Top 10 python packages for Data Analysis

Python packages are a collection of modules. Modules that are related to each other are mainly put in the same package. When a module from an external package is required in a program, that package can be imported and its modules can be put to use. Python packages streamline many significant processes, like analyzing and visualizing data, building machine learning models, capturing unstructured data from the web, and processing image and text information efficiently. Here is the list of top 10 python packages for data scientists.

TensorFlow: TensorFlow is one of the most famous machine learning libraries for some very good reasons. It specializes in numerical computation using data flow graphs. It works like a computational library for writing new algorithms that involve a large number of tensor operations.

NumPy: NumPy is the primary tool for scientific computing in Python. It combines the flexibility and simplicity of Python with the speed of languages like C and Fortran. It is a valuable Python package for a variety of general-purpose programming tasks.

SciPy: SciPy library contains modules for optimization, linear algebra, integration, and statistics. It is a gigantic library of data science packages mainly focused on mathematics, science, and engineering. SciPy uses NumPy arrays as the basic data structure and comes with modules for various commonly used tasks in scientific programming.

Pandas: Pandas is an ML library in Python that provides data structures of high-level and a wide variety of tools for analysis. It is known as a fast, efficient, and easy-to-use tool for data analysis and manipulation. It works with data frame objects; a data frame is a dedicated structure for two-dimensional data.

Matplotlib: Matplotlib is a Python 2D plotting library that makes it easy to produce cross-platform charts and figures. It is to create basic graphs like line plots, histograms, scatter plots, bar charts, and pie charts.

Keras: Keras is built for fast experimentation. It's capable of running on top of other frameworks. It provides an easier mechanism to express neural networks. Keras contains numerous implementations of commonly used neural network building blocks such as layers, objectives, activation functions, and optimizers.

SciKit-Learn: Scikit-Learn has such a gentle learning curve, even the people on the business side of an organization can use it. It is considered one of the best libraries for working with complex data. It contains numerous algorithms for implementing standard machine learning and data mining tasks like reducing dimensionality, classification, regression, clustering, and model selection.

PyTorch: PyTorch is the largest machine learning library that allows developers to perform tensor computations wan with the acceleration of GPU, create dynamic computational graphs, and calculate gradients automatically. It builds dynamic neural networks on a tape-based autograd system.

Caffe: Caffe stands for Convolutional Architecture for Fast Feature Embedding. It is one of the fastest implementations of a convolutional network, making it ideal for image recognition. Caffe's image processing is quite astounding.

Theano: Theano is a Python library that allows you to define, optimize, and efficiently evaluate mathematical expressions involving multi-dimensional arrays. It is one of the earliest open-source software libraries for deep-learning development. It's best for high-speed computation.