

Proyecto 2 — Detección de Fraude usando Técnicas de Aprendizaje Supervisado

INFO1185: Inteligencia Artificial

Prof. Dr. Ricardo Soto Catalán

10 de octubre de 2025

1. Descripción general

La detección automática de transacciones fraudulentas es un problema real y crítico en el ámbito financiero. Las organizaciones deben identificar actividades sospechosas entre millones de transacciones legítimas, en contextos donde los casos de fraude son extremadamente poco frecuentes.

En este proyecto, la tarea es **entrenar y evaluar diferentes modelos de clasificación supervisada** para detectar transacciones fraudulentas en un conjunto de datos reales.

Este problema presenta desafíos importantes:

- El dataset está **altamente desbalanceado** (menos del 1 % de las transacciones son fraudulentas).
- Las variables han sido transformadas mediante PCA, por lo que no existe un significado semántico directo. No tiene sentido aplicar PCA u otro algoritmo de reducción de dimensión a este conjunto de datos.
- Las métricas estándar como la *accuracy* no son adecuadas en este contexto.

Para abordar este problema, los equipos deberán diseñar un **pipeline completo de modelado**, desde el preprocesamiento hasta la evaluación, justificando las decisiones técnicas en cada etapa.

2. Objetivos

- 1) Comprender y explorar un conjunto de datos reales y desbalanceados.
- 2) Implementar estrategias adecuadas de preprocesamiento y preparación de datos.
- 3) Seleccionar un subconjunto de características predictivas usando una técnica de **selección de características basada en modelos o tests estadísticos** (no reducción tipo PCA).
- 4) Entrenar, ajustar y comparar el desempeño de cuatro clasificadores supervisados:

- k-Nearest Neighbors (k-NN)
 - Árbol de Decisión
 - Support Vector Machine (SVM)
 - Random Forest
- 5) Realizar búsqueda y selección de hiperparámetros para cada modelo, documentando el procedimiento y la configuración óptima obtenida.
 - 6) Evaluar el desempeño de los modelos usando métricas adecuadas para problemas desbalanceados y discutir críticamente los resultados.

3. Conjunto de datos

- **Nombre original:** Credit Card Fraud Detection Dataset
- **Fuente:** Kaggle
- **Número de registros:** 284 807
- **Número de fraudes:** 492 (0,172 %)
- **Atributos:**
 - Time, Amount
 - V1 – V28: componentes principales generadas por PCA
 - Class: variable objetivo (0 = legítima, 1 = fraude)

4. Instrucciones técnicas

- 1) **Preprocesamiento:**
 - Escalado o normalización de variables numéricas.
 - División en conjuntos de entrenamiento (70 %) y prueba (30 %).
- 2) **Selección de características:**
 - Aplicar al menos una técnica de selección de características, por ejemplo, Sequential Forward Selection.
 - Reporte el vector de características seleccionado.
- 3) **Entrenamiento y ajuste de modelos:**
 - Entrenar los cuatro clasificadores especificados, considerando minimizar la tasa de falsos negativos (evitar fraudes clasificados como caso normal).
 - Realizar selección de hiperparámetros mediante validación cruzada (por ejemplo, Grid Search o Random Search).
 - Documentar el espacio de búsqueda y la configuración óptima obtenida para cada modelo.

4) Evaluación:

- Reportar las siguientes métricas:
 - Sensibilidad, Especificidad, Precisión, F1-score, Exactitud.
- Discutir por qué métricas como la *exactitud* no son representativas en este contexto.
- Analizar falsos positivos y falsos negativos.

5) Comparación crítica:

- Comparar resultados entre modelos y justificar cuál es más adecuado para este problema.
- Discutir ventajas y limitaciones de cada técnica en el contexto del desbalance.

5. Entregables

Entregable	Descripción	Fecha
Informe técnico (PDF)	6–10 páginas. Debe incluir descripción del problema, selección de características, tuning, resultados, análisis crítico y conclusiones.	22 de octubre a las 11 hrs. en Blackboard
Código	Código completo y bien documentado, que permita reproducir todos los resultados.	22 de octubre a las 11 hrs. en Blackboard
Presentación oral	10–12 minutos por equipo + preguntas. Enfocada en decisiones técnicas y hallazgos.	22 y 23 de octubre

6. Rúbricas de evaluación

La evaluación se compone de tres partes: Informe, código y entrevista. De modo que la nota del proyecto (NP2) se calcula como

$$NP2 = FP \cdot \text{Nota informe} + \text{codigo}. \quad (1)$$

En las Tablas 1, 2 y 3 se presentan las rúbricas del informe (85 pts.), código (15 pts.) y entrevista, respectivamente.

La entrevista tiene por propósito validar el conocimiento del estudiante sobre el desarrollo del proyecto y de aspectos teóricos presentes en él. El profesor realizará 5 preguntas. Cada pregunta tiene un puntaje máximo de 10 puntos, tal que el puntaje total de la entrevista son 50 puntos, tal que el factor de ponderación FP se calcula como $FP = \frac{PE}{50}$, donde PE es el puntaje obtenido en la entrevista.

