Un dibujo de una persona

Descripción generada automáticamente con confianza baja



**TECNOLOGICO NACIONAL DE MEXICO**

**INSTITUTO TECNOLOGICO DE NUEVO LAREDO**

**INTELIGENCIA ARTIFICIAL 2**

INGENIERIA EN SISTEMAS COMPUTACIONALES

DOCENTE: LUIS DANIEL CASTILLO GARCIA

***U2 – PRACTICA 2 – REGRESION LOGISTICA***

JOEL RODRIGUEZ MUÑOZ

NUMERO DE CONTROL: 19100244

NUEVO LAREDO TAMAULIPAS.

13 Octubre de 2024

**Utilizando 20 correos de cualquier cuenta de correo electrónico, 15 no spam y 5**

**de spam**

**• Mostrar el contenido del correo**

**• Procesar el correo atreves del Parseador**

**• Etiquetar el correo**

**• Utilizar el modelo ya creado para predecir dichos 20 correos**

**• Comparar la predicción con la etiqueta**

**En este ejercicio se muestran los fundamentos de la Regresión Logística planteando uno de los primeros problemas que fueron solucionados mediante el uso de técnicas de Machine Learning: la detección de SPAM.**

**1.- FUNCIONES COMPLEMENTARIAS**

Captura de pantalla de un celular

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

LECTURA DE UN CORREO EN FORMATO RAW

Texto

Descripción generada automáticamente

PARSING DEL CORREO ELECTRONICO

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

LECTURA DEL INDICE

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

2.- PREPROCESAMIENTO DE LOS DATOS DEL CONJUNTO DE DATOS

Con las funciones presentadas anteriormente se permite la lectura de los correos electrónicos de manera programática y el procesamiento de estos para eliminar aquellos componentes que no resultan de utilidad para la detección de correos de SPAM. Sin embargo, cada uno de los correos sigue estando representado por un diccionario de Python con una serie de palabras.

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

El algoritmo de Regresión Logística no es capaz de ingerir texto como parte del conjunto de datos. Por lo tanto, deben aplicarse una serie de funciones adicionales que transformen el texto de los correos electrónicos parseados en una representación numérica.

APLICACIÓN DEL COUNTVECTORIZER

Texto

Descripción generada automáticamente

Imagen que contiene Tabla

Descripción generada automáticamente

APLICACIÓN DE OneHotEncoding

Texto

Descripción generada automáticamente

Funciones auxiliares para preprocesamiento del conjunto de datos

Pantalla de computadora con letras

Descripción generada automáticamente con confianza media

3.- Entrenamiento del algoritmo

Texto

Descripción generada automáticamente

Aplicamos la vectorización a los datos

Texto

Descripción generada automáticamente

Texto

Descripción generada automáticamente

Imagen que contiene Texto

Descripción generada automáticamente

Se genera el modelo de regresión logistica

Texto

Descripción generada automáticamente

4.- Predicción

Texto

Descripción generada automáticamente

Preprocesamiento de los correos con el vectorizador creado anteriormente

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente con confianza media

Predicción del tipo de correo

Una captura de pantalla de un celular

Descripción generada automáticamente con confianza media

Captura de pantalla de un celular

Descripción generada automáticamente

Repositorio de GitHub

<https://github.com/YoelRM/IA2.git>