Nama : Ihsan Arif Rahman

NIM : G64144025

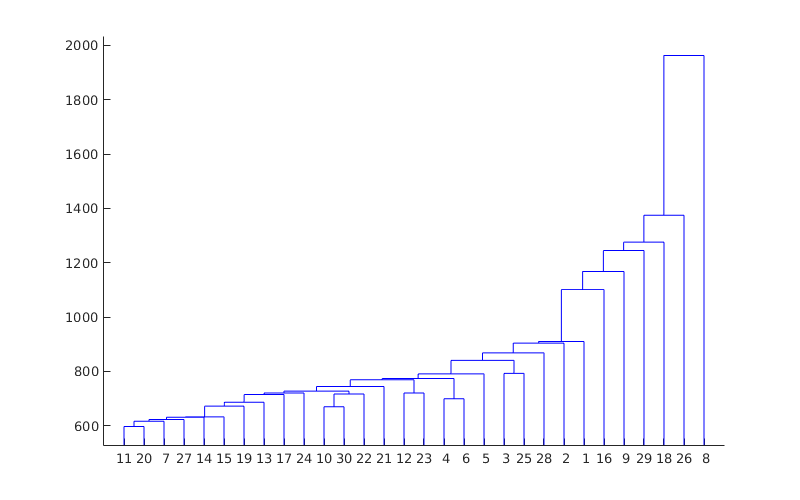
Analisis kluster adalah teknik statistik yang berguna untuk megelompokkan objek atau variabel ke dalam beberapa kelompok tertentu dimana setiap objek memiliki sifat dan karakteristik yang berdekatan. Pada riset pemasaran, analisis kluster biasanya digunakan untuk melakukan proses segmentasi sejumlah responden dalam hal ini konsumen berdasarkan ciri-ciri sejumlah atribut yang ada. Analisis kluster dapat dibagi menjadi 2 jenis, yaitu : *Hierarchical Cluster* dan *K-Mean Cluster*. *Hierarchical Cluster* biasanya digunakan untuk jumlah sampel (data) yang relatif sedikit (< 100) sedangkan *K-Mean Cluster* digunakan untuk data yang relatif banyak (> 100). Hierarchical berupaya mengelompokkan responden berdasar kemiripan yang ada (persepsi) mereka. Hal ini disebabkan kluster secara hierarki akan melakukan proses dengan membandingkan setiap pasang kasus yang tentunya untuk jumlah kasus yang sedikit.

Contoh kasus yang diuji coba adalah data citra buah-buahan yang diambil dari internet. Buah-buahan tersebut diambil 5 kelas atau jenis yaitu apel merah, apel hijau, kiwi, pisang dan buah pear. Masing masing foto memiliki karakteristik tersendiri. Pengolahan citra ini menggunakan perangkat lunak Matlab dengan cara :

1. mengubah citra dari RGB menjadi mode grayscale.
2. Mengubah ukuran resolusi dari semua citra yang tidak memiliki kesamaan resolusi sehingga diubah menjadi ukuran 30x30.

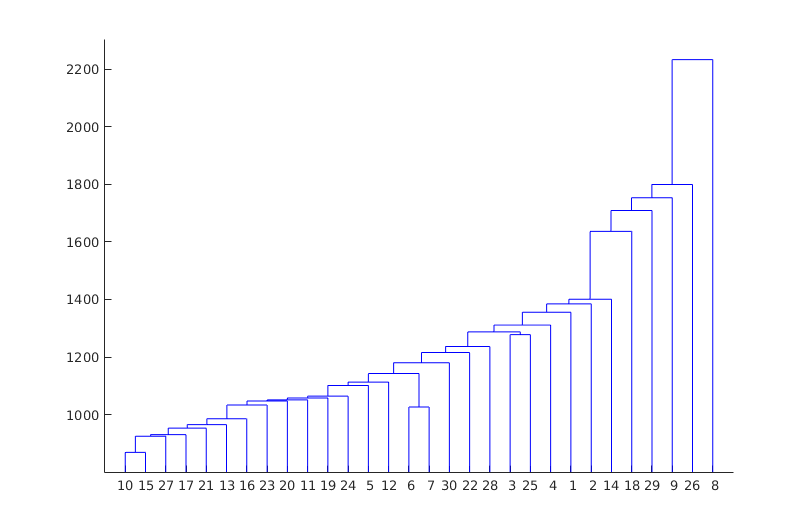
Pengolahan citra ini sebelum dilakukan kluster terlebih dahulu olah menggunakan metode PCA (Principal Component Analysis). Metode ini adalah sebuah teknik untuk membangun variabel baru yang merupakan kombinasi linear dari variabel asli. Jumlah maximum dari variabel baru ini akan sama dengan jumlah dari variabel lama, dan variabel baru ini tidak saling berkorelasi satu sama lain.

Dari data citra sebanyak 50 dan memiliki ukuran resolusi 30x30 maka kita memiliki 50 sampel dan variabel 900 dari perkalian pixel 30x30. Dengan variabel 900 sangat banyak sekali dengan menggunakan PCA sehingga variabel tersebut bisa direduksi atau disederhanakan dengan tingkat keberhasilan tertentu. Pada percobaan ini akan mengambil sampel 70% dan 90%. Dengan sudah ditentukannya tingkat keberhasilan dari data citra tersebut maka untuk tingkat keberhasilan 70% menggunakan 7 variabel sedangkan data citra yang memiliki tingkat keberhasilan 90% menggunakan 20 variabel dari 900 variabel.



Gambar 1 Dendrogram dengan tingkat keberhasilan 70%

Pada Gambar 1 menunjukkan bahwa sebaran nilai data tersebut beragam dari 50 sampel tersebut. Seperti pada data citra ke-8 memiliki nilai yang tinggi. Hasil dari kluster ini menyebar menyebar di bebagai kelompok. Dan ada juga data citra yang memiliki hasil yang saling berdekatan. Dari tingkat keberhasilan 70% secara detail tidak terlalu kelihatan pengelompokkan data citranya.



Gambar 2 Dendrogram dengan tingkat keberhasilan 90%

Pada Gambar 2 menunjukkan sebaran data citra yang lebih terlihat mengelompokkan. Data citra hasil kluster dengan tingkat keberhasilan 90% ini hampir sama keragaman datanya walaupun berbeda dari segi nilai.

Dari hasil kedua tersebut pada tingkat keberhasilan 90% memiliki kesamaan yang bisa dilihat dengan data berjejeran saling berdekatan nilainya pada satu kelas yang sama. Sedangkan pada tingkat keberhasilan 70%, nilai yang memiliki satu kelas itu saling menyebar di berbagai titik sehingga pengelompokkan hasil kluster ini tidak terlalu mencirikan suatu kelompok yang sama.