

PROGRAM APLIKASI PENGELOMPOKAN OBJEK DENGAN METODE *SELF ORGANIZING MAP* MENGUNAKAN BAHASA R

Siti Kania¹, Dewi Rachmatin² dan Jarnawi Afgani Dahlan³
¹²³Departemen Pendidikan Matematika Fakultas FPMIPA
Universitas Pendidikan Indonesia
*surel. sitikania@student.upi.edu¹ dewirachmatin@upi.edu²

ABSTRAK

Analisis kluster merupakan salah satu teknik analisis statistika multivariat yang bertujuan untuk mengalokasikan sekelompok objek pada suatu kelompok-kelompok yang saling bebas, yang disebut sebagai kluster, sehingga objek-objek di dalam satu kelompok homogen, sedangkan objek-objek di dalam kelompok yang berbeda heterogen. Pada penelitian ini, proses analisis kluster dilakukan dengan menggunakan metode *Self Organizing Map*. *Self Organizing Map* merupakan salah satu metode dalam jaringan syaraf tiruan yang menggunakan pembelajaran tak terawasi. Akan tetapi proses pengelompokan dengan metode *Self Organizing Map* memerlukan waktu yang cukup lama serta dapat terjadi kesalahan dalam perhitungannya apabila dilakukan secara manual, sehingga pada penelitian ini dibuat program aplikasi untuk proses pengelompokan objek dengan metode *Self Organizing Map* menggunakan bahasa pemrograman R. *Output* dari program aplikasi tersebut berupa proses *clustering* yang terdiri dari hasil perhitungan tiap iterasi dan hasil pengelompokan objek yang termuat dalam lembar kerja 'Console' pada *software* R. Setelah program aplikasi selesai dibuat, kemudian diaplikasikan pada data Indeks Pembangunan Manusia Provinsi Aceh tahun 2013. Dari program aplikasi tersebut dengan menentukan terlebih dahulu kluster yang akan dibentuk yaitu 4 kluster diperoleh jumlah anggota kluster ke 1,2,3, dan 4 secara berturut-turut adalah 10, 2, 1, dan 11.

Kata kunci : Analisis Kluster, *Self-Organizing Map*, R

OBJECTGROUPING APPLICATION PROGRAM WITH SELF ORGANIZING MAP METHOD USING R

ABSTRACT

Cluster analysis is one of multivariate analysis technique that is purposed to allocate a group of object of independent groups, the so-called cluster, so as to each object in a same group is homogeneous, whilst each object in different group is heterogeneous. In this research, the process of cluster analysis is conducted by employing Self Organizing Map method. Self Organizing Map is one of method in artificial neural network that use unsupervised learning. However, the process of grouping of Self Organizing Map took a long time and it can generate mistakes in its calculation if it is conducted manually, for that concern this research provides a program application for the grouping process with Self Organizing Map using R programming language. Output of the application program is a clustering process that consist of calculation result for each iteration and object grouping result provided in Console worksheet in R software. After the application program is completely created, finally it is applied to data of Human Development Index of Aceh province in 2013. From the application program, by first determining the cluster to be formed ie 4 clusters obtained by the number of cluster members to 1,2,3 and 4 respectively are 10, 2, 1, and 11.

Key Words: *Cluster Analysis, Self-Organizing Map, R-language*

1. PENDAHULUAN

Analisis kluster merupakan salah satu teknik analisis statistika multivariat yang bertujuan untuk mengalokasikan sekelompok objek pada suatu kelompok-kelompok yang saling bebas, yang disebut sebagai kluster, sehingga objek-objek didalam satu kelompok bersifat homogen, sedangkan objek-objek di dalam kelompok yang berbeda bersifat heterogen. Ada bermacam-macam metode untuk melakukan analisis kluster mulai dari metode yang sederhana diantaranya metode hierarki dan non-hierarki, hingga metode yang kompleks yaitu dengan menggunakan kecerdasan buatan, seperti jaringan syaraf tiruan (*artificial neural network*). Metode jaringan syaraf tiruan untuk melakukan *clustering* adalah metode jaringan syaraf (*neural network*) yang menggunakan *unsupervised learning* (pembelajaran tak terawasi), yaitu *Self-Organizing Map*.

