

ANALÍTICA DE DATOS Y HERRAMIENTAS DE  
INTELIGENCIA ARTIFICIAL II

# DIMEMEX

Hector Julian Zarate Ramirez A01027743  
Andres Eduardo Gomes Brandt A01781321

# DIMEMEX - 2025

Es en conjunto de datos que esta conformado por al rededor de **3,000** memes, los cuales fueron recolectados en grupos de Facebook en México que ya fueron manualmente etiquetados

**Etiquetado simple :** *none, inappropriate content, and hate-speech*

**Etiquetado complejo :** *none, inappropriate content, Sexism, Racism, Classicism and other (hate-speech)*



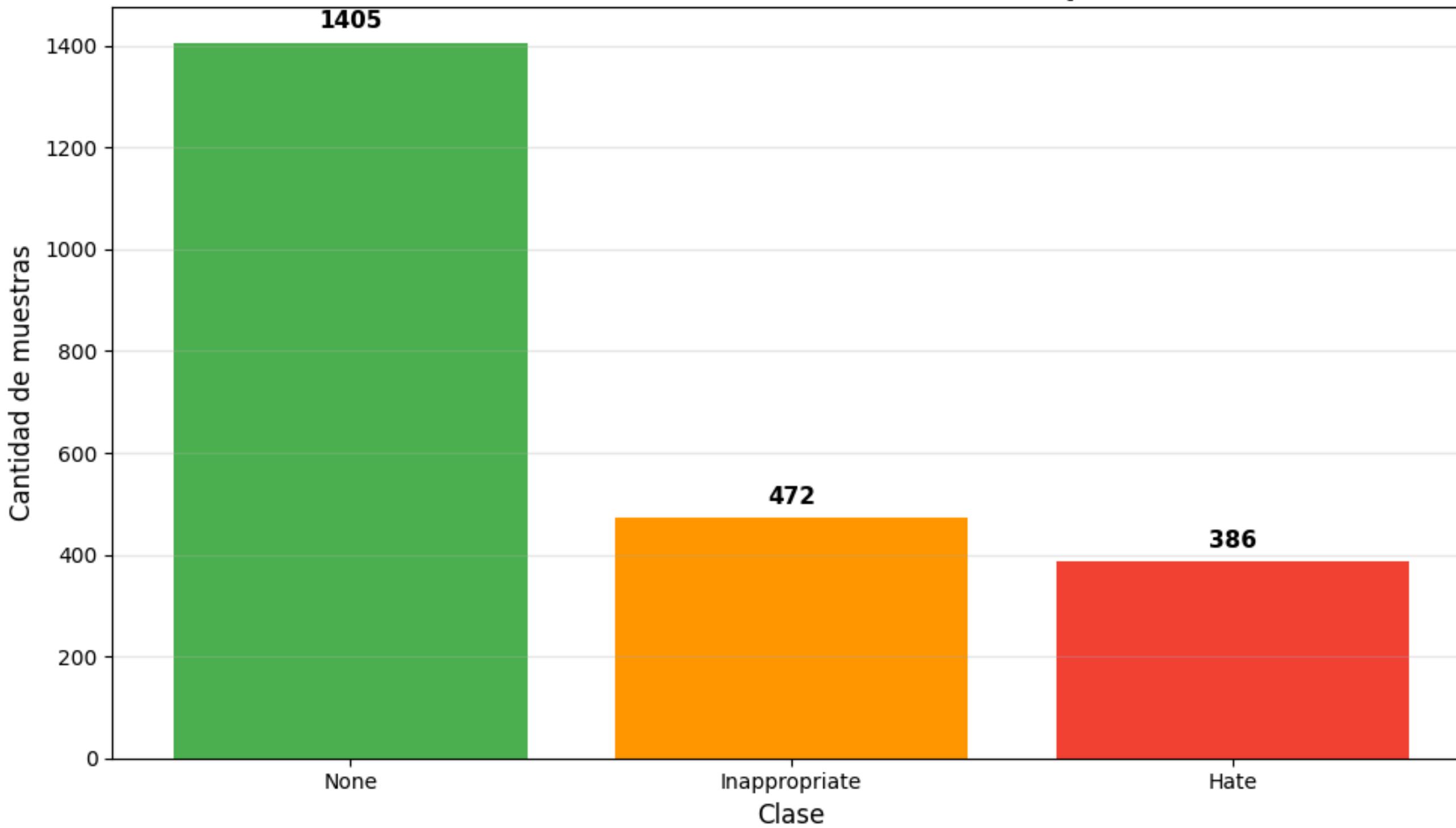
