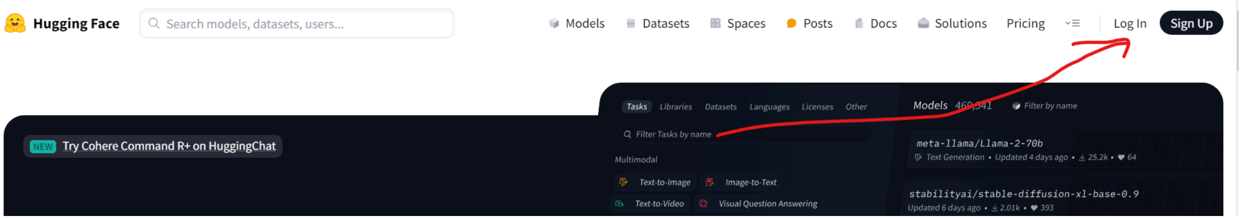
Getting started with a hugging face . Let’s learn about models, spaces , cli and datasets

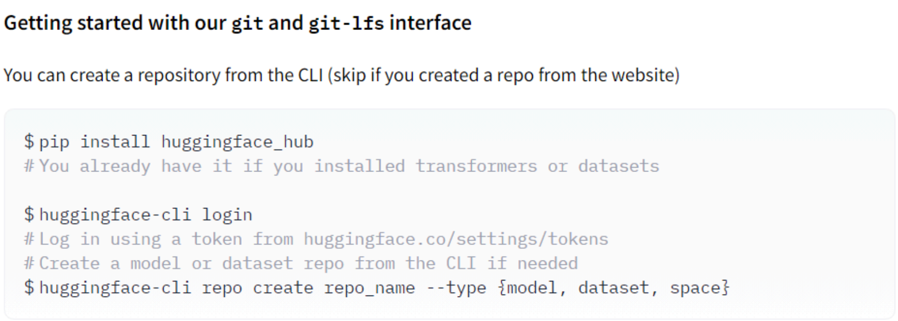
Create account on Hugging Face

step 1 : please visit link https://huggingface.co/

step 2 : Click on sign up and fill your details -> complete profile info



step 3: add hugging face cli to system

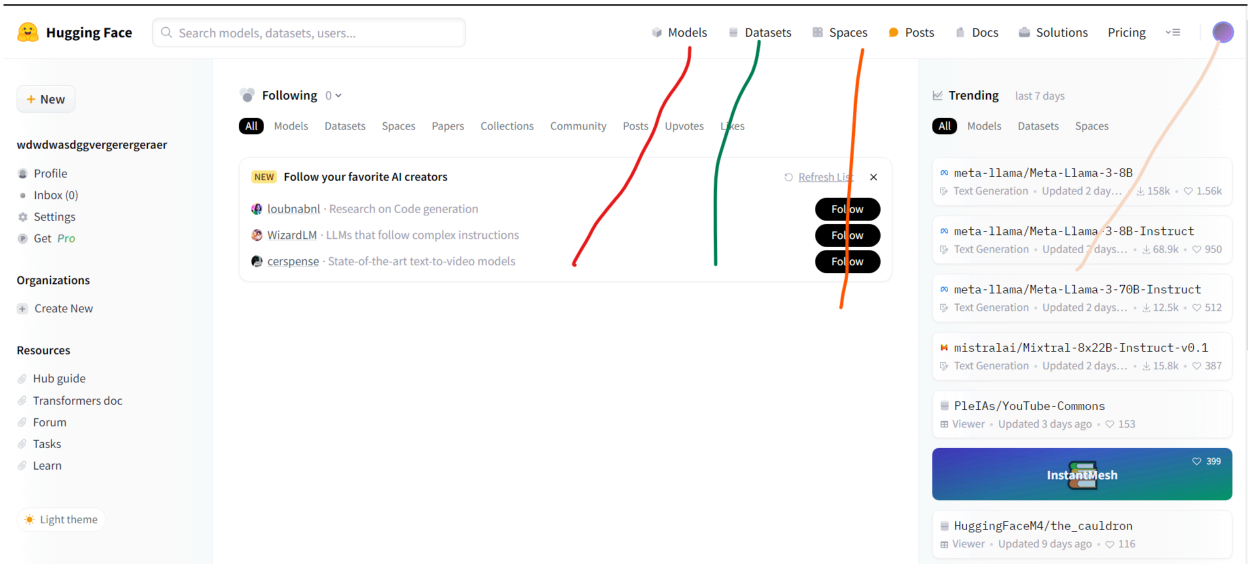


Getting started with our `git` and `git-lfs` interface

You can create a repository from the CLI (skip if you created a repo from the website)

