

REPORTE DE IMPORTANCIA DE FEATURES

Sistema de Detección de Spam - Regresión Lineal

INFORMACIÓN DEL MODELO

- Algoritmo: Regresión Lineal
- Instancias de entrenamiento: 800
- Instancias de prueba: 200
- Total de features: 1000
- Precisión: 1.0000
- R^2 Score: 0.9992

TOP 10 FEATURES MÁS IMPORTANTES

Rank	Feature	Coeficiente	Dirección	Importancia
1	su servicio	-0.4358	HAM	Alta
2	hemos	-0.1651	HAM	Alta
3	verifique	0.1464	SPAM	Alta
4	contáctenos	0.1450	SPAM	Alta
5	actualizado su	-0.1398	HAM	Alta
6	hemos actualizado	-0.1398	HAM	Alta
7	cuenta inmediatamente	0.1368	SPAM	Alta
8	está	-0.1319	HAM	Alta
9	descubrimiento increíble	0.1287	SPAM	Alta
10	su revisión	-0.1191	HAM	Alta

INTERPRETACIÓN

- Coeficiente positivo: La feature indica SPAM
- Coeficiente negativo: La feature indica HAM
- Valor absoluto mayor: Mayor importancia en la decisión
- Las features se obtuvieron mediante CountVectorizer con n-gramas (1,2)