

# 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 for Task 1 (Assignment 1) will be the **20th of February, 2024** and For Task 2 (Assignment 2) will be the **29th of February 2024**.

## 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 1:** Review of Relevant Literature on Empathy Detection and Emotion Classification. The main objective of this task is for you to gain an understanding of the current state-of-the-art approaches in empathy detection and emotion classification. Therefore, your main tasks will be to:

1. Conduct a literature review on empathy detection and emotion classification in NLP(review at least 6 papers, 2 per each group member ).
2. Summarize the key theoretical frameworks, methodologies, and evaluation metrics used in each of the papers.
3. Point out the strengths and limitations of existing approaches and discuss the challenges and future directions of research in this area.

The final deliverables from this task will be:

1. A well-structured, easy-to-read, and understandable literature review report, organized into clear sections and subsections.
2. A table summarizing the key features, advantages, and disadvantages of various empathy detection and emotion classification methods.
3. A discussion section that evaluates the reviewed papers and identifies future research direction.

**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. Select a 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 the chosen dataset, 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.