# ANALISIS KLUSTER MENGGUNAKAN BAHASA PEMOGRAMAN R UNTUK KAJIAN EKOLOGI

**Muhammad Wiharto**
Jurusan Biologi, FMIPA, Universitas Negeri Makassar
Gunung Sari Baru, Jl. A.P.Pettarani Makassar 90222
*e-mail*: wiharto09@gmail.com

**Abstract: The Analysis of Cluster by Using R Program for Ecology Study.** The aims of cluster analysis is to grouping the variables into cluster based on certain similarity. Programming with R language can processing the data to cluster classification with hierarchy method and visualizing in the form of dendogram. R packages that are needed where hclust, plot, and cutree while function that contained in Vegan that is use is vegdist.

**Abstrak: Analisis Kluster Menggunakan Bahasa Pemograman R untuk Kajian Ekologi.** Analisis kluster bertujuan untuk mengelompokkan variabel ke dalam kelompok berdasarkan kesamaan tertentu. Pemograman dengan bahasa R dapat mengolah data untuk klasifikasi kluster dengan metode hirarki dan ditampilkan dalam bentuk diagram dendogram. Package R yang di perlukan adalah Vegan. Fungsi-fungsi penting yang digunakan adalah hclust, plot, dan cutree sedangkan fungsi yang terdapat dalam Vegan yang digunakan adalah vegdist.

**Kata kunci:** *kluster, pemograman R, metode hirarki, kajian ekologi*

## A. PENDAHULUAN

Tujuan dari Analisis kluster adalah mengelompokkan variabel ke dalam kelompok berdasarkan kesamaan tertentu. Menurut Kindt & Coe (2005), penggunaan analisis kluster dalam kajian ekologi dilakukan untuk menyatukan tegakan-tegakan ke dalam kelompok. Tegakan-tegakan yang berkumpul dalam satu kelompok memiliki kesamaan dalam komposisi spesies, dan hal ini ditentukan berdasarkan jarak ekologi yang dipilih. Selanjutnya menurut Leps & Smilauer (2003), kesamaan ekologi tersebut juga dapat berupa kesamaan dalam tingkah laku ekologi, yang nampak pada kesamaan dalam distribusi. Santoso (2002); Santoso & Tjiptono (2002) mengatakan bahwa analisis kluster merupakan bagian dari statistik multivariat.

Analisis kluster dalam kajian ekologi banyak dimanfaatkan untuk kegiatan klasifikasi (Krebs, 1989). Analisis kluster merupakan salah satu dari metode klasifikasi numerik dalam ekologi, dan tujuan utama penggunaan metode ini adalah untuk mereduksi data (Kent & Coker, 1997).

R adalah suatu suite perangkat lunak yang digunakan untuk manipulasi data, perhitungan, simulasi, penayangan grafik, dan sekaligus sebagai bahasa pemograman yang bersifat interpreter. R diturunkan dari bahasa S, suatu bahasa pemograman yang dikembangkan di Laboratorium Bell. Oleh karena R berlisensi open source maka ia dapat diperoleh dan diedarkan secara cuma-cuma di bawah lisensi publik GNU.

R dapat dijalankan pada sistem operasi Window, Mac OS X, Unix, maupun Linux. Perangkat lunak R dapat diunduh dari situs CRAN (Comprehensive R Archive Network) website: http://lib.stat.cmu.edu/R/CRAN/. R adalah bahasa pemograman yang bersifat *object oriented*, sehingga seluruh variable, data, fungsi, dan lain sebagainya disimpan dalam memori komputer sebagai objek. R merupakan bahasa pemograman jenis interpreter yang bersifat *case sensitive* (Cohen & Cohen, 2008; Torgo, 2003; Ohri, 2012).

Program merupakan serangkaian instruksi yang memberitahu komputer sesuatu yang harus dilakukan (Zelle, 2002). Selanjutnya dikatakan bahwa, sebuah program adalah sejumlah instruksi-instruksi yang berisi baris perintah-

perintah dalam bahasa pemograman komputer untuk menyelesaikan masalah dengan bantuan komputer. Masalah-masalah komputasi tersebut mungkin berupa permasalahan seperti matematika, berupa penyelesaian beragam fungsi dan rumus, namun juga dapat berupa beragam permasalahan lainnya (Hendri, 2003).

