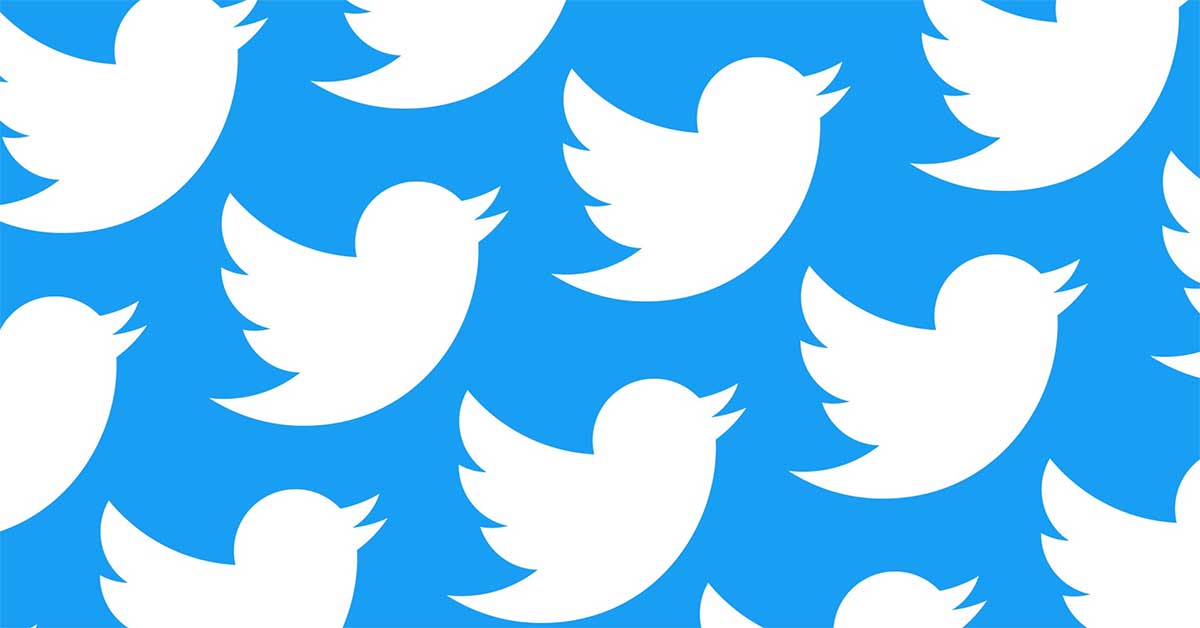
TFM

Política



Alejandro Vaqueiro Vaqueiro

Ester Sánchez Sánchez

Inês Bolaños Pais

Silvia Lendinez Fernández

**PROPUESTA DE VALOR**

**Análisis de idea**

Vamos a analizar las dimensiones, influencia y conexiones entre los usuarios de Twitter con influencia política en España. Representando la distribución política en un gráfico que muestre un panorama fidedigno de la realidad existente en Twitter.  El siguiente paso será identificar el sesgo entre la red social y los resultados electorales ocurridos en el pasado y en el futuro próximo. Prestando especial atención a las futuras elecciones generales del 28 de abril.

**Estrategia**

Aportamos una visión mucho más objetiva de la representación existente en la red social Twitter. El primer objetivo es de carácter descriptivo de la red social, el segundo de contraste de la validez de la muestra obtenida con la realidad existente en la población y el tercero medir la influencia que tienen en la votación final las encuestas previas que se muestra a la población.

Nuestra hipótesis de inicio es que los datos obtenidos de Twitter no tienen representatividad (no son fieles a la realidad). Creemos que podemos obtener una visión más representativa de la realidad con el estudio que vamos a realizar.

Nos centraremos en conseguir datos sobre las relaciones, los seguidores etc. de los usuarios con influencia política.

**Análisis técnico y desarrollo de la solución**

Utilizaremos la API de Twitter para importar los datos de seguidores, interacciones y hashtags de temas políticos en España.

Crearemos una base de datos en SQL Server con todos los datos necesarios obtenidos de la API.

