ELBOW METHOD



**LAPORAN PRAKTIKUM  
Disusun untuk Memenuhi Tugas Laporan Praktikum Pertemuan 4  
Mata Kuliah Pembelajaran Mesin**

**DISUSUN OLEH:  
LINGGAR MARETVA CENDANI  
24060117120031**

**PROGRAM STUDI STRATA 1 INFORMATIKA  
DEPARTEMEN ILMU KOMPUTER/INFORMATIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG  
2019**

# BAB I RUMUSAN MASALAH DAN TUJUAN

## 1.1 Rumusan Masalah

1. Bagaimana cara meng-mport library?
2. Bagaimana cara memuat data?
3. Bagaimana cara menentukan nilai yang tepat untuk cluster?
4. Bagaimana cara plot nilai k?
5. Bagaimana cara menentukan nilai k berdasarkan elbow graph?

## 1.2 Tujuan

1. Mengetahui cara meng-mport library.
2. Mengetahui cara memuat data.
3. Mengetahui cara menentukan nilai yang tepat untuk cluster.
4. Mengetahui cara plot nilai k.
5. Mengetahui cara menentukan nilai k berdasarkan elbow graph.

# BAB II DASAR TEORI

## 2.1 Scikit-learn

## Scikit-learn adalah pustaka Python open source yang mengimplementasikan serangkaian pembelajaran mesin, pra-pemrosesan, validasi silang, dan algoritma visualisasi menggunakan antarmuka terpadu.

## 2.2 Matplotlib

Matplotlib adalah library Python 2D yang dapat menghasilkan plot dengan kualitas tinggi dalam berbagai format dan dapat digunakan di banyak platform.

## Matplotlib dapat digunakan sebagai pembuat grafik dalam berbagai platform, seperti Python dan Jupyter. Grafik yang dapat dibuat beragam, seperti grafik garis, batang, lingkaran, histogram, dsb.

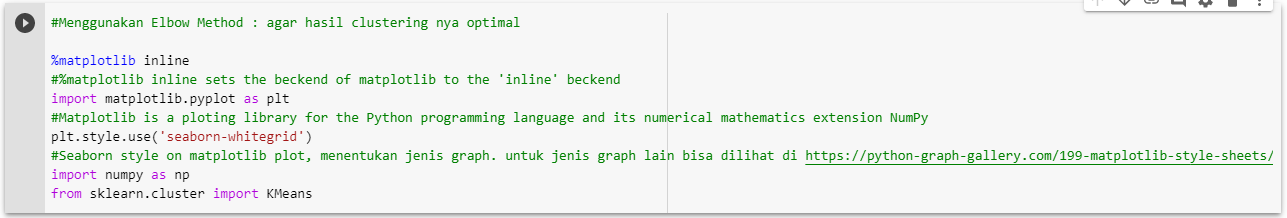
## 2.3 Numpy

NumPy adalah paket pemrosesan array tujuan umum. Ini menyediakan objek array multidimensi berkinerja tinggi, dan alat untuk bekerja dengan array ini. Ini adalah paket mendasar untuk komputasi ilmiah dengan Python. Ini berisi berbagai fitur termasuk yang penting ini seperti Objek array N-dimensi yang kuat, Fungsi-fungsi canggih (penyiaran), Alat untuk mengintegrasikan kode C / C ++ dan Fortran, Aljabar linier yang bermanfaat, transformasi Fourier, dan kemampuan angka acak.

# BAB III PEMBAHASAN

## 3.1 Import Library

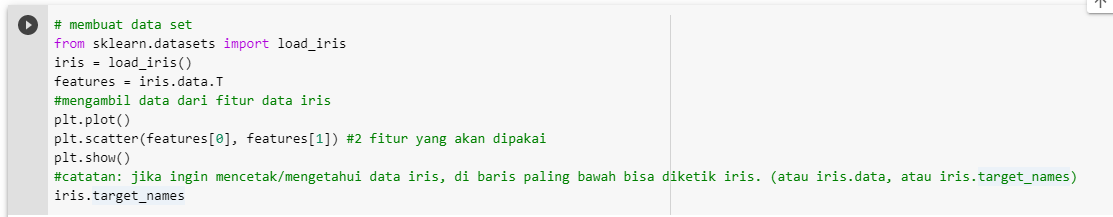
Pertama, kita import terlebih dahulu library - library yang akan kita gunakan pada program python kita, untuk mengolah dataset yang kita miliki.



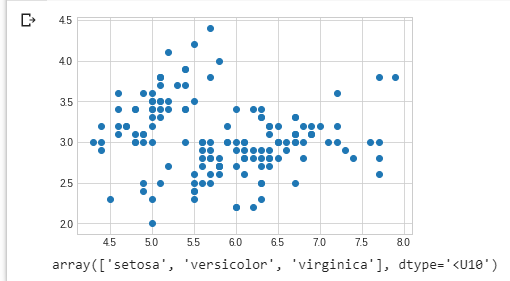
Seperti yang ditunjukkan pada gambar di atas, library yang kita import adalah Numpy, Matplotlib, dan Sklearn.