R merupakan bahasa pemograman komputer yang memungkinkan pengguna untuk memrogramkan algoritme dan menggunakan alat yang telah dikembangkan melalui R oleh pengguna lainnya (Zuur *et al.*, 2009). R merupakan bahasa pemograman tingkat tinggi dan juga merupakan lingkungan untuk analisis data dan grafik. Desain R dipengaruhi sangat kuat oleh dua bahasa pemograman komputer yang telah ada sebelumnya, yaitu bahasa S yang dikembangkan oleh Becker, Chamber, dan Wilk, serta bahasa pemograman Scheme yang dikembangkan oleh Sussman. Dengan demikian bahasa ini sangat mirip dengan bahasa S, tetapi implementasi dan semantiknya ditopang oleh Scheme (Crawley, 2007). Hal penting yang terdapat pada R adalah suatu lingkungan pemograman yang bersifat belajar sendiri (Bocard *et al.*, 2011).

Kajian ekologi saat ini banyak memanfaatkan analisis secara kuantitatif, selain itu ukuran data yang harus diolah juga sangat besar. Untuk memudahkan didalam melakukan perhitungan kuantitatif dan menganalisis data maka diperlukan perangkat lunak yang dapat digunakan keperluan tersebut. Tujuan penelitian ini adalah untuk membangun program berbasis bahasa R untuk analisis kluster dalam kajian ekologi.

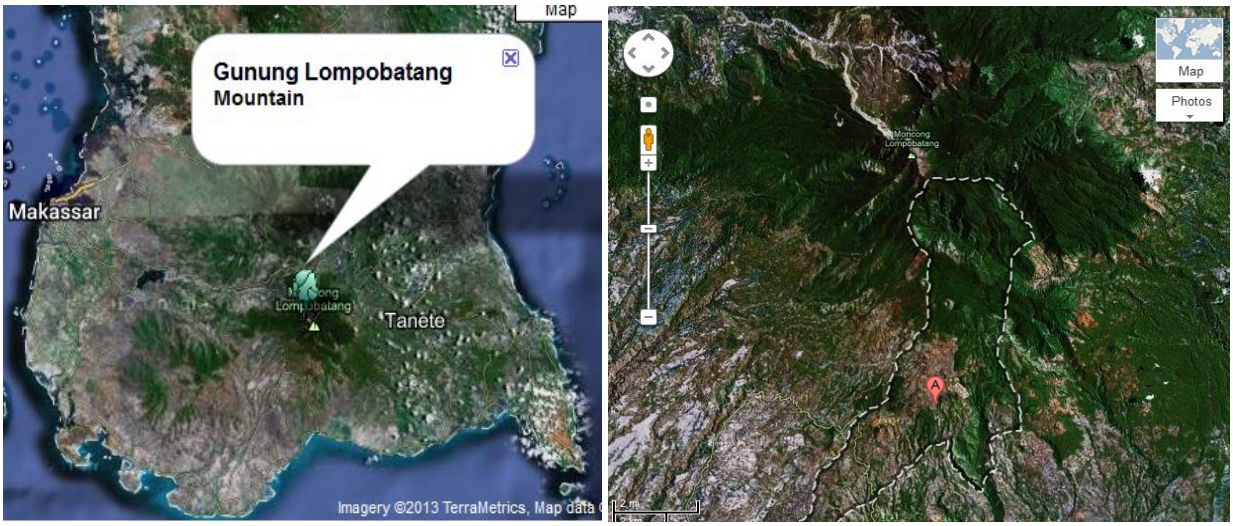
## B. METODE

Penelitian dilaksanakan di Laboratorium Jurusan Biologi FMIPA UNM, berlangsung selama bulan Mei dan Juli 2013. Data dari Alfian (2013), berupa data organisme hewan tanah yang diambil di kaki Gunung Lompobattang, daerah Lannyi, Desa Bontolojong, Kecamatan Uluere, Kabupaten Bantaeng, Provinsi Sulawesi Selatan (Gambar 1).

Langkah-langkah yang ditempuh adalah menyusun data ke dalam format yang dapat dibaca oleh program R dan di simpan dalam format *comma separated file* (csv). Kegiatan ini dilakukan dengan menggunakan Excel versi 2007. Script dengan program dengan R ini juga memanfaatkan package Vegan (2011). Versi perangkat lunak yang digunakan adalah R version 3.0.0. Metode yang digunakan untuk analisis kluster adalah metode hirarki.