Metode *Self-Organizing Map* (SOM) menghasilkan akurasi yang lebih baik dalam mengelompokkan objek ke dalam kelompok yang cocok dari pada algoritma *K-Means* dan metode hirarki serta menunjukkan hasil yang baik saat menggunakan himpunan data yang kecil daripada himpunan data yang besar [9]. Selain itu, pada penelitian yang dilakukan oleh Yiheng Chen dan kawan-kawan [8] dengan membandingkan algoritma SOM dan algoritma *K-Means* dalam kasus *clustering* dokumen (data sampel yang digunakan adalah 420 artikel dari topik yang berbeda), teridentifikasi bahwa secara keseluruhan kinerja algoritma SOM lebih bagus daripada algoritma *K-Means* untuk kasus *clustering* dokumen

Proses *clustering* dengan menggunakan metode *Self Organizing Map* memerlukan waktu yang cukup lama serta dapat terjadi kesalahan dalam perhitungannya apabila dilakukan secara manual. Oleh karena itu, perhitungan *Self Organizing Map* memerlukan bantuan *software* yang juga dapat menerjemahkan algoritma *Self Organizing Map* ke dalam bahasa pemrograman sehingga perhitungan *Self Organizing Map* dapat dikerjakan dengan program yang telah dibuat sesuai algoritma. Wiharto [11] menjelaskan bahwa R adalah suatu perangkat lunak (*software*) yang digunakan untuk manipulasi data, perhitungan, simulasi, penayangan grafik, dan sekaligus sebagai bahasa pemrograman yang bersifat *interpreter* (penerjemah). Lebih lanjut Zuur,*et. al* [1] menjelaskan bahwa R merupakan bahasa pemrograman komputer yang memungkinkan pengguna untuk memrogramkan algoritma dan menggunakan alat yang telah dikembangkan melalui R oleh pengguna lainnya.

Rachmatin [7] menjelaskan bahwa tidak seperti halnya dengan analisis multivariat yang lain (contohnya analisis diskriminan) memerlukan asumsi seperti normalitas, dalam analisis kluster asumsi yang harus diperhatikan adalah data bebas dari pencilan dan tidak ada multikolinieritas. Dalam melakukan pemilihan objek ke dalam kluster-kluster, analisis kluster peka terhadap pencilan. Kluster-kluster yang diperoleh akan tidak sesuai dengan struktur yang sebenarnya dari populasi jika pencilan dilibatkan dalam pengolahan data. Sedangkan jika data terdapat multikolinearitas antar variabel sebelum dilakukan analisis kluster, data awal terlebih dahulu ditransformasi melalui teknik komponen utama menjadi *zscore*.

2. METODOLOGI

Metode penelitian yang digunakan pada skripsi ini merupakan metode penelitian dasar yang mengkaji teori tentang analisis kluster, *Self Organizing Map* (SOM) kemudian mengkonstruksi algoritma *Self Organizing Map* ke dalam bahasa pemrograman R dan diaplikasikan pada suatu data.

2.1 *Self Organizing Map*

Metode Self-Organizing Map (SOM) merupakan suatu metode Jaringan Syaraf Tiruan yang diperkenalkan oleh Professor Teuvo Kohonen pada tahun 1980-an, sebagai salah satu bentuk topologi dari *Unsupervised Artificial Neural Network* (Unsupervised ANN) dimana dalam proses pelatihannya tidak memerlukan pengawasan atau target keluaran [1]. Jaringan SOM terdiri dari dua lapisan (*layer*), yaitu lapisan *input* dan lapisan *output*. Setiap neuron dalam lapisan *input* terhubung dengan setiap neuron pada lapisan *output*. Setiap neuron dalam lapisan *output* merepresentasikan kelas (kluster) dari *input* yang diberikan. Setiap neuron *output* mempunyai bobot untuk masing-masing neuron *input*. Selama proses *clustering*, kluster (neuron pada lapisan *output*) yang memiliki jarak paling dekat dengan pola input akan terpilih sebagai pemenang dan beserta neuron tetangganya akan memperbaiki bobotnya.

