



Informe de estado: Clasificación tipológica de barrios

Procesamiento del Lenguaje Natural

Pablo Vicente · 02/05/2024

Tareas a realizar



- Extracción de datos estructurados



- Extracción de datos no estructurados



- Creación modelo de reglas



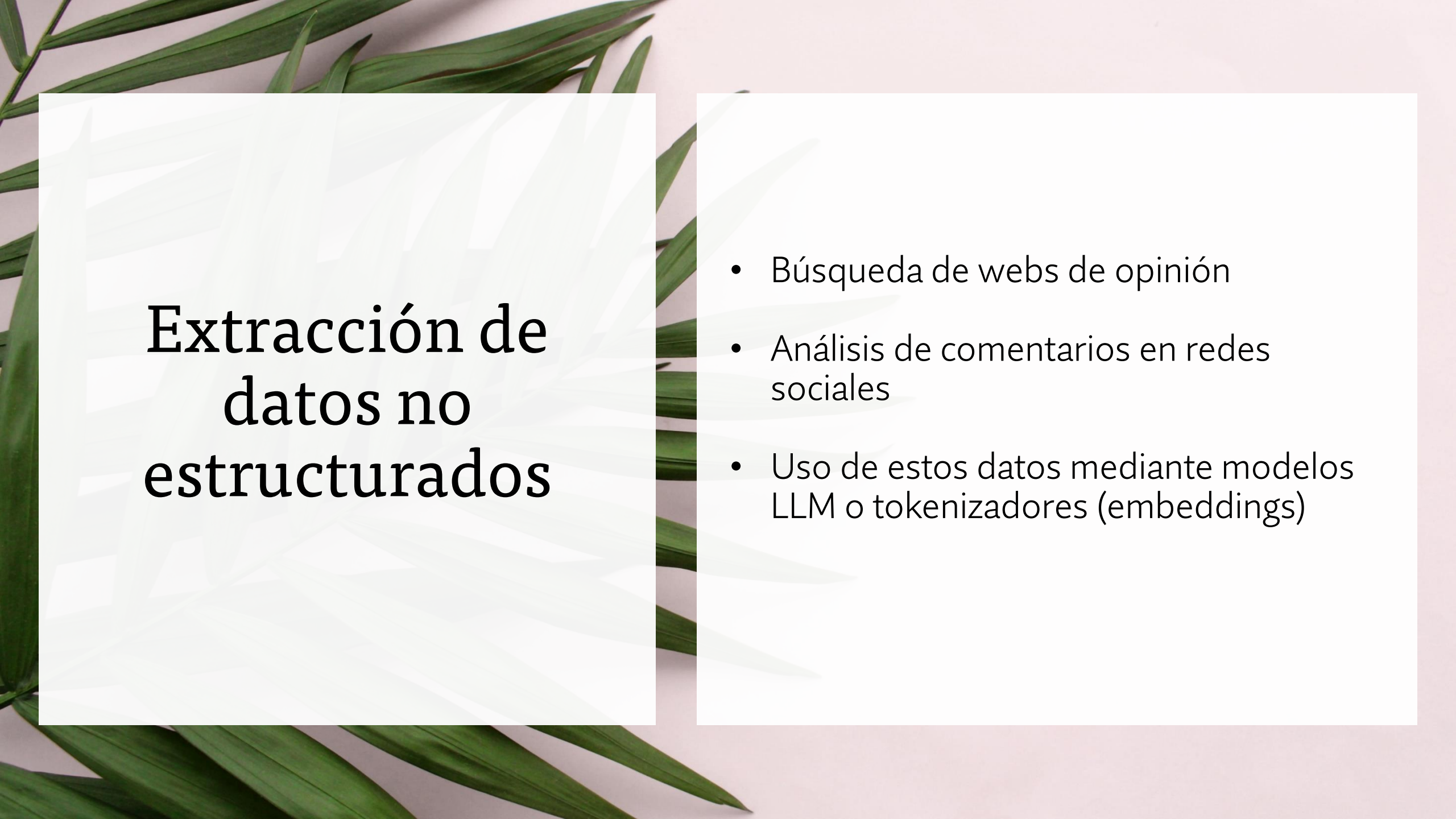
- Uso de LLM para clasificación



- Embeddings

Extracción de datos estructurados

- Oficina de estadística - Ayuntamiento de Valencia
- Por distrito
- Extranjería, convivencia, discriminación, seguridad...
- Creación “barrios referencia” para posterior clustering



Extracción de datos no estructurados

- Búsqueda de webs de opinión
- Análisis de comentarios en redes sociales
- Uso de estos datos mediante modelos LLM o tokenizadores (embeddings)

Modelos para la clasificación de barrios

- LLM: utilizar las descripciones obtenidas de los distintos sitios web para que el modelo relacione dichas descripciones con una de las categorías de barrio que consideremos.
- Reglas: a partir de los datos estructurados, establecer un modelo que según determinados parámetros asigne una o varias etiquetas a un barrio o distrito.
- Embeddings: vectorizar los textos obtenidos a partir del web scrapping para realizar determinadas operaciones, como por ejemplo agrupar textos realizando un cluster o calcular la distancia coseno para observar qué barrios son similares según los textos que se les asocian



Gracias

P. Sherman

Calle Wallaby 42

www.consultatubarrio.com