

## ▶ 학습 데이터 소개

## ① 데이터 수집 및 저장

### - Hugging Face Datasets

#### - Python\_code\_instructions\_18k\_alpaca

- Instruction
- Input
- Output
- Prompt

### - Column 소개

- Instruction (Task 지시사항)
- Input (Task 수행시 필요한 데이터)
- Output (Python 코드 출력)
- Prompt (질문 – 응답 텍스트)

### - 예시

- Instruction = 시퀀스 합계를 계산하는 함수 만들어줘
- Input "[1, 2, 3, 4, 5, 6, ...]"
- Output "def\_sum\_sequence(sequence)..."
- Prompt

작업에 대한 요청이 있습니다.

요청을 완성하는 응답을 작성하십시오..

Hugging Face

Search models, datasets, users...

Models Datasets

Datasets: iamtarun/python\_code\_instructions\_18k\_alpaca like 227

Tasks: Question Answering Text2Text Generation Text Generation Modalities: Text Formats: parquet

Libraries: Datasets pandas Croissant +1

Dataset card Viewer Files and versions Community

Dataset Viewer

Split (1) train · 18.6k rows

Search this dataset SQL Console

instruction	input	output	prompt
Create a function to calculate the sum of a sequence of integers.	[1, 2, 3, 4, 5]	# Python code def sum_sequence(sequence): sum = 0 for num in sequence: sum += num return sum	Below is an instruction that describes a task. Write a response that appropriately completes the...
Generate a Python code for crawling a website for a specific type of data.	website: www.example.com data to crawl: phone numbers	import requests import re def crawl_website_for_phone_numbers(website): response...	Below is an instruction that describes a task. Write a response that appropriately completes the...
Create a Python list comprehension to get the squared values of a list [1, 2, 3, 5, 8, 13].		[x*x for x in [1, 2, 3, 5, 8, 13]]	Below is an instruction that describes a task. Write a response that appropriately completes the...
Generate a python script to perform this action.	Given a string, remove all the consecutive duplicates from the string. Input: "AAABBBCCCD"	def remove_duplicates(string): result = "" prev = '' for char in string: if char != prev: result += char	Below is an instruction that describes a task. Write a response that appropriately completes the...
Write a python script to generates random numbers between 0 and 9 that are divisible by...		def generate_random_divisible_number(): import random while True: # Generate a random number...	Below is an instruction that describes a task. Write a response that appropriately completes the...
Write a Python code to get the third largest element in a given row.	[12, 13, 13, 45, 22, 99]	def third_largest(lst): if len(lst) < 3: return distinct = [] for i in lst: if i not in distinct...	Below is an instruction that describes a task. Write a response that appropriately completes the...
Generate a REST API with Python and Flask that allows users to create, read, update, and...	Not applicable	from flask import Flask, request from flask_sqlalchemy import SQLAlchemy app = _	Below is an instruction that describes a task. Write a response that appropriately completes the...
Create a Python function that takes in a string and a list of words and returns true if the...	'This is a test', ['test', 'this', 'is']	def contains_words(input_string, words): for word in words: if word not in input_string: return False	Below is an instruction that describes a task. Write a response that appropriately completes the...
What should this Python program do?	def is_palindrome(s): 'Write your code here. ...	def is_palindrome(s): 'This function returns True if the string s is a palindrome and False if it...	Below is an instruction that describes a task. Write a response that appropriately completes the...

Hugging Face is way more fun with friends and colleagues! Join an organization Dismiss this message

Datasets: iamtarun/python\_code\_instructions\_18k\_alpaca like 227

Dataset card Viewer Files and versions Community

Split (1) train · 18.6k rows

Search this dataset SQL Console

instruction	input	output	prompt
Create a function to calculate the sum of a sequence of integers.	[1, 2, 3, 4, 5]	# Python code def sum_sequence(sequence): sum = 0 for num in sequence: sum += num return sum	Below is an instruction that describes a task. Write a response that appropriately completes the...
Generate a Python code for crawling a website for a specific type of data.	website: www.example.com data to crawl: phone numbers	import requests import re def crawl_website_for_phone_numbers(website): response...	Below is an instruction that describes a task. Write a response that appropriately completes the...
Create a Python list comprehension to get the squared values of a list [1, 2, 3, 5, 8, 13].		[x*x for x in [1, 2, 3, 5, 8, 13]]	Below is an instruction that describes a task. Write a response that appropriately completes the...
Generate a python script to perform this action.	Given a string, remove all the consecutive duplicates from the string. Input: "AAABBBCCCD"	def remove_duplicates(string): result = "" prev = '' for char in string: if char != prev: result += char	Below is an instruction that describes a task. Write a response that appropriately completes the...
Write a python script to generates random numbers between 0 and 9 that are divisible by...		def generate_random_divisible_number(): import random while True: # Generate a random number...	Below is an instruction that describes a task. Write a response that appropriately completes the...
Write a Python code to get the third largest element in a given row.	[12, 13, 13, 45, 22, 99]	def third_largest(lst): if len(lst) < 3: return distinct = [] for i in lst: if i not in distinct...	Below is an instruction that describes a task. Write a response that appropriately completes the...
Generate a REST API with Python and Flask that allows users to create, read, update, and...	Not applicable	from flask import Flask, request from flask_sqlalchemy import SQLAlchemy app = _	Below is an instruction that describes a task. Write a response that appropriately completes the...
Create a Python function that takes in a string and a list of words and returns true if the...	'This is a test', ['test', 'this', 'is']	def contains_words(input_string, words): for word in words: if word not in input_string: return False	Below is an instruction that describes a task. Write a response that appropriately completes the...
What should this Python program do?	def is_palindrome(s): 'Write your code here. ...	def is_palindrome(s): 'This function returns True if the string s is a palindrome and False if it...	Below is an instruction that describes a task. Write a response that appropriately completes the...

## - Hugging Face Datasets

- Python\_code\_instructions\_18k\_alpaca
- 출처
  - Hugging Face Dataset
- 수집 방법
  - 프로젝트에 맞는 Code 생성을 위해 Hub에서 가장 최적화된 데이터셋에서 채택함

## - 데이터 출처에 대한 법적 규제

- 하단의 이미지는 Hugging Face Docs Datasets의 내용입니다.

## - 데이터 수집 목적

- 모델의 학습을 위해서 데이터 수집을 해야 했고 방법에 기술 되어 있듯이 Code 생성을 위해 (메인인 Python을 위해) 최대한 잘 정제되어있는 데이터셋을 가져옴

### Datasets



😊 Datasets is a library for easily accessing and sharing datasets for Audio, Computer Vision, and Natural Language Processing (NLP) tasks.

Load a dataset in a single line of code, and use our powerful data processing methods to quickly get your dataset ready for training in a deep learning model. Backed by the Apache Arrow format, process large datasets with zero-copy reads without any memory constraints for optimal speed and efficiency. We also feature a deep integration with the [Hugging Face Hub](#), allowing you to easily load and share a dataset with the wider machine learning community.

Find your dataset today on the [Hugging Face Hub](#), and take an in-depth look inside of it with the live viewer.

#### Tutorials

Learn the basics and become familiar with loading, accessing, and processing a dataset. Start here if you are using 😊 Datasets for the first time!

#### How-to guides

Practical guides to help you achieve a specific goal. Take a look at these guides to learn how to use 😊 Datasets to solve real-world problems.

#### Conceptual guides

High-level explanations for building a better understanding about important topics such as the underlying data format, the cache, and how datasets are generated.

#### Reference

Technical descriptions of how 😊 Datasets classes and methods work.