## C. HASIL DAN PEMBAHASAN

Script program untuk analisis kluster dapat dilihat pada Tabel 1. Terdapat 53 buah baris program, dimana 30 buah baris program merupakan keterangan. Pada program ini terdapat satu user defined program yang diperoleh dari Sanches (tanpa tahun). Baris 6 pada program digunakan untuk membersihkan memori. Program kemudian diarahkan ke direktori tempat penyimpanan data (baris 9). Data yang dalam file bernama cacing.csv




**Gambar 1. Lokasi Pengambilan Data. Sumber: Google Map (Satelite) dalam Alfian (2013)**

kemudian dibaca oleh komputer, dan disimpan dengan nama data1 (baris 10).

Tabel 2 bagian a, menampilkan 6 baris pertama dari data, dan bagian b, untuk enam baris terakhir dari data. Perintah yang terdapat pada fungsi head berguna untuk menampilkan baris pertama, dan fungsi tail untuk menampilkan 6 baris terakhir. Jumlah enam baris data yang ditampilkan oleh kedua fungsi ini adalah jumlah default, sehingga dapat diubah-ubah sesuai dengan tujuan program. Perintah nampak pada baris 12 dan 14.

Pada baris 16, dimanfaatkan untuk mengaktifkan package Vegan. Baris 18 digunakan untuk menentukan jarak antara tegakan, dan metode yang digunakan adalah metode Jaccard. Agglomerasi untuk pengelompokan tegakan-tegakan yang memiliki kesamaan yang tinggi dilakukan dengan metode Ward (baris 21). Jumlah kelompok tegakan yang dibentuk ditetapkan maksimal 4. Fungsi yang digunakan untuk menampilkan jumlah kelompok tegakan ini adalah cutree yang terdapat pada baris 23.

Kelompok-kelompok tegakan yang terbentuk dapat dilihat pada Tabel 3. Tegakan 1, 3, 5, 7, 8, dan 10 nampak tergabung ke dalam kelompok tegakan 1. Tegakan 2 dan 6 tergabung ke dalam kelompok tegakan 2, dan tegakan 4 dan 5 tergabung ke dalam kelompok tegakan 3. Tegakan-tegakan lainnya yang tersisa tergabung ke dalam kelompok tegakan 4. Baris-baris program yang digunakan untuk menghasilkan kelompok tegakan ini adalah baris 23 sampai 25.

**Tabel 1. Script Program Menggunakan Bahasa R untuk Mengelompokkan Tegakan dengan Menggunakan Analisis Kluster.**

```
 1  .  #-------------------------------------------------------
 2  .  #---- Programmer: Dr. Ir. Muhammad Wiharto Caronge, M.Si -
 3  .  #-----                                                  --
 4  .  #-- (1): Makassar 19 Mei 2013 ----------------------------
 5  .  #-------------------------------------------------------
 6  .  rm(list=ls(all=TRUE))
 7  .  #-------------------------------------------------------
 8  .  #--- Mengambil data ---------------------------------------
 9  .  setwd('d:/Download/Biodiversity/Seminar Biologi FMIPA UNM/)
       data1  <-read.csv('cacing.csv',header=TRUE, row.names=1,
10  .  stringsAsFactors = FALSE)
11  .  #--- Menampilkan enam data pertama ----------------------
12  .  head(data1)
13  .  #--- Menampilkan enam data terakhir ----------------------
14  .  tail(data1)
15  .  #--- Mengaktifkan package Vegan --------------------------
16  .  library(vegan)
17  .  #--- program jarak indeks Jaccard  ----------------------
18  .  data <- vegdist(data1, method = 'jaccard')
19  .  #--- metode pengelompokan yang dipakai ----
20  .  #--- adalah metode Ward -------------------
21  .  kluster_hirarki <- hclust(dist(data), method = 'ward')
22  .  #--- Melihat pengelompokan ----------------
23  .  group <-cutree(kluster_hirarki,4)
24  .  kelompok <-cbind(group)
25  .  Kelompok
26  .  #----------------------------------------
27  .  kluster_hirarki.a <- as.dendrogram(kluster_hirarki)
28  .  #----------------------------------------
29  .  #--- membuat warna ------------------------
30  .  #--- Warna yang digunakan adalah merah, ---
```

```
31  .  #--- merah,biru,hitam,hijau ----
32  .  #--- dan ungu ----------------------------
Sambungan Tabel 1.
33  .  warna = c("red", "blue", "black", "green")
34  .  #--- menentukan 4 cluster -----------------
35  .  jlhkelompok = cutree(kluster_hirarki, 4)
36  .  #--- Fungsi untuk memberi warna -----------
37  .  #--- diperoleh dari:    -------------------
38  .  #--- http://rpubs.com/gaston/dendrograms ---
39  .  #--- Gaston Sanches -- Visualizing Dendrograms in R ---
40  .  #----------------------------------------
41  .  warnaklp <- function(n) {
           if (is.leaf(n)) {
           a <- attributes(n)
           warnagbr <- warna[jlhkelompok[which
                                    (names(jlhkelompok)
           == a$label)]]
           attr(n, "nodePar") <- c(a$nodePar,
           lab.col = warnagbr)
           }
           N
       }
42  .  #----------------------------------------
43  .  #------- memanfaatkan fungsi dendrapplym -----
44  .  gambar <- dendrapply(kluster_hirarki.a, warnaklp)
45  .  #------------------------------------------------------
46  .  # -- membuat plot dendogram ------------------
       plot(gambar, col='blue',main='',sub='',xlab='Tegakan',
47  .  ylab='Indeks Jaccard')
48  .  #----------------------------------------
49  .  #------------------------------------------------------
       plot2 <- plot(kluster_hirarki, hang=-1,col='black',
50  .  main='',sub='',xlab='Tegakan',ylab='Indeks Jaccard')
51  .  rect.hclust(kluster_hirarki, k=4, border="red")
52  .  plot2
53  .  #------------------------------------------------------
```

