

## Cài đặt

Trước khi bắt đầu, bạn cần thiết lập môi trường của mình và cài đặt các gói thích hợp. Bộ dữ liệu được thử nghiệm trên Python 3.9+.

[!TIP]

Nếu bạn muốn sử dụng Bộ dữ liệu với TensorFlow hoặc PyTorch, bạn sẽ cần cài đặt chúng riêng. Tham khảo trang cài đặt TensorFlow hoặc trang cài đặt PyTorch để biết lệnh cài đặt cụ thể cho khung của bạn.

### Môi trường ảo

Bạn nên cài đặt Datasets trong môi trường ảo để mọi thứ gọn gàng, tránh tình trạng xung đột phụ thuộc.

1. Tạo và điều hướng đến thư mục dự án của bạn:

```
mkdir ~/dự án của tôi  
cd ~/dự án của tôi
```

2. Bắt đầu một môi trường ảo bên trong thư mục của bạn:

```
python -m venv .env
```

3. Kích hoạt và hủy kích hoạt môi trường ảo bằng các lệnh sau:

```
# Kích hoạt môi trường ảo  
nguồn .env/bin/kích hoạt  
  
# Vô hiệu hóa môi trường ảo  
nguồn .env/bin/hủy kích hoạt
```

Sau khi tạo môi trường ảo của mình, bạn có thể cài đặt Bộ dữ liệu trong đó.

pip

Cách đơn giản nhất để cài đặt Bộ dữ liệu là bằng pip:

tập dữ liệu cài đặt pip

Chạy lệnh sau để kiểm tra xem Bộ dữ liệu đã được cài đặt đúng chưa:

```
python -c "from datasets import load_dataset; print(load_dataset('rajpurkar/squad', split='train'))"
```

This command downloads version 1 of the Stanford Question Answering Dataset (SQuAD), tải phần đào tạo và in ví dụ đào tạo đầu tiên. Bạn nên xem:

```
{'answers': {'answer_start': [515], 'text': ['Saint Bernadette Soubirous']}, 'context': 'Architect
```

Âm thanh

Để làm việc với tập dữ liệu âm thanh, bạn cần cài đặt tính năng Âm thanh dưới dạng phần phụ thuộc bổ sung:

```
pip install datasets[audio]
```

Tầm nhìn

Để làm việc với tập dữ liệu hình ảnh, bạn cần cài đặt tính năng Hình ảnh dưới dạng phần phụ thuộc bổ sung:

```
pip install datasets[vision]
```

nguồn

Xây dựng Bộ dữ liệu từ nguồn cho phép bạn thực hiện các thay đổi đối với cơ sở mã. Để cài đặt từ nguồn, sao chép kho lưu trữ và cài đặt bằng các lệnh sau:

```
bản sao git https://github.com/huggingface/datasets.git
tập dữ liệu cd
cài đặt pip -e .
```

Một lần nữa, bạn có thể kiểm tra xem Bộ dữ liệu đã được cài đặt đúng cách hay chưa bằng lệnh sau:

```
python -c "from datasets import load_dataset; print(load_dataset('rajpurkar/squad', split='train'))"
```

chung cư

Bộ dữ liệu cũng có thể được cài đặt từ conda, một hệ thống quản lý gói:

```
cài đặt conda -c ôm mặt -c bộ dữ liệu conda-forge
```