Más tarde, a través de la utilización de Python obtendremos los datos, los limpiaremos y realizaremos el análisis y la representación de éstos.

Vamos a realizar un análisis descriptivo, uno de clustering por comunidades y un análisis de grafos en el que representaremos las relaciones entre los usuarios ligados a la política española.

**¿A quién va dirigido?**

A medios de comunicación social, partidos políticos, periodistas, influencers políticos y no políticos y cualquier persona interesada en tendencias políticas. Cuantificar el sesgo existente entre las redes sociales (twitter) y los resultados finales de las elecciones para poder tener una visión más aproximada de lo que va a pasar en las elecciones. Tener cuantificado el sesgo les permite a los partidos políticos y a los periodistas aproximarse de una forma más real al resultado, los primeros podrían corregir acciones de la campaña y los segundos entender mejor el contexto a la hora de publicar sus artículos.

**REUNIONES DE TUTORÍA**

***1 febrero***

**Trabajo realizado:**

* Cambio de tema e identificación del nuevo tema de forma general
* Identificación de las herramientas necesarias para extracción de los datos (API de Twitter)

**Próximos pasos:**

* Entender la API de twitter y sus limitaciones
* Hacer la propuesta de valor
* Definir el tipo de muestra que vamos a usar

***26 febrero***

**Trabajo realizado:**

* Extracción de muestra de datos de la API de Twitter (*Prueba Importar tweets.ipynb*) e intento de convertirlos en un dataframe (falta convertir la parte del usuario)
* Definición de las limitaciones de los datos
* Necesidad de definir de los criterios de análisis
* Identificación de una [herramienta](https://www.scrapehero.com/how-to-scrape-historical-search-data-from-twitter/) de web scrapping que permite sacar los datos históricos de Twitter

**Próximos pasos:**

1. Definir el ámbito de análisis y el problema inicial (enviar a Ángel cuando esté)
   1. Identificar las métricas (KPIs) de referencia de twitter y RRSS (benchmark y case studies)
   2. Identificar los principales hechos políticos en España en los últimos 4/5 años (importante que esté por lo menos un ciclo político completo)

De ese período, seleccionar cual vamos a analizar

* 1. Identificar las principales encuestas y sondeos, cuando se realizan e ir sacando datos
  2. Listar todos los partidos y representantes políticos (hacerlo extensivo en una primera etapa) para el **período en análisis**
     + Identificar sus cuentas de Twitter
     + Identificar los usuarios que más comentan esas cuentas políticas o que más los mencionan (ver si es posible)
     + Hacer un contexto evolutivo de # y cuentas
     + Sacar los # más usados
       - De las cuentas oficiales
       - De los usuarios que más comentan
     + Analizar los # más usados por:
       - Período (puede ser mensual, por hecho político – a definir)
       - Tipo de cuenta (ver si hay diferencias entre hastags oficiales y de influencers)
     + Según los # más usados, definir un criterio de análisis
  3. Estudiar análisis de sentimiento

Leer artículos y encontrar case studies para entender mejor el análisis de sentimiento: usos, limitaciones, métodos, herramientas

Github, médium, otras fuentes

* 1. Herramientas de análisis de sentimiento

Buscar principales herramientas (ej: Stanford)

Buscar herramientas de análisis específico de tweets y de política (junto o por separado)

Usar herramientas en inglés, traducir una muestra de tweets con una API (ej: Google) y comparar con herramientas en español

Hacer tests con datasets hasta encontrar la herramienta más adecuada para nuestro análisis

* 1. Estudiar series temporales

Buscar teoría sobre series temporales

Ver si hay ejemplos de series temporales aplicadas a twitter

1. Creación del ETL **(29 marzo!!!)**
   1. Definir el tipo de base de datos que vamos a usar (relacional vs no relacional)
   2. Definir requisitos y necesidades para montar el ETL (podemos hacer una extracción histórica con la herramienta de webscrapping, pero para sacar los datos incrementales, debemos usar la API de Twitter