Proyecto Final

Grupo 7

Propuesta de Proyecto de IA: Sistema de Detección de Fraude para Pagos Digitales.

Elaborado por:

Ana Lucia Espinoza Aguilera

Andrea Katherine Ordoñez Roldan

Andrea Paola Tapia Nicolalde

César Alejandro Cabrera Vázquez

Maestría en Inteligencia de Negocios y Ciencia de Datos

Materia: Inteligencia Artificial

Docente: PHD. Gladys Villegas Rugel

20 de diciembre del 2025

Contenido

Contenido *2*

1. Título y Resumen Ejecutivo. *5*

Título del Proyecto **5**

Resumen Ejecutivo **5**

2. Definición del Problema *6*

Contexto **6**

Problema Específico **6**

Justificación **6**

Stakeholders **7**

Primarios 7

Secundarios 7

3. Objetivos del Proyecto *7*

Objetivo General **7**

Objetivos Específicos (SMART): **7**

Alcance y Limitaciones **8**

Incluye. 8

No Incluye. 8

4. Solución Propuesta con IA *8*

Tipo de Problema **8**

Arquitectura Preliminar **9**

Figura 1. Arquitectura Simplificada de Detección de Fraude Financiero en Tiempo Real Basada en Inteligencia Artificial 9

Alternativas Consideradas **10**

5. Datos ***10***

Fuente de Datos **10**

Estructura y Variables del Dataset **11**

Identificación 11

Información del Usuario 11

Características de la Transacción 11

Variables Temporales 11

Contexto Geográfico y Tecnológico 12

Variables Derivadas 12

Variable Objetivo 12

Justificación del Diseño del Dataset **12**

Estrategia de Recolección y Gestión de Datos **13**

Naturaleza de los Datos 13

Variables Clasificadas 14

Desbalance Intencional de Clases 14

Pipeline de Procesamiento y Gestión de Datos 14

Pre procesamiento e Ingeniería de Características **16**

6. Metodología ***17***

Fases del Proyecto (Cronograma de 14 semanas) **17**

Métricas de Evaluación Clave **17**

Métrica de Negocio Principal 17

Métricas Técnicas 17

Herramientas y Tecnologías **17**

Lenguajes de Programación 17

ML/Data. 17

Plan de Validación 18

7. Viabilidad y Recursos 18

Recursos Técnicos 18

Hardware. 18

Software. 18

Recursos Humanos (Equipo Unipersonal con Rol Ampliado) 18

Presupuesto Estimado 18

Riesgos Identificados y Mitigación 19

Riesgo Técnico-Alto. 19

Plan de Mitigación 19

8. Referencias 19

1. Título y Resumen Ejecutivo.

Título del Proyecto

Implementación de un Sistema de IA para la Detección de Fraude para Pagos Digitales.

Resumen Ejecutivo

Las Empresas Financieras, como Bancos, cooperativas y en general, las fintech de pagos digitales enfrentan un desafío crítico de rentabilidad y experiencia de usuario: aunque solo el 0.1% de las transacciones son fraudulentas, estas representan aproximadamente el 60% de las pérdidas financieras. Paralelamente, los sistemas tradicionales generan una alta tasa de falsos positivos, bloqueando transacciones legítimas y deteriorando la satisfacción del cliente. Este proyecto propone una solución integral de Inteligencia Artificial (IA) diseñada para atacar ambos frentes. Lo innovador es que el sistema no solo detecta patrones anómalos en tiempo real, sino que aprende continuamente de cada interacción, reduciendo progresivamente los errores y adaptándose a nuevas modalidades de fraude sin necesidad de reprogramación manual.

La solución consiste en un sistema de detección en tiempo real que combina múltiples enfoques técnicos. Un ensamblado (ensemble) de modelos supervisado (XGBoost, LightGBM y Redes Neuronales) analizará más de 200 características de comportamiento transaccional para identificar patrones de fraude conocidos. La implementación de técnicas de *IA Explicable* (XAI), como *SHAP*, garantizará la transparencia de las decisiones, facilitando el cumplimiento regulatorio y la auditoría interna.

El valor entregado incluirá un pipeline de ingeniería de características reproducible, un modelo en producción para scoring en tiempo real, un dashboard de monitoreo de métricas y un simulador de transacciones para pruebas (Dataset Sintético). Se espera una reducción significativa tanto en pérdidas por fraude como en la tasa de falsos positivos, fortaleciendo la

posición competitiva de la Institución Financiera. El simulador permitirá recrear escenarios de ataque altamente sofisticados, ofreciendo a la organización la capacidad de anticiparse a nuevas tácticas de fraude y validar la resiliencia del sistema antes de que ocurran en la práctica.

