Natural Language Processing assignment 1 and 2

General guidelines:

- The tasks are targeted at groups of three students. Please ensure the load is well divided: every student should contribute.
- Please take advantage of the practical sessions to ask your questions about the tasks.
- A reviewer should be able to understand plots independently; be sure to label axes, a legend for colors, use easily readable font size, etc.
- Please carefully read the instructions and deliverables for each of the assignments.
- Late submissions will incur a 5% penalty per day, up to a maximum of three days, beyond which submissions will be rejected.

- The submission deadline: **23:59, March 2, 2024.**

- Dataset: You can find the dataset on Brightspace: Go to the "Assignment" section and then "Assignment 2".

- For the report (Assignment 2): The report can be submitted using Jupyter Notebook (.ipynb file) alongside your implementation or create your report in your preferred word processor and save it as a PDF. Please include code as .py or .ipynb files as attachments in Brightspace.

- Datasplit: Divide your training data into two sets: 70% for training your model and 30% for validation. Use the provided test set to evaluate your model's performance.

Empathy Detection and Emotion Classification

(The Data and the task are from the WASSA 2022 Shared Task on Empathy Detection and Emotion Classification)

Task Description

You are given the empathic reactions to news stories dataset which contains essays and Batson empathic concern and personal distress scores in reaction to news articles where there is harm to a person, group, or other (for more details about the task see https://aclanthology.org/D18-1507.pdf). Your task is to perform a review of relevant literature and build an **Emotion Classification (EMO) model**, which consists of predicting the emotion at the essay level. The specific details for each of the tasks are given below.

Task 2: Implementing an Emotion Classifier Your main task or objective here is to develop and evaluate an emotion classifier model. Therefore your main tasks here will be:

1. Select a suitable NLP toolkit or framework for building the emotion classifier.
2. Upload the provided dataset for training and evaluating the classifier and perform data preprocessing and feature engineering.
3. Implement an appropriate machine learning or deep learning algorithm for classifying emotions from textual data.
4. Train and evaluate the emotion classifier on the chosen dataset, using appropriate performance metrics. Use your knowledge and skills from the introduction to machine learning and neural network courses.
5. Analyze the results of the model and make recommendations for improving the classifier's performance. Use visualization in the analysis of the results.

The final deliverables from this task will be a working implementation of the emotion classifier mode and should contain the following.

1. The working implementation should contain code for Data Preprocessing, Feature Engineering (TF-IDF, Word2vec), Model Selection, Training and Evaluation.
2. A detailed description of training procedure, and evaluation metrics.
3. A comprehensive analysis of the classifier's performance, including confusion matrices, precision, recall, and F1 scores.
4. Suggestions for improving the classifier's performance, considering factors such as data preprocessing, feature engineering, and model architecture.