

## Tarea 2: Aprendizaje bayesiano

### Problema

Considere el problema de predicción de palabras que todos utilizamos diariamente en servicios como el buscador de Google o los teclados de los celulares.

Se propone implementar una solución a dicho problema utilizando el algoritmo Naive Bayes, asumiendo independencia en el orden de aparición de las  $N$  últimas palabras de la frase que se viene escribiendo ( $N$  es un hiperparámetro de la implementación).

Para entrenar el modelo cada grupo deberá descargar de su propio Whatsapp<sup>1</sup> el contenido del grupo que mas conversaciones tenga. Dicho contenido no debe ser parte de la entrega; es sólo a los efectos de que las pruebas que hagan tengan sentido para ustedes, y para estandarizar el formato de carga de datos. Con dichos datos se podrán aproximar las probabilidades requeridas por el método.

Para probarlo deberán utilizar un simulador de cliente implementado en un cuaderno provisto<sup>2</sup>. Este simulador permite crear frases *palabra por palabra*, mostrar las palabras recomendadas (resultado del algoritmo a implementar), aceptar la recomendación (con ENTER) o ignorarla (escribiendo una nueva palabra para la frase), y recomenzar con una frase nueva (con PUNTO).

La entrega deberá ser realizada en una copia de dicho cuaderno, donde deberán:

1. Modificar la función `recomendacion_bayesiana` con su implementación del algoritmo descrito.
2. Permitir cargar el archivo de entrenamiento con un CSV provisto por Whatsapp y considerar solo el texto de los mensajes (ignorando las columnas fecha y autor)
3. Comparar el desempeño del algoritmo variando el hiperparámetro  $N$  (con valores 1, 2, 3 y 4)
4. Reentrenar el modelo al finalizar cada frase (cada vez que se ingresa un PUNTO). Esto puede ser computacionalmente pesado pero muestra la capacidad del método para adaptarse a nueva evidencia.

Este algoritmo será intenso en uso de CPU. Se deberán utilizar las estructuras de datos más adecuadas provistas por el lenguaje Python para que la implementación sea lo más limpia y eficiente posible.

### Entregables

El informe a entregar debe ser una copia del cuaderno provisto, donde se incluirá:

- el código escrito para resolver el problema
- un informe con las pruebas realizadas y los resultados obtenidos

### Fecha límite de entrega

Miércoles 13 de setiembre (inclusive).

1 En Whatsapp: Configuración → Chats → Exportar chat

2 [https://gitlab.fing.edu.uy/aprendaut2/aa23/-/raw/main/notebooks/Cliente\\_Lab2\\_2023.ipynb?](https://gitlab.fing.edu.uy/aprendaut2/aa23/-/raw/main/notebooks/Cliente_Lab2_2023.ipynb?inline=false)  
`inline=false`