



Implementation of Self Organizing Map Algorithm to Identify the Grouping Patterns of Family Welfare Level in Siak Regency

Implementasi Algoritma Self Organizing Map untuk Identifikasi Pola Pengelompokan Tingkat Kesejahteraan Keluarga Kabupaten Siak

Syarfi Aziz^{1*}, Mustakim²

¹Amik Tri Dharma Pekanbaru

²Puzzle Research Data Technology (Predatech) Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

E-Mail: ¹syarfi.aziz@gmail.com, ²mustakim@uin-suska.ac.id

Makalah: Diterima 30 Agustus 2022; Diperbaiki 5 September 2022; Disetujui 11 September 2022
Corresponding Author: Syarfi Aziz

Abstrak

Dinas Sosial Kabupaten Siak dalam penentuan kelompok kesejahteraan keluarga menggunakan aspek kesehatan dan pendidikan, serta tidak mementingkan aspek ekonomi. Perlunya dilakukan pengelompokan untuk ketiga aspek tersebut dalam menentukan kesejahteraan keluarga. Penelitian ini dimaksudkan untuk melakukan *clustering* data kesejahteraan keluarga menggunakan algoritma SOM. Penelitian melakukan 36 percobaan menggunakan data kesejahteraan tahun 2016 sebanyak 22.047 data. Parameter SOM *cluster* 3, 4 dan 5, learning rate 0,05; 0,10; 0,15 dan 0,20 dengan iterasi 500, 750 dan 1.000. Hasil Validasi menggunakan DBI menunjukkan *cluster* 3, learning rate 0,20 dan iterasi 500 pada percobaan ke-10 dengan nilai DBI 0,9398 merupakan *cluster* optimal dengan 2 *cluster* dan 1 outlier. Hasil pada *cluster* 0 dari aspek kesehatan sebanyak 25,22% tidak memiliki cacat, dari aspek pendidikan sebanyak 20,92% tidak bersekolah lagi, dari aspek ekonomi sebanyak 16,59% tidak memiliki pekerjaan utama. Kemudian pada *cluster* 2 dari aspek kesehatan sebanyak 72,89% tidak memiliki cacat, dari aspek pendidikan sebanyak 28,49% memiliki ijazah SD/ sederajat dan aspek ekonomi sebanyak 33,51% pekerjaan utama adalah berkebun. Hasil dari ketiga aspek tersebut dijadikan sebagai rekomendasi untuk Dinas Sosial Kabupaten Siak agar lebih optimal mengelompokkan kesejahteraan keluarga dalam memberikan bantuan.

Kata Kunci: *Cluster*, DBI, Kabupaten Siak, Kesejahteraan Keluarga, SOM

Abstract

Siak District Social Service in determining family welfare groups using health and education aspects, and not concerned with economic aspects. The need for grouping for these three aspects in determining family welfare. This research is intended to conduct clustering of family welfare data using SOM algorithm. The study conducted 36 trials using data welfare year 2016 much 22.047 data. SOM cluster parameters 3, 4 and 5, learning rate 0,05; 0,10; 0,15 and 0.20 with 500, 750 and 1,000 iterations. Validation results using DBI showed cluster 3, learning rate 0,20 and 500 iteration in the tenth experiment with a value of DBI 0.9398 is an optimal cluster with 2 clusters and 1 outlier. The result of cluster 0 from health aspect much 25,22% has no defect, from education aspect much 20,92% no schooling, from economic aspect much 16,59% do not have main job. Then in cluster 2 from the health aspect much 72,89% have no defect, from education aspect much 28,49% have certificate / equal and economic aspect 33,51% main job is gardening. The results of these three aspects are used as recommendations for the Social Service of Siak District to more optimally classify the welfare of families in providing assistance.