2. Definición del Problema

Contexto

El proyecto se enmarca en el ámbito de una Institución Financiera de pagos digitales, un sector caracterizado por altos volúmenes transaccionales, expectativas de experiencia de usuario fluidas y una creciente presión regulatoria en materia de prevención de fraude y transparencia. La innovación tecnológica se convierte en un factor diferenciador clave, la adopción de soluciones basadas en Inteligencia Artificial no solo permite optimizar la detección de fraudes en tiempo real, sino también mejorar la confianza del cliente y cumplir con estándares de seguridad.

Problema Específico

La Institución Financiera sufre una asimetría crítica: una mínima fracción de transacciones fraudulentas (0.1%) es responsable de la mayor parte de sus pérdidas financieras (60%). Además, su sistema actual genera una alta tasa de falsas alarmas, lo que resulta en la interrupción innecesaria de pagos legítimos, insatisfacción del cliente, aumento en los costos de servicio al cliente y desgaste de la marca.

Justificación

Resolver este problema es imperativo para la sostenibilidad del negocio. La optimización del sistema de detección impacta directamente en el margen de utilidad (reduciendo pérdidas) y en el crecimiento (mejorando la retención y adquisición de clientes mediante una experiencia más fluida). La capacidad de explicar las decisiones de IA es, además, un requisito cada vez más común de supervisores financieros.

Stakeholders

Primarios.

Equipo de Gestión de Riesgo y Fraude, Departamento de Producto y Tecnología de la Institución financiera.

Secundarios.

Establecimiento y Comercios, Usuarios finales (experiencia mejorada), Equipo de Cumplimiento Normativo (Compliance), la Superintendencia de Bancos del Ecuador y la SEPS (Superintendencia de Economía Popular y Solidaria).

3. Objetivos del Proyecto

Objetivo General

Diseñar y validar un sistema supervisado de IA, explicable y adaptable, para la detección de fraudes en transacciones digitales, que maximice la captura de fraudes (recall) mientras minimiza las interferencias a clientes legítimos (falsos positivos). El sistema integrará técnicas de aprendizaje continuo, lo que permitirá anticiparse a nuevas modalidades de fraude y mantener un desempeño robusto incluso en escenarios cambiantes del mercado financiero.

Objetivos Específicos (SMART):

Desarrollar un pipeline automatizado de ingeniería de características que genere y seleccione un conjunto robusto de más de 200 variables a partir de datos transaccionales crudos, incluyendo agregados temporales y de comportamiento por usuario.

Implementar un marco de IA Explicable (XAI) basado en SHAP para el modelo final, produciendo reportes interpretables que justifiquen por lo menos el 95% de las alertas de fraude de alto riesgo generadas.

Diseñar la arquitectura de un sistema de *scoring* en tiempo real y un dashboard de monitoreo que visualice KPIs clave (fraudes detectados, falsos positivos, *drift* de

características) y proponer un esquema de online learning para actualización periódica del modelo.

Alcance y Limitaciones

Incluye.

Desarrollo de pipelines de datos y modelos, explicabilidad (SHAP), dashboard prototipo, simulador de transacciones y documentación técnica completa.

No Incluye.

Despliegue en infraestructura productiva de alta disponibilidad, integración directa con los sistemas *core* de la fintech, ni la gestión de datos sensibles reales en producción.

