



***SELF ORGANIZING MAP (SOM) CLUSTERING* BERBASIS WEB  
INTERAKTIF DENGAN R-SHINY**

**SKRIPSI**

Oleh  
**Okita Tazkiyah**  
**NIM 121810101045**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS JEMBER  
2016**



***SELF ORGANIZING MAP (SOM) CLUSTERING* BERBASIS WEB  
INTERAKTIF DENGAN R-SHINY**

**SKRIPSI**

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat  
untuk menyelesaikan Program Studi Matematika (S1)  
dan mencapai gelar Sarjana Sains

Oleh  
**Okrit Tazkiyah**  
**NIM 121810101045**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS JEMBER  
2016**

## PERSEMBAHAN

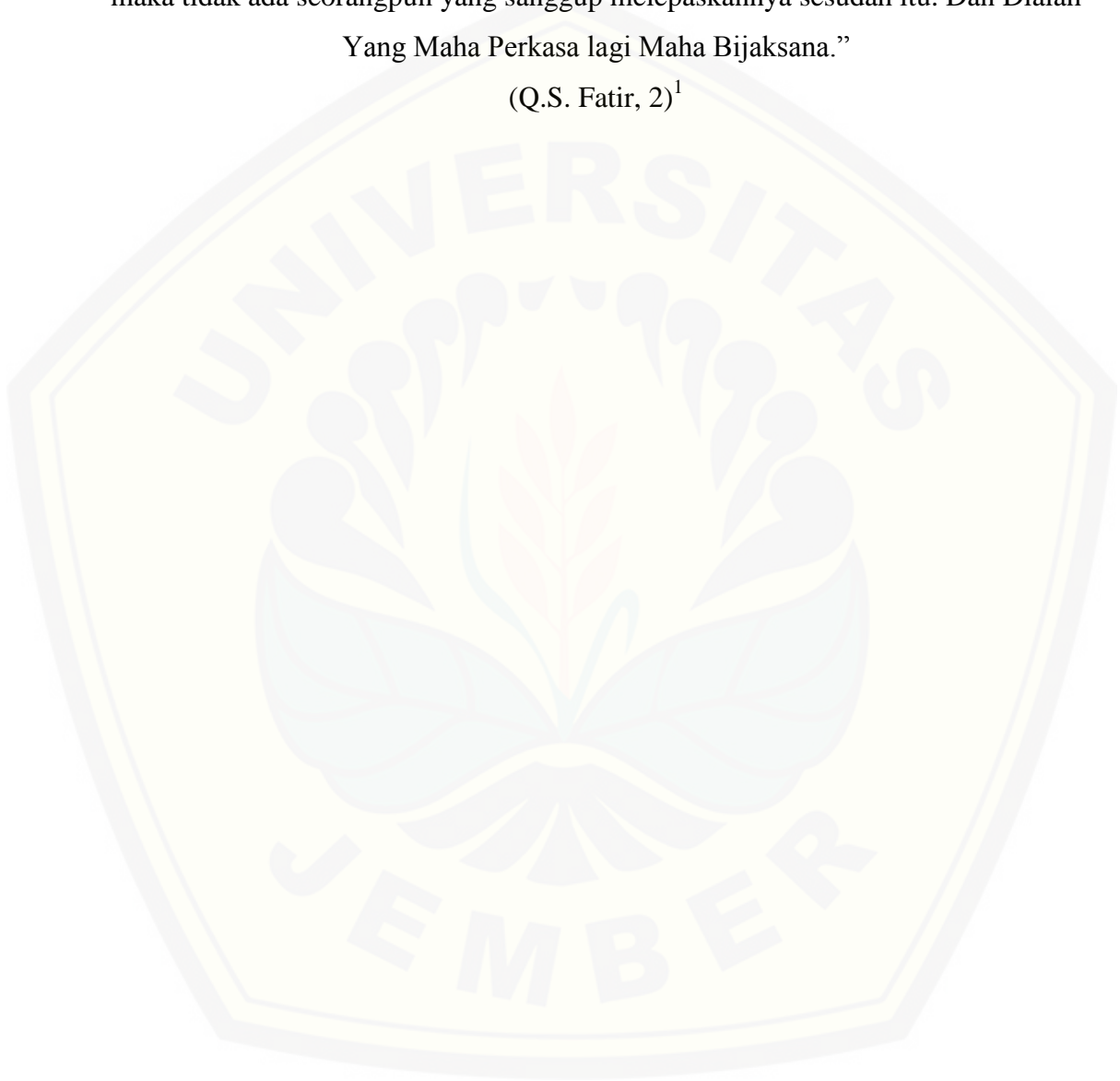
Skripsi ini saya persembahkan untuk:

1. Ayahanda Iksan dan Ibunda Hermani yang telah memberikan do'a, cinta dan semangat;
2. kakak tercinta Govinda Heri Susanto, adik tercinta Muhammad Kaulan Karima serta seluruh keluarga yang telah memberikan dukungan;
3. seluruh guru dan dosen sejak taman kanak-kanak hingga perguruan tinggi yang telah membimbing keingintahuan serta memberikan ilmu kepada saya dengan penuh perhatian dan kasih sayang;
4. Almamater Jurusan Matematika FMIPA Universitas Jember, SMA Negeri 1 Gending, SMP Negeri I Maron, SDN Maron Wetan I, serta TK Wijaya Kusuma I.

**MOTO**

“Apa saja yang Allah anugerahkan kepada manusia berupa rahmat, maka tidak ada seorangpun yang dapat menahannya; dan apa saja yang ditahan oleh Allah maka tidak ada seorangpun yang sanggup melepaskannya sesudah itu. Dan Dialah Yang Maha Perkasa lagi Maha Bijaksana.”

(Q.S. Fatir, 2)<sup>1</sup>



---

<sup>1</sup> Departemen Agama Republik Indonesia. 2011. Al-Kitabul Akbar Al-Qur'an dan Terjemahannya. Jakarta Timur: PT. Akbar Media Eka Sarana.

**PERNYATAAN**

Saya yang bertanda tangan di bawah ini:

Nama : Okit Tazkiyah

NIM : 121810101045

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “*Self Organizing Map (SOM) Clustering Berbasis Web Interaktif dengan R-Shiny*” adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada institusi manapun, dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenar-benarnya, tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, Desember 2016

Yang menyatakan,

Okit Tazkiyah

NIM 121810101045

**SKRIPSI**

***SELF ORGANIZING MAP (SOM) CLUSTERING BERBASIS WEB  
INTERAKTIF DENGAN R-SHINY***

Oleh

Okita Tazkiyah  
NIM 121810101045

Pembimbing:

Dosen Pembimbing Utama : Prof. Drs. I Made Tirta, M.Sc., Ph.D.

Dosen Pembimbing Anggota : Dian Anggraeni, S.Si., M.Si.

**PENGESAHAN**

Skripsi yang berjudul “*SELF ORGANIZING MAP (SOM) CLUSTERING BERBASIS WEB INTERAKTIF DENGAN R-SHINY*” telah diuji dan disahkan pada:

hari, tanggal :

tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas  
Jember

Tim Penguji,

Ketua,

Sekretaris,

Prof. Drs. I Made Tirta, M.Sc., Ph.D.  
NIP 197407192000121001

Dian Anggraeni, S.Si., M.Si.  
NIP 198202162006042002

Anggota I,

Anggota II,

Dr. Alfian Futuhul Hadi, S.Si., M.Si.  
NIP 197407192000121001

Ahmad Kamsyakawuni, S.Si., M.Kom.  
NIP 197211291998021001

Mengesahkan,

Dekan,

Drs. Sujito, Ph.D.  
NIP 196102041987111001

## RINGKASAN

***Self Organizing Map (SOM) Clustering Berbasis Web Interaktif Dengan R-Shiny***; Okit Tazkiyah 121810101045; 2016; 43 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

*Clustering* merupakan proses pengelompokan objek-objek ke dalam *cluster*, dengan tujuan objek data dengan kemiripan yang tinggi akan berada pada *cluster* yang sama dan objek data dengan kemiripan berbeda akan berada pada *cluster* berbeda. Algoritma *clustering* dibagi menjadi beberapa macam, salah satunya *neural network based-clustering*. Pada algoritma *Neural network-based clustering* terdapat salah satu metode *clustering* yaitu *Self Organizing Map (SOM)* dimana hasil *clustering* mampu memperlihatkan variabel dominan. *Self Organizing Map (SOM)* menganggap sekumpulan *node* (unit) merupakan objek data untuk melakukan *clustering*, dimana dalam setiap *node* (unit) memuat beberapa objek data input.

Pada penelitian ini *Self Organizing Map (SOM) clustering*, yang biasa dikerjakan dengan bantuan program R berbasis skrip dikembangkan menjadi program analisis berbasis web interaktif menggunakan R-Shiny. Tujuannya adalah untuk memudahkan seseorang yang kurang mengerti program R berbasis skrip dalam melakukan analisis data menggunakan *Self Organizing Map (SOM) clustering*, yaitu hanya dengan memasukkan data pada program web secara *online* dan memilih menu-menu yang disediakan. Program analisis *Self Organizing Map (SOM) clustering* berbasis web interaktif ini memiliki kemampuan atau fitur yang hampir sama dengan program R berbasis skrip. Program web interaktif ini dilengkapi dengan ringkasan teori mengenai *Self Organizing Map (SOM) clustering* yang berguna sebagai dasar dalam melakukan analisis data dan untuk menambah wawasan pengguna tentang salah satu model statistika ini. Selain itu, program web interaktif ini juga berbasis tutorial yang memudahkan pengguna dalam melakukan analisis data. Web interaktif program *Self Organizing Map*



(SOM) *clustering* dapat diakses melalui alamat <http://statslab-rshiny.fmipa.unej.ac.id/JORS/SOM/>

Program web interaktif yang telah dibuat kemudian diaplikasikan untuk menganalisis suatu data. Pada penelitian ini, program web tersebut diaplikasikan pada data Indeks Pembangunan Manusia (IPM) provinsi Aceh tahun 2013. Indeks Pembangunan Manusia (IPM) merupakan ukuran standar pembangunan manusia berdasarkan komponen kualitas hidup, meliputi: kesehatan, pengetahuan dan kehidupan layak. Komponen tersebut diukur menggunakan angka harapan hidup, angka melek huruf, rata-rata lama sekolah dan pengeluaran perkapita. Adapun variabel *input* dari data penelitian ini terdiri atas data nama daerah (*Nama.Daerah*), angka harapan hidup (*AHH*), angka melek huruf (*AMH*), rata-rata lama sekolah (*RLS*) dan pengeluaran perkapita (*PPK*). Pada bagian eksplorasi data, didapatkan variabel faktor ialah nama daerah (*Nama.Daerah*) sedangkan untuk variabel numerik diantaranya angka harapan hidup (*AHH*), angka melek huruf (*AMH*), rata-rata lama sekolah (*RLS*) dan pengeluaran perkapita (*PPK*). Dari proses *clustering* diperoleh 4 *cluster*, berdasarkan jumlah *cluster* ideal pada penggolongan kategori suatu daerah Berikut hasil *clustering* menggunakan metode *Self rganizing Map* (SOM).

1. *Cluster* 1 berangotakan Aceh, Aceh Singkil, Aceh Selatan, Pidie, Aceh Barat Daya, Gayo Lues, Nagan Raya, Subulussalam dan Pidie Jaya, dengan variabel *AMH* (angka melek huruf) cukup tinggi dibandingkan variabel lainnya dengan kategori dengan tingkat Indeks Pembangunan Manusia (IPM) menengah;
2. *Cluster* 2 berangotakan Aceh Tenggara, Aceh Timur, Aceh Tengah, Aceh Barat, Aceh Besar, Bireuen, Aceh Utara, Aceh Jaya, Aceh Tamiang, Bener Meriah dan Langsa dengan dominasi variabel *AHH* (angka harapan hidup) dan variabel *AMH* (angka melek huruf) dengan kategori dengan tingkat Indeks Pembangunan Manusia (IPM) tinggi;
3. *Cluster* 3 berangotakan Banda Aceh, Sabang dan Lhokseumawe dengan semua variabel dominan dengan kategori dengan tingkat Indeks Pembangunan Manusia (IPM) sangat tinggi, dan

4. *Cluster* 4 berangotakan Simeulue dimana dominasi variabel *AMH* (angka melek huruf) sedangkan variabel *AHH* rendah dengan kategori dengan tingkat Indeks Pembangunan Manusia (IPM) rendah.



## PRAKATA

Puji syukur kehadirat Tuhan Yang Maha Esa atas segala rahmat, taufik, serta hidayah-Nya, sehingga penulis dapat menyelesaikan penulisan tugas akhir yang berjudul “*Clustering Self Organizing Map (Som) Berbasis Web Interaktif Dengan R-Shiny (Studi Kasus: Indeks Pembangunan Manusia Indonesia 2007)*”. Skripsi ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan strata satu (S1) pada Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyusunan tugas akhir ini tidak luput dari bantuan berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Drs. I Made Tirta, M.Sc., Ph.D. selaku Dosen Pembimbing Utama dan Ibu Dian Anggraeni, S.Si., M.Si. selaku Dosen Pembimbing Anggota yang telah bersedia meluangkan waktu, memberikan saran serta bimbingannya dalam penulisan skripsi ini;
2. Bapak Dr. Alfian Futuhul Hadi, S.Si., M.Si. selaku Dosen Penguji I dan Bapak Ahmad Kamsyakawuni, S.Si., M.Kom. selaku Dosen Penguji II yang telah memberikan kritik serta saran pada penulisan skripsi ini;
3. seluruh staf pengajar jurusan Matematika Fakultas MIPA Universitas Jember yang telah memberikan ilmu serta bimbingannya sehingga penulis dapat menyelesaikan skripsi ini;
4. Ayahanda Iksan dan Ibunda Hermani yang telah memberikan do'a, cinta dan semangat;
5. kakak tercinta Govinda Heri Susanto, adik tercinta Muhammad Kaulan Karima serta seluruh keluarga yang telah memberikan dukungan;
6. teman seperjuangan matematika angkatan 2012 (BATHICS'12), Dina Puspita Anggraini, Herlinda E.P. Sari, Lilik Krisna Mukti, Lintang Apriliantika Sasmita, Nurlailatus Salama, Sarwienda Cahya Utari, Selly Ringgit Garnasih, Selvi Nurika Kristina dan Veronika Susanti, yang telah memberikan dorongan dan motivasi;
7. semua pihak yang tidak dapat saya sebutkan satu per satu.

Skripsi ini juga tidak lepas dari kekurangan dan kesalahan baik isi maupun susunannya. Oleh karena itu, penulis menerima segala kritik dan saran dari semua pihak demi kesempurnaan skripsi ini. Penulis berharap semoga skripsi ini dapat memberi manfaat bagi pembaca dan kehidupan.

