

REI603M - Assignment 4

Please return this assignment online on Canvas. **You need to hand in both your notebooks and your presentation slides (.pptx or .pdf).** Please submit them as a .zip file.

If your project doesn't align with any of the specified categories below, please consult the instructor directly through personal communications for further instructions.

Few-shot Learning

This week, the objective is to establish a strong baseline model using the data that was collected and feature-engineered in your second and third assignments.

The goal is to determine the performance of one or more models with minimal effort.

You have several options for the approach you can take:

- If you are working solely with text data and doing something that pertains to information retrieval, you should explore text embedding models such as the ones available on the [Massive Text Embedding Benchmark](#). You can convert your text into embedding vectors and consequently use those vectors to train models and/or to explore structure in your dataset.
- For image-based data, examine models such as [CLIP](#) or more recent models like [SigLIP](#). The image encoder can be utilized for traditional learning, while the text encoder can be used for zero-shot learning. If you're working with tensor data, you can still use CLIP-style models on each slice of your tensor and combine the vectors through averaging or some other means.

It is important to keep in mind that the quality of data is more crucial than the choice of algorithm. In this exercise, you should not be training complex models such as neural networks, only use them as feature extractors.

The Task

Information Retrieval

If you are doing information retrieval, such as looking up information in hospital pamphlets, segments from a textbook or pub-quiz questions then you should explore different retrieval methods. Here is an outline of what you should do.

1. Start by selecting an appropriate embedding model from MTEB. You can use the opportunity to compare 2-5 models if you have the time.

2. Embed your dataset. If it is very large, you should consider using only a subset of your dataset.
3. Optional. To prepare for using a large dataset, you may want to look into similarity search options such as FAISS. If you want to outsource your similarity search, you may want to consider a service such as Pinecone.
4. Set up 40 or more queries to evaluate your system. Ideally, you should have more than a 100 queries to get a robust performance estimate.
5. Perform retrieval using KNN and SVMs. [This](#) good example by Andrej Karpathy shows how you can compare the two in the setting when you store your embeddings as numpy arrays.
6. Remember to compute relevant performance metrics for your setup.

You should present your approach and the results in your presentation.

Supervised Learning

If you are doing more traditional supervised learning¹ then here are some options you should explore. This is only relevant if you have at least 30 labelled examples for each class (ideally, much more).

- With text and image data, you should explore different foundation models as feature extractors.
 - For image processing, please explore models such as CLIP, SigLIP and EVA-CLIP.
 - For text processing, compare your embedding models of choice from the massive text embedding benchmark².
- Use the embedding models to embed your images as vectors³. Then use the extracted vectors to train a traditional model such as logistic regression. Report the appropriate performance measures for your task and compare the results for different models.
- To evaluate how well your method makes use of your data, you should plot a training curve. This means that you will train on 10%, 20%, . . . , 80% of your data, and test on the remaining 20% (assuming you don't have much training data). You can repeat this experiment using cross-validation in order to obtain a standard deviation estimate of your model's performance. With the resulting data, you should plot a training curve that illustrates how the test performance changes as you increase the amount of training data used. This can help you understand the value of collecting more data.

Summarization and Airway Segmentation

For summarization, you should start training models in the summarization task. Select an appropriate base model (such as [mT5-base](#) or any other capable model) and fine-tune it in the summarization task.

¹Such as mole classification, food allergen classification, severity detection of bug reports or detection of gender bias in text.

²If you have time, you can, in addition, also explore more traditional approaches such as using a bag-of-words representation or word2vec.

³If you have tensor data (such as videos) you should consider concatenating samples of vectors or averaging them in this feature extraction exercise.

As these language models can be prohibitively large, you may want to use parameter efficient fine-tuning methods such as low rank adaption (LoRA). Such methods allow you to perform fine-tuning without the requirement of an extensive GPU memory.

For airway segmentation, please scale up the training you did in your last presentation and expand on the performance metrics you use to evaluate your findings.

Submission Requirements

For this assignment, you are required to submit the script used, as well as slides to present your results. The script can be in the form of a notebook or a collection of .py files.

In terms of the presentation, it is recommended to use slides for ease of switching between presenters on the lecture room computer. Keep in mind that the presentation should be concise and easily understandable within the given time frame. **Your presentation should be 5-7 minutes.**

Peer Evaluation

In this assignment, you are required to conduct a peer evaluation of your classmates' work. The rubric is available below. Note that 20 points are enough to get a full grade. Please refrain from giving 5 points unless you consider the performance exceptional.

1. Approach and Methodology

- **0 Points:** No clear methodology or approach is described; the submission lacks coherence in its chosen methods.
- **1-2 Points:** The methodology is mentioned but lacks detail or clarity; the approach is somewhat relevant but poorly explained.
- **3-4 Points:** The approach is well-defined and relevant to the task; some details may be missing, but the overall methodology is sound.
- **5 Points (Exceptional):** A comprehensive and clearly articulated methodology; demonstrates a thorough understanding and application of appropriate techniques for the task.

2. Use of Data

- **0 Points:** Ineffective or no use of data; fails to demonstrate how data was applied in the project.
- **1-2 Points:** Limited or basic use of data; some attempt at leveraging data but with significant room for improvement.
- **3-4 Points:** Good use of data; the submission shows a solid attempt at utilizing data effectively, though there may be minor issues or missed opportunities.
- **5 Points (Exceptional):** Excellent use of data; demonstrates innovative and effective application of data in addressing the project's objectives.

3. Results and Performance Analysis

- **0 Points:** No results or performance analysis provided.
- **1-2 Points:** Results are mentioned but lack detail or meaningful analysis; limited insight into performance.

- **3-4 Points:** Results are well-presented with a clear performance analysis; demonstrates a good understanding of outcomes with some critical analysis.
- **5 Points (Exceptional):** Exceptional presentation of results and comprehensive performance analysis; provides deep insights and thorough examination of the project's outcomes.

4. Presentation Quality

- **0 Points:** The presentation is unclear, disorganized, or significantly lacks critical information.
- **1-2 Points:** The presentation is somewhat organized but lacks clarity or polish; difficult to follow at times.
- **3-4 Points:** Good presentation quality; clear and well-organized, though there may be minor issues with flow or clarity.
- **5 Points (Exceptional):** Exceptional presentation quality; highly organized, clear, engaging, and effectively communicates the project's details and results.

5. Innovation and Creativity

- **0 Points:** No evidence of creativity or innovation; the project is entirely derivative.
- **1-2 Points:** Limited creativity or innovation; some attempts at originality but largely follows conventional approaches.
- **3-4 Points:** Good display of creativity or innovation; shows original thinking in some aspects of the project.
- **5 Points (Exceptional):** Highly creative and innovative approach; demonstrates a novel application of methods or techniques that significantly enhance the project's value.

Please note that the deadline for submitting your assignments is at 18:00 on the 14th of February.

We wish you the best of luck with this assignment. If you have any questions or concerns, please do not hesitate to reach out to through Ed or by sending an email to hafsteinne@hi.is.