

# Fundamentación de Instrumentos y Procedimiento Diagnóstico

## Implementación de un Modelo de Machine Learning para la Detección de Anomalías y Fraude en Pagos Transaccionales en TechSport (2024–2025)

### 1. Propósito y marco metodológico

El diagnóstico se orienta por el método científico planteado por Hernández Sampieri (2014), asegurando la coherencia entre problema, objetivos y medición de variables. Tal como se estableció en la *Tarea 2 – Selección de métodos del nivel teórico*, se articulan los razonamientos analítico-sintético, inductivo-deductivo, de modelación e hipotético-deductivo. El corpus conceptual compilado en la *Tarea 3 – Fundamentos teóricos referenciales* proporciona el soporte epistemológico que permite operacionalizar cada instrumento dentro de las fases de observación, experimentación, contrastación y retroalimentación descritas por Sampieri.

### 2. Fundamentación de los instrumentos seleccionados

- **Observación técnica directa y revisión de logs.** Siguiendo los lineamientos de observación estructurada de Sampieri, se analizan los logs transaccionales y reportes de fraude documentados en `capitulos/03_metodologia.tex:88-93`. Este instrumento caracteriza el desempeño del sistema actual, identifica patrones de error y establece la línea base que el nuevo modelo debe superar, aportando validez externa al vincular el marco teórico con la realidad operativa de TechSport.
- **Dataset anonimizado de transacciones.** Constituye el instrumento central para las variables independiente y dependiente, conforme a las definiciones y operacionalización descritas en `capitulos/03_metodologia.tex:31-86`. Incluye atributos (monto, país, canal, pasarela, *timestamp*) y etiquetas de clasificación, cumpliendo con el criterio de Sampieri de emplear instrumentos estandarizados que aseguren confiabilidad dentro de la delimitación temporal y espacial del estudio.
- **Guía de evaluación del modelo.** Derivada de la matriz de indicadores (precisión, *recall*, F1-score, tasa de falsos positivos, tiempo de respuesta), formaliza los criterios de comparación entre el modelo inteligente y el sistema basado en reglas. Opera como instrumento de control de calidad que incrementa la validez interna del experimento comparativo e integra los aportes de la *Tarea 3* sobre métricas y estándares de ciberseguridad (NIST CSF 2.0, PCI DSS).

- **Scripts de procesamiento y análisis.** Los cuadernos en Python (Pandas, Scikit-learn, Matplotlib) garantizan trazabilidad y replicabilidad del análisis. Permiten ejecutar validación cruzada k-fold, generar matrices de confusión y curvas ROC, conforme al principio sampieriano de documentar procedimientos para sostener el método hipotético-deductivo.

### 3. Procedimiento diagnóstico alineado al método científico

1. **Planeación conceptual.** Integrar antecedentes, variables y objetivos recogidos en el perfil (`capitulos/00_perfil.tex`) y en la *Tarea 3* para formular hipótesis diagnósticas sobre las causas del bajo desempeño actual.
2. **Levantamiento de información.** Ejecutar la observación técnica directa mediante extracción de logs y solicitar al equipo de TechSport el dataset anonimizado siguiendo los criterios éticos definidos en la delimitación temporal y espacial (`capitulos/03_metodologia.tex:14-28`).
3. **Preparación y estructuración de datos.** Aplicar los scripts de limpieza, balanceo y partición 70/15/15 establecidos en la metodología, registrando cada transformación en una bitácora técnica para preservar la confiabilidad del instrumento.
4. **Experimentación y contraste.** Entrenar los modelos propuestos, ejecutar validación cruzada k-fold ( $k=5$ ) y comparar resultados con el sistema actual mediante la guía de evaluación. Documentar métricas en tablas y visualizaciones reproducibles.
5. **Síntesis y retroalimentación.** Elaborar el informe diagnóstico que articule hallazgos cuantitativos con el análisis teórico de las Tareas 2 y 3, identificando brechas, riesgos y oportunidades de mejora. Incorporar recomendaciones alineadas a los estándares NIST CSF y PCI DSS.

### 4. Garantías de validez y confiabilidad

- **Validez interna.** La guía de evaluación y la comparación controlada con el sistema basado en reglas permiten aislar el efecto del modelo de Machine Learning, cumpliendo la lógica experimental de Sampieri.
- **Validez externa.** El uso de datos reales de TechSport y la observación directa de logs aseguran que los hallazgos sean transferibles al entorno operativo, reforzando la pertinencia del diagnóstico frente a los objetivos institucionales.

- **Confiabledad.** La estandarización del dataset, el versionamiento de scripts y la bitácora de transformaciones facilitan la repetición del estudio y la auditoría de resultados, en concordancia con la recomendación de Sampieri sobre instrumentos reproducibles.

## 5. Articulación con entregables previos y próximos pasos

- La matriz de métodos y la estructura conceptual trabajadas en la *Tarea 2* respaldan la selección de instrumentos al demostrar su alineación con los objetivos específicos y la hipótesis de trabajo.
- El compendio bibliográfico y la matriz de referencias de la *Tarea 3* justifican la elección de métricas, técnicas de aprendizaje y estándares de ciberseguridad considerados en la guía de evaluación.
- El material de la *Tarea 4* (marco contextual) refuerza la necesidad de situar el diagnóstico en las características institucionales y tecnológicas de TechSport, evitando sesgos por información periférica.

**Próximos pasos inmediatos:** consolidar el dataset anonimizado con metadatos completos, formalizar la guía de evaluación en formato *checklist* y preparar el primer informe de observación basado en los logs históricos para presentar al tutor.