4. Solución Propuesta con IA

Tipo de Problema

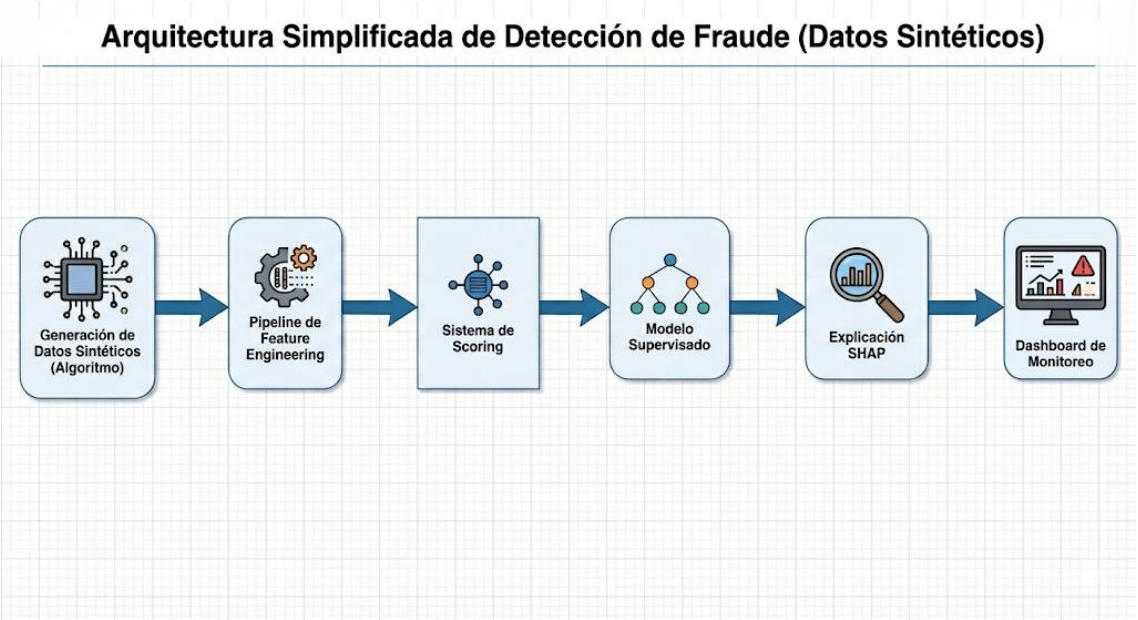
Clasificación binaria supervisada (1 = fraude; 0 = legítimo) con el Enfoque Técnico Modelado Supervisado *XGBoost* por su velocidad, eficiencia con datos tabulares y alto rendimiento en competencias. Se complementarán con una *Red Neuronal Artificial (ANN)* para capturar interacciones no lineales complejas. Un ensemble (promedio ponderado o stacking) combinará sus fortalezas.

Diseño e implementación de un pipeline automatizado para el reentrenamiento periódico de modelos, utilizando datos actualizados que aseguren su capacidad de adaptación. Este sistema no solo optimiza la actualización continua de los modelos, sino que también establece un ciclo robusto de mejora continua mediante la automatización del flujo de datos, la validación constante del rendimiento y la integración eficiente de nuevas muestras.

Uso de SHAP (Shapley Additive exPlanations) para asignar un valor de contribución a cada característica en la decisión de un modelo individual, garantizando transparencia y cumplimiento (Molnar, 2022).

Arquitectura Preliminar

Figura 1. Arquitectura Simplificada de Detección de Fraude Financiero en Tiempo Real Basada en Inteligencia Artificial



Nota: Diagrama del flujo de datos en tiempo real (Autores, 2025).

Se propone una arquitectura escalable para la detección de fraude financiero en tiempo real, diseñada para maximizar la precisión, interpretabilidad y cobertura de patrones anómalos. El sistema inicia con un flujo transaccional continuo, que es procesado por un pipeline de ingeniería de características para enriquecer los datos con variables derivadas y normalizadas. Posteriormente, las transacciones son evaluadas por: un modelo ensemble supervisado que genera un score de riesgo y explicaciones interpretables mediante valores SHAP (Lundberg & Lee, 2017), facilitando la trazabilidad de decisiones. Los resultados de este componente son consolidados en un sistema de alertas unificadas, que alimenta un dashboard de monitoreo para la visualización en tiempo real, facilitando la toma de decisiones operativas y estratégicas. Esta arquitectura supera enfoques tradicionales al integrar múltiples perspectivas analíticas, promoviendo un ciclo de mejora continua y una defensa robusta contra el fraude financiero (Bhattacharyya et al., 2011).

Alternativas Consideradas

Sistemas basados únicamente en reglas (descartados por inflexibilidad) o en un solo algoritmo (descartado por menor robustez). El enfoque híbrido propuesto ofrece redundancia y cobertura complementaria.

5. Datos

Fuente de Datos

La base de datos utilizada en el presente proyecto corresponde a un conjunto de datos sintéticos que simula transacciones financieras realizadas en una fintech de pagos digitales. Su generación responde a la necesidad de preservar la confidencialidad de la información, al tiempo que se reproducen patrones estadísticos y comportamientos realistas observados en sistemas financieros reales.