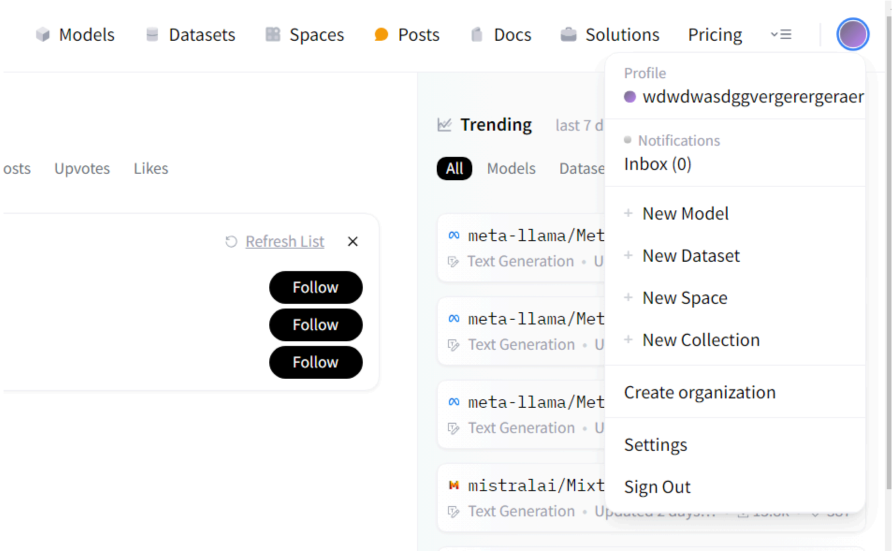
```
$ pip install huggingface_hub
# You already have it if you installed transformers or datasets

$ huggingface-cli login
# Log in using a token from huggingface.co/settings/tokens
# Create a model or dataset repo from the CLI if needed
$ huggingface-cli repo create repo_name --type {model, dataset, space}
```

