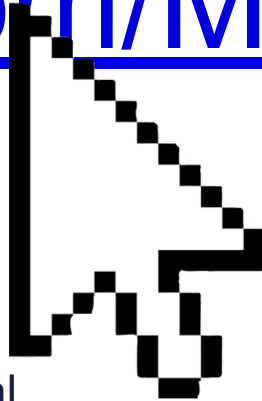


# Second Term Project Assignment

BEIR CQADupStack and ARQMath Collection

[https://github.com/MIR-MU/  
pv211-utils](https://github.com/MIR-MU/pv211-utils)



MUNI  
FI



Vít Starý Novotný et al.

PV211: Introduction to Information Retrieval

Spring 2024

# Introduction

- **CRANFIELD** (1950s) evaluates retrieval of scientific article metadata.
- **CQADupStack** (2015) evaluates the retrieval of text+math+code questions.
- **ARQMath 1–3** (2020–2) evaluates retrieval of text+math questions/answers.

## Query

How can I evaluate  $\sum_{n=0}^{\infty} (n + 1)x^n$ ?

Asked 8 years, 5 months ago Active 4 months ago Viewed 34k times

How can I evaluate

384

$$\sum_{n=1}^{\infty} \frac{2n}{3^{n+1}}$$

★

146

I know the answer thanks to [Wolfram Alpha](#), but I'm more concerned with how I can derive that answer. It cites tests to prove that it is convergent, but my class has never learned these before so I feel that there must be a simpler method.

In general, how can I evaluate

$$\sum_{n=0}^{\infty} (n + 1)x^n?$$

## Search Results

1

No need to use Taylor series, this can be derived in a similar way to the formula for geometric series. Let's find a general formula for the following sum:

$$S_m = \sum_{n=1}^m nr^n.$$

...

2

It is equivalent to  $x(x + 1)(x + 5)(x + 6) + 96 = 0$

Now

...

$$(x^2 + 6x)(x^2 + 6x + 5) + 96 = 0$$

3

If you want a solution that doesn't require derivatives or integrals, notice that

$$1 + 2x + 3x^2 + 4x^3 + \dots = 1 + x + x^2 + x^3 + \dots$$

...

$$+ x + x^2 + x^3 + \dots$$

$$+ x^2 + x^3 + \dots$$

# Requirements

- Implement a *supervised* ranked retrieval system using relevance judgements.
  - Experiment with techniques such as [weighted zone scoring](#) and [large language models](#).
  - NVIDIA GPUs are available at [JupyterHub](#):
    - NVIDIA A10 (24G VRAM), A40 (48G VRAM), and A100 (80G VRAM)
- Document your code in accordance with [PEP 8](#) and [PEP 257](#).
- Reach at least 25% MAP for **CQADupStack** and 10% MAP for **ARQMath**.
- Submit .ipynb file by 2024-05-07 for either **CQADupStack** or **ARQMath**.
- You will be awarded  $\leq 20$ pt for project,  $\leq 6$ pt for explanation,  $\leq 6$  pt for review.
- You can get extra 20 / 10 / 9 / 8 / ... / 1 points for 1st / 2nd / ... / 11th final place in the leaderboard.