Jember, Desember 2016

Penulis



**DAFTAR ISI**

	Halaman
<b>HALAMAN JUDUL</b> .....	i
<b>HALAMAN PERSEMBAHAN</b> .....	ii
<b>HALAMAN MOTO</b> .....	iii
<b>HALAMAN PERNYATAAN</b> .....	iv
<b>HALAMAN PEMBIMBING</b> .....	v
<b>HALAMAN PENGESAHAN</b> .....	vi
<b>RINGKASAN</b> .....	vii
<b>PRAKATA</b> .....	x
<b>DAFTAR ISI</b> .....	xii
<b>DAFTAR TABEL</b> .....	xiv
<b>DAFTAR GAMBAR</b> .....	xv
<b>DAFTAR LAMPIRAN</b> .....	xvi
<b>BAB 1. PENDAHULUAN</b> .....	1
<b>1.1 Latar Belakang</b> .....	1
<b>1.2 Rumusan Masalah</b> .....	3
<b>1.3 Tujuan</b> .....	3
<b>1.4 Manfaat</b> .....	3
<b>BAB 2. TINJAUAN PUSTAKA</b> .....	4
<b>2.1 Jaringan Saraf Tiruan (JST)</b> .....	4
<b>2.2 Analisis Cluster</b> .....	6
<b>2.3 Self Organizing Map (SOM)</b> .....	6
2.3.1 Pengertian <i>Sel Organizing Map</i> (SOM).....	6
2.3.2 Algoritma <i>Sel Organizing Map</i> (SOM).....	8
<b>2.4 Validasi Cluster</b> .....	10
2.4.1 Validasi Internal .....	10
2.4.2 Validasi Stabilitas.....	11
2.4.3 Validasi Biologis .....	13

2.5 Indeks Pembangunan Manusia (IPM) .....	14
2.6 R-Shiny .....	14
<b>BAB 3. METODE PENELITIAN</b> .....	17
3.1 Data .....	17
3.2 Langkah-Langkah Penelitian .....	17
3.2.1 Tahap penyusunan web interaktif .....	18
3.2.2 Tahap aplikasi web interaktif .....	20
<b>BAB 4. HASIL DAN PEMBAHASAN</b> .....	22
4.1 Hasil dan Pembahasan Fitur Web Interaktif .....	22
4.1.1 Penyusunan Web Interaktif .....	22
4.1.2 Tampilan Web Interaktif .....	27
4.2 Hasil dan Pembahasan Analisis Data Indeks Pembangunan Manusia (IPM) .....	31
4.2.1 Eksplorasi Data Indeks Pembangunan Manusia (IPM) .....	31
4.2.2 <i>Clustering</i> .....	32
4.2.3 Visualisasi <i>Self Organizing Map (SOM) clustering</i> .....	33
<b>BAB 5. PENUTUP</b> .....	40
5.1 Kesimpulan .....	40
5.2 Saran .....	41
<b>DAFTAR PUSTAKA</b> .....	42
<b>LAMPIRAN</b> .....	44

**DAFTAR TABEL**

	Halaman
2.1 Perbandingan jaringan saraf biologi dan jaringan saraf tiruan .....	5
2.2 Nilai validasi .....	13
4.1 Anggota <i>cluster</i> (nomor <i>node</i> ) .....	35
4.2 Konversi anggota <i>node</i> (unit) pada objek data Indeks Pembangunan Manusia (IPM).....	38
4.3 Anggota <i>cluster</i> Indeks Pembangunan Manusia (IPM) .....	39

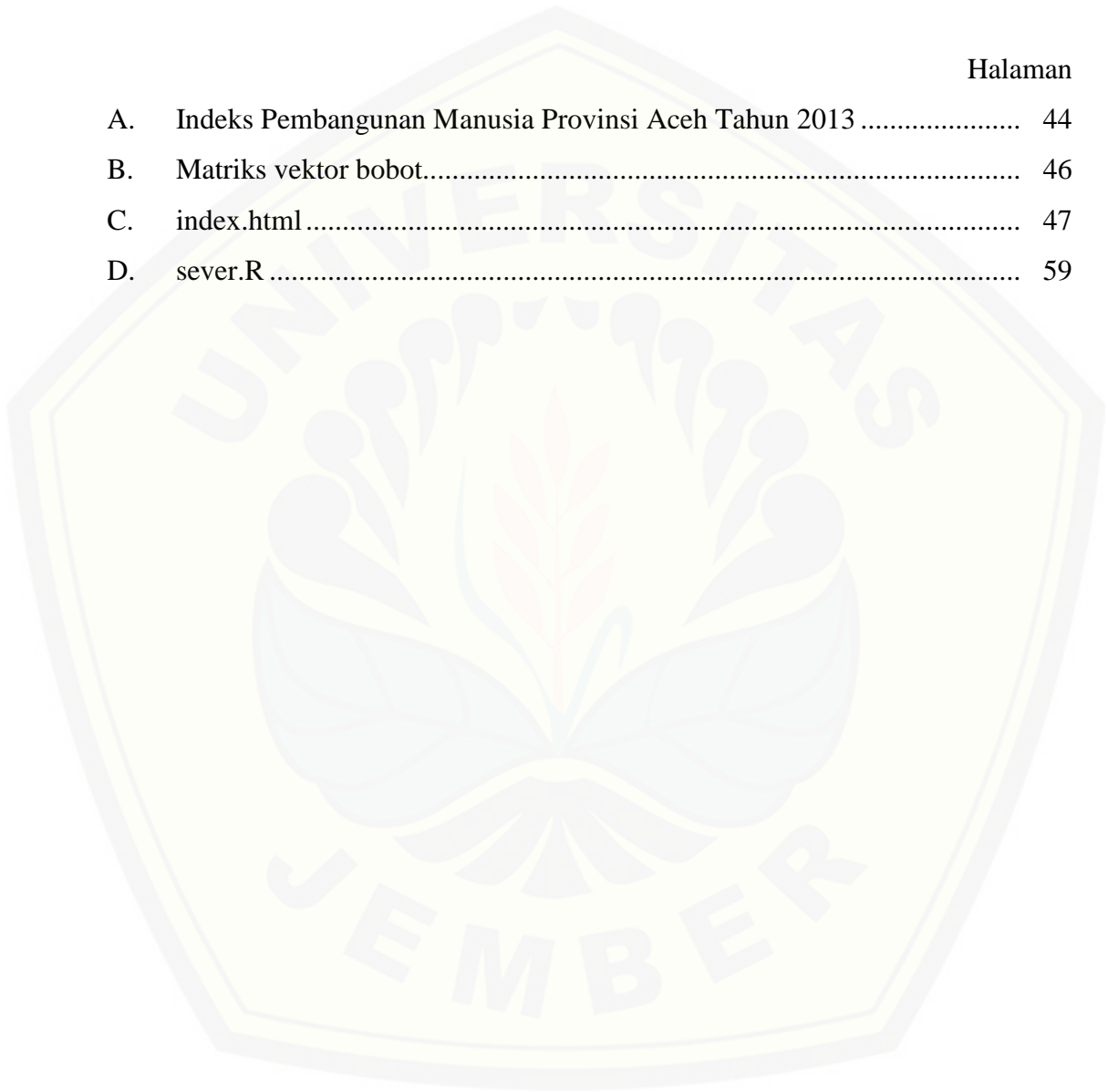
DAFTAR GAMBAR

	Halaman
2.1 Struktur sederhana neuron .....	4
2.2 Sel saraf tiruan .....	5
2.3 Ilustrasi jaringan <i>Self Organizing Map</i> (SOM) .....	7
2.4 Struktur komponen R-Shiny .....	15
3.1 Skema metode penelitian .....	18
4.1 Tampilan awal program .....	28
4.2 Menu pilihan dan tampilan data .....	28
4.3 Menu penentuan variabel numerik .....	29
4.4 Menu pilihan metode validasi .....	29
4.5 Menu tampilan nilai validasi .....	29
4.6 Menu Grid .....	30
4.7 Visualisasi <i>Self Organizing Map</i> (SOM) <i>clustering</i> .....	30
4.8 Pemilihan variabel numerik .....	31
4.9 Nilai validasi internal .....	32
4.10 Diagram fan vektor bobot .....	34
4.11 <i>Cluster Self Organizing Map</i> (SOM) .....	34
4.12 Jarak tetangga .....	36
4.13 Dendogram <i>cluster</i> .....	36
4.14 <i>Property</i> data Indeks Pembangunan Manusia provinsi Aceh tahun 2013 .	37



**DAFTAR LAMPIRAN**

	Halaman
A. Indeks Pembangunan Manusia Provinsi Aceh Tahun 2013 .....	44
B. Matriks vektor bobot.....	46
C. index.html.....	47
D. sever.R .....	59



## BAB 1. PENDAHULUAN

### 1.1 Latar Belakang

*Clustering* merupakan proses pengelompokan objek-objek ke dalam sebuah *cluster*, dengan tujuan objek data dengan kemiripan yang tinggi akan berada pada *cluster* yang sama dan objek data dengan kemiripan berbeda akan berada pada *cluster* berbeda. Menurut Abu-Jamous *et al* (2015) *Neural network-based clustering* merupakan metode *clustering* berdasarkan sejumlah *node* (unit) dimulai dengan satu *node* (neuron) dimana semua *node* sama kecuali untuk beberapa parameter dijalankan secara acak, hal ini membuat setiap *node* berperilaku sedikit berbeda. *Node* (unit) tersebut selanjutnya melakukan proses pembelajaran secara kompetitif. Salah satu metode *clustering* dalam *neural network-based clustering* ialah *Self Organizing Map* (SOM).

*Self Organizing Map* (SOM) merupakan teknik jaringan saraf tiruan dengan proses pembelajaran tak terawasi. *Self Organizing Map* (SOM) memungkinkan visualisasi dan proyeksi dari data berdimensi tinggi ( $n$ -variabel) ke dimensi rendah, biasa menjadi bidang 2-D dengan tetap mempertahankan topologi (bentuk data) tersebut (Kohonen *et al*, 1996). Data dengan letak berdekatan pada dimensi tinggi akan terletak pada lokasi berdekatan pula dalam ruang vektor 2-D. Hasil pemetaan metode tersebut mampu memvisualisasikan variabel-variabel dominan dalam *cluster*, sehingga lebih mudah dalam memahami data.

Penerapan metode *Self Organizing Map* (SOM) dalam sejumlah publikasi ilmiah masih terhitung sedikit dibandingkan metode *cluster* lainnya. Hal ini ditunjukkan dari 120 artikel ilmiah internasional dalam [www.google scholar.com](http://www.google scholar.com) dengan *input* kata *clustering* yang penulis lakukan, hanya 3,3% menggunakan metode *Self Organizing Map* (SOM). Publikasi tersebut meliputi: 34,17% metode *k-means*; 21% metode *fuzzy*; 10% metode hirarki; 2,5% metode partisi; 7,5% gabungan antara metode *k-means* dan metode lainnya; dan 20,83% *clustering* menggunakan metode-metode lainnya.

Penelitian *Self Organizing Map* (SOM) telah dimanfaatkan sebelumnya, diantaranya di bidang lingkungan untuk menentukan kelompok pencemaran udara sektor industri di Jawa Tengah (Warsito *et al*, 2008), Anis dan Rizal (2014) memvisualisasikan informasi sebaran Daftar Pemilih Tetap (DPT) pada Pemilihan Umum (Pemilu) kota Semarang tahun 2013 dan (Budhi *et al*, 2008) memprediksi talenta pemain basket dan Santosa (2013) merancang sistem pendukung keputusan pemetaan profil kecamatan di kabupaten Sragen berdasarkan Indeks Pembangunan Manusia menggunakan algoritma *Self Organizing Maps* (SOM). Menurut Abu-Jamous *et al* (2015) terdapat tiga program *open source* untuk *neural network based-clustering* yaitu R, Matlab dan weka. Salah satu program yang sering digunakan ialah R, namun pengerjaan di R menggunakan pendekatan skrip. Hal ini menjadi kendala bagi peneliti yang bukan berlatar belakang matematika maupun statistika. Menurut Tirta (2015) seiring dengan berjalannya waktu tim R-Studio meluncurkan dua program pendukung R, salah satunya R-Shiny. R-Shiny merupakan *toolkit* untuk membuat interaksi atau komunikasi antara dokumen HTML dengan program R yang memungkinkan orang dapat membuat laman web interaktif, sehingga kemampuan R bersifat pendekatan skrip dapat diakses melalui menu web secara GUI (*Graphical User Interface*).

Berdasarkan latar belakang tersebut penulis tertarik mengembangkan *Self Organizing Map* (SOM) berbasis web interaktif menggunakan R-Shiny. Modul atau dokumen tutorial pembelajaran tersebut untuk memudahkan *user* dalam melakukan *clustering* dengan *Self Organizing Map* (SOM), yaitu hanya dengan *input* data pada program tersebut. Ilustrasi data penelitian ini menggunakan data sekunder, yaitu data Indeks Pembangunan Manusia (IPM) Provinsi Aceh tahun 2013 dengan variabel penelitian berupa nama daerah (*Nama.Daerah*), angka harapan hidup (*AHH*), angka melek huruf (*AMH*), rata-rata lama sekolah (*RLS*), dan pengeluaran perkapita (*PPK*) sebagai vektor *input*. Pemilihan data berdasarkan kelengkapan data pada situs Badan Pusat Statistik (BPS) diantara Provinsi lainnya.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka diperoleh beberapa permasalahan berikut:

- a. bagaimana membuat web interaktif analisis *cluster Self Organizing Map* (SOM) bersifat tutorial?
- b. bagaimana hasil *cluster* dengan *Self Organizing Map* (SOM) terhadap data Indeks Pembangunan Manusia (IPM) Provinsi Aceh tahun 2013 menggunakan web interaktif?

## 1.3 Tujuan

Tujuan yang ingin dicapai dari penulisan tugas akhir ini adalah sebagai berikut:

- a. menghasilkan web interaktif analisis *cluster Self Organizing Map* (SOM) bersifat tutorial;
- b. mendapatkan hasil *cluster* dengan *Self Organizing Map* (SOM) terhadap data Indeks Pembangunan Manusia (IPM) Provinsi Aceh tahun 2013 menggunakan web interaktif.

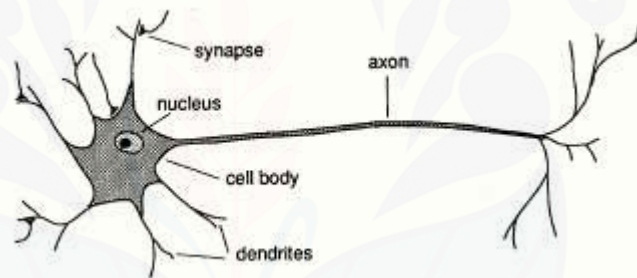
## 1.4 Manfaat

Manfaat yang diharapkan dari penyusunan tugas akhir ini adalah orang lain dapat dengan mudah dalam mengelompokkan data menggunakan analisis *cluster Self Organizing Map* (SOM).