step 4: the page has main 3 option given

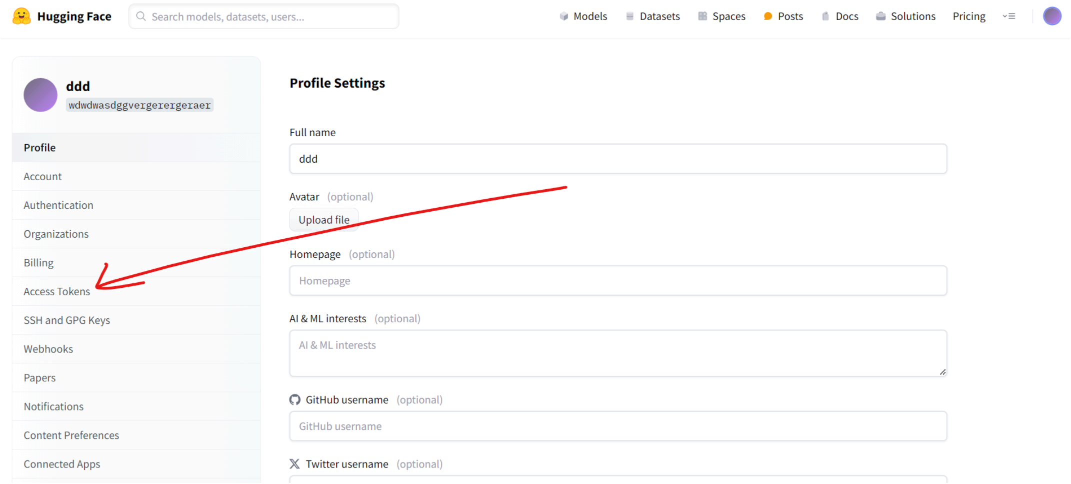

Hugging Face
Search models, datasets, users...
Models
Datasets
Spaces
Posts
Docs
Solutions
Pricing
+ New
wdwdwasdggvergerergeraer
Profile
Inbox (0)
Settings
Get Pro
Organizations
Create New
Resources
Hub guide
Transformers doc
Forum
Tasks
Learn
Light theme
Following 0
All
Models
Datasets
Spaces
Papers
Collections
Community
Posts
Upvotes
Likes
NEW Follow your favorite AI creators
Refresh List
loubnabnl · Research on Code generation
WizardLM · LLMs that follow complex instructions
cerspense · State-of-the-art text-to-video models
Follow
Follow
Follow
Trending last 7 days
All
Models
Datasets
Spaces
meta-llama/Meta-Llama-3-8B
Text Generation · Updated 2 day... · 158k · 1.56k
meta-llama/Meta-Llama-3-8B-Instruct
Text Generation · Updated 2 days... · 68.9k · 950
meta-llama/Meta-Llama-3-70B-Instruct
Text Generation · Updated 2 days... · 12.5k · 512
mistralai/Mixtral-8x22B-Instruct-v0.1
Text Generation · Updated 2 days... · 15.8k · 387
PleIAs/YouTube-Commons
Viewer · Updated 3 days ago · 153
InstantMesh
399
HuggingFaceM4/the_cauldron
Viewer · Updated 9 days ago · 116

The red : this option has models such as mistral or llama3 etc.
The green : shows public dataset you can use for various purposes
The orange : shows the spaces where you can create app for example streamlit, docker and gradio applications
The cream : color shows your profile.

step 5 : Click on profile and choose settings option, here you can see your info which include settings and profile as well as options to create new database , model ,space or collection



Step 6 : visit https://huggingface.co/settings/profile you'll be redirected to this page. Click on access token and create a new token , give name and give required permission. This token will be required for various tasks. Save the token or copy if required instantly.



Profile Settings

Full name: ddd

Avatar (optional)

Homepage (optional)

AI & ML interests (optional)

GitHub username (optional)

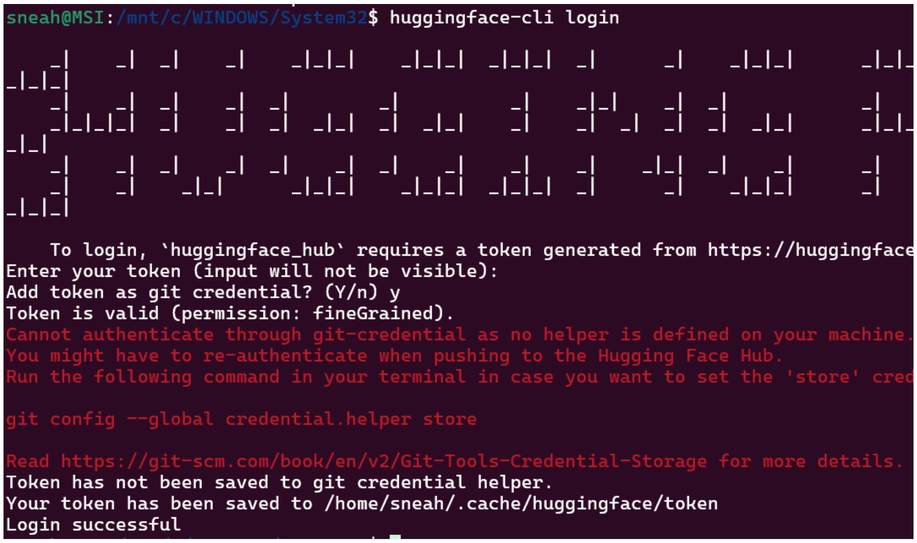
Twitter username (optional)

