

# Detector de Spam basado en ML

Proyecto para el DS Bootcamp de  
The Bridge



# El Problema del Spam

## 1. Porcentaje Promedio de Spam en el Tráfico Mundial de Correo Electrónico (2021)

En 2021, el porcentaje promedio de spam en el tráfico mundial de correo electrónico fue del 45,56%.

Fuente: Kaspersky

## 2. España: Líder Mundial en Recepción de Spam

Durante el segundo trimestre de 2020, España fue el país que más ataques de spam recibió.

Representó un 9,3% del total de estas amenazas a nivel mundial.

Fuente: Europa Press

## 3. Tasa de Quejas de Spam Aceptable

La tasa de quejas de spam estándar aceptable en la industria es inferior al 0,1%.

Esto equivale a recibir una queja por cada 1.000 mensajes enviados.

Fuente: ActiveCampaign



# Descripción del Proyecto

**Objetivo:** Clasificar textos como spam o ham (legítimos) en inglés y español.

## Resultados:

- Creación de modelo/s entrenados para la detección de spam en español e inglés a partir de textos en lenguaje natural, así como de los procesos que los habiliten.
- Aplicación: Desarrollada en Streamlit para proporcionar una interfaz web fácil de usar.

## Bienvenido a nuestro servicio de detección de spam

Esta aplicación permite detectar si un mensaje es spam o no en inglés y español. Seleccione el idioma, ingrese el texto del mensaje y haga clic en el botón de - Detectar Spam.

Seleccione el idioma del mensaje:

Español



## Ingrese el texto del mensaje a evaluar:

Texto del mensaje:

Recuerda llamar a tu madre

Detectar Spam

## Resultado: Ham

Probabilidad de Spam: 23.65%

Probabilidad de No Spam (Ham): 76.35%

---

Desarrollado por Fernando Manzano. Utilizamos técnicas de Machine Learning para ofrecer una detección precisa de spam.

# Datos utilizados

## Español:

Fuente: Hugging Face

[https://huggingface.co/datasets/softcapps/spam\\_ham\\_spanish/tree/main](https://huggingface.co/datasets/softcapps/spam_ham_spanish/tree/main)

- Descripción: +1000 mensajes de texto en español etiquetados como spam o ham en train.csv y test.csv

- Estructura:

- 'Mensaje': Texto del mensaje.
- 'tipo': Indicación de spam o ham.

## Inglés:

Fuente: Kaggle



<https://www.kaggle.com/datasets/venky73/spam-mails-dataset>

- Descripción: 4993 mensajes únicos de texto en inglés etiquetados como spam o ham en spam\_ham\_dataset.csv

- Estructura:

- 'Mensaje': Texto del mensaje.
- 'label': Indicación de spam o ham.
- 'label\_num': Indicación de spam = 1 o ham = 0.

mensaje	tipo
string	string
Descubre como perder peso rapidamente	spam
Necesitas ayuda con tu tarea	ham
Gana dinero desde casa sin esfuerzo	spam
Reclama tu herencia de un pariente lejano	spam
Mejora tu rendimiento sexual con este producto	spam
Por favor responde a esta encuesta	ham

#	label	text	# label_num
	Labels of Emails which can be either Spam or Ham	Emails data	if spam it's 1, or else it's 0
	ham 71% spam 29%	4993 unique values	
605	ham	Subject: enron methanol ; meter # : 988291 this is a follow up to the note i gave you on monday , 4...	0
2349	ham	Subject: hpl nom for january 9 , 2001 ( see attached file : hplnol 09 . xls ) - hplnol 09 . xls	0
3624	ham	Subject: neon retreat ho ho ho ,	0