2.2 Algoritma *Self Organizing Map*.

Proses *Self Organizing Map* (SOM) dilakukan dengan mengikuti algoritma berikut [4] :

a. Masukan :

1. Neuron pada lapisan input $X = \{\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n\}$; $\vec{x}_i \in R^n$, di mana $\vec{x}_i = \langle x_{i1}, x_{i2}, \dots, x_{ip} \rangle$
2. Bobot $W = (w_{ij})$; $i = 1, 2, \dots, m$ dan $j = 1, 2, \dots, p$ di mana p adalah banyaknya variabel dan m adalah banyak kluster diinginkan.
3. *Learning rate* (α) bernilai $0 < \alpha < 1$
4. Jumlah iterasi (T).
5. Galat maksimum (ϵ)

b. Keluaran : $W = (w_{ij})$; $i = 1, 2, \dots, m$ dan $j = 1, 2, \dots, p$

Objek yang sudah tercluster

c. Langkah :

1. Inisialisasi bobot koneksi antara neuron input dan output (w_{ij}) dengan bilangan acak 0 sampai 1, di mana $i = 1, 2, \dots, p$ dan $j = 1, 2, \dots, m$ di mana p adalah banyaknya variabel dan m adalah banyak kluster diinginkan.
2. Hitung jarak vektor input terhadap bobot koneksi untuk masing-masing neuron output dengan menggunakan rumus :

$$d(j) = \sum_{i=1}^p (w_{ij} - x_{ki})^2, k = 1, 2, \dots, n \quad (2.1)$$

3. Tentukan indeks j sedemikian sehingga d_j minimum.
4. Perbaharui nilai bobot W_{ij} untuk setiap unit j dengan menggunakan rumus:

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + h(t)(x_{ni} - w_{ij})(\text{old}) \quad (2.2)$$

5. Ulangi langkah dua sampai langkah empat untuk setiap vektor input \mathbf{X}
6. Perbaharui *learning rate* (α) pada saat iterasi ke t ($t=1,2,3,\dots,T$) dengan persamaan :

$$\alpha(t+1) = \alpha(1)\left(1 - \frac{t}{T}\right) \quad (2.3)$$

7. Ulangi langkah kelima sampai keenam sampai maksimum iterasi tercapai dan *learning rate* konvergen menuju nol.
8. Selanjutnya dilakukan proses pengelompokan atau clusterisasi, yaitu dengan cara mencari jarak Euclid antara bobot terakhir yang didapat dengan seluruh objek pengamatan dengan menggunakan persamaan (2.1). Kemudian cari jarak terkecil pada setiap objek pengamatan.

2.3 Learning Rate ($\alpha(t)$)

Learning Rate digunakan untuk menunjukkan bagaimana adaptasi pembelajaran terhadap data. Fungsi skalar adaptasi bernilai $0 < \alpha(t) < 1$. Semakin besar nilai $\alpha(t)$, semakin cepat bobot koneksi beradaptasi/semakin besar pengaruh vektor input terhadap perubahan bobot koneksi yang terjadi. *Learning rate* ini semakin lama akan semakin mengecil, berkurang seiring berjalannya waktu/iterasi. Semakin dekat *learning rate* $\alpha(t)$ mendekati 0, perubahan bobot akan semakin kecil dan vektor-vektor input dapat dipetakan dengan baik [3]. Fungsi *Learning Rate* yang digunakan adalah fungsi linear, berikut persamaan dari fungsi linear :

$$\alpha(t+1) = \alpha(1)\left(1 - \frac{t}{T}\right) \quad (2.4)$$

di mana $\alpha(1)$ merupakan nilai *learning rate* awal yang ditentukan peneliti.

2.4 Neighborhood Function

Fungsi $h(t)$ adalah fungsi lingkungan (*Neighborhood Function*), yang bernilai $0 \leq h(t) \leq 1$. Fungsi ini memberi pengaruh perubahan bobot secara proporsional dari neuron *best matching* ke neuron - neuron tetangganya. Fungsi lingkungan merupakan fungsi posisi dari dua unit neuron (neuron pemenang dengan neuron lainnya) dengan *Neighborhood Radius* (r) yang diberikan. Terdapat beberapa fungsi lingkungan, dua fungsi lingkungan yang paling umum adalah *Gaussian Neighborhood* dan *bubble neighborhood*. *Gaussian Neighborhood* cenderung lebih dapat diandalkan (inisialisasi yang berbeda cenderung konvergen ke peta yang sama), sedangkan *bubble neighborhood* memiliki kuantisasi kesalahan yang lebih kecil, dan secara komputasi jauh lebih cepat [6]. Dalam penelitian ini, fungsi lingkungan yang akan digunakan adalah *bubble neighborhood* dan *Neighborhood Radius* sama dengan nol. Berikut fungsi lingkungan *bubble neighborhood*.