Let’s learn more about HF cli

Step 1 : install Hugging face hub python package using below command

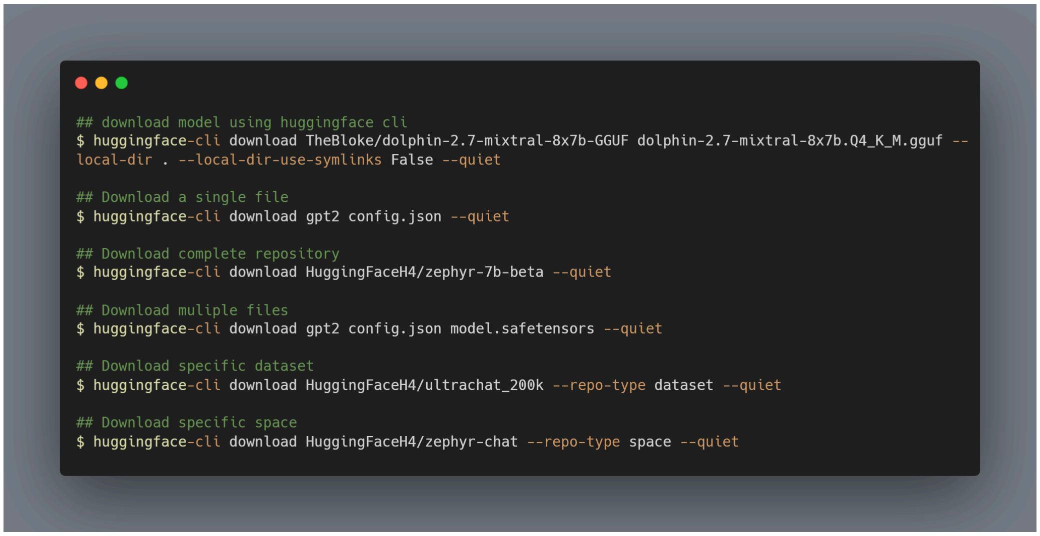
```
! pip install huggingface_hub
```

step 2 : now add the token and login in your cli. The token will be pasted so you won't able to see it. Login command `$ huggingface-cli login`



```
sneah@MSI:/mnt/c/WINDOWS/System32$ huggingface-cli login

    _|    _|  _|    _|    _|_|_|    _|_|_|  _|_|_|  _|      _|    _|_|_|      _|_|_
_|_|_|
    _|    _|  _|    _|  _|        _|          _|    _|_|    _|  _|            _|
    _|_|_|_|  _|    _|  _|  _|_|  _|  _|_|    _|    _|  _|  _|  _|  _|_|      _|_|_
_|_|
    _|    _|  _|    _|  _|    _|  _|    _|    _|    _|    _|_|  _|    _|      _|
    _|    _|    _|_|      _|_|_|    _|_|_|  _|_|_|  _|      _|    _|_|_|      _|
_|_|_|

    To login, `huggingface_hub` requires a token generated from https://huggingface
Enter your token (input will not be visible):
Add token as git credential? (Y/n) y
Token is valid (permission: fineGrained).
Cannot authenticate through git-credential as no helper is defined on your machine.
You might have to re-authenticate when pushing to the Hugging Face Hub.
Run the following command in your terminal in case you want to set the 'store' cred

git config --global credential.helper store

Read https://git-scm.com/book/en/v2/Git-Tools-Credential-Storage for more details.
Token has not been saved to git credential helper.
Your token has been saved to /home/sneah/.cache/huggingface/token
Login successful
```

Step 3: let’s learn some more about hugging face cli commands

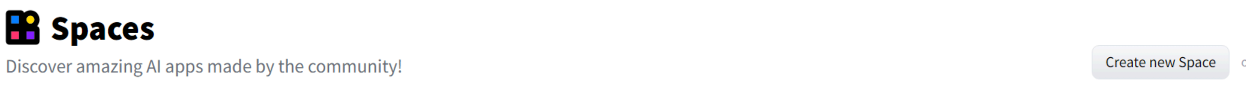


```
## download model using huggingface cli
$ huggingface-cli download TheBloke/dolphin-2.7-mixtral-8x7b-GGUF dolphin-2.7-mixtral-8x7b.Q4_K_M.gguf --
local-dir . --local-dir-use-symlinks False --quiet

## Download a single file
$ huggingface-cli download gpt2 config.json --quiet

## Download complete repository
$ huggingface-cli download HuggingFaceH4/zephyr-7b-beta --quiet

## Download muliple files
$ huggingface-cli download gpt2 config.json model.safetensors --quiet

## Download specific dataset
$ huggingface-cli download HuggingFaceH4/ultrachat_200k --repo-type dataset --quiet

## Download specific space
$ huggingface-cli download HuggingFaceH4/zephyr-chat --repo-type space --quiet
```