# Preprocesamiento de datos

- Eliminación de duplicados y valores repetidos.
- Eliminación de palabras vacías usando stopwords de NLTK.
- Eliminación de caracteres no alfanuméricos.
- Conversión a minúsculas.
- Eliminación de espacios extra y prefijos como "Subject:".

```
5 nltk.download('stopwords')
6 stopwords_sp = stopwords.words('spanish')
7
8 # Agregar palabras adicionales que no aporten significado en este contexto específico
9 stopwords_sp.extend(['este', 'nuestro', 'con', 'para', 'esta']) # Podemos ajustar esta
10
11 # Función para limpiar el texto y filtrar stopwords
12 def limpiar_texto_con_stopwords(texto):
13     texto = re.sub(r'\W', ' ', texto) # Eliminar caracteres no alfanuméricos
14     texto = texto.lower() # Convertir a minúsculas
15     texto = re.sub(r'\s+', ' ', texto) # Eliminar espacios extra
16     palabras = texto.split()
17     palabras_filtradas = [palabra for palabra in palabras if palabra not in stopwords_sp]
18     return ' '.join(palabras_filtradas)
19
20 # Aplicar limpieza con filtro de stopwords a los mensajes
21 test_df['mensaje_limpio_stopwords'] = test_df['mensaje'].apply(limpiar_texto_con_stopwords)
22
```

# Modelos utilizados

## - **TfidfVectorizer:**

- Convierte textos en una matriz de características TF-IDF.
- Asigna peso a cada palabra basado en su frecuencia en el documento y su frecuencia inversa en el conjunto de documentos.

## - Modelos Evaluados:

- Español: Naive Bayes, Random Forest y SVM. Seleccionado: **Naive Bayes**.

Naive Bayes es adecuado para el tratamiento de textos y detección de spam debido a su simplicidad, eficiencia computacional, capacidad para manejar grandes volúmenes de datos, y su uso eficaz de la independencia condicional entre palabras para calcular probabilidades.

- Inglés: Naive Bayes y SVM. Seleccionado: **SVM optimizado con GridSearchCV**.

SVM para el tratamiento de textos y detección de spam destaca por su capacidad para manejar espacios de alta dimensión, como los textos, encontrar un hiperplano óptimo que maximiza la separación entre clases y su robustez ante datos ruidosos.

## Resultados y Evaluación

	<b>NB Español</b>	<b>SVM Inglés</b> Optimizado: {'C': 10, 'class_weight': None, 'kernel': 'rbf'}
<b>Precisión</b>	0.89	0.98
<b>Recall</b>	0.89	0.98
<b>F1 - Score</b>	0.89	0.98
<b>Matriz de Confusión</b>	[[94 15]  [ 8 92]]	[[719 13]  [ 7 260]]

# Con los usuarios en mente. Una app construida con Streamlit

## Detección de Spam en Mensajes

### Bienvenido a nuestro servicio de detección de spam

Esta aplicación permite detectar si un mensaje es spam o no en inglés y español. Seleccione el idioma, ingrese el texto del mensaje y haga clic en el botón de - Detectar Spam.

Seleccione el idioma del mensaje:

Español

### Ingrese el texto del mensaje a evaluar:

Texto del mensaje:

Quiero conocerte

Detectar Spam

### Resultado: Spam

Probabilidad de Spam: 64.75%

Probabilidad de No Spam (Ham): 35.25%

### Bienvenido a nuestro servicio de detección de spam

Esta aplicación permite detectar si un mensaje es spam o no en inglés y español. Seleccione el idioma, ingrese el texto del mensaje y haga clic en el botón de - Detectar Spam.

Seleccione el idioma del mensaje:

Inglés

### Ingrese el texto del mensaje a evaluar:

Texto del mensaje:

Amazon prime days

Detectar Spam

### Resultado: Ham

Probabilidad de Spam: 10.91%

Probabilidad de No Spam (Ham): 89.09%

---

Desarrollado por Fernando Manzano. Utilizamos técnicas de Machine Learning para ofrecer una detección precisa de spam.



## Lecciones

- Importancia del **orden** desde el primer momento.
- Existen **distintos enfoques** de solución para un mismo problema.
- Hay mucho que descubrir en la potencia y variedad de las **bibliotecas** existentes de ML y DL.



## Mejoras

- Incluir **explicación** del funcionamiento de la app.
- Aplicar a **nuevos casos** como phishing y SMS fraudulentos.
- Crear un **dataset de spam en español**.
- Mejorar modelos para **predicciones erróneas**.

