Informe de Clasificación de SPAM

Modelo: Regresión Logística (scikit-learn)
Autor: Leny Lopez - Curso 802 - Materia: Machine Learning

Se entrenó un modelo de regresión logística para clasificar correos como SPAM o HAM usando 10 características. Este modelo fue elegido porque ofrece interpretabilidad y un buen balance entre simplicidad y rendimiento.

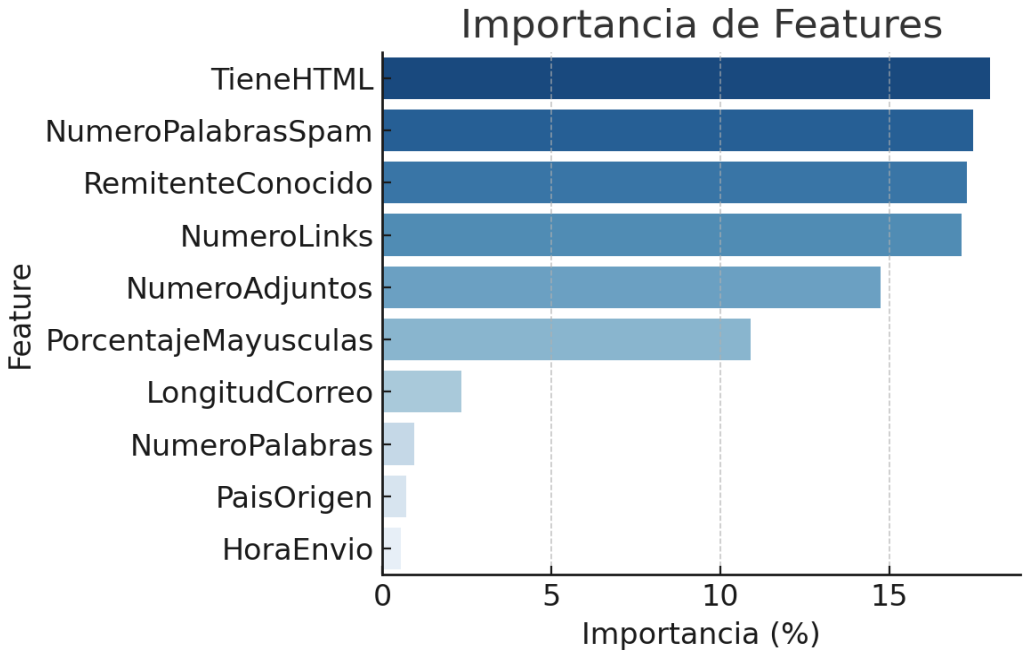
Número de observaciones: 1000
Número de features usados: 10
Umbral óptimo (máx F1): 0.58
Mejor F1-Score: 0.941
ROC-AUC: 0.969

Matriz de confusión (ham=0, spam=1):
TN=94 FP=11
FN=12 TP=183

El modelo asigna distinta importancia a cada característica. La siguiente tabla muestra su peso relativo expresado en porcentaje.

