



**TECNOLÓGICO  
NACIONAL DE MÉXICO**

**INSTITUTO TECNOLÓGICO DE CULIACÁN**  
**INGENIERIA EN SISTEMAS COMPUTACIONALES**

**INTELIGENCIA ARTIFICIAL**  
**PARADIGMAS DE LA INTELIGENCIA ARTIFICIAL Y**  
**EL PROCESO DE APRENDIZAJE AUTOMÁTICO**

**MAESTRO**

**ZURIEL DATHAN MORA FELIX**

**ALUMNOS**

**AYALA RODRÍGUEZ JOSÉ ERNESTO**  
**LIZARRAGA VALENZUELA JESUS EDUARDO**

**HORA:**

**11:00-12:00**

**09 de marzo del 2025**

**Culiacán, Sinaloa**

## INTRODUCCIÓN

**Este programa trata sobre el aprendizaje automatico para la clasificacion de correos spam y no spam por el cual utilizamos una dataset para clasificar los correos en el que eliminamos duplicados, convertimos todo a letra chica, tokenizamos y utilizamos el teorema de bayes para medir la probabilidad de que el siguiente correo sea spam**

## LIBRERÍAS QUE UTILIZAMOS

```
# Estas librerías se deben instalar si no las tiene. Use pip install pandas numpy nltk scikit-learn
import tkinter as tk # Para la interfaz gráfica
from tkinter import filedialog, messagebox, ttk # Componentes adicionales de la interfaz
import pandas as pd # Para manipular el dataset en formato tabla
import numpy as np # Para operaciones matemáticas y de arrays
import re # Para expresiones regulares, limpieza de texto
import unicodedata # Para eliminar acentos
import nltk # Para procesamiento de lenguaje natural
from nltk.corpus import stopwords # Palabras vacías (stopwords) en español
from sklearn.feature_extraction.text import TfidfVectorizer # Vectorización TF-IDF
```

## FUNCIONES PARA HACER LIMPIA DEL TEXTO, UTILIZANDO STOPWORDS PARA ELIMINAR LAS PALABRAS VACIAS DEL TEXTO, LOWER PARA PASAR TODO A MINÚSCULAS, QUITAMOS ACENTOS Y CARACTERES ESPECIALES

```
# Cargar las stopwords en español
stop_words = set(stopwords.words('spanish'))

# Función para eliminar acentos del texto
def quitar_acentos(texto):
    texto = unicodedata.normalize('NFKD', texto)
    texto = texto.encode('ascii', 'ignore').decode('utf-8')
    return str(texto)

# Función para limpiar el texto: pasar a minúsculas, quitar acentos y símbolos no alfanuméricos
def limpiar_texto(texto):
    texto = texto.lower().strip()
    texto = quitar_acentos(texto)
    texto = re.sub(r"[^a-z0-9\s]", "", texto)
    return texto
```

## LIMPIAMOS EL TEXTO, CLASIFICAMOS POR ETIQUETA Y MENSAJE

```
# Normalizar etiquetas ('spam' o 'ham') y limpiar los mensajes
df["etiqueta"] = df["etiqueta"].astype(str).str.lower().str.strip()
df["mensaje_limpio"] = df["mensaje"].astype(str).apply(limpiar_texto)
```

## SEPARAMOS POR SPAM Y NO SPAM

```
# Separar mensajes spam y no spam (ham)
spam = df[df["etiqueta"] == "spam"]
ham = df[df["etiqueta"] == "ham"]
```

## CALCULAMOS LA PROBABILIDAD DE SPAM Y NO SPAM POR MEDIO DE LA TEORÍA DE BAYES

```
# Calcular probabilidades base de spam y ham
P_spam = len(spam) / len(df)
P_no_spam = len(ham) / len(df)

# Vectorizar ambos subconjuntos
X_spam = vectorizer.transform(spam["mensaje_limpio"])
X_ham = vectorizer.transform(ham["mensaje_limpio"])

# Suavizado (Laplace) para evitar ceros
alpha = 1
P_caracteristicas_spam = (np.sum(X_spam.toarray(), axis=0) + alpha) / (np.sum(X_spam.toarray()) + alpha * len(palabras))
P_caracteristicas_no_spam = (np.sum(X_ham.toarray(), axis=0) + alpha) / (np.sum(X_ham.toarray()) + alpha * len(palabras))

# Función para clasificar un correo como spam o ham usando Naive Bayes con log-probabilidades
def clasificar_correo(texto):
    texto = limpiar_texto(texto)
    texto_vectorizado = vectorizer.transform([texto]).toarray()[0]
    log_P_spam = np.log(P_spam) + np.sum(np.log(P_caracteristicas_spam) * texto_vectorizado)
    log_P_no_spam = np.log(P_no_spam) + np.sum(np.log(P_caracteristicas_no_spam) * texto_vectorizado)
    return "spam" if log_P_spam > log_P_no_spam else "ham"
```

