**Ejercicios de evaluación Módulo III**

**Parte I (clasificación)**

**Ejercicio 1.** Realizar una modificación, al algoritmo visto en clase en el libro de Jupyter “ArbolesBinariosProcedimiento\_HojaTrabajo.ipynb, para que se puede realizar un procedimiento de clasificación mediante la función de entropía.

**Ejercicio 2.** Con base en el ajuste realizado en el Ejercicio 1, aplicar este algoritmo de clasificación a la base “EvaluacionModulo3\_Clasificacion.csv”. Realizar este procedimiento en un libro de Jupyter mostrando su matriz de confusión y su curva ROC.

**Ejercicio 3.** Implementar en la base “EvaluacionModulo3\_Clasificacion.csv” los algoritmos de clasificación mediante las librerías de Python, y estudiados en este módulo, descritos a continuación:

1. Regresión logística
2. Maquinas de vectores de soporte.
3. Árboles binarios.
4. Árboles binarios con *bagging* simplificado.
5. Bosque aleatorio.
6. Red neuronal.

Los resultados deben ser reportados en un libro de Jupyter, indicando con qué algoritmo se obtiene un mejor desempeño. Como medida de selección del mejor algoritmo, deben presentarse la **matriz de confusión**, la gráfica de las **curvas ROC** y el **área bajo la curva** de dichas curvas ROC. Se sugiere realizar una partición de la base de datos para considerar un 80% como base de entrenamiento y un 20% como base de verificación.

**Parte II (regresión)**

**Ejercicio 1.** Usando el dataset “MedicalPremium.csv”, aplicar los algoritmos de regresión estudiados en el módulo usando las librerías de Python de SK Learn.

1. Regresión simple
2. Regresión Ridge
3. Regresión Lasso

La tarea consiste en **predecir el costo del seguro médico anual** para nuevos pacientes de una compañía de seguros con base en la información de los pacientes con los que cuenta.

El resultado a entregar debe ser reportado mediante el uso de la métrica de **“Precisión”** para cada algoritmo e **indicar qué algoritmo presentó mejor resultado**. **Indicar los parámetros o hiperparámetros utilizados** para cada algoritmo con los que obtuvieron el mejor resultado.

Para obtener un modelo más adecuado, aplicar las técnicas vistas en las sesiones que consideren adecuadas. (Cross-validation, selección de variable no correlacionadas, eliminación de duplicados).

Favor de entregar la Parte I y la parte II de la tarea en libros de Jupyter diferentes.