FUNDAMENTOS TEÓRICOS REFERENCIALES

Implementación de un Modelo de Machine Learning para la Detección de Anomalías y Fraude en Pagos Transaccionales

Introducción

El presente capítulo desarrolla los fundamentos teóricos referenciales que sustentan conceptual y metodológicamente la investigación orientada a la implementación de un modelo de Machine Learning para la detección de anomalías y fraude en pagos transaccionales. Este marco integra tres componentes esenciales: el marco referencial, que contextualiza los antecedentes y estado del arte; el marco teórico, que expone las teorías y enfoques científicos aplicables; y el marco conceptual, que delimita terminológicamente los constructos centrales del estudio.

La construcción de estos fundamentos responde a la necesidad de comprender la evolución del fraude en medios de pago digitales, los principios del aprendizaje automático supervisado y su aplicación en entornos transaccionales complejos. Asimismo, permite identificar las brechas teóricas y prácticas que justifican la presente investigación en el contexto de plataformas fintech multicanal como TechSport.

1. Marco Referencial: Antecedentes y Estado del Arte

1.1. Panorama global del fraude en pagos digitales

El fraude en transacciones electrónicas representa uno de los desafíos más significativos para la industria fintech contemporánea. Hernandez Aros et al. (2024) documentan que el crecimiento exponencial de las transacciones digitales ha generado un incremento proporcional en las actividades fraudulentas, lo que demanda sistemas de detección progresivamente más sofisticados. Esta tendencia se intensifica particularmente en plataformas que operan con múltiples pasarelas de pago y arquitecturas distribuidas, donde la complejidad del ecosistema transaccional dificulta la identificación oportuna de patrones anómalos.

En el ámbito internacional, Hafez et al. (2025) realizaron una revisión sistemática de técnicas de inteligencia artificial aplicadas a la detección de fraude con tarjetas de crédito, evidenciando que los modelos de aprendizaje automático han demostrado superiores niveles de precisión en comparación con los sistemas tradicionales basados en reglas estáticas. Los autores reportan que ciertos algoritmos supervisados alcanzan tasas de precisión del 94.3 % en la clasificación de transacciones fraudulentas, manteniendo simultáneamente

una tasa reducida de falsos positivos, lo cual resulta crítico para preservar la experiencia del usuario legítimo.

Por su parte, Bello y Olufemi (2024) analizan las técnicas de inteligencia artificial aplicadas a la prevención del fraude, destacando que los sistemas inteligentes ofrecen ventajas sustanciales en términos de adaptabilidad ante nuevos patrones de ataque y capacidad de procesamiento en tiempo real. No obstante, los autores también identifican desafíos relacionados con el desbalance de clases en los conjuntos de datos, la interpretabilidad de los modelos y los requerimientos computacionales para implementaciones a escala industrial.

1.2. Estudios aplicados a detección de fraude mediante Machine Learning

En el contexto latinoamericano, diversas investigaciones han explorado la viabilidad de implementar modelos de aprendizaje automático para la detección de fraude transaccional. Rayo Mondragón (2020) desarrollaron un prototipo basado en inteligencia artificial aplicado a un banco peruano, empleando algoritmos Random Forest para analizar transacciones de comercio electrónico. Los resultados demostraron una mejora significativa en la detección temprana de intentos fraudulentos en comparación con el sistema previo basado en reglas heurísticas.

De manera similar, Pérez González (2021) implementaron un modelo de Machine Learning para detectar transacciones fraudulentas en tarjetas de crédito en el contexto colombiano, enfocándose en técnicas de detección de anomalías. El estudio evidenció que la combinación de múltiples algoritmos supervisados mediante ensamblaje (ensemble learning) incrementa la robustez del sistema y reduce la variabilidad en el desempeño frente a diferentes tipologías de fraude.

