

# Evaluación 2

Diabb Zegpi Delgado

Esta evaluación es de carácter individual y tiene una ponderación del 35% sobre la nota final de la asignatura. Los contenidos a evaluar son: stemming, lematización, análisis de sentimientos y análisis de redes. Se solicita entregar el código junto con las respuestas a las preguntas, en formato de notebook (ipynb, rmd, qmd, html, etc.). La entrega deberá ser realizada en classroom, y tendrá plazo hasta el día viernes 30 de septiembre para entregar sus resultados.

Para el desarrollo de esta evaluación, tendrá que escoger un dataset único (no usado por sus compañeros), cuyo contenido principal sea texto. Este dataset debe ser validado por el profesor. Se sugiere que visite el compendio de datasets en [www.kaggle.com](http://www.kaggle.com).

## Ejercicio 1 (65 puntos)

1. A partir de los datos que seleccionó, proponga un plan para desarrollar un producto de uso privado o comercial. Este plan debe considerar las tareas específicas que realizaría en cada una de las etapas del proyecto de text mining (detección de outliers, EDA, tokenización, entre otros): obtención de datos, limpieza, procesamiento, modelización y monitoreo del desempeño (30 puntos).
2. Ejecute las tareas que declaró en obtención de datos y limpieza (10 puntos).
3. Ejecute el procesamiento, empleando listas de stopwords y lematización/stemming. Justifique sus decisiones de diseño, por ejemplo: ¿por qué unigramas?, ¿por qué lematización/stemming?, ¿por qué lista de stopwords personalizada? (25 puntos).

## Ejercicio 2 (35 puntos)

1. Para el proyecto esbozado en el ejercicio 1, escoja: análisis de sentimientos o análisis de redes. Justifique (10 puntos).
2. Desarrolle el análisis escogido y presente sus resultados mediante visualizaciones. Explique sus visualizaciones (25 puntos).