## Y LA DEMÁS PARTE DEL CÓDIGO ES DE LA INTERFAZ GRÁFICA

```
# Limpia el texto ingresado y los resultados
def limpiar_entrada():
    entrada_texto.delete("1.0", "end")
    salida_resultado.config(state="normal")
    salida_resultado.delete("1.0", "end")
    salida_resultado.config(state="disabled")

# Botones para analizar y limpiar
tk.Button(panel_centro, text="Analizar", command=analizar_texto).pack(pady=5)
tk.Button(panel_centro, text="Limpiar", command=limpiar_entrada).pack(pady=5)

# Caja de texto para mostrar resultados
salida_resultado = tk.Text(panel_centro, height=10, width=50, state="disabled")
salida_resultado.pack(pady=10)

# Panel derecho: muestra una tabla con parte del dataset#
frame_tabla = tk.Frame(frame)
frame_tabla.pack(side="right", padx=10, pady=10, fill="both", expand=True)

# Tabla con mensajes y etiquetas (sin predicción)
tabla = ttk.Treeview(frame_tabla, columns=("mensaje", "etiqueta"), show="headings")
tabla.heading("mensaje", text="Mensaje")
tabla.heading("etiqueta", text="Etiqueta")
tabla.column("mensaje", width=400)
tabla.column("etiqueta", width=80)

# Scroll para la tabla
scrollbar = ttk.Scrollbar(frame_tabla, orient="vertical", command=tabla.yview)
tabla.configure(yscrollcommand=scrollbar.set)
scrollbar.pack(side="right", fill="y")
tabla.pack(fill="both", expand=True)

# Llenar la tabla con los primeros 100 mensajes del dataset
for _, fila in df.head(100).iterrows():
    tabla.insert("", "end", values=(fila["mensaje"][:100], fila["etiqueta"]))

# Ejecutar la aplicación
ventana.mainloop()
```

CODIGO EN EJECUCIÓN

Asunto: ¡Felicitaciones! ¡Ganaste un iPhone 15 GRATIS!

De: promociones@superpremiosonline.com

Cuerpo del mensaje:

Estimado usuario,

¡Ha sido seleccionado al azar como el GRAN GANADOR de un iPhone 15 totalmente GRATIS!

Para reclamar su premio, simplemente haga clic en el siguiente enlace y complete sus datos:

👉 Haz clic aquí para reclamar tu premio

¡Date prisa! Esta oferta expira en 24 horas.

Atentamente,

El equipo de Super Premios Online

Analizar

Limpiar

Resultado del análisis:

Clasificación del correo: SPAM

Precisión del modelo: 0.9922

Recall (solo spam): 0.9968

| Mensaje                                                                    | Etiqueta |
|----------------------------------------------------------------------------|----------|
| Tu paquete llegará mañana a las 5:00 PM.                                   | ham      |
| ¡Estamos muy felices de que te unas a nuestro programa!                    | ham      |
| Nos gustaría saber cómo fue tu experiencia con nosotros.                   | ham      |
| Recuerda que tienes una reunión con el equipo de ventas el martes.         | ham      |
| Tu cuenta ha sido verificada correctamente.                                | ham      |
| Gracias por tu compra, ¡disfruta de tus productos!                         | ham      |
| Recuerda que tienes un examen el viernes.                                  | ham      |
| No olvides traer todos los documentos necesarios para tu cita.             | ham      |
| Recuerda revisar tus correos para obtener más información.                 | ham      |
| ¡Nos encantaría verte en nuestro próximo evento!                           | ham      |
| Gracias por tu tiempo en nuestra conferencia.                              | ham      |
| Nos alegra saber que te gustó el producto.                                 | ham      |
| Recuerda que el pago de tu suscripción es automático.                      | ham      |
| Nos vemos en el evento de este sábado.                                     | ham      |
| Tu inscripción en el taller ha sido confirmada.                            | ham      |
| Recuerda que tu pedido será entregado en 48 horas.                         | ham      |
| Tu cuenta de correo ha sido suspendida.                                    | ham      |
| Recuerda que la tienda cierra a las 8:00 PM.                               | ham      |
| Gracias por ser parte de nuestro proyecto.                                 | ham      |
| ¡Felicidades! Has ganado un premio de \$5000.                              | spam     |
| ¡Haz clic aquí para reclamar tu premio ahora mismo!                        | spam     |
| ¡Gana un iPhone nuevo solo por registrarte!                                | spam     |
| Última oportunidad para ganar un auto nuevo, solo responde a este mensaje. | spam     |
| ¿Quieres ganar \$1000? Haz clic aquí para participar.                      | spam     |
| ¡Felicitaciones, has sido seleccionado para una beca!                      | spam     |
| Participa en nuestra encuesta y gana un premio.                            | spam     |
| ¡Gana dinero desde casa con solo hacer clic en este enlace!                | spam     |
| Te hemos seleccionado para un sorteo, haz clic aquí para participar.       | spam     |
| ¡Haz clic aquí y gana una tarjeta de regalo de \$100!                      | spam     |
| Recuerda que solo quedan 24 horas para recibir un descuento del 50%.       | spam     |
| ¡Compra ahora y recibe un descuento de \$100!                              | spam     |
| Haz clic aquí para obtener tu descuento exclusivo.                         | spam     |
| Te has ganado un viaje a París, haz clic para más detalles.                | spam     |
| ¿Quieres ganar \$5000? Solo necesitas responder este mensaje.              | spam     |
| Última oportunidad para obtener una suscripción gratuita.                  | spam     |
| Haz clic para ver tu descuento exclusivo de tiempo limitado.               | spam     |
| ¡Felicidades! Has ganado un celular de última generación.                  | spam     |
| Participa y gana una laptop nueva.                                         | spam     |
| Este es tu último aviso, haz clic para reclamar tu premio.                 | spam     |
| ¡Gana dinero fácil con solo hacer clic aquí!                               | spam     |