

Perbandingan Hasil Pengelompokan Menggunakan Metode Self Organizing Maps dan Metode Average Linkage (Studi Kasus: Data PDRB tiap provinsi di Indonesia)

Venn Yan Ishak Ilwaru¹, Gabriella Haumahu^{2*},

^{1,2}Jurusan Matematika, Fakultas MIPA, Universitas Pattimura, Ambon-Indonesia.

Email: gbhaumahu@gmail.com

Manuscript submitted : Oktober 2020.

Accepted for publication : Desember 2020.

Abstract: Gross domestic regional product (GDRP) is one of important indicators that can be used to determine economic conditions in some areas. This makes GDRP is very interesting to be studied. In this paper we cluster 33 provinces in Indonesia based on GDRP in 2013. We use self organizing and average linkage method of cluster analysis to investigate it. We also compare the results of both methods. We found that Average linkage method has better performance than self organizing method because it has smaller ratio.

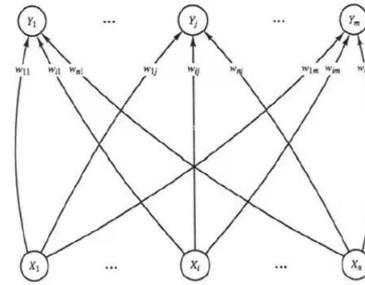
2010 Mathematical Subject Classification : 91C20, 92B20.

Keywords: average linkage method, GDRP, self organizing method.

1. Pendahuluan

Istilah analisis cluster pertama kali digunakan oleh Tryon pada tahun 1939. Analisis cluster meliputi sejumlah algoritma dan metode yang berbeda untuk mengelompokkan objek serupa ke dalam masing-masing kategori. Analisis cluster adalah teknik yang digunakan untuk mengidentifikasi objek atau individu yang serupa dengan memperhatikan beberapa kriteria [2]. Analisis cluster merupakan analisis statistik yang digunakan untuk mengelompokkan n objek ke dalam k kelompok, dengan setiap objek dalam kelompok memiliki keragaman yang besar dibandingkan antar kelompok [1].

Jaringan SOM menggunakan metode pembelajaran. Jaringan ini pertama kali di perkenalkan oleh Prof Teuvo kohonen pada tahun 1982. Jaringan *Self Organizing Maps* dianggap sebagai salah satu jaringan saraf yang disukai untuk analisis pengelompokan. Titik kunci dari algoritma SOM adalah dengan perhitungan geometrik. Jaringan Kohonen SOM (*Self-Organizing Maps*) merupakan salah satu model jaringan syaraf yang menggunakan metode pembelajaran *unsupervised*. Jaringan Kohonen 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 dari input yang diberikan.



Gambar 1. Struktur JST Kohonen

Beberapa langkah dalam penerapan algoritma sebagai berikut:

- Langkah 0 : Inisialisasi pembobotan w_{ij} dengan nilai random. Menentukan parameter laju pelatihan (α), dan radius tetangga (R)
- Langkah 1 : apabila kondisi selesai belum terpenuhi, lakukan langkah 2-8
- Langkah 2 : untuk tiap vektor input x ($x_i, i = 1, 2, \dots, n$) lakukan langkah 3-5
- Langkah 3 : untuk tiap j ($j = 1, 2, \dots, m$), hitung jarak Euclidean
- Langkah 4 : mencari indeks j dengan jarak $D(j)$ terdekat (minimum)
- Langkah 5 : Melakukan perbaikan nilai w_{ij} dengan nilai tertentu.
yaitu: $w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha [x_i - w_{ij}(\text{old})]$
- Langkah 6 : melakukan *update learning rate* $\alpha(t+1) = 0.5 \alpha t$
- Langkah 7 : Mereduksi radius dari fungsi tetangga pada waktu tertentu (epoch)
- Langkah 8 : Menentukan kondisi STOP.

