

# Tweetments

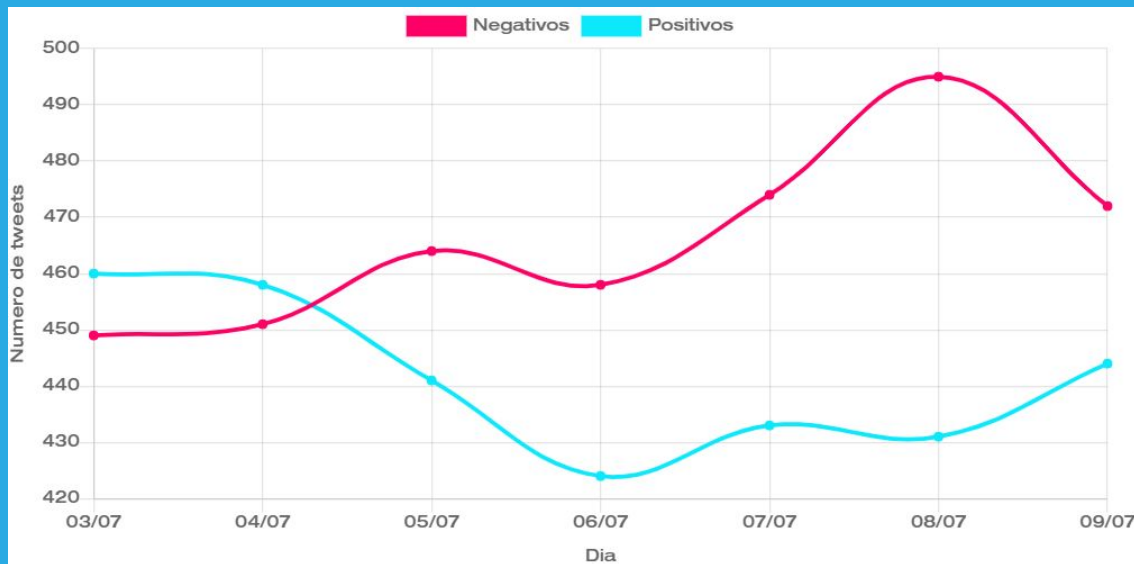
David Andrés Martín-Pozuelo

Pablo Tirado Hernández

Daniel Ortega Conesa

# Objetivos

Partiendo de un conjunto de tweets, pertenecientes a una localización y temática específica, determinar si cada uno de los tweet es de carácter positivo o negativo representando el conjunto en una gráfica.



# Los datos

La fuente de datos que hemos usado para entrenar al modelo ha sido proporcionada por la Sociedad Española de Procesamiento del Lenguaje Natural (SEPLN)



Los datos obtenidos de esta sociedad provenían en formato xml, el cual hemos adaptado para que nuestro modelo lo comprenda, el resultado ha sido un total de 45.745 tweets

	content	polarity	polarity_bin
	Saludos a #KoldoMeabe y otros tantos padres que ahora luchan por la custodia compartida y la igualdad	P	1
	Buenas noches! Descansar;-)	P	1
	Hicimos los deberes, en la epoca de crecimiento reducimos el déficit para en tiempos de crisis no hacer ni un sólo recorte #debateRTVA	N	0
	Buenos días ;:))	P	1
	Borja Sarasola, inauguración de la nueva sede del PP de Brunete <a href="http://t.co/GTxPW3XW">http://t.co/GTxPW3XW</a>	P	1

# ¿Cómo se ha realizado el entrenamiento?

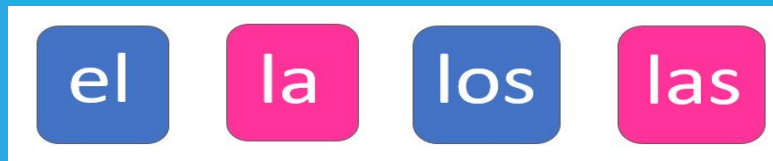
Partimos de la frase: "como siempre, mis mejores deseos para este día! ;-)"

Eliminamos los caracteres que no son palabras, obteniendo como resultado:  
"como siempre mis mejores deseos para este día"

? ! ? !  
" ; : !