A nivel internacional, AlEmad (2022) compararon el desempeño de diversos algoritmos de clasificación (K-Nearest Neighbors, Support Vector Machines y Regresión Logística) en la detección de fraude con tarjetas de crédito, concluyendo que no existe un algoritmo universalmente superior, sino que la selección depende de las características específicas del conjunto de datos y del contexto operativo. Este hallazgo refuerza la necesidad de realizar experimentación empírica contextualizada en cada escenario de aplicación.

Lucas (2019) abordaron la integración de conocimiento contextual en modelos de detección de fraude mediante el análisis de datasets secuenciales. Su investigación demostró que la incorporación de variables temporales, geográficas y comportamentales mejora sustancialmente la capacidad predictiva de los modelos, especialmente en la identificación de patrones fraudulentos emergentes que no han sido observados previamente en los datos de entrenamiento.

1.3. Enfoques avanzados: Redes neuronales y procesamiento de lenguaje natural

Investigaciones recientes han explorado técnicas más sofisticadas para la detección de fraude. Al-Khasawneh (2025) propusieron métodos híbridos basados en redes neuronales que combinan arquitecturas de aprendizaje profundo con técnicas convencionales, logrando mejoras en la detección de fraudes complejos que involucran múltiples transacciones coordinadas.

Rodríguez et al. (2023) introdujeron un enfoque innovador que incorpora procesamiento de lenguaje natural para analizar información textual asociada a las transacciones (descripciones de productos, mensajes de confirmación, etc.), demostrando que la información no estructurada contiene señales valiosas para la identificación de comportamientos fraudulentos que los modelos tradicionales basados exclusivamente en variables numéricas no logran capturar.

En el ámbito de las redes neuronales de grafos, Cheng et al. (2025) desarrollaron una revisión sistemática sobre la aplicación de Graph Neural Networks (GNN) en la detección de fraude financiero, destacando que estas arquitecturas son particularmente efectivas para identificar patrones de colusión y fraude organizado que involucran redes de múltiples actores interconectados.

1.4. Marcos normativos y estándares de seguridad

La implementación de sistemas de detección de fraude debe alinearse con marcos regulatorios y estándares internacionales de ciberseguridad. National Institute of Standards and Technology (2024) publicaron la versión 2.0 del Marco de Ciberseguridad del NIST (CSF 2.0), que incorpora la función "Govern" como eje transversal, enfatizando que la ciberseguridad constituye una fuente crítica de riesgo empresarial que debe gestionarse desde la alta dirección. Este marco proporciona lineamientos específicos para organizaciones de todos los tamaños, incluyendo plataformas fintech que procesan información financiera sensible.

En el contexto de América Latina, Organización de los Estados Americanos (OEA) y Banco Interamericano de Desarrollo (BID) (2020) documentaron las brechas existentes en capacidades de monitoreo, análisis de amenazas y respuesta operativa ante incidentes de ciberseguridad. El informe identifica que la fragmentación regulatoria, la diversidad de medios de pago y los niveles heterogéneos de madurez tecnológica crean un entorno propicio para la proliferación de fraudes que evolucionan más rápidamente que los controles implementados.

2. Marco Teórico: Fundamentos del Aprendizaje Automático y Detección de Fraude

2.1. Fundamentos del aprendizaje automático supervisado

El aprendizaje automático supervisado constituye un paradigma computacional en el cual los algoritmos aprenden a partir de ejemplos etiquetados para realizar predicciones sobre nuevas observaciones (Géron, 2022). En el contexto de la detección de fraude, esto implica entrenar modelos con conjuntos de datos históricos donde cada transacción ha sido clasificada como legítima o fraudulenta, permitiendo al sistema aprender los patrones distintivos de cada categoría.

Bishop (2006) establece que los modelos de clasificación supervisada buscan aproximar una función desconocida que mapea características de entrada hacia categorías de salida, minimizando el error de predicción mediante procesos iterativos de optimización. Este principio fundamental sustenta el diseño de sistemas inteligentes de detección de fraude, donde las características de entrada corresponden a atributos transaccionales (monto, ubicación, hora, etc.) y las categorías de salida representan la clasificación binaria (fraude/no fraude).