Keyword: *Cluster*, DBI, Family Welfare, SOM, Siak District

1. Pendahuluan

Kesejahteraan adalah rasa tenang rakyat karena terpenuhinya hajat hidup lahir dan batin. Kesejahteraan lahir didasarkan pada standar universal menyangkut kesehatan, sandang, pangan, dan papan. Kesejahteraan batin menyangkut persepsi yang bersifat intelektual, emosional dan spiritual[1]. Kesejahteraan keluarga adalah suatu kondisi yang harus diciptakan oleh keluarga dalam membentuk keluarga yang sejahtera. Keluarga sejahtera merupakan model yang dihasilkan dari usaha kesejahteraan keluarga[2].

Berdasarkan data Badan Pusat Statistik (BPS) Kota Pekanbaru dalam website-nya menyebutkan bahwa jumlah penduduk di Kabupaten Siak pada tahun 2016 mencapai 453.052 jiwa dengan jumlah penduduk kategori sejahtera Berdasarkan data Tim Nasional Program Penanganan Kemiskinan (TNP2K) dalam website-nya juga menyebutkan bahwa jumlah penduduk sejahtera di Kabupten Siak pada tahun 2016 mencapai 87.059 jiwa dengan 22.050 jiwa Kartu Keluarga. Jumlah penduduk tersebut juga Berdasarkan Kepmensos Nomor 57/HUK/2017. Tidak termasuk keluarga Program Keluarga Harapan (PKH) yang belum memiliki status kesejahteraan di kabupaten Siak.

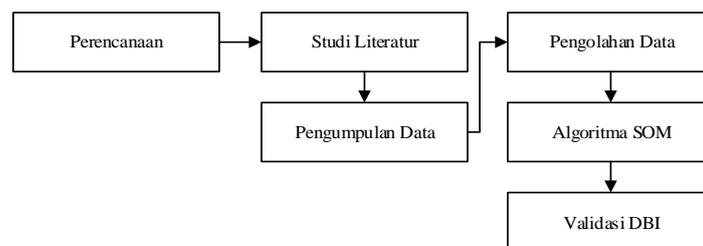
Secara Keilmuan, Dinas Sosial Kabupaten Siak kurang tepat dalam menentukan pola kesejahteraan keluarga, dilihat dari aturan penentuan kesejahteraan menurut Cahyat terdiri dari aspek kesehatan, aspek pendidikan dan aspek ekonomi. Namun pada kenyataannya aspek ekonomi tidak diperhitungkan oleh pemertintah setempat. Mereka menentukan tingkat kesejahteraan berdasarkan musyawarah dengan memnentukan kesejahteraan berdasarkan aspek kesehatan dan pendidikan. Oleh karena itu, perlunya pengelompokan dan pemodelan analisis *cluster* untuk mencari format baku dalam penentuan kesejahteraan keluarga di Kabupaten Siak.

Dalam data mining terdapat suatu cara untuk *cluster* data yaitu teknik *Clustering*. Prinsip dari *clustering* ini proses pengelompokan sekumpulan objek data ke dalam beberapa kelompok atau *cluster* sehingga objek dalam sebuah *cluster* memiliki kemiripan yang tinggi, tetapi sangat berbeda dengan objek dalam *cluster* lain[3]. Salah satu algoritma Data Mining dengan metode *clustering* adalah Self Organizing Map (SOM).

Berdasarkan uraian di atas dan didukung oleh beberapa penelitian sebelumnya serta acuan dasar penetapan kesejahteraan yang dikemukakan oleh Cahyat, maka pada penelitian ini akan mengaplikasikan algoritma SOM untuk mendapatkan sebuah pola tingkat kesejahteraan keluarga dari aspek Pendidikan, Kesehatan dan Ekonomi yang dapat digunakan sebagai pengetahuan yang bermanfaat oleh Dinas Sosial Kabupaten Siak.

2. Bahan dan Metode

Metodologi yang diterapkan dalam melakukan penelitian ini ditunjukkan seperti pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1 Data Mining

Data mining merupakan proses menemukan sebuah pola yang menarik dan pengetahuan dari sejumlah besar data. Sumber data dapat mencakup *database*, *data warehouse*, *web*, repositori informasi lainnya atau data yang dialirkan ke dalam sistem dinamis[3].

