**LAPORAN TUGAS 2 KECERDASAN KOMPUTASIONAL**



**Kelas : B**

**Alie Husaini R. 05111840000097**

**Ammar Alifian 05111184000007**

**Nodas Uziel 05111840007007**

**Dosen:**

**Dr.Diana Purwitasari , S.Kom., M.Sc**

**Departemen Infomatika**

**Fakultas Teknologi Elektro dan Informatika Cerdas**

**Institut Teknologi Sepuluh Nopember (ITS) Surabaya**

**2020**

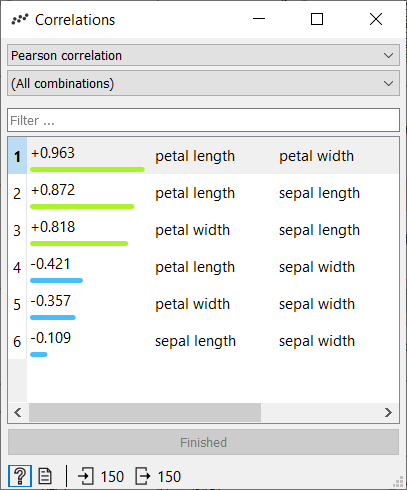
**Rumusan Masalah**

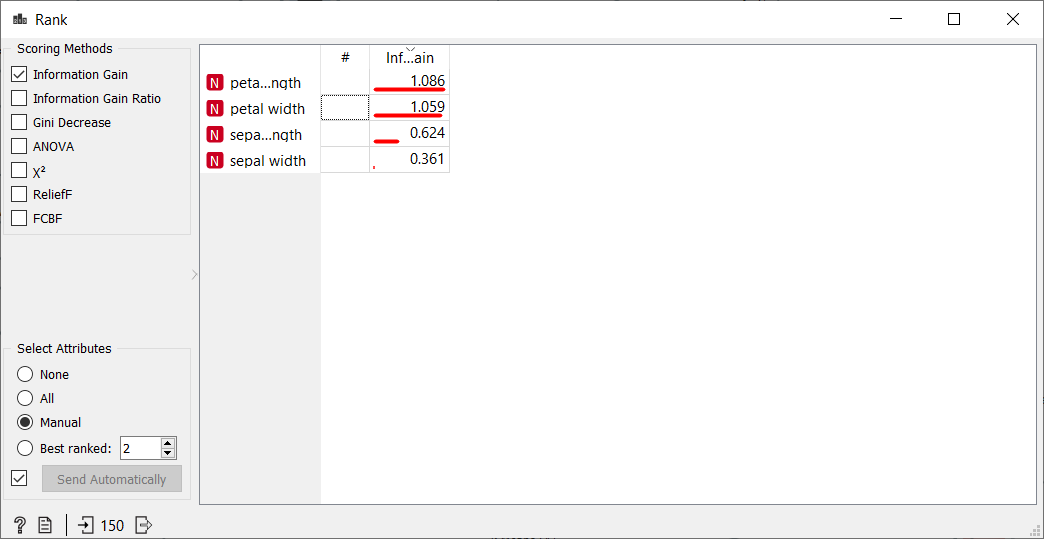
1. Apa kombinasi fitur yang memberikan cluster terbaik menurut indikator Sillhoutte score?
2. Bagaimana hubungan korelasi antar fitur2 yang terpilih?
3. Apakah ada proses cleaning data (impute, dll)?
4. Bagaimana hasil cluster jika centroid terbentuk secara random atau dipiihkan yang terbaik kmeans ++?
5. Bandingkan KMeans dgn Hiearchical (divisive, agglomerative)!
6. Bandingkan dengan algoritma jika menggunakan SkLearn!

*Dalam pengerjaan tugas kami memakai data default iris dari orange*

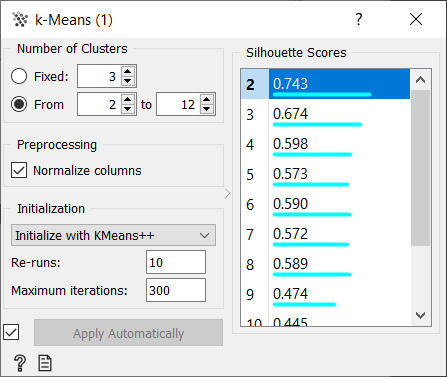
1. Apa kombinasi fitur yang memberikan cluster terbaik menurut indikator Sillhoutte score?