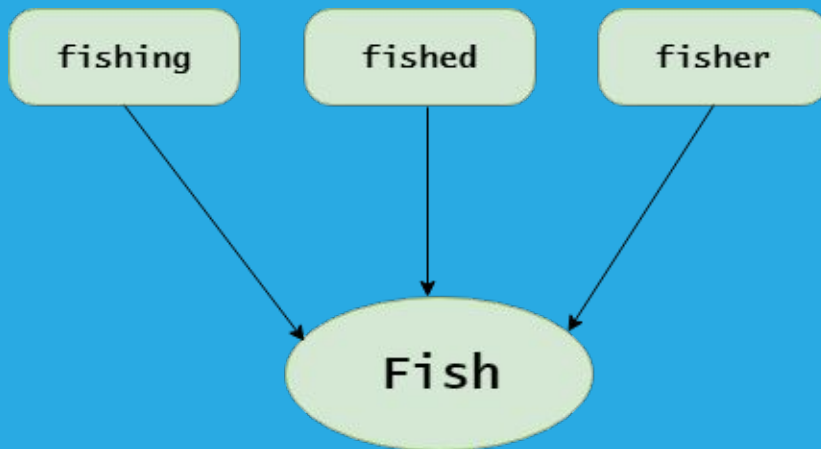
# ¿Cómo se ha realizado el entrenamiento?

Eliminamos las "Palabras vacías", estas son las palabras sin significado como artículos, pronombres, preposiciones, etc..., obteniendo como resultado: "siempre mejores deseos día"



# ¿Cómo se ha realizado el entrenamiento?

Reducimos cada palabra a su raíz: "siempre mejor deseo día"



# ¿Cómo se ha realizado el entrenamiento?

- Repetimos el proceso anterior para cada frase obtenida.

Al finalizar, tendremos un conjunto de frases cuyo contenido es totalmente significativo

- Entrenamos nuestro modelo con este conjunto de frases y creamos una bolsa de palabras, la bolsa de palabras es la unión de todas las frases junto con la cantidad de veces que aparece cada palabra.

	it	is	puppy	cat	pen	a	this
it is a puppy	1	1	1	0	0	1	0
it is a kitten	1	1	0	0	0	1	0
it is a cat	1	1	0	1	0	1	0
that is a dog and this is a pen	0	2	0	0	1	2	1
it is a matrix	1	1	0	0	0	1	0

# ¿Cómo funciona la predicción?

Finalmente, a la hora de realizar una predicción utilizamos un algoritmo support-vector machine de tipo clasificador que utilizando la bolsa de palabras predecirá si una frase introducida es positiva o negativa