$$h(t) = \begin{cases} \alpha(t), & \text{untuk setiap bobot pada neuron pemenang } j \\ 0, & \text{lainnya} \end{cases} \quad (2.5)$$

2.5 Normalisasi Data

Sebelum dilakukan proses pembelajaran (*training*), data *input* harus dinormalisasi terlebih dahulu. Normalisasi adalah penskalaan terhadap nilai-nilai input sedemikian sehingga data-data input masuk dalam suatu range tertentu. Pada pembelajaran algoritma *Self Organizing Map* proses normalisasi perlu dilakukan agar rentang nilai pada masing-masing variabel tidak terpaut jauh. Lebih lanjut Larose [5] menjelaskan bahwa salah satu kelemahan dari jaringan syaraf tiruan adalah semua nilai variabel dari suatu objek harus dikodekan yaitu mengambil nilai antara nol dan satu, bahkan untuk variabel kategori. Oleh karena hal tersebut, proses normalisasi dilakukan dengan metode *Min-Max Normalization*. Metode *Min-Max Normalization* didefinisikan sebagai berikut :

$$x_i^* = \frac{x_i - \min_{x_i}}{\max_{x_i} - \min_{x_i}} \quad (2.6)$$

dengan :

- x_i^* : nilai variabel ke-i yang baru setelah dinormalisasi
- x_i : nilai variabel ke-i yang lama sebelum dinormalisasi
- \min_{x_i} : nilai minimum dari variabel ke-i
- \max_{x_i} : nilai maksimum dari variabel ke-i

3. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari *website* resmi Badan Pusat Statistik Provinsi Aceh [2]. Data tersebut adalah data Indeks Pembangunan Manusia (IPM) Provinsi Aceh tahun 2013 yang disajikan pada tabel berikut ini.

Tabel 3. 1 Data Indeks Pembangunan Manusia (IPM) Provinsi Aceh Tahun 2013

No.	Wilayah	Angka Harapan Hidup (Tahun)	Angka Melek Huruf (%)	Rata-Rata Lama Sekolah (Tahun)	Pengeluaran Perkapita (ribu rupiah ppp)
1	Aceh	69,4	97,04	9,02	621,4
2	Simeulue	63,32	99,79	8,97	628,09
3	Aceh Singkil	65,58	96,27	7,83	620,4
4	Aceh Selatan	67,54	96,6	8,51	616,71
5	Aceh Tenggara	69,69	98,08	9,38	609,76
6	Aceh Timur	70,26	98,33	8,58	599,27
7	Aceh Tengah	70,26	98,98	9,77	628,15
8	Aceh Barat	70,55	95,12	8,81	610,22
9	Aceh Besar	71,17	97	9,86	619,61
10	Pidie	70,34	96,32	8,75	622,8
11	Bireuen	72,63	98,55	9,31	603,75
12	Aceh Utara	70,26	97,87	9,26	616,46
13	Aceh Barat Daya	67,78	96,51	8,35	628,41
14	Gayo Lues	67,62	88,19	8,76	610,93
15	Aceh Tamiang	68,75	98,38	8,89	610,37
16	Nagan Raya	70,26	92,12	8,4	615,23
17	Aceh Jaya	68,53	95,67	8,77	607,36
18	Bener Meriah	68,04	98,94	8,98	617,52
19	Subulussalam	66,63	96,57	7,66	624,12
20	Pidie Jaya	69,76	95,54	8,75	634,12
21	Banda Aceh	71,72	99,39	12,27	643,83
22	Sabang	71,89	99,14	10,63	636,22
23	Langsa	71,23	99,36	10,6	613
24	Lhokseumawe	72,03	99,69	10,67	641,23

Proses *clustering* diawali dengan pengujian asumsi yaitu data bebas pencilan dan tidak adanya multikolinearitas. Untuk menguji data bebas dari pencilan maka diperlukan jarak Mahalanobis pada setiap wilayah dan nilai khi-kuadrat. Untuk mencari jarak Mahalanobis, peneliti menerjemahkan rumus jarak mahalanobis ke dalam bahasa pemrograman R. Berikut hasil dari perhitungan jarak Mahalanobis dengan bantuan bahasa pemrograman R dan nilai khi-kuadrat dengan bantuan *software excel*.

