**Laporan Tugas Akhir MK. Data Mining (KOM332), Semester Genap 2014/2015**

**Penerapan Teknik K-Means Clustering pada data Abalone**

Ika Putri Puspitasari (G64154027)

**PENDAHULUAN**

**Latar Belakang**

Abalon (berasal dari [bahasa Spanyol](https://id.wikipedia.org/wiki/Bahasa_Spanyol), *Abulón*) ialah suatu spesies [kerang](https://id.wikipedia.org/wiki/Kerang)-kerangan ([moluska](https://id.wikipedia.org/wiki/Moluska)) dari [familia](https://id.wikipedia.org/wiki/Familia) Haliotidae dan [genus](https://id.wikipedia.org/wiki/Genus) *Haliotis*. Ia dikenal pula sebagai kerang mata tujuh atau siput balik batu, *ormer* di [Jersey](https://id.wikipedia.org/wiki/Jersey) dan [Guernsey](https://id.wikipedia.org/wiki/Guernsey), *perlemoen* di [Afrika Selatan](https://id.wikipedia.org/wiki/Afrika_Selatan), dan *pāua* di [Selandia Baru](https://id.wikipedia.org/wiki/Selandia_Baru).

Abalon tergolong dalam kelas [Gastropoda](https://id.wikipedia.org/wiki/Gastropoda) yang besar. Terdapat hanya satu genus dalam famili [Haliotidae](https://id.wikipedia.org/w/index.php?title=Haliotidae&action=edit&redlink=1) dan kira-kira 4 - 7 subgenus. [Taksonominya](https://id.wikipedia.org/wiki/Taksonomi) agak membingungkan. Spesiesnya berjumlah antara kira-kira 100 hingga 130 (karena adanya [hibrida](https://id.wikipedia.org/wiki/Hibrida)).

Abalon memiliki ciri-ciri permukaan kulit sebelah dalam yang berwarna-warni yang terbuat dari [nakre](https://id.wikipedia.org/w/index.php?title=Nakre&action=edit&redlink=1). Daging [moluska](https://id.wikipedia.org/wiki/Moluska) ini dianggap sebagai salah satu makanan istimewa di sebagian [Amerika Latin](https://id.wikipedia.org/wiki/Amerika_Latin) (khususnya [Chili](https://id.wikipedia.org/wiki/Chili)), [Asia Tenggara](https://id.wikipedia.org/wiki/Asia_Tenggara), dan [Asia Timur](https://id.wikipedia.org/wiki/Asia_Timur) (khususnya di [Republik Rakyat Tiongkok](https://id.wikipedia.org/wiki/Republik_Rakyat_Tiongkok), [Jepang](https://id.wikipedia.org/wiki/Jepang), dan [Korea](https://id.wikipedia.org/wiki/Korea)).

Memprediksi usia abalone dari pengukuran fisik. Usia abalone adalah ditentukan dengan memotong shell melalui kerucut, pewarnaan, dan menghitung jumlah dering melalui mikroskop - tugas yang membosankan dan memakan waktu. pengukuran lain, yang mudah untuk mendapatkan, digunakan untuk memprediksi usia. Informasi lebih lanjut, seperti pola cuaca dan Lokasi (maka ketersediaan pangan) mungkin diperlukan untuk memecahkan masalah.

**Tujuan**

Tujuan dari tugas akhir ini adalah :

1. Mendapatkan centroid dari data abalone.

2. Mengetahui berapa perbandingan nilai bila memiliki nilai k=3.

**Ruang Lingkup**

Ruang lingkup dari pengambilan data adalah:

Teknik yang digunakan dalam mengolah data adalah teknik *K-Mean Clustering* dan *Hierarchical Cluster.*

**Manfaat**

Manfaat dari hasil yang diperoleh dari tugas akhir ini adalah dapat mendapatkan pola dari sebuah data abalone dan untuk mengerti cara mengolah data menggunakan K-mean Clustering dan Hierarchical Cluster.

**TINJAUAN PUSTAKA**

Analisis kluster adalah teknik statistik yang berguna untuk mengelompokan objek atau variabel ke dalam beberapa kelompok tertentu dimana setiap objek memiliki sifat dan karakteristik yang berdekatan. Analisis kluster biasaya 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.

**METODOLOGI**

**Data**

Data yang digunakan dalam tugas akhir ini adalah Abalone. Data tersebut memiliki 9 atribut dan 100 *records*.

