Parcial 1 de analítica de datos parte 1.

Usando la base de datos de cancer de mama disponible en: https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic) realizar un modelo de regresión logística que permita categorizar correctamente los registros entre pacientes sanos y no sanos. Para ello se debe seguir las siguientes indicaciones:

1. Se debe visualizar relaciones entre variables usando los métodos vistos en clase.
2. Encontrar valores atípicos o faltantes entre las columnas de los registros. En caso de que existan, cuantificarlos (contarlos) para conocer el número total de ellos.
3. Encontrar los índices de los atípicos (si existen) y cuantificar cuántos registros contienen estos tipos de datos. NOTA: es diferente encontrar el número total vs el número de registros. Por ejemplo: Únicamente la fila 7 tiene un atípico en la columna 7,10 y 15. Del anterior ejemplo podemos afirmar que un registro contiene 3 datos atípicos y para solucionar dicho problema el proceso que debemos seguir es eliminar dicho registro.
4. En caso que la cantidad de registros que se eliminarían supera el 40 % del total del conjunto de datos, se procederá a eliminar únicamente el 20 % de forma aleatoria los registros atípicos.
5. El modelo debe ser validado por una validación cruzada con K=10. Dicho proceso debe ser programado de manera manual, es decir NO SE PERMITE LAS LIBRERÍAS EXTERNAS. NOTA: DEBEN GARANTIZAR LA HOMOGENEIDAD DE LOS DATOS AL MOMENTO DE ENTRENAR EL MODELO.
6. Se debe imprimir la matriz de confusión por cada validación del numeral anterior.
7. El desempeño de este modelo debe ser encontrado al calcular las métricas de sensibilidad, especificidad y precisión.
8. Deben desarrollar al menos 4 modelos diferentes en busca del mejor desempeño posible(Se debe evidenciar la numerosas pruebas realizadas) y decidir en base de una curva ROC cual de ellos es el mejor.

   NOTA 1: CADA MODELO NO DEBE SUPERAR MÁS DE 7 ENTRADAS.
   NOTA 2: Se debe entregar un notebook o paper que se evidencia todo lo realizado aquí. El código debe estar debidamente comentado.
   NOTA 3: El parcial se entregará el LUNES 20 de marzo del 2023 en parejas.