

The background of the slide is a photograph of a church with a bell tower and a circular window, situated on a grassy hill. The image is faded and serves as a backdrop for the text.

La influencia del lenguaje en la popularidad de las personalidades políticas de un pequeño municipio barcelonés

Evaluación final para pasar al proyecto de *Data science*

Cristina Cosma
Febrero 2024

Objetivo general:

- Analizar el sentimiento en el lenguaje de los políticos para averiguar si éste influye en su popularidad.

Objetivos específicos:

- Descubrir cuál es el sentimiento principal que se desprende de cada texto por persona, partido y fecha
- Observar su evolución en fechas claves y a través de los años
- Construir un conjunto de datos propio desde cero para los fines del proyecto
- Adaptar los datos (textos) para producir un conjunto de datos cuya calidad y cantidad sea suficiente para poderse analizar
- Procesar el lenguaje natural (*NLP*) de los textos importados al conjunto de datos
- Saber aplicar las funciones *tokenizer*, *lemmatizer* y *stemming*
- Analizar las palabras más frecuentes y el sentimiento que revela cada texto en su conjunto
- Comparar estadísticamente y relacionar los datos, en tiempo y en cantidad
- Hacer observaciones y sacar conclusiones relevantes de los datos recogidos, transformados y analizados
- Comparar los hallazgos relativos al lenguaje con la popularidad de estas personas y de las formaciones políticas que representan a nivel municipal

Metodología, procesos y herramientas:

- Los textos serán los originales en catalán o castellano y se guardarán por fecha, por persona y partido, para formar un conjunto de datos temporal mensual sobre varios años.
- Muchas herramientas de NLP funcionan mejor con el idioma inglés, por lo que cada texto se traducirá y se revisará, aparte de mantener el original en una columna aparte.
- Se hará el EDA (estudio descriptivo y analítico del conjunto de datos)
- Se procederá a algunas transformaciones del conjunto de datos inicial
- Se usarán algoritmos para conocer las palabras y conceptos más utilizados en sus comunicaciones con los ciudadanos
- Tras conocer las palabras más usadas por persona y mes, se procederá al análisis de sentimiento para resaltar polaridades (positivo, negativo, neutro, etc.)
- Tras el análisis temporal mensual y anual del lenguaje utilizado por cada uno de las personalidades, se harán comparaciones con los datos de popularidad de dichas personas o partidos, en una línea del tiempo que cubra los últimos años
- La popularidad se entiende por los resultados en las elecciones
- El lenguaje de programación para este trabajo será Python y las librerías de análisis y procesamiento de texto Nltk, Spacy, Stanza, así como Pandas y Matplotlib para tablas y gráficos.

Presentación del conjunto de datos:

Se toman los textos de los regidores y regidoras que representan las principales formaciones políticas y también los de los alcaldes que hubo entre 2012 y 2023 en Sant Boi de Llobregat, un municipio del Área Metropolitana de Barcelona de unos 85 mil habitantes. Los textos se dirigen a los ciudadanos del municipio y son descargados desde su versión en línea del periódico municipal “Viure Sant Boi”. A parte de encontrarse en internet a través de la web del ayuntamiento, dicho informativo es distribuido a principios de cada mes a todos los hogares de Sant Boi de Llobregat a través de correos, en formato papel.

El conjunto está formado por **5 variables** : Fecha, Partido, Persona, Original, Traducción. Hay **5 columnas** (las variables) y **360 filas** (las observaciones).

Definición de las variables y columnas:

Fecha: mes y año de la publicación. Variable de tipo temporal.

Partido: formación política a la que pertenece el orador. Variable de tipo *string*

Persona: nombre del orador. Variable de tipo *string*

Original: texto completo en el idioma original. Variable de tipo *string*.

Traducción: texto completo traducido al inglés, para facilitar algunos procesos de NLP. Variable de tipo *string*.

Fuente de los datos: descargados desde su versión en línea del periódico mensual “Viure Sant Boi”. Conjunto de datos de elaboración propia.