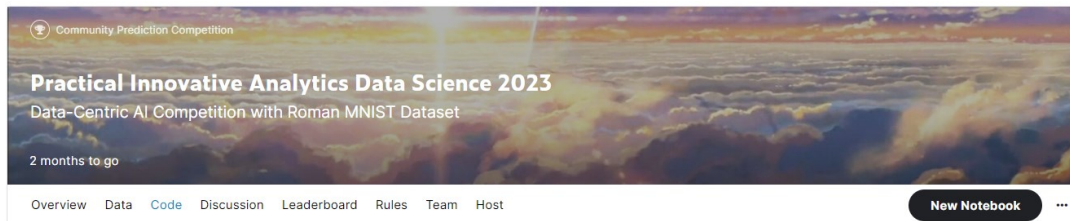


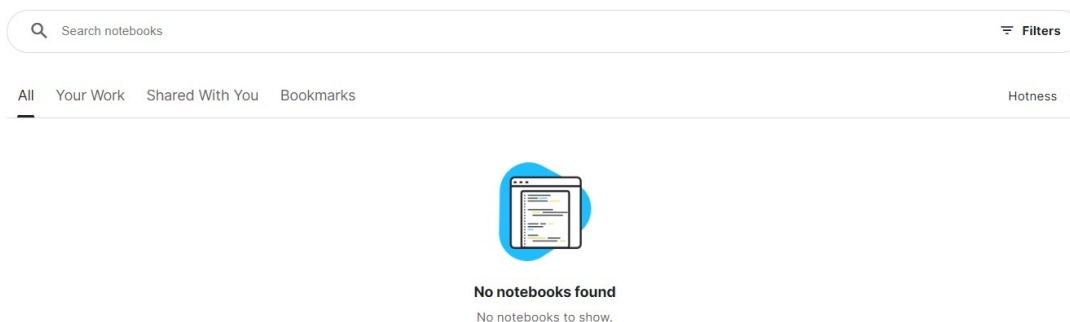
# Kaggle tutorial

## 1. How to get the data?

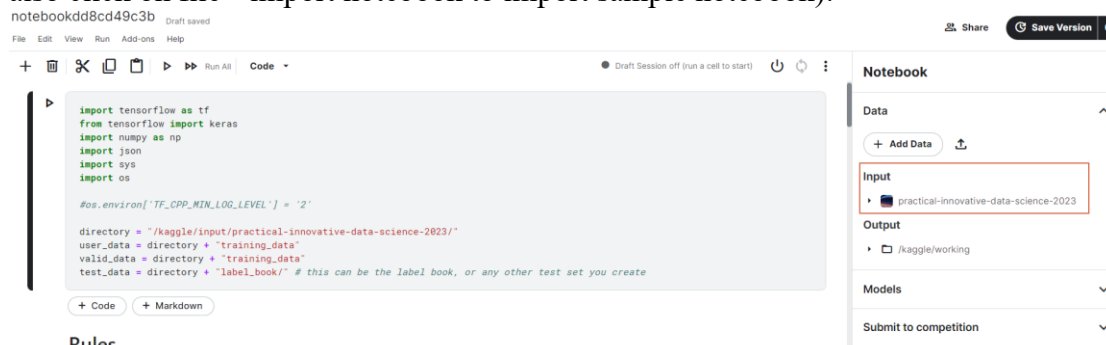
Register [Kaggle](#) first and on the [competition website](#), in the code panel click “New Notebook”:



### Notebooks



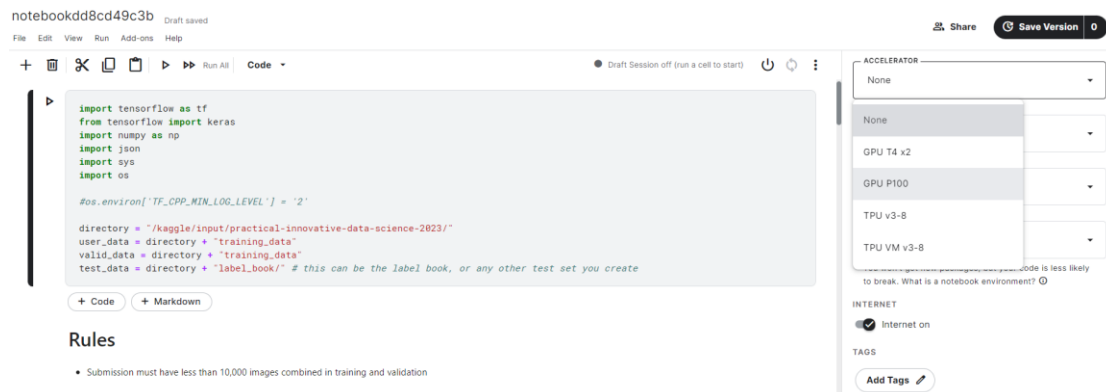
The data will be automatically downloaded to the “/kaggle/input” directory (You can also click on file->import notebook to import sample notebook):