El dataset está conformado por 200.000 registros de transacciones, asociados a aproximadamente 5.000 usuarios únicos, y contempla variables de carácter demográfico, transaccional, temporal y contextual, permitiendo un análisis integral del comportamiento de los usuarios y la detección de eventos anómalos.

El dataset fue generado de forma desbalanceada, con una proporción de transacciones fraudulentas cercana al 0.1%, con el fin de replicar escenarios reales del sistema financiero. Este desbalance es característico de los problemas de detección de fraude y permite evaluar modelos bajo condiciones realistas, priorizando métricas como recall, precision y AUC, en lugar de accuracy, además de habilitar técnicas de detección de anomalías.

Adicionalmente, el dataset incorpora ruido controlado y variaciones en los patrones de comportamiento, lo que permite evaluar la robustez de los modelos frente a escenarios adversos y evita el sobreajuste. Esto asegura que las soluciones propuestas no solo funcionen en condiciones ideales, sino que también mantengan un desempeño estable ante cambios inesperados en el entorno financiero.

Estructura y Variables del Dataset

Identificación

- **transaction_id:** Identificador único de la transacción.
- **user_id:** Identificador único del usuario asociado a la transacción.

Información del Usuario

- **user_age:** Edad del usuario (en años).
- **account_age_days:** Antigüedad de la cuenta del usuario en días.
- **avg_amount:** Monto promedio histórico de consumo del usuario USD.

Estas variables permiten caracterizar el perfil del cliente y detectar desviaciones respecto a su comportamiento habitual.

Características de la Transacción

- **amount:** Monto de la transacción en USD.
- **channel:** Canal utilizado para la transacción (web, mobile_app, pos).
- **merchant_category:** Categoría del comercio (alimentación, tecnología, viajes, moda, servicios).

Estas variables son fundamentales para identificar patrones de consumos normales y atípicos.

Variables Temporales

- **timestamp:** Fecha y hora exacta de la transacción.
- **hour:** Hora del día en que se realiza la transacción.
- **day_of_week:** Día de la semana (0 = lunes, 6 = domingo).

Permiten analizar hábitos de consumo y detectar actividades inusuales en horarios de riesgo.

Contexto Geográfico y Tecnológico

- **country:** País desde donde se origina la transacción.
- **device:** Tipo de dispositivo utilizado (android, ios, desktop).
- **is_foreign:** Variable binaria que indica si la transacción se realiza fuera del país base.

Estas variables son clave para identificar transacciones de alto riesgo asociadas a ubicaciones o dispositivos no habituales.

Variables Derivadas

- **high_amount:** Indicador binario que señala si el monto supera un umbral elevado.

Este tipo de variables facilita la interpretación de los modelos y la aplicación de reglas de negocio.

Variable Objetivo

- **is_fraud:** Variable binaria que indica si la transacción es fraudulenta (1) o legítima (0).

El dataset presenta una distribución altamente desbalanceada, con aproximadamente 0.1% de transacciones fraudulentas, replicando escenarios reales de fraude financiero.

Justificación del Diseño del Dataset

El diseño desbalanceado del dataset responde a la naturaleza del fraude financiero, donde los eventos fraudulentos son poco frecuentes pero altamente costosos. Esta estructura permite evaluar modelos bajo condiciones realistas, priorizando métricas como recall,

precisión y AUC, y habilita el uso de técnicas de detección de anomalías y modelos explicables, alineadas con requerimientos regulatorios.

Estrategia de Recolección y Gestión de Datos

Naturaleza de los Datos

El presente proyecto utiliza un dataset sintético generado mediante un código para simular las condiciones operativas de una fintech de pagos digitales. Esta aproximación metodológica responde a dos imperativos fundamentales:

- **Preservación de la confidencialidad:** El uso de datos sintéticos elimina riesgos asociados a la exposición de información sensible de clientes y transacciones reales, cumpliendo con regulaciones de protección de datos como GDPR, PCI-DSS y normativas locales de privacidad financiera.
- **Reproducibilidad experimental:** Los datos sintéticos permiten la replicabilidad del estudio y facilitan la validación de resultados por parte de terceros sin comprometer información propietaria.