Di sisi lain, kita juga mengenal analisis cluster metode hirarki, Metode hirarki dimulai dengan pengelompokan dua atau lebih objek yang mempunyai kesamaan yang paling dekat. Kemudian proses di teruskan ke objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sampai cluster akan membentuk semacam "pohon" hirarki (tingkatan yang jelas antar objek dari yang paling mirip sampai yang paling tidak mirip. Dendogram biasanya digunakan untuk membantu memperjelas proses hirarki tersebut. Metode-metode pengelompokan hirarki dibedakan berdasarkan konsep jarak antar kelompok, penentuan jarak, antar kelompok untuk metode-metode tersebut adalah:

1. Single-linkage (pautan tunggal), metode dengan prinsip jarak minimum.
2. Complete linkage (pautan lengkap), metode dengan prinsip jarak maksimum.
3. Average linkage (pautan rata-rata), metode dengan prinsip jarak rata-rata
4. Ward's method, metode ini menggunakan perhitungan yang lengkap dan memaksimalkan homogenitas di dalam satu kelompok.
5. Centroid method (metode titik pusat), metode yang menggunakan rata-rata jarak pada sebuah kelompok yang diperoleh dengan cara menghitung rata-rata pada setiap variabel untuk semua objek.

Metode yang akan digunakan dalam penelitian ini adalah metode average linkage dan SOM. Secara khusus dua metode ini akan digunakan dalam mengklaster 33 provinsi di Indonesia berdasarkan produk domestic regional bruto (PDRB). PDRB pada dasarnya merupakan jumlah nilai tambah yang dihasilkan oleh seluruh unit usaha dalam suatu daerah tertentu, atau merupakan jumlah nilai barang dan jasa akhir yang dihasilkan oleh seluruh unit ekonomi pada suatu daerah. PDRB atas dasar harga konstan menunjukkan nilai tambah barang dan jasa tersebut yang dihitung menggunakan harga yang berlaku pada satu tahun tertentu sebagai tahun dasar. PDRB konstan digunakan untuk mengetahui pertumbuhan ekonomi secara riil dari tahun ke tahun atau pertumbuhan ekonomi yang tidak dipengaruhi oleh faktor harga.

2. Hasil

Penelitian ini menggunakan data yang diperoleh dari BPS (Badan Pusat Statistik) Kota Ambon yaitu data tentang Produk Domestik Regional Bruto (PDRB). Data ini terdiri dari 7 variabel yaitu X_1 (pengeluaran konsumsi rumah tangga), X_2 (pengeluaran konsumsi lembaga swasta nirlaba yang mewakili rumah tangga), X_3 (pengeluaran konsumsi pemerintah), X_4 (pembentukan modal tetap bruto), X_5 (perubahan persediaan), X_6 (ekspor barang dan jasa), dan X_7 (dikurangi impor barang dan jasa). Sedangkan yang menjadi objeknya yaitu 33 Provinsi di Indonesia.

2.1. Proses Pengelompokan Menggunakan Metode *Self Organizing Maps* (SOM)

Menggunakan metode *Self Organizing Maps* (SOM) pada *software* Matlab adalah untuk mengelompokkan 33 provinsi berdasarkan data indikator Produk Domestik Regional Bruto. Adapun pengelompokan dilakukan dengan menentukan jumlah *cluster* sama seperti pada metode *Average Linkage* [3].

Berikut ini merupakan parameter yang digunakan dalam jaringan SOM yang digunakan untuk mengelompokkan 33 provinsi berdasarkan indikator PDRB.

Inisialisasi bobot w_{ij}	=	$2,5 \times 10^8$
<i>Learning rate</i> (α)	=	0,6
Radius tetangga (R)	=	0
<i>Learning rate</i> fase update	=	0,01

1. Pengelompokan untuk 4 *cluster* dengan 1000 iterasi. Formula untuk pengelompokan dengan 4 *cluster* disajikan pada Lampiran 2 dan hasilnya seperti yang ditampilkan pada Tabel 1 berikut ini.

