# 10 BEST PYTHON LIBRARIES FOR MACHINE LEARNING

With the increase in the markets for smart products, auto-pilot cars and other smart products, the ML industry is on a rise. Machine Learning is also one of the most prominent tools of cost-cutting in almost every sector of industry nowadays.

ML libraries are available in many programming languages, but python being the most user-friendly and easy to manage language, and having a large developer community, is best suited for machine learning purposes and that's why many ML libraries are being written in Python.

Also, the python works seamlessly with C and C++ and so, the already written libraries in C/C++ can be easily extended to Python. In this tutorial, we will be discussing the most useful and **best machine-learning libraries in Python** programming language.

## 1. TENSORFLOW :

Website: https://www.tensorflow.org/

GitHub Repository: https://github.com/tensorflow/tensorflow

Developed By: Google Brain Team

Primary Purpose: Deep Neural Networks



TensorFlow is a library developed by the Google Brain team for the primary purpose of Deep Learning and Neural Networks. It allows easy distribution of work onto multiple CPU cores or GPU cores, and can even distribute the work to multiple GPUs. TensorFlow uses Tensors for this purpose. Tensors can be defined as a container that can store N-dimensional data along with its linear operations. Although it is production-ready and does support reinforcement learning along with Neural networks, it is not commercially supported which means any bug or defect can be resolved only by community help.

## 2. NUMPY:

Website: https://numpy.org/
Github Repository: https://github.com/numpy/numpy
Developed By: Community Project (originally authored by Travis Oliphant)
Primary purpose: General Purpose Array Processing



Created on the top of an older library Numeric, the Numpy is used for handling multi-dimensional data and intricate mathematical functions. Numpy is a fast computational library that can handle tasks and functions ranging from basic algebra to even Fourier transforms, random simulations, and shape manipulations. This library is written in C language, which gives it an edge over standard python built-in sequencing. Numpy arrays are better than pandas series in the term of indexing and Numpy works better if the number of records is less than 50k. The NumPy arrays are loaded into a single CPU which can cause slowness in processing over the new alternatives like Tensorflow, Dask, or JAX, but still, the learning of Numpy is very easy and it is one of the most popular libraries to enter into the Machine Learning world.
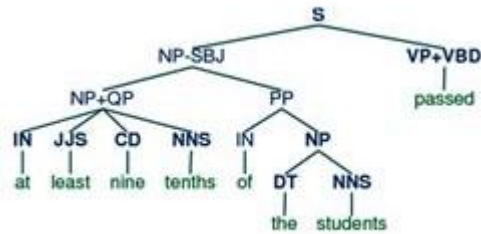
# 3. NATURAL LANGUAGE TOOLKIT (NLTK):

Website:  https://www.nltk.org/
Github Repository:https://github.com/nltk/nltk
Developed By: Team NLTK
Primary Purpose: Natural Language Processing



NLTK is the widely used library for Text Classification and Natural Language Processing. It performs word Stemming, Lemmatizing, Tokenization, and searching a keyword in documents. The library can be further used for sentiment analysis, understanding movie reviews, food reviews, text-classifier, checking and censoring the vulgarised words from comments, text mining, and many other human language-related operations. The wider scope of its uses includes AI-powered chatbots which need text processing to train their models to identify and also create sentences important for machine and human interaction in the upcoming future.

# 4.PANDAS

Website: https://pandas.pydata.org/

Github Repository: https://github.com/pandas-dev/pandas

Developed By: Community Developed (Originally Authored by Wes McKinney)

Primary Purpose: Data Analysis and Manipulation



The Library is written in Python Web framwork and is used for data manipulation for numerical data and time series. It uses data frames and series to define three-dimensional and two-dimensional data respectively. It also provides options for indexing large data for quick search in large datasets. It is well known for the capabilities of data reshaping, pivoting on user-defined axis, handling missing data, merging and joining datasets, and the options for data filtrations. Pandas is very useful and very fast with large datasets. Its speed exceeds that of Numpy when the records are more than 50k. It is the best library when it comes to data cleaning because it provides interactiveness like excel and speed like Numpy. It is also one of the few ML libraries that can deal with DateTime without any help from any external libraries and also with a bare minimum code with **python coding standard** quality. As we all know the most significant part of data analysis and ML is the data cleaning, processing, and analyzing where Pandas helps very effectively.