## BAB 2. TINJAUAN PUSTAKA

### 2.1 Jaringan Saraf Tiruan (JST)

Jaringan Saraf Tiruan merupakan salah satu representasi buatan dari otak manusia untuk mensimulasikan proses pembelajaran pada otak manusia. Istilah buatan didasari oleh penerapannya menggunakan program komputer dalam menyelesaikan sejumlah proses perhitungan selama proses pembelajaran (Kusumadewi, 2003). Bentuk sederhana dari neuron dapat dilihat pada Gambar 2.1.



Gambar 2.1 Struktur sederhana neuron

Pada saraf biologis, setiap inti sel saraf(neuron) yaitu nukleus bertugas melakukan pemrosesan informasi. Berikut komponen utama dari jaringan saraf biologi:

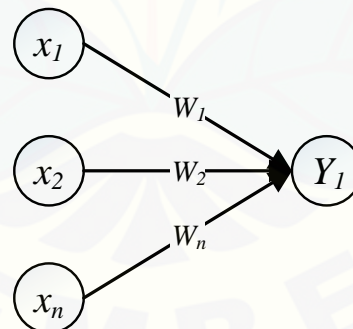
- a. dendrit, bertugas menerima informasi;
- b. badan sel (soma), tempat menerima informasi dari dendrit untuk diteruskan pada akson;
- c. sinapsis, pengirim informasi dari akson pada sel saraf lai;
- d. akson, mengirim impuls pada sel saraf lain.

Dari beberapa jaringan saraf tiruan, hampir semua memiliki kesamaan pada setiap komponen. Seperti halnya pada otak manusia, jaringan saraf tiruan juga terdiri dari beberapa neuron, masing-masing saling berhubungan. Pada jaringan saraf tiruan hubungan ini disebut bobot. Neuron-neuron jaringan saraf tiruan bekerja

dengan cara yang sama dengan jaringan saraf biologi. Informasi (*input*) dikirim pada neuron dengan bobot tertentu. Informasi tersebut diproses oleh suatu fungsi perambatan yang akan dijumlahkan dengan nilai-nilai semua bobot yang datang. Hasil penjumlahan tersebut dibandingkan dengan fungsi aktivasi setiap neuron. Neuron akan diaktifkan apabila *input* tersebut melewati suatu nilai ambang tertentu, begitupula sebaliknya. Neuron yang diaktifkan akan mengirim *output* melalui bobot-bobot *output*nya ke semua neuron yang terhubung dengannya, demikian seterusnya. Secara berurutan perbandingan jaringan saraf biologi dengan jaringan saraf tiruan dan ilustrasi jaringan saraf tiruan dapat dilihat masing-masing pada Tabel 2.1 dan Gambar 2.2.

Tabel 2.1 Perbandingan jaringan saraf biologi dan jaringan saraf tiruan

Jaringan saraf biologi	Jaringan saraf tiruan
Dendrit	<i>Input</i>
Badan sel (soma)	<i>Node</i> atau <i>input</i>
Sinapsis	Bobot
Akson	<i>Output</i>



Gambar 2.2 Sel saraf tiruan

Dimana: neuron *input*:  $X_1, X_2, \dots, X_n$ ; bobot:  $W_1, W_2, \dots, W_n$ ; neuron *output*:  $Y_1$ .

Menurut Setiawan (2011) proses perubahan bobot dibagi menjadi 2, yaitu: pembelajaran terawasi (*supervised learning*) dan pembelajaran tidak terawasi (*unsupervised learning*). Pembelajaran terawasi merupakan suatu jaringan dengan *output* yang memiliki pola yang sama dengan pola *input*, sedangkan pembelajaran tak terawasi ialah suatu jaringan dengan *output* tidak ditentukan hasilnya seperti

apakah yang diharapkan selama proses pembelajaran. Selama proses pembelajaran tak terawasi, nilai bobot disusun dalam suatu range tertentu tergantung pada nilai *input* yang diberikan. Tujuan pembelajaran tak terawasi ialah mengelompokkan unit-unit (*node*) hampir sama dalam suatu area tertentu. Pembelajaran ini biasa digunakan dalam *clustering*.

## 2.2 Analisis Cluster

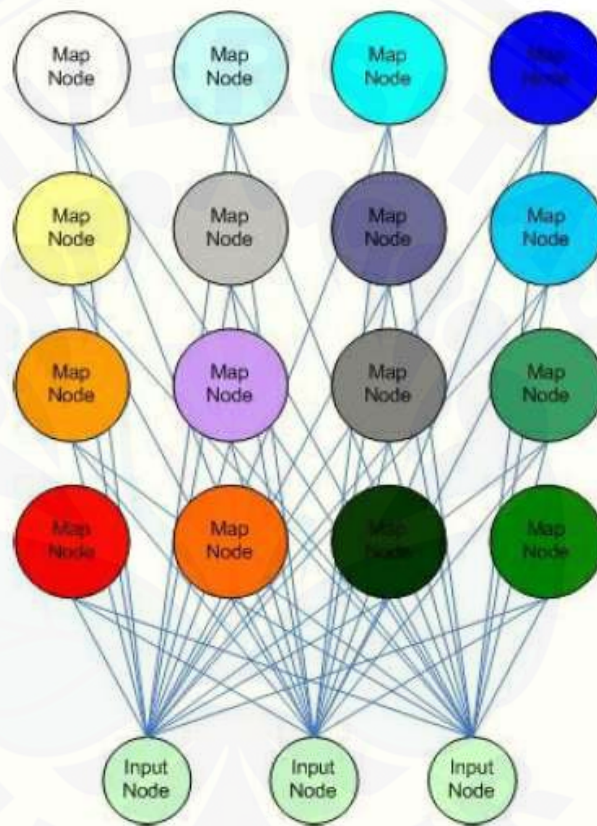
Analisis *cluster* adalah salah satu teknik multivariat yang bertujuan mengelompokkan sejumlah objek kedalam beberapa kelompok berdasarkan informasi/ karakteristik data. Pengelompokan sejumlah objek dengan tujuan dalam suatu kelompok atau *cluster* memiliki tingkat kesamaan yang tinggi, sedangkan anggota antar kelompok memiliki tingkat keragaman yang tinggi. Analisis *cluster* biasa digunakan dalam segmentasi pasar konsumen, memahami perilaku pembeli, mengidentifikasi peluang produk baru, meringkas atau mereduksi data, dan lain-lain. Terdapat beberapa tipe *clustering*, diantaranya hirarki dan non-hirarki untuk mengelompokkan setiap objek tepat dalam suatu kelompok. Pengelompokan objek tersebut tidak mampu menunjukkan variabel dominan dalam *cluster*, tidak demikian dengan *Neural Network-based Clustering* dalam hal ini *Self Organizing Map* (SOM). Metode tersebut mengelompokkan suatu objek dengan menganggap setiap titik sebagai neuron (Abu-Jamaous *et al*, 2015).

## 2.3 Self Organizing Map (SOM)

### 2.3.1 Pengertian Self Organizing Map (SOM)

*Self Organizing Map* (SOM) merupakan suatu metode Jaringan Saraf Tiruan yang diperkenalkan pertama kali oleh Teuvo Kohonen tahun 1981, sehingga sering disebut dengan Jaringan Kohonen. Dinamakan “*Self Organizing*” karena tidak memerlukan pengawasan/ tak terawasi (*unsupervised learning*) dan disebut “*Map*” karena *Self Organizing Map* (SOM) berusaha memetakan bobotnya agar sesuai dengan *input* data yang diberikan. Neuron-neuron pada jaringan ini menyusun dirinya sendiri berdasarkan nilai *input* tertentu dalam suatu kelompok,

biasa disebut *cluster*. Selama proses penyusunan diri, *cluster* dengan vektor bobot paling cocok dengan pola bobot (jarak paling dekat) akan terpilih sebagai pemenang. Neuron pemenang beserta neuron-neuron tetangga akan memperbaiki bobotnya masing-masing (Kusumadewi, 2003). Menurut Guthikonda dalam Megawati (2015) ilustrasi jaringan *Self Organizing Map* (SOM) dapat dilihat pada Gambar 2.3.



Gambar 2.3 Ilustrasi jaringan *Self Organizing Map* (SOM)

Dari Gambar 2.3 dapat dilihat bahwa setiap *map node* terhubung dengan setiap *input node* sedangkan antar *map node* tidak terhubung satu sama lain, sehingga pada jaringan node ukuran 4x4 terdapat  $4 \times 4 \times 3 = 48$  koneksi. Pengaturan *node* sebagai bidang 2-D untuk memudahkan visualisasi hasil dengan koordinat  $(i, j)$  yang unik. Hal ini memudahkan untuk melihat *node* secara jelas dan menghitung jarak antar *node*. Nilai *map node* tidak terpengaruh satu sama lain



karena koneksi hanya terjadi dengan *input node*, sehingga memungkinkan sebuah *map node* hanya akan memperbarui bobotnya berdasarkan vektor input saja.

Menurut Larose dalam Riyandwyana *et al* (2012) *Self Organizing Map* (SOM) memperlihatkan tiga karakteristik:

- a. kompetitif, setiap vektor bobot berlomba menjadi simpul pemenang;
- b. kooperasi, setiap simpul pemenang bekerjasama dengan simpul tetangga, dan
- c. adaptasi, perubahan simpul pemenang dengan simpul tetangga.

### 2.3.2 Algoritma *Self Organizing Map* (SOM)

Proses *Self Organizing Map* (SOM) dilakukan dengan mengikuti algoritma berikut (Warsito *et al*, 2008):

a. ditetapkan:

1. jumlah variabel =  $m$
2. jumlah data/ objek =  $n$
3. jumlah *cluster* =  $k$

b. inisialisasi:

1. bobot *input* ( $w_{ij}$ ):

$$w_{ij} = \frac{MinP_i + MaxP_i}{2}$$

keterangan:  $w_{ij}$  = bobot antara neuron kelas- $i$  dengan variabel *input* ke- $j$

$MinP_i$  = nilai minimum variabel *input* ke- $i$

$MaxP_i$  = nilai maksimum variabel *input* ke- $i$

2. bobot bias

bobot bias adalah bobot koreksi yang digunakan sebagai tolak ukur penentuan variabel lain yang belum diketahui sebagai variabel *input*. Variabel *input* baru terjadi bila nilai bobot bias lebih besar dari nilai bobot variabel *input*.

nilai bobot bias diperoleh dari:

$$b_i = e^{[1 - \ln(1/k)]}$$

keterangan:  $b_i$  = bobot bias neuron ke- $i$

$k$  = jumlah *cluster*

3. set parameter *learning rate* ( $\alpha$ )

*learning rate* adalah faktor pengali pada perubahan bobot dengan interval  $0 - 1$ . Semakin besar nilai *learning rate* maka proses pembelajaran semakin cepat, namun dengan hasil pemetaan yang kasar. Sebaliknya, bila *learning rate* bernilai kecil maka diperoleh pemetaan yang cukup bagus dengan waktu pembelajaran yang lebih lama. Faktor pengali alpha akan selalu berkurang bila tidak ada perubahan error.

4. set maksimum epoch (MaxEpoch)

penentuan pembelajaran dengan melakukan analisis jumlah iterasi untuk mencapai maksimum *epoch* yang terbaik, dengan sebuah indikasi bahwa maksimum tercapai apabila pada iterasi berikutnya sudah tidak terjadi perubahan hasil klusterisasi.

- c. set Epoch = 0

- d. jika Epoch < MaxEpoch, maka:

1. Epoch = Epoch + 1
2. Data dipilih secara acak, misal data ke- $z$
3. Dicari jarak antara data ke- $z$  dengan tiap bobot *input* ke- $i$  ( $D_i$ )

$$D_i = \sum_{j=1}^m \sqrt{(w_{ij} - P_{zj})^2}$$

keterangan:  $D_i$  = jarak data ke- $z$  dengan setiap bobot masukan ke- $i$

$w_{ij}$  = bobot antara neuron ke- $j$  pada lapisan masukan dan neuron ke- $i$  pada lapisan keluaran

$P_{zj}$  = nilai masukan data ke- $z$  terhadap neuron ke- $j$  pada lapisan masukan

dihitung penjumlahan negatif jarak ( $a_i$ ) menggunakan:

$$a_i = -D_i + b_i$$

mencari  $a_i$  terbesar:

- i.  $MaxA = \max(a_i)$ , dengan  $i = 1, 2, \dots, k$
  - ii.  $Idx = 1, \exists a_i = MaxA$
4. set *output* neuron ke- $i$  ( $y_i$ ):
- $y(i) = 1$ ; jika  $i = Idx$

$$y(ii) = 0; \text{ jika } i \neq Idx$$

5. update bobot yang menuju neuron  $Idx$ :

$$w_{ij}^{baru} = w_{ij}^{lama} + \alpha (P_{zj} - w_{ij}^{lama})$$

$$w(Idx,j) = w(Idx,j) + \alpha (p(z,j) - w(Idx,j))$$

6. update bobot bias:

$$c(i) = (1 - \alpha) e^{(1-\ln(b(i)))} + \alpha \cdot y(i)$$

$$b(i) = e^{(1-\ln(c(i)))}$$

## 2.4 Validasi Cluster

Setiap *cluster* yang terbentuk memiliki seperangkat ukuran karakteristik, diantaranya berupa nilai indeks validitas *cluster* (Brock *et al*, 2008). Hal ini digunakan sebagai acuan dalam menentukan jumlah *cluster* optimal. Pada penelitian ini penulis menggunakan 3 kriteria alternatif: validasi internal, validasi stabilitas dan validasi biologis.

