NLP Course

Graded Project Instructions

Fall 2023

About the dataset:

List of tweet texts with emotion labels like *joy*, *sadness*, *fear*, *anger*…

Dataset is split into *train*, *test* and *validation* sets for building the machine learning model. At first, you are given only *train* and *test* sets. The *validation* one will be given in the end of the project for you to check the final performance of your algorithm (to make sure there is no overfitting over the *test* data).

You can work on this project by **group** of **one**, **two** or **three students**. This exercise is **mandatory**, not giving it back is equivalent to getting to lowest grade.

Goal:

- Train different kind of models able to classify each text according to the sentiment mainly present in it
- Compare the results of your different models and try to analyze and explain the differences

Instructions:

Train different classification models relying mainly on

1. A Fully Connected Neural Network (see **Course 2**) **5 points**
2. A Recurrent Neural Network, based on LSTM or GRU (see **Course 3**) **5 points**
3. A fine-tuned Transformer Architecture from a pretrained model that can be found on sites like HuggingFace (see **Course 4**) **5 points**
4. Compare the different models to find the best approach and explained what you have learned from this exercise and how would you proceed with another text classification use case **5 points**

Deadline:

The project is due after **Course 5** (the last one) on **February the 5th**. No late submission will be accepted unless you get an explicit permission from me.

You can work on this project on group of **one**, **two** or **three** students. This exercise is **mandatory**, not giving it back is equivalent to getting the lowest grade.

Sources:

- https://aclanthology.org/D18-1404/
- https://twitter.com/omarsar0?s=20
- https://www.kaggle.com/datasets/praveengovi/