

# Proyecto 5: Procesamiento del Lenguaje Natural

## Instrucciones:

1. El proyecto debe codificarse en Python y debe ser desarrollado en grupo
2. El grupo debe presentar en Google Colab dos aspectos: (1) documentación de la implementación detallada del Algoritmo A Priori para detección de reglas entre tokens y (2) aplicación de bag-of-words con 4 algoritmos baselines de aprendizaje supervisado para clasificación de textos.
3. Aplicar las métricas de desempeño vistas en clase: F1, precisión, recall y accuracy para aprendizaje supervisado.
4. La rúbrica de evaluación del proyecto puede ser consultada en Canvas: <https://drive.google.com/file/d/1is10V6Hh3u2Z7b0J3fJi-13EPHe5HMxy/view?usp=sharing>
5. La fecha de entrega es vía Canvas a las 2pm del jueves 10 de Diciembre del 2020 hora UTC/GMT -5 horas

## 1. Respetto al Proyecto en Python

Tomar en cuenta las siguientes secciones para implementar el proyecto:

**Dataset:** el dataset a utilizar en el proyecto está aquí.

**Algoritmo a Apriori:** El objetivo es implementar el algoritmo a priori para la obtención de reglas en términos de tokens de los textos analizados. Emplear al menos 10 thresholds distintos para la obtención de dichas reglas.

**Aprendizaje Supervisado:** se debe emplear al menos 4 algoritmos baselines para clasificación de textos basándose en la representación bag-of-words.

**Rúbrica:** todos los aspectos solicitados en la rúbrica deben reflejarse en el Google Colab. Es importante mantener una documentación organizada y con un nivel de detalle apropiado.

## 2. Respetto al Proyecto General

**Cumplimiento del informe y proyecto en Python:** el no cumplimiento de la documentación dentro del Google Colab del presente documento afectan a la calificación en la rúbrica y viceversa. Por lo tanto, el grupo debe estar comprometido a leer tanto el enunciado como la rúbrica del proyecto