Top 18 Python Libraries for Data Scientists

(DOWNLOAD THE FREE PDF)

Scrapy

Zyte created Scrapy in 2008, a Python library for web scraping. The library includes a broad variety of capabilities, including data extraction from websites or multiple pages, data export to several formats, and more.

Here is the official page of Scrapy (https://scrapy.org/).

BeautifulSoup

Leonard Richardson created BeautifulSoup in 2004 as a Python toolkit to extract data from HTML and XML files.

It is compatible with request and other scraping libraries. The various functions of BeautifulSoup include browsing and searching through HTML documents as well as extracting data from tags and attributes.

Here is the official page of BeautifulSoup (https://www.crummy.com/software/BeautifulSoup/bs4/doc/).

Numpy

NumPy is a numerical computing library for Python. It was created by Travis Oliphant in 2005. NumPy provides functions for performing operations on arrays, including mathematical, logical, shape manipulation, basic linear algebra, basic statistical operations, and more.

Here is the official page of NumPy (https://numpy.org/).

SciPy

SciPy is a scientific computing library for Python. It was created by Eric Jones and Travis Oliphant in 2001. SciPy builds on top of NumPy and provides a wide range of numerical and scientific computing functions such as numerical integration, optimization, signal and image processing, linear algebra, statistics, and more.

Here is the official page of SciPy (https://scipy.org/).

math

It is a built-in Python library that offers mathematical functions. There are functions for more complex mathematical operations like trigonometric functions, logarithms, and exponentials, as well as functions for simpler mathematical operations like addition, subtraction, multiplication, and division.

Here is the official page of math (https://docs.python.org/3/library/math.html).

scikit-learn

scikit-learn is a machine-learning library developed by David Cournapeau in 2007.

It has many different features to build classification, regression, and clustering algorithms.

Here is the official page of scikit-learn (https://scikit-learn.org/).

Keras

François Chollet developed the machine-learning library Keras in 2015 for using in Machine Learning.

It offers various capabilities for creating and improving neural networks, as well as for processing images and texts, and more.

Here is the official page of Keras (https://keras.io/).

PyTorch

PyTorch is a machine-learning library developed by MetaAI in 2016. You can do many things with PyTorch, like building deep learning models, image classification, natural language processing, and more.

Here is the official page of PyTorch (https://pytorch.org/).

Tensorflow

Tensorflow is a machine learning library developed by Google in 2015. You can do many things with TensorFlow, like image classification, natural language processing, or generative modeling.

Here is the official page of Tensorflow (https://www.tensorflow.org/).

Matplotlib

Matplotlib is a popular python data visualization library that enables users to create a range of visualizations in 2D.

It was developed by John D. Hunter in 2002.

Here is the official web page of Matplotlib (https://matplotlib.org/).

seaborn

seaborn is a data visualization library for Python. It was created by Michael Waskom in 2014. seaborn is also built on top of Matplotlib, and often they worked together.

Here is the official web page of seaborn (https://seaborn.pydata.org/).

django

django is a web framework, developed in 2003 by Python programmers Adrian Holovaty and Simon Willison.

Here is the official web page of django (https://www.djangoproject.com/).

Flask

Flask is a micro web framework. Like django, by using Flask you can also develop your own API or Web app. It was founded in 2004, by Armin Ronacher of Pocoo.

Here is the official web page of Flask (https://flask.palletsprojects.com/en/2.2.x/).

FastAPI

FastAPI is a web framework, that allows users to create applications quickly, developed in 2018 by Sebastián Ramírez.

Here is the official web page of FastAPI (https://fastapi.tiangolo.com/).

plotly

plotly is a data visualization library for Python and other programming languages. It was created by Alex Johnson, Chris Parmer, Jack Parmer, and others in 2012.

plotly is often used for its interactive visualizations, including line plots, scatter plots, bar plots, and more.

Here is the official web page of plotly (https://plotly.com/).

pandas

pandas is a data manipulation and analysis library for Python, but it's also heavily used in data visualization.

It was created by Wes McKinney in 2008. pandas provide functions for reading and writing data, handling missing data, and performing data analysis tasks such as aggregation and reshaping.

Due to its functionalities, pandas are equally popular when it comes to manipulating data, performing mathematical operations, and visualizing data.

Here is the official web page of Pandas (https://pandas.pydata.org/docs/user_guide/index.html).

Selenium

Thoughtworks created the Selenium Python package in 2002 to be used for browser automation, testing, and scraping. It includes a wide range of functions, including the ability to fill out forms and automate browser actions. It can also be used to scrape websites.

Here is the official page of Selenium (https://selenium-python.readthedocs.io/).

Requests

It was created in 2011 by Kenneth Reitz.

Requests can be used to interact with APIs, send HTTP requests, and handle HTTP errors.

Here is the official page of Requests (https://requests.readthedocs.io/).