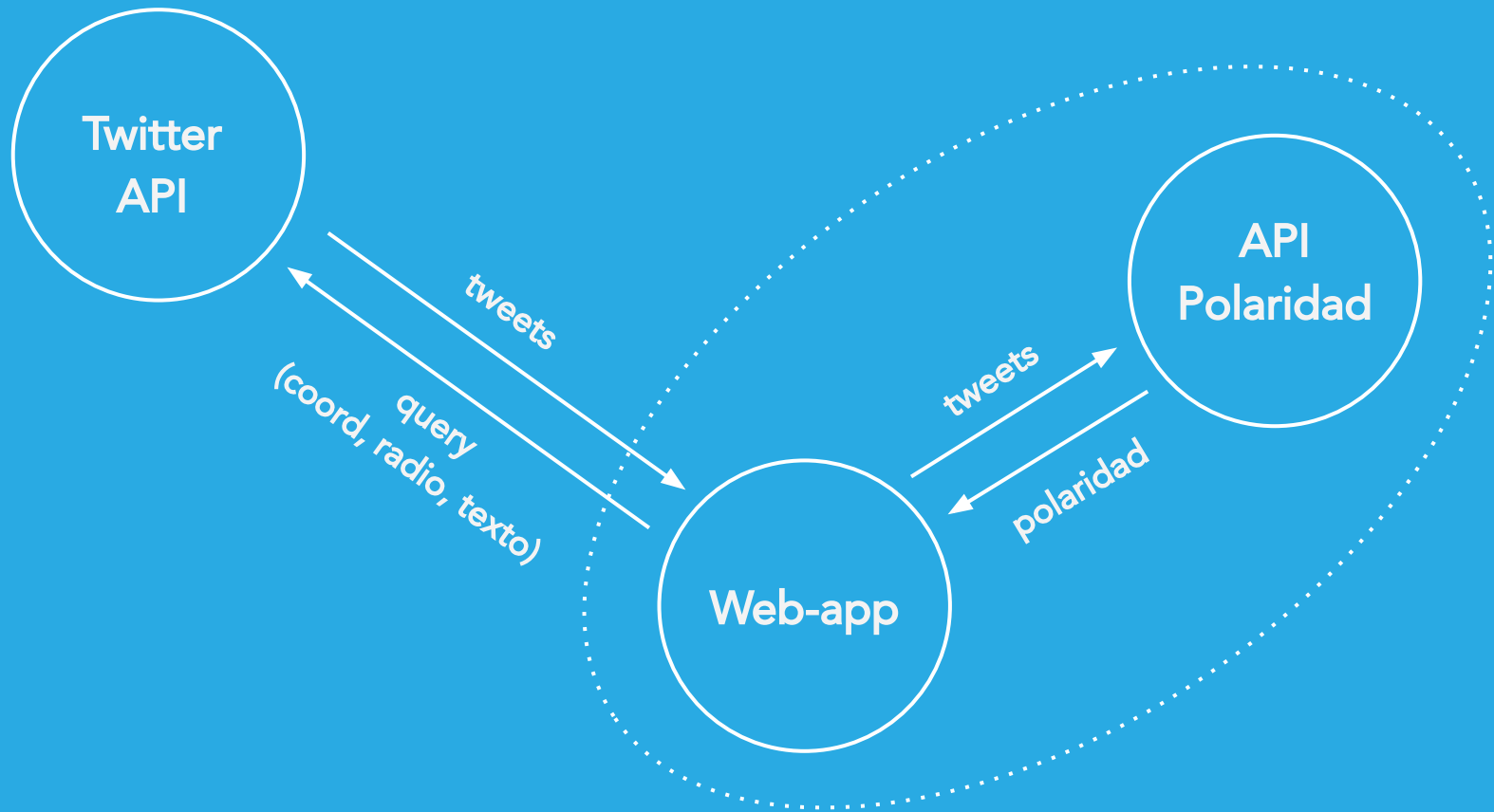
# Web-app con Twitter API

Este algoritmo se integra con una interfaz web que permite seleccionar por el usuario los parámetros para realizar la búsqueda de tweets en la API de Twitter:

- Punto geográfico
- Radio de búsqueda
- Término/s de búsqueda

Con estos parámetros se hace una búsqueda a través de la API de Twitter, cuyos resultados son procesados por el algoritmo para valorar su polaridad y los datos resultantes representados en una gráfica.

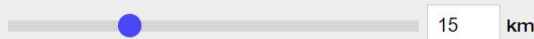
# Web-app con Twitter API



# Web-app con Twitter API

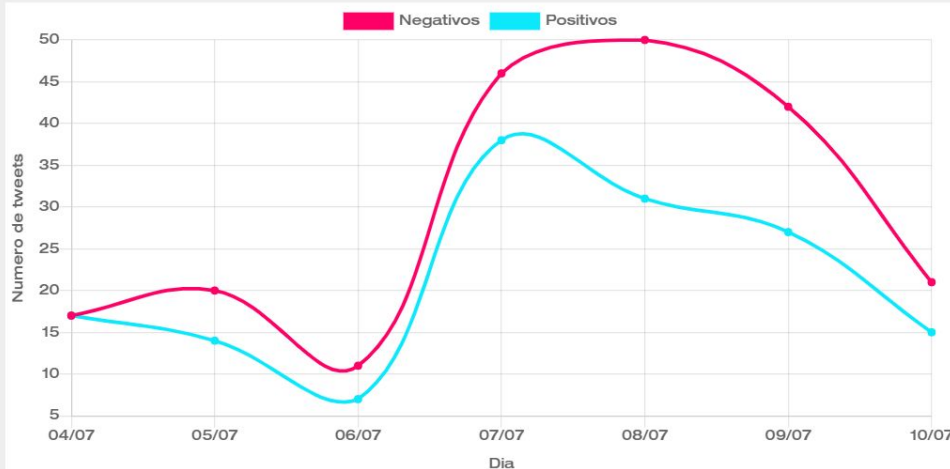


RADIO DE BÚSQUEDA (ORIGEN DE TWEETS):



TÉRMINOS DE BÚSQUEDA EN TWITTER:

BUSCAR



175 tweets procesados.

01-07-2019 15:43:44 |

RT @J\_Warclimb: Hoy en el telediario de Telecinco estaban muy contentos porque por el calor ha subido el turismo. En los grupos de Whatsap...

01-07-2019 16:00:39 | Alicante, España

#Viajes | Cinco impresionantes destinos para huir del turismo de masas en vacaciones  
<https://t.co/ojb59UCet6> <https://t.co/WRJDbW4EZI>