



# Open Source

We are proud to distribute and contribute to a variety of open source projects.



## Support OSS

You can help support ongoing innovation on projects like these in the open source community. NumFOCUS is a non-profit organization supporting numerous open-source projects in support of an inclusive scientific and research community making impactful discoveries for a better world. Your donation goes directly to them and supports their work.

This website uses cookies to ensure you get the best experience on our website.

[Privacy Policy](#)

[Accept](#)



## How to get started

More than 250 of the most commonly used open-source data science and machine learning packages are automatically installed when you download [Anaconda Individual Edition](#), and thousands of others can be installed by simply typing `conda install [package-name]`.

## The Fundamentals

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)

Jupyter is an open-source project created to support interactive data science and scientific computing across programming languages. Jupyter offers a web-based environment for working with notebooks containing code, data, and text. Jupyter notebooks are the standard workspace for most Python data scientists.



A library for tabular data structures, data analysis, and data modeling tools, including built-in plotting using Matplotlib. pandas aims to be the fundamental high-level building block for doing practical, real world data analysis in Python



The SciPy library consists of a specific set of fundamental scientific and numerical tools for Python that data scientists use to build their own tools and programs. It provides many user-friendly and efficient numerical routines, such as routines for numerical integration, interpolation, optimization, linear algebra, and statistics.

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)

A core package for scientific computing with Python. Numpy enables array formation and basic operations with arrays. Numpy is used for indexing and sorting but can also be used for linear algebra and other operations. Many other data-science libraries for Python are built on NumPy internally, including Pandas and SciPy.

# Machine Learning

## K Keras

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

Accept

TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.

## PYTORCH

An open-source deep learning framework using GPUs and CPUs that consists of fundamental tools and libraries for Python AI and machine learning development.



A powerful and versatile machine learning library for machine learning basics like classification, regression, and clustering. It includes both supervised and unsupervised ML algorithms with important functions like cross-validation and feature extraction. Scikit-learn is the most frequently downloaded machine learning library.

This website uses cookies to ensure you get the best experience on our website.

[Privacy Policy](#)

[Accept](#)

# Data Visualization



Matplotlib is the most well-established Python data visualization tool, focusing primarily on two-dimensional plots (line charts, bar charts, scatter plots, histograms, and many others). It works with many GUI interfaces and file formats, but has relatively limited interactive support in web browsers.



Bokeh is an interactive visualization library for modern web browsers. It provides elegant, concise construction of versatile graphics, and affords high-performance interactivity over large or streaming datasets. Bokeh can help anyone who would like to quickly and easily make interactive plots, dashboards, and data applications.



This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)



HoloViz is an Anaconda project to simplify and improve Python-based visualization by adding high-performance server-side rendering (Datashader), simple plug-in replacement for static visualizations with interactive Bokeh-based plots (hvPlot), and declarative high-level interfaces for building large and complex systems (HoloViews and Param).

# Dashboarding



Panel is an open-source Python library that lets you create custom interactive web apps and dashboards by connecting user-defined widgets to

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)



Dash is a productive Python framework for building web applications. Through a couple of simple patterns, Dash abstracts away all of the technologies and protocols that are required to build an interactive web-based application.



Voilà turns Jupyter notebooks into standalone web applications. Unlike the usual HTML-converted notebooks, each user connecting to the Voilà tornado application gets a dedicated Jupyter kernel which can execute the callbacks to changes in Jupyter interactive widgets.



Streamlit is an open-source Python library that makes it easy to build beautiful custom web-apps for machine learning and data science. It runs on a simple and powerful app model that lets you build rich UIs incredibly quickly.

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)

# Image Processing

## Pillow

Pillow (a “friendly fork” of the older PIL library) is a Python imaging library and a general image processing tool with support for opening, manipulating, and saving images in many different file formats.



scikit-image  
image processing in python

Scikit-Image is an open-source Python package containing a collection of image-processing algorithms, including segmentation, geometric transformations, color space manipulation, and feature detection. It uses NumPy arrays as image objects.



This website uses cookies to ensure you get the best experience on our website.

[Privacy Policy](#)

[Accept](#)

infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products.

# Scalable Computing



Numba is a high-performance Python compiler. It makes Python faster and optimizes the performance of Numpy arrays, reaching the speed of FORTRAN and C without a an additional compilation step.



This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)

pandas and NumPy API, making it more intuitive for Python data scientists.

## RAPIDS

The RAPIDS data science framework is a collection of libraries for running end-to-end data science pipelines completely on the GPU. The interaction is designed to have a familiar look and feel to working in Python, but utilizes optimized NVIDIA® CUDA® primitives and high-bandwidth GPU memory under the hood.



A fault-tolerant cluster computing framework and interface for programming clusters launched by UC Berkeley. Developed for Java/Hadoop ecosystem but with support for Python. PySpark is the Python API for Spark.

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)

# Data Pipelines / ETL



An open-source workflow automation tool by Apache for creating data workflows, scheduling tasks and monitoring results. It integrates with multiple cloud providers, including AWS, Azure, and Google Cloud.



A data ingest/loading library for a wide variety of file formats and data services, with hierarchical cataloguing, searching, and interactivity with remote storage platforms under a single interface.

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)

# Natural Language Processing (NLP)

## NLTK

An open-source Python natural language toolkit for symbolic and statistical NLP. It includes a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning in multiple languages.

## gensim

A Python library for topic modeling, document indexing, and similarity retrieval for large bodies of text with efficient multicore implementations of NLP algorithms.

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

Accept

spaCy is an open-source Python library for NLP and one of the fastest, if not the fastest, syntactic parser. spaCy excels at large-scale information extraction tasks. It's written from the ground up in carefully memory-managed Cython.

## Looking Ahead: AI Frontiers We're Watching



ONNX

An open neural network exchange making machine learning models portable between frameworks and platforms. Microsoft and Facebook started this community in 2017 to create an open ecosystem for interchangeable models.

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)

A burgeoning project by open-source developers at Microsoft. FairLearn is a Python package for assessing fairness and mitigating unfairness in ML models and AI systems.

## **AI Fairness (360)**

A comprehensive open-source Python toolkit of metrics that checks for and measures bias in datasets and ML models. It also included algorithms to mitigate bias. This toolkit was developed by IBM's open-source team.

## **InterpretML**

An open-source Python package that makes it easy to compare algorithms for interpretability.

It provides a "scikit-learn style uniform API" and includes an interactive visualization platform and dashboard so data scientists can compare algorithms with ease.

## **LIME**

This website uses cookies to ensure you get the best experience on our website.  
[Privacy Policy](#)

[Accept](#)

# Supercharge your data science efforts with **Anaconda**.

[Get Started](#)



By data scientists, for **data scientists**

## PRODUCTS

[Individual Edition](#)

[Commercial Edition](#)

## SOLUTIONS

[Use Cases](#)

[Industries](#)

This website uses cookies to ensure you get the best experience on our website.

[Privacy Policy](#)

[Accept](#)