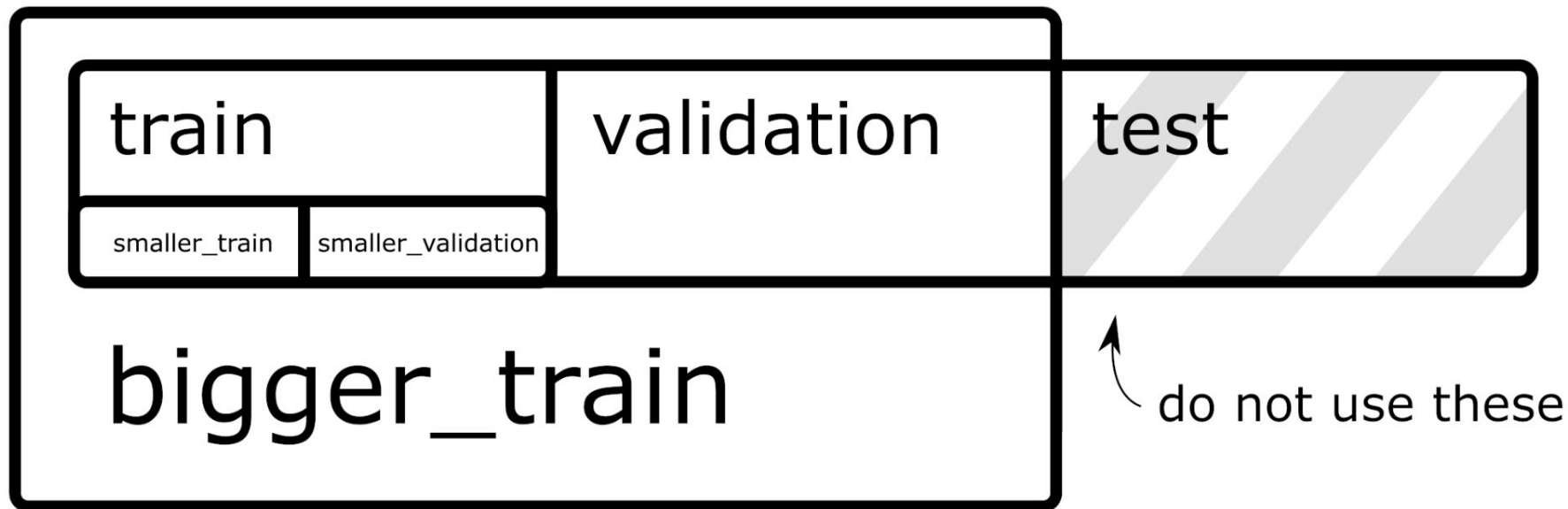
# Retrieval Units

- **CRANFIELD** contains 225 queries over 1398 abstracts of journal articles.
  - **Task:** Finding journal articles using short queries.
- **CQADupStack** contains 300 queries over 457,188 questions from SE forums.
  - **Task:** Finding old duplicate questions to new questions.
- **ARQMath** contains 200 queries over 1,445,495 answers from MSE forums.
  - **Task:** Finding old answers to new math questions.
  - Each answer is connected to one of 1,020,585 questions, which can be used for enrichment.
  - Queries, questions, and answers are available with seven different *math representations*.
- **CQADupStack** and **ARQMath** don't fit in 12G RAM at Google Colab.  
Use [JupyterHub](#), where we provide up to 64G RAM.

# Relevance Judgements

- **CRANFIELD** contains 314,550 judgements with 1,837 (0.58%) relevant.
  - The relevance judgements are *exhaustive*: all 1398 abstracts × 225 queries have been judged
- **CQADupStack** contains 23,692 judgements, all of which (100%) are relevant.
- **ARQMath** contains 70,912 judgements with 3,765 (5.31%) relevant.
  - The relevance judgements are *graded* from 0 (non-relevant) to 3 (relevant).
- For evaluation, we use *MAP*, which uses exhaustive non-graded judgements.
  - We made judgements for **CQADupStack** & **ARQMath** exhaustive by non-judged = non-relevant.
  - We made judgements for **ARQMath** non-graded by making 0–1 non-relevant and 2–3 relevant.