Tabel 3 juga menunjukkan model data yang digunakan untuk analisis kluster dengan R, sehingga melalui model data yang demikian ini data antar baris dapat dikorelasikan. Menurut Santosa (2002) model data yang demikian ini adalah model data untuk Q analisis dan analisis kluster merupakan bagian dari Q analisis.

Dendogram dihasilkan melalui baris-baris program 27 sampai 44. Baris-baris program ini merupakan modifikasi dari Sanches (tanpa tahun). Grafik dendogram dihasilkan melalui baris-baris program 46 sampai 52. Dendogram yang dihasilkan pada Gambar 1 ditampilkan melalui perintah pada baris 47, sedangkan pada Gambar 2 dihasilkan melalui perintah pada baris 50 sampai 52. Baris 41 merupakan fungsi yang akan menghasilkan warna-warna yang berbeda pada setiap kelompok tegakan. Baris 51 merupakan fungsi yang akan menghasilkan kotak pada kelompok tegakan yang berbeda. Fungsi plot merupakan fungsi bawaan dari R dan digunakan untuk menampilkan dendogram.

Dendogram pada Gambar 1 menunjukkan adanya 4 buah kelompok tegakan yang terbentuk pada jarak indeks Jaccard di bawah 5. Hal yang serupa nampak pada Gambar 2. Pada Gambar 1 kelompok tegakan dibedakan berdasarkan warna, yakni warna merah merupakan kelompok tegakan 1, warna biru merupakan kelompok tegakan 2, warna hitam

**Tabel 2. Data Penelitian. (a) Enam Baris Data Pertama, (b) Enam Baris Data Kedua.**

a. Enam baris data pertama

| | cacing.tanah | cacing.merah | cacing.lumbricus | cacing.pheretima | cacing.perionyx |
|---|---|---|---|---|---|
| T1 | 2 | 0 | 0 | 0 | 0 |
| T2 | 1 | 0 | 0 | 0 | 0 |
| T3 | 2 | 1 | 0 | 0 | 0 |
| T4 | 0 | 1 | 0 | 0 | 0 |
| T5 | 2 | 0 | 0 | 0 | 0 |
| T6 | 1 | 0 | 0 | 0 | 0 |

