

# Análisis de Datos para Ciberseguridad: Trabajo práctico 2

Alonso Araya Calvo

Pedro Soto

Sofia Oviedo

Instituto Tecnológico de Costa Rica,  
Escuela de Ingeniería en Computación,  
Programa de Maestría en Ciberseguridad

14 de setiembre de 2025

En este trabajo practico se realiza un análisis y implementación de un árbol de decisión para el set de datos KDD99.

Para efectos del proyecto se tomo un dataset reducido de KDD99, el cual contiene 10 % de los datos originales, en el cual se va tomar en cuenta solamente las clases 'normal' que indica tráfico normal y 'backdoor' que indica tráfico malicioso. Además de eso se creo una versión filtrada del dataset, sin duplicados, filas sin valores y datos categóricos convertidos a numéricos mediante one-hot-encoding como lo son 'protocol type', 'service' y 'flag'.

Se realizó un análisis de los momentos estadísticos de las características del dataset, así como el calculo de la distancia Jensen-Shannon, histogramas comparativos y la creación, evaluación y análisis de un árbol de decisión.

Se utilizo librerías como numpy, pandas, matplotlib, scipy, pytorch y sklearn para los métodos creados para el trabajo.

## **1. Análisis descriptivo de las características en el conjunto**