Notice that all the intermediate output will be stored in the default location “/kaggle/working”. You can directly download the data in this directory by the GUI.

## 2. How to use GPU?

Turn on the accelerator as follows:

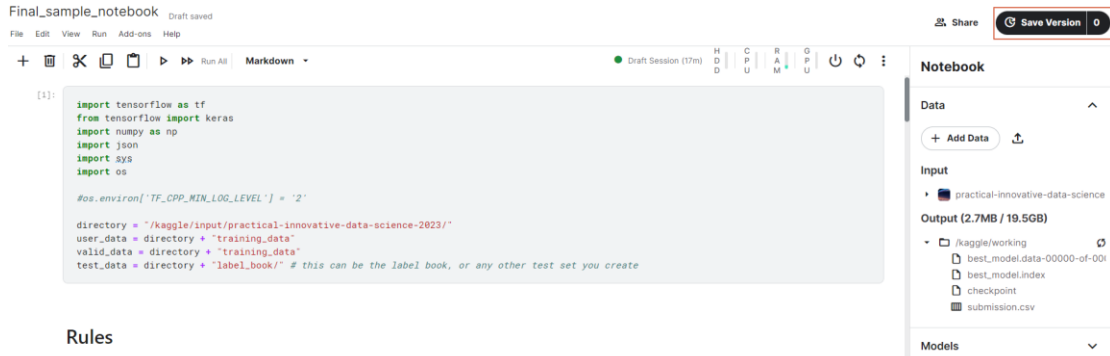


The Resources are listed below

- Kaggle GPU: 16G NVIDIA TESLA P100
  - <https://www.kaggle.com/docs/efficient-gpu-usage>
- Limited to 30+ hrs/week depending on usage.
- Limited to 12hrs/run

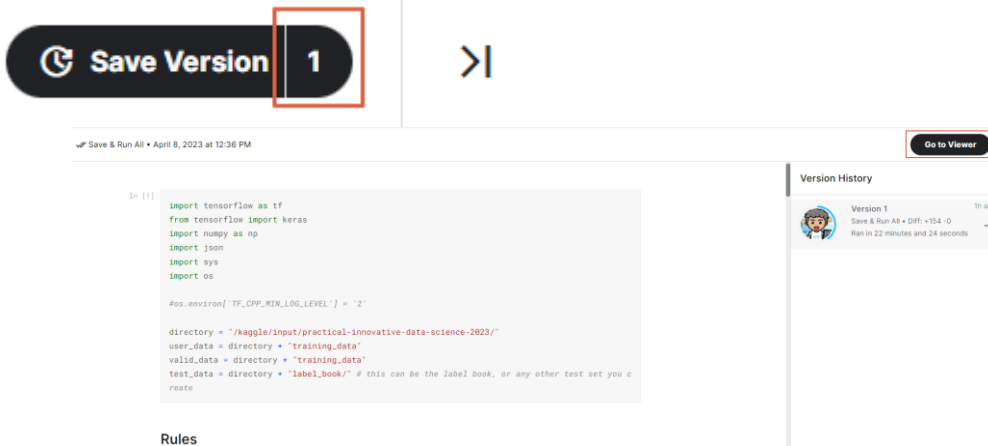
### 3. Run the code in the background

You can also run the code in the background with Kaggle. Firstly, make sure your code is bug-free, as any error in any code block would result in early stopping. Click the “Save Version” button as follows:



Rules

Notice that the output is limited to 20G and the max run time is limited to 12hrs. You can then see the results by clicking the version number:



Rules

The log will be shown in the Notebook panel

**notebookdd8cd49c3b**

Python · Practical Innovative Analytics Data Science 2023

Notebook Input Output **Logs** Comments (0) Settings

#### Logs

Download Logs

Successfully ran in 1344.7s

Accelerator  
None

Environment  
Latest Container Image

Output  
2.8 MB

Time	#	Log Message
1264.0s	1	Epoch 73/75 1/421 [.....] - ETA: 36s - loss: 0.1584 - accuracy: 1.0000 3/421 [.....] - ETA: 15s - loss: 0.0922 - accuracy: 1.0000 5/421 [.....] - ETA: 14s - loss: 0.0734 - accuracy: 1.0000 7/421 [.....] - ETA: 14s - loss: 0.0991 - accuracy: 0.9821 9/421 [.....] - ETA: 15s - loss: 0.0886 - accuracy: 0.9861 11/421 [.....] - ETA: 14s - loss: 0.0917 - accuracy: 0.9866 13/421

And your output can be accessed through the “Data” panel, where you can download your data.:

## notebookdd8cd49c3b

Python · Practical Innovative Analytics Data Science 2023

Notebook Input **Output** Logs Comments (0) Settings

### Output Data

submission.csv (4.46 kB) Submit ↓ >

Id	Predicted
0000.png	8
0001.png	1
0002.png	10
0003.png	10
0004.png	10
0005.png	5
0006.png	2

Output :

- best\_model.data-00000-of-0
- best\_model.index
- checkpoint
- submission.csv**

## Colab tutorial

### 1. How to get the data?

Register [Kaggle](#) first. In the “Data” panel of the [competition website](#), click Download all:

Host Overview **Data** Code Discussion Leaderboard Rules Team Submissions ...

label\_book (10 directories, 1 files) 🗨 >

About this directory ✎

This file does not have a description yet.

i 6 files

ii 6 files

iii 7 files

iv 6 files

ix 6 files

v 7 files

vi 6 files

vii 6 files

viii 6 files

x 6 files

desktop.ini 244 B

Data Explorer 8.86 MB

- label\_book
- test\_data
- training\_data

Summary

- 4815 files

Download All

+ New Version

```
>_ kaggle competitions download -c practical-innovative-data-science-2023
```

You can follow the sample notebook first. After exploring, store your data in your google drive and get the shareable link. Execute the following commands:

```
!gdown -fuzzy https://drive.google.com/file/d/1HiLwQWkzeAICXz-sXqNHtCMiEyl9Q7F/view?usp=sharing # Replace the link with your file
```

Remember, DO NOT store input data in your drive and load from there. The input/output is very slow (store at ./ instead). Your output data should be stored in

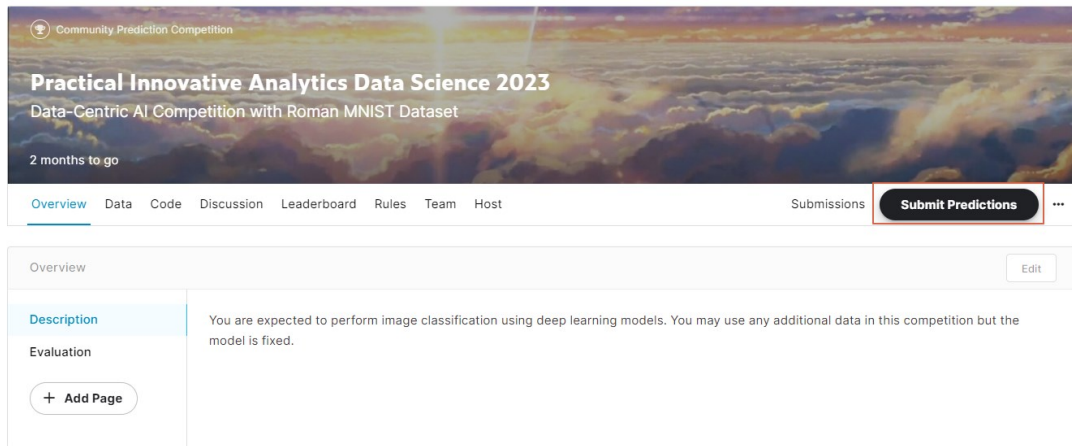
your google drive so that it can be accessed next time.

2. For other tips on using Colab refer to the Colab tutorial at our course website <https://phonchi.github.io/nsysu-math608/schedule/>.

## Tips

---

1. You can use the sample notebook to prepare your *submission.csv* and Upload your submission to get an estimate of your score



2. Some tips that may help you to improve your accuracy
  - Data cleaning
  - Data preprocessing
  - Fix the label issue
  - Data augmentation
  - Collect or generate more data