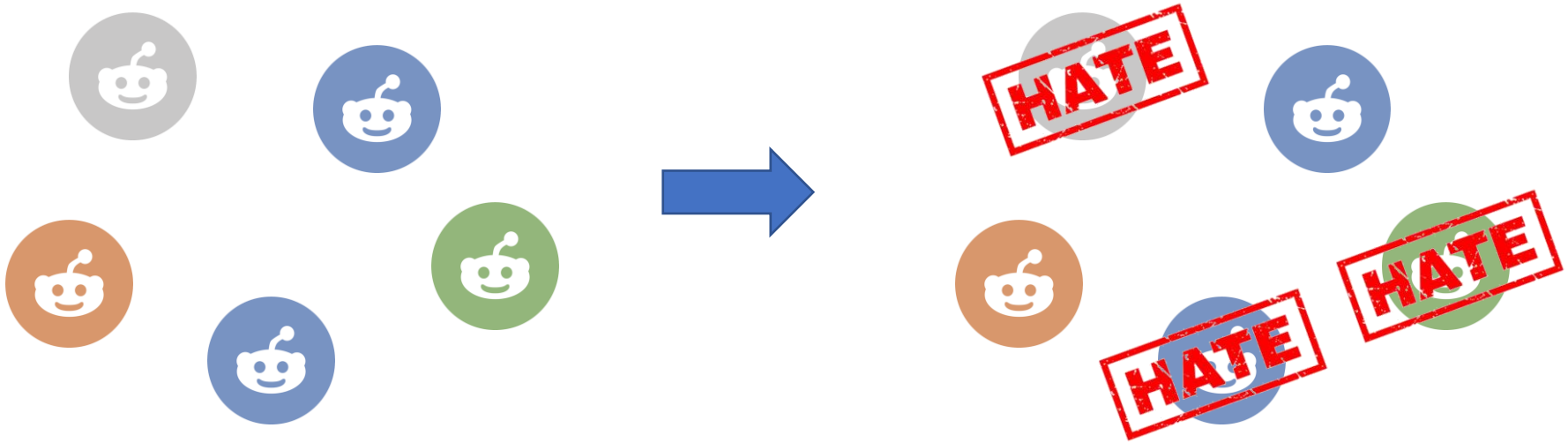
Procesamiento de Lenguaje Natural / TP

Detección de *hate speech* en medios sociales

Medios sociales & hate speech

Cuál es el objetivo?

Detectar aquellos comentarios de Reddit que contienen hate speech distinguiéndolos de aquellos que no!



Medios sociales & hate speech

Y los datos?

- **“A Benchmark Dataset for Learning to Intervene in Online Hate Speech”**
    - El objetivo del trabajo no fue detectar hate, sino diseñar estrategias de intervención con respuestas automáticas a conversaciones quqe tienen hate speech.

- Aproximadamente 5k conversaciones, con 22k comentarios.
    - Actualmente se encuentran disponibles alrededor de 4k con 18k comentarios.

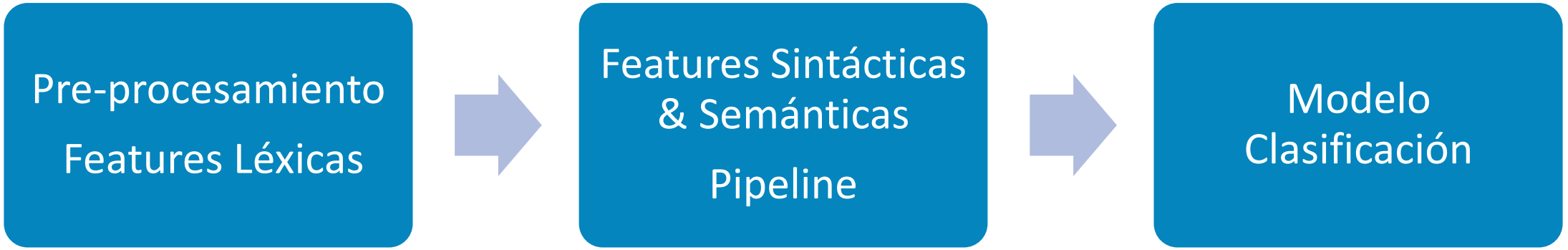
- Comentarios pertenecientes a 10 subreddits:
    - `r/DankMemes`
    - `r/Imgoingtohellforthis`
    - `r/KotakuInAction`
    - `r/MensRights`
    - `r/MetaCanada`
    - `r/MGTOW`
    - `r/PussyPass`
    - `r/PussyPassDenied`
    - `r/The Donald`
    - `r/TumblrInAction`

- Para cada subreddit, recolectarons los 200 posts más “hot”.
- Buscaron posts con keywords de hate.
- Recolectaron las conversaciones.
- No hay repetidos.

https://github.com/jing-qian/A-Benchmark-Dataset-for-Learning-to-Intervene-in-Online-Hate-Speech

Medios Sociales & Desinformación

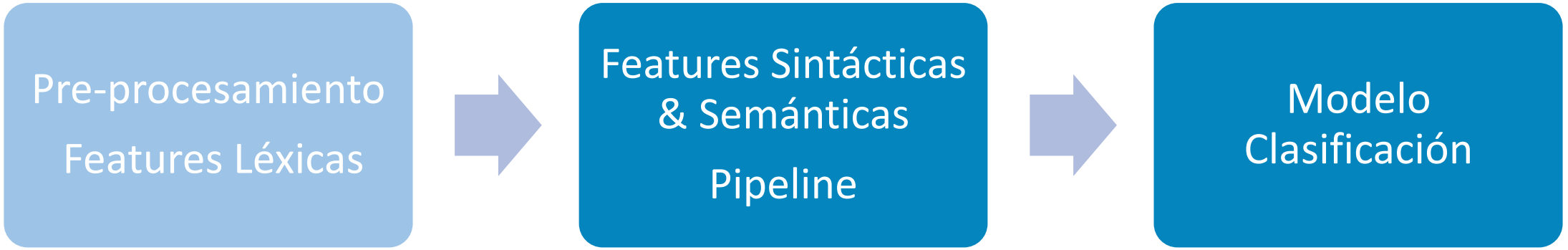
Los prácticos!