2.2 Pengelompokan (*cluster*)

Pengkluteraan merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas obyek-obyek yang memiliki kemiripan[4].

2.3 Self Organizing Map (SOM)

Algoritma SOM atau yang sering disebut dengan Jaringan Syaraf Tiruan Kohonen merupakan suatu metode jaringan syaraf tiruan yang diperkenalkan oleh Professor Teuvo Kohonen pada tahun 1981. Jaringan Kohonen/SOM digunakan untuk mengelompokkan (*clustering*) data berdasarkan karakteristik/fitur-fitur data. Jaringan Kohonen termasuk dalam pembelajaran tak terawasi (*unsupervised learning*)[5]. Adapun tahapan algoritma SOM sebagai berikut:

Sebelumnya data dinormalisasi terlebih dahulu dengan persamaan sebagai berikut:

$$\frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

1. Inisialisasi bobot.

Pada tahap ini menentukan secara acak bobot awal secara random sebagai w_{ij} . Inisialisasi nilai bobot juga dapat menggunakan nilai tengah (*middle point/mid point*) atau menggunakan nilai acak[6].

2. Repeat

a. Menentukan data

Pada algoritma tahap ini adalah menentukan data, selanjutnya

b. Menentukan *weight vector* dari obyek tersebut

Untuk setiap data terhadap bobot dihitung menggunakan *Euclidean Distance matrix*.

$$D_j = \sqrt{\sum_{i=1}^m (w_{ij} - x_i)^2} \quad (2)$$

c. Menentukan bobot terbaru

Dalam menentukan bobot terbaru pada waktu t , maka diasumsikan obyek saat ini $x(i)$ dan centroid yang terbentuk w_j . Kemudian untuk menentukan centroid yang baru untuk waktu berikutnya $t+1$

$$W_{ij}(\text{baru}) = W_{ij}(\text{lama}) + \alpha [x_i - W_{ij}(\text{lama})] \quad (3)$$

α adalah learning rate, tiap kenaikan epoch (iterasi) maka dilakukan pembaruan *learning rate*, dengan cara;

$$\text{Decrement } \alpha = \alpha \text{ awal} * \text{decrement} \quad (4)$$

$$\alpha \text{ baru} = \alpha \text{ awal} - \text{Decrement } \alpha \quad (5)$$

d. Sampai tidak ada perubahan *weight vector* atau threshold sudah terpenuhi.

e. Iterasi pada langkah ke-2 akan berhenti apabila threshold terpenuhi,

f. Menetapkan setiap obyek terhadap *weight vector* dan menentukan letak *Cluster* tersebut.

2.4 Davies Bouldin Index (DBI)

Davies Bouldin Indeks (DBI) merupakan salah satu metode validasi *cluster* untuk evaluasi kuantitatif dari hasil *clustering*. Pengukuran ini bertujuan memaksimalkan jarak *intercluster* antara satu *cluster* dengan *cluster* yang lain. Dalam penelitian ini DBI akan digunakan untuk mendeteksi outlier pada masing-masing *cluster* yang terbentuk[7]. Langkah-langkah perhitungan DBI sebagai berikut[8]:

1. Mengelompokan data berdasarkan jumlah *cluster*.

2. Menghitung rata-rata jarak S_i dalam setiap *cluster*

$$S_i = \left\{ \frac{1}{T_i} \sum_{j=i}^{T_i} |X_j - A_i|^q \right\}^{1/q} \quad (6)$$

Dimana T_i jumlah vector dalam *cluster* i

A_i adalah centroid *cluster* i

3. Menghitung jarak antar setiap centroid/ *weight vector*

$$M_{ij} = \left\{ \sum_{k=i}^N |a_{ki} - a_{kj}|^p \right\}^{1/p} \quad (7)$$

dimana a_{ki} merupakan komponen k dari n -dimensi vektor a_i , yang merupakan *centroid cluster* i .

