



UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS
CARRERA DE INGENIERÍA EN SISTEMAS
Asignatura Inteligencia Artificial
Ejercicio N°2 Detector de SPAM

Descripción

Se desea crear un programa que sea capaz de entrenar un modelo de regresión logística que pueda clasificar si un correo electrónico es SPAM (correo no deseado) o no de acuerdo su contenido.

Requisitos de Presentación

- El ejercicio constará de dos entregables: el primer entregable consistirá en un directorio que contenga 40 correos electrónicos utilizando el formato que se indica en el fichero proporcionado: spam_detector.py. De éstos 40 correos electrónicos 20 serán SPAM y 20 no. Los archivos de los correos serán nombrados de la siguiente forma:

g<n>01.txt

Donde <n> se reemplaza por el número de grupo y 01 representa el número de correo el cual llegará hasta 40.

- Después de la entrega de la primera parte se pondrá a disposición de toda la clase un directorio conjunto con los datos proporcionados por todos los grupos para el entrenamiento del modelo.
- El segundo entregable consistirá en el archivo spam_detector.py (no enviar los archivos de los correos electrónicos), con todos sus métodos propuestos implementados, consulte la cadena de documentación de cada método.
- Las fechas máximas de entrega de ambas partes se definirán en el campus virtual.
- El ejercicio se desarrollará de forma grupal.
- La entrega se realizará por uno sólo de los miembros del grupo, utilizando el enlace habilitado en el campus virtual.
- Para cada uno de los entregables deberá enviar únicamente un directorio comprimido que llevará por nombre:

Grupo n

Donde n es reemplazado por el número de grupo al que pertenece, por ejemplo Grupo 3. Asegúrese de poner este nombre tanto al directorio (antes de comprimirlo) como al archivo comprimido, de modo que al descomprimirlo siga manteniendo el nombre Grupo n.

Incluya dentro del directorio un fichero llamado "integrantes.txt", donde se listen sólo los miembros del grupo que colaboraron, incluyendo sus números de cuenta.

- No está permitido cambiar el nombre o el número de parámetros a las clases y métodos proporcionados. En caso de hacerlo el software evaluador no podrá asignarle los puntos correspondientes.
- Sí es permitido (y necesario) añadir atributos y métodos privados a las clases proporcionadas, así como importar módulos estándar del lenguaje (los mismos que se usan en el laboratorio).
- Se anunciará durante la clase al grupo o grupos que logren una precisión de entrenamiento más alta en los datos de prueba.
- El plagio se penalizará con una nota de 0.
- No cumplir con alguno de los puntos anteriores implicará una disminución en su nota, favor si alguno no está claro, consultar en clase.

Elaborado por: Raúl José Palma Mendoza.