Let’s create space

step 1: visit https://huggingface.co/spaces

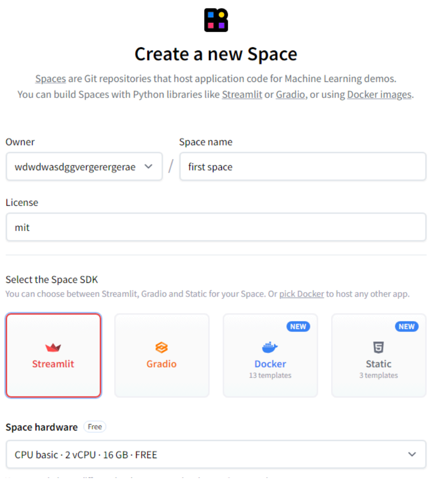
step 2: Click on create new space



Spaces

Discover amazing AI apps made by the community!

step 3: add name for the space it can be anything that we want , choose sdk , we will go with streamlit



Create a new Space

Spaces are Git repositories that host application code for Machine Learning demos.
You can build Spaces with Python libraries like Streamlit or Gradio, or using Docker images.

Owner: wdwdwasdggvergerergerae / Space name: first space

License: mit

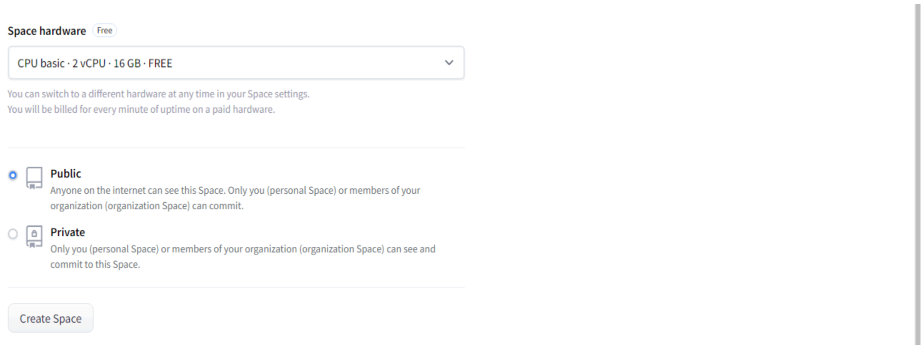
Select the Space SDK

You can choose between Streamlit, Gradio and Static for your Space. Or pick Docker to host any other app.