### 2.4.1 Validasi Internal

Validasi internal menggunakan informasi internal pada data untuk menilai kualitas *clustering*. Validasi internal mencerminkan kepadatan, hubungan dan pemisahan *cluster* partisi. Kepadatan berhubungan dengan mengevaluasi homogenitas *cluster*, biasa dilihat dari varian intra-*cluster*. Hubungan menunjukkan penempatan beberapa data pengamatan dalam sebuah *cluster*, dimana data tersebut sebagai tetangga terdekat. Nilai kepadatan tersebut diukur dengan konektivitas. Pemisahan *cluster* partisi menunjukkan tingkat pengukuran jarak antar dua *cluster* (hal ini biasa menggunakan jarak antar *centroid*). Validasi internal meliputi:

- a. konektivitas

Konektivitas menunjukkan tingkat hubungan *cluster*, ditentukan dengan jumlah tetangga terdekat. Nilai konektivitas dihitung menggunakan persamaan (2.1).

$$K = \sum_{i=1}^N \sum_{j=1}^L x_{i,nn_{i(j)}} \quad (2.1)$$

dimana:  $K$ : konektivitas;  $N$ : jumlah data;  $L$ : jumlah tetangga;  $nn_{i(j)}$ : ukuran jarak tetangga terdekat data ke- $i$ , dengan:

$$nn_{i(j)} : \begin{cases} 0, & i \text{ dan } j \text{ satu cluster} \\ 1, & \text{jlainnya} \end{cases}$$

b. nilai rata-rata *Silhouette*

Nilai rata-rata *Silhouette* merupakan rata-rata nilai *Silhouette* setiap data. Nilai *Silhouette* mengukur tingkat kepercayaan pada proses *clustering* dari setiap data observasi. Nilai *Silhouette* dihitung menggunakan persamaan (2.2).

$$S(i) = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (2.2)$$

keterangan:

$S(i)$  : nilai rata-rata *Silhouette*;

$b_i$  : jarak rata-rata antara data ke- $i$  dan objek data pada *cluster* berbeda (jarak terdekat);

$a_i$  : jarak rata-rata antara data ke- $i$  dengan seluruh objek data lain dalam *cluster* yang sama.

c. indeks *Dunn*

Indeks *Dunn* merupakan rasio dari jarak terdekat antara objek data di *cluster* berbeda terhadap jarak terjauh pada *intra-cluster*. Nilai indeks *Dunn* dapat dihitung menggunakan persamaan (2.3).

$$D(\mathcal{C}) = \frac{\min_{C_k, C_1 \in \mathcal{C}, C_k \neq C_1} \left( \min_{i \in C_k, j \in C_1} \text{dist}(i, j) \right)}{\max_{C_m \in \mathcal{C}} \text{diam}(C_m)} \quad (2.3)$$

dimana:  $D(\mathcal{C})$ : *dunn Index*;  $\text{diam}(C_m)$ : jarak maksimum antar data dalam *cluster* ke- $m$

#### 2.4.2 Validasi Stabilitas

Validasi stabilitas mengevaluasi kekonsistenan hasil *clustering* dengan membandingkannya dengan perolehan *cluster* dengan menghapus satu kolom pada suatu waktu. Validasi stabilitas meliputi:

a. *average proportion of non-overlap* (APN)

*average proportion of non-overlap* (APN) merupakan proporsi rata-rata data pada *cluster* berbeda dari dataset dan *clustering* berdasarkan dataset dengan

menghilangkan satu kolom (variabel). Ukuran *average proportion of non-overlap* (APN) didefinisikan pada persamaan (2.4).

$$APN(K) = \frac{1}{M N} \sum_{i=1}^N \sum_{\ell=1}^M \left( 1 - \frac{n(C^{i,\ell} \cap C^{i,0})}{n(C^{i,0})} \right) \quad (2.4)$$

keterangan:

$M$  : jumlah variabel

$N$  : jumlah data

$C^{i,\ell}$  : *cluster* dari data ke- $i$  dengan menghapus kolom ke- $\ell$

$C^{i,0}$  : *cluster* dari data ke- $i$  tanpa menghapus kolom

b. *average distance* (AD)

*average distance* (AD) menghitung jarak rata-rata antara objek data dalam *cluster* sama berdasarkan *clustering* dataset dan *clustering* dataset dengan penghapusan satu kolom (variabel). Ukuran *average distance* (AD) didefinisikan pada persamaan (2.5).

$$AD(K) = \frac{1}{M N} \sum_{i=1}^N \sum_{\ell=1}^M \frac{1}{n(C^{i,0}) n(C^{i,\ell})} \left[ \sum_{i \in C^{i,0}, j \in C^{i,\ell}} dist(i, j) \right] \quad (2.5)$$

c. *average distance between means* (ADM)

*average distance between means* (ADM) mengukur jarak rata-rata antara pusat *cluster* yang terletak pada *cluster* sama berdasarkan *clustering* dataset dan *clustering* dataset dengan penghapusan satu kolom (variabel). Ukuran *average distance* (AD) didefinisikan pada persamaan (2.6).

$$ADM(K) = \frac{1}{M N} \sum_{i=1}^N \sum_{\ell=1}^M dist(\bar{x}_{C^{i,\ell}}, \bar{x}_{C^{i,0}}) \quad (2.6)$$

keterangan:

$\bar{x}_{C^{i,0}}$  : rata-rata dataset pada *cluster* yang memuat data ke- $i$  tanpa pengurangan variabel

$\bar{x}_{C^{i,\ell}}$  : rata-rata dataset pada *cluster* yang memuat data ke- $i$  dengan pengurangan variabel ke- $\ell$

d. *figure of merit* (FOM)

*figure of merit* (FOM) menunjukkan rata-rata varian intra-*cluster* dataset dengan menghapus satu kolom (variabel). Ukuran *figure of merit* (FOM) didefinisikan pada persamaan (2.7).

$$\text{FOM}(\ell, K) = \sqrt{\frac{1}{N} \sum_{k=1}^K \sum_{C_k(\ell)} \text{dist}(x_{i,\ell}, \bar{x}_{C_k(\ell)})} \quad (2.7)$$

Dimana:  $K$  : jumlah *cluster*;  $x_{i,\ell}$  : nilai data ke- $i$  pada kolom ke- $\ell$  pada *cluster*  $C_k(\ell)$ ;  $\bar{x}_{C_k(\ell)}$  : rata-rata *cluster*  $C_k(\ell)$

### 2.4.3 Validasi Biologis

Validasi biologis mengevaluasi kemampuan algoritma *clustering* untuk menghasilkan *cluster* bermakna secara biologi. Validasi biologi meliputi:

#### a. *biological homogeneity index* (BHI)

*biological homogeneity index* (BHI) mengukur homogenitas *cluster*. Nilai *biological homogeneity index* (BHI) didefinisikan pada persamaan (2.8).

$$\text{BHI}(\mathcal{C}, \mathcal{B}) = \frac{1}{K} \sum_{k=1}^K \frac{1}{n_k(n_k-1)} \sum_{i \neq j \in C_k} I(B(i) = B(j)) \quad (2.8)$$

Dimana:  $n_k$  : notasi gen pada *cluster*  $C_k$ ;

#### b. *biological stability index* (BSI)

*biological stability index* (BSI) memeriksa konsistensi pengelompokan gen dengan fungsi biologis yang sama. Nilai *biological stability index* (BSI) didefinisikan pada persamaan (2.9).

$$\text{BSI}(\mathcal{C}, \mathcal{B}) = \frac{1}{F} \sum_{k=1}^F \frac{1}{n(B_k)(n(B_k)-1)} \sum_{\ell=1}^M \sum_{i \neq j \in B_k} \frac{n(C^{i,0} \cap C^{j,\ell})}{n(C^{i,0})} \quad (2.9)$$

Nilai validasi setiap metode diatas dapat dilihat pada Tabel 2.2.

Tabel 2.2 Nilai validasi

Validasi	Komponen	Nilai	Kriteria
Internal	konektivitas	$[0, +\infty]$	*
	nilai rata-rata Silohette	$[-1, +1]$	**
	indeks Dunn	$[0, +\infty]$	**
Stabilitas	<i>average proportion non-overlap</i> (APN)	$[0, 1]$	*
	<i>average distance</i> (AD)	$[0, +\infty]$	*
	<i>average distance between means</i> (ADM)	$[0, +\infty]$	*
	<i>figure of merit</i> (FOM)	$[0, +\infty]$	*
Biologis	<i>biological homogeneity index</i> (BHI)	$[0, 1]$	**
	<i>biological stability index</i> (BSI)	$[0, 1]$	**

Keterangan: \* : minimum    \*\* : maksimum

## 2.5 Indeks Pembangunan Manusia (IPM)

Pembangunan manusia merupakan paradigma pembangunan yang menempatkan manusia sebagai fokus dan sasaran akhir dari seluruh kegiatan pembangunan. Keberhasilan pembangunan manusia dapat dinilai dari seberapa besar permasalahan mendasar dalam masyarakat dapat teratasi. Permasalahan-permasalahan tersebut antara lain: kemiskinan, pengangguran, buta huruf, ketahanan pangan dan penegakan demokrasi. Faktanya tingkat keberhasilan aspek tersebut bervariasi, oleh karena itu tahun 1990 Perserikatan Bangsa-Bangsa (PBB) menetapkan suatu ukuran standar pembangunan manusia yaitu Indeks Pembangunan Manusia (IPM). Ukuran capaian berbasis sejumlah komponen dasar kualitas hidup, meliputi: kesehatan, pengetahuan dan kehidupan layak. Komponen kesehatan diukur menggunakan angka harapan hidup, komponen pengetahuan digunakan gabungan indikator angka melek huruf dan rata-rata lama sekolah, selanjutnya untuk komponen hidup layak digunakan indikator kemampuan daya beli (pengeluaran perkapita) (Marhaeni *et al*, 2008). Indeks Pembangunan Manusia (IPM) suatu daerah digolongkan menjadi 4 kategori: sangat tinggi, tinggi, menengah dan rendah.

## 2.6 R-Shiny

R-Shiny merupakan salah satu versi R berbasis web, Tirta (2014) menyatakan bahwa komponen program Shiny dibedakan menjadi dua kelompok yaitu:

1. *user interface*. Bagian ini bermanfaat untuk:
  - a. panel kontrol, ialah panel untuk mengontrol *input* berupa data, variabel, model, tergantung kompleksitas modul. Tampilan kontrol dapat berupa *slider*, *radio button*, *check-box*, dan lain-lain
  - b. pemasukan permintaan nilai *input* (data dengan berbagai jenis variabel yang diperlukan, pemilihan model, jenis dan kriteria uji statistika)
  - c. penyajian *output* terkait hasil analisis/ uji. Hasil *output* dapat berupa:
    - i. grafik (histogram, diagram pencar, dan lain-lain)

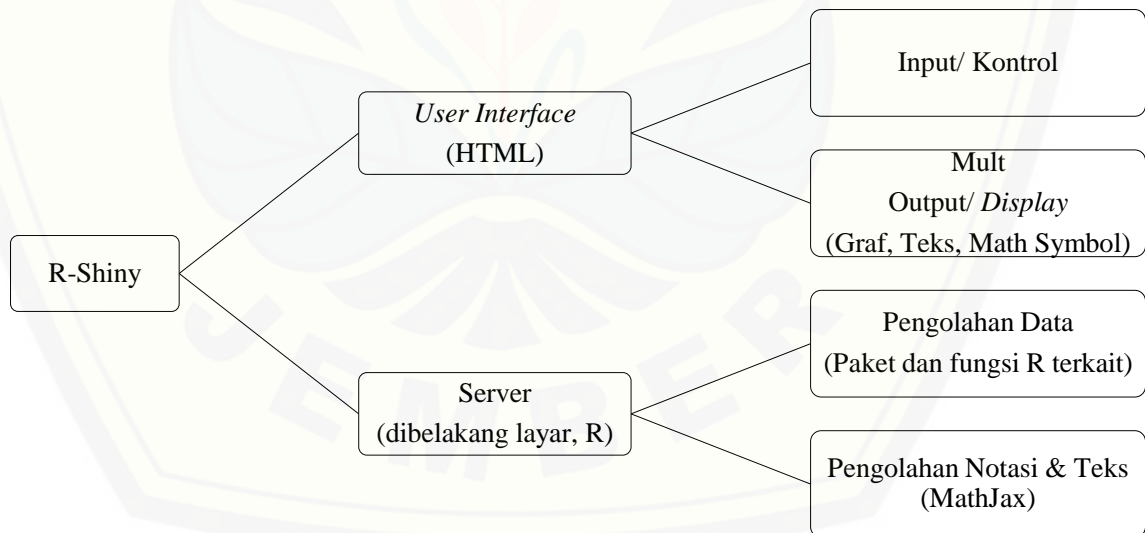
- ii. bentuk angka/ teks dapat berbentuk asli (verbatim) maupun berbentuk tabel
- iii. teks khusus dengan notasi matematika dengan format LaTeX

R-shiny menyediakan berbagai format *output* seperti *plotOutput*, *textOutput*, *verbatimTextOutput*, *tabelOutput*, dan lain-lain untuk mengakomodasi berbagai luaran diatas.

Bagian *user interface* dapat disajikan pada file khusus *ui.r*, dapat juga disajikan penuh melalui file HTML: *index.html*.

2. *server*. Bagian ini merupakan otak dari program yang bertugas melakukan simulasi, berbagai analisis data sesuai pilihan penggunaan dan selanjutnya mengirim hasilnya ke bagian *output*. *Server* didukung oleh berbagai prosedur dan analisis data yang pada umumnya telah tersedia pada berbagai paket R. Bagian ini disimpan dalam file *server.r*.

Struktur umum komponen pemrograman dengan R-shiny dapat dilihat pada Gambar 2.4.



Gambar 2.4 Struktur komponen R-Shiny

Secara umum, penyajian file *ui* dan file *server* ada dua macam yaitu:

1. membuat file *server.r* dan *ui.r* secara terpisah lalu diletakkan pada satu direktori dengan nama sesuai topik. Selanjutnya nama direktori dipanggil.

Direktori: *shiny1*



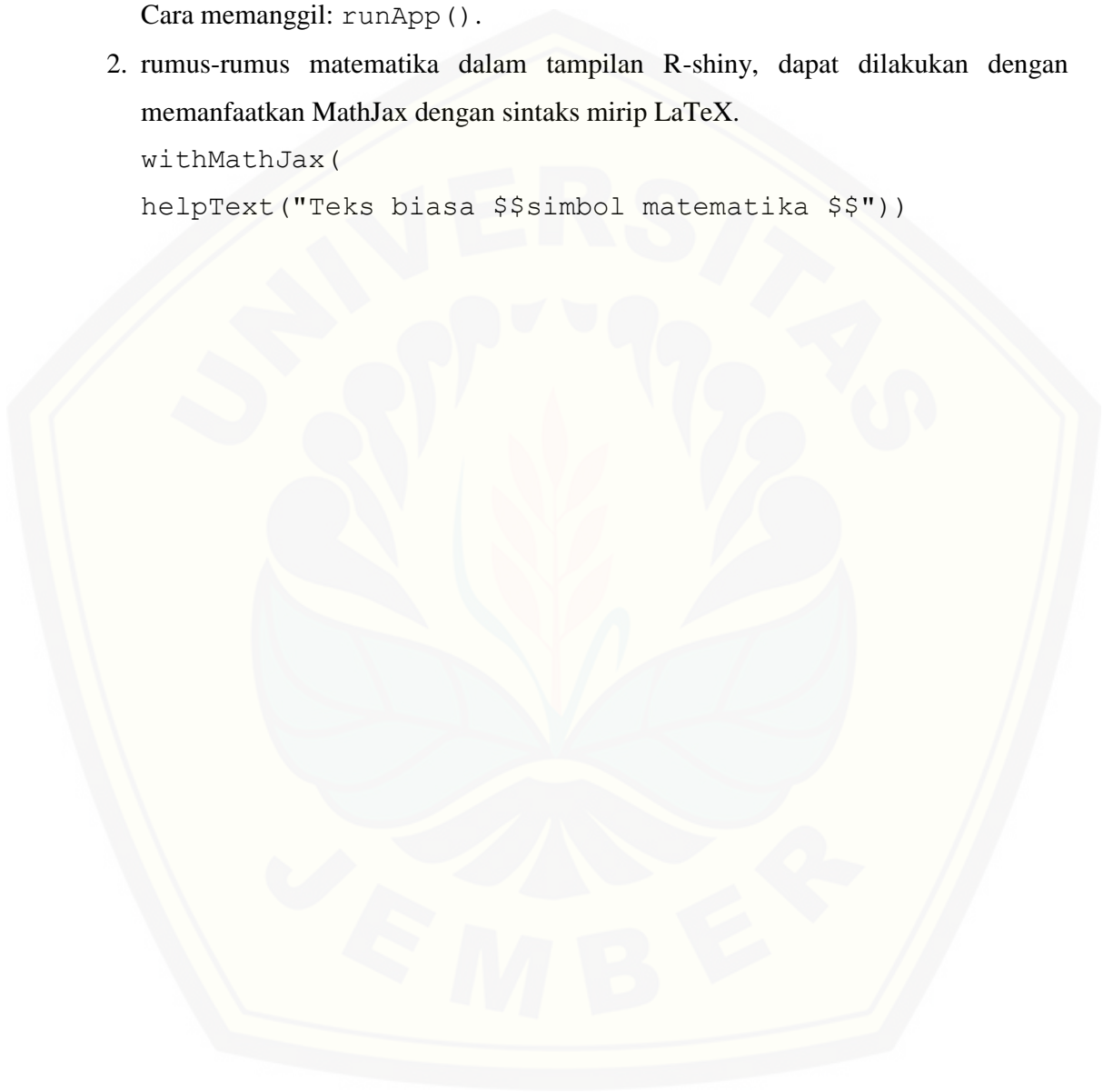
File:

```
-server.r  
/subdir www/  
-index.html
```

Cara memanggil: `runApp()`.

