



Curso de especialización inteligencia artificial y big data

## Ejercicios

Profesor: Sebastián Rubio Valero

Noviembre 2025



Sistema aprendizaje automático

Ejercicio 1: El Conjunto de Datos de Evaluación de Riesgos Globales de Terremotos y Tsunamis (The Global Earthquake-Tsunami Risk Assessment Dataset) es un conjunto de datos integral, listo para el aprendizaje automático, que contiene **\*\*características sísmicas e indicadores de potencial de tsunami de 782 terremotos significativos registrados a nivel mundial desde 2001 hasta 2022.** Este conjunto de datos está diseñado específicamente para aplicaciones de predicción de riesgo de tsunami, análisis de terremotos y evaluación de peligros sísmico.

Consta de las siguientes características Este conjunto de datos no tiene valores perdi-

Table 1: Placeholder Caption

Feature	Type	Description	Range/Values
magnitude	Float	Magnitud del terremoto(Richter scale)	6.5 - 9.1
cdi	Integer	Intensidad decimal comunitario (felt intensity)	0 - 9
mmi	Integer	Medida instrumental (instrumental)	1 - 9
sig	Integer	Puntuación de la importancia	650 - 2910
nst	Integer	Número de estaciones de monitorización	0 - 934
dmin	Float	Distancia más cercana a un estación sísmica (degrees)	0.0 - 17.7
gap	Float	Ángulo horizontal	0.0-239
depth	Float	Profundidad en Km	2.7-670.8
latitud	Float	Latitud	-61.85-71.63
year	Integer	Año del suceso	2001-2002
month	Integer	Mes del suceso	1-12
tsunami	Binario	variable objetivo	0-1

dos y consta de 782 registros. La variable tsunami es la variable objetivo. Se quiere crear un aplicación de predicción de tsunami, que constará de los siguientes modelos de aprendizaje automático:

- Predicción de riesgo de tsunami: Modelo de clasificación
- Aviso de alerta temprana de tsunami:
- Estimación de la magnitud: Modelo de regresión donde la variable objetivo es la magnitud del terremoto , teniendo en cuenta las siguientes variables mmi,nst,dmin,gap,depth,latitud,month
- Detección de zonas de riesgo

Entrena los modelos con los siguientes métodos y eligen el que mejor resultado tengas

- Hold out validation
- K-fold Cross Validation
- K-fold stratified (sólo en el que se pueda)