

Semana 3: (27 al 31 de marzo):

Tareas realizadas:

- ☒ Redacción del documento: introducción y trabajos relacionados.
- ☒ Exploración y análisis del dataset en español¹, dicho dataset posee ejemplos de tweets etiquetados según la polaridad del mismo (positivo, negativo, neutro).
- ☒ Mejoras del avance 1: Redefinición de la problemática y los posibles aportes.
- ☒ Lectura del artículo: "Twitter Sentiment Analysis using Naïve Bayes Classifier with Mutual Information Feature Selection"² by: Maria Arista Ulfa, Budi Irmawati, y Ario Yudo Husodo.

Puntos importantes:

- Se experimentan con modelos de análisis de sentimientos en tweets implementando Naïve Bayes con y sin la selección de características de información mutua (MI).
 - MI permite descartar los features que no aporten información en un texto en cuestión.
 - Se obtiene un accuracy de entre el 96.2% y el 97.9% en el experimento implementado con MI.
 - Los puntos mencionados anteriormente son parafraseados en el documento del avance 1, en el apartado de trabajos relacionados, bajo la referencia [5].
- ☒ Lectura del artículo: "Emotion and sentiment analysis of tweets using BERT"³ by: Andrea Chiorrini, Claudia Diamantini, Alex Mircoli y Domenico Potena.

Puntos importantes:

- Se implementa un modelo basado en la arquitectura BERT, lo cual permite un análisis del contexto de los tweets.
 - Para el análisis de los sentimientos se recurre a un dataset con 1.6 millones de tweets. Los cuales son etiquetados según la polaridad.
 - Se obtiene un accuracy del 92% y un valor de F1 correspondiente al 92%.
 - Los autores plantean realizar una mejora del rendimiento del modelo, optimizando la cantidad idónea de capas y neuronas.
 - Los puntos mencionados anteriormente son parafraseados en el documento del avance 1, en el apartado de trabajos relacionados, bajo la referencia [6].
- Tareas para la próxima semana:
 - Experimentación de los algoritmos utilizando el dataset en español.
 - Documentación de los resultados de los experimentos realizados.

¹ Enlace al dataset: <https://www.kaggle.com/datasets/lexandermorales/acoso-ciberntico>

² Enlace al artículo: <https://jcosine.if.unram.ac.id/index.php/jcosine/article/view/120/26>

³ Enlace al artículo: https://www.researchgate.net/publication/350591267_Emotion_and_sentiment_analysis_of_tweets_using_BERT

- (Si es posible) realizar integración entre diferentes datasets para tener más filas.