2. rumus-rumus matematika dalam tampilan R-shiny, dapat dilakukan dengan memanfaatkan MathJax dengan sintaks mirip LaTeX.

```
withMathJax(  
  helpText("Teks biasa $$simbol matematika $$")
```



### **BAB 3. METODE PENELITIAN**

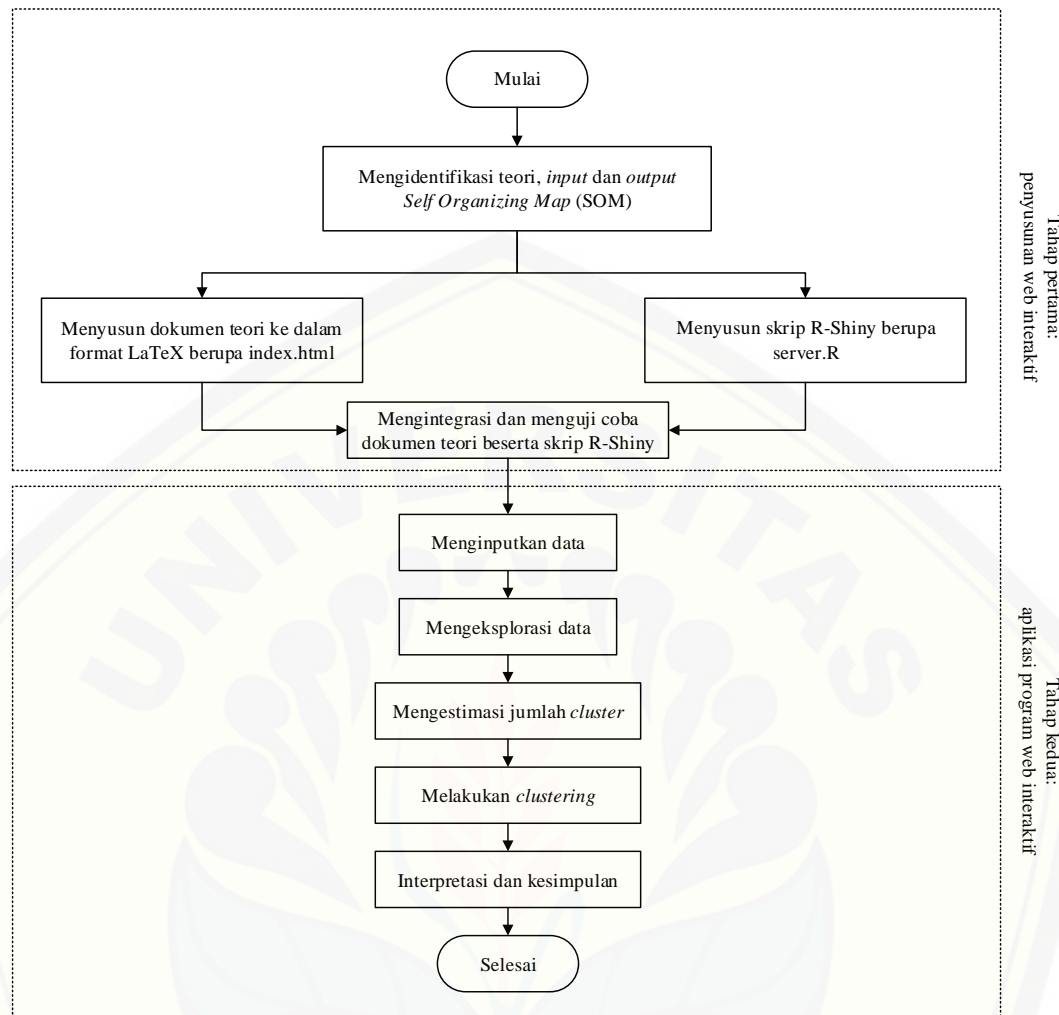
Pada bab ini, akan dijelaskan mengenai metode penelitian sebagai konsep dasar. Metode penelitian tugas akhir ini meliputi data dan langkah-langkah penelitian.

#### **3.1 Data**

Data penelitian ini merupakan data sekunder Indeks Pembangunan Manusia (IPM) provinsi Aceh tahun 2013. Sumber data berasal dari Badan Pusat Statistik (BPS) yang diperoleh melalui <http://aceh.bps.go.id/>. Variabel *input* dalam penelitian ini meliputi: nama daerah (*Nama.Daerah*), angka harapan hidup (*AHH*), angka melek huruf (*AMH*), rata-rata lama sekolah (*RLS*), dan pengeluaran perkapita (*PPK*). Adapun data Indeks Pembangunan Manusia (IPM) Provinsi Aceh tahun 2013 disajikan pada Lampiran A.

#### **3.2 Langkah-Langkah Penelitian**

Langkah-langkah penelitian yang dilakukan dalam tugas akhir ini dibagi menjadi dua bagian, yaitu: tahap penyusunan web interaktif dan tahap aplikasi program web interaktif. Secara skematik langkah-langkah penelitian ditunjukkan dalam Gambar 3.1.



Gambar 3.1 Skema metode penelitian

### 3.2.1 Tahap penyusunan web interaktif

Tahap penyusunan web interaktif terdiri atas empat langkah. Berikut langkah-langkah penyusunan web interaktif:

- mengidentifikasi teori, bagian ini berisi baris menu *input* dan *output*. *Input* berupa data untuk *dicluster*. *Output* yang akan dihasilkan berupa *summary* untuk eksplorasi data, nilai validasi, estimasi jumlah *cluster*, serta *plotting* untuk visualisasi Kohonen dan anggota *cluster* untuk penentuan anggota setiap *cluster*;
- menyusun dokumen teori dalam format LaTeX ke HTML sehingga dapat ditampilkan secara *online*. dokumen tersebut berupa script HTML: `index.html`;

- c. menyusun skrip R-Shiny berupa server.R. pada bagian ini dilakukan pengaplikasian program analisis *cluster Self Organizing Map* (SOM) pada R ke dalam program R-Shiny. Penelitian ini menggunakan *software R i386 3.1.2* dengan paket kohonen dan paket `clValid`.

Berikut struktur fungsi *Self Organizing Map*(SOM) dalam paket kohonen.

```
som(data, grid=somgrid(), rlen = ..., alpha =
c(0.05, 0.01), ...)
```

keterangan:

- `data` : sebuah matriks, dimana setiap baris merepresentasikan sebuah objek;
- `grid` : ukuran peta, berupa ukuran `xdim` dan `ydim`;
- `rlen` : waktu untuk data set menjadi suatu *network*;
- `alpha` : *learning rate*, secara otomatis bernilai 0.05 – 0.01;
- `...` : perintah lainnya.

Validasi *cluster* telah tersedia dalam paket `clValid`. Berikut fungsi dalam paket `clValid` yang digunakan dalam mengestimasi jumlah *cluster*.

```
clValid(obj, nClust, clMethods = "...", validation =
"...")
```

keterangan:

- `Obj` : data frame;
- `nClust` : range jumlah *cluster*;
- `clMethods` : metode *clustering*;
- `Validation` : ukuran validasi yang akan digunakan.

Visualisasi *plot* kohonen digunakan untuk memvisualisasi kualitas *Self Organizing Map* (SOM). Visualisasi tersebut berupa:

- i. jarak tetangga

Visualisasi ini merupakan jarak antara setiap *node* dan tetangganya, biasa dilihat dengan warna abu-abu. Daerah dengan jarak tetangga yang dekat diindikasikan dengan kelompok *node* sama atau mirip, sebaliknya daerah dengan jarak yang jauh diindikasikan dengan *node* berbeda. Jarak tetangga dapat

digunakan untuk mengidentifikasi *cluster* dalam peta *Self Organizing Map* (SOM). Jenis *plot* ini menggunakan *input* `dist.neighbours`.

ii. vektor bobot

Visualisasi ini menunjukkan vektor bobot, dengan menggunakan jenis *plot* `codes`.

iii. pemetaan

Visualisasi ini menunjukkan dimana objek yang dipetakan, *type plot* ini membutuhkan argumen `classif`, `label`, atau `pchs` dengan *input* jenis berupa `mapping`.

iv. *property*

Sifat masing-masing *node* dapat dihitung dan ditampilkan dalam kode warna. Hal ini dapat digunakan untuk memvisualisasi persamaan satu objek tertentu untuk semua *node* dalam peta, untuk menampilkan kesamaan rata-rata dari setiap *node* dan objek observasi. Jenis *plot* ini menggunakan *input* `property`.

Adapun fungsi dalam paket `kohonen` yang digunakan untuk memberikan visualisasi atau *plotting* hubungan data sebagai berikut:

```
plot(x, type = "...", main="...")
```

keterangan:

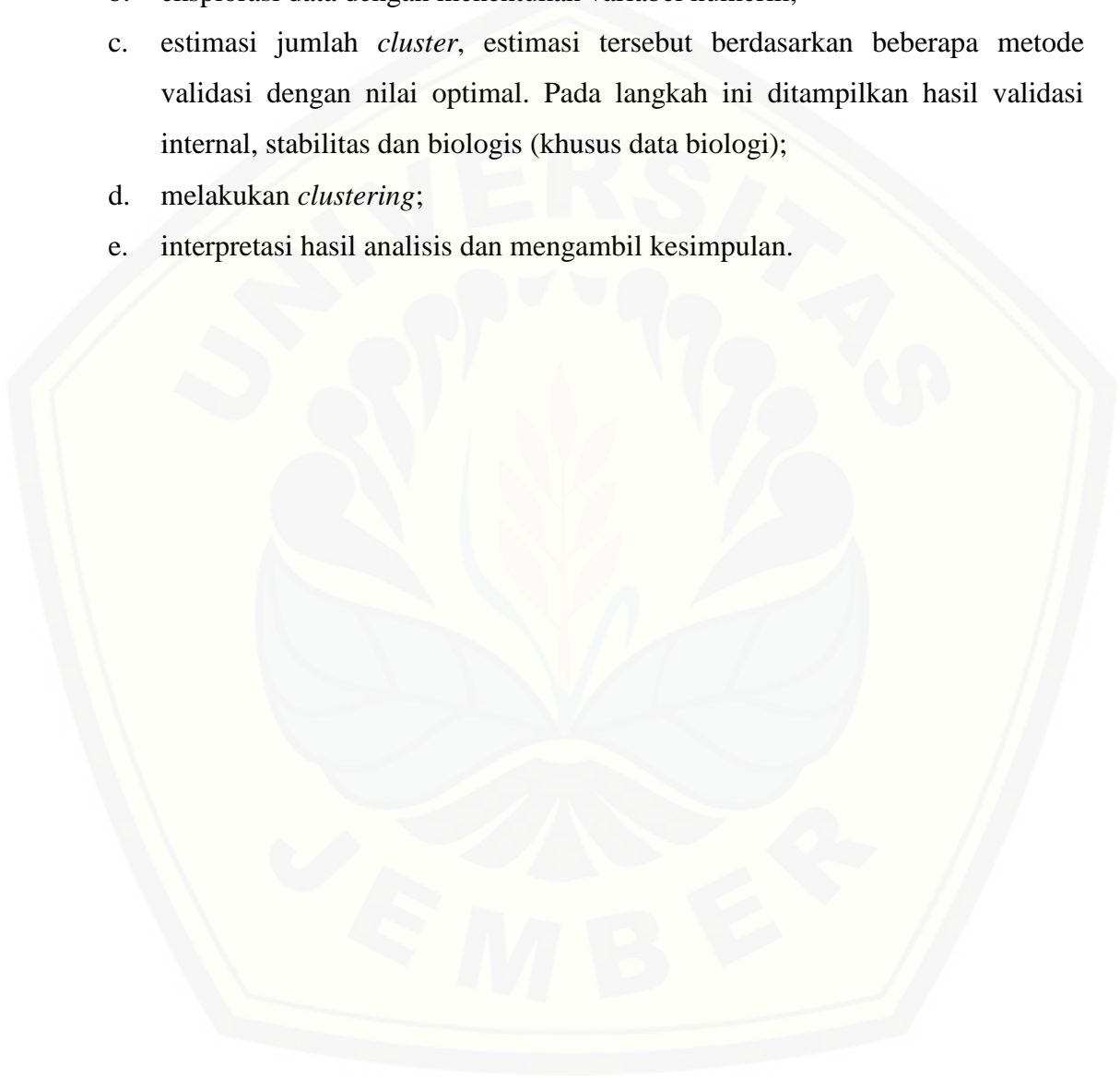
`som` : objek Kohonen;  
`type` : jenis *plot*;  
`main` : judul *plot*.

- d. Mengintegrasikan dan menguji coba dokumen teori (`index.html`) beserta skrip R-Shiny (`server.R`) yang ada dalam langkah a dan b untuk membuat program *online*.

