Outils de développement pour l'apprentissage automatique

Club Intelligence Artificielle - Université Laval

Loïc Baret

Novembre 2023

Les outils

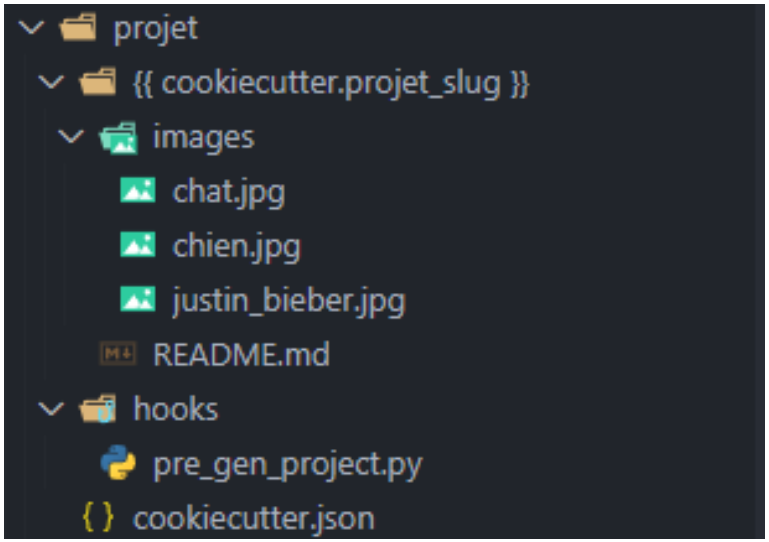
Aujourd'hui nous parlerons des outils suivants :

Rôle	Outils
Initialisation du projet	Cookiecutter
Gestion de l'environnement	pyenv + poetry
Configuration d'expérimentations	Hydra
Gestion des expériences	MLFlow
Encapsulation des modèles	PyTorch Lightning
Versionnage des données	Data Version Control

Cookiecutter

Cookiecutter

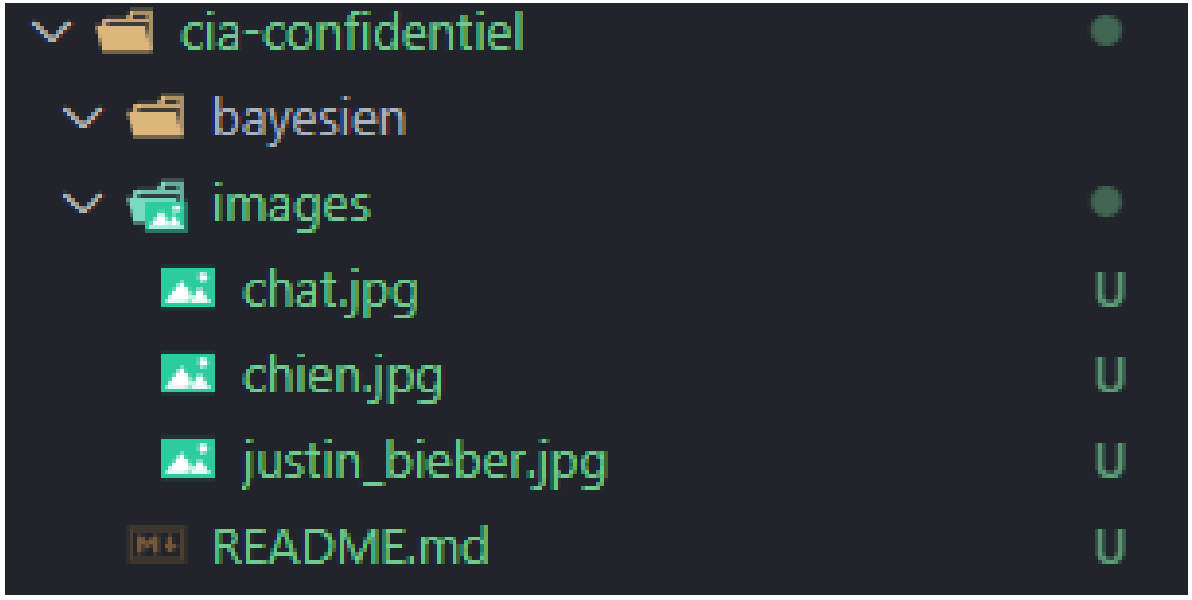
"Une image vaut milles mots" - Confucius :




```
[1/5] nom_projet (Exemple tres simple): cia-confidentiel
[2/5] projet_slug (cia-confidentiel):
[3/5] auteur (CIA Ulaval):
[4/5] Select prefere
  1 - chat
  2 - chien
  3 - justin bieber
  4 - aucun
  Choose from [1/2/3/4] (1): 4
[5/5] Select bayesien_ou_frequentiste
  1 - bayesien
  2 - frequentiste
  3 - autre
  Choose from [1/2/3] (1): 1
```