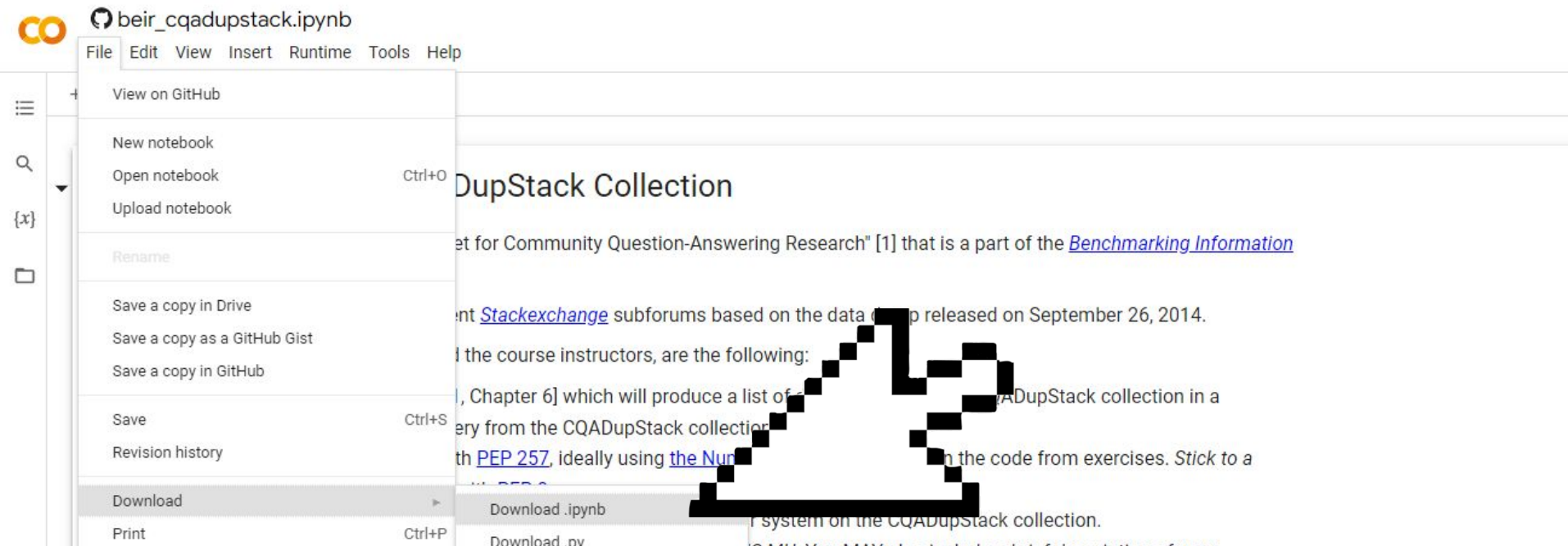
## Dataset Splits



Dataset splits for supervised training on **CQADupStack** and **ARQMath**.

# How to Use JupyterHub

1. Download notebook with example solution for [CQADupStack](#) / [ARQMath](#):



The image shows a Jupyter Notebook interface. At the top left, there is a Jupyter logo and the filename "beir\_cqadupstack.ipynb". Below the filename is a menu bar with options: File, Edit, View, Insert, Runtime, Tools, Help. A "File" menu is open, displaying several options: "View on GitHub", "New notebook", "Open notebook" (with a keyboard shortcut "Ctrl+O"), "Upload notebook", "Rename", "Save a copy in Drive", "Save a copy as a GitHub Gist", "Save a copy in GitHub", "Save" (with a keyboard shortcut "Ctrl+S"), "Revision history", "Download" (highlighted), and "Print" (with a keyboard shortcut "Ctrl+P").

The main content area of the notebook shows the title "DupStack Collection" and the beginning of a paragraph: "et for Community Question-Answering Research" [1] that is a part of the [Benchmarking Information](#)". Below this, there is a large black redaction box covering several lines of text. The visible text includes "ent [Stackexchange](#) subforums based on the data (p released on September 26, 2014.", "d the course instructors, are the following:", "l, Chapter 6] which will produce a list of", "ADupStack collection in a", "ery from the CQADupStack collection", "th [PEP 257](#), ideally using [the Nun](#)", "n the code from exercises. *Stick to a*", and "r system on the CQADupStack collection."

# How to Use JupyterHub

2. Set up (CPU, GPU, RAM) and launch a server at [JupyterHub](#):

Server Options

**PV 211**

Home

Erase if home exists

Take care of checking this button, it removes whole home directory and previous data will be lost. Use in case only when notebook is broken so it does not start, in other cases, remove data from terminal.

Resources

CPU

By default, 1 CPU is assigned to notebooks.

Select number of CPU (1-32):

Memory

Please choose upper memory limit (in GB) which will be assigned to notebook (default 4):

GPU

By default, no GPU is assigned. Would you like to use GPU?

