



CURSO: CMP 5002 - DATA MINING
COLEGIO: POLITÉCNICO
Semestre: 1er Semestre 2024/2025

Tarea 7: Ejercicio sobre detección de anomalías con técnicas no supervisadas.

Objetivo:

- Implementar un módulo de un sistema de detección de anomalías para la detección de fraude en transacciones bancarias.

Modalidad: Trabajo en equipo. Por tanto, cada integrante debe preparar su parte (asignada por el líder) y dominar las técnicas evaluadas completamente.

Problema:

1. Dado el conjunto de datos **Credit_card.zip** (se encuentra en la misma carpeta de orientación del proyecto en el D2L). Se desea:
 - Implementar tres algoritmos de reducción de dimensionalidad (isomap, mds, AE) para reducir el espacio original a uno latente de menor dimensión. Se debe realizar una optimización de parámetros tanto de los algoritmos como del espacio latente para minimizar el error de reconstrucción. La variación de parámetros debe producir al menos 8 escenarios de comparación. **(3 puntos)**
 - Mostrar los gráficos de los espacios reducidos para cada combinación de parámetros (Un subplot tipo grid) e imprimir el modelo y parámetros del espacio seleccionado de acuerdo a la función loss empleada. **(2 puntos)**
 - Para el espacio seleccionado, mostrar la reconstrucción visual y el error de reconstrucción. **(2 puntos)**
 - Aplicar un modelo de clustering (libre elección) y un modelo de algoritmo de detección de anomalías (isolation forest, lof, one-svm) para corroborar la calidad del espacio latente seleccionado. Mostrar los resultados de las métricas de clasificación escogidas. **(2 puntos)**.
 - Aplicar el modelo completo (reductor+clasificador) al conjunto test externo e imprimir las métricas de validación escogidas. **(1 punto)**.
 - Cargar al D2L los códigos implementados (archivo compactado) dentro del plazo de entrega.