

Semana 4: (10 al 14 de marzo)

Tareas realizadas:

- ☒ Se define el dataset a utilizar durante la experimentación¹. Este dataset contiene un millón de tweets etiquetados en cuatro clases: positive, negative, litigious y uncertainty. No obstante, nos interesa experimentar únicamente con los tweets en español, por lo tanto, durante la experimentación se filtran solo los datos que cumplen con dicho requerimiento.
- ☒ Se realizó experimentación con los siguientes algoritmos: Naive Bayes, LSTM Bidireccional, BERT y Regresión Logística.

Detalle importante:

Durante la experimentación con BERT notamos que la etapa de training llevaba más de 10 horas por época, por lo tanto, decidimos sustituirlo por regresión logística el cual era más rápido.

Para la próxima semana:

- Realizar reporte del análisis con los resultados de la experimentación.
- Corrección del artículo, considerando la utilización de Regresión Logística y la retroalimentación brindada por los docentes con respecto a los presentado en el avance 1.
- Investigar sobre la metodología y redacción de la misma.
- Realizar gráficos útiles que muestran el comportamiento de los modelos ejecutados.

Los códigos utilizados para la experimentación fueron obtenidos de los siguientes enlaces:

Regresión logística: <https://www.kaggle.com/code/mateenmureed/sentiment-analysis-with-logistic-regression>

LSTM Bidireccional: <https://www.kaggle.com/code/rashmimanwani/twitter-sentiment-analysis-bidirectional-lstm>

Naïve Bayes: <https://www.kaggle.com/code/karanseghal13/twitter-sentiment-analysis-using-naive-bayes-nlp>

BERT: <https://www.kaggle.com/code/matyasmacudzinski/sentiment-analysis-using-bert>

Todos los códigos fueron adaptados para funcionar con el dataset seleccionado para la experimentación. Igualmente, puede accederse a los notebooks con los modelos implementados en la carpeta "Notebooks" ubicada en nuestro repositorio, los notebooks se encuentran ejecutados y con los resultados con respecto a su accuracy.

¹ Enlace al dataset:

<https://www.kaggle.com/datasets/tariqsays/sentiment-dataset-with-1-million-tweets>