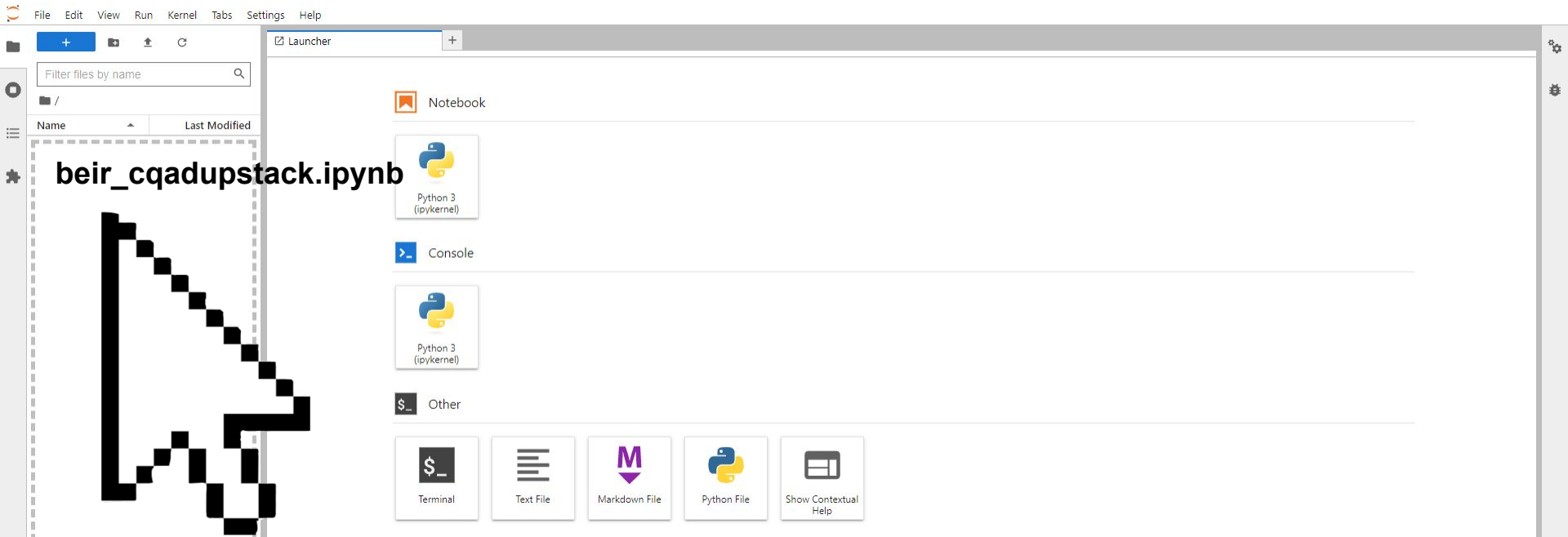
Start





# How to Use JupyterHub

## 3. Drag'n'drop notebook with example solution to JupyterHub:



# How to Use JupyterHub

4. Work on the assignment as you would at Google Colab or elsewhere.

The screenshot shows a JupyterLab environment. On the left is a file browser with a search bar and a table of files. The main area displays a notebook with the following content:

**Second Term Project: CQADupStack Collection**

The CQADupStack is "[a] Benchmark Data Set for Community Question-Answering Research" [1] that is a part of the *Benchmarking Information Retrieval (BEIR)* collection.

CQADupStack contains data from 12 different *Stackexchange* subforums based on the data dump released on September 26, 2014.

Your tasks, reviewed by your colleagues and the course instructors, are the following:

1. *Implement a ranked retrieval system*, [1, Chapter 6] which will produce a list of documents from the CQADupStack collection in a descending order of relevance to a query from the CQADupStack collection.
2. *Document your code* in accordance with PEP 257, ideally using the *NumPy style guide* as seen in the code from exercises. *Stick to a consistent coding style* in accordance with PEP 8.
3. *Reach at least 25% mean average precision at 10* [1, Section 8.4] with your system on the CQADupStack collection.
4. *Upload an .ipynb file with this Jupyter notebook to the homework vault in IS MU*. You MAY also include a brief description of your information retrieval system and a link to an external service such as *Google Collaboratory*, *DeepNote*, or *JupyterHub*.

[1] Hoogeveen, Doris and Verspoor, Karin M. and Baldwin, Timothy, *CQADupStack: A Benchmark Data Set for Community Question-Answering Research*. ACM, 2015.

**Import the utility tools from the git repository.**

First, we will install our library.