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \quad (8)$$

$$\bar{R} = \frac{1}{N} \sum_{i=1}^N R_i \quad (9)$$

dimana $R_i = \text{maximum dari } R_{ij} \text{ } i \neq j$. *clustering* optimal berada pada DBI terendah.

2.5 Sistem Informasi

Sistem informasi merupakan seperangkat komponen saling berhubungan dan berintegrasi yang berfungsi memproses, mendistribusikan, serta menyimpan informasi guna mendukung keputusan dan pengawasan di dalam suatu organisasi[9].

2.6 Kesejahteraan

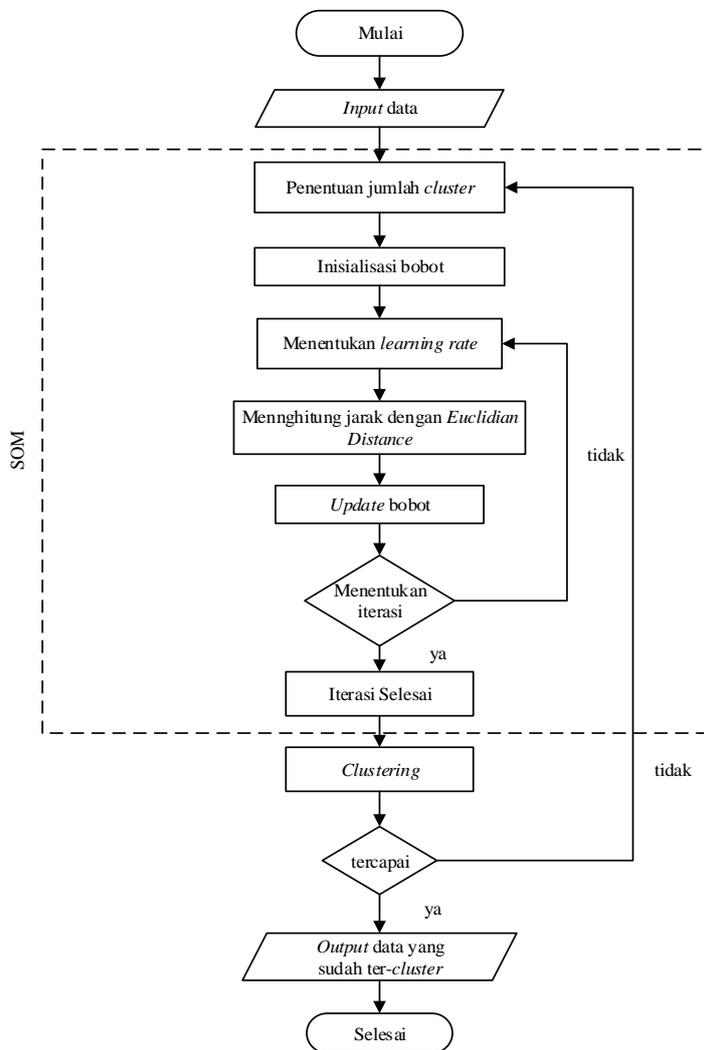
Kesejahteraan merupakan kondisi dapat memenuhi kebutuhan dasar yang bersifat material (kebendaan) maupun bukan material, yang menyangkut aspek gizi dan kesehatan (aspek kesehatan), pengetahuan (aspek pendidikan) dan kekeayaan materi (aspek ekonomi)[10].

2.7 Matlab

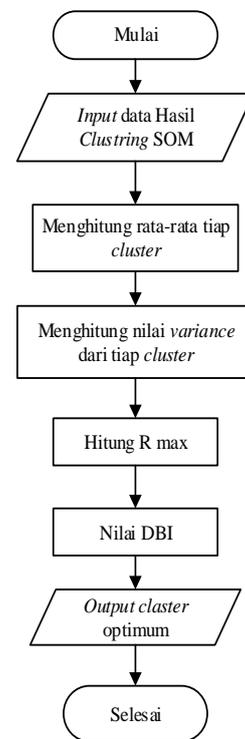
Matlab merupakan merupakan suatu bahasa pemrograman matematika lanjutan, yang dibentuk dengan dasar pemikiran yang menggunakan sifat dan bentuk matriks. Matlab hadir dengan fungsi dan karakteristik yang berbeda dengan bahasa pemrograman lain yang sudah ada lebih dahulu seperti Delphi, Basic maupun C++[11].

