

# INTRODUCCIÓN A LA CIENCIA DE DATOS

## RETO 2

Seleccionar uno de los siguientes wrappers de python para trabajar con el API de twitter (<https://developer.twitter.com>).

- **Tweepy**  
<https://github.com/tweepy/tweepy>  
<https://www.tweepy.org/>
- **Python Twitter Tools**  
<https://github.com/sixohsix/twitter>  
<https://mike.verdone.ca/twitter/>
- **Python-twitter**  
<https://github.com/bear/python-twitter>  
<https://python-twitter.readthedocs.io/en/latest/>
- **Twython**  
<https://github.com/ryanmcgrath/twython>  
<https://twython.readthedocs.io/en/latest/>

Utilice uno o varios paquetes de python que permiten hacer NLP, como por ejemplo:

- **NLTK**  
<https://www.nltk.org/>
- **TextBlob**  
<https://textblob.readthedocs.io/en/dev/index.html>
- **Spacy**  
<https://spacy.io/>

Utilizando el wrapper y el paquete seleccionado haga una análisis NLP de tweets relacionados a los resultados de las elecciones.

1. Desarrolle un pre-procesamiento que involucre tareas típicas, tales como:

- Tokenizar
- Eliminar stop-words (artículos, pronombres, preposiciones, etc)
- Convertir a minúscula
- Eliminar acentos
- Eliminar caracteres especiales
- Tratar con hashtags y otra simbología propia de Twitter
- Procesar negaciones
- Tratar URLs
- Etc...

2. Desarrolle una transformación que involucre tareas típicas, tales como:

- Lematizar
- Stemming
- Etc...

3. Minería de datos e Interpretación

- Estadística descriptiva
- Análisis de sentimientos