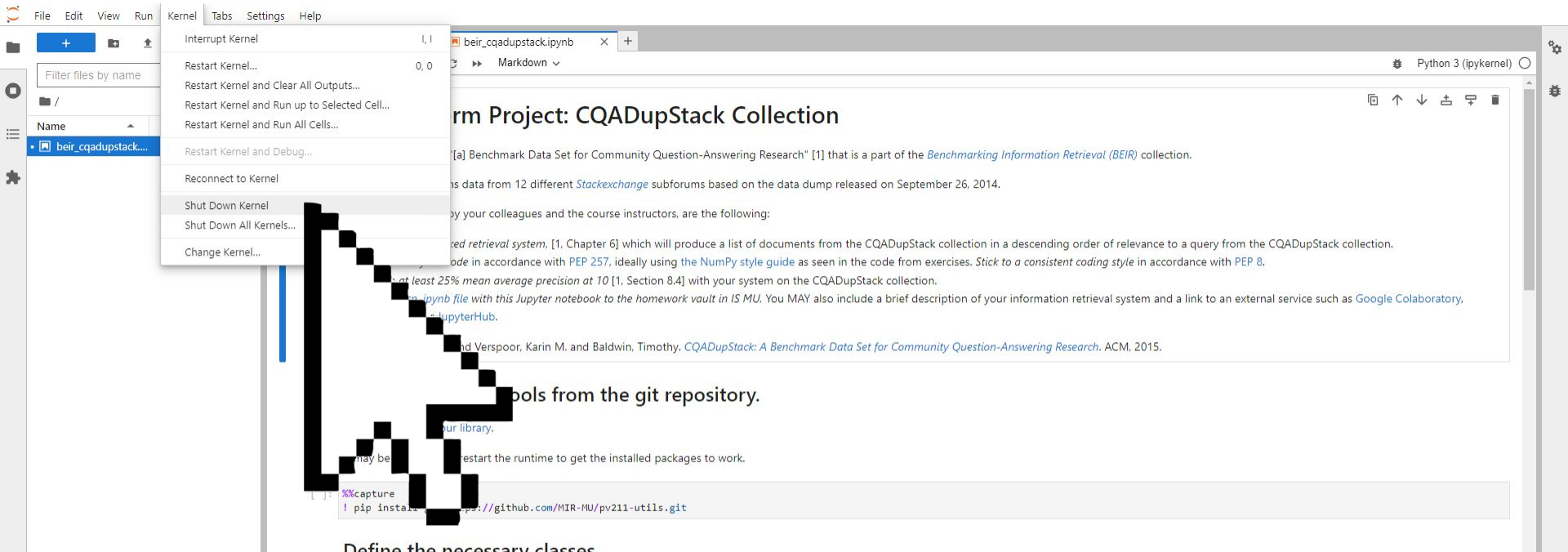
It may be necessary to restart the runtime to get the installed packages to work.

```
[ ]: %capture
! pip install git+https://github.com/MIR-MU/pv211-utils.git
```

**Define the necessary classes**

# How to Use JupyterHub

5. When you are done for the day, shut down Python, so the server can be reaped.



# Peer Review

- **Same Questions as in the first peer review**

**(1) Reproducible submission**

*Your colleague uploaded a Jupyter notebook file (in .ipynb format) to the homework vault in IS MU. After hitting "Run all cells", the notebook will run without issues all the way down to successful evaluation.*

**(2) Transparency**

*You are able to understand the methodology of the reviewed implementation, including the details, such as used constants, or data resources. Where needed, the code is documented and typed in accordance with PEP 257, ideally using the NumPy style guide as seen in the code from exercises.*

**(3) Correctness**

*The reviewed implementation is fair; It does not use the test data, or does not adjust the shared evaluation. The score reported in the leaderboard matches the evaluation score that you obtain from running the script. The submission passes minimum score threshold. In case you think the solution could be plagiarised or copy-pasted without a reference to the source, please give more details in the description below.*

**(4) Innovativeness**

*You personally think this solution is exciting, creative or in any way inspiring. This can include, for instance, innovativeness in the method design, or efficiency. Alternatively, you learned something new from this solution, including small things, like the impact of specific design decisions.*

# Peer Review

- **Same** Questions as in the **first peer review**
- **Extra (6pts)**: To help your reviewers to assess the quality of your solution, we'll ask you to submit your notebook with a **short Technical Report** (max 400 words):

## **Report Objectives & Evaluation criteria:**

- To allow your reviewer to **understand your solution**, even if it uses technologies outside the scope of the course
- To **present findings** from the development that you do not submit, but led you to final solution
  - comparison of scores with different choice of preprocessing, encoding, distance measure, ...
- We will provide an optional Report template 1 week before the deadline