

Train model on Kaggle

Mì AI

Content

- Remind about Colab
- What is Kaggle?
- Kaggle Notebook & Dataset
- Kaggle Input
- Kaggle Models
- Kaggle Output
- Handson

What is Colab?

What is Colaboratory?

Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

Is it really free of charge to use?

Yes. Colab is free of charge to use.

Seems too good to be true. What are the limitations?

Colab resources are not guaranteed and not unlimited, and the usage limits sometimes fluctuate. This is necessary for Colab to be able to provide resources free of charge. For more details, see [Resource Limits](#)

Users who are interested in more reliable access to better resources may be interested in [Colab Pro](#).

Colab Limit

- RAM: 12.7GB (11GB usable)
- GPU: 12GB / 16GB (Nvidia K80 / T4)
- HDD: 80GB (72GB usable)

How long can notebooks run in Colab?

Colab prioritizes interactive compute. Runtimes will time out if you are idle.

In the version of Colab that is free of charge notebooks can run for at most 12 hours, depending on availability and your usage patterns. Colab Pro, Pro+, and Pay As You Go offer you increased compute availability based on your compute unit balance.

<https://research.google.com/colaboratory/faq.html#resource-limits>

Colab Limit

- RAM: 12.7GB (11GB usable)
- GPU: 12GB / 16GB (Nvidia K80 / T4)
- HDD: 80GB (72GB usable)

The usage limit is pretty dynamic and depends on how much/long you use colab. I was able to use the GPUs after 5 days; however, my account again reached usage limit right after 30mins of using the GPUs (google must have decreased it further for my account). The situation really became normal after months of not using colab from that account. My suggestion is to have multiple google accounts for colab, so you could use the other accounts when facing usage limits. Sorry for not replying to the comments.

Kaggle

What is Kaggle?

Kaggle is an online community platform for data scientists and machine learning enthusiasts. Kaggle allows users to collaborate with other users, find and publish datasets, use GPU integrated notebooks, and compete with other data scientists to solve data science challenges. The aim of this online platform (founded in 2010 by Anthony Goldbloom and Jeremy Howard and acquired by Google in 2017) is to help professionals and learners reach their goals in their data science journey with the powerful tools and resources it provides. As of today (2021), there are over 8 million registered users on Kaggle.

Kaggle

Is Kaggle free?

Yes, everything on Kaggle is completely free: courses, certificates obtained from courses, datasets, participation in competitions, discussion sections, etc.

Efficient GPU Usage Tips and Tricks

Kaggle provides free access to NVIDIA TESLA P100 GPUs. These GPUs are useful for training deep learning models, though they do not accelerate most other workflows (i.e. libraries like pandas and scikit-learn do not benefit from access to GPUs).

You can use up to a quota limit per week of GPU. The quota resets weekly and is 30 hours or sometimes higher depending on demand and resources

- Thông thường hay nghe đến Kaggle với các cuộc thi ML
- Cung cấp nhiều Dataset, Notebook Sample hay!

Kaggle Notebook

The screenshot displays the Kaggle IDE interface. At the top, the title bar reads "Titanic Data_RandomForest_by R" with a "Draft saved" status. The main menu includes "File", "Edit", "Insert", "Run", and "Help". On the right, there are buttons for "+ Add Data", "Commit", and a navigation arrow.

