



Instituto Tecnológico  
de Buenos Aires

VIOLETA SAGUIER

**TP 1**

**KYA - HERRAMIENTA PARA YOUTUBE**

82.18 Procesamiento del lenguaje natural

03/10/2022

# AGENDA

**01** **Introducción**  
Caso de negocio, objetivo y desafío del trabajo.

**02** **Idea sobre el modelo**  
Elección del corpus y tratamiento posterior.

**03** **Normalización y ejecución**  
Presentación del modelo para la implementación

**04** **Resultados**  
Aplicación del modelo y usos de negocio.

# 01

## CASO DE NEGOCIO

Herramienta para que los creadores de contenido puedan conocer a su audiencia de forma automatizada.

## OBJETIVO

Brindar información clara y útil para los clientes respectivos.

## DESAFÍO

¿Qué busco conocer? ¿Cómo conseguirlo?

# 01

## Introducción al corpus

- ❖ Comentarios en video de YouTube.
- ❖ Scrapeo con App Script Google.
- ❖ Carga en Google Sheets:
  - Límite máximo de google sheets: No Row Limit (10 million cells or 18,278 columns)
  - Vs. máximo en Excel: 1.048.576 registros.
- ❖ Caso de análisis:



## 02 Idea sobre el modelo

Elección del corpus y tratamiento posterior.

- ❖ Uso de VADER
  - Aclaración por emojis

It can efficiently handle vocabularies, abbreviations, capitalizations, repeated punctuations, emoticons (😭, 😊, 😭, etc.), etc. usually adopted on social media platforms to express one's sentiment, which makes it a great fit for social media sentiment text analysis.



- ❖ Palabras más repetidas - tokenización y nube de palabras.

# 03 Normalización y ejecución

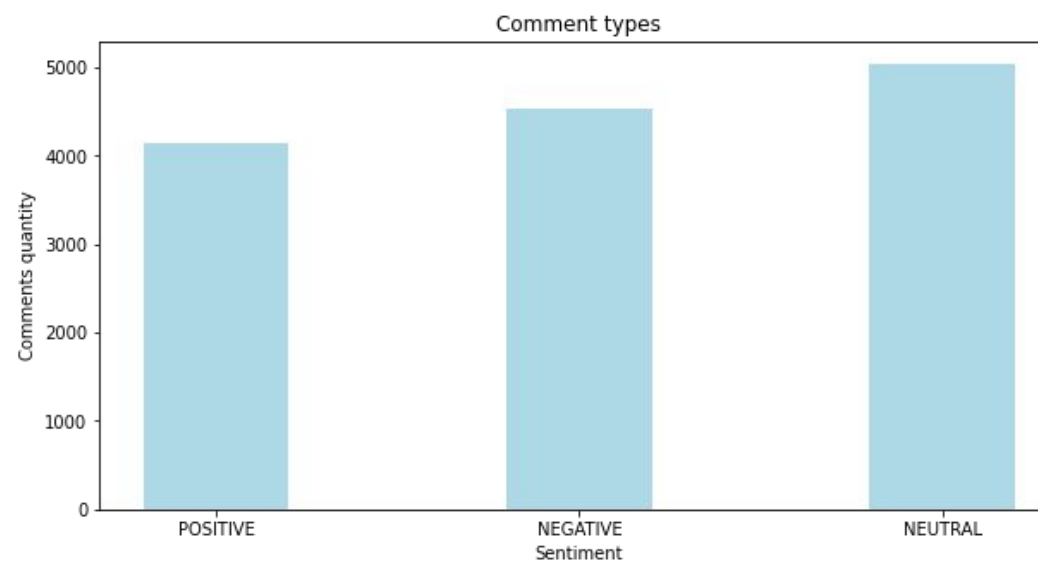
- ❖ Normalización de caracteres que distraen el análisis.
  - Limpieza de signos de exclamación.
  - Limpieza de Emojis.
  - Transformación a minúscula.
  - Exclusión de caracteres que no sean letras.
  - Limpieza de links entre los comentarios.
- ❖ Tratamiento en Python:
  - Análisis de sentimiento.

# CREACIÓN DEL MODELO

- ❖ Uso de VADER: sentimientos positivos, negativos y neutros

Comment	vader_prediction
Bro focus on your body bro your dying	neutral
😎👍 Ur making me hungry 😊	neutral
do ur thing its ur life :)	positive
You cant even walk no more thats sad bro  And y'all supporting this	positive
stop doing this, your going to eat yourself to death, your just doing for 🇲🇪 only One pack of olives with cheese and chips you koen you can those pringles on amazon	negative

- ❖ Bar chart informativo:



# CREACIÓN DEL MODELO

- ❖ Palabras más repetidas con tokenización y WordCloud

eat	2802
stop	1854
healthy	1520
less	1448





## Implementaciones a futuro

- ❖ Trabajar todo en python (no depender de google sheets)
- ❖ Interpretación de Emojis
- ❖ Análisis en tiempo real (live straming)
- ❖ Análisis de múltiples sentimientos.
- ❖ Personalización de los resultados: métricas custom-made.
  - Entrenamiento para generación de comentarios (bots).

**GRACIAS.**

**Espacio de  
consulta.**