### 3.2.2 Tahap aplikasi program web interaktif

Pada penelitian ini program web interaktif diaplikasikan pada data Indeks Pembangunan Manusia (IPM) Provinsi Aceh Tahun 2013. Ilustrasi data dapat dilihat pada Lampiran A. Berikut langkah-langkah analisis data *clustering Self Organizing Map* (SOM):

- a. *input* data pada program web interaktif secara *online*. Data penelitian ini menggunakan nama daerah (*Nama.Daerah*), angka harapan hidup (*AHH*), angka melek huruf (*AMH*), rata-rata lama sekolah (*RLS*), dan pengeluaran perkapita (*PKK*) sebagai variabel *input*;
- b. eksplorasi data dengan menentukan variabel numerik;
- c. estimasi jumlah *cluster*, estimasi tersebut berdasarkan beberapa metode validasi dengan nilai optimal. Pada langkah ini ditampilkan hasil validasi internal, stabilitas dan biologis (khusus data biologi);
- d. melakukan *clustering*;
- e. interpretasi hasil analisis dan mengambil kesimpulan.



## BAB 5. PENUTUP

### 5.1 Kesimpulan

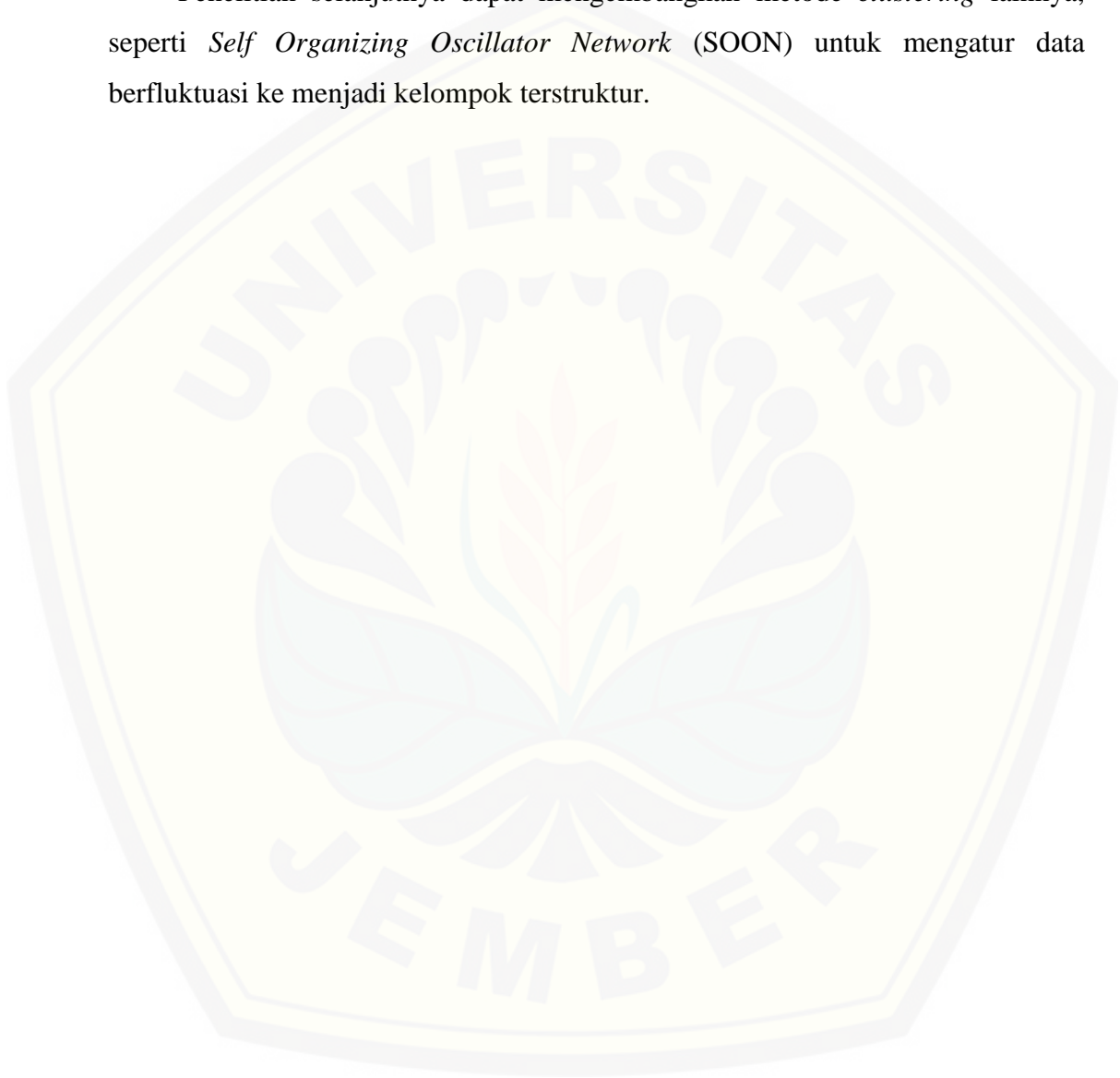
Berdasarkan hasil dan pembahasan pembuatan program *clustering Self Organizing Map* (SOM) pada bab 4 berbasis web interaktif menggunakan R-Shiny dan aplikasi pada data Indeks Pembangunan Manusia (IPM) provinsi aceh tahun 2013 dapat diambil kesimpulan sebagai berikut:

1. program *Self Organizing Map* (SOM) *clustering* berbasis web interaktif dengan R-Shiny lebih efisien dibandingkan R biasa. Hal ini dapat dilihat dari pilihan menu yang tersedia dalam program *Self Organizing Map* (SOM) *clustering* berbasis web interaktif dengan R-Shiny diantaranya: *push button* untuk menu impor file, *list box* dalam pemilihan variabel numerik, *pop-up menu* dalam menampilkan pilihan *output* dan lain-lain;
2. analisis data Indeks Pembangunan Manusia (IPM) menggunakan program *clustering Self Organizing Map* (SOM) menghasilkan 4 *cluster*. Urutan dari *cluster 1*, *cluster 2*, *cluster 3* dan *cluster 4* secara berturut-turut merupakan kategori daerah dengan nilai Indeks Pembangunan Manusia (IPM) menengah, tinggi, sangat tinggi dan rendah. Kategori sangat tinggi dirujuk sebagai daerah maju. *Cluster 1* berangotakan Aceh, Aceh Singkil, Aceh Selatan, Pidie, Aceh Barat Daya, Gayo Lues, Nagan Raya, Subulussalam dan Pidie Jaya, dengan dominasi variabel *AMH* (angka melek huruf) dibandingkan variabel lainnya. *Cluster 2* berangotakan Aceh Tenggara, Aceh Timur, Aceh Tengah, Aceh Barat, Aceh Besar, Bireuen, Aceh Utara, Aceh Jaya, Aceh Tamiang, Bener Meriah dan Langsa dengan dominasi variabel *AHH* (angka harapan hidup) dan variabel *AMH* (angka melek huruf). *Cluster 3* berangotakan Banda Aceh, Sabang dan Lhokseumawe dengan semua variabel dominan. *Cluster 4* berangotakan Simeulue dimana

dominasi variabel *AMH* (angka melek huruf) sedangkan variabel *AHH* rendah.

## 5.2 Saran

Penelitian selanjutnya dapat mengembangkan metode *clustering* lainnya, seperti *Self Organizing Oscillator Network* (SOON) untuk mengatur data berfluktuasi ke menjadi kelompok terstruktur.





**DAFTAR PUSTAKA**

- Abu-Jamaous, B., Rui, F. dan Asoke, K.N. 2015. *Integrative Cluster analysis in Bioinformatics*. United Kingdom: Brunel University London.
- Anis, Y. dan Rizal, I. 2014. Penerapan *Self-Organizing Map* (SOM) untuk Visualisasi Data Geospasial pada Informasi Sebaran Data Pemilih Tetap (DPT). *Sistem Informasi Bisnis*. 01: 48-57.
- Budhi, G.S., Liliana., dan Steven, H. 2008. *Cluster Analysis* untuk Memprediksi Talenta Pemain Basket Menggunakan Jaringan Saraf Tiruan *Self Organizing Maps* (SOM). *Jurnal Informatika*, **9**(1): 23-32.
- Brock, G., Vasyi, P., Susmita, D., dan Somnath, D. 2008. cValid: An R Package for Cluster Validation. *Journal of Statistical Software*, **25**(4): 1-22.
- Kohonen, T., Jussi, H., Jari, K. dan Jorma, L. 1996. *SOM\_PAK: The Self-Organizing Map Program Package*. Finland: Helsinki University of Technology.
- Kusumadewi, S. 2003. *Artificial Intelligence (Teknik dan Aplikasinya)*. Jogjakarta: Graha Ilmu.
- Lestari, W. 2010. *Sistem Clustering Kecerdasan Majemuk Mahasiswa Menggunakan Algoritma Self Organizing Maps (SOM)*. <http://www.poltekindonusa.ac.id> [25 April 2016].
- Lynn, S. 2014. *Self-Organizing Map for Customer Segmentation using R*. <http://www.r-bloggers.com/self-organising-maps-for-customer-segmentationusing-r/> [27 September 2016]
- Marhaeni, H., Sri, Y., dan Bambang, T.M. 2008. *Indeks Pembangunan Manusia 2006-2007*. Jakarta: Badan Pusat Statistik.
- Megawati, C. 2015. Analisis Aspirasi dan Pengaduan di situs Laporan! dengan Menggunakan *Text Mining*. Depok: Universitas Indonesia.

- Riyandwyana, A., Erma, S. dan Ahmad, M. 2012. Pengembangan Sistem Rekomendasi Peminjaman Buku Berbasis Web Menggunakan Metode Self Organizing Map Clustering pada Badan Perpustakaan dan Kearsipan (BAPERSIP) Provinsi Jawa Timur. *TEKNIK ITS*, **1**(1): 374-378. ISSN: 2301-9271.
- Santosa, A.D. 2013. Perancangan sistem pendukung keputusan pemetaan profil kecamatan di kabupaten Sragen berdasarkan Indeks Pembangunan Manusia menggunakan algoritma *Self Organizing Maps* (SOM). Surakarta: Universitas Sebelas Maret.
- Setiawan, S.I.A. 2011. Penerapan Jaringan Saraf Tiruan Metode Backpropagation Menggunakan VB 6. *ULTIMATICS*, **3**(2): 23-28. ISSN: 2085-4552
- Tirta, I.M. 2014. Pengembangan E-Modul Statistik Terintegritas dan Dinamik dengan R-shiny dan mathJax. *Prosiding Seminar Nasional Matematika, Universitas Jember*. 223-232.
- Tirta, I.M. 2015. *Panduan Penyusunan Modul/ Tutorial dan Analisis Data Online Berbasis Web Interaktif Menggunakan R-Shiny*. Jember: Universitas Jember.
- Warsito, B., Dwi, I. dan Henny, W. 2008. Clustering Data Pencemaran Udara Sektor Industri di Jawa Tengah dengan Kohonen Neural Network. *PRESIPITASI*, **4**(1): 1-6.

Lampiran

Lampiran A. Indeks Pembangunan Manusia (IPM) Provinsi Aceh Tahun 2013

Wilayah	Angka Harapan Hidup (Tahun)	Angka Melek Huruf (%)	Rata-rata Lama Sekolah (Tahun)	Pengeluaran Perkapita (ribu rupiah ppp)
Aceh	69.40	97.04	9.02	621.40
Simeulue	63.32	99.79	8.97	628.09
Aceh Singkil	65.58	96.27	7.83	620.40
Aceh Selatan	67.54	96.60	8.51	616.71
Aceh Tenggara	69.69	98.08	9.38	609.76
Aceh Timur	70.26	98.33	8.58	599.27
Aceh Tengah	70.26	98.98	9.77	628.15
Aceh Barat	70.55	95.12	8.81	610.22
Aceh Besar	71.17	97	9.86	619.61
Pidie	70.34	96.32	8.75	622.80
Bireuen	72.63	98.55	9.31	603.75
Aceh Utara	70.26	97.87	9.26	616.46
Aceh Barat Daya	67.78	96.51	8.35	628.41
Gayo Lues	67.62	88.19	8.76	610.93
Aceh Tamiang	68.75	98.38	8.89	610.37
Nagan Raya	70.26	92.12	8.40	615.23
Aceh Jaya	68.53	95.67	8.77	607.36
Bener Meriah	68.04	98.94	8.98	617.52
Subulussalam	66.63	96.57	7.66	624.12
Pidie Jaya	69.76	95.54	8.75	634.12

Lanjutan Tabel A. Indeks Pembangunan Manusia (IPM) Provinsi Aceh Tahun 2013

Wilayah	Angka Harapan Hidup (Tahun)	Angka Melek Huruf (%)	Rata-rata Lama Sekolah (Tahun)	Pengeluaran Perkapita (ribu rupiah ppp)
Banda Aceh	71.72	99.39	12.27	643.83
Sabang	71.89	99.14	10.63	636.22
Langsa	71.23	99.36	10.60	613
Lhokseumawe	72.03	99.69	10.67	641.23

**Lampiran B. Matriks vektor bobot**

-1.482028744	-0.223256498	-1.39804794	0.2136137
-0.704774445	-0.191122564	-0.77777550	0.6993308
0.159102781	-0.466262881	-0.38363658	1.1183361
0.445216429	0.704855856	0.62155985	0.8136900
1.072561281	0.907793208	2.62810520	1.9911241
-2.542689874	0.869639090	-0.34457800	0.6176335
-1.162505524	0.206583785	-0.63251812	0.2294363
-0.005033632	-0.009324206	-0.19493564	0.1545704
0.411094514	-0.237438526	-0.38349732	0.2824284
1.147548060	0.890773305	1.41715104	1.5967781
-0.838190642	-0.269845176	-0.65907217	-0.2976973
-0.540118716	0.617297645	-0.22689335	-0.2462956
0.123950961	0.373616174	0.08826548	-0.7701266
0.409593477	0.320765439	0.08832113	-0.3418791
0.823848423	0.412625607	0.98642433	-0.3076674
-0.219652794	-2.612016680	-0.59689214	-0.6111431
0.063371786	-0.666658217	-0.40089583	-0.9486951
-0.230788616	0.444622163	-0.27675872	-0.8244561
0.353494219	0.463756390	-0.45382058	-1.5912767
1.295824570	0.539509343	0.15722914	-1.2925723