2.7 Model SOM dan DBI

Langkah-langkah dalam metode SOM dan DBI dapat dilihat pada Gambar 2 dan Gambar 3[12]:



Gambar 2. Alur Metode SOM



Gambar 3. Alur DBI

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Pengumpulan data adalah tahapan untuk mendapatkan data yang digunakan dalam melakukan penelitian setelah tahap perencanaan. Pengumpulan data penelitian ini diperoleh dari Dinas Sosial Kabupaten Siak sebanyak 22.047 record seperti pada Tabel 1.

3.2 Penentuan Atribut

Pemilihan kriteria atau atribut dalam melakukan proses *clustering* ditentukan berdasarkan data yang berhubungan dengan penelitian yang sesuai dengan teori Cahyat seperti aspek kesehatan, pendidikan dan ekonomi seperti pada Tabel 2.

Tabel 1. Data Kesejahteraan Tahun 2016

No	Nama	Jenis Cacat	Penyakit Kronis	Partisipasi Sekolah	Ijazah tertinggi	Bekerja/Membantu bekerja	...	Televisi
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	...	<i>16</i>
1	N-1	0	0	2	1	1	...	4
2	N-2	0	3	2	1	2	...	3
3	N-3	0	1	2	0	1	...	3
...
22.047	N-22047	0	0	2	0	1	...	4

Tabel 2. Penentuan Atribut

Nama Variabel	Simbol	Keterangan
Status kepemilikan bangunan tempat tinggal	SBT	0. Tidak Memiliki
		1. Milik sendiri
		2. Kontrak/sewa
		3. Bebas sewa
		4. Dinas
Status kepemilikan lahan tempat tinggal	SLTT	5. Lainnya
		0. Tidak Memiliki
		1. Milik sendiri
		2. Milik orang lain
		3. Tanah negara
Izazah tertinggi	IT	4. Lainnya
		0. Tidak punya ijazah
		1. SD/ sederajat
		2. SMP/ sederajat
		3. SMA/ sederajat
		4. D1/D2/D3
Emas/Perhiasan/Tabungan	MS	5. D4/S1
		6. S2/S3
		0. Tidak Memiliki
Lapangan usaha dari pekerjaan utama	LUPU	1. >=10 gr
		2. <10 gr
		0. Tidak Memiliki Lapangan Pekerjaan
		1. Pertanian (padi & palawija)
		2. Hortikultura
		3. Perkebunan
		4. Perikanan tangkap
		5. Perikanan budidaya
		6. Peternakan
		7. Kehutanan & pertanian lain
		8. Pertambangan/penggalian
		9. Industri pengolahan
		10. Listrik & gas
		11. Bangunan/konstruksi
		12. Perdagangan
		13. Hotel & rumah makan
		14. Transportasi & pergudangan
		15. Informasi & komunikasi
16. Keuangan & asuransi		
17. Jasa Pendidikan		
18. Jasa kesehatan		

19. Jasa kemasyarakatan, pemerintahan & perorangan
20. Pemulung
21. Lainnya

Tabel 2. Penentuan Atribut (Lanjutan)