Tabel 3. 2 Hasil Perhitungan Jarak Mahalanobis dan Nilai Khi-Kuadrat

Wilayah	Jarak Mahalanobis	$\chi^2_{0,0001;3}$	Wilayah	Jarak Mahalanobis	$\chi^2_{0,0001;3}$
Aceh	0,1619723	23,51274	Aceh Tamiang	1,4600912	23,51274
Simeulue	11,8492578	23,51274	Nagan Raya	4,5409142	23,51274
Aceh Singkil	3,2836656	23,51274	Aceh Jaya	1,6895884	23,51274
Aceh Selatan	0,7693794	23,51274	Bener Meriah	1,2921040	23,51274
Aceh Tenggara	1,5618799	23,51274	Subulussalam	4,0871401	23,51274
Aceh Timur	4,6133486	23,51274	Pidie Jaya	4,6095191	23,51274
Aceh Tengah	0,9281464	23,51274	Banda Aceh	10,8247936	23,51274
Aceh Barat	1,5183797	23,51274	Sabang	3,1865807	23,51274
Aceh Besar	0,8228997	23,51274	Langsa	4,3856087	23,51274
Pidie	1,8326953	23,51274	Lhokseumawe	4,8427876	23,51274
Bireuen	5,1823637	23,51274			
Aceh Utara	0,4233160	23,51274			
Aceh Barat Daya	2,4371833	23,51274			

Dapat dilihat pada Tabel 3.2, karena jarak Malahanobis $< \chi^2_{((\alpha<0,001);p)} = \chi^2_{0,0001;3}$ hal ini berarti data bebas pencilan.

Sedangkan untuk menguji tidak adanya multikolinearitas antar variabel maka diperlukan koefisien korelasi antar variabel. Berikut koefisien korelasi antar variabel dengan bantuan bahasa pemrograman R.

Tabel 3. 3 Koefisien Korelasi Antar Variabel

Variabel	AHH	AMK	HLS	PP
AHH	1,00000000	0,1923716	0,6195756	0,04822058
AMK	0,19237156	1,00000000	0,4767187	0,28528569
HLS	0,61957559	0,4767187	1,00000000	0,48175330
PP	0,04822058	0,2852857	0,4817533	1,00000000

Berdasarkan Tabel 3.3, dapat dilihat bahwa nilai koefisien korelasi antara variabel Angka Harapan Hidup (AHH) dan Harapan Lama Sekolah (HLS) tinggi yaitu sebesar 0,61957559. Karena data mengandung korelasi maka dilakukan proses analisis komponen utama, yaitu dilakukan transformasi data awal menjadi *zscore*. Untuk melakukan transformasi data menggunakan analisis komponen utama, peneliti menerjemahkan rumus analisis komponen utama ke dalam bahasa pemrograman R.

Seperti yang telah dijelaskan sebelumnya bahwa salah satu kelemahan pada jaringan saraf tiruan yaitu semua nilai variabel dari suatu objek harus dikodekan yaitu mengambil nilai antara nol dan satu, maka data *zscore* hasil analisis komponen utama harus dinormalisasi terlebih dahulu.

Untuk memperlihatkan proses *clustering* yang terjadi dalam program aplikasi, peneliti mengambil 10 data yaitu 10 wilayah pertama pada provinsi Aceh untuk simulasi program aplikasi tersebut.

Tabel 3. 4 Data Masukan untuk Simulasi Proses Clustering Menggunakan Program Aplikasi

Data	Notasi	Value
Data olah	X	[,1] [,2] [,3] [,4]
		[1,] 0,4921277 0,35994947 0,333410023 0,285264545
		[2,] 0,6441351 0,47352936 1,000000000 0,770728779
		[3,] 0,4665696 0,59706537 0,639889961 0,271278181
		[4,] 0,3858517 0,45301093 0,487250027 0,374591427
		[5,] 0,2343667 0,19642066 0,388909892 0,533886945
		[6,] 0,0000000 0,10531494 0,392318820 0,255278676
		[7,] 0,6466631 0,21312183 0,343262570 0,322589843
		[8,] 0,2397621 0,38046876 0,152832344 0,308694736
		[9,] 0,4533578 0,25868835 0,168562241 0,471309304
		[10,] 0,5222319 0,38324005 0,212600120 0,064942105
Bobot	w	[,1] [,2] [,3] [,4]
		[1,] 0,010 0,020 0,030 0,040
		[2,] 0,050 0,060 0,070 0,080
<i>Learning Rate</i>	a	0,6
Jumlah Iterasi	banyakT	3
Galat Maksimum	Galat	0,00001