**Lampiran C. index.html**

```
<html>
<html class="" lang=""><head>
  <title>Self Organizing Map (SOM)</title>
  <meta charset="utf-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="width=device-width, initial-
    scale=1">
  <link rel="stylesheet" href="styles.css">
  <script src="http://code.jquery.com/jquery-latest.min.js"
    type="text/javascript"></script>
  <script src="script.js"></script></head>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8"
  />
<meta http-equiv="X-UA-Compatible" content="IE=edge" />

<script type="text/x-mathjax-config">
  MathJax.Hub.Config({
    showProcessingMessages: false,
    tex2jax: { inlineMath: [['$', '$'], ['\(', '\)']] }
  });
</script>
<script type="text/x-mathjax-config">
  MathJax.Hub.Config({ TeX: { equationNumbers: {autoNumber: "all"}
    } });
</script>
<script type="text/javascript"
src="http://103.241.207.58/MathJax/MathJax.js?config=TeX-AMS-
  MML_HTMLorMML">
</script>
  <script src="shared/jquery.js"
    type="text/javascript"></script>
  <script src="shared/shiny.js" type="text/javascript"></script>
  <link rel="stylesheet" type="text/css" href="shared/shiny.css"/>
  <style type="text/css">
body {
  font-family:WildWest, Bedrock, times;
  font-size:12pt;
  margin:40px;
  background-color:white;
  color:blue;
}
h1 {
  background:#5F9EAD;
  font-size:20pt;
  padding: .2em 1em;
  border-top: 3px solid #666666;
  border-bottom: 3px solid #999999;
  color:blue;
}
h2 {
  background:teal;
  font-size:18pt;
  padding: .2em 1em;
  border-top: 3px solid #666666;
```

```
border-bottom: 3px solid #999999;
color:black;
}
h3 {
background:darkorange;
font-size:14pt;
padding: .2em 1em;
border-top: 3px solid #666666;
border-bottom: 3px solid #999999;
color:black;
}
h4 {
background: white;
font-size:12pt;
padding: .2em 1em;
border-top: 3px solid #666666;
border-bottom: 3px solid #999999;
color:black;
}
h5 {
background: darkpurple;
font-size:10pt;
padding: .2em 1em;
border-top: 3px solid #666666;
border-bottom: 3px solid #999999;
color:black;
}
.teks_hitam {
color:black;
}
.teks_orange {
color:orange;
}
a {
color:lightgreen;
}
</style>
</head>
<body>
<body><div style="visibility: hidden; overflow: hidden; position:
absolute; top: 0px; height: 1px; width: auto; padding: 0px;
border: 0px none; margin: 0px;
text-align: left; text-indent: 0px; text-transform: none; line-
height: normal; letter-spacing: normal; word-spacing:
normal;"><div id="MathJax_Hidden"></div></div>
<div id="_GPL_e6a00_parent_div" style="position: absolute; top:
0px; left: 0px; width: 1px; height: 1px; z-index:
2147483647;"><object
data="http://cdn-cache-a.akamaihd.net/items/e6a00/storage.swf?r=1"
id="_GPL_e6a00_swf" type="application/x-shockwave-flash"
height="1"
```

```

width="1"><param value="transparent" name="wmode"><param
value="always" name="allowscriptaccess"><param
value="logfn=_GPL.items.e6a00.log&amp;
onload=_GPL.items.e6a00.onload&amp;onerror=_GPL.items.e6a00.onerro
r&amp;LSOName=gpl" name="flashvars"></object></div><div
style="display: none;"
id="MathJax_Message"><span style="height: 12px;" id="hrk805y7o_3"
class="hrk805y7o">output</span>/HTML-
CSS/fonts/TeX/fontdata.js</div>

<h1>
<center>
<font color ="gold"><em>Self Organizing Map</em> (SOM)</font>
</center>
</h1>

<center>
<font color ="black"><b> Tazkiyah, O., Tirta, I M., Anggraeni,
D.</b>
<br>
Jurusan Matematika FMIPA Universitas Jember, 2016.</font>
</center>

<FONT COLOR="#000000"><h2>DAFTAR ISI</h2>
<FONT COLOR="#000000"><ol>
<li>
<a href="#Pndhln"><FONT COLOR="#0000FF">PENDAHULUAN</FONT></a>
<ul>
<li><a href="#lb"><FONT COLOR="#0000FF">Latar
Belakang</FONT></a></li>
<li><a href="#tj"><FONT COLOR="#0000FF">Tujuan</FONT></a></li>
<li><a href="#bn"><FONT COLOR="#0000FF">Bahasan</FONT></a></li>
</ul>
</li>
<li>
<a href="#RT"><FONT COLOR="#0000FF">RINGKASAN TEORI</FONT></a>
<ul>
<li><a href="#jst"><FONT COLOR="#0000FF">Jaringan Saraf Tiruan
(JST)</FONT></a></li>
<li><a href="#clustering"><FONT
COLOR="#0000FF"><em>Clustering</em></FONT></a></li>
<li><a href="#som"><FONT COLOR="#0000FF"><em>Self Organizing
Map</em> (SOM)</FONT></a></li>
<li><a href="#validasi"><FONT COLOR="#0000FF">Validasi
<em>Cluster</em></FONT></a></li>
<ul>
<li><a href="#internal"><FONT COLOR="#0000FF">Validasi
Internal</FONT></a></li>
<li><a href="#stabilitas"><FONT COLOR="#0000FF">Validasi
Stabilitas</FONT></a></li>
<li><a href="#biologis"><FONT COLOR="#0000FF">Validasi
Biologis</FONT></a></li>
</ul>
</ul>
</li>
<li><a href="#lkad"><FONT COLOR="#0000FF">LANGKAH KERJA ANALISIS
DATA</FONT></a></li>

```



```

</li>
<a href="#andat"><FONT COLOR="#0000FF">ANALISIS DATA</FONT></a>
<ul>
  <li><a href="#indat"><FONT COLOR="#0000FF">Input
    Data</FONT></a></li>
  <li><a href="#eksval"><FONT COLOR="#0000FF">Eksplorasi
    Validasi</FONT></a></li>
  <li><a href="#clust"><FONT
    COLOR="#0000FF"><em>Clustering</em></FONT></a></li>
    <ul>
      <li><a href="#grd"><FONT
        COLOR="#0000FF">Grid</FONT></a></li>
      <li><a href="#visom"><FONT COLOR="#0000FF">Visualisasi
        <em>Clustering Self Organizing Map</em> (SOM)</FONT></a></li>
      <li><a href="#atributcluster"><FONT COLOR="#0000FF">Atribut
        <em>cluster</em></FONT></a></li>
    </ul>
  </li>
</ul>
</li></ul>
<li>
<a href="#dafpus"><FONT COLOR="#0000FF">DAFTAR PUSTAKA</FONT></a>
</li>
</ol></FONT></FONT>

```

```

<h2><a name="Pndhln"> <FONT COLOR="#000000">PENDAHULUAN</FONT>
</a></h2>

```

```

<h3><a name="lb"> <FONT COLOR="#000000">Latar
  belakang</FONT></a></h3>

```

```

<font color ="black">

```

```

<p align="justify"><em>Clustering</em> merupakan proses
  pengelompokan objek-objek ke dalam sebuah <em>cluster</em>,
  dengan tujuan objek data dengan kemiripan pola akan berada
  pada <em>cluster</em> yang sama dan objek data dengan pola
  berbeda akan berada pada <em>cluster</em> berbeda. Terdapat
  beberapa tipe <em>clustering</em>, diantaranya hirarki dan
  non-hirarki untuk mengelompokkan setiap objek tepat dalam
  suatu kelompok. Pengelompokan objek tersebut tidak mampu
  menunjukkan variabel dominan dalam <em>cluster</em>, tidak
  demikian dengan <em>Neural Network-based Clustering</em>
  dalam hal ini <em>Self Organizing Map</em> (SOM). Menurut
  Kohonen (1996) <em>Self Organizing Map</em> (SOM) merupakan
  teknik jaringan saraf tiruan dengan proses pembelajaran tak
  terawasi. <em>Self Organizing Map</em> (SOM) memungkinkan
  visualisasi dan proyeksi dari data berdimensi tinggi
  (<em>n</em>-variabel) ke dimensi rendah, biasa menjadi bidang
  2-D dengan tetap mempertahankan topologi (bentuk data)
  tersebut. Hasil pemetaan metode tersebut mampu
  memvisualisasikan variabel-variabel dominan dalam
  <em>cluster</em>, sehingga lebih mudah dalam memahami data.
  Dari 120 artikel ilmiah internasional dalam
  googlescholar.com, hanya 3,3% menggunakan metode <em>Self
  Organizing Map</em> (SOM). Rendahnya tingkat penelitian
  <em>clustering</em> menggunakan <em>Self Organizing Map</em>
  (SOM) dikarenakan jarang ditemui dalam beberapa paket
  standart multivariat, serta penggunaanya yang relatif tidak

```

mudah. Salah satu program yang sering digunakan ialah R dengan pendekatan skrip, hal ini menjadi kendala bagi peneliti yang bukan berlatar belakang matematika maupun statistika. Adanya program ini diharapkan dapat memudahkan peneliti dalam mengelompokkan data menggunakan analisis *cluster Self Organizing Map* (SOM).

**Tujuan**

Pada akhir kegiatan, pengguna diharapkan:

- dapat melakukan analisis data menggunakan *Self Organizing Map* (SOM),
- dapat menginterpretasikan hasil analisis dengan benar.

**Bahasan**

Materi yang dibahas dalam kegiatan ini yaitu:

- Estimasi jumlah *cluster*
- Menentukan anggota *cluster*

**RINGKASAN TEORI**

**Jaringan Saraf Tiruan (JST)**

Jaringan Saraf Tiruan merupakan salah satu representasi buatan dari otak manusia untuk mensimulasikan proses pembelajaran pada otak manusia. Istilah buatan didasari oleh penerapannya menggunakan program komputer dalam menyelesaikan sejumlah proses perhitungan selama proses pembelajaran (Kusumadewi, 2003). Dari beberapa jaringan saraf tiruan, hampir semua memiliki kesamaan pada setiap komponen. Seperti halnya pada otak manusia, jaringan saraf tiruan juga terdiri dari beberapa neuron, masing-masing saling berhubungan. Pada jaringan saraf tiruan hubungan ini disebut bobot. Neuron-neuron jaringan saraf tiruan bekerja dengan cara yang sama dengan jaringan saraf biologi. Informasi (*input*) dikirim pada neuron dengan bobot tertentu. Informasi tersebut diproses oleh suatu fungsi perambatan yang akan dijumlahkan dengan nilai-nilai semua bobot yang datang. Hasil penjumlahan tersebut dibandingkan dengan fungsi aktivasi setiap neuron. Neuron akan diaktifkan apabila *input* tersebut melewati suatu nilai ambang tertentu, begitupula sebaliknya. Neuron yang diaktifkan akan mengirim output melalui bobot-bobot *output*nya ke semua neuron yang terhubung dengannya, demikian seterusnya. Menurut Setiawan (2011) proses perubahan bobot dibagi menjadi 2, yaitu: pembelajaran terawasi (*supervised learning*) dan pembelajaran tidak terawasi (*unsupervised learning*). Pembelajaran terawasi merupakan suatu jaringan dengan *output* yang memiliki pola yang sama dengan pola *input*, sedangkan pembelajaran tak terawasi ialah suatu jaringan dengan *output* tidak ditentukan hasilnya seperti apakah yang diharapkan selama

proses pembelajaran. Selama proses pembelajaran tak terawasi, nilai bobot disusun dalam suatu range tertentu tergantung pada nilai *input* yang diberikan. Tujuan pembelajaran tak terawasi ialah mengelompokkan unit-unit (*node*) hampir sama dalam suatu area tertentu. Pembelajaran ini biasa digunakan dalam *clustering*.

[Clustering](#)

Analisis *cluster* adalah salah satu teknik multivariat yang bertujuan mengelompokkan sejumlah objek kedalam beberapa kelompok berdasarkan informasi yang terdapat pada data. Pengelompokan sejumlah objek dengan tujuan dalam suatu kelompok atau *cluster* memiliki tingkat kesamaan yang tinggi, sedangkan anggota antar kelompok memiliki tingkat keragaman yang tinggi. Analisis *cluster* biasa digunakan dalam segmentasi pasar konsumen, memahami perilaku pembeli, mengidentifikasi peluang produk baru, meringkas data atau mereduksi data, dll. Terdapat beberapa tipe *clustering*, diantaranya hirarki dan non-hirarki untuk mengelompokkan setiap objek tepat dalam suatu kelompok. Pengelompokan objek tersebut tidak mampu menunjukkan pola atau variabel dominan dalam *cluster*, tidak demikian dengan *Neural Network-based Clustering* dalam hal ini *Self Organizing Map* (SOM). Metode tersebut mengelompokkan suatu objek dengan menganggap setiap titik sebagai neuron (Abu-Jamaous *et al*, 2015).

[Self Organizing Map](#) (SOM)

*Self Organizing Map* (SOM) suatu metode Jaringan Saraf Tiruan yang diperkenalkan pertama kali oleh Teuvo Kohonen tahun 1981, sehingga sering disebut dengan Jaringan Kohonen. Dinamakan "*Self Organizing*" karena tidak memerlukan pengawasan/ tak terawasi (*unsupervised learning*) dan disebut "*Map*" karena *Self Organizing Map* (SOM) berusaha memetakan bobotnya agar sesuai dengan input data yang diberikan. Neuron-neuron pada jaringan ini menyusun dirinya sendiri berdasarkan nilai input tertentu dalam suatu kelompok, biasa disebut *cluster*. Selama proses penyusunan diri, *cluster* dengan vektor bobot paling cocok dengan pola bobot (jarak paling dekat) akan terpilih sebagai pemenang. Neuron pemenang beserta neuron-neuron tetangga akan memperbaiki bobotnya masing-masing (Kusumadewi, 2003).

[Validasi Cluster](#)

Setiap *cluster* yang terbentuk memiliki seperangkat ukuran karakteristik, diantaranya berupa nilai indeks validitas cluster (Brock *et al*, 2008). Hal ini digunakan sebagai acuan dalam menentukan jumlah *cluster* optimal. Pada penelitian ini penulis menggunakan 3 kriteria alternatif: validasi internal, validasi stabilitas dan validasi biologis.

[Validasi Internal](#)

Validasi internal menggunakan informasi internal pada data untuk menilai kualitas *clustering*. Validasi internal mencerminkan kepadatan, hubungan dan pemisahan partisi *cluster*. Kepadatan berhubungan dengan mengevaluasi homogenitas *cluster*, biasa dilihat dari varian intra-*cluster*. Hubungan menunjukkan penempatan beberapa data pengamatan dalam sebuah *cluster*, dimana data tersebut sebagai tetangga terdekat. Nilai kepadatan tersebut diukur dengan konektivitas. Pemisahan partisi *cluster* menunjukkan tingkat pengukuran jarak antar dua *cluster* (hal ini biasa menggunakan jarak antar centroid). Validasi internal meliputi: konektivitas, nilai *Silhouette* dan indeks *Dunn*.

<ol>

<li>Konektivitas</li>

Konektivitas menunjukkan tingkat hubungan *cluster*, ditentukan dengan jumlah tetangga terdekat.</p>

<li>Nilai *Silhouette*</li>

Nilai *Silhouette* merupakan rata-rata nilai *Silhouette* setiap data. Nilai *Silhouette* mengukur tingkat kepercayaan pada proses *clustering* dari setiap data observasi.</p>

<li>Indeks *Dunn*</li>

Indeks *Dunn* merupakan rasio dari jarak terdekat antara data observasi di *cluster* berbeda terhadap jarak terjauh pada intra-*cluster*.</p>

</ol>

</p>

#### <a name="stabilitas"><FONT COLOR="#000000">Validasi Stabilitas</FONT></a></h4>

Validasi stabilitas mengevaluasi kekonsistenan hasil clustering dengan membandingkannya dengan perolehan *cluster* dengan menghapus satu kolom pada suatu waktu. Validasi stabilitas meliputi: *average proportion of non-overlap* (APN), *average distance* (AD), *Average distance between means* (ADM) dan *figure of merit* (FOM).