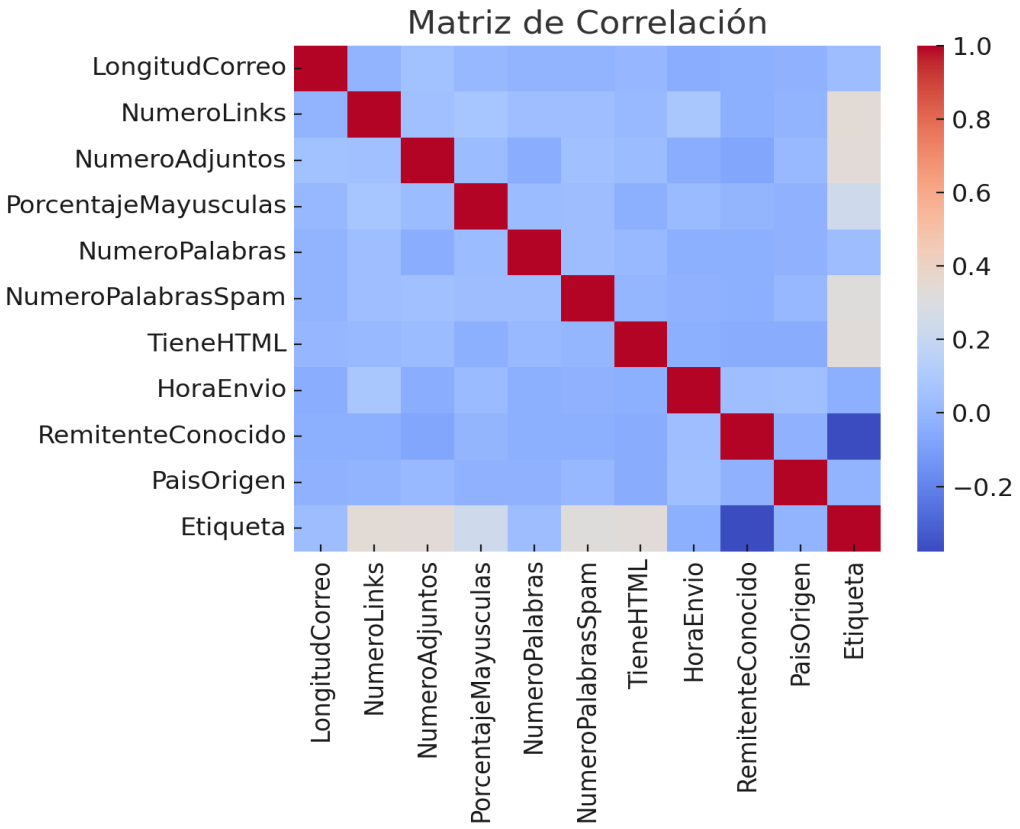
Feature	Importancia (%)
TieneHTML	17.98%
NumeroPalabrasSpam	17.48%
RemitenteConocido	17.30%
NumeroLinks	17.13%
NumeroAdjuntos	14.74%
PorcentajeMayusculas	10.89%
LongitudCorreo	2.33%
NumeroPalabras	0.94%

PaisOrigen	0.69%
HoraEnvio	0.54%



Correlación de cada característica con la etiqueta SPAM y matriz completa de correlaciones:

Feature	Correlación
RemitenteConocido	-0.38
NumeroLinks	0.34
NumeroAdjuntos	0.34
TieneHTML	0.33
NumeroPalabrasSpam	0.32
PorcentajeMayusculas	0.24
LongitudCorreo	0.03
HoraEnvio	-0.03
NumeroPalabras	0.03
PaisOrigen	-0.02



El análisis revela que no todos los atributos influyen de igual manera. El número de palabras sospechosas es el factor más relevante, seguido por el número de enlaces y el porcentaje de mayúsculas. Esto refleja que el lenguaje llamativo, la insistencia en redirecciones y el abuso de recursos visuales son señales claras de SPAM.

Las demás variables también aportan: la inclusión de HTML, adjuntos o el país de origen añaden matices que ayudan al modelo a diferenciar un correo legítimo de uno no deseado. La combinación de todas ellas permite alcanzar un F1 de 0.94, mostrando que cada detalle suma valor.

Los 10 features se mantuvieron porque todos mostraron información relevante. En particular, 'PaisOrigen' fue codificado con un LabelEncoder para mantener el diseño original del dataset. Aunque esto simplifica la representación, en un escenario real sería recomendable un One-Hot Encoding.

Más allá de las métricas, este proyecto nos permite comprender mejor cómo funciona el SPAM. Los atacantes buscan captar atención con mayúsculas, repetir palabras clave y redirigir a múltiples enlaces. El modelo simplemente aprende estos patrones y los utiliza para tomar decisiones rápidas y eficientes.

De este modo, el aprendizaje automático no solo filtra correos: también nos ofrece una forma de entender el comportamiento digital y protegernos mejor en entornos cotidianos.