## Midterm project - 2021 Fall

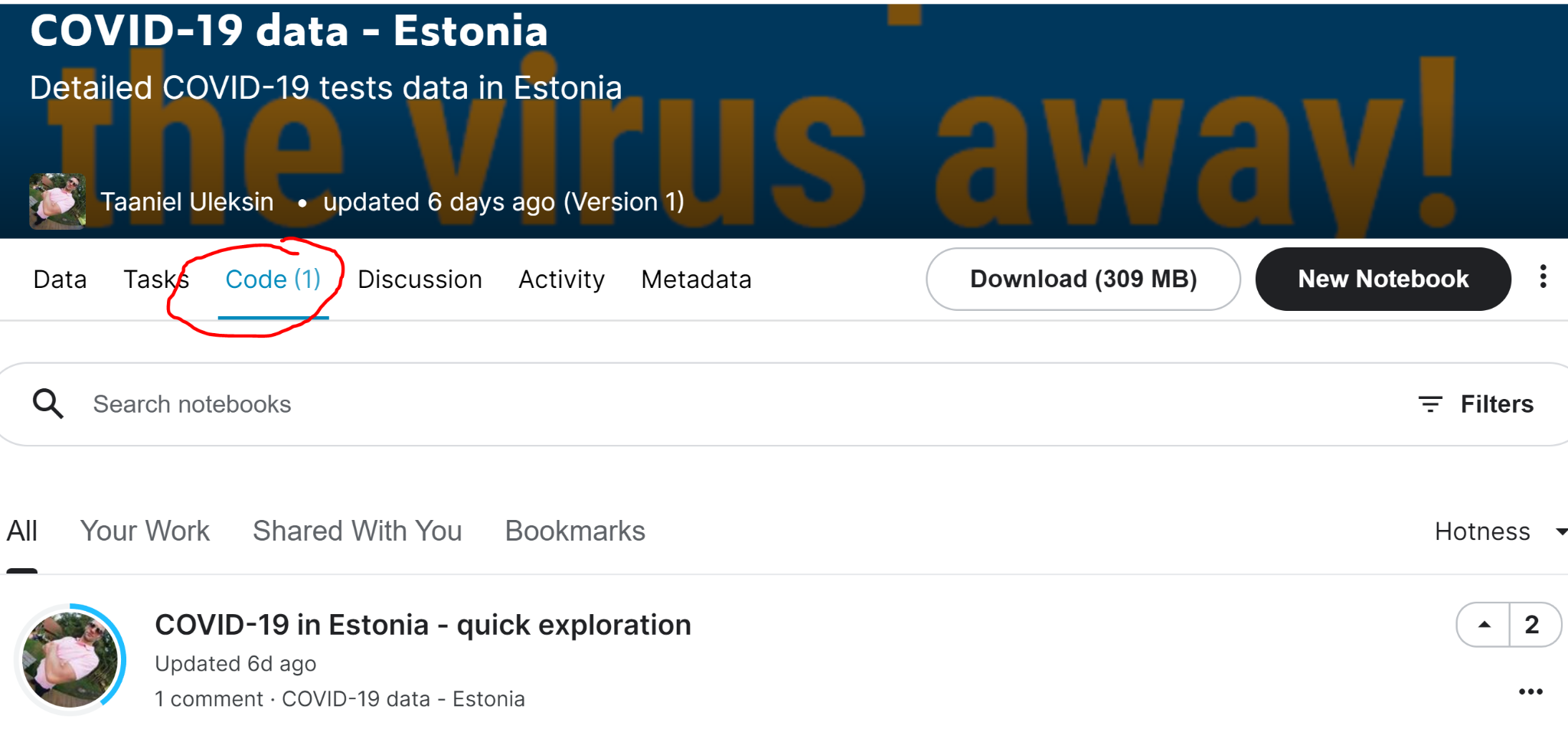
### **0. General guidelines**

* Write your own codes by following the instructions below.
* Submit your 1) code, 2) written report (doc or pdf) and 3) video presentation file (TOTAL length of 5 mins or shorter) to [hyunyang@postech.ac.kr](mailto:hyunyang@postech.ac.kr) with the title “[AI lecture] Student ID Midterm project”.
  + Put “AI lecture” on your title. This is very important.
* Deadline: by **23:59 pm, Nov. 21.**
* Code cheating will be checked. If you cheat, you will get 0 grades on any occasion.

### **I. Use a unique dataset from Kaggle**

**[Rules for determining your dataset]**

* It can be any type of dataset, videos, images, texts, signals, etc.
* You need to use the dataset that nobody has made a learning program for.



For instance, there should be no deep learning codes published for the dataset as shown in the snapshot above. In the above snapshot, the code is just about plotting the dataset samples, so this kind of code does not account. I mean classification or regression codes: there should be no such deep learning codes for your dataset.

* You need to use a distinct dataset. Once you determine to use a certain dataset, go visit the google sheet [**here**](https://docs.google.com/spreadsheets/d/1KrzZW4u0C8TnbnNxYqxzAtAwAyThvuiWwuk8oRM_Ews/edit?usp=sharing). If your dataset is not on the sheet already, you can use that dataset, so write down your dataset in the sheet [**here**](https://docs.google.com/spreadsheets/d/1KrzZW4u0C8TnbnNxYqxzAtAwAyThvuiWwuk8oRM_Ews/edit?usp=sharing). First-come, first-served. You can change your decision later on the same non-overlapping usage constraint.

**[What you need to do]**

* Write a code for regression or classification yourself by designing custom fully-connected neural networks.
* Cut down some of the input features or reduce the dataset size if you struggle to deal with the dataset in Colab.
* Show your more than three efforts to improve the inference accuracy. The efforts include
  + controlling your learning rate
  + modifying your neural network architecture
  + changing the optimizer
  + pre-processing of data (rotating, cropping, transforming, etc)
  + etc

**[Grading policy: each item will be graded as one of A, B, and C]**

* Is the dataset distinct? (no other students are using, and there are no DL codes in Kaggle)
* Does the dataset have abundant samples for each training and testing (say more than 100 samples or so)
* Do your codes run well without errors?
* Are there more than three efforts to improve the inference performance?
* Is the report including insightful results and discussions?
* Is the presentation video well explaining the work?

### **II. Generate and use your own custom dataset**

**[Rules for determining your dataset]**

* Make arbitrary multi-dimensional inputs (the dimension of the input should be larger than 10).
  + It can be limited/unlimited, discrete/continuous, whatever.
* Make an arbitrarily complicated (use polynomials, exponentials, square-root, etc) input-output function, and generate training (more than 10,000 samples) and testing datasets (more than 2,000 samples) based on your function.

**[What you need to do]**

* Write a code for regression to approximate your function by designing custom fully-connected neural networks.
* Show your more than two efforts to improve the inference accuracy of the deep learning approach. The efforts include
  + controlling your learning rate
  + modifying your neural network architecture
  + etc

**[Grading policy: each item will be graded as one of A, B, and C]**

* Is the dataset following the instructions? (size, dimension, and the complexity of the function)
* Do your codes run well without errors?
* Are there more than two efforts to improve the DL inference performance?
* Is the report including insightful results and discussions?
* Is the presentation video well explaining the work?