

Capítulo 3 (Tarea)

Fundamentación de los Instrumentos y Procedimiento Diagnóstico

Introducción

El presente capítulo forma parte de las actividades formativas orientadas a dominar la estructura metodológica de la tesis “*Implementación de un modelo de Machine Learning para la detección de anomalías y fraude en pagos transaccionales en TechSport (2024–2025)*”. Siguiendo el enfoque científico propuesto por Hernández Sampieri (2014), se fundamenta la selección de los instrumentos de investigación y se detalla el procedimiento para elaborar el diagnóstico empírico que sustenta la validación del modelo inteligente.

1. Enfoque metodológico y criterios de selección

El estudio adopta un enfoque cuantitativo con diseño empírico-analítico. Bajo esta lógica, los instrumentos deben cumplir tres criterios esenciales: (i) coherencia con las variables operacionalizadas (modelo de Machine Learning y detección de fraude), (ii) capacidad para generar datos confiables y comparables, y (iii) trazabilidad que permita replicar los resultados. Estos lineamientos derivan del método científico y de las recomendaciones para estudios aplicados en analítica de fraude (Baesens, 2015; Lucas, 2019). En consecuencia, los instrumentos se integran para cubrir el ciclo de observación, experimentación y verificación que exige el diagnóstico previo a la implementación del sistema inteligente.

2. Fundamentación académica de los instrumentos

2.1 Observación técnica directa y revisión de logs

La observación estructurada de registros operativos responde a la necesidad de comprender el comportamiento real del sistema actual de detección. Permite identificar patrones de falsos positivos, tiempos de respuesta y errores recurrentes que la literatura considera críticos para diseñar modelos predictivos robustos (Hernández Aros et al., 2024). Además, provee la línea base con la que se contrastará el desempeño del nuevo modelo, asegurando pertinencia externa al diagnóstico.

2.2 Dataset anonimizado de transacciones

El dataset constituye el instrumento central para operacionalizar las variables cuantitativas. Su anonimización garantiza el cumplimiento ético y normativo, mientras que la selección de atributos (monto, canal, pasarela, geolocalización, etiqueta fraude/no fraude) permite medir el fenómeno con indicadores verificables. La estructura tabular facilita la aplicación de algoritmos supervisados y respeta las recomendaciones de curado de datos para aprendizaje automático (Géron, 2022). La estandarización y balance de clases fortalecen la confiabilidad del instrumento.

2.3 Guía de evaluación del modelo

Este documento establece los criterios objetivos para comparar el sistema actual con el modelo propuesto. Incluye métricas reconocidas en la literatura de fraude financiero (precisión, *recall*, F1-score, tasa de falsos positivos) y parámetros operativos relevantes para TechSport, como el tiempo promedio de respuesta. Siguiendo a Sampieri, la guía funciona como instrumento de control que preserva la validez interna del experimento, al limitar el sesgo subjetivo durante la interpretación de resultados.

2.4 Scripts reproducibles de procesamiento y análisis

Los cuadernos en Python (Pandas, Scikit-learn, Matplotlib) documentan cada etapa de preparación, entrenamiento, validación cruzada y visualización. Su empleo se justifica porque garantizan trazabilidad, replicabilidad y auditoría externa del diagnóstico, condiciones exigidas en proyectos de ingeniería de datos críticos. Asimismo, permiten incorporar técnicas de evaluación robusta (validación k-fold, matrices de confusión, curvas ROC), contribuyendo a la comprobación empírica del supuesto de mejora frente al sistema basado en reglas.

3. Procedimiento para elaborar el diagnóstico

El procedimiento diagnóstico articula los instrumentos anteriores en un flujo secuencial que responde al método científico:

1. **Planeación y delimitación.** Revisar los objetivos específicos y las variables definidas en el perfil de tesis para formular hipótesis diagnósticas sobre las causas del bajo desempeño actual.
2. **Recolección estructurada.** Ejecutar la observación técnica directo sobre los logs y reportes históricos, y consolidar el dataset anonimizado conforme a los criterios éticos y de calidad de datos.

3. **Preparación y curado.** Aplicar los scripts para limpiar, transformar y balancear la muestra (por ejemplo, proporcionalidad entre casos legítimos y fraudulentos) y documentar cada ajuste en una bitácora técnica.
4. **Experimentación controlada.** Entrenar los modelos seleccionados, aplicar validación cruzada k-fold y registrar las métricas obtenidas mediante la guía de evaluación.
5. **Contraste y síntesis.** Comparar los resultados con la línea base del sistema actual, identificar brechas y formular recomendaciones operativas para la mejora de la gestión de riesgo transaccional.

4. Consideraciones de validez, confiabilidad y ética

La combinación de instrumentos garantiza triangulación de datos cuantitativos, fortaleciendo la validez interna y externa del diagnóstico. La estandarización del dataset y la automatización de los scripts aseguran confiabilidad, mientras que la anonimización y el uso controlado de logs atienden los principios éticos de confidencialidad. Estos elementos alinean el diagnóstico con las buenas prácticas de gobernanza de datos y con los estándares de seguridad transaccional (PCI DSS, NIST CSF 2.0).

Conclusiones

La fundamentación presentada evidencia que cada instrumento cumple una función específica dentro del enfoque empírico-analítico propuesto y que el procedimiento diagnóstico se sostiene metodológicamente. Este capítulo, elaborado como ejercicio académico, demuestra cómo estructurar una defensa técnica y científica de los instrumentos y actividades necesarias para validar el modelo de Machine Learning en TechSport, sirviendo como guía para la redacción definitiva del capítulo metodológico.