Hugging Face

Hugging Face is repository of models, datasets and spaces(ready to run apps using the models)

            Models

Models include millions of ML models for various tasks such as text-generation , NLP tasks, Image , vide

We can filter and select a model and see it outputs there itself and also see which all spaces use this mo

* You can either copy this model locally , by cloning it using git clone and run locally (sometimes require a

* Sometimes you didn't want to clone it rather use it via an API , for that you can use a library provided by

   1.
■- Sample code
■from transformers import AutoModel,AutoTokenizer
■model_name= "path to model"
■model = AutoModel.from_pretrained(model_name)
■tokenizer = AutoTokenizer.from_pretrained(model_name)
■
■#now we can tokenize the input the same way model used to train on
■inp = tokenizer.encode(input_data)

■#Generate output
■out = model.generate(inp)
■■OR
■out = model(inp)

■#decode back to natural language
■print(tokenizer.decode(out))

   2. Another way is using pipeline
■
■from transformers import pipeline
■model = pipeline("mention the task",model= "")

■if model didn't specified then default model will be chosen

* You can also use these models in the code by using HuggingFaceHub class in Langchain

All the models are open-source but some are private & gated for accessing them we need to get an acce

----
            Datasets

Datasets consists of lots of Datasets and we can download it locally and use or simply use them in our m

Hugging Face provides a library named datasets , simply import it and use load_dataset() class from it

■-Sample
■from datasets import load_dataset
■
■data = load_dataset("path")

■#saving it
■datasets.save_to_disk(data)

* There are various ways to load data and also split them , normally it is split into training and testing data

----
            Spaces

Spaces consists of multiple apps/spaces where we can simply click on them and run them there itself

So based on our likes we can locally clone those spaces and use it