

# Análisis del Filtro de Spam usando Modelos de Markov

May 10, 2025

## 1 Introducción

El enfoque para implementar un filtro de spam con un modelo de Markov se basa en analizar las **transiciones de palabras** dentro de los mensajes, en lugar de simplemente contar ocurrencias independientes como en el modelo de **Naive Bayes**. Esto permite capturar el contexto entre palabras, algo fundamental en el procesamiento del lenguaje natural.

## 2 Modelo de Markov

Un **Modelo de Markov** de primer orden asume que la probabilidad de que una palabra aparezca depende únicamente de la palabra que la precede, sin importar el resto del contexto previo. Esto se expresa matemáticamente como:

$$P(w_n|w_{n-1}, C) = \frac{P(w_{n-1}, w_n, C)}{P(w_{n-1}, C)} \quad (1)$$

donde:

- $w_n$  es la palabra actual,
- $w_{n-1}$  es la palabra anterior,
- $C$  es la clase del mensaje (Spam o Ham).

## 3 Matrices de Transición

Los siguientes son los resultados de las matrices de transición para los mensajes de entrenamiento:

### 3.1 Matriz de Transición para Spam

	Gana	dinero	Reclama	tu
Gana	0	0.167	0	0
dinero	0	0	0	0.167
Reclama	0	0	0	0.167
tu	0	0.167	0	0

 (2)

### 3.2 Matriz de Transición para Ham

	Reunión	a	las	3PM	Actualización	del
Reunión	0	0.133	0	0	0	0
a	0	0	0.133	0	0	0
las	0	0	0	0.133	0	0
Actualización	0	0	0	0	0	0.133
del	0	0	0	0	0	0.133
proyecto	0	0	0	0	0	0.133

 (3)

## 4 Cálculo de las Probabilidades de Transición

Aplicando **suavizado de Laplace** para evitar probabilidades cero:

$$P(\text{siguiente palabra}|\text{palabra actual}, C) = \frac{\text{conteo del par} + 1}{\text{conteo total de la palabra actual} + \text{tamaño del vocabulario}} \quad (4)$$

## 5 Cálculo para el Mensaje de Prueba

Para el mensaje **"Reclama tu dinero premio"**:

$$\begin{aligned} P(\text{Spam}|\text{Mensaje}) &= P(\text{Reclama}|\text{Spam}) \cdot P(\text{tu}|\text{Reclama}, \text{Spam}) \cdot P(\text{dinero}|\text{tu}, \text{Spam}) \cdot P(\text{premio}|\text{dinero}, \text{Spam}) \\ &= 0.167 \cdot 0.167 \cdot 0.167 \cdot 0.167 \\ &= 0.000078 \end{aligned}$$

$$\begin{aligned} P(\text{Ham}|\text{Mensaje}) &= P(\text{Reclama}|\text{Ham}) \cdot P(\text{tu}|\text{Reclama}, \text{Ham}) \cdot P(\text{dinero}|\text{tu}, \text{Ham}) \cdot P(\text{premio}|\text{dinero}, \text{Ham}) \\ &= \frac{1}{15} \cdot \frac{1}{15} \cdot \frac{1}{15} \cdot \frac{1}{15} \\ &= 0.00003 \end{aligned}$$

## 6 Decisión Final

Comparando las dos probabilidades:

$$0.000078 > 0.00003 \quad (5)$$

El mensaje se clasifica como **Spam**.

## 7 Ventajas del Enfoque de Markov sobre Naive Bayes

- **Considera el Contexto:** Incluye dependencias entre palabras.
- **Más Realista para Lenguaje Natural:** Captura la estructura del lenguaje mejor que un enfoque totalmente independiente.