## 3.2 Memuat Data

Data yang digunakan dalam eksperimen ini adalah data iris. Dataset ini terdiri atas empat fitur yaitu Sepal Length, Sepal Width, Petal Length, dan Petal Width. Serta terdiri atas tiga kelas, yaitu Iris-setosa, Iris-versicolor, dan Iris-virginica. Pada Iris dataset, jumlah keseluruhan data adalah 150 baris dengan masing-masing kelas memiliki 50 baris data.

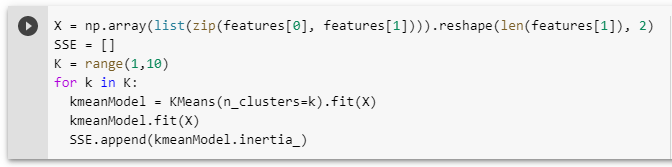


Dalam eksperimen ini kita akan memakai 2 fitur dari data iris, tanpa menggunakan label dari data tersebut untuk melakukan clustering. Data iris terdiri dari data pengukuran fitur (iris.data) dan label (iris.target\_names). Setelah eksekusi plt.show() maka akan ditampilkan data sebagai berikut.



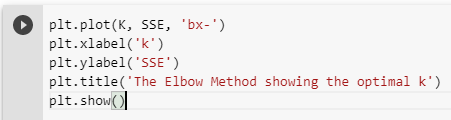
## 3.3 Menentukan Nilai Yang Tepat Untuk Cluster

Untuk menentukan nilai yang tepat untuk K, kita dapat menghitung Sum of Squared error dari cluster yang dihasilkan. Dalam python hal tersebut dapat dilakukan dengan menyimpan nilai inertia dalam variable SSE sebagai berikut.

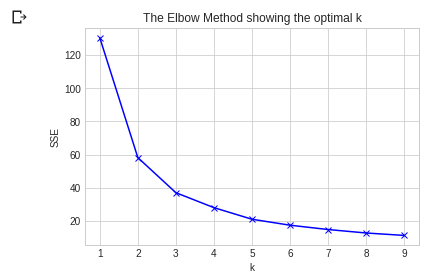


## 3.4 Plot Nilai K

Lakukan Plot nilai K dengan sumbu x berupa jumlah K dan sumbu y berupa SSE-nya.

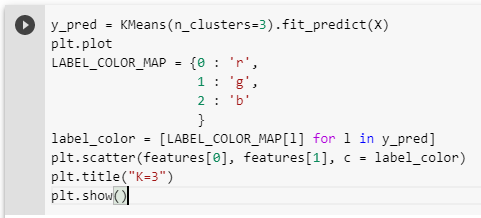


Maka hasilnya akan sebagai berikut.

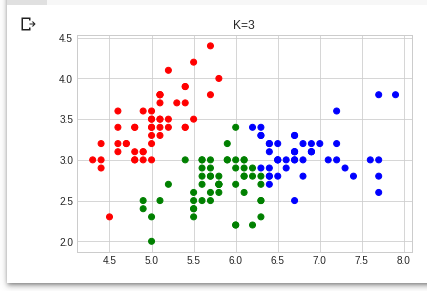


## 3.5 Menentukan Nilai K Berdasarkan Elbow Graph

Berdasarkan nilai K yang didapatkan, dapat dilakukan clustering dengan nilai K tersebut. Sebagai contoh jika kita ambil K adalah 3.



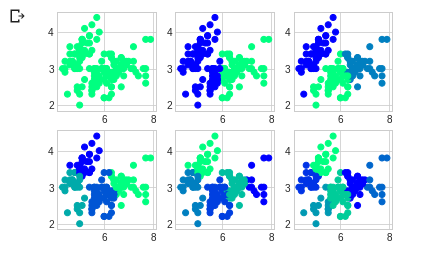
Maka akan tercetak hasil sebagai berikut.



Percobaan tambahan dapat dilakukan untuk melihat bentuk cluster untuk semua kemungkinan K yang kita inginkan. Misalkan ketika kita ingin melihat hasil cluster untuk K dari 1 sampai 6 maka :



Maka akan tercetak hasil sebagai berikut.



# BAB IV KESIMPULAN

## 1 Kesimpulan

Dari beberapa percobaan jumlah K yang kita lihat pada Elbow Graph, didapatkan bahwa K yang paling optimal adalah K dengan jumlah 3. Ditambah grafik yang dihasilkan pada percobaan tambahan dari K 1 sampai 6.

Maka dari itu, dapat ditarik kesimpulan bahwa jumlah K terbaik yang dapat kita pakai adalah 3.

# 

# DAFTAR PUSTAKA

Informatics (21 Oktober 2019). "*Modul Praktikum 4. Issue dalam K-Means Clustering*". Makalah disajikan dalam Praktikum Pembelajaran Mesin di laboratorium komputer gedung E Universitas Diponegoro. Semarang, 4 November 2019.

GeeksforGeeks (8 April 2017). "*Learning Model Building in Scikit-learn : A Python Machine Learning Library*". Retrieved 4 November 2019. from GeeksforGeeks : https://www.geeksforgeeks.org/learning-model-building-scikit-learn-python-machine-learning-library/

GeeksforGeeks (10 Juli 2018). "*NumPy in Python | Set 1 (Introduction)*". Retrieved 4 November 2019. from GeeksforGeeks : https://www.geeksforgeeks.org/numpy-in-python-set-1-introduction/