Cookiecutter

Le résultat de l'arborescence :



pyenv

pyenv

"On ne peut pas se battre contre une image" - Jacques Lacan :

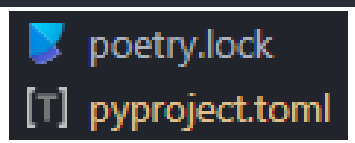
```
lbaret@DESKTOP-58IJQVU:~/projects/CIA_exemple$ which python
/home/lbaret/.pyenv/shims/python
lbaret@DESKTOP-58IJQVU:~/projects/CIA_exemple$ python --version
Python 3.6.15
lbaret@DESKTOP-58IJQVU:~/projects/CIA_exemple$ pyenv versions
  system
* 3.6.15 (set by /home/lbaret/projects/CIA_exemple/.python-version)
  3.10.11
  3.10.12
lbaret@DESKTOP-58IJQVU:~/projects/CIA_exemple$ pyenv local 3.10.12
lbaret@DESKTOP-58IJQVU:~/projects/CIA_exemple$ python --version
Python 3.10.12
lbaret@DESKTOP-58IJQVU:~/projects/CIA_exemple$ pyenv install 3.9.17
Downloading Python-3.9.17.tar.xz...
-> https://www.python.org/ftp/python/3.9.17/Python-3.9.17.tar.xz
Installing Python-3.9.17...
Installed Python-3.9.17 to /home/lbaret/.pyenv/versions/3.9.17
lbaret@DESKTOP-58IJQVU:~/projects/CIA_exemple$ pyenv versions
  system
  3.6.15
  3.9.17
  3.10.11
* 3.10.12 (set by /home/lbaret/projects/CIA_exemple/.python-version)
```

Poetry

Poetry

"Le style a le mouvement et l'image" - Anatole France

```toml
[tool.poetry]
name = "cia-mlops-tools"
version = "0.1.0"
description = "CIA MLOps tutorial"
authors = ["Loïc <baret.loic.1@gmail.com>"]
license = "MIT License"
readme = "README.md"

[tool.poetry.dependencies]
python = ">=3.10,<4.0"
cookiecutter = "^2.3.1"
mlflow = "^2.7.1"
lightning = "^2.0.9"
hydra-core = "^1.3.2"
tensorboard = "^2.14.0"
ipykernel = "^6.25.2"
poethepoet = "^0.24.1"

[tool.poe.tasks]
torch-cuda = "pip install torch torchvision torchaudio --index-url https://download.pytorch.org/whl/cu118"


[build-system]
requires = ["poetry-core"]
build-backend = "poetry.core.masonry.api"
```



Hydra


CONFIGURATION
HYDRA
READERS
CUSTOMS
EN DUR
DANS LE CODE
imgflip.com

Hydra

"Image, ma seule, mon unique passion" - Charles Beaudelaire


```yaml
defaults:
  # - base_config
  - membre: loic
  - projet: manga
  - _self_

experience:
  nom_modele: "manga_translator"
  poids: "./checkpoints/model_weights.pt"
  epoques: 50
  taux_apprentissage: 0.1
  optimiseur: "Adam"
  perte: "CrossEntropy"
```