**Tahapan Penelitian**

Hierachy -Clustering

K-Mean

Cluster Data

Penyiapan data

Evaluasi

Kinerja

**Lingkungan Pengembangan**

Perangkat lunak yang digunakan dalam project tugas akhir ini adalah:

* Sistem operasi Microsoft Windows 10 (64-bit).
* RStudio sebagai proses pengolahan data.

Perangkat keras yang digunakan dalam project tugas akhir ini dengan spesifikasi:

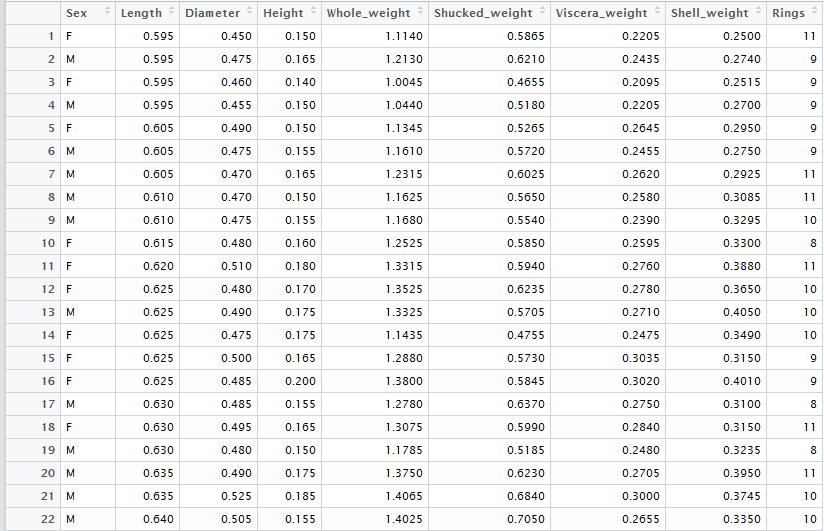
* Prosesor : Intel Core i3-380M ( 2.53GHz 1066MHz 3MB )
* Memory : 6 GB RAM
* VGA : Intel HD Graphics

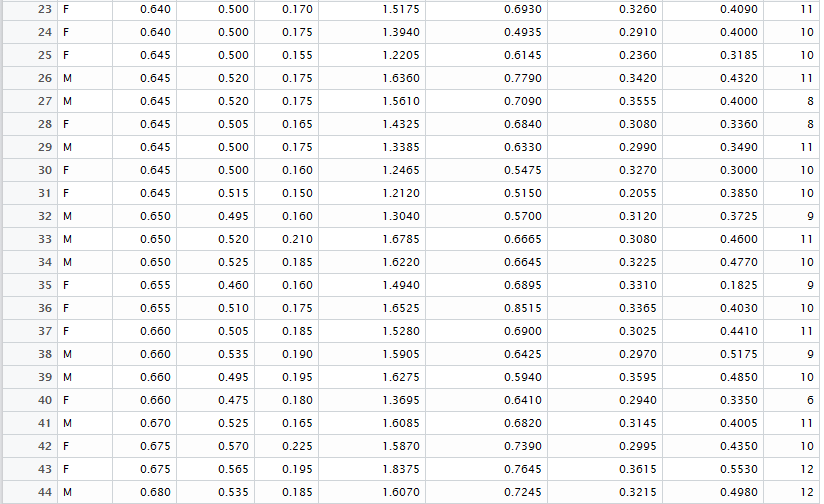
Perangkat keras yang digunakan dalam project tugas akhir ini dengan spesifikasi:

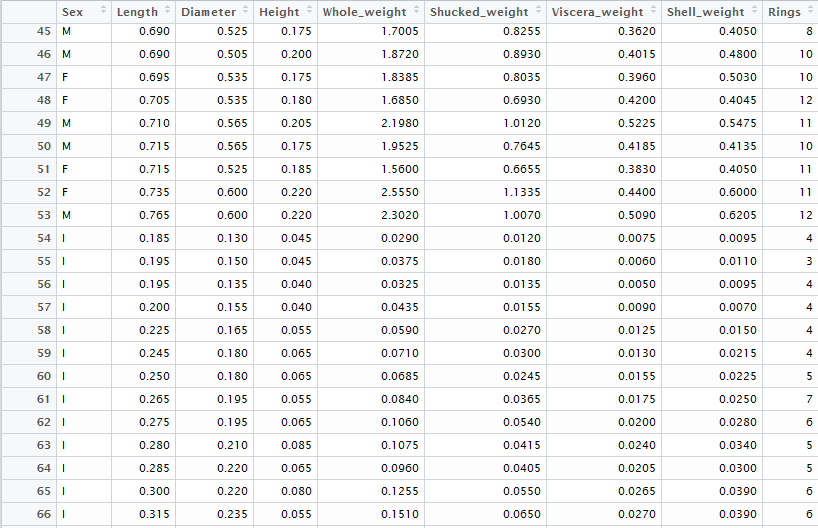
* Prosesor : Intel Core i3-380M ( 2.53GHz 1066MHz 3MB )
* Memory : 6 GB RAM
* VGA : Intel HD Graphics

