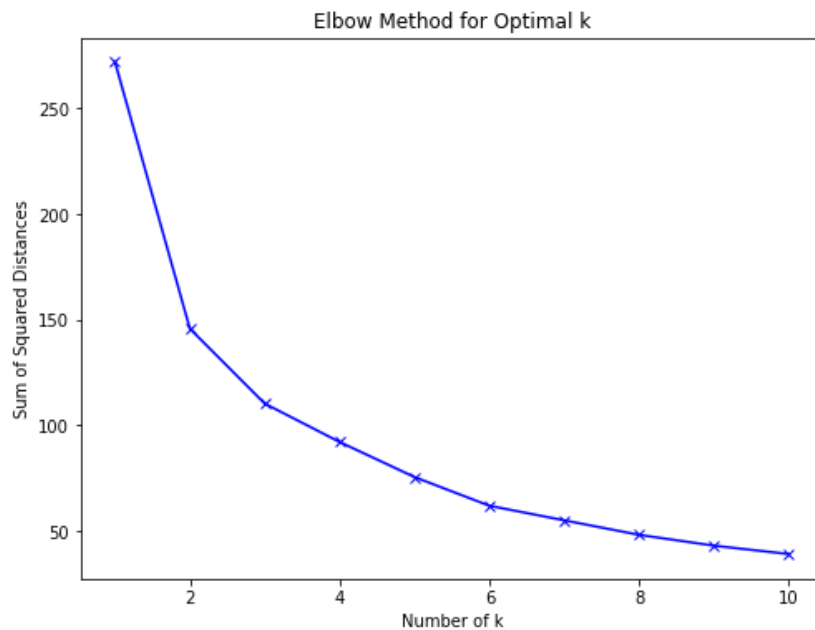


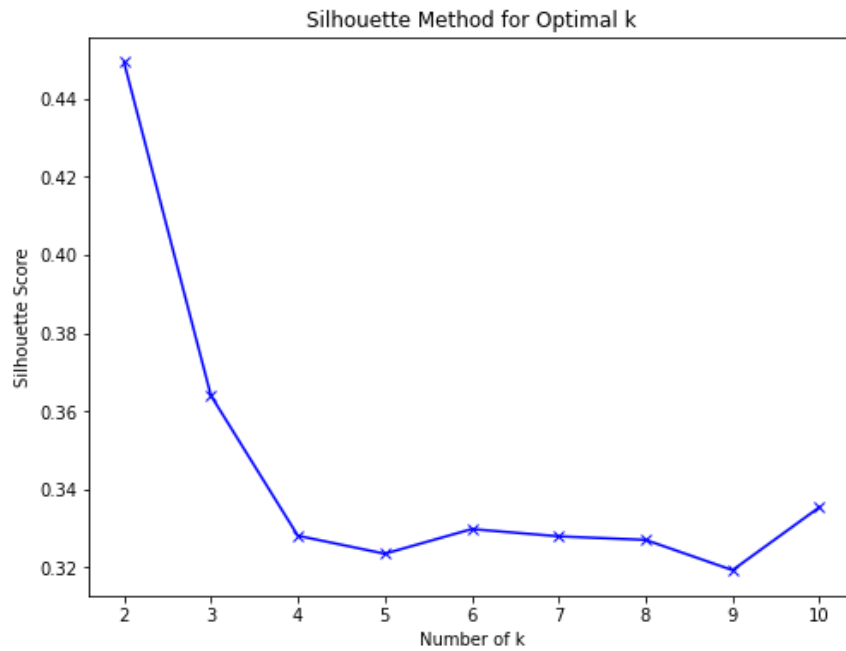
TUGAS 12

Pada latihan hari ini penulis ingin melakukan K-Means Clustering terhadap data di pokedex. Data yang diambil adalah dari 1-500 (Emboar) dengan total baris sebanyak 598, dan 10 kolom. Kolom yang akan dilakukan clustering adalah kolom Defense dan Attack. Hal pertama sebelum melakukan clustering adalah transformasi data. Kali ini penulis melakukan transformasi data menjadi nilai logaritmik. Pada saat melakukan clustering, kita harus menentukan banyaknya cluster atau k yang akan dibagi. Beberapa metode untuk menentukan nilai k yang terbaik diantaranya adalah elbow method, dan silhouette method.