Nama Variabel	Simbol	Keterangan
Mobil	MBL	Unit
Status kedudukan dari pekerjaan utama	SKPU	0. Tidak Memiliki Pekerjaan 1. Berusaha sendiri 2. Berusaha dibantu buruh tidak tetap/tidak dibayar 3. Berusaha dibantu buruh tetap/dibayar 4. Buruh/karyawan/pegawai swasta 5. PNS/TNI/Polri/BUMN/BUMD/Anggota legislative 6. Pekerja bebas pertanian 7. Pekerja bebas non-pertanian 8. Pekerja keluarga/tidak dibayar
Jenis cacat	JC	0. Tidak cacat 1. Tuna daksa/cacat tubuh 2. Tuna netra/buta 3. Tuna rungu 4. Tuna wicara 5. Tuna rungu & wicara 6. Tuna netra & cacat tubuh 7. Tuna netra, rungu & wicara 8. Tuna rungu, wicara & cacat tubuh 9. Tuna rungu, wicara, netra & cacat tubuh 10. Cacat mental retardasi 11. Mantan penderita gangguan jiwa 12. Cacat fisik & mental
Penyakit kronis/menahun	PK	0. Tidak ada 1. Hipertensi 2. Rematik 3. Asma 4. Masalah jantung 5. Diabetes 6. Tuberculosis 7. Stroke 8. Kanker/tumor 9. Lainnya (gagal ginjal, paru-paru, flek dan sejenisnya)
Bekerja/Membantu Bekerja	BMB	1. Bekerja 2. Membantu Bekerja
Sepeda Motor	MTR	Unit
Komputer	KOM	Unit
Televisi	TV	Unit
Partisipasi Sekolah	PS	0. Tidak / belum pernah sekolah 1. Masih sekolah 2. Tidak bersekolah lagi

3.3 Pembersihan Data

Pembersihan data dilakukan untuk mengurangi kerancuan dan noise yang dapat mempengaruhi perhitungan. Pembersihan data dilakukan dengan cara menghapus data yang tidak memiliki nilai pada setiap atribut. Data yang akan digunakan dalam penelitian ini sebanyak 22.047 record seperti pada Tabel 3.

Tabel 3. *Cleaning Data*

No	Nama	Jenis Cacat	Penyakit Kronis	Partisipasi Sekolah	Ijazah tertinggi	Bekerja/Membantu bekerja	...	Televisi
1	N-1	0	0	2	1	1	...	4
2	N-2	0	3	2	1	2	...	3
3	N-3	0	1	2	0	1	...	3
4	N-4	0	0	2	0	2	...	4
5	N-5	0	0	2	1	1	...	4
...
22.047	N-22047	0	0	2	0	1	...	4

3.4 Transformasi Data

Tahap selanjutnya adalah melakukan proses *cleaning* data kemudian dilakukan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Dalam penelitian ini tidak ada atribut yang diubah dari variabel menjadi angka numerik, karena data sudah ditransformasikan pada Tabel 1.

3.5 Normalisasi Data

Langkah selanjutnya yaitu melakukan normalisasi, tujuannya untuk menghasilkan nilai keseimbangan antara nilai yang rendah dengan nilai yang tinggi dapat dilihat pada Tabel 4.

Tabel 4. Normalisasi Data

Nama	JC	PK	PS	IT	BMB	LUPU	SKPU	SBT	...	TV
N-1	0,000	0,000	0,222	0,111	0,500	0,143	0,875	0,000	...	1,000
N-2	0,000	0,333	0,222	0,111	1,000	0,000	0,000	0,000	...	0,750
N-3	0,000	0,111	0,222	0,000	0,500	0,524	0,875	0,000	...	0,750
N-4	0,000	0,000	0,222	0,000	1,000	0,000	0,000	0,250	...	1,000
N-5	0,000	0,000	0,222	0,111	0,500	0,190	0,125	0,000	...	1,000
...
N-22047	0,000	0,000	0,222	0,000	0,500	0,143	0,500	0,000	...	1,000

3.6 Pengolahan Data

Penelitian ini akan di lakukan 36 kombinasi percobaan *clustering* menggunakan SOM. Percobaan dilakukan dengan menggunakan *cluster* 3, 4 dan 5, *learning rate* 0,05; 0,10; 0,15 dan 0,20 dengan iterasi 500, 750 dan 1000. Setelah mendapatkan hasil *cluster* pada tiap percobaan, selanjutnya dilakukan analisis pada hasil *cluster* untuk menemukan pola pada data kesejahteraan sehingga dapat diketahui kedalam kelompok kesejahteraan keluarga.