MLFlow

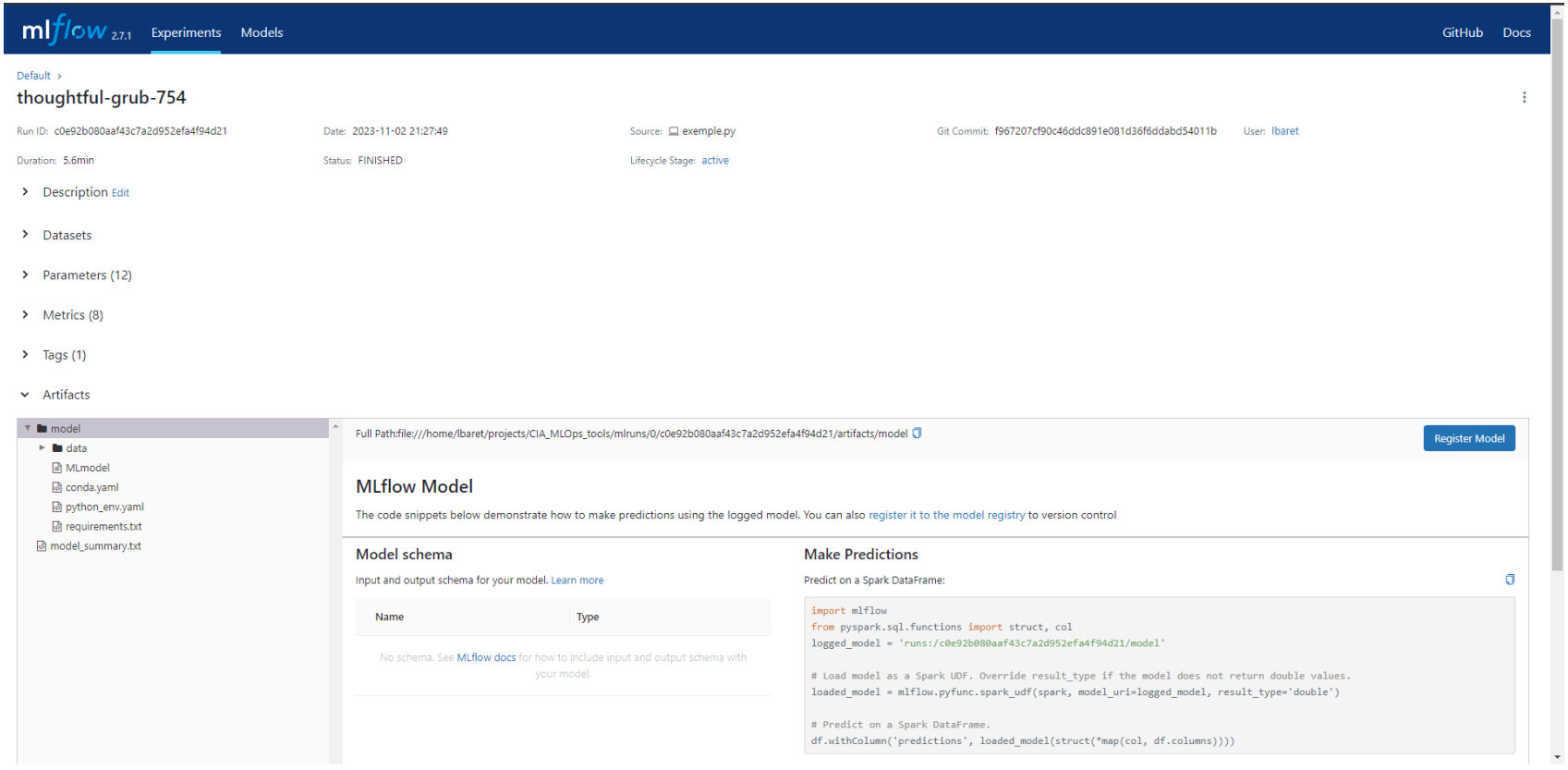
Godiche (Adjectif)

[go.diʃ] / Masculin et féminin identiques

- Benêt, maladroit.

MLFlow

"Nous sommes dans un siècle de l'image. pour le bien comme pour le mal, nous subissons plus que jamais l'action de l'image." - Gaston Bachelard



