
TEXT MINING & IMAGE RECOGNITION
LABORATORIO # 2

Instrucciones: A continuación verá una lista de ejercicios que debe completar para poder entregar el laboratorio #2. Al finalizar, todos sus archivos deben estar contenidos en un archivo lab2-sucarnet.zip. Este archivo debe entregarlo en el link del GES. Por favor cree una carpeta para cada ejercicio que usted realice.

Problema #1:

Utilizando el código trabajado durante la sesión de laboratorio construya una función que le permita realizar la clasificación de un texto aleatorio proporcionado como argumento a dicha función. Su función debe utilizar los procedimientos y modelos de Machine Learning que se generaron en clase, si usted desea puede reescribir los procedimientos para volverlos más eficientes, el objetivo es que se pueda realizar la predicción de *ham* o *spam* para cualquier string dado.

Problema #2:

Para el dataset llamado **imdb_rev.txt** desarrolle un clasificador de reviews positivos o negativos, la variable que debe predecir está codificada como 0 y como 1, donde 0 es un review malo y 1 es un review bueno. Tome en cuenta que este archivo no es un csv y usted deberá procesarlo para poder formar una estructura adecuada para la clasificación, además note que está en inglés, por tal motivo debe utilizar el modelo de lenguaje adecuado. Recuerde usar todos los recursos que conoce sobre Machine Learning para determinar si su programa clasifica adecuadamente, puede usar cualquier clasificador que desee incluso varios si lo considera necesario.