

---

---

# Clase N°21

## — Análisis de Sentimiento —

---

---

# Motivación

Vimos que existen herramientas para abordar el estudio cuantitativo de textos.

## Clase 20: Procesamiento del lenguaje natural (NPL)

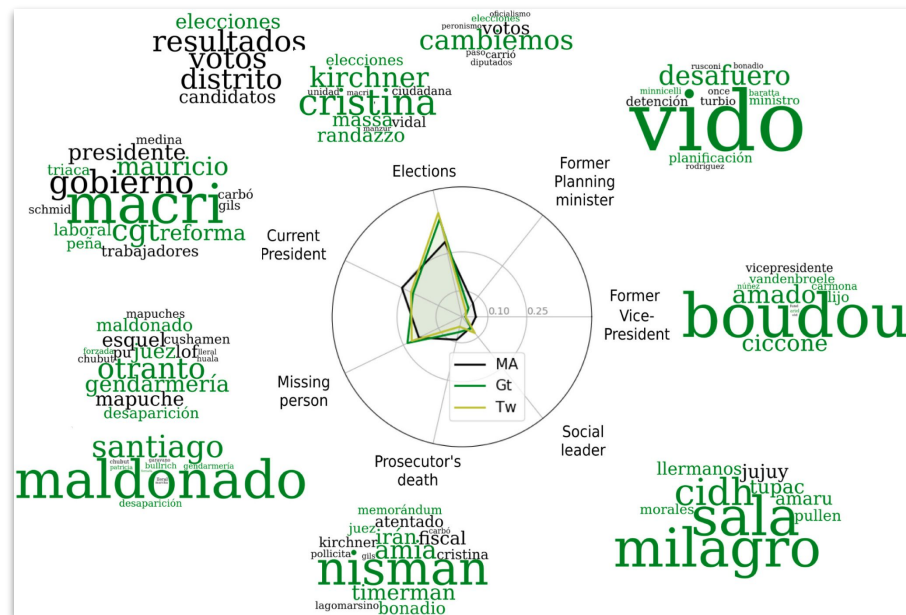
Laboratorio de datos, FCEyN, 4/6/2021



# Motivación

Vimos que existen herramientas para abordar el estudio cuantitativo de textos.

Con varias metodologías, se busca clasificar textos según el tema del que hablan; o bien, vincularlos según métricas de similitud.

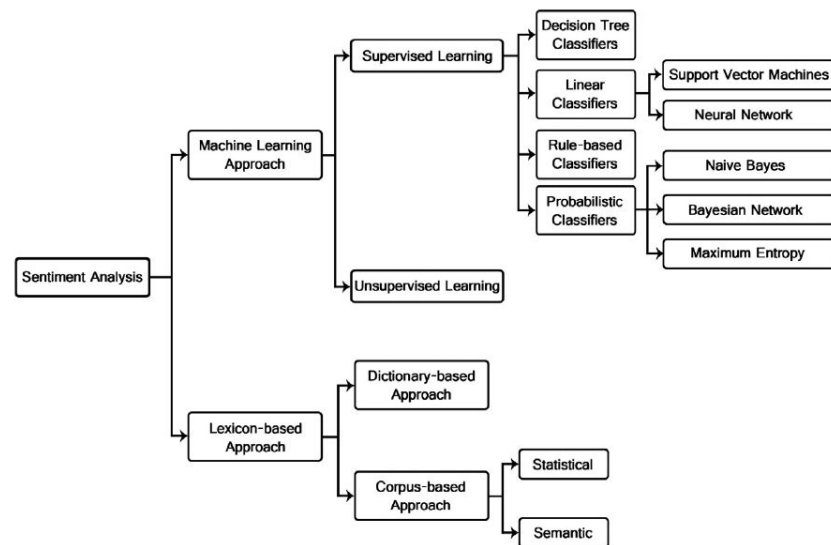


# Motivación

Vimos que existen herramientas para abordar el estudio cuantitativo de textos.

Con varias metodologías, se busca clasificar textos según el tema del que hablan; o bien, vincularlos según métricas de similaridad.

No obstante, existe un universo gigante de técnicas que buscan atacar el problema de la connotación sentimental de los textos, y eso es lo que veremos hoy.



# Generalidades

El Análisis de Sentimiento (*Sentimental Analysis - Opinion Mining*) es un área dentro del PLN que busca estudiar el sentimiento expresado en un texto, a partir de las palabras y expresiones contenidas en el mismo.

# Generalidades

El Análisis de Sentimiento (*Sentimental Analysis - Opinion Mining*) es un área dentro del PLN que busca estudiar el sentimiento expresado en un texto, a partir de las palabras y expresiones contenidas en el mismo.

Suele utilizarse en (y pareciera haber nacido a partir de) contextos de opiniones de consumidores sobre artículos, o de clientes sobre servicios.

# Generalidades

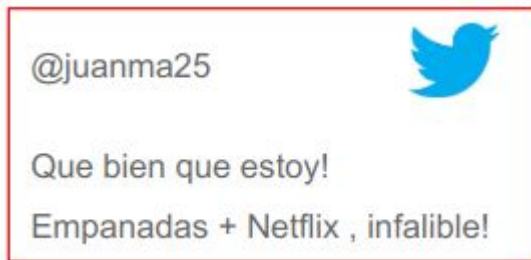
El Análisis de Sentimiento (*Sentimental Analysis - Opinion Mining*) es un área dentro del PLN que busca estudiar el sentimiento expresado en un texto, a partir de las palabras y expresiones contenidas en el mismo.

