

Universidad del Valle de Guatemala

Security Data Science

Jorge Yass

Sección 10



# Proyecto

## ***Métricas custom para reducción de falsos positivos en clasificación binaria-fraude***

Fernanda Esquivel, 21542

Guatemala, 02 de junio de 2025

## Resumen

Este trabajo explora y evalúa estrategias para mejorar la detección de fraudes mediante la implementación de métricas personalizadas en *LightGBM*, con el objetivo específico de reducir los falsos positivos sin comprometer significativamente la capacidad de detección. Utilizando un *dataset* simulado de transacciones con tarjeta de crédito del período 2019-2020 con enfoque en transacciones de madrugada, se desarrollaron cinco funciones de evaluación personalizadas: False Positive Ratio, Precision Oriented, Weighted FP Ratio, Night Optimized y Cost Sensitive.

Los resultados demuestran que las primeras cuatro funciones alcanzaron un desempeño idéntico con un *ratio* de falsos positivos de 2.88, precisión de 34.72% y *recall* de 84.46%, mientras que la función *Cost Sensitive* mostró un desempeño inferior (*ratio* FP: 4.63). El modelo seleccionado logró detectar 652 casos de fraude con 1,226 falsos positivos, representando una mejora significativa en la reducción de alertas innecesarias comparado con métricas tradicionales.

## Metodología

### Análisis Exploratorio de Datos (EDA)

Se realizó un análisis exhaustivo del *dataset* que reveló un fuerte desbalance de clases, con 98.77% de transacciones legítimas versus 1.23% fraudulentas en entrenamiento, y 99.31% versus 0.69% en el conjunto de prueba (diciembre 2020).

### Ingeniería de Variables

1. Eliminación de variables no predictivas (números de tarjeta, información personal, fechas)
2. Conversión de variables categóricas mediante one-hot encoding
3. Manejo de variables booleanas convirtiéndolas a formato numérico
4. Preparación de 77 características finales para el modelado

### Implementación del Modelo Base

Se estableció un modelo *LightGBM* con parámetros optimizados:

- Objective: binary classification
- Scale\_pos\_weight: 80.35 (para manejar desbalance)
- Learning rate: 0.05
- Num\_leaves: 31

- Early stopping: 50 rondas

## Desarrollo de Métricas Personalizadas

Se diseñaron cinco funciones de evaluación (feval) específicas:

- False Positive Ratio: Minimización directa del ratio  $(TP + FP) / TP$
- Precision Oriented: Optimización de precisión con umbrales dinámicos y penalización por recall < 90%
- Weighted FP Ratio: Ratio ponderado con énfasis en mantenimiento de recall  $\geq 90\%$
- Cost Sensitive: Asignación de costos asimétricos (FP cost: 5.0, FN cost: 1.0)
- Night Optimized: Función especializada para transacciones nocturnas con umbral de recall del 85%

## Optimización de Umbrales

Para cada modelo se implementó una búsqueda exhaustiva de umbrales (0.1 a 0.99) priorizando el mantenimiento de alta sensibilidad ( $\geq 85\text{-}90\%$  recall) mientras se minimizaba el ratio de falsos positivos.

## Metodología

### Arquitectura del Modelo

- **Framework:** LightGBM con callbacks de early stopping y logging
- **Validación:** Split estratificado 80/20 para entrenamiento/validación
- **Evaluación:** Conjunto de prueba independiente (diciembre 2020)
- **Métricas de monitoreo:** AUC-ROC, precisión, recall, ratio FP

### Proceso de Entrenamiento

1. **Preparación de datos:** Datasets categorizados para LightGBM con manejo automático de variables categóricas
2. **Entrenamiento iterativo:** Cada función feval entrenada por separado con 1,000 boost rounds máximo
3. **Selección de modelos:** Guardado automático del mejor modelo por función
4. **Optimización de umbrales:** Búsqueda grid sobre conjunto de validación

## Configuración Computacional

- Implementación en Python con pandas, numpy, scikit-learn y lightgbm
- Visualizaciones con matplotlib y seaborn
- Almacenamiento de modelos en formato nativo LightGBM (.txt)

## Análisis de los Resultados de la Evaluación

### Comparativo de Estrategias

Los resultados revelan un patrón interesante donde cuatro de las cinco funciones convergieron al mismo desempeño óptimo:

Función	Umbral	Test de Precisión	Test de Recall	Ratio FP	FP	TP
False Positivo Ratio	0.32	34.72%	84.46%	2.88	1,226	652
Precision Oriented	0.32	34.72%	84.46%	2.88	1,226	652
Weighted FP Ratio	0.32	34.72%	84.46%	2.88	1,226	652
Night Optimized	0.32	34.72%	84.46%	2.88	1,226	652
Cost Sensitive	0.10	21.56%	84.97%	4.64	2,386	656

### Análisis de Convergencia

La convergencia de cuatro funciones al mismo resultado sugiere que existe un óptimo global robusto para este problema específico. El umbral de 0.32 representa un punto de equilibrio efectivo entre sensibilidad y especificidad para el contexto de transacciones nocturnas.

### Desempeño vs Modelo Base

Comparado con métricas tradicionales (AUC-ROC), las funciones personalizadas lograron:

- Reducción significativa de falsos positivos por fraude detectado
- Mantenimiento de alta sensibilidad (>84%)

- Mejora en la precisión operacional del sistema de alertas

## Conclusiones

1. **Efectividad de las Métricas Personalizadas:** Las funciones de evaluación personalizadas demostraron ser superiores a las métricas tradicionales para el objetivo específico de reducir falsos positivos. La convergencia de cuatro funciones al mismo óptimo valida la robustez del enfoque.
2. **Selección del Modelo Óptimo:** La función **False Positive Ratio** se identifica como la más efectiva debido a su implementación directa y simple de la métrica objetivo, alcanzando un ratio de 2.88 falsos positivos por cada verdadero positivo detectado.
3. **Balance Operacional:** El modelo seleccionado logra un equilibrio práctico detectando 84.46% de los fraudes con una tasa de falsos positivos manejable, lo que representa una mejora significativa para sistemas de alerta en tiempo real.
4. **Especialización Temporal:** La optimización específica para transacciones de madrugada demostró ser efectiva, sugiriendo que enfoques temporalmente segmentados pueden mejorar el desempeño en detección de fraude.
5. **Recomendaciones de Implementación:** Se recomienda implementar la función False Positive Ratio con umbral 0.32 en sistemas productivos, complementado con monitoreo continuo de la deriva del modelo y reentrenamiento periódico.
6. **Escalabilidad:** La metodología desarrollada es escalable a otros contextos temporales y tipos de transacción, proporcionando un framework robusto para optimización de métricas específicas en detección de fraude.

El trabajo demuestra exitosamente que métricas personalizadas pueden mejorar significativamente el desempeño operacional de sistemas de detección de fraude, reduciendo la carga de trabajo de analistas y mejorando la experiencia del cliente al minimizar alertas falsas.

## Anexos

### Notebooks de la Implementación

Puede acceder a los notebooks del proyecto en el [Repositorio de GitHub](#).