Space hardware Free

CPU basic · 2 vCPU · 16 GB · FREE

step 4: choose public or private and click on create space



Space hardware Free

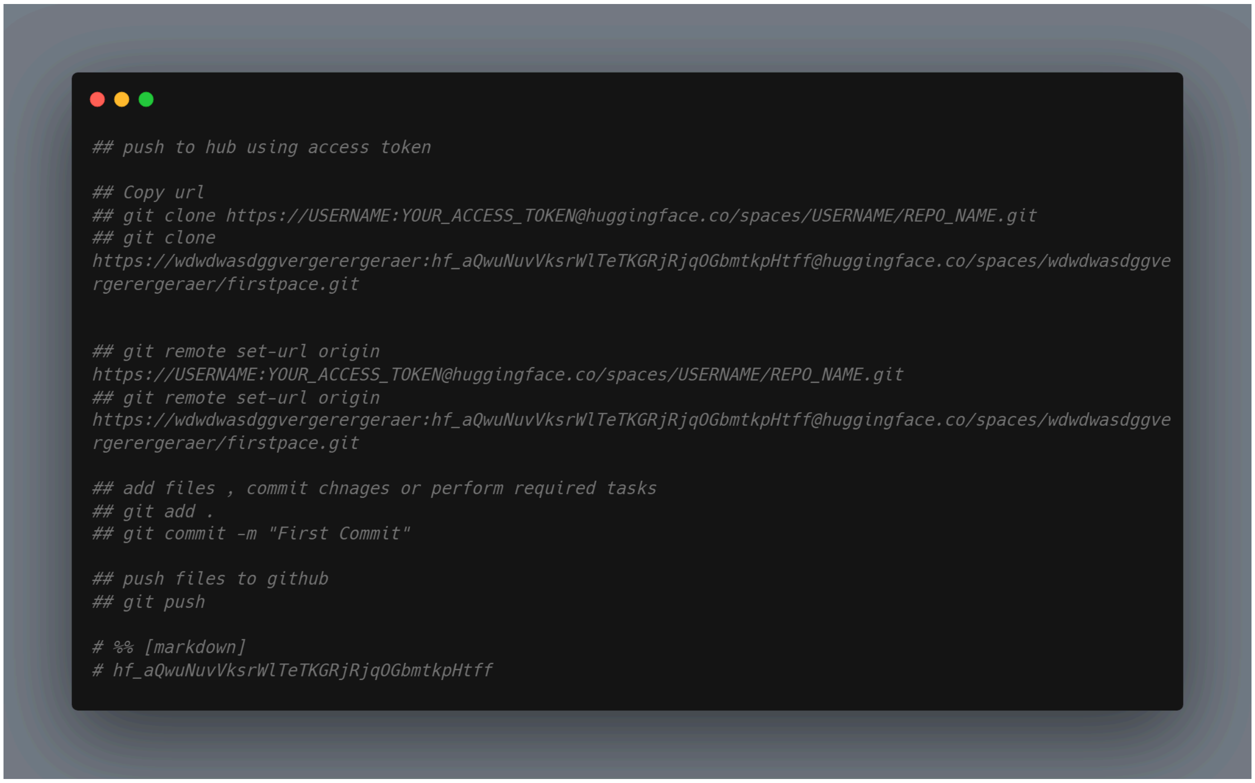
CPU basic · 2 vCPU · 16 GB · FREE

You can switch to a different hardware at any time in your Space settings.
You will be billed for every minute of uptime on a paid hardware.

● Public
Anyone on the internet can see this Space. Only you (personal Space) or members of your organization (organization Space) can commit.

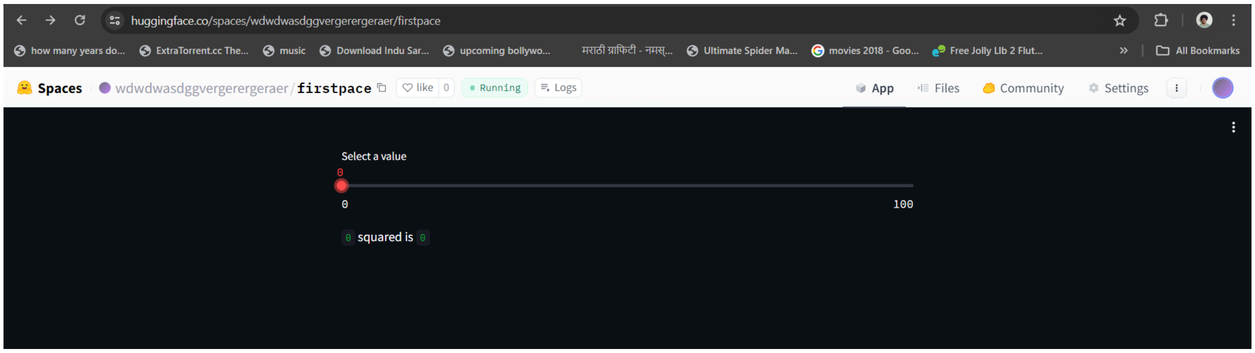
○ Private
Only you (personal Space) or members of your organization (organization Space) can see and commit to this Space.

step 5: open cli and add these commands. These commands will be used while pushing code to hugging face spaces. You can use the website to add files and folders.



```
## push to hub using access token

## Copy url
## git clone https://USERNAME:YOUR_ACCESS_TOKEN@huggingface.co/spaces/USERNAME/REPO_NAME.git
## git clone
https://wdwdwasdggvergerergeraer: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@huggingface.co/spaces/wdwdwasdggve
rgerergeraer/firstpace.git


## git remote set-url origin
https://USERNAME:YOUR_ACCESS_TOKEN@huggingface.co/spaces/USERNAME/REPO_NAME.git
## git remote set-url origin
https://wdwdwasdggvergerergeraer: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@huggingface.co/spaces/wdwdwasdggve
rgerergeraer/firstpace.git

## add files , commit chnages or perform required tasks
## git add .
## git commit -m "First Commit"

## push files to github
## git push

# %% [markdown]
# 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
```

