

# Proyecto – Procesamiento y Clasificación de Datos

## Tarea 1. Definición del proyecto

Nombre: Paulina Aldape Bretado

---

### 1. Tema

Extracción de opiniones de texto para la aplicación de modelos de análisis de sentimientos y su relación con indicadores tradicionales económicos.

### 2. Objetivo principal

Utilizar información en formato de texto proveniente de cuentas de twitter/columnas de opinión de especialistas que brindan su sentir en materia económica para estimar índices de sentimiento sobre sus opiniones y compararlos con indicadores tradicionales, publicados por organismos oficiales, sobre los que dan su opinión.

- **Objetivos secundarios**

Utilizar dos metodologías de análisis de sentimientos, basado en reglas y aprendizaje automático para ver con cual se obtiene un mejor resultado de predicción de clasificación.

### 3. Opciones de bases de datos a utilizar

- a) API de twitter: Obtener la información proveniente de tweets realizados por los especialistas (Pendiente acceso a Twitter Developer).
- b) Columnas de opinión: Obtener la información de periódicos en línea donde los especialistas dan su opinión.

### 4. Justificación del uso de los datos

Las redes sociales son una fuente moderna masiva para detectar las emociones y sentimientos de las personas. Estudios recientes han demostrado que con Twitter es posible obtener información de las personas a partir de sus perfiles.

Por ser información en formato de texto, poco sistematizada, y sobre la cual las metodologías tradicionales de análisis de datos estructurados no podrían aplicarse, se impulsa la adopción de técnicas capaces de aprovechar esta información, siendo una de ellas, el análisis de sentimiento.

La relevancia de utilizar este tipo de información reside en que los índices de sentimiento se pueden generar tan pronto la persona da su perspectiva u opinión adelantada, mientras que algunos indicadores tradicionales que se publican normalmente suelen traer cierto rezago. Así, este ejercicio permite evaluar la capacidad de los primeros para complementar la información provista por los indicadores que se publican con rezago, o bien, reforzar la señal que deriva de indicadores que se publican sin retraso.

## 5. Literatura

Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine learning-based sentiment analysis for twitter accounts. *Mathematical and Computational Applications*, 23(1), 11.

- La aportación de este trabajo incluye el análisis de los sentimientos electorales recogidos de los perfiles de Twitter. Además, este artículo presenta la validación de los resultados obtenidos de cada analizador basados en reglas con clasificadores de aprendizaje automático.

Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2022). Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods. *Cognitive computation*, 14(1), 372-387.

- Este estudio se basa en el análisis del sentimiento financiero de cuentas influyentes de Twitter y su relación con el comportamiento de importantes índices financieros.

Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, 55(2016), 16-24.

- En este estudio, se demuestra la aplicación del análisis de sentimientos usando datos de Twitter de las actitudes/percepciones de los clientes en cuestiones de hospitalidad. Las métricas de sentimiento resultantes se utilizan luego para comparar estas empresas entre sí y comparar su desempeño a lo largo del tiempo. Las métricas también se cotejan con las clasificaciones externas de hoteles y casinos (es decir, TripAdvisor) para establecer la validez externa.

Damarta, R., Hidayat, A., & Abdullah, A. S. (2021). The application of k-nearest neighbors classifier for sentiment analysis of PT PLN (Persero) twitter account service quality. In *Journal of Physics: Conference Series* (Vol. 1722, No. 1, p. 012002). IOP Publishing.

- A través de una empresa que provee servicios de electricidad, se usa la información en Twitter para poder saber qué piensan los consumidores sobre los servicios brindados y dar el mejor servicio a la comunidad para satisfacer estas necesidades primarias. En este estudio, los datos recopilados pasarán la etapa de preprocesamiento, utilizando el algoritmo k-Nearest Neighbors para clasificar los datos en clases negativas, neutras o positivas.

Torre, L., González, E., Casillas, R., & Alvarado, J. (2022). Índices de Sentimiento e Indicadores de Actividad Económica en México 2016-2021.

- Este trabajo utiliza, por vez primera, información en formato de texto proveniente de 9,802 entrevistas realizadas, para estimar índices de sentimiento regionales y nacionales. Estos índices son posteriormente asociados con diferentes indicadores de actividad económica.