

Nombre(s): _____

Matrícula(s): _____

En la actividad de esta semana trabajarás en equipos mediante el uso de Transformers.

HuggingFace (HF) es una plataforma de una comunidad de IA que promueve las contribuciones de código abierto, en particular para compartir modelos de aprendizaje profundo, entre ellos los Transformers o modelos basados en el mecanismo de Attention: <https://huggingface.co/>

La plataforma de HF está muy bien documentada, es de acceso abierto, pero deberás generar una cuenta sin costo para su uso. En ocasiones deberás generar una clave para el uso de algunos modelos, pero es solo para el control del uso de dicha plataforma y no generan algún costo para su uso educativo o de investigación, al menos hasta ahora.

En particular en esta actividad estarás usando varios de los modelos Transformer que hacen tareas diversas, tanto en inglés como en español. El objetivo es que te familiarices con dicha plataforma y experimentes posteriormente con ella para las tareas o problemáticas particulares que te interesen.

1. Descarga el archivo **amazonbaby5000.csv** que se encuentra en Canvas. Este archivo está formado de comentarios en inglés sobre productos para bebé adquiridos en la plataforma de Amazon.
2. Realiza una partición de los datos en el porcentaje que consideres adecuado, entrenamiento, validación y prueba.
3. Utiliza un modelo Transformer de HuggingFace para análisis de sentimiento en inglés y lleva a cabo la predicción de los comentarios en los siguientes casos:
 - a. Utilizando un modelo predeterminado de HF de manera directa, es decir, sin ajuste de parámetros del modelo. Indicar el desempeño (accuracy) obtenido.
 - b. Utiliza ahora un ajuste de parámetros (fine-tuning) con el modelo Transformer DistilBERT, para generar un nuevo modelo y clasificación de los comentarios. Es decir, el objetivo es utilizar un modelo Transformer pre-entrenado y adaptarlo (fine-tuning) o ajustarlo a las características de tus datos de entrenamiento. Indicar desempeño (accuracy) obtenido.
 - c. Compara los resultados e incluye tus comentarios.

NOTA: Puedes apoyarte en el siguiente tutorial de la misma comunidad de HF:

<https://huggingface.co/blog/sentiment-analysis-python>

4. Utiliza el modelo Transformer pre-entrenado para diferentes idiomas llamado Helsinki-NLP de HF, para hacer una traducción del Inglés al Español de los 5000 comentarios. Observa algunos de los comentarios traducidos y comenta qué tan bien consideras dichas traducciones. Por el momento es

un simple comentario subjetivo, aunque existen métricas y bases de datos para evaluar el desempeño de un Transformer.

NOTA: Guarda el archivo de los comentarios traducidos para facilitar su uso en las siguientes preguntas de esta actividad.

Existen pocos modelos pre-entrenados con corpus en Español, pero dos de los más conocidos son BETO (<https://github.com/dccuchile/beto>) y MarIA (<https://arxiv.org/abs/2107.07253>). En particular para análisis de sentimiento se puede utilizar alguno de los modelos derivados de BETO:

5. Utiliza el modelo Transformer “pysentimiento”, basado en BETO y este a su vez basado en BERT, para llevar a cabo un análisis de sentimiento (sentiment analysis) y hacer la predicción de los comentarios traducidos al español. Incluye tus comentarios del resultado obtenido.
6. Resume los resultados obtenidos en esta actividad de manera tabular e incluye tus comentarios finales.

NOTA: A partir de esta actividad, no dudes en experimentar con otros modelos Transformer pre-entrenados tanto para traducción al español, como para análisis de sentimiento.