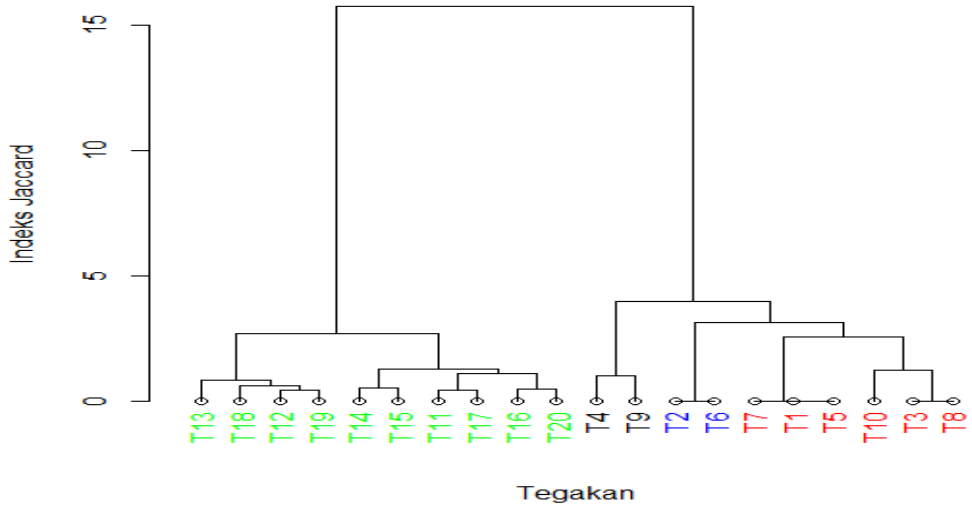
b. Enam baris data terakhir.

| | cacing.tanah | cacing.merah | cacing.lumbricus | cacing.pheretima | cacing.perionyx |
|---|---|---|---|---|---|
| T15 | 0 | 0 | 5 | 1 | 0 |
| T16 | 0 | 0 | 3 | 1 | 1 |
| T17 | 0 | 0 | 4 | 0 | 0 |
| T18 | 0 | 0 | 1 | 2 | 1 |
| T19 | 0 | 0 | 2 | 2 | 1 |
| T20 | 0 | 0 | 3 | 0 | 1 |

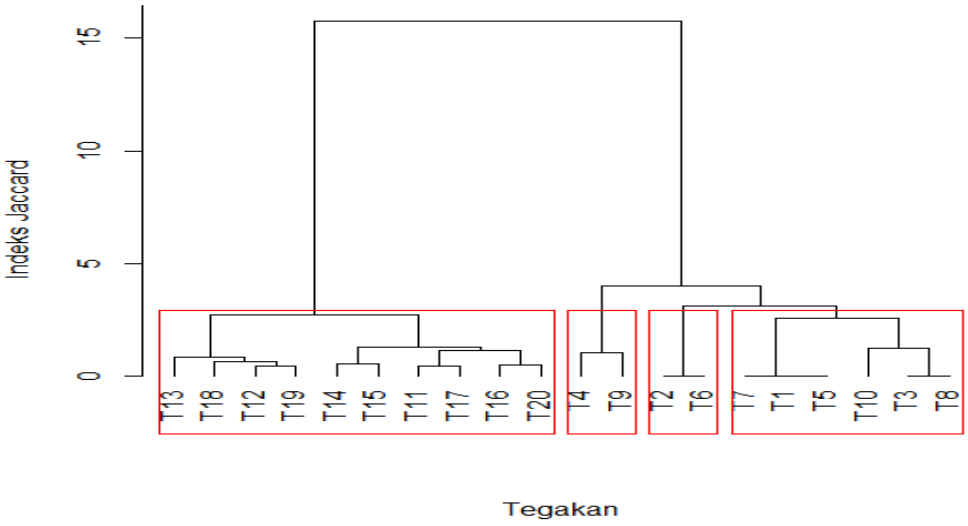
**Tabel 3. Kelompok Tegakan yang Terbentuk melalui Analisis Kluster.**

| Tegakan* | Kelompok | Tegakan* | Kelompok |
|---|---|---|---|
| T1 | 1 | T11 | 4 |
| T2 | 2 | T12 | 4 |
| T3 | 1 | T13 | 4 |
| T4 | 3 | T14 | 4 |
| T5 | 1 | T15 | 4 |
| T6 | 2 | T16 | 4 |
| T7 | 1 | T17 | 4 |
| T8 | 1 | T18 | 4 |
| T9 | 3 | T19 | 4 |
| T10 | 1 | T20 | 4 |

Keterangan: T: Tegakan




**Gambar 2. Dendoram Hasil Analisis Kluster dengan Warna sebagai Pembeda Kelompok Tegakan. Keterangan: T: Tegakan**




**Gambar 3. Dendoram Hasil Analisis Kluster dengan Kotak sebagai Pembeda Kelompok Tegakan. Keterangan: T: Tegakan.**

merupakan kelompok tagakan 3, dan warna hijau merupakan kelompok tegakan 4. Kelompok tegakan pada Gambar 2 dibedakan berdasarkan kotak, dimana setiap kotak meupakan suatu kelompok tegakan.