**Category:** Neither/Harmless

**OCR text:** Yo viendo que son las 1:59 AM ////  
Y de pronto cambia a las 03:00 AM

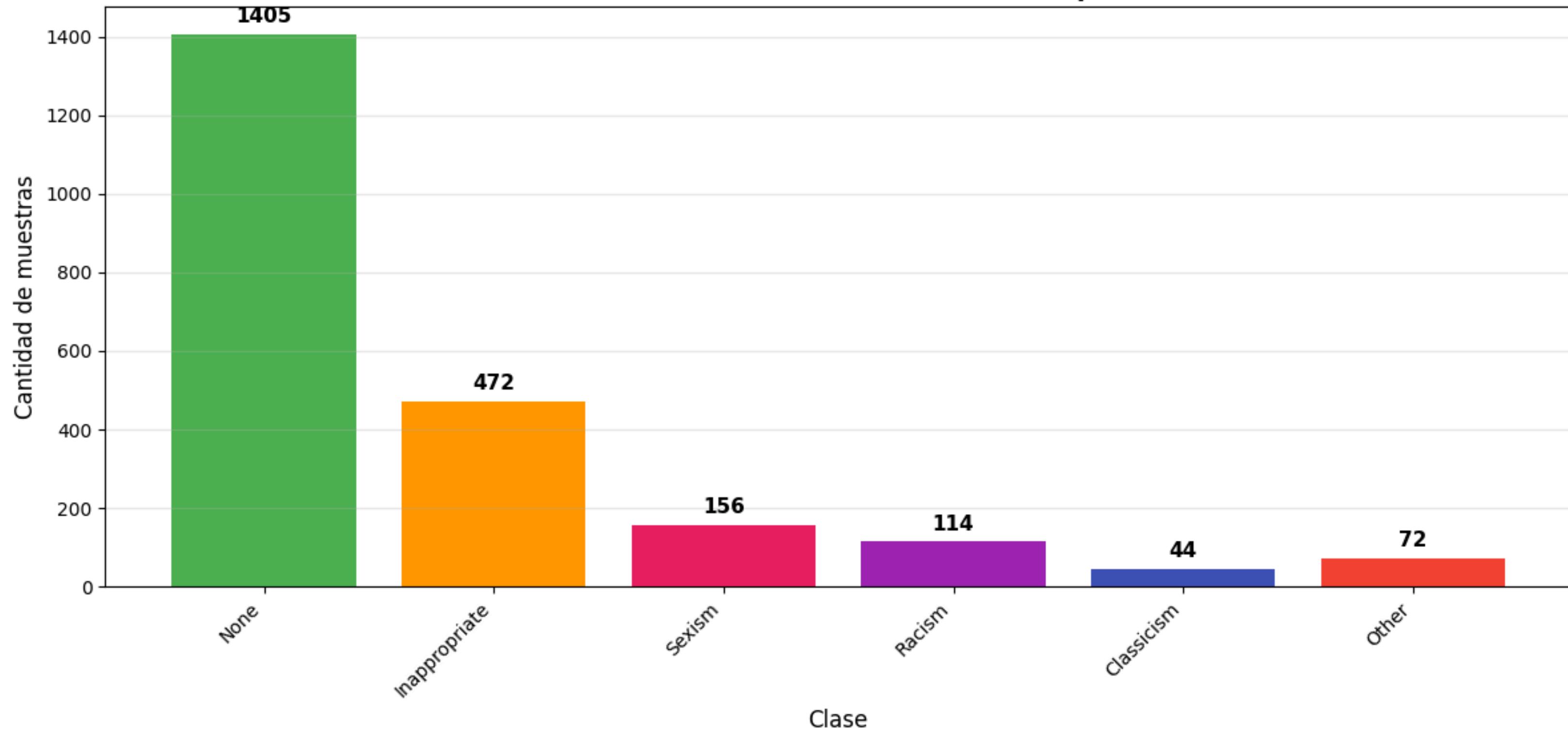
**Image Caption:** La imagen es un meme que consta de dos paneles extraídos de una escena de un programa de televisión. El personaje en el primer panel tiene una expresión relajada y sonriente, mientras que en el segundo panel su rostro muestra una sorpresa extrema, con los ojos muy abiertos.