Dalam mecari kombinasi fitur yang terbaik, kami memakai widget Rank dan Correlations. Pada widget Rank petal length dan petal width memiliki nilai informasi yang cukup tinggi dan pada widget Correlations petal length dan petal width memiliki kolerasi yang cukup tinggi dari kolerasi yang lain, sehingga kombinasi fitur yang memiliki cluster terbaik adalah petal length dan petal width

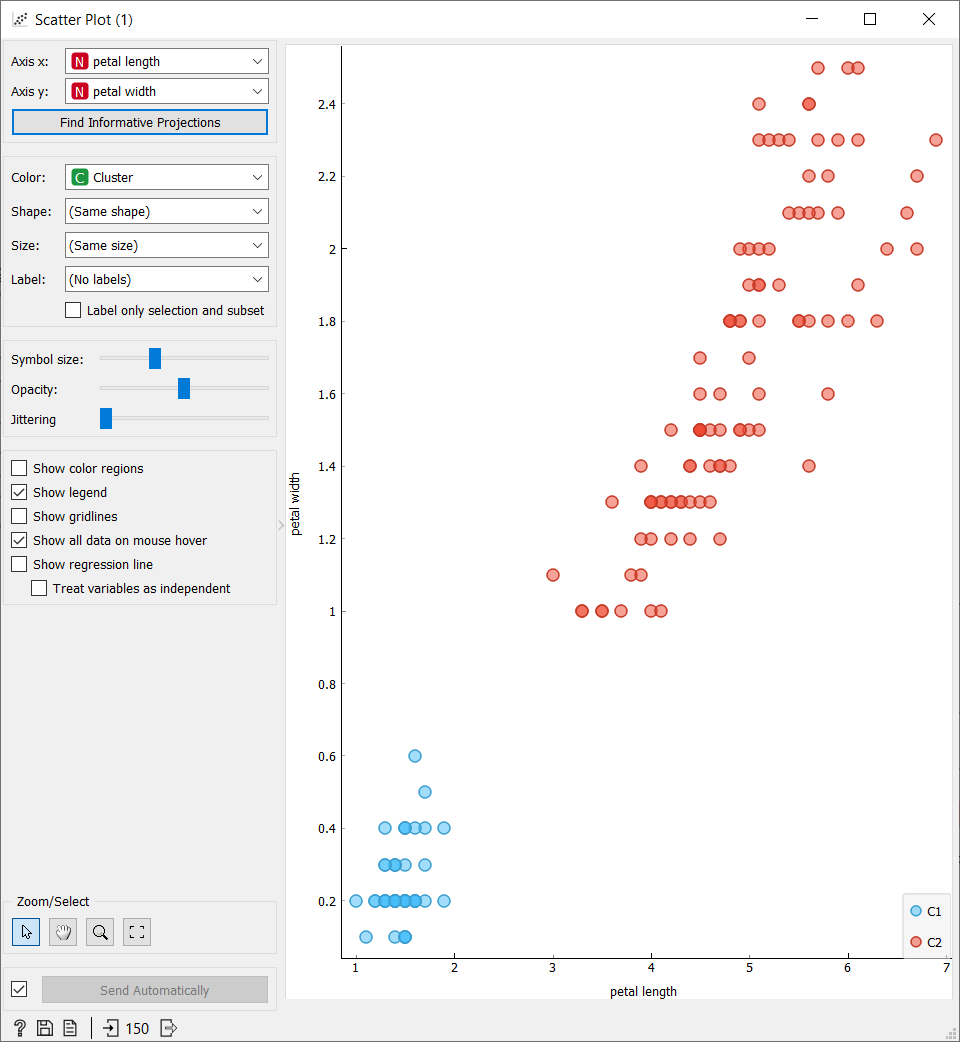




Dan pada saat mencari K-means dengan memilih kolom petal width dan petal length, dapat terlihat bahwa cluster yang dimiliki cukup besar dari cluster lain.



*Scatter plot petal length dan petal width*



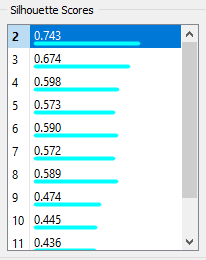
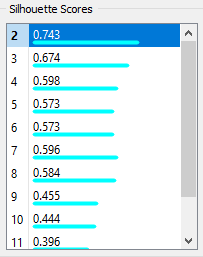
1. Bagaimana hubungan korelasi antar fitur2 yang terpilih?

Dapat disimpulkan dari nomor 1 bahwa, nilai informasi pada widget Rank dan nilai kolerasi pada widget Colerrations mempengaruhi cluster terbaik yang nantinya di cari dalam indicator Silhouette Scores.

1. Apakah ada proses cleaning data (impute, dll)?

Tidak ada impute data yang terjadi pada data Iris yang kami pakai karena tidak terdapat nilai Null.