Mediante una generación Sintética Controlada, los datos fueron creados mediante técnicas de simulación estadística que replican las distribuciones, correlaciones y patrones comportamentales observados en sistemas financieros reales. El proceso de generación contempló el acceso y almacenamiento de la siguiente manera:

- Los datos sintéticos se encuentran almacenados en formato CSV, facilitando su portabilidad e interoperabilidad con diferentes herramientas de análisis.
- La estructura de almacenamiento sigue principios de organización jerárquica, con separación lógica entre datos crudos (raw), procesados (processed) y resultados (results).

Variables Clasificadas

La estrategia de recolección contempló la generación de variables en cuatro dimensiones principales:

- **Dimensión demográfica:** Edad del usuario y antigüedad de cuenta, permitiendo perfilamiento de clientes.
- **Dimensión transaccional:** Montos, canales, categorías de comercio y variables derivadas de comportamiento histórico.
- **Dimensión temporal:** Timestamps, hora del día y día de la semana, habilitando análisis de patrones temporales y detección de actividades en horarios anómalos.
- **Dimensión contextual:** País de origen, dispositivo utilizado e indicadores de transacciones internacionales, fundamentales para la detección de actividades de alto riesgo.

Desbalance Intencional de Clases

Como se mencionó en la descripción del dataset, se implementó un desbalance intencional con aproximadamente 0.1% de transacciones fraudulentas. Esta decisión metodológica responde a:

- Replicar condiciones reales de sistemas financieros donde el fraude es un evento raro pero de alto impacto.
- Priorizar métricas más informativas que accuracy (recall, precision, F1-score, AUC-ROC, AUC-PR) que son críticas en escenarios de detección de fraude.

Pipeline de Procesamiento y Gestión de Datos

Una vez generados los datos sintéticos, se implementa un pipeline estructurado que garantiza la calidad, consistencia y preparación de los datos para el modelado:

1. **Carga y Validación Inicial**

- Lectura del dataset desde archivos CSV con validación de integridad (verificación de columnas esperadas, tipos de datos, valores faltantes).
- Análisis exploratorio inicial para caracterizar distribuciones, identificar outliers y validar la coherencia de las variables sintéticas.

2. Limpieza y Pre procesamiento

- Tratamiento de valores faltantes mediante imputación o eliminación según criticidad de la variable.
- Detección y tratamiento de outliers mediante análisis estadístico (IQR, z-score) con decisiones basadas en conocimiento del dominio financiero.
- Validación de rangos de valores para variables numéricas y categorías para variables nominales.

3. Feature Engineering

El pipeline incorpora transformaciones de variables para maximizar el poder predictivo:

- **Variables derivadas:** Generación de ratios, agregaciones temporales, desviaciones respecto a comportamiento histórico del usuario.
- **Codificación categórica:** One-hot encoding para variables nominales (canal, categoría de comercio, país) y label encoding para variables ordinales cuando sea aplicable.
- **Escalamiento:** Normalización o estandarización de variables numéricas para algoritmos sensibles a escalas.
- **Transformaciones temporales:** Extracción de componentes cíclicos (hora del día, día de semana) mediante transformaciones seno-coseno para capturar periodicidad.

4. División y Estratificación

- Separación en conjuntos de entrenamiento, validación y prueba con estratificación para mantener la proporción de fraude en cada conjunto.

- Validación cruzada estratificada para evaluación robusta del desempeño del modelo.
- Consideración temporal en la división cuando sea relevante (evitar data leakage por información futura).

La estrategia de recolección fundamentada en datos sintéticos posibilita el desarrollo, entrenamiento y validación de modelos de detección de fraude dentro de un entorno controlado y reproducible, estableciendo las bases metodológicas para su futura implementación con datos reales. El pipeline de procesamiento diseñado se caracteriza por ser modular, escalable y altamente adaptable a las particularidades de los sistemas financieros en producción, lo que garantiza una transición eficiente desde la fase experimental hacia la operación, incorporando los ajustes necesarios en infraestructura, gobernanza de datos y mecanismos de supervisión continua.

Lineamientos Éticos

El empleo de datos sintéticos generados para simulación, reduce significativamente los riesgos asociados a la privacidad y el manejo de información sensible. Este enfoque garantiza mayor transparencia, confiabilidad y responsabilidad ética en el sistema de detección de fraude.

Pre procesamiento e Ingeniería de Características

1. Unión de tablas transaction e identity.
2. Imputación avanzada (por mediana, moda o modelos predictivos) y codificación de variables categóricas de alta cardinalidad.
3. Creación de más de 200 características derivadas, agregados temporales (ej., número de transacciones por usuario en las últimas 1h, 24h), ratios (monto/promedio histórico), diferencias (tiempo desde la última transacción), y características de riesgo basadas en combinaciones de dispositivo y ubicación.