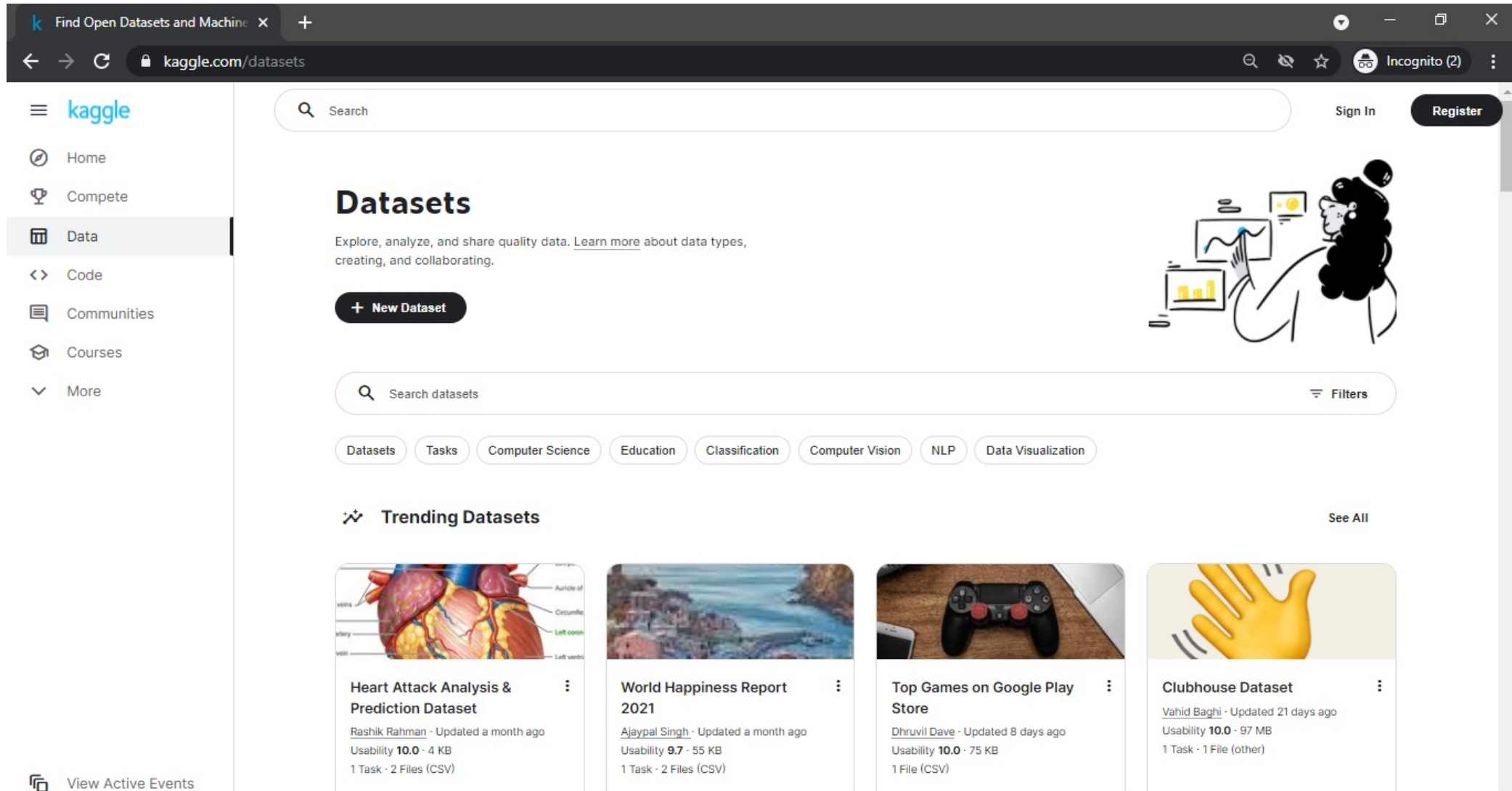
The central workspace shows two R code snippets. The first snippet, labeled "In[]:", contains the code to read a CSV file: `## read data file` followed by `titanic = read.csv('../input/titanic/train.csv', header = TRUE)`. The second snippet, also labeled "In[]:", contains the code to view column names: `## column names` followed by `colnames(titanic)`.

Below the code editor, a file explorer shows the "train.csv" file selected. The file path is displayed as `../input/titanic/train.csv`. The file size is noted as 59.76KB. The interface indicates that 7 of 12 columns are visible.

On the right side, a sidebar provides session and workspace information. The "Sessions" section shows a "Draft Session" with a duration of 12m/9h and GPU status "Off". The "Workspace" section lists the files: "input (read-only data)", "Titanic: Machine Learning from..." (containing "gender_submission.csv 418 x 2", "test.csv 418 x 11", and "train.csv 891 x 12").

At the bottom, a table preview is visible with columns: "# PassengerId", "# Survived", "# Pclass", and "A Name". The first row shows a passenger with ID 1, who survived, and is in Pclass 3.

Kaggle Dataset



The screenshot shows the Kaggle Datasets homepage in a web browser. The browser's address bar displays 'kaggle.com/datasets'. The page features a left sidebar with navigation links: Home, Compete, Data (highlighted), Code, Communities, Courses, and More. The main content area has a search bar, a 'Sign In' link, and a 'Register' button. Below the search bar is a 'Datasets' section with a description: 'Explore, analyze, and share quality data. Learn more about data types, creating, and collaborating.' A '+ New Dataset' button is present. A second search bar is followed by a 'Filters' button. Below these are category tags: Datasets, Tasks, Computer Science, Education, Classification, Computer Vision, NLP, and Data Visualization. The 'Trending Datasets' section displays four dataset cards: 'Heart Attack Analysis & Prediction Dataset' by Rashik Rahman (Usability 10.0, 4 KB, 1 Task, 2 Files (CSV)), 'World Happiness Report 2021' by Ajaypal Singh (Usability 9.7, 55 KB, 1 Task, 2 Files (CSV)), 'Top Games on Google Play Store' by Dhruvil Dave (Usability 10.0, 75 KB, 1 File (CSV)), and 'Clubhouse Dataset' by Vahid Baghi (Usability 10.0, 97 MB, 1 Task, 1 File (other)). A 'View Active Events' link is at the bottom left.