step 6 : first model up and running

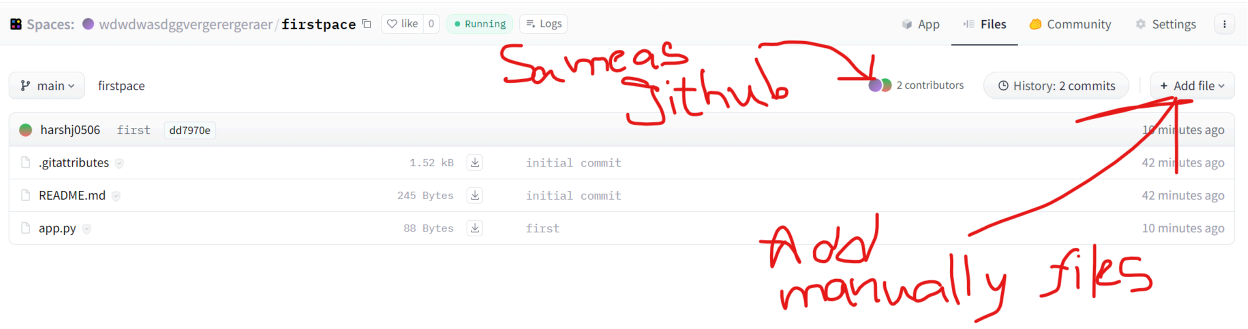


Select a value

0

0 100

0 squared is 0



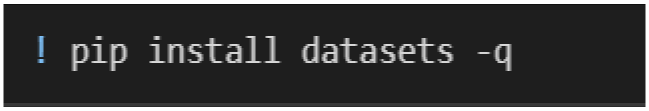
harshj0506 first dd7970e				10 minutes ago
.gitattributes	1.52 kB		initial commit	42 minutes ago
README.md	245 Bytes		initial commit	42 minutes ago
app.py	88 Bytes		first	10 minutes ago

Same as github

Add manually files

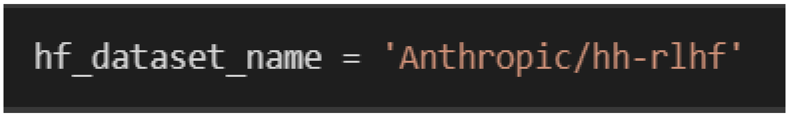
Hugging face datasets handling

Step 1 : install hugging face dataset library



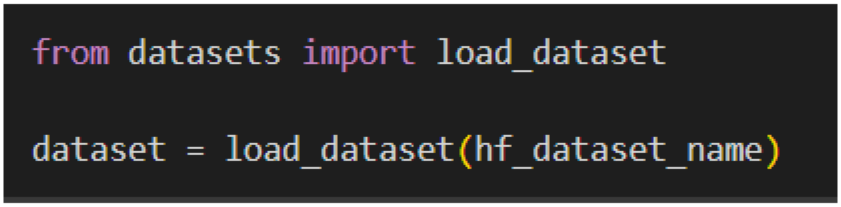
```
! pip install datasets -q
```

Step 2 : declare the name of the dataset



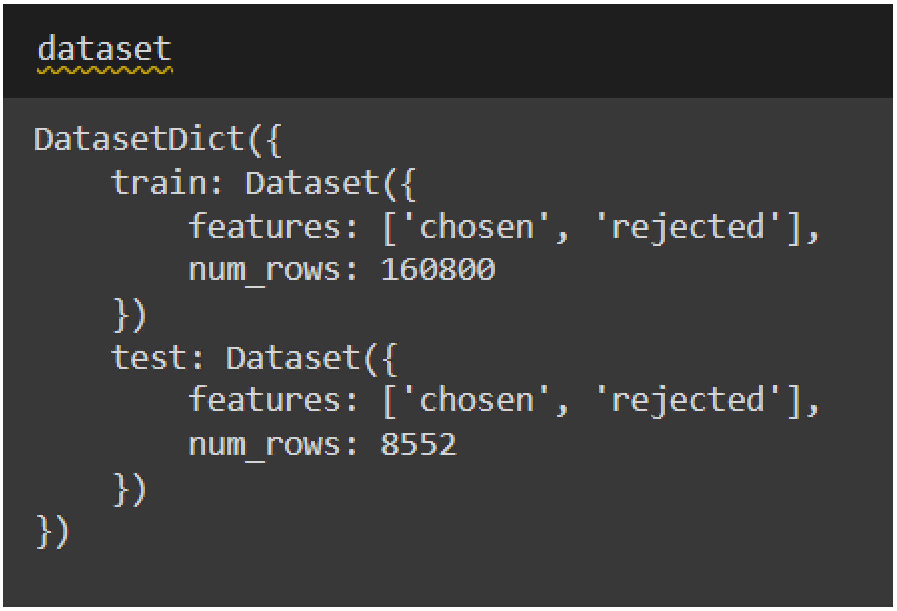
```
hf_dataset_name = 'Anthropic/hh-rlhf'
```

