

Muhammad Rafly Alfarizqy Maulana

Python for Data Science

Numpy adalah sebuah library yang berbasis Bahasa python dan berfungsi untuk membantu memudahkan perhitungan numerik yang banyak dan rumit dalam Bahasa python tanpa menggunakan syntax dasar. Numpy merupakan library yang dapat diakses secara gratis oleh siapapun dapat digunakan dalam berbagai macam bidang seperti perhitungan dasar, data science, data analyst, machine learning, dan banyak bidang lain.

Numpy adalah singkatan dari Numerical Python, berdasarkan Namanya python berfungsi untuk melakukan operasi perhitungan numeric. Penggunaan Numpy dilakukan karena kecepatan pengerjaan perhitungan numerik jauh jika dibandingkan dengan pengerjaan secara manual dengan bahasa python. Numpy menggunakan pendekatan secara array programming sehingga user tidak perlu melakukan perulangan dalam operasi numerik yang banyak dan besar. Numpy juga mampu menangani operasi yang lebih kompleks seperti aljabar linear, perkalian dan matriks, transformasi fourier, dll.

Numpy dirilis pada tahun 2006 dan berfungsi untuk memperbarui proyek pemrosesan data numerik yang ada sebelumnya dan bernama Numeric, Numeric kemudian dikembangkan dan digantikan oleh Numarray sebagai library yang lebih fleksibel. Numeric yang kemudian menjadi Numarray diintegrasikan dengan proyek NumPy sebagai Pustaka yang memuat semua pemrosesan numerik dalam Bahasa python, NumPy sendiri awalnya diajukan untuk dimasukkan sebagai fitur dalam python, namun hal ini tidak diwujudkan dan dibuatlah library NumPy tersendiri.

Untuk menggunakan NumPy, kita dapat melakukan instalasi library dalam environment yang didukung seperti Anaconda dan Pip. Untuk mengoperasikan library numpy, dapat menggunakan Jupyter Notebook atau Jupyter lab. Setelah membuat notebook baru harus dilakukan import Numpy dengan code “import numpy as np”, dan fitur yang disediakan library numpy sudah bisa dioperasikan.

Array adalah struktur data yang umum dalam semua Bahasa pemrograman, dan bahasa python memiliki list yang mirip dengan array biasa tetapi lebih dinamis dari array, namun array lebih efektif dalam operasi numerik karena bersifat statis dan hanya menampung satu tipe data. List dalam python menggunakan kurung kurawal ({ ..., ..., }) sedangkan Array menggunakan library numpy dibuat dengan function np.array([..., ..., ..., ..., ...,]) misal seperti arr =

`np.array([1, 2, 3, 4, 5])`. Array juga dapat dibuat sebagai array multidimensi dengan menuliskan `np.array([[1,2,3,4,5], [6,7,8,9,10], [11,12,13,14,15]])`. Array multidimensi tidak memiliki batas, tapi umumnya hanya digunakan sampai tiga dimensi.

Array memiliki system pengalamatan dengan indexing, dimana index selalu dimulai dari index ke-0, untuk memanggil suatu elemen dalam array dapat menggunakan `array[x]` dan juga `array[x][y]` untuk array multidimensi. Cara indexing juga dapat digunakan untuk operasi secara langsung untuk mendapat elemen dalam array yang dibutuhkan oleh operasi numerik. Array slicing adalah Teknik untuk mengambil beberapa elemen dalam array dengan cara seperti memotong array, Teknik slicing dapat dilakukan dengan indexing seperti biasanya.

Dalam struktur list, untuk menyalin suatu list kita hanya perlu membuat list baru yang memiliki nilai sama dengan list yang sudah ada, dan list baru ini dapat dimodifikasi dan diubah sesuai keperluan tanpa mengubah list awal. Berbeda untuk array, array menggunakan cara menyalin yang sama dengan list biasa tetapi juga dapat menggunakan function `view` untuk melakukan proses salin, tetapi kedua array akan saling terhubung sehingga nilai yang diubah pada array Salinan akan merubah nilai pada array awal. Array juga dapat difilter untuk mendapatkan sejumlah elemen yang memiliki ketentuan sesuai kebutuhan tanpa memilikinya secara manual.

Menggunakan Array dalam NumPy kita dapat menyelesaikan berbagai permasalahan yang dapat dipecahkan oleh aljabar linear menggunakan function NumPy `np.linalg.solve(x,y)` dengan array dua dimensi sebagai bentuk matriks nya. Selain function aljabar linear, ada beberapa fungsi universal yang umum seperti operasi aritmatika, operasi matriks, function `repeat`, `add`, dan function-function lain sehingga tidak perlu dilakukan looping untuk perhitungan array. Operasi Matriks adalah operasi yang menggunakan beberapa function yang memiliki kemampuan untuk melakukan perhitungan dalam matriks (array dua dimensi), beberapa contoh operasi diantaranya `sum`, `reduce`, `add`, dll.

Visualisasi data dapat dilakukan menggunakan suatu library yang bernama `matplotlib`, sebelum menggunakannya perlu lakukan instalasi terlebih dahulu. Untuk menggunakannya lakukan import dengan `"import matplotlib.pyplot as plt"`. bentuk visualisasi data dengan `matplotlib` dapat dengan grafik batang, pie chart, grafik garis, histogram, box plot dan bentuk-bentuk lain.