Penggunaan *package* Vegan dalam program ini karena memiliki beragam fungsi-fungsi yang terkait dengan kajian ekologi. Pada *package* ini, metode jarak dapat diakses melalui fungsi *vegdist*, sedangkan fungsi yang digunakan untuk mengelompokkan tegakan adalah melalui fungsi *hclust*, yang merupakan fungsi bawaan dari R. Dixon (2009) mengatakan bahwa, Vegan memiliki fungsi-fungsi dari untuk analisis vegetasi sampai kepada beragam program statistik, dan package ini dapat diunduh secara gratis pada situs R.

Metode jarak yang tersedia dan digunakan untuk indeks similaritas di dalam fungsi *vegdist* adalah: (1) manhattan, (2) euclidean, (3) canberra, (4) bray, (5) kulczynski, (6) jaccard, (7) gower, (8) altGower, (9) morisita, (10) horn, (11) mountford, (12) raup, (13) binomial, dan (14) chao, dimana yang menjadi default adalah 'bray' atau Bray-Curtis.

Indeks kesamaan sering digunakan untuk mengkaji kehadiran bersama suatu spesies atau kesamaan dari tegakan-tegakan sampel. Suatu matriks koefisien kesamaan baik di antara spesies atau lokasi, dapat dianalisa dengan dua cara, yaitu, ordinasi, yakni kegiatan yang bertujuan untuk mengatur lokasi-lokasi atau tegakan atau spesies ke dalam suatu urutan yang secara teoritis kontinu. Cara lainnya adalah melalui klasifikasi, yaitu kegiatan yang bertujuan untuk menempatkan lokasi atau tegakan atau spesies ke dalam kelompok-kelompok yang bersifat diskontinu (McCoy *et al*., 1986 *dalam* Real & Vargas, 1996).

Indeks Jaccard merupakan salah satu indeks kesamaan yang digunakan dalam kajian ekologi. Rumus indeks ini adalah sebagai berikut: J = C/ A + B - C, dimana J adalah nilai Indeks Jaccard, C = jumlah spesies umum yang hadir bersama pada suatu tegakan, A = jumlah spesies yang unik terhadap tegakan pertama, dan B = jumlah spesies yang unik terhadap tegakan kedua (Real & Vargas, 1996). Indeks ini sering juga digunakan untuk kajian konservasi spesies, oleh karena dapat digunakan untuk *power function* dalam menentukan keterkaitan antara spesies dan area, dan untuk menentukan ukuran optimal suatu cagar alam (Higgs & Usher, 1980).

Fungsi hclust menghasilkan objek yang mengandung informasi yang penting untuk menjelaskan hasil klusterisasi sepenuhnya (Borcard *et al*., 2011), fungsi ini merupakan fungsi bawaan dari R. *Complete linkage* mungkin merupakan metode agglomerasi terbaik untuk sebagian besar data-data paleontologi atau data geologi; namun demikian metode *average linkage* ataupun juga metode Ward banyak digunakan untuk bidang akjian lainnya dan hal ini tergantung kepada data. Fungsi hclust memungkinkan untuk penggunaan metode agglomerasi berikut ini, (1) ward, (2) single, (3) complete, (4) average, (5) mcqutty, (6) median, (7) centroid, dengan cara menspesifikasikan

metode yang akan digunakan pada fungsi ini. Kent & Coker (1997) mengatakan bahwa metode Ward mungkin meupakan metode yang paling optimal untuk analisis kesamaan.

