**Pandas Vs NumPy**

**What is Pandas?**

Pandas is defined as an open-source library that provides high-performance data manipulation in Python. It is built on top of the NumPy package, which means Numpy is required for operating the Pandas. The name of Pandas is derived from the word Panel Data, which means an Econometrics from Multidimensional data. It is used for data analysis in Python and developed by Wes McKinney in 2008.

Before Pandas, Python was capable for data preparation, but it only provided limited support for data analysis. So, Pandas came into the picture and enhanced the capabilities of data analysis. It can perform five significant steps required for processing and analysis of data irrespective of the origin of the data, i.e., load, manipulate, prepare, model, and analyse.

**What is NumPy?**

NumPy is mostly written in C language, and it is an extension module of Python. It is defined as a Python package used for performing the various numerical computations and processing of the multidimensional and singledimensional array elements. The calculations using Numpy arrays are faster than the normal Python array.

The NumPy package is created by the Travis Oliphant in 2005 by adding the functionalities of the ancestor module Numeric into another module Numarray. It is also capable of handling a vast amount of data and convenient with Matrix multiplication and data reshaping.

Both the Pandas and NumPy can be seen as an essential library for any scientific computation, including machine learning due to their intuitive syntax and high-performance matrix computation capabilities. These two libraries are also, best suited for data science applications.

**Difference between Pandas and NumPy:**

There are some differences between Pandas and NumPy that is listed below:

* The Pandas module mainly works with the tabular data, whereas the NumPy module works with the numerical data.
* The Pandas provides some sets of powerful tools like DataFrame and Series that mainly used for analysing the data, whereas in NumPy module offers a powerful object called Array.
* Instacart, SendGrid, and Sighten are some of the famous companies that work on the Pandas module, whereas NumPy is used by SweepSouth.
* The Pandas covered the broader application because it is mentioned in 73 company stacks and 46 developer stacks, whereas in NumPy, 62 company stacks and 32 developer stacks are being mentioned.
* The performance of Pandas is better than the NumPy for 500K rows or more. Between 50K to 500K rows, performance depends on the kind of operation.
* NumPy library provides objects for multi-dimensional arrays, whereas Pandas is capable of offering an in memory 2D table object called DataFrame.
* NumPy consumes less memory as compared to Pandas.
* Indexing of the Series objects is quite slow as compared to NumPy arrays.

The below table shows the comparison chart between the Pandas and NumPy:

|  |  |  |
| --- | --- | --- |
| Basis for Comparison | Pandas | NumPy |
| Works with | Pandas module works with the tabular data. | NumPy module works with numerical data. |
| Powerful Tools | Pandas has powerful tools like Series, DataFrame etc. | NumPy has a powerful tool like Arrays. |
| Organizational usage | Pandas is used in popular organizations like Instacart, SendGrid, and Sighten. | NumPy is used in the popular organization like SweepSouth. |
| Performance | Pandas has a better performance for 500K rows or more. | NumPy has a better performance for 50K rows or less. |
| Memory Utilization | Pandas consume large memory as compared to NumPy. | NumPy consumes less memory as compared to Pandas. |
| Industrial Coverage | Pandas is mentioned in 73 company stacks and 46 developer stacks. | NumPy is mentioned in 62 company stacks and 32 developer stacks. |
| Objects | Pandas provides 2D table object called DataFrame. | NumPy provides a multi-dimensional array. |