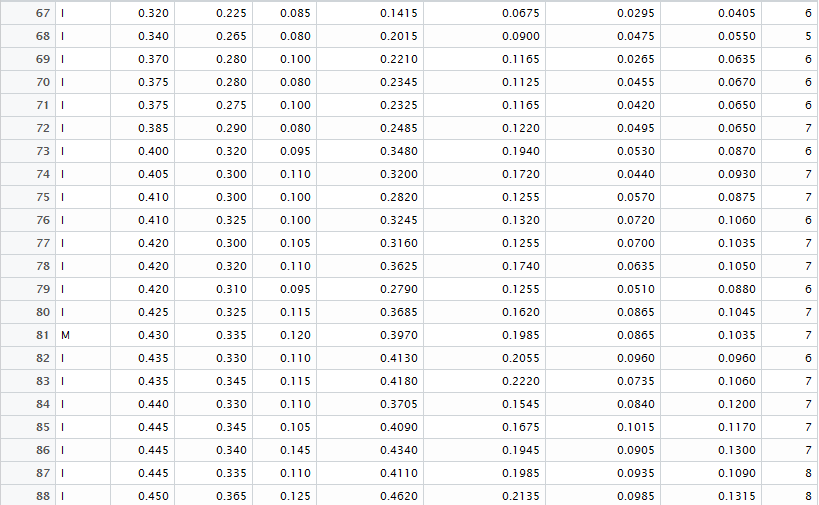
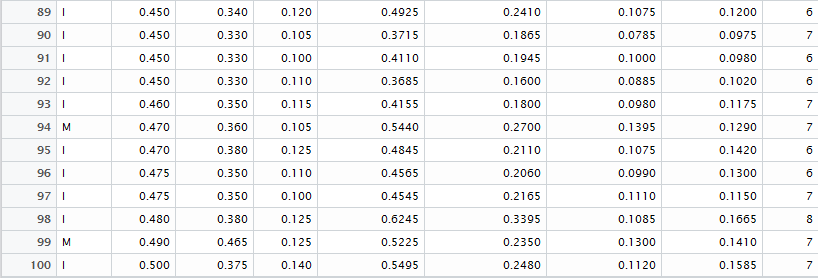
**HASIL DAN PEMBAHASAN**

Data yang digunakan adalah data primer yang diambil dari data mentah abalone. Atribute yang digunakan sebanyak 9 yaitu length, Diameter, Height, whole\_weight, Shucked\_weight, Viscera\_weight, Shell\_weight dan rings. Data record yang ada sebanyak 100. Sebagai berikut data abalone :







Data diatas mulanya dalam bentuk csv lalu di masukkan dalam Rstudio agar dapat diolah sebagai berikut :

> Data27 <- read.csv(file.choose(), header=TRUE)

> View(Data27)

> dataika<-Data27

> dataika$Sex<-NULL

> (kmeans.result <- kmeans(dataika,3))

K-means clustering with 3 clusters of sizes 46, 27, 27

Cluster means:

Length Diameter Height Whole\_weight

1 0.6494565 0.5082609 0.1755435 1.4829891

2 0.4814815 0.3714815 0.1222222 0.6210370

3 0.3462963 0.2588889 0.0850000 0.2557407

Shucked\_weight Viscera\_weight Shell\_weight

1 0.6645435 0.31182609 0.39089130

2 0.2926296 0.13338889 0.16407407

3 0.1188889 0.05355556 0.06892593

Rings

1 10.282609

2 7.333333

3 5.370370

Clustering vector:

[1] 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1

[25] 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 2 1 1 1

[49] 1 1 1 1 1 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 2

[73] 3 2 2 3 2 2 3 2 2 3 2 2 2 2 2 2 3 2 3 3 2 2 3 3

[97] 2 2 2 2

Within cluster sum of squares by cluster:

[1] 43.85182 13.32354 25.35083

(between\_SS / total\_SS = 85.2 %)

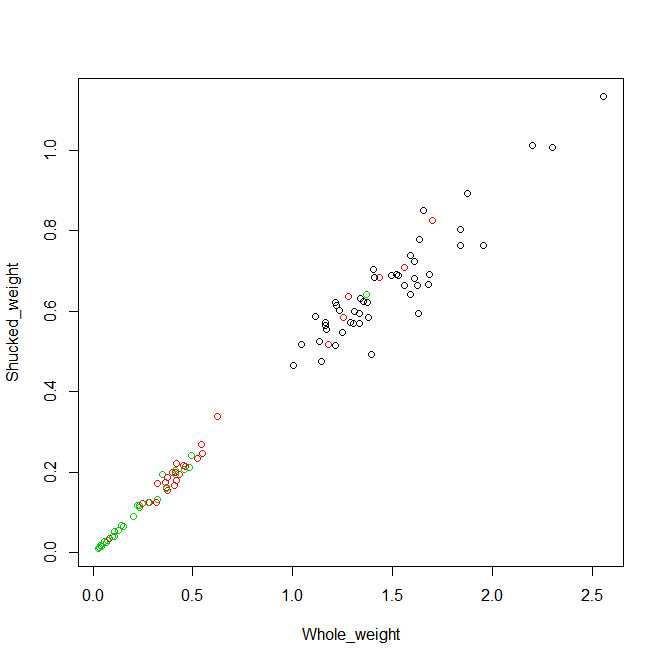
Available components:

[1] "cluster" "centers" "totss"

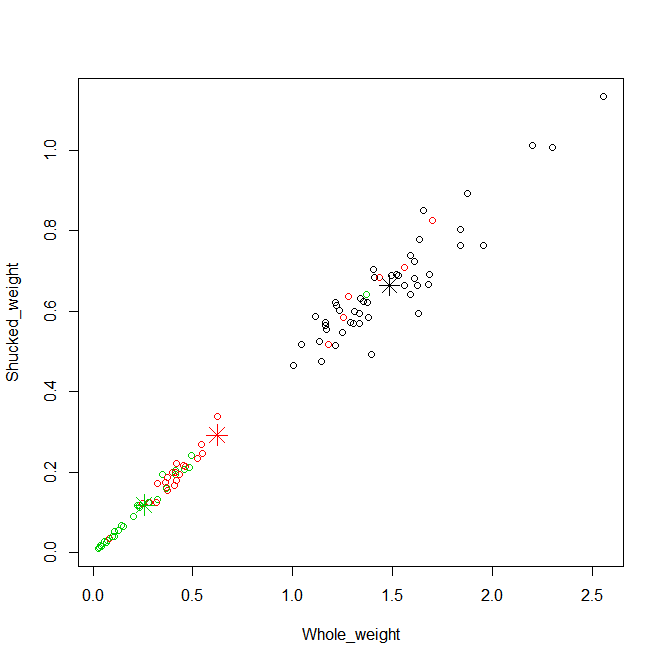
[4] "withinss" "tot.withinss" "betweenss"

[7] "size" "iter" "ifault"

plot(dataika[c("Whole\_weight","Shucked\_weight")],col=kmeans.result$cluster)

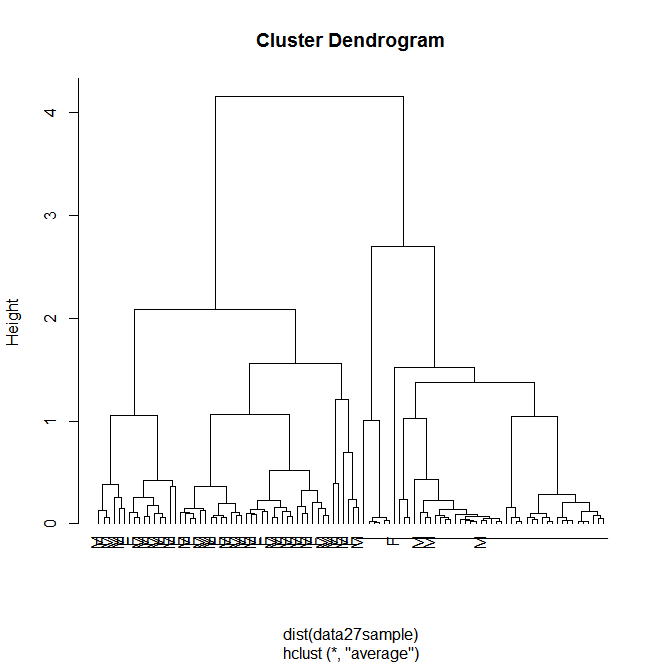


points(kmeans.result$centers[,c("Whole\_weight","Shucked\_weight","Viscera\_weight")],col =1:3,pch=8,cex=2)