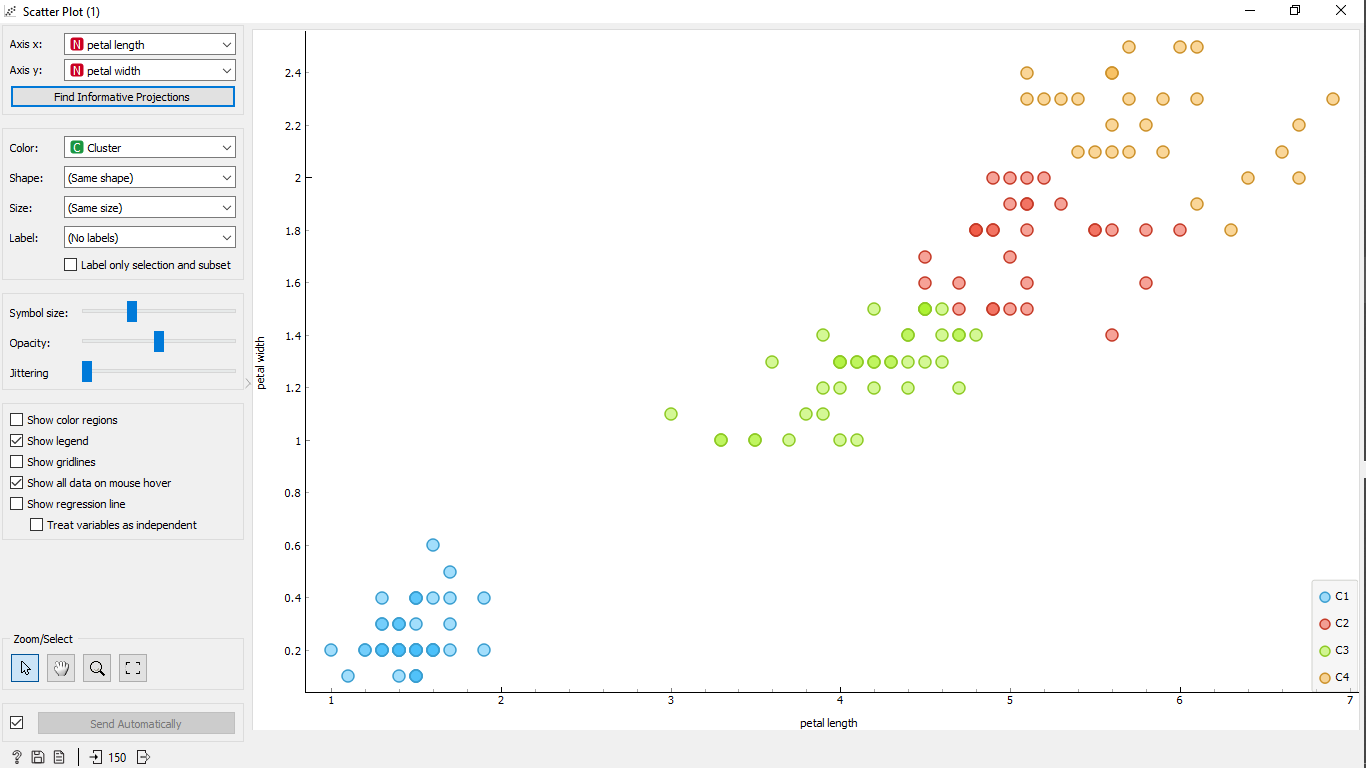
1. Bagaimana hasil cluster jika centroid terbentuk secara random atau dipiihkan yang terbaik kmeans ++?

