



## DESAFÍO CLASIFICACIÓN TEXTO

Una empresa de marketing ha contratado a Brain Food para clasificar comentarios extraídos de diversas redes sociales. La empresa desea mejorar su capacidad de análisis de sentimientos. Actualmente, los comentarios se clasifican manualmente, lo que es un proceso lento y propenso a errores. La empresa desea implementar un modelo capaz de procesar grandes cantidades de comentarios en tiempo real y clasificarlos automáticamente, sin embargo, necesitan que este sea rápido, por lo que se nos ha pedido que no construyamos modelos basados en redes neuronales.

La empresa ya ha clasificado manualmente alrededor de 48 comentarios, con 28 emociones diferentes.

Se entregan los siguientes archivos: train.csv, test.csv, emotions.txt. En emotions.txt, la lista es correlativa, i.e., la primera emoción corresponde al número 0 y así sucesivamente.

El objetivo del ejercicio es:

1. Construir un modelo que sea capaz de clasificar las distintas emociones. El modelo queda a criterio de la/el Data Scientist.
2. Describir paso a paso qué se está haciendo y por qué.
3. Entregar un análisis de las diferentes métricas utilizadas (accuracy, F1-score, entre otras).
4. En base a la clasificación de los comentarios, construir un indicador o KPI capaz de medir el nivel de satisfacción global, explicar significado y justificar la selección.

Luego del desarrollo, necesitamos que respondas lo siguiente:

1. ¿De qué manera se puede complementar la solución? Pensar en propuestas para el cliente.
2. ¿Cómo se podría simplificar la tarea?
3. ¿Cuáles pueden ser las limitaciones, riesgos, sesgos de los modelos al implementar este tipo de soluciones?
4. ¿Qué otras cosas hay que considerar al momento de implementar un proyecto como este?

Entregables del desafío:

1. Clasificación final del dataset "text\_classification\_test.csv", para poder comparar, en formato csv. Que incluya la columna de clasificación con la etiqueta de la emoción.
2. El desarrollo, resultados y análisis deben ser entregados en un Jupyter Notebook autocontenido.
3. Este debe contener las preguntas especificadas arriba

Tendrás un plazo de 1 semana para entregar el caso.