Sentiment Analysis en review de productos

Materia: Análisis y Curación de datos

Análisis del dataset.

Comunicación de resultados y conclusiones

A partir de lo visto en la teoría de la materia y en los laboratorios trabajados, se busca explotar más las habilidades aprendidas. Para ello diagramar una comunicación en formato textual o la que el alumno considere adecuada, describiendo las reviews contra algún aspecto en particular del dataset provisto.

Identificar aquellas variables más correlacionadas y posibles interpretaciones.

- Distribución de los ratings con respecto a la clase target.

Cuestiones interesante que deberían ser exploradas y respondidas:

- Analizar la consistencia/inconsistencia de los datos.
    - Siempre es bueno tratar de trabajar sin duplicados en un dataset, esto nos evita problemas futuros en el entrenamiento. Verifique si existen datos repetidos y proceda en consecuencia.
    - En caso de presentarse faltantes en las features de interés, que decisión tomaría para su tratamiento. ¿Porqué?. Aplique su decisión al dataset.
    - Muchas veces cuando importamos un dataset pandas infiere que valor podría ser, de no encontrar un valor conocido pone uno por defecto. Validar que los tipos de datos de las features después de importarse correspondan con su valor intrínseco. De no ser así cambie aquellos que pueden resultar de interés para la exploración.
- Explorando los datos.
    - ¿Los rating en este dataset, tiene una distribución análoga respecto al del primer trabajo? ¿Qué conclusiones pueden arribar observando la distribución de ratings de ambos dataset? Presente los resultados.
    - Exploren correlaciones interesante respecto a los ratings (stars) en el dataset. En particular con la feature "**useful**". Elija al menos una más que le parezca relevante.
    - Haga una rápida exploración de valores atípicos (outliers) del conjunto de datos. Realice los gráficos que considere pertinente.

La comunicación debe estar apuntada a un público técnico pero sin conocimiento del tema particular, como por ejemplo, sus compañeros de clase etc.

Se evaluarán los siguientes aspectos:

- El informe debe contener un mensaje claro y presentado de forma concisa.
- Los gráficos deben aplicar los conceptos de percepción visual vistos en clase.
- Se debe describir o estimar la significancia estadística de su trabajo.

Entrega

En lo posible mantener el mismo repositorio en el cual se trabajó en la primer entrega. Para el informe mantenemos la misma idea, generarlo junto con la presentación (Google Drive) y compartir el link.

Presentación

A definirlo entre todos. Estaría bueno tomarnos 30’ minutos, para que puedan exponer los resultados y charlamos un rato sobre ellos. Podemos ver de hacerlo por video llamada o presencial como la primer reunión.