Find Open Datasets and Machine Learning x +

kaggle.com/datasets

Incognito (2)

Search

Sign In Register

Datasets

Explore, analyze, and share quality data. [Learn more](#) about data types, creating, and collaborating.

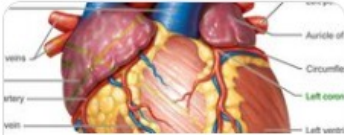
+ New Dataset

Search datasets Filters

Datasets Tasks Computer Science Education Classification Computer Vision NLP Data Visualization


Trending Datasets

See All




Heart Attack Analysis & Prediction Dataset

[Rashik Rahman](#) · Updated a month ago
Usability **10.0** · 4 KB
1 Task · 2 Files (CSV)




World Happiness Report 2021

[Ajaypal Singh](#) · Updated a month ago
Usability **9.7** · 55 KB
1 Task · 2 Files (CSV)



Top Games on Google Play Store

[Dhruvil Dave](#) · Updated 8 days ago
Usability **10.0** · 75 KB
1 File (CSV)



Clubhouse Dataset

[Vahid Baghi](#) · Updated 21 days ago
Usability **10.0** · 97 MB
1 Task · 1 File (other)

View Active Events

Kaggle Folders

Kaggle Datasets allows you to publish and share datasets privately or publicly. We provide resources for storing and processing datasets, but there are certain technical specifications: **100GB per dataset limit**. 100GB max private datasets (if you exceed this, either make your datasets public or delete unused datasets)



Kaggle

<https://www.kaggle.com/docs/datasets>

Datasets - Kaggle

- **Input: /kaggle/input** (100GB)
- **Output:**
 - **/kaggle/working**: output lưu ở đây sẽ được lưu lại để sử dụng cho những lần sau. Quota cho output là 20GB.
 - **/kaggle/temp**: output lưu ở đây chỉ sử dụng được trong nội bộ session đang chạy. Nếu lưu và thoát ra thì sẽ biến mất.

Kaggle Input

- Public DS
- Private Upload (ZIP)
- Remote URL



Remote URL

Create a dataset from remote URLs.
URLs must point to a file.



Import GitHub repository

Create from a GitHub repository archive.
Use the repo URL or any deep link.



Import Google Cloud Storage

Create a dataset from a public Google Cloud Storage bucket.
URLs must point to a bucket or path. Ex: gs://bucket/path

Kaggle Input

- Download from Gdrive

```
# kaggle - download zip file using id
```

```
!pip install gdown
```

```
!gdown --id 1xjl-AwJPmIsIzr4miur8XgG_1WIkEhyB
```

Collecting gdown

Downloading gdown-4.7.1-py3-none-any.whl (15 kB)

Requirement already satisfied: filelock in /opt/conda/lib/python3.10/site-packages (from gd

Requirement already satisfied: requests[socks] in /opt/conda/lib/python3.10/site-packages (

Requirement already satisfied: six in /opt/conda/lib/python3.10/site-packages (from gdown)

Requirement already satisfied: tqdm in /opt/conda/lib/python3.10/site-packages (from gdown)

Requirement already satisfied: beautifulsoup4 in /opt/conda/lib/python3.10/site-packages (f

Requirement already satisfied: soupsieve>1.2 in /opt/conda/lib/python3.10/site-packages (fr

Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.10/site-p

Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-packages (fro

Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/conda/lib/python3.10/site-package

Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.10/site-package

Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /opt/conda/lib/python3.10/site-pac

Installing collected packages: gdown

Successfully installed gdown-4.7.1



Kaggle Model

- Add public model

Models

+ Add Models



No models added

Add a Kaggle model

Add Models

Search keyword or URL

Text

Image

Computer Vision

Video

306 Models

Hotness ▾



mistral

Mistral AI team is proud to release Mistral, th...

Mistral AI · 3 Variations



llama-2

Llama 2 is a collection of pretrained and fine...



Request access on the [model's page](#)



Download Output

Output (72.5MB / 19.5GB)

▼  /kaggle/working



 a.h5



Download

Models



Handson