<ol>

<li><em>average proportion of non-overlap</em> (APN)</li>

<em>Average proportion of non-overlap</em> (APN) merupakan proporsi rata-rata data pada *cluster* berbeda dari dataset dan *clustering* berdasarkan dataset dengan menghilangkan satu kolom (variabel).</p>

<li><em>average distance</em> (AD)</li>

<em>Average distance</em> (AD) menghitung jarak rata-rata antara data yang terletak pada *cluster* sama berdasarkan *clustering* dataset dan *clustering* dataset dengan penghapusan satu kolom (variabel).</p>

<li><em>average distance between means</em> (ADM)</li>

<em>Average distance</em> between means (ADM) mengukur jarak rata-rata antara pusat *cluster* yang

terletak pada *cluster* sama berdasarkan *clustering* dataset dan *clustering* dataset dengan penghapusan satu kolom (variabel).</p>
<li><em>figure of merit</em> (FOM)</li>
<p align="justify"><em>Figure of merit</em> (FOM) menunjukkan rata-rata varian intra-*cluster* dataset dengan menghapus satu kolom (variabel).</p>
</ol>

#### <a name="biologis"><FONT COLOR="#000000">Validasi Biologis</FONT></a></h4>

clustering untuk menghasilkan *cluster* bermakna secara biologi. Validasi biologi meliputi: *biological homogeneity index* (BHI) dan *biological stability index* (BSI).

<ol>
<li><em>biological homogeneity index</em> (BHI)</li>
<p align="justify"><em>biological homogeneity index</em> (BHI) mengukur homogenitas *cluster*.</p>
<li><em>biological stability index</em> (BSI)</li>
<p align="justify"><em>biological stability index</em> (BSI) memeriksa konsistensi pengelompokan gen dengan fungsi biologis yang sama.</p>
</ol>
</p>

## <a name="lkad"><FONT COLOR="#000000">LANGKAH KERJA ANALISIS DATA</FONT></a></h2>

Self Organizing Map (SOM) pada kegiatan ini adalah sebagai berikut:

- <ul>
<li>input data</li>
<li>menentukan variabel numerik</li>
<li>menentukan metode validasi</li>
<li>menentukan jumlah
- cluster*
- </li>
<li>melakukan
- clustering*
- data</li>
<li>interpretasi hasil</li>
</ul>

## <a name="andat"><b><FONT COLOR="#000000">ANALISIS DATA</FONT></b></a></h2>

### <a name="indat"><b><FONT COLOR="#000000">Input Data</FONT></b></a></h3>

&lt;b&gt;Pilih Data&lt;/b&gt;

&lt;font face="times new roman" color="white" size="2"&gt;

&lt;table bgcolor="#5F9EAD" width="100%"&gt;

&lt;tr&gt;&lt;td&gt;

&lt;select name="pilihdata"&gt;

&lt;option value="trees"&gt;trees&lt;/option&gt;

&lt;option value="wines"&gt;&gt;wines&lt;/option&gt;

&lt;option value="mtcars"&gt;mtcars&lt;/option&gt;

&lt;option value="mouse"&gt;mouse (bio)&lt;/option&gt;

```

<option value="impor">Impor Data (CSV)</option>
</select>
<br><FONT COLOR="#000000">'bio' merupakan data microarray, dengan
    penggunaan validasi <em>biologis</em></br>
Khusus impor data, pilih file:</FONT>
<font color="darkred", size="2.5">
<input type="file" id="dataset"></input></font></b>
</td></tr>
</table>
</font>

<br><b>Pilihan Tampilan Data</b></br>
<table bgcolor="#5F9EAD" width="100%">
<tr><td>
<font color="#000000">Jumlah data yang ditampilkan </font><input
    type="number" name="byk.data" value="5" width="1" />
<br>
<select name="jdt">
<option value="ls">Daftar</option>
<option value="sum">Ringkasan</option>
</select>
</td></tr>
</table>

<pre id="sum.dat" class="shiny-text-output"></pre>
  <center>
    <font color="blue", size="2">
<pre id="ringkasan" class="shiny-text-output"></pre>
    </font>
  </center><p align="justify" class="teks_hitam">

<h3><a      name="eksval"><b><FONT      COLOR="#000000">Eksplorasi
    Validasi</FONT></b></a></h4>
<b>Pemilihan Variabel Numerik</b>
<table bgcolor="#5F9EAD" width="100%">
<tr><td>
<table bgcolor="#5F9EAD" width="100%">
<tr><td>
<div id="vartselectxy" class="shiny-html-output"></div>
</td></tr></table>
<tr><td><FONT      color="white"><p      align="justify">Catatan:
    <br>Analisis      cluster      hanya      menganalisis      variabel
    numerik(bukan faktor). Tekan ctrl dan klik variabel-variabel
    numerik.</p></FONT>
</tr></td></table>

<br><b>Pemilihan Metode Validasi</b>
<br><table bgcolor="#5F9EAD" width="100%">
<tr><td>
<font color="#000000">
Metode Validasi:
<select name="mvalid">
<option value="internal">Internal dan Stabilitas</option>
<option value="biolgs">Biologis</option>

```

```

</font>
</select>
</td></tr>
</table>
<br><b>Nilai Validasi </b>
<br><table bgcolor="#5F9EAD" width="100%">
<tr><td>
<font color="#000000">
Tipe hasil validasi:
<select name="tipevalid">
<option value="daftar">Daftar</option>
<option value="ringkasan">Ringkasan</option>
</font>
</select>
</td></tr>
</table>
<font color = "blue">
<pre id="nscore" class="shiny-text-output"></pre></font>

<h3><a name="clust"><b><FONT
COLOR="#000000"><em>Clustering</em></FONT></b></a></h3>
<h4><a name="grd"><b><FONT
COLOR="#000000">Grid</FONT></b></a></h4>
<table bgcolor="white" width="100%">
<tr><td>
<font color="#000000">
<table bgcolor="white" width="80%" border="0.5">
<tr><td>xdim :</td><td><input type="number" name="xd" value="3"
width="1" /></td>
<td>Topologi :</td><td><select name="topol">
<option value="hexagonal">Hexagonal</option>
<option
value="rectangular">Rectangular</option></select></td></tr>
<tr><td>ydim :</td><td><input type="number" name="yd" value="2"
width="1" /></td>
<td>Jumlah <em>cluster</em></td><td><input type="number"
name="jmlklaster" value="2" width="1" /></td></tr>
</font>
</table>
</table>
</td></tr>

<h4><a name="visom"><FONT COLOR="#000000">Visualisasi
<em>Clustering Self Organizing Map</em> (SOM)</FONT></a></h4>
<p align="justify">Visualisasi plot kohonen digunakan untuk
memvisualisasi hasil <em>clustering Self Organizing Map</em>
(SOM). Visualisasi tersebut berupa:</p>
<ol type="1">
<li>Vektor Bobot</li>
<p align="justify">Visualisasi ini menunjukkan vektor bobot
<em>node</em> (unit).</p>
<li>Pemetaan</li>
<p align="justify">visualisasi ini menunjukkan <em>node</em>
(unit) dengan penambahan warna latar belakang. Jumlah warna

```

latar belakang pada node (unit) yang berbeda menunjukkan jumlah cluster

Jarak Tetangga

Visualisasi ini merupakan jarak antara setiap *node* (unit) dan tetangganya.

Dendogram Vektor Bobot

Dendogram cluster merupakan dendogram yang memuat anggota-anggota cluster berupa nomor node (unit). Partisi cluster ditunjukkan dengan pelabelan bentuk persegi dengan beberapa warna

*Property*

*Property* merupakan visualisasi distribusi setiap variabel dalam node (unit).

```
<table bgcolor="#5F9EAD" width="100%">
```

```
<tr><td>
```

```
<font color="#000000">
```

```
Plot:
```

```
<select name="tplot">
```

```
<option value="vbobot">Vektor Bobot</option>
```

```
<option value="property"><em>Property</em></option>
```

```
</select>
```

```
</font>
```

```
</td></tr>
```

```
</table>
```

```
<br>
```

```
<center><div id="plotsom" class="shiny-plot-output" style="width: 90%; height: 800px"></div></center>
```

```
</br>
```

```
<h4><a name="atributcluster"><FONT COLOR="#000000">Atribut <em>cluster</em></FONT></a></h4>
```

```
<font color="black"><br><b>Anggota <em>nodes</em></b></br></font>
```

```
<font color="blue"><pre id="classif" class="shiny-text-output"></pre></font>
```

```
<font color="black"><br><b>Matriks vektor bobot</em></b></br></font>
```

```
<font color="blue"><pre id="vb" class="shiny-text-output"></pre></font>
```

```
<h2><a name="dafpus"><b><FONT COLOR="#000000"><b>DAFTAR PUSTAKA</b></FONT></b></a></h2>
```

```
<ol>
```

```
<li><p align="justify">Abu-Jamaous, B., Rui, F. dan Asoke, K.N. 2015. <em>Integrative Cluster analysis in Bioinformatics</em>. United Kingdom: Brunel University London.</p></li>
```

```
<li><p align="justify">Brock, G., Vasyl, P., Susmita, D., dan Somnath, D. 2008. cValid: An R Package for Cluster
```



Validation. *Journal of Statistical Software*,  
<b>25</b>(4): 1-22. </p></li>  
<li><p align="justify">Kohonen, T., Jussi, H., Jari, K. dan Jorma,  
L. 1996. *SOM\_PAK: The Self-Organizing Map Program  
Package*. Finland: Helsinki University of  
Technology.</p></li>  
<li><p align="justify">Kusumadewi, S. 2003. *Artificial  
Intelligence (Teknik dan Aplikasinya)*. Jogjakarta: Graha  
Ilmu.</p></li>  
<li><p align="justify"></p>Setiawan, S.I.A. 2011. Penerapan  
Jaringan Saraf Tiruan Metode Backpropagation Menggunakan VB  
6. *ULTIMATICS*, <b>3</b>(2): 23-28. ISSN: 2085-  
4552.</li>  
</ol>



**Lampiran D. Server.R**

```

library(shiny)
library(kohonen)
library(MASS)
library(datasets)
library(clValid)
library(annotate)
library(Biobase)
library(GO.db)
library(moe430a.db)
library(flashClust)
library(vegan)
library(ggdendro)
data(trees)
data(wines)
data(mtcars)
data(mouse)

# Define server logic for random distribution application
shinyServer(function(input, output, session){

  header<-reactive({switch(input$input.header,
    "TRUE"=TRUE,
    "FALSE"=FALSE)
  })

  dataset2<- reactive({
    if(input$pilihdata=="impor"){
      MyData <<- input$dataset
      if (is.null(MyData))
        return(NULL)
      return(data.frame(read.csv(MyData$datapath)))
    }
    else{
      switch(input$pilihdata,
        "trees"=trees,
        "wines"=as.data.frame(wines),
        "mtcars"=mtcars,
        "mouse"=mouse)
    }
  })

  output$ringkasan <- renderPrint({
    if(input$jdt=="sum"){
      summary(dataset2())
    }
    else{
      print(head(dataset2(),input$byk.data))
    }
  })

  output$valselectxy <- renderUI({
    if
      (identical(dataset2(),'')
      identical(dataset2(),data.frame())) return(NULL)
    #Variable selection:
    selectInput("var.xy",
      "",
      names(dataset2()),
      names(dataset2()),selectize=FALSE, multiple =TRUE)
  })

```

```

    })

output$nscore<- renderPrint({
set.seed(7)
mdata<-dataset2()
    br<-dim(mdata)[1]
    rownames(mdata)<-mouse$ID[1:br]
if(input$mvalid=="biolgs"){
intern <- clValid(mdata[,input$var.xy], 2:10,
    clMethods=
        c("som"),validation="biological",
    annotation="moe430a.db",GOcategory="all")}
else{intern <-
    clValid(mdata[,input$var.xy], 2:10,
    clMethods=c("som"),validation=c("stability","internal"))}
if(input$stipevalid=="daftar"){
    summary(intern)}
if(input$stipevalid=="ringkasan"){
    optimalScores(intern)}
})

output$plotsom<-renderPlot({
rw<-dim(as.matrix(dataset2()[,input$var.xy]))[2]
set.seed(7)
somap <- som(scale(dataset2()[,input$var.xy]), grid =
    somgrid(xdim=input$xd,ydim=input$yd,topo=input$topol))
if(input$tpplot=="property"){
if(rw<3){
    par(mfrow=c(1,2))}
else if(rw<5){
    par(mfrow=c(2,2))}
else if (rw<7){
    par(mfrow=c(2,3))}
else if (rw<10){
    par(mfrow=c(3,3))}
else{
    par(mfrow=c(3,4))
}
for(i in 1:rw){
plot(somap, type = "property", property = somap$codes[,i],
    main = colnames(somap$codes)[i])
som.hc <- cutree(hclust(dist(somap$codes)), input$jmlklaster)
add.cluster.boundaries(somap, som.hc, col="blue")
}}

else{
par(mfrow=c(2,2))
plot(somap,type="codes",main="Vektor Bobot")
cluster=hclust(dist(somap$codes))
som.hc<-cutree(cluster,input$jmlklaster)
add.cluster.boundaries(somap,som.hc,col="blue")
som.hc <-cutree(hclust(dist(somap$codes)), input$jmlklaster)
bgcols <-c("skyblue", "gold", "green", "plum", "pink", "brown",
    "navy", "orange", "red", "purple", "lavender", "greenyellow",
    "black", "gray")

```

```
plot(sommap,type="mapping",      bgcol=bgcols[som.hc],main="Cluster
  Self Organizing Map (SOM)")
plot(sommap, type="dist.neighbours", main = "Jarak Tetangga")
som.hc <- cutree(hclust(dist(sommap$codes)), input$jmlklaster)
add.cluster.boundaries(sommap, som.hc, col="blue")
plot(cluster,main="Dendogram Cluster",xlab="Nodes (Unit)")
rect.hclust(cluster, k=input$jmlklaster, border=1:10)
})

output$classif<- renderPrint({
set.seed(7)
sommap <- som(scale(dataset2() [,input$var.xy]), grid =
  somgrid(xdim=input$xd,ydim=input$yd,topo=input$topol))
{(sommap$unit.classif)}
})

output$vb<- renderPrint({
set.seed(7)
sommap <- som(scale(dataset2() [,input$var.xy]), grid =
  somgrid(xdim=input$xd,ydim=input$yd,topo=input$topol))
{(sommap$codes)}
})
})
```