Berdasarkan hasil perhitungan menggunakan program aplikasi yang telah dibuat, dapat dilihat perhitungan tiap iterasi dan dalam satu iterasi terlihat perubahan bobot untuk setiap objek serta dapat dilihat hasil dari pengelompokan. Bobot akhir untuk 10 objek pengamatan yang diperoleh dari hasil penerapan program aplikasi adalah

$$w = \begin{bmatrix} 0,00192 & 0,08893447 & 0,3227536 & 0,2139452 \\ 0,4476119 & 0,3425724 & 0,3068789 & 0,3213787 \end{bmatrix}$$

dan hasil *clustering* untuk kesepuluh objek pengamatan adalah

- Objek ke- 1 masuk ke klaster: 2
- Objek ke- 2 masuk ke klaster: 2
- Objek ke- 3 masuk ke klaster: 2
- Objek ke- 4 masuk ke klaster: 2
- Objek ke- 5 masuk ke klaster: 2
- Objek ke- 6 masuk ke klaster: 1
- Objek ke- 7 masuk ke klaster: 2
- Objek ke- 8 masuk ke klaster: 2
- Objek ke- 9 masuk ke klaster: 2
- Objek ke- 10 masuk ke klaster: 2

Pada penelitian ini, hasil *clustering* data IPM Provinsi Aceh Tahun 2013 akan dibandingkan dengan hasil penelitian Okit Tazkiyah [10] yang menggunakan *package R* di mana *neighborhood* dan *learning function* yang digunakan adalah *gaussian* dan *inverse*.

Tabel 3. 5 Hasil Clustering Perbandingan Program Aplikasi yang telah Dibuat dengan yang Sudah Ada

No	Kaster	Hasil <i>Clustering</i> dengan Menggunakan Pogram yang Telah Dibuat	Hasil <i>Clustering</i> dengan Menggunakan Pogram yang Sudah Ada
1	Klaster 1	Aceh, Aceh Singkil, Aceh Tengah , Pidie, Aceh Barat Daya, Nagan Raya, Subulussalam, Pidie Jaya, Sabang , dan Lhokseumawe	Aceh Tengah , Aceh Tenggara, Aceh Timur, Aceh Barat, Aceh Besar, Bireuen, Aceh Utara, Aceh Tamiang, Aceh Jaya, Bener Meriah, dan Langsa
2	Klaster 2	Gayo Luwes dan Banda Aceh	Sabang , Lhokseumawe , dan Banda Aceh
3	Klaster 3	Simeulue	Aceh, Aceh Singkil, Aceh Selatan , Pidie, Gayo Lues , Aceh Barat Daya, Nagan Raya, Subulussalam dan Pidie Jaya
4	Klaster 4	Aceh Selatan , Aceh Tenggara, Aceh Timur, Aceh Barat, Aceh Besar, Bireuen, Aceh Utara, Aceh Tamiang, Aceh Jaya, Bener Meriah, dan Langsa	Simelue

Dapat dilihat untuk jumlah anggota tiap kluster tidak terlalu jauh berbeda. Sedangkan untuk keanggotaannya, terdapat wilayah yang dikelompokkan ke dalam kluster yang berbeda, antara dengan menggunakan program aplikasi yang telah dibuat dengan program yang sudah ada, yaitu wilayah Aceh Tengah, Sabang, Lhokseumawe, Gayo Lues, dan Aceh Selatan. Perbedaan tersebut kemungkinan dikarenakan pemilihan *neighborhood function*, fungsi *learning rate*, dan inialisasi awal yang berbeda sehingga akan menghasilkan algoritma yang berbeda dan akan berdampak pada hasil *clustering*.