Dari data yang digunakan pada abalone kita menggunakan k=3 maka di dapatkan hasil total yang cukup baik yaitu sebesar 85.2%.

Berikut ini hasil dari sebuah Hierarcy Clustering

> idx <- sample(1:dim(Data27)[1],100)

> data27sample<-Data27[idx,]

> data27sample$Sex<-NULL

> hc <- hclust(dist(data27sample),method = "ave")

> plot(hc,hang = -1,labels = Data27$Sex[idx])

> clusterCut <- cutree(hc, 3)

> table(clusterCut, Data27$Sex)

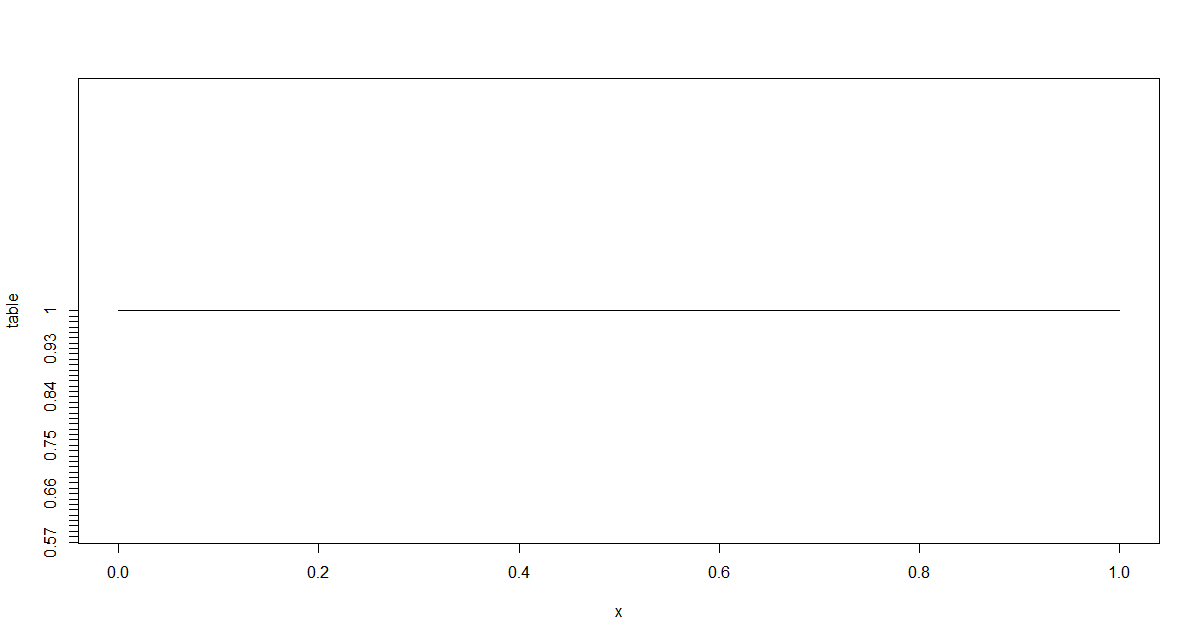
clusterCut F I M

1 9 20 13

2 13 23 16

3 4 1 1

* plot(table)



Dengan menggunakan *Hierarchical Cluster* dapat diketahui bahwa nilai yang terbanyak terdapat pada nilai height adalah 8.

**KESIMPULAN DAN SARAN**

Dari analisis di atas, dapat disimpulkan bahwa hasil dari clustering sebesar 82.5 % dapat diambil kesimpulan yaitu bahwa "Whole\_weight" dengan "Shucked\_weight" memiliki nilai yang hampir mirip. Dari hasil tersebut bisa dipastikan bahwa whole\_weight dengan shucked\_weight memiliki kesamanaan yang berbanding lurus. Dan hasil dari tabel cluster belum didapatkan sebuah keputusan karena nilai yang didapatkan seragam.