6. Metodología

Fases del Proyecto (Cronograma de 14 semanas)

- **Fase 1 (Semanas 1-2):** Análisis exploratorio (EDA) profundo de los datasets a ser utilizados y diseño del esquema de ingeniería de características.
- **Fase 2 (Semanas 3-5):** Desarrollo del pipeline de feature engineering y división estratificada-temporal de datos (train/validation/test).
- **Fase 3 (Semanas 6-9):** Entrenamiento y evaluación comparativa de modelos (Ensemble vs. Anomaly Detection). Optimización de hiperparámetros con *Optuna*.
- **Fase 4 (Semanas 10-11):** Implementación de explicabilidad con SHAP y exploración preliminar de GNNs para análisis de conexiones.
- **Fase 5 (Semanas 12-13):** Desarrollo del prototipo de dashboard (con `Streamlit` del simulador de transacciones.
- **Fase 6 (Semana 14):** Documentación integral y preparación de la presentación final.

Métricas de Evaluación Clave

Métrica de Negocio Principal.

Precisión a un Recall fijo (ej., Precisión cuando se detecta el 95% de los fraudes). Esto mide directamente el impacto en falsos positivos.

Métricas Técnicas

Curva Precision-Recall (AUC-PR), F1-Score, Matriz de Confusión detallada.

Herramientas y Tecnologías

Lenguajes de Programación.

Python 3 y Jupyter Notebook.

ML/Data.

Pandas, NumPy, Scikit-learn, XGBoost, LightGBM, TensorFlow/PyTorch (para Autoencoders y GNNs), Imbalanced-learn, Feature-engine, Optuna. Explicabilidad com SHAP y

Lime. Visualización/Dashboard: Matplotlib, Seaborn, Plotly, Streamlit y Power BI. Infraestructura con Google Colab Pro y Jupyter Notebook.

Plan de Validación.

Se utilizará una división temporal estricta para simular condiciones reales. El modelo se entrenará con los primeros `X` días, se validará en el período siguiente y se probará de forma final en el período más reciente, garantizando que no haya *data leakage*.

7. Viabilidad y Recursos

Recursos Técnicos

Hardware.

Acceso a instancias cloud con GPU (NVIDIA T4 o superior) para el entrenamiento eficiente de Autoencoders y GNNs (aproximadamente 50-100 horas de uso).

Software.

Entorno de desarrollo Python y Jupyter Notebook, con las librerías antes mencionadas, y Github para control de versiones.

Recursos Humanos (Equipo Unipersonal con Rol Ampliado)

Rol: Científico de Datos / Ingeniero de ML.

Responsabilidades: Cubrir todo el ciclo: análisis, ingeniería de características, modelado (supervisado), explicabilidad, desarrollo del dashboard y documentación.

Presupuesto Estimado

Se establece un tentativo de $150 - $500. Principalmente para créditos de cloud computing (AWS SageMaker o Google Vertex AI) y posible uso de APIs premium de generación de datos sintéticos.

Riesgos Identificados y Mitigación

Riesgo Técnico-Alto.

Complejidad al integrar múltiples enfoques (Ensemble, Anomaly, GNN).

Plan de Mitigación.

- **Enfoque modular.** Priorizar el modelo ensamblado supervisado como núcleo, e incorporar los demás como módulos de investigación y validación.
- **Riesgo de Cumplimiento:** Que las explicaciones SHAP no sean lo suficientemente claras para auditores.
- **Mitigación:** Desarrollar plantillas de reporte estandarizadas que traduzcan los valores SHAP a razones de negocio (ej., "transacción marcada por alto riesgo debido a un dispositivo nunca antes visto combinado con un monto 5x superior al promedio del usuario").

8. Referencias

- American Psychological Association. (2020). Publication manual of the American Psychological Association (7th ed.).
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. https://doi.org/10.1145/2939672.2939785
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.
- Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable (2nd ed.). https://christophm.github.io/interpretable-ml-book.

- Zhou, Z., Liu, Y., & Li, G. (2021). Fraud detection using graph neural networks: A survey. IEEE Access, 9, 159657-159670. https://doi.org/10.1109/ACCESS.2021.3131840
- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613.
- Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv pre-print arXiv:1901.03407.