***Observación: El proyecto se debe realizar en parejas.**

Tabla 1: Rúbrica del informe.

Sección	Criterio	Excelente (1)	Bien (0.66)	Insuficiente (0.33)	Deficiente (0)
Resumen (10 pts)	Claridad y síntesis (7 pts)	Resume claramente qué se hizo, cómo y principales resultados en un párrafo preciso.	Resume parcialmente, faltan algunos elementos clave.	Resumen confuso o incompleto.	No presenta resumen o es irrelevante.
	Pertinencia (3 pts)	Incluye solo información esencial, sin detalles innecesarios.	Incluye algunos detalles no esenciales.	Contiene información poco relacionada.	No cumple el objetivo del resumen.
Planteamiento (9 pts)	Contextualización	Explica claramente el problema, relevancia, desafíos y métricas.	Presenta el problema de forma general o poco profunda.	Menciona el problema superficialmente.	No presenta planteamiento claro o es incorrecto.
	Objetivos y justificación	Objetivos claros y bien justificados.	Objetivos correctos pero justificación débil.	Objetivos vagos o poco relacionados.	No presenta objetivos claros.
Metodología (27 pts)	Pre-procesamiento (9 pts)	Describe claramente normalización, división y estrategias para desbalance con justificación.	Describe correctamente pero justifica parcialmente.	Menciona etapas superficialmente.	No describe correctamente las etapas.
	Selección de características (9 pts)	Explica técnica aplicada, justifica elección, presenta vector final.	Técnica bien descrita pero con detalles incompletos.	Técnica superficial, sin justificación adecuada.	No se describe la técnica o es incorrecta.
	Entrenamiento y ajuste (9 pts)	Explica validación, espacio de hiperparámetros, umbral y criterio de selección claramente.	Procedimiento presentado parcialmente.	Entrenamiento incompleto o poco claro.	No detalla el procedimiento o es incorrecto.
Resultados (27 pts)	Reporte de métricas (9 pts)	Presenta métricas completas en tablas/figuras claras y etiquetadas.	Métricas correctas pero formato mejorable.	Métricas incompletas o poco claras.	Métricas incorrectas o ausentes.
	Análisis de desempeño (9 pts)	Analiza críticamente resultados, interpreta métricas y limita exactitud.	Análisis parcial con algunos elementos críticos.	Interpretación superficial.	No hay análisis o es erróneo.
	Comparación entre modelos (9 pts)	Compara modelos de forma fundamentada, considerando métricas, ventajas y limitaciones.	Comparación parcial o poco profunda.	Comparación débil, solo numérica.	No hay comparación real.
Conclusiones (12 pts)	Síntesis de hallazgos (6 pts)	Resume claramente hallazgos y justifica elección del mejor modelo.	Resume hallazgos con argumentación débil.	Hallazgos poco claros o incompletos.	No hay conclusiones o son irrelevantes.
	Reflexión crítica (6 pts)	Presenta ventajas, limitaciones y mejoras futuras fundamentadas.	Menciona algunos puntos críticos parcialmente.	Reflexión superficial.	No hay reflexión crítica.

Tabla 2: Rúbrica del código.

Nivel	Descripción	Puntaje
Excelente	Código modular, limpio y bien documentado. Uso de semillas y parámetros claros.	12–15
Satisfactorio	Código ejecutable y organizado, con documentación parcial.	8–11
Básico	Código desordenado o con dificultades de ejecución. Poca documentación.	3–7
Insuficiente	Código incompleto, ilegible o no ejecutable.	0–2

Tabla 3: Rúbrica de entrevista.

Nivel	Descripción	Puntaje
Excelente	Responde correctamente, con seguridad y profundidad; demuestra dominio sólido del tema, integra fundamentos teóricos y los conecta con el desarrollo del proyecto. La respuesta tarda menos de 1 minuto.	9 – 10
Muy bueno	Responde de forma adecuada; cubre la mayor parte del contenido esperado, con explicaciones correctas y fundamentadas, presenta detalle menor. La respuesta tarda menos de 1 minuto.	7 – 8
Satisfactorio	Respuesta parcial; identifica conceptos clave pero con explicaciones incompletas o poco claras. Muestra comprensión básica, pero sin profundización ni conexión explícita con el proyecto. Tarda entre 1 y 1.5 minutos en responder.	5 – 6
Insuficiente	Respuesta vaga, con errores conceptuales o desconexión con el proyecto. Evidencia conocimiento limitado del tema. Tarda entre 1 y 1.5 minutos en responder.	3 – 4
Deficiente	No responde, o la respuesta es incorrecta, irrelevante o demuestra desconocimiento del tema. No logra responder en un tiempo máximo de 1.5 minutos.	0 – 2