Semana 1 | Semana 2 | Semana 3

Medios Sociales & Desinformación

Qué hicieron hasta ahora?

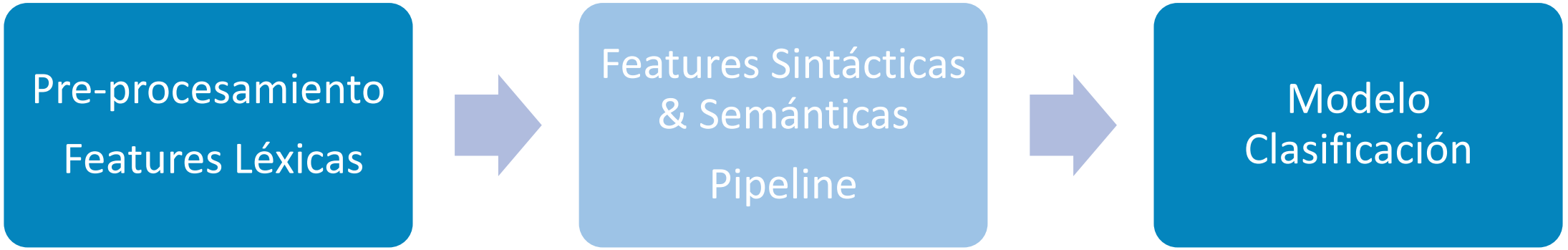


Con este práctico deberían haber:

- Procesado el json con los comentarios y almacenarlos en alguna estructura.
- Decidido si considerar el hilo de conversaciones.
- Elegido algunas características para representar los comentarios.
- Aplicado pasos de pre-procesamiento sobre el texto.
- Pensado en alguna estrategia para representar los comentarios (opcional).
- Calculado estadísticas sobre los comentarios (por ejemplo, palabras más frecuentes).

Medios Sociales & Desinformación

Qué hicieron hasta ahora?



Semana 1 Semana 2 Semana 3

Con este práctico deberían haber integrado:

- Procesamiento de los json con los comentarios y almacenarlos en alguna estructura.
- Elección de las características para representar los comentarios.
- Elección de la representación para los comentarios.

Medios Sociales & Desinformación

Qué tienen que hacer?

Modelo Clasificación

- Entrenar modelos de clasificación de comentarios!
- Elegir una técnica de selección de características para aplicar.
  - Mostrar la relevancia de las características de acuerdo a la técnica seleccionada.
- Elegir al menos dos clasificadores para entrenar.
  - En lo posible, realizar optimización de parámetros.
- Reportar resultados la clasificación con los diferentes clasificadores:
  - Aplicando o no pre-procesamiento.
  - Aplicando o no feature selection (sobre la variante con pre-procesamiento).

Medios Sociales & Desinformación

Qué tienen que entregar?

Modelo Clasificación

- Notebook con:
  - Aplicación de la técnica elegida de feature selection.
  - Entrenamiento de los clasificadores elegidos + optimización de parámetros.
    - Explicar brevemente la elección de los clasificadores.
  - Reporte de los resultados.
    - Elegir las métricas que consideren adecuadas. Explicar brevemente la elección.
    - Incluir algún gráfico comparativo.

Medios Sociales & Desinformación

Qué tienen que entregar?

Debe incluir:

- Carga de datos
- Pre-procesamiento (TP-1)
- Selección de features (TP-1, TP-2)
- Creación de la representación de los comentarios (TP-2)

Modelo Clasificación

- Notebook con:
  - Aplicación de la técnica elegida de feature selection.
  - Entrenamiento de los clasificadores elegidos + optimización de parámetros.
    - Explicar brevemente la elección de los clasificadores.
  - Reporte de los resultados.
    - Elegir las métricas que consideren adecuadas. Explicar brevemente la elección.
    - Incluir algún gráfico comparativo.

Medios Sociales & Desinformación

Qué tienen que entregar?

Fecha de entrega: **2 de Agosto 2020**

Modelo Clasificación

- Notebook con:
  - Aplicación de la técnica elegida de feature selection.
  - Entrenamiento de los clasificadores elegidos + optimización de parámetros.
    - Explicar brevemente la elección de los clasificadores.
  - Reporte de los resultados.
    - Elegir las métricas que consideren adecuadas. Explicar brevemente la elección.
    - Incluir algún gráfico comparativo.

Medios Sociales & Desinformación

Qué tienen que entregar?

Fecha de entrega: **2 de Agosto 2020**

La notebook debe poder ejecutarse sin errores y debe incluir los outputs generados!

Modelo Clasificación

- Notebook con:
  - Aplicación de la técnica elegida de feature selection.
  - Entrenamiento de los clasificadores elegidos + optimización de parámetros.
    - Explicar brevemente la elección de los clasificadores.
  - Reporte de los resultados.
    - Elegir las métricas que consideren adecuadas. Explicar brevemente la elección.
    - Incluir algún gráfico comparativo.

Procesamiento de Lenguaje Natural / TP

Detección de *hate speech* en medios sociales