## 5. SCIKIT-LEARN:

Website: https://scikit-learn.org/
Github Repository: https://github.com/scikit-learn/scikit-learn
Developed By: SkLearn.org
Primary Purpose: Predictive Data Analysis and Data Modeling



Scikit-learn is mostly focused on various data modeling concepts like regression, classification, clustering, model selections, etc. The library is written on the top of Numpy, Scipy, and matplotlib. It is an open-source and commercially usable library that is also very easy to understand. It has easy integrability which other ML libraries like Numpy and Pandas for analysis and Plotly for plotting the data in a graphical format for visualization purposes. This library helps both in supervised as well as unsupervised learnings.

## 6. KERAS:

Website: https://keras.io/
Github Repository: https://github.com/keras-team/keras
Developed By: various Developers, initially by Francois Chollet
Primary purpose: Focused on Neural Networks



Keras provides a Python interface of Tensorflow Library especially focused on AI

neural networks. The earlier versions also included many other backends like

Theano, Microsoft cognitive platform, and PlaidMl. Keras contains standard blocks

of commonly used neural networks, and also the tools to make image and text

processing faster and smoother. Apart from standard blocks of neural networks, it

also provides re-occurring neural networks.

# 7. PYTORCH:

Website: https://pytorch.org/
Github Repository: https://github.com/pytorch/pytorch
Developed By: Facebook AI Research lab (FAIR)
Primary purpose: Deep learning, Natural language Processing, and Computer Vision



Pytorch is a Facebook-developed ML library that is based on the Torch Library (an open-source ML library written in Lua Programming language). The project is written in Python Web Development, C++, and CUDA languages. Along with Python, PyTorch has extensions in both C and C++ languages. It is a competitor to Tensorflow as both of these libraries use tensors but it is easier to learn and has better integrability with Python. Although it supports NLP, but the main focus of the library is only on developing and training deep learning models only.
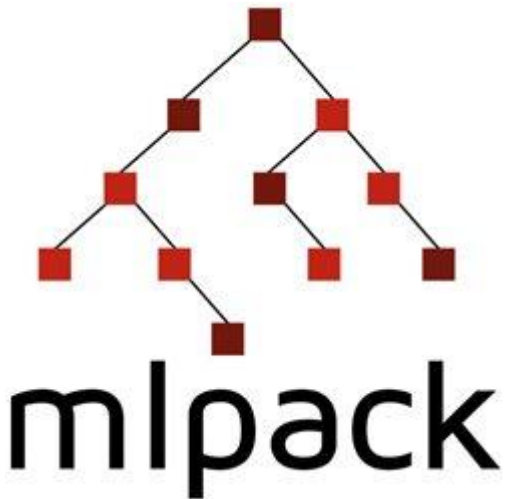
## 8. MLPACK:

Github Repository:  https://github.com/mlpack/mlpack

Developed By: Community, supported by Georgia Institute of technology

Primary purpose: Multiple ML Models and Algorithms



MlPack is mostly C++-based ML library that has bindings to Python other languages including R programming, Julia, and GO. It is designed to support almost all famous ML algorithms and models like GMMs, K-means, least angle regression, Linear regression, etc. The main emphasis while developing this library was on making it a fast, scalable, and easy-to-understand as well as an easy-to-use library so that even a coder new to programming can understand and use it without any problem. It comes under a BSD license making it approachable as both open source and also proprietary software as per the need.

## 9. OPENCV:

Website: https://opencv.org/
Github Repository: https://github.com/opencv/opencv
Developed By: initially by Intel Corporation
Primary purpose: Only focuses on Computer Vision



OpenCV is an open-source platform dedicated to computer vision and image processing. This library has more than 2500 algorithms dedicated to computer vision and ML. It can track human movements, detect moving objects, extract 3d models, stitch images together to create a high-resolution image, exploring the AR possibilities. It is used in various CCTV monitoring activities by many governments, especially in China and Isreal. Also, the major camera companies in the world use OpenCv for making their technology smart and user-friendly.

## 10. MATPLOTLIB:

Website: https://matplotlib.org/
Github Repository: https://github.com/matplotlib/matplotlib
Developed By: Micheal Droettboom, Community
Primary purpose: Data Visualization



Matplotlib is a library used in Python for graphical representation to understand the data before moving it to data-processing and training it for Machine learning purposes. It uses python GUI toolkits to produce graphs and plots using object-oriented APIs. The Matplotlib also provides a MATLAB-like interface so that a user can do similar tasks as MATLAB. This library is free and open-source and has many extension interfaces that extend matplotlib API to various other libraries.