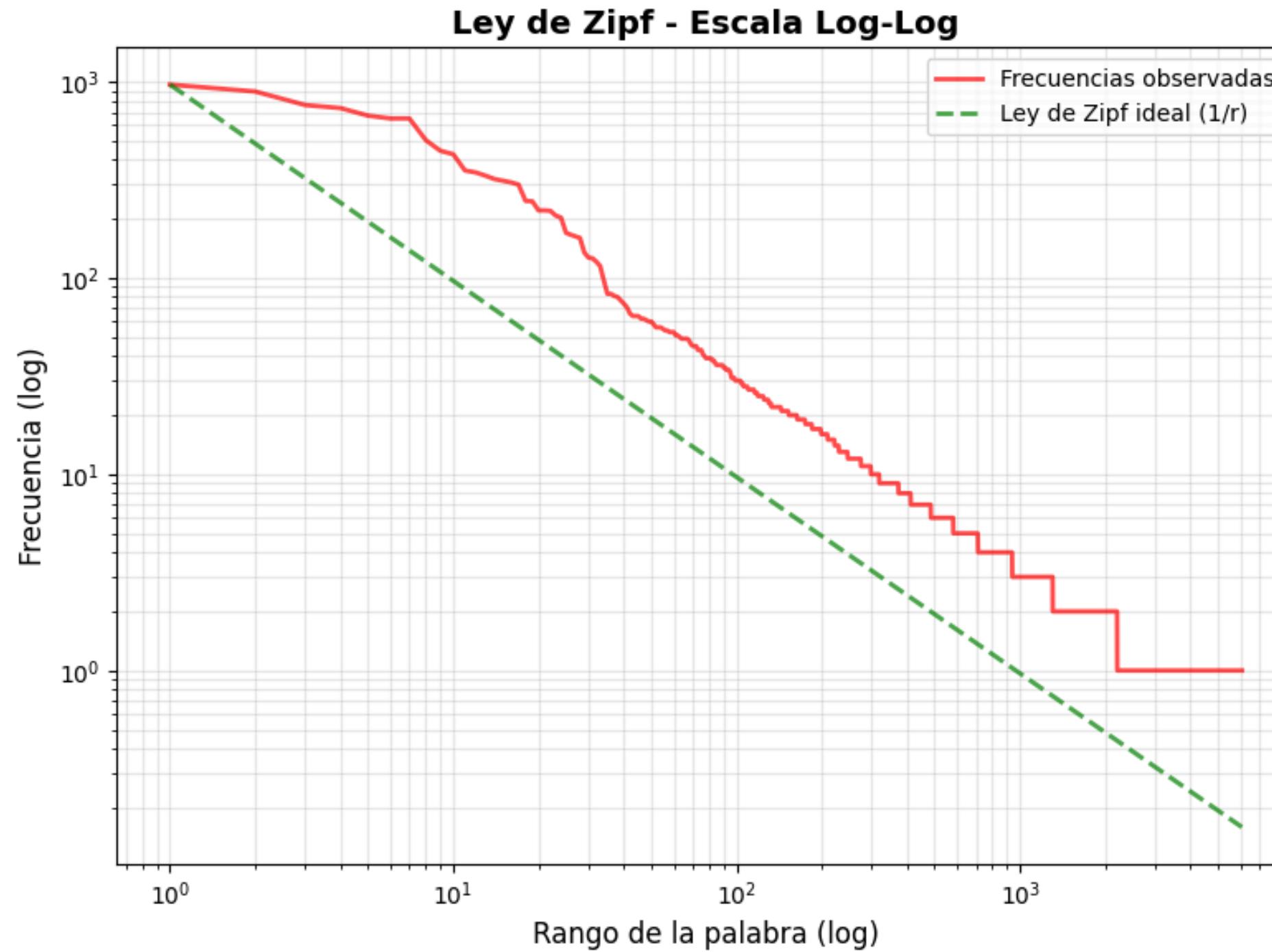
# **EXPLORACIÓN DE DATOS**

### Distribución de clases - Dataset Simple