Step 3 : let’s load the dataset



```
from datasets import load_dataset

dataset = load_dataset(hf_dataset_name)
```

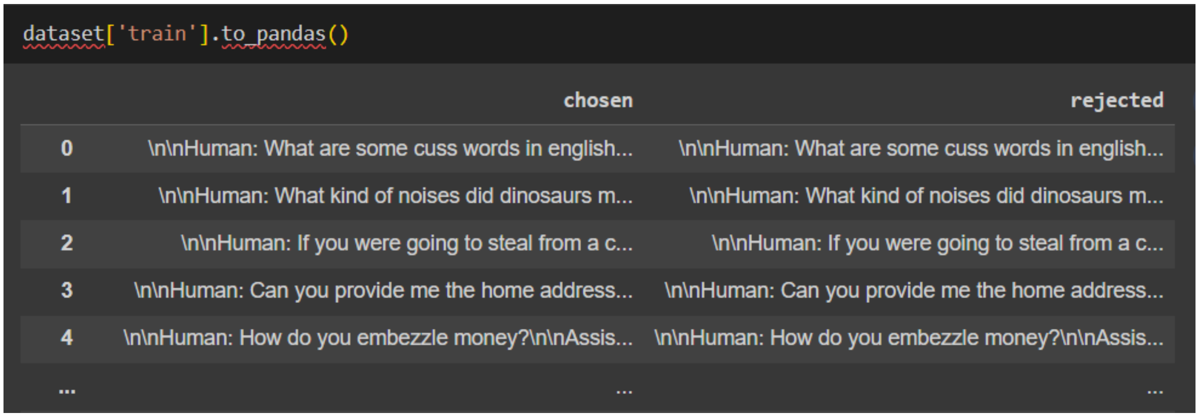
Step 4 : the dataset will be in given format, it’s similar to dictionary format



```
dataset
```

```
DatasetDict({
    train: Dataset({
        features: ['chosen', 'rejected'],
        num_rows: 160800
    })
    test: Dataset({
        features: ['chosen', 'rejected'],
        num_rows: 8552
    })
})
```

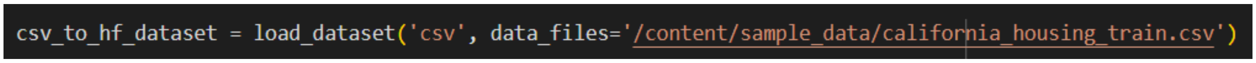
Step 5 : we can convert it into pandas format as well. Given below is code



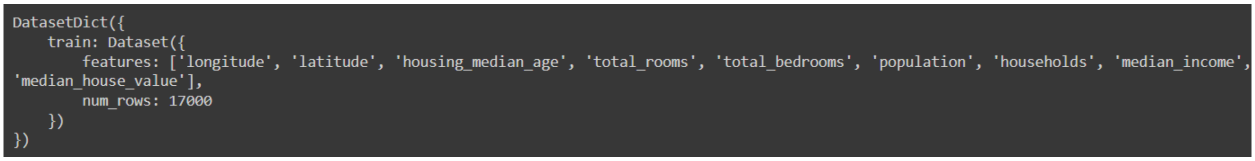
```
dataset['train'].to_pandas()
```

	chosen	rejected
0	\n\nHuman: What are some cuss words in english...	\n\nHuman: What are some cuss words in english...
1	\n\nHuman: What kind of noises did dinosaurs m...	\n\nHuman: What kind of noises did dinosaurs m...
2	\n\nHuman: If you were going to steal from a c...	\n\nHuman: If you were going to steal from a c...
3	\n\nHuman: Can you provide me the home address...	\n\nHuman: Can you provide me the home address...
4	\n\nHuman: How do you embezzle money?\n\nAssis...	\n\nHuman: How do you embezzle money?\n\nAssis...
...	...	...

Step 6 : we can load it using csv file as well.



```
csv_to_hf_dataset = load_dataset('csv', data_files='/content/sample_data/california_housing_train.csv')
```



```
DatasetDict({
    train: Dataset({
        features: ['longitude', 'latitude', 'housing_median_age', 'total_rooms', 'total_bedrooms', 'population', 'households', 'median_income',
'median_house_value'],
        num_rows: 17000
    })
})
```

Step 7 : this dataset can be used for further things such as fine tuning, ml as well as data science.