

Traitement Automatique du Langage Naturel

TP 2 Sentiment Analysis



Dataset link : https://huggingface.co/datasets/M-A-D/Mixed-Arabic-Datasets-Repo/viewer/Ara--cardiffnlp--tweet_sentiment_multilingual

Model link

Multilingual-MiniLM-L12-H384 :

<https://huggingface.co/microsoft/Multilingual-MiniLM-L12-H384>

<https://arxiv.org/pdf/2002.10957.pdf>

Arabert : <https://huggingface.co/aubmindlab/bert-base-arabert>

Using the dataset in the link and transformer-based model like MiniLM , **Arabert** build a sentence analysis using the Transformer library from Huggingface and train it in Kaggle GPU.

You can use this material to start :

Simple Training with the Transformers Trainer

<https://www.youtube.com/watch?v=u--UVvH-LIQ>

<https://huggingface.co/docs/transformers/training>