Table 1. Pengelompokkan Provinsi ke Dalam 4 Cluster

Pengelompokan	Anggota
Cluster 1	DKI Jakarta, Jawa Barat, Jawa Timur
Cluster 2	Sumatra Barat, Jawa Tengah, Banten, Kalimantan Timur
Cluster 3	Riau, Sumatra Selatan, Lampung, Kepulauan Riau, Bali, Sulawesi Selatan
Cluster 4	Aceh, Sumatra Utara, Jambi, Bengkulu, Kepulauan Bangka Belitung, DI Yogyakarta, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua

2. Pengelompokan untuk 3 *cluster* dengan 1000 iterasi. Hasil pengelompokan dengan 3 *cluster* ditampilkan pada Tabel 2 berikut ini.

Tabel 2. Pengelompokkan Provinsi ke Dalam 3 Cluster

Pengelompokan	Anggota
Cluster 1	DKI Jakarta, Jawa Barat, Jawa Timur
Cluster 2	Sumatra Barat, Riau, Jawa Tengah, Banten, Kalimantan Timur

Cluster 3	Aceh, Sumatra Utara, Jambi, Sumatra Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua
-----------	--

3. Pengelompokan untuk 2 cluster dengan 1000 iterasi. Hasil pengelompokan dengan 2 cluster ditampilkan pada Tabel 3 berikut ini.

Tabel 3. Pengelompokan Provinsi ke Dalam 2 Cluster

Pengelompokan	Anggota
Cluster 1	DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur
Cluster 2	Aceh, Sumatra Barat, Sumatra Utara, Jambi, Riau, Sumatra Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua

2.2. Pemilihan Metode Pengelompokan Terbaik

Untuk mengetahui metode mana yang mempunyai kinerja terbaik, maka digunakan kriteria dua nilai simpangan baku, yaitu rata-rata simpangan baku-baku dalam cluster (S_w) dan simpangan baku antar cluster (S_B). Metode terbaik mempunyai nilai rasio antara simpangan baku-baku dalam cluster dengan simpangan baku antar cluster yang terkecil. Semakin kecil nilai S_w dan semakin besar nilai S_B maka akan memiliki kinerja yang baik, artinya mempunyai homogenitas atau kesamaan yang tinggi antar anggota dalam satu cluster (Bunkers dkk, 1996).

Tabel 4. Pengelompokan Provinsi Dengan Metode SOM

No	Propinsi	2 Cluster	3 Cluster	4 Cluster
1	Aceh	2	3	4
2	Sumatera Barat	2	2	2
3	Sumatera Utara	2	3	4
4	Riau	2	2	3
5	Jambi	2	3	4
6	Sumatera Selatan	2	3	3
7	Bengkulu	2	3	4
8	Lampung	2	3	3
9	Kepulauan Bangka Belitung	2	3	4
10	Kepulauan Riau	2	3	3
11	DKI Jakarta	1	1	1

12	Jawa Barat	1	1	1
13	Jawa Tengah	1	2	2
14	DI Yogyakarta	2	3	4
15	Jawa Timur	1	1	1
16	Banten	2	2	2
17	Bali	2	3	3
18	Nusa Tenggara Barat	2	3	4
19	Nusa Tenggara Timur	2	3	4
20	Kalimantan Barat	2	3	4
21	Kalimantan Tengah	2	3	4
22	Kalimantan Selatan	2	3	4
23	Kalimantan Timur	2	2	2
24	Sulawesi Utara	2	3	4
25	Sulawesi Tengah	2	3	4
26	Sulawesi Selatan	2	3	3
27	Sulawesi Tenggara	2	3	4
28	Gorontalo	2	3	4
29	Sulawesi Barat	2	3	4
30	Maluku	2	3	4
31	Maluku Utara	2	3	4
32	Papua Barat	2	3	4
33	Papua	2	3	4

Berdasarkan hasil penelitian yang dilakukan oleh Lessil [3] yakni pengelompokkan 33 provinsi di Indonesia berdasarkan indikator PDRB dengan menggunakan metode hirarki *Average Linkage* diperoleh hasil pengelompokkan seperti yang disajikan pada Tabel 5 berikut.

