Ciencia de Datos
Certamen Práctico 1

Profesor: Diego Ramírez
Ayudante: Abner Astete

14 de mayo de 2023

El conjunto de datos de mama es un conjunto de datos completo que contiene casi todos los datos del estudio PLCO disponibles para los análisis de incidencia y mortalidad por cáncer de mama. Para muchas mujeres el ensayo documenta múltiples cánceres de mama, sin embargo, este archivo solo tiene datos sobre el primer cáncer de mama diagnosticado en el ensayo.

Las características se calculan a partir de una imagen digitalizada de un aspirado con aguja fina (FNA) de una masa mamaria. Describen las características de los núcleos celulares presentes en la imagen.

Se entrega el dataset en el archivo `breast-cancer.csv`.

1. Realice con **Python** un preprocesamiento de los datos y descubra los valores faltantes del conjunto de datos. Reemplácelos utilizando una técnica de minería de datos. Debe mostrar el antes y el después del dataset. (25 ptos)

2. Descubra con **Python**, desde el dataset original, las muestras que son candidatos a outliers. Además utilizando alguno de los enfoques vistos en clases, describa cuales de ellos son los 5 más probables a ser outliers. (25 ptos)

3. Descubra con **Python**, desde el dataset original, los atributos que mejor describen la clase (si posee o no eventos recurrentes). (25 ptos)

4. Descubra en **Python** las reglas de asociación para este dataset con una confianza mayor o igual a 0,9 . (25 ptos)

CONDICIONES DE ENTREGA:

1. Adjunte el Notebook con el código ordenado y comentado en AULA.