Python libraries for data processing

Data processing services are available in various encodings, including CSV, XML, HTML, SQL, and JSON. Each situation requires a unique processing format. There are numerous programming languages. Python is frequently recommended as a viable alternative for machine learning applications due to its implementation of major libraries and cutting-edge technologies. Machine learning is built on data processing, and model success is highly dependent on the ability to read and transform data into the format required for the task at hand. Let us examine the various Python libraries in terms of the data types they provide.

Below, we have covered the Python libraries used for processing different types of data:

Tabular Data

Most of the large data is available in the tabular format, with rows referring to records and columns corresponding to features. Pandas in Python can handle such type data very perfectly. The advent of tabular data has evolved into a full-featured library that can handle both series and tabular data.

Text data

First, it's worth noting Python's extensive built-in text-processing capabilities. However, many natural language processing techniques, such as tokenization and lemmatization, may be done using NLTK. Along with that, Spacy is a good choice for advanced natural language processing and optimised pipelines.

Audio and musical data

Audio processing is enabled via libraries like librosa and essentia. Mido and pretty midi are good choices for symbolic music, like MIDI. Finally, music21 is a sophisticated library targeted at musicology analysis.

Images

Pillow is an image processing library in Python. Opencv is a computer vision library that can process videos or camera data. Because of its vast range of supported formats, imageio can give image data to the python script.

Python, in particular, is a highly regarded data processing language for a variety of reasons, including the following:

- Prototypes and experimentation with code are incredibly simple. Processing data, especially from less-than-clean sources, necessitates a great deal of tweaking, back and forth, and a struggle to capture all options.
- Python3 significantly improved multi-language support by making every string in the system UTF-8, which enables the processing of data encoded in different character sets by different languages.
- The standard library is quite strong and packed with essential modules that provide native support for common file types such as CSV files, zip files, and databases.
- The Python third-party library is enormous, and it has a wealth of excellent modules that enable it to increase the capabilities of a programme. There are also modules for geospatial data analysis, creating command-line interfaces, graphical interfaces, parsing data, and everything in between.
- Jupyter Notebooks allows you to execute code and receive immediate feedback. Python is quite agnostic about the development environment required, allowing it to function with anything from a simple text editor to more complex alternatives such as Visual Studio.

Python Libraries for Data mining

- NumPy (Numerical Python)
    - N dimensional array
    - Basic linear algebra functions, Fourier transforms, and advanced random number capabilities.
- SciPy (Scientific Python)
    - Variety of high level science and engineering modules.
    - Discrete fourier transform, linear algebra, optimization and sparse matrices
- Matplotlib
    - For plotting vast variety of graphs (histograms, line plots, heat plots, and so on)
- Pandas
    - Structured data operations and manipulations.
    - Extensively used for data manipulation.

Python libraries for data visualization

Python programming language has different types of libraries for all kind of projects. Likewise, python has various libraries for visualization of Data, so that user can understand the dataset in very detailed way and analyze it properly.

Each library of visualization has its own specification. Using the particular libraries for specific task helps the user to complete the task in more easy and accurate way. Some liberates work better than the others. In this article we will discuss pros and cons of the libraries to understand which library would be better for data visualization.

The Libraries for Data Visualization in Python programming are given below:

- Matplotlib
- Ggplot
- Pygal
- Missingno
- Seaborn
- Plotly
- Gleam
- Leather
- Geoplotlib
- Bokeh
- Folium