Goodfellow et al. (2016) amplían esta perspectiva al abordar arquitecturas de aprendizaje profundo, que permiten la extracción automática de representaciones jerárquicas de los datos. En el dominio de la detección de fraude, estas técnicas posibilitan identificar patrones complejos y relaciones no lineales que los métodos tradicionales no logran capturar eficientemente.

2.2. Métricas de evaluación en contextos desbalanceados

La evaluación de modelos de detección de fraude presenta desafíos particulares derivados del desbalance inherente en los datos: las transacciones fraudulentas representan típicamente una fracción minoritaria del total de transacciones. Murphy (2022) enfatiza que en estos contextos, métricas convencionales como la exactitud (accuracy) resultan engañosas, requiriéndose indicadores alternativos como precisión (precision), exhaustividad (recall) y F1-score.

La precisión mide la proporción de transacciones clasificadas como fraude que efectivamente lo son, minimizando falsos positivos que deterioran la experiencia del usuario. El recall cuantifica la proporción de fraudes reales que el sistema logra detectar, siendo crítico para minimizar pérdidas económicas. El F1-score proporciona una media armónica que equilibra ambos criterios, resultando particularmente útil en la comparación de modelos alternativos (Géron, 2022).

2.3. Teorías sobre detección de anomalías

La detección de fraude puede conceptualizarse como un problema de identificación de anomalías, entendidas como observaciones que se desvían significativamente del comportamiento esperado (Baesens et al., 2015). Esta perspectiva teórica distingue entre anomalías puntuales (transacciones individuales atípicas), anomalías contextuales (transacciones normales en circunstancias inusuales) y anomalías colectivas (secuencias de transacciones que conjuntamente resultan sospechosas).

Baesens et al. (2015) desarrollan un marco analítico integral para fraud analytics que combina técnicas descriptivas, predictivas y de análisis de redes sociales. Los autores argumentan que la detección efectiva de fraude requiere aproximaciones multidimensionales que incorporen no solamente características individuales de las transacciones, sino también patrones de comportamiento temporal, relaciones entre entidades y análisis de desviaciones respecto a perfiles históricos.

3. Marco Conceptual: Definiciones Operacionales

Para garantizar la precisión terminológica y facilitar la comprensión del estudio, se presentan las definiciones operacionales de los constructos centrales:

3.1. Machine Learning (Aprendizaje Automático)

Disciplina de la inteligencia artificial que desarrolla algoritmos capaces de aprender patrones a partir de datos sin ser explícitamente programados para cada tarea específica (Géron, 2022). En esta investigación, se refiere específicamente a técnicas supervisadas de clasificación aplicadas a la detección de fraude transaccional.

3.2. Fraude en pagos transaccionales

Acción deliberada destinada a obtener beneficios económicos ilícitos mediante el uso no autorizado de medios de pago, información financiera o identidades, comprometiendo la integridad de sistemas de comercio electrónico y plataformas fintech (Baesens et al., 2015).

3.3. Anomalía transaccional

Evento o patrón en los datos de transacciones que se desvía significativamente del comportamiento esperado o establecido, pudiendo indicar potenciales intentos de fraude, errores operativos o comportamientos legítimos atípicos que requieren verificación adicional.

3.4. Modelo supervisado

Algoritmo de aprendizaje automático entrenado mediante ejemplos etiquetados, donde cada instancia del conjunto de entrenamiento incluye tanto las características descriptivas como la categoría de clasificación correspondiente (Bishop, 2006).

3.5. Precisión (Precision)

Métrica que cuantifica la proporción de instancias clasificadas como positivas (fraude) que efectivamente pertenecen a esa categoría, calculada como el cociente entre verdaderos positivos y la suma de verdaderos positivos más falsos positivos.

3.6. Exhaustividad (Recall/Sensitivity)

Métrica que mide la proporción de instancias positivas reales que el modelo logra identificar correctamente, calculada como el cociente entre verdaderos positivos y la suma de verdaderos positivos más falsos negativos.