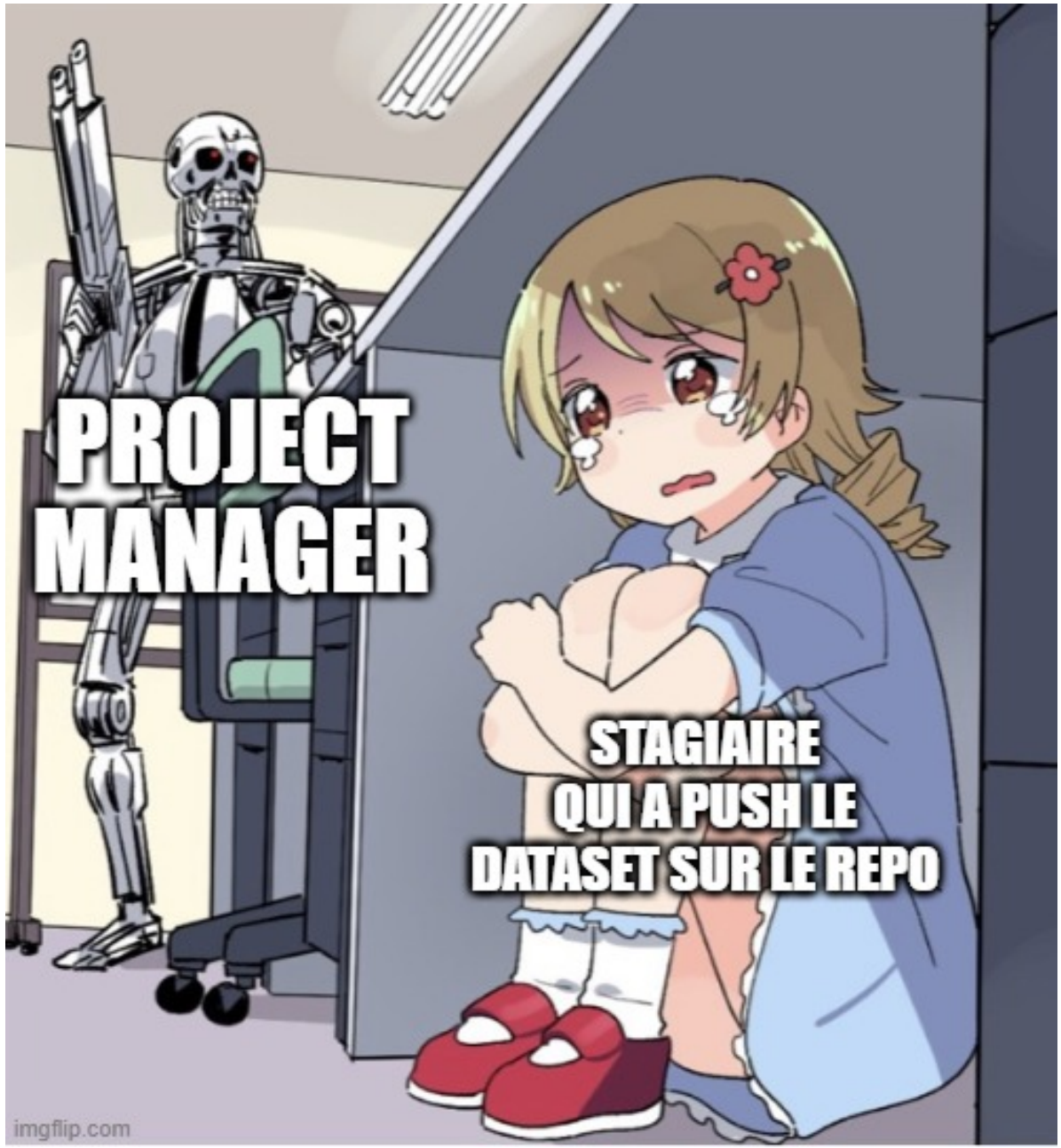
PyTorch Lightning

PyTorch Lightning

"L'image a une force que la parole n'a pas forcément." - Marceline Loridan-Ivens

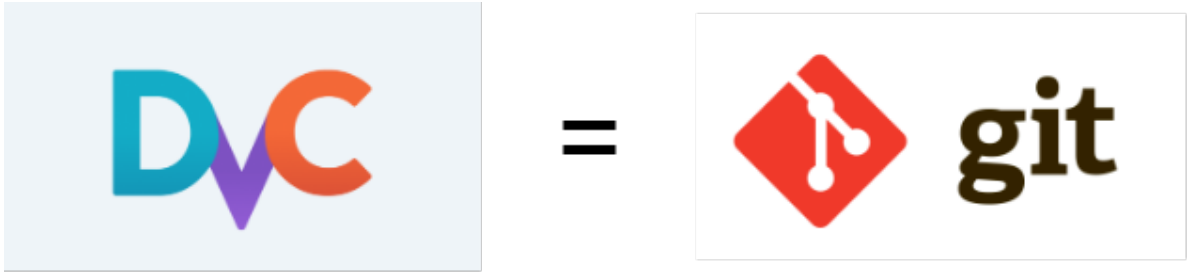
```
from models import LightningWrappedModel
from pytorch_lightning import Trainer

model = LightningWrappedModel(
    **hyperparameters
)
Trainer(
    max_epochs=100
).fit(
    model, train_loader, val_loader
)
```

Data Version Control



Data Version Control

"Créer des memes et faire des screenshots c'est quand même moins chiant que de devoir écrire un README.md"- Loïc Baret



Sans surprise : DvC c'est (grossièrement, ce n'est pas exactement le cas) un git pour les données. On y retrouve les mêmes mécanismes (pull, commit, push, etc...)

Autres outils intéressants

Rôle	Outils
Gérer le déploiement et le scaling	Kubernetes
Développement d'applications en environnement contrôllé	Docker
Exécution de scripts python en command line	Click
Génération de documentation	Sphinx