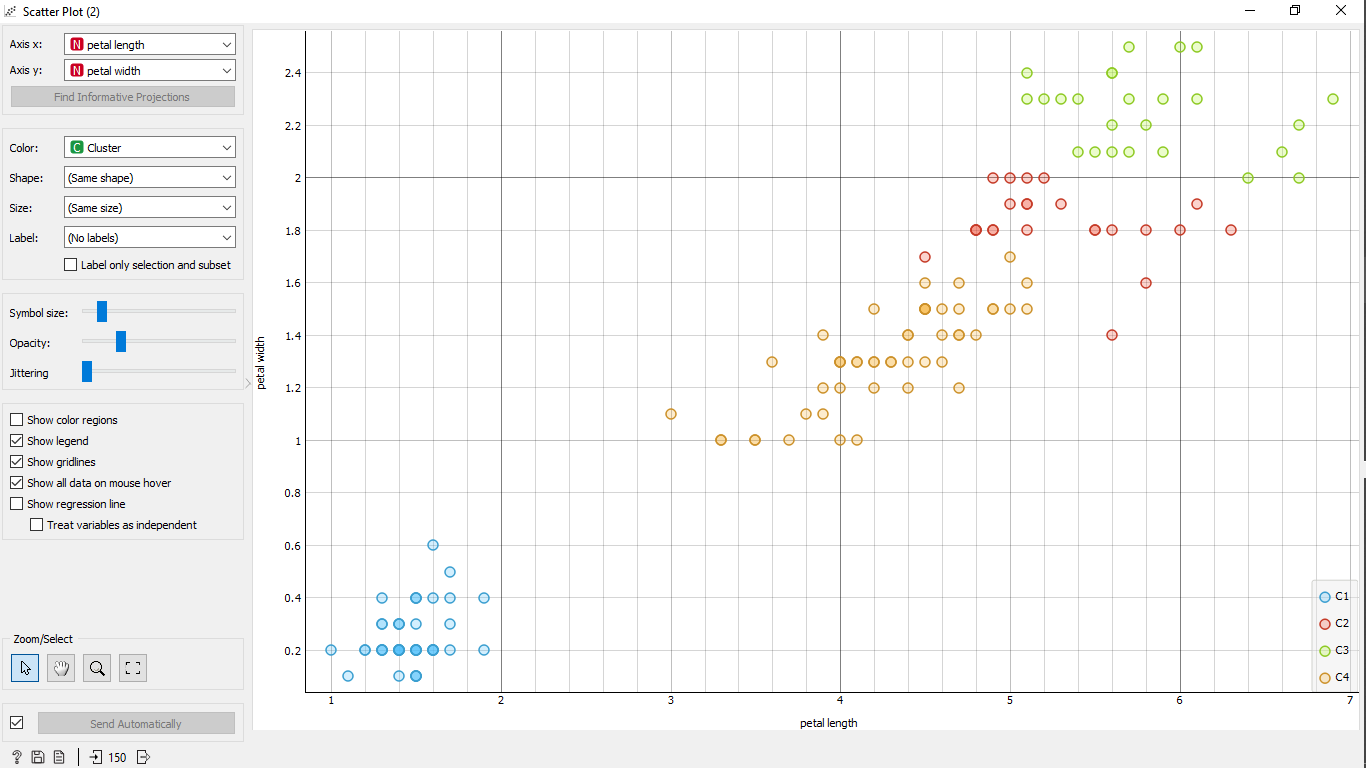
*KMeans++ Inisialisasi random*

Pada jumlah cluster yang silhouette score-nya tinggi kurang lebih tidak ada perbedaan, namun untuk jumlah cluster dengan silhouette score yang lebih rendah ada sedikit selisih antara KMeans++ dan insialisasi random dimana KMeans++ cenderung memiliki hasil yang lebih baik daripada inisialisasi random. Hal ini bisa saja terjadi karena data yang kami gunakan sudah cukup meng-cluster.

1. Bandingkan KMeans dgn Hiearchical (divisive, agglomerative)!



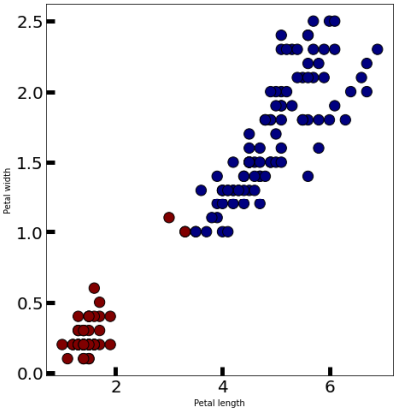
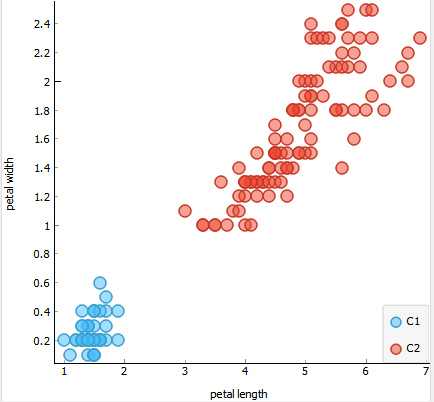
*KMeans*



*Hierarchical*

Setelah mencoba semua linkage Hierarchical, kami menemukan bahwa hasil terbaik diperoleh dengan menggunakan average linkage. Meski demikian, jika dibandingkan dengan menggunakan KMeans, hasilnya sama untuk 2 cluster dan cenderung lebih baik KMeans untuk jumlah cluster lebih dari 2.

1. Bandingkan dengan algoritma jika menggunakan SkLearn!

*SkLearn Orange*

Jika kita perhatikan, pada scatter plot SkLearn terdapat beberapa titik yang harusnya bukan bagian dari cluster-nya. Sedangkan untuk Orange, cluster-nya lebih sesuai. Sehingga, kami menyimpulkan bahwa menggunakan Orange lebih akurat daripada menggunakan SkLearn.