4. KESIMPULAN

Hasil penerapan *clustering* dengan metode *self organizing map* dengan bantuan bahasa pemrograman R untuk data Indeks Pembangunan Manusia Provinsi Aceh tahun 2013 secara lengkap dapat dilihat pada skripsi penulis. Berdasarkan hasil penerapan dari program aplikasi diperoleh 4 kluster sebagai berikut :

1. Kluster pertama terdiri dari wilayah Aceh, Aceh Singkil, Aceh Tengah, Pidie, Aceh Barat Daya, Nagan Raya, Subulussalam, Pidie Jaya, Sabang dan Lhokseumawe. Kluster pertama merupakan wilayah-wilayah dengan Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Rata-Rata Lama Sekolah (RLS), dan Pengeluaran Perkapita (PP) menengah di antara kluster lainnya.
2. Kluster kedua terdiri dari wilayah Gayo Lues dan Banda Aceh. Kluster kedua merupakan wilayah-wilayah dengan Angka Melek Huruf (AMH) lebih kecil dari pada kluster lainnya serta angka Rata-Rata Lama Sekolah (RLS) lebih tinggi dari pada kluster lainnya.
3. Kluster ketiga terdiri dari wilayah Simeulue yang merupakan wilayah dengan Angka Harapan Hidup (AHH) dan angka Rata-Rata Lama Sekolah (RLS) lebih kecil dari pada kluster lainnya serta Angka Melek Huruf (AMH) dan Pengeluaran Perkapita (PP) lebih tinggi dari pada kluster lainnya.
4. Kluster Keempat terdiri dari wilayah Aceh Selatan, Aceh Tenggara, Aceh Timur, Aceh Barat, Aceh Besar, Bireueun, Aceh Utara, Aceh Tamiang, Aceh Jaya, Bener Meriah, dan Langsa. Kluster keempat merupakan wilayah dengan Angka Harapan Hidup (AHH) lebih tinggi dari pada kluster lainnya serta Pengeluaran Perkapita (PP) lebih rendah dari pada kluster lainnya.

DAFTAR PUSTAKA

- [1] Anis, Y. & Isnanto, R. (2014). Penerapan metode Self-Organizing Map (SOM) untuk visualisasi data geospasial pada informasi sebaran Data Pemilih Tetap (DPT). *Jurnal Sistem Informasi Bisnis*, 48-57. Doi : 10.21456/vol4iss1pp48-57.
- [2] Badan Pusat Statistik Provinsi Aceh (BPS). [Online]. Tersedia di : <http://www.aceh.bps.go.id/> [20 April 2018]
- [3] Budhi, S, G. Cluster Analysis untuk meprediksi talenta pemain basket menggunakan jaringan safaf tiruan Self Organizing Maps (SOM). *Jurnal Informatika*, 9(1), 23-32. Doi : <https://doi.org/10.9744/informatika.9.1.23-32>
- [4] Fausett, L. (1994). *Fundamentals of Neural Networks Architectures, Algorithms, and Applications*. New Jersey : Prentice-Hall, Inc.
- [5] Larose, T. & Larose, D. (2014). *Discovering Knowledge In Data An Introduction To Data Mining*. New Jersey : John Wiley & Sons, Inc.
- [6] Lobo, S,A. (2009). Application of Self-Organizing Maps to the Maritime Environment. *Information fusion and geographic information systems*, 19-36.
- [7] Rachmatin, D. (2014). Aplikasi Metode-Metode Agglomerative dalam Analisis Klaster pada Data Tingkat Populasi Udara. *Jurnal Ilmiah Program Studi Matematika STKIP Siliwangi Bandung*, 3(2), 133-149.
- [8] Suryaningsih, V. (2015). *Clustering* Dokumen Menggunakan Algoritma *Self Organizing Maps* (SOM). [Online]. Tersedia di : <https://digilib.uns.ac.id/dokumen/download/181498/MTgxNDk4> [28 Februari 2018].
- [9] Syaripudin, U. dkk. (2013). Studi Komparatif Penerapan Metode Hierarchical, K-Means dan Self Organizing Maps (SOM) Clustering pada Basis Data. *Jurnal Kajian Islam, Sains dan Teknologi*, 7(1), 132-149.
- [10] Tazkiyah, O. (2016). *Self Organizing Map (SOM) Clustering Berbasis WEB Interaktif Dengan R-Shiny*.
- [11] Wiharto, M. (2013). Analisis Kluster Menggunakan Bahasa Pemrograman R untuk Kajian Ekologi. *Jurnal Bionature*, 14(2), 73-79.