### Distribución de clases - Dataset Complex

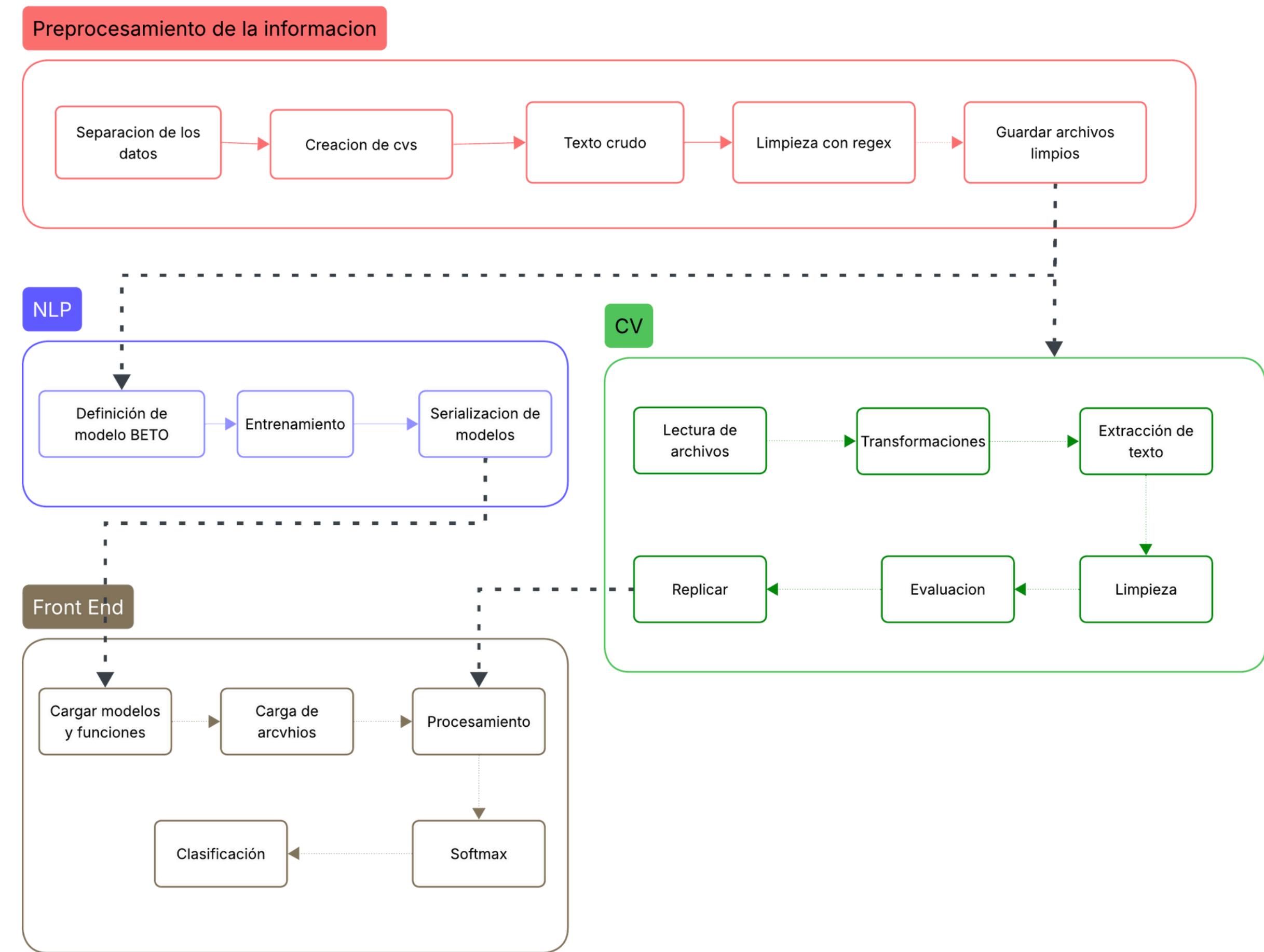




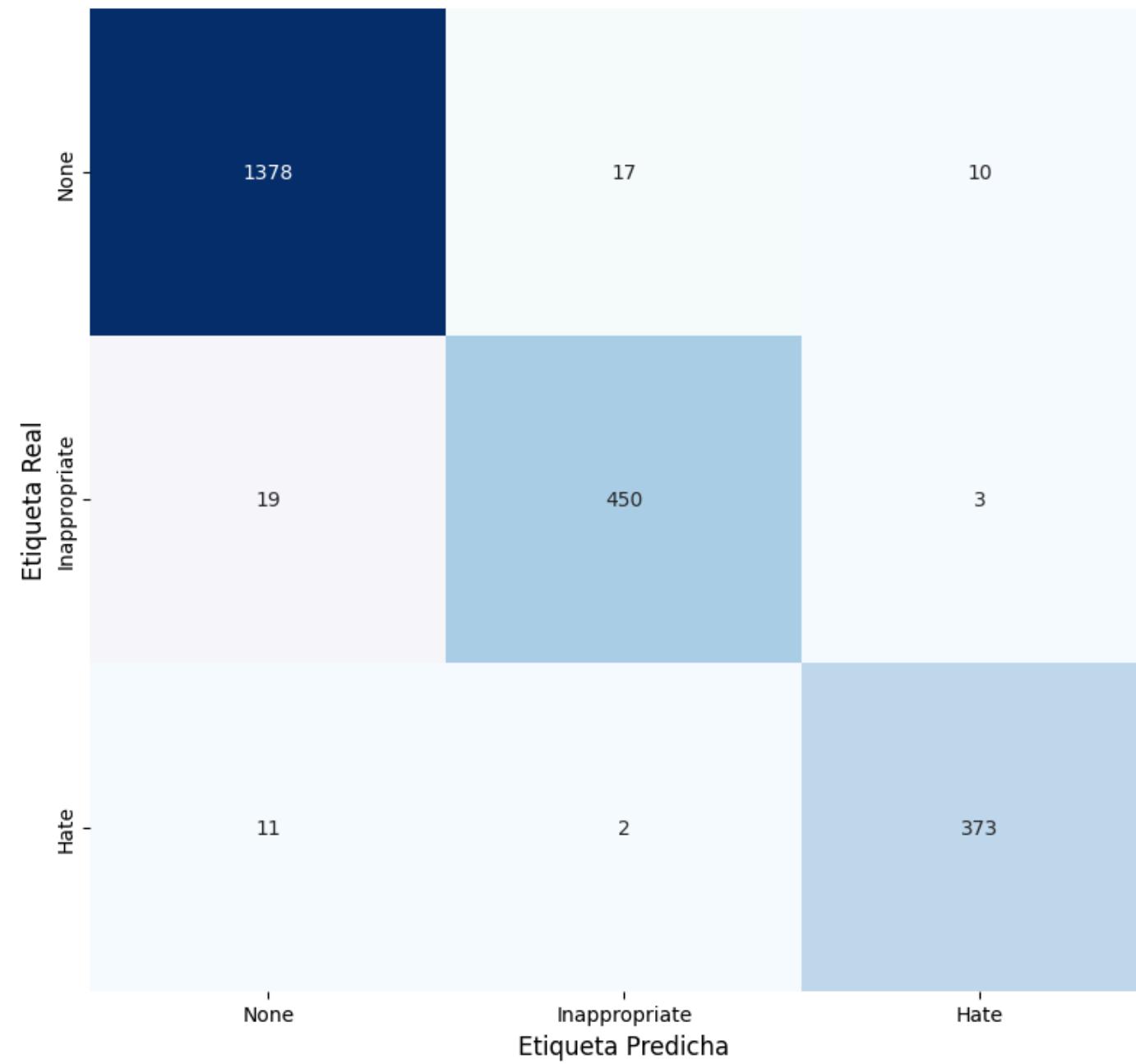
## VOCABULARIO

- 28,990 palabras en total
- 6,416 palabras unicas
- palabras mas comunes “de”,  
“que”, “a”, “la” y “el”