Suele utilizarse en (y pareciera haber nacido a partir de) contextos de opiniones de consumidores sobre artículos, o de clientes sobre servicios.

Pero también puede aplicarse, entre otras cosas, a estudiar la cobertura mediática sobre determinados temas y posiciones valorativas en torno a las mismas.

# El ejemplo

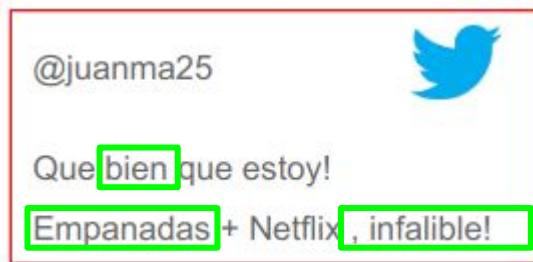
Sea un conjunto de palabras que conocemos con valoración positiva y negativa, podemos determinar de forma frecuentista la valoración de una frase:





# El ejemplo

Sea un conjunto de palabras que conocemos con valoración positiva y negativa, podemos determinar de forma frecuentista la valoración de una frase:



# Generalidades

En la literatura, hay consenso sobre que existen tres grandes enfoques:

- Basados en léxico *-lexicon based-*
- Basados en Aprendizaje Automático *-automatic learning-*
- Enfoques híbridos *-hybrid-*

# Método Basado en Léxico

Utiliza listados de palabras/expresiones cuya connotación positiva/negativa es conocida, para clasificar sentencias y textos.

Existen varios métodos para la construcción de estos léxicos:

- **a mano**

- diccionarios

- corpus

a favor	en contra
<ul style="list-style-type: none"><li>- preciso</li><li>- controlado</li></ul>	<ul style="list-style-type: none"><li>- demanda demasiado tiempo</li></ul>

# Método Basado en Léxico

Utiliza listados de palabras/expresiones cuya connotación positiva/negativa es conocida, para clasificar sentencias y textos.

Existen varios métodos para la construcción de estos léxicos:

- a mano
- **diccionarios**
- corpus

a favor	en contra
<ul style="list-style-type: none"><li>- automático</li><li>- accesible en términos de facilidad de encontrar diccionarios, etc</li></ul>	<ul style="list-style-type: none"><li>- no puede distinguir connotaciones que varíen de un contexto a otro</li></ul>

# Método Basado en Léxico

Utiliza listados de palabras/expresiones cuya connotación positiva/negativa es conocida, para clasificar sentencias y textos.

Existen varios métodos para la construcción de estos léxicos:

- a mano
- diccionarios
- **corpus**

a favor	en contra
<ul style="list-style-type: none"><li>- automático</li><li>- flexible a los cambios de contexto según los textos que le pasemos</li></ul>	<ul style="list-style-type: none"><li>- ausencia de determinadas palabras en los pasos de entrenamiento</li></ul>

# Método Basado en Aprendizaje Automático

Se usan técnicas de ML (algunas ya vistas durante la cursada) para clasificar textos.

En definitiva, el problema de clasificación termina siendo:

- Existe un conjunto de textos, cada uno asignado a una clase (positivo, negativo, estrellas, etc)
- Se entrena el modelo de forma tal de relacionar las features relevantes a las clases
- Se utiliza el modelo ya entrenado para predecir el sentimiento de nuevos textos

# Método Basado en Aprendizaje Automático - NB

Dentro del universo de métodos de clasificación, uno de los utilizados es el de Naïve-Bayes

La idea es computar la probabilidad de que, dadas ciertas features, se obtenga cierta clase, basándonos en las probabilidades inversas obtenidas del entrenamiento.

$$P(label|features) = \frac{P(label) \cdot P(features|label)}{P(features)} \quad \text{Bayes}$$

$$\begin{array}{l} \text{Naïve} \\ \text{independencia de las f} \end{array} = \frac{P(label) \cdot P(f_1|label) \cdot P(f_2|label) \cdot \dots \cdot P(f_m|label)}{P(features)}$$

## ¿Hay más?

No sólo de sentimientos valorativos se trata, sino que aparecen campos como el análisis de emociones.

Además, sobre la construcción listados de palabras/sentencias suele haber mucho trabajo.

Así también como lo que refiere a la utilización de ciertos contextos bastante exhaustivos para trabajar en otros, ej: Wikipedia ---> Twitter



# Problemas abiertos

Todavía quedan misterios sin resolver:

- El de las bases de datos
- El del idioma
- El del procesamiento de lenguaje natural como preprocesamiento para encarar el análisis de sentimiento

# Métodos de la notebook

Si bien existen herramientas gratuitas e incluso online para análisis de sentimiento en inglés, para el español no parece haber tanta suerte.

Acá dos librerías que vienen a aportar en este sentido:

- <https://github.com/FernanOrtega/SentiLeak> (lexicon based)
- <https://github.com/sentiment-analysis-spanish/sentiment-spanish> (ml based)

# Bibliografía

- MEDHAT, Walaa; HASSAN, Ahmed; KORASHY, Hoda. **Sentiment analysis algorithms and applications: A survey**. *Ain Shams engineering journal*, 2014, vol. 5, no 4, p. 1093-1113.
- ZHANG, Lei; WANG, Shuai; LIU, Bing. **Deep learning for sentiment analysis: A survey**. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2018, vol. 8, no 4, p. e1253.