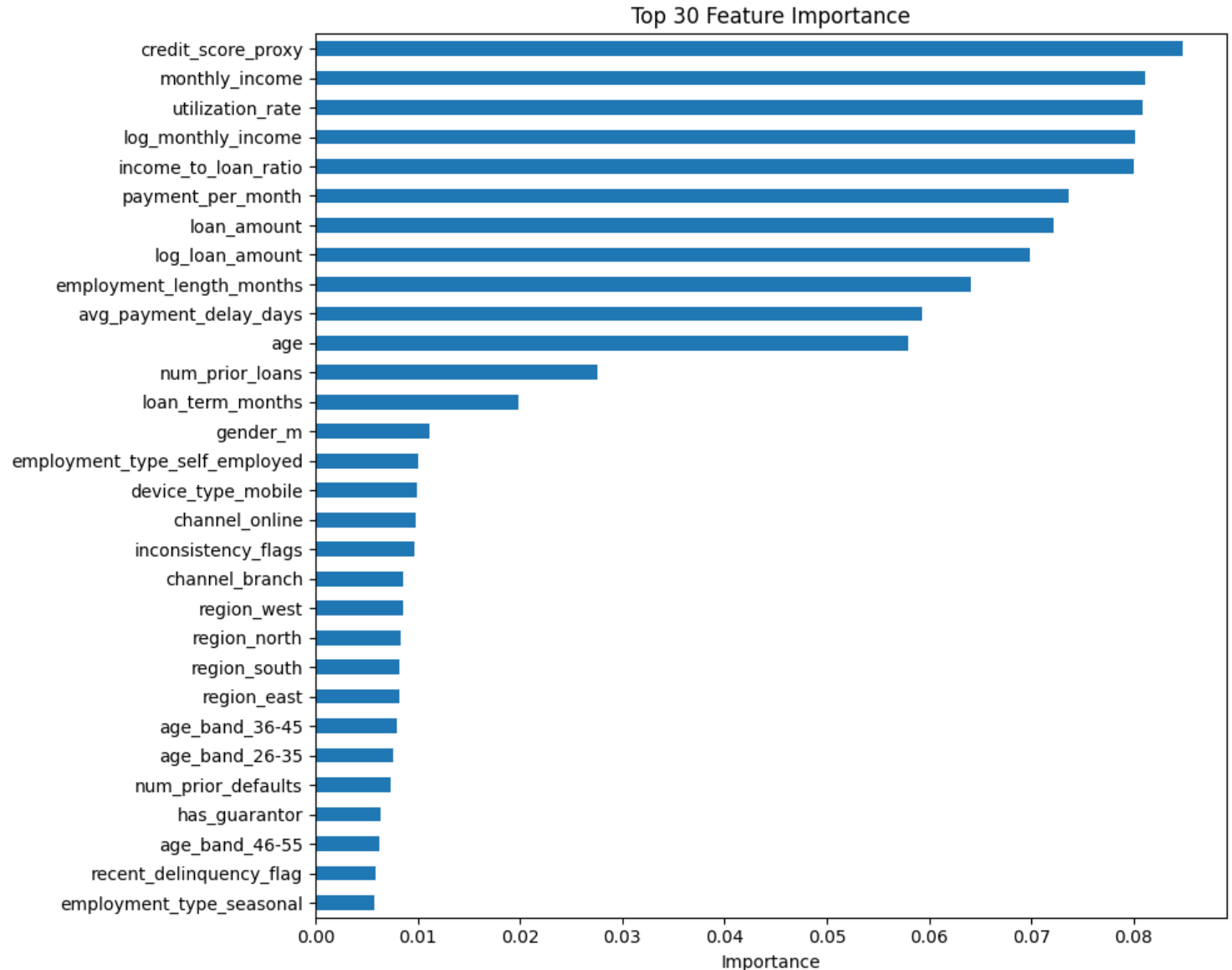
PRECISION- RECALL CURVE AUC-PR: 0.2066

- Modelo tiene desempeño bajo en detectar positivos.
- La precisión cae rápido al aumentar recall - patrón típico de alta confusión entre clases.
- El resultado está solo ligeramente por encima del baseline.
- Necesita mejoras en balanceo y modelado antes de poder usarse en producción.



•FEATURE IMPORTANCE (TOP 30)

- Las variables más importantes están relacionadas con score crediticio, ingresos, relación deuda/ingreso y monto del préstamo.
- Factores de estabilidad laboral y comportamiento de pago también son muy influyentes.
- Variantes demográficas y de canal tienen impacto menor.
- Las features creadas (como ratios y logs) están entre las más relevantes, confirmando un buen feature engineering.



```
Falsos Negativos: 1268 (100.00% de activos)
Falsos Positivos: 0 (0.00% de inactivos)
```

```
=== IMPACTO EN NEGOCIO ===
```

```
Threshold: 0.5
```

```
Solicitudes aceptadas: 7121 (100.0%)
```

```
Solicitudes rechazadas: 0 (0.0%)
```

```
Defaults evitados: 0 / 1268 (0.0%)
```

```
Clientes buenos rechazados erróneamente: 0
```

```
=== VALIDACIÓN CRUZADA TEMPORAL ===
```

```
Fold AUC: 0.5366
```

```
Fold AUC: 0.5441
```

```
Fold AUC: 0.5351
```

```
Fold AUC: 0.5414
```

```
Fold AUC: 0.5494
```

```
Mean CV AUC: 0.5413 ± 0.0052
```

```
=== RESUMEN DE MÉTRICAS ===
```

```
AUC-ROC: 0.5659
```

```
AUC-PR: 0.2066
```

```
Recall: 0.0000
```

```
Precision: nan
```

```
Specificity: 1.0000
```

```
F1-Score: 0.0000
```

ANÁLISIS DE ERRORES

- El modelo no detecta ningún caso positivo (recall = 0%).
- 100% de los defaults pasan desapercibidos.
- AUC en CV = 0.54 - desempeño apenas superior al azar.
- Umbral 0.5 inutiliza el modelo para decisiones.
- Causa principal: desbalance + modelo demasiado simple + threshold inapropiado.

Conclusión General

- El pipeline de MLOps se implementó de forma completa y coherente con la metodología CRISP-DM, permitiendo un flujo reproducible desde la ingesta de datos hasta el registro del modelo.
- Las fases de preprocesamiento, modelado y tracking con MLflow fueron ejecutadas con éxito, logrando trazabilidad, experimentación estructurada y control de versiones de los modelos.
- El análisis revela que el dataset —aunque sintético— contiene patrones valiosos que pueden ser explotados con algoritmos más avanzados y técnicas específicas para clasificación desbalanceada.
- El proyecto establece una base sólida para escalar hacia un sistema real de scoring crediticio, integrando principios de MLOps, experimentación continua y monitoreo.