Tabel 5. Pengelompokkan Provinsi Dengan Metode *Average Linkage*

No	Propinsi	2 Cluster	3 Cluster	4 Cluster
1	Aceh	2	3	4
2	Sumatera Barat	2	3	3
3	Sumatera Utara	2	3	4
4	Riau	2	3	4
5	Jambi	2	3	4
6	Sumatera Selatan	2	3	4
7	Bengkulu	2	3	4
8	Lampung	2	3	4
9	Kepulauan Bangka Belitung	2	3	4
10	Kepulauan Riau	2	3	4
11	DKI Jakarta	1	1	1
12	Jawa Barat	1	2	2
13	Jawa Tengah	2	3	3
14	DI Yogyakarta	2	3	4

15	JawaTimur	1	2	2
16	Banten	2	3	3
17	Bali	2	3	4
18	Nusa Tenggara Barat	2	3	4
19	Nusa Tenggara Timur	2	3	4
20	Kalimantan Barat	2	3	4
21	Kalimantan Tengah	2	3	4
22	Kalimantan Selatan	2	3	4
23	Kalimantan Timur	2	3	4
24	Sulawesi Utara	2	3	4
25	Sulawesi Tengah	2	3	4
26	Sulawesi Selatan	2	3	4
27	Sulawesi Tenggara	2	3	4
28	Gorontalo	2	3	4
29	Sulawesi Barat	2	3	4
30	Maluku	2	3	4
31	Maluku Utara	2	3	4
32	Papua Barat	2	3	4
33	Papua	2	3	4

Tabel 6. Pemilihan Metode Pengelompokan Terbaik

Cluster	Rasio		Metode Terbaik
	Average Linkage	SOM	
4 Cluster	0,1258	0,2855	Average Linkage
3 Cluster	0,1210	0,2027	Average Linkage
2 Cluster	0,3001	0,3700	Average Linkage

Berdasarkan nilai rasio simpangan baku dalam *cluster* dan simpangan baku antar *cluster* seperti yang disajikan dalam Tabel 4.6 menunjukkan bahwa untuk semua pengelompokan 33 provinsi di Indonesia berdasarkan indikator PDRB baik pengelompokan menjadi 2 *cluster*, 3 *cluster* maupun 4 *cluster* terlihat bahwa metode *Average Linkage* memiliki kinerja lebih baik dari pada metode SOM karena memiliki nilai rasio yang lebih kecil yaitu 0,3001; 0,1210 dan 0,1258, dibandingkan dengan nilai rasio SOM yaitu 0,3700; 0,2027 dan 0,2855.

3. Kesimpulan

Nilai rasio simpangan baku dalam *cluster* dan simpangan baku antar *cluster* menunjukkan bahwa untuk semua pengelompokan 33 Provinsi di Indonesia berdasarkan indikator PBRB baik pengelompokan menjadi 2 *cluster*, 3 *cluster* maupun 4 *cluster* diperoleh bahwa metode *Average Linkage* memiliki kinerja lebih baik dari pada metode *Self Organizing Maps* (SOM) karena memiliki nilai rasio yang lebih kecil yaitu: 0,3001; 0,1210 dan 0,1258, dibandingkan dengan nilai rasio SOM yaitu 0,3700; 0,2027 dan 0,2855.

References

(Book)

- [1] D. Gil, M. Johnsson, J.M.G. Chamizo, P.A. Soriano, & F. D. Ruiz, (2009). Application of artificial neural networks in the diagnosis of urological dysfunctions. *Expert Systems with Applications*, 36(3), 5754-5760.
- [2] M. Kuncoro, (2013). *Metode Riset Untuk Ekonomi Bisnis*. Penerbit Erlangga, Jakarta
- [3] P. Lessil, (2015). *Pengelompokan Provinsi di Indonesia berdasarkan Produk Domestik Regional Bruto (PDRB) atas dasar harga konstan dengan menggunakan analisis Cluster*.