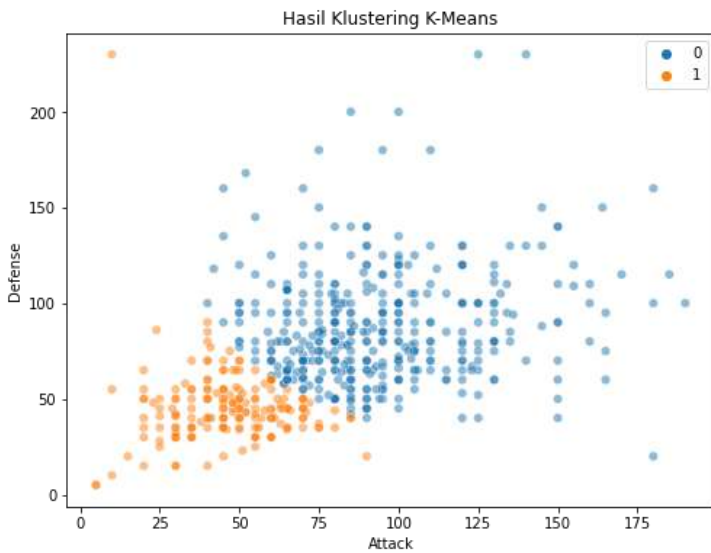
Gambar 1. Grafik elbow method untuk menentukan nilai k terbaik

Elbow method adalah salah satu metode untuk menentukan jumlah cluster yang tepat melalui persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik. Penulis melakukan uji untuk menentukan yang mana nilai k terbaik dari 1 sampai 10 dengan data yang telah ditransformasi. Berdasarkan Gambar (1), tidak terlalu jelas terlihat siku yang terbentuk apakah bernilai 2 atau 3. Untuk itu penulis melakukan uji dengan metode lain, yaitu silhouette method.

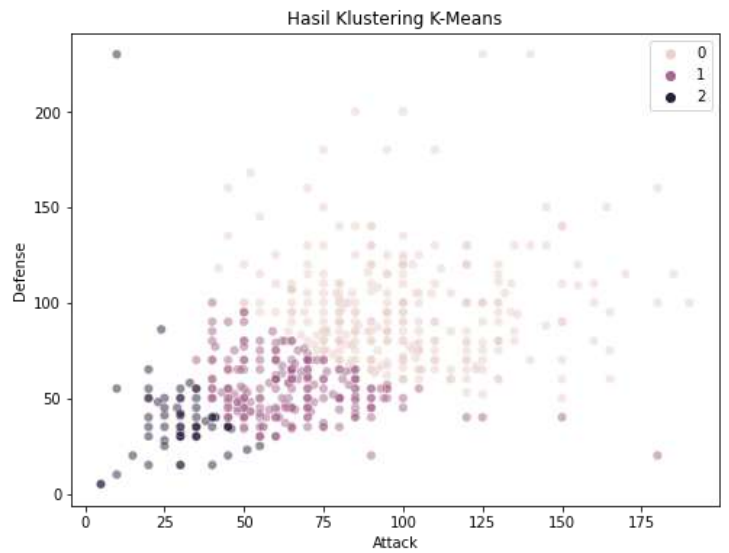


Gambar 2. Grafik silhouette method untuk menentukan nilai k terbaik

Silhouette method digunakan untuk melihat kualitas dan kekuatan cluster, seberapa baik atau buruknya suatu objek ditempatkan dalam suatu cluster. Penulis melakukan uji untuk menentukan yang mana nilai k terbaik dari 2 sampai 10 dengan data yang telah ditransformasi. Berdasarkan Gambar (2), nilai puncak berada di $k = 2$, sehingga dapat disimpulkan nilai k terbaik dari data yang diolah adalah 2.



(a)



(b)

Gambar 3. Grafik hasil clustering K-Means

Pada silhouette method sebelumnya didapatkan nilai k terbaik adalah 2. Penulis melakukan beberapa kali percobaan dengan nilai k yang berbeda untuk melihat visualisasinya. Silhouette score untuk nilai $k = 2$ adalah 0.45, dan nilai $k = 3$ adalah 0.36. Kesimpulan untuk percobaan kali ini adalah nilai k terbaik untuk data yang diolah adalah 2 dengan silhouette score sebesar 0.45.