

Top 20 Python Libraries for Data Science

 SciPy	Data wrangling
 NumPy	Data wrangling
 pandas	Data wrangling
 StatsModels Statistics in Python	Statistics

 ockit learn	Machine learning
XGBoost LightGBM CatBoost	Machine learning
 eli5	Machine learning

 matplotlib	Visualization
 Bokeh	Visualization
 plotly	Visualization
 Seaborn	Visualization
 pydot	Visualization

 Natural Language Toolkit	NLP
 spaCy	NLP
 gensim	NLP
 Scrapy	Data scraping

 TensorFlow	Deep learning
 PYTORCH	Deep learning
 Keras	Deep learning
dist-keras elephas spark-deep-learning	Distributed deep learning

Core Libraries & Statistics

NumPy (<http://www.numpy.org/>)

- It is intended for processing large multidimensional arrays and matrices, and an extensive collection of high-level mathematical functions and implemented methods makes it possible to perform various operations with these objects.



SciPy (<https://scipy.org/scipylib/>)

- It is based on NumPy and therefore extends its capabilities. SciPy main data structure is again a multidimensional array, implemented by Numpy. The package contains tools that help with solving linear algebra, probability theory, integral calculus and many more tasks.



Pandas (<https://pandas.pydata.org/>)

- Pandas provides high-level data structures and a vast variety of tools for analysis. The great feature of this package is the ability to translate rather complex operations with data into one or two commands. Pandas contains many built-in methods for grouping, filtering, and combining data, as well as the time-series functionality.



Visualization

Matplotlib (<https://matplotlib.org/index.html>)

- Matplotlib is a low-level library for creating two-dimensional diagrams and graphs. With its help, you can build diverse charts, from histograms and scatterplots to non-Cartesian coordinates graphs. Moreover, many popular plotting libraries are designed to work in conjunction with matplotlib.

Seaborn (<https://seaborn.pydata.org/>)

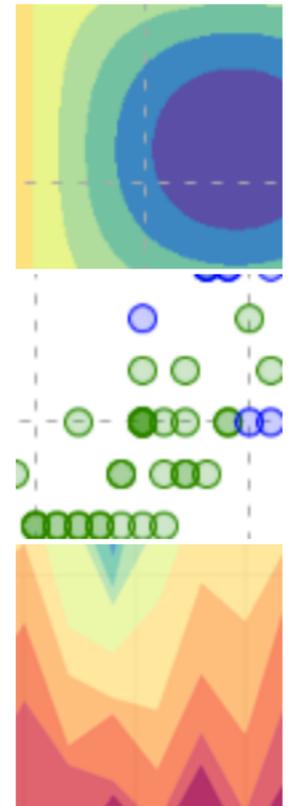
- Seaborn is essentially a higher-level API based on the matplotlib library. It contains more suitable default settings for processing charts. Also, there is a rich gallery of visualizations including some complex types like time series, jointplots, and violin diagrams.

Plotly (<https://plot.ly/python/>)

- Plotly is a popular library that allows you to build sophisticated graphics easily. The package is adapted to work in interactive web applications. Among its remarkable visualizations are contour graphics, ternary plots, and 3D charts.

Bokeh (<https://bokeh.pydata.org/en/latest/>)

- The Bokeh library creates interactive and scalable visualizations in a browser using JavaScript widgets. The library provides a versatile collection of graphs, styling possibilities, interaction abilities in the form of linking plots, adding widgets, and defining callbacks, and many more useful features.



Data Mining & Machine Learning

Scikit-learn (<https://scikit-learn.org/stable/>)

- This Python module based on NumPy and SciPy is one of the best libraries for working with data. It provides algorithms for many standard machine learning and data mining tasks such as clustering, regression, classification, dimensionality reduction, and model selection.



PyFim (<http://www.borgelt.net/pyfim.html>)

- PyFIM is an extension module that makes several frequent item set mining implementations available as functions. Currently apriori, eclat, fpgrowth, sam, relim, carpenter, ista, accretion and apriacc are available as functions, although the interfaces do not offer all of the options of the command line program.



Eli5 (<https://eli5.readthedocs.io/en/latest/>)

- Often the results of machine learning models predictions are not entirely clear, and this is the challenge that eli5 library helps to deal with. It is a package for visualization and debugging machine learning models and tracking the work of an algorithm step by step. It provides support for scikit-learn, XGBoost, LightGBM, lightning, and sklearn-crfsuite libraries and performs the different tasks for each of them.

Weight	Feature
+11.493	<BIAS>
+6.280	x
-14.140	y

Deep Learning

TensorFlow (<https://www.tensorflow.org/>)

- TensorFlow is a popular framework for deep and machine learning, developed in Google Brain. It provides abilities to work with artificial neural networks with multiple data sets. Among the most popular TensorFlow applications are object identification, speech recognition, and more.



PyTorch (<https://pytorch.org/>)

- PyTorch is a large framework that allows you to perform tensor computations with GPU acceleration, create dynamic computational graphs and automatically calculate gradients. Above this, PyTorch offers a rich API for solving applications related to neural networks.



Keras (<https://keras.io/>)

- Keras is a high-level library for working with neural networks, running on top of TensorFlow, Theano, and now as a result of the new releases. It simplifies many specific tasks and greatly reduces the amount of monotonous code. However, it may not be suitable for some complicated things.



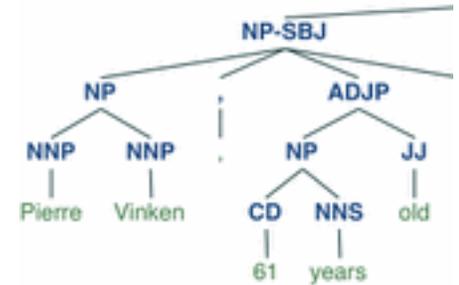
Dist-keras (<https://joerihermans.com/work/distributed-keras/>)

- dist-keras and others are gaining popularity and developing rapidly, and it is very difficult to single out one of the libraries since they are all designed to solve a common task. These packages allow you to train neural networks based on the Keras library directly with the help of Apache Spark.

Natural Language Processing & Data Scraping

NLTK (<https://www.nltk.org/>)

- NLTK is a set of libraries, a whole platform for natural language processing. With the help of NLTK, you can process and analyze text in a variety of ways, tokenize and tag it, extract information, etc. NLTK is also used for prototyping and building research systems.



Gensim (<https://radimrehurek.com/gensim/>)

- Gensim is a Python library for robust semantic analysis, topic modeling and vector-space modeling, and is built upon Numpy and Scipy. It provides an implementation of popular NLP algorithms, such as word2vec. Although gensim has its own models.wrappers.fasttext implementation, the fasttext library can also be used for efficient learning of word representations.



Scrapy (<https://scrapy.org/>)

- Scrapy is a library used to create spiders bots that scan website pages and collect structured data. In addition, Scrapy can extract data from the API. The library happens to be very handy due to its extensibility and portability.



Thank you

<https://www.kdnuggets.com/2018/06/top-20-python-libraries-data-science-2018.html>