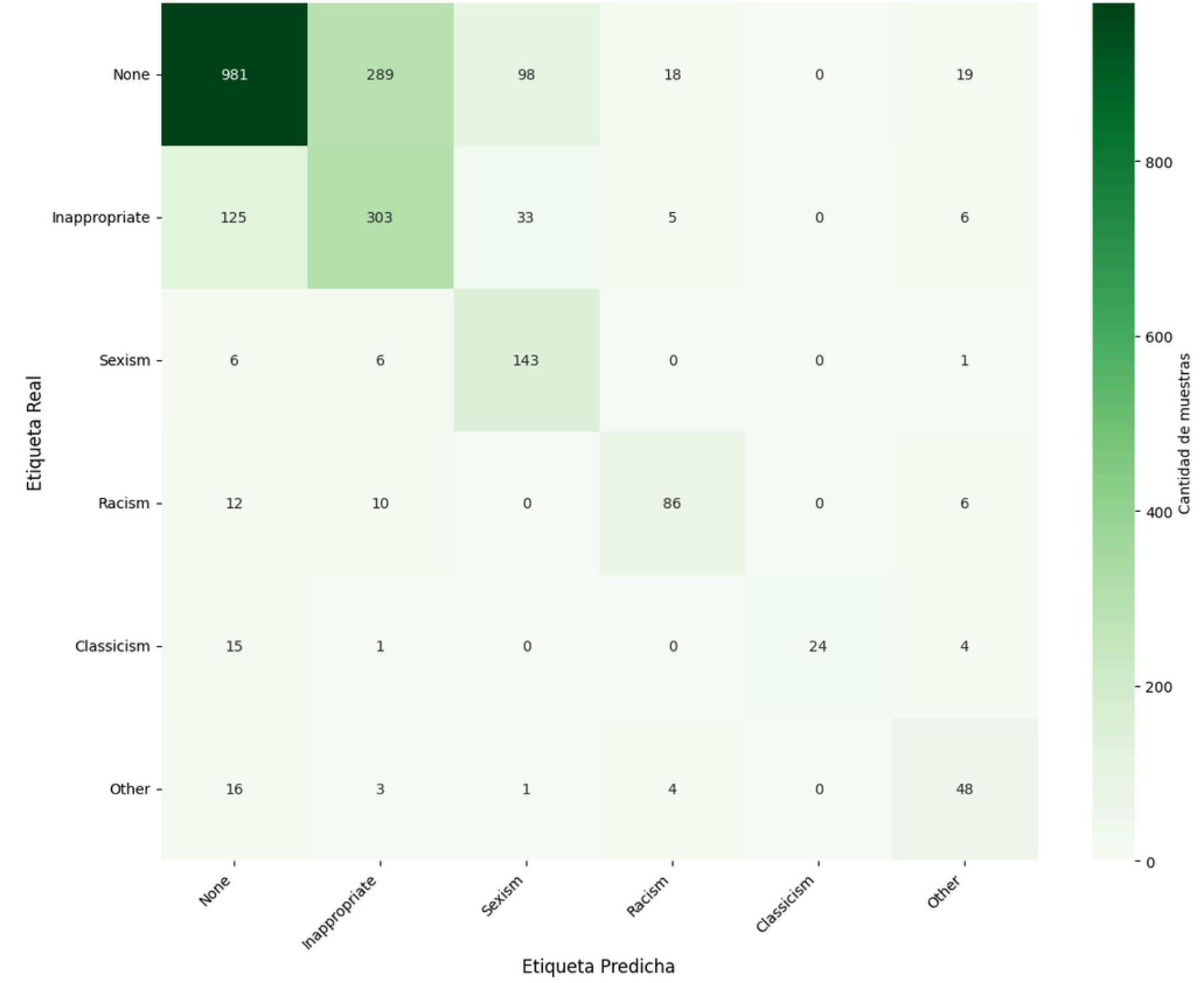
# PIPELINE



**Matriz de Confusión - Modelo Simple**



**Matriz de Confusión - Modelo Complex**



# RESULTADOS



## ¡ REGLAS !

1. Memes que tengan texto
2. Memes con el texto en español

**GRACIAS**