Dendogram pada Gambar 1 dan 2 merupakan hasil pengelompokan data, dalam hal ini tegakan dengan metode hirarki. Dendogram menurut Mayr *et al*., (1953) *dalam* Clifford & Stevenson (1975) adalah ilustrasi diagramatik dari suatu relasi berdasarkan tingkat kesamaan. Santoso (2005) mengatakan bahwa melalui konsep ini, dua data pada awalnya digabungkan, dan penggabungan didasarkan pada kesamaan yang ada pada data. Penggabungan terus berlanjut terhadap data lainnya yang memiliki kemiripan. Penggabungan ini membentuk tampilan seperti struktur pohon, sehingga oleh Krebs (1989) disebut juga dengan metode *agglomerative*, yakni suatu metode klasifikasi yang dimulai dari satu perangkat tegakan dan kemudian menggabungkannya mengelompok dengan tegakan lain ke dalam bentuk kelas-kelas. Kent & Coker (1997) mengatakan bahwa metode ini juga bersifat *polythethic*, yaitu proses pengelompokan tegakan berdasarkan ke-seluruhan data yang ada.

Proses klusterisasi merupakan prosedur yang bersifat *heuristic*, dan bukan merupakan suatu uji statistik. Pilihan kepada suatu koefisien asosiasi dan metode klustering akan mempengaruhi hasil yang di peroleh. Hal ini menekankan pada pentingnya metode yang dipilih yang konsisten dengan tujuan analisis (Borcard *et al*., 2011).

## D. KESIMPULAN

Pemograman dengan bahasa R dapat mengolah data untuk klasifikasi kluster dengan metode hirarki dan ditampilkan dalam bentuk diagram dendogram. Package R yang di perlukan adalah Vegan. Fungsi-fungsi penting yang digunakan adalah hclust, plot, dan cutree sedangkan fungsi yang terdapat dalam Vegan yang digunakan adalah vegdist.

## E. DAFTAR PUSTAKA


Alfian, A. 2013. *Laporan PKL Ekologi Hewan*. Jurusan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar.

Borcard, D., F. Gillet., & P. Legendre. 2011. *Numerical Ecology with R*. Springer. New York, Dordrecht, London, Heidelberg.

Clifford, H.T., & W. Stepenson. 1975. *An Introduction to Numerical Classification*. Academic Press, New York, San Francisco, London.

Cohen, Y., & J.Y. Cohen. 2008. *Statistics and Data with R:An applied approach through examples*. John Wiley & Sons Ltd. United Kingdom.

Crawley, M. J. 2007. *The R Book*. John Wiley & Sons Ltd, England.

Dixon, P. 2009. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, Vol. 14, Issue 6, pages 927–930

Hendri. 2003. *Cepat Mahir Python*. Kuliah Umum IlmuKomputer.com. IlmuKomputer.com.

Higgs, A. J., & M.B. Usher. 1980. Should reserves be large or small? *Nature* 285:568-569.

Kent, M., & P. Coker. 1997. *Vegetation Description and Analysis. A Practical Approach*. CRC Press & Belhaven Press, Boca Alton, Ann Arbor, London.

Kindt, R. & R. Coe. 2005. *Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies*. World Agroforestry Centre (ICRAF), Nairobi.

Krebs, C.J. 1989. *Ecological Methodology. Harper and Row Publishers*, New York, Singapore, Sidney.

Leps, J., & P. Smilauer. 2003. *Multivariate Analysis of Ecological Data Using CANOCO*. Cambridge University Press, Cambridge, New York, Melbourn, Madrid, Cape Town, Singapore, Sao Paolo.

Ohri, A. 2012. *R for Business Analytics*. Springer, New York, Heidelberg, Dordrecht, London

Real, R., & J.M. Vargas. 1996. The Probabilistic Basis of Jaccard's Index of Similarity. *Syst. Biol*. 45(3): 380-385.

Sanchez. G. Tanpa Tahun. Visualizing dendogram in R. RPubs brougth to you by RStudio. http://rpubs.com/gaston/dendrograms [1 Januari 2013]

Santoso, S. 2002. *Buku Latihan SPSS Statistik Multivariat*. PT Elex Media Komputindo, Jakarta.

Santoso, S., & F. Tjiptono. 2002. *Riset Pemasaran*. Konsep dan aplikasi dengan SPSS. PT Elex Media Komputindo, Jakarta.

Torgo, L. 2003. *Data Mining with R: learning by case studie*. LIACC-FEP, University of Porto, Porto.

Vegan. 2011. Vegan Utils version 3.0. Jari Oksanen.

Zelle, J.M. 2002. *Python Programming: An Introduction to Computer Science*. Version 1.0rc2. Wartburg College Printing Services.

Zuur, F.A., E. N. Ienol., & E. H.W.G. Meesters. 2009. *A Beginner's Guide to R*. Springer, Dordrecht, Heidelberg, London, New York.