3.7. F1-Score

Media armónica entre precisión y exhaustividad, proporcionando una métrica balanceada del desempeño del modelo que resulta especialmente útil en contextos con clases desbalanceadas (Géron, 2022).

Conclusiones del capítulo

Los fundamentos teóricos referenciales presentados evidencian la solidez del corpus científico que sustenta la implementación de modelos de Machine Learning para la detección de fraude en pagos transaccionales. El estado del arte demuestra que los enfoques basados en aprendizaje automático supervisado superan consistentemente a los sistemas tradicionales basados en reglas estáticas, ofreciendo mayor capacidad de adaptación, precisión y escalabilidad.

No obstante, la literatura también revela desafíos metodológicos relevantes, particularmente relacionados con el desbalance de clases, la necesidad de interpretabilidad de los modelos y la integración efectiva en arquitecturas transaccionales complejas. Estos elementos justifican la necesidad de investigación contextualizada en escenarios específicos como el de la plataforma TechSport, donde la multiplicidad de pasarelas de pago y canales transaccionales introduce complejidades adicionales que requieren soluciones adaptadas.

La integración de los marcos referencial, teórico y conceptual proporciona una base epistemológica sólida para el diseño, desarrollo y evaluación del modelo propuesto, asegurando la coherencia científica y relevancia práctica de la investigación.

Referencias Bibliográficas

AlEmad, M. (2022). *Credit Card Fraud Detection Using Machine Learning* [Master's Project]. Rochester Institute of Technology.

Al-Khasawneh, M. (2025). Hybrid Neural Network Methods for the Detection of Credit Card Fraud. *Security and Privacy*. https://doi.org/10.1002/spy2.500

Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection*. Wiley.

Bello, O. A., & Olufemi, K. (2024). Artificial intelligence in fraud prevention: Exploring techniques and applications challenges and opportunities. *Computer Science & IT Research Journal*, *5*(6), 1505-1520. https://doi.org/10.51594/csitrj.v5i6.1252

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag New York.

Cheng, D., Zou, Y., Xiang, S., & Jiang, C. (2025). Graph neural networks for financial fraud detection: a review. *Frontiers of Computer Science*. https://doi.org/10.1007/s11704-024-40474-y

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (3.ª ed.). O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data*, *12*(6). https://doi.org/10.1186/s40537-024-01048-8

Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: a literature review. *Humanities and Social Sciences Communications*, *11*, 1130. https://doi.org/10.1057/s41599-024-03606-0

Lucas, Y. (2019). *Credit card fraud detection using machine learning with integration of contextual knowledge* [Tesis doctoral, INSA de Lyon].

Murphy, K. P. (2022). *Probabilistic Machine Learning: An Introduction*. MIT Press.

National Institute of Standards and Technology. (2024). *The NIST Cybersecurity Framework (CSF) 2.0* (NIST Cybersecurity White Paper N.º CSWP 29). National Institute of Standards y Technology. https://doi.org/10.6028/NIST.CSWP.29

Organización de los Estados Americanos (OEA) & Banco Interamericano de Desarrollo (BID). (2020). *Ciberseguridad: Riesgos, avances y el camino a seguir en América Latina y el Caribe* (Informe técnico). Organización de los Estados Americanos y Banco Interamericano de Desarrollo. Washington, D.C.

Pérez González, G. A. (2021). *Detección de transacciones fraudulentas en tarjetas de crédito mediante el uso de modelos de Machine Learning* [Trabajo de grado]. Universidad de los Andes.

Rayo Mondragón, C. A. (2020). *Prototipo de detección de fraudes con tarjetas de crédito basado en inteligencia artificial aplicado a un banco peruano* [Trabajo de suficiencia profesional]. Universidad de Lima.

Rodríguez, J. F., Papale, M., Carminati, M., & Zanero, S. (2023). Fraud detection with natural language processing. *Machine Learning*. https://doi.org/10.1007/s10994-023-06354-5