Installation

Before you start, you'll need to setup your environment and install the appropriate packages. 🤗 Datasets is tested on **Python 3.9+**.

> [!TIP]
> If you want to use 🤗 Datasets with TensorFlow or PyTorch, you'll need to install them separately. Refer to the TensorFlow installation page or the PyTorch installation page for the specific install command for your framework.

Virtual environment

You should install 🤗 Datasets in a virtual environment to keep things tidy and avoid dependency conflicts.

1. Create and navigate to your project directory:

```
mkdir ~/my-project
cd ~/my-project
```

2. Start a virtual environment inside your directory:

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

3. Activate and deactivate the virtual environment with the following commands:

```
# Activate the virtual environment
source .env/bin/activate

# Deactivate the virtual environment
source .env/bin/deactivate
```

Once you've created your virtual environment, you can install 🤗 Datasets in it.

pip

The most straightforward way to install 🤗 Datasets is with pip:

```
pip install datasets
```

Run the following command to check if 🤗 Datasets has been properly installed:

```
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), loads the training split, and prints the first training example. You should see:

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

Audio

To work with audio datasets, you need to install the Audio feature as an extra dependency:

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

Vision

To work with image datasets, you need to install the Image feature as an extra dependency:

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

source

Building 🤗 Datasets from source lets you make changes to the code base. To install from the source, clone the repository and install with the following commands:

```
git clone https://github.com/huggingface/datasets.git
cd datasets
pip install -e .
```

Again, you can check if 🤗 Datasets was properly installed with the following command:

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

conda

🤗 Datasets can also be installed from conda, a package management system:

```
conda install -c huggingface -c conda-forge datasets
```