3.7 Perhitungan SOM

Adapun tahapan dalam algoritma SOM adalah sebagai berikut:

1. Inisialisasi

Tabel 5. Inisialisasi

1	Jumlah <i>cluster</i>	m	3
2	<i>Max</i> iterasi	Iterasi	1000
3	<i>Learning rate</i> awal	LR	0,05
4	Penurunan <i>learning rate</i>	<i>Decrament</i>	0,01

Pada penelitian ini, nilai *decrement* dipakai untuk semua jenis percobaan, yang merupakan parameter statis. Sedangkan *cluster*, *learning rate* dan *iterasi* pada penelitian ini akan menjadi parameter pada setiap percobaan yang bersifat dinamis.

2. Bobot awal (*weight vector*) = w_{ij}

Pada tahap ini menentukan bobot awal secara random sebagai w_{ij} dapat dilihat pada Tabel 6.

Tabel 6. Weight vector random

	D ₀	D ₁	D ₂	D ₃	D ₄	D ₅	D ₆	...	D ₁₃
w ₁	0,390	0,076	0,893	0,570	0,893	0,798	0,238	...	0,247
w ₂	0,001	0,013	0,211	0,160	0,651	0,328	0,313	...	0,798
w ₃	0,243	0,951	0,537	0,621	0,581	0,803	0,829	...	0,806

3. Repeat

Data yang digunakan pada Percobaan ini yaitu data kesejahteraan sebanyak 22047 record seperti pada Tabel 7.

Tabel 7. Data Simulasi

Nama	JC (D ₀)	PK (D ₁)	PS (D ₂)	IT (D ₃)	SB (D ₄)	LUPU (D ₅)	SKPU (D ₆)	...	TV (D ₁₃)
N-1	0,000	0,000	0,222	0,111	0,500	0,143	0,875	...	1,000
...
N-22047	0,000	0,000	0,222	0,000	0,500	0,143	0,500	...	1,000

Iterasi 1/ epoch 1

Observasi 1

Menghitung jarak antara data observasi ke-1 dengan wij menggunakan rumus *euclidian distance*.

$$O_1 W_1 = \sqrt{(0,000-0,390)^2+(0,000-0,076)^2+(0,222-0,893)^2+(0,111-0,570)^2+(0,500-0,893)^2+(0,143-0,798)^2+(0,875-0,238)^2+(0,000-0,029)^2+(0,250-0,706)^2+(1,000-0,019)^2+(1,000-0,855)^2+(1,000-0,010)^2+(1,000-0,732)^2+(1,000-0,247)^2}$$

$$=2,150$$

$$O_1 W_2 = \sqrt{(0,000-0,001)^2+(0,000-0,013)^2+(0,222-0,211)^2+(0,111-0,160)^2+(0,500-0,651)^2+(0,143-0,328)^2+(0,875-0,313)^2+(0,000-0,036)^2+(0,250-0,263)^2+(1,000-0,996)^2+(1,000-0,991)^2+(1,000-0,815)^2+(1,000-0,000)^2+(1,000-0,798)^2}$$

$$=0,673$$

$$O_1 W_3 = \sqrt{(0,000-0,243)^2+(0,000-0,951)^2+(0,222-0,537)^2+(0,111-0,621)^2+(0,500-0,581)^2+(0,143-0,803)^2+(0,875-0,829)^2+(0,000-0,638)^2+(0,250-0,376)^2+(1,000-0,772)^2+(1,000-0,926)^2+(1,000-0,382)^2+(1,000-0,552)^2+(1,000-0,806)^2}$$

$$=1,693$$

Update weight vector

w_{ij} pemenang diperoleh dari kuadrat jarak minimum *euclidian distance*, w_{ij} pemenang berasal dari perhitungan *euclidian distance* observasi pertama dengan w₂, maka bobot yang akan diperbarui adalah w₂.

$$w_2 D_0 = 0,001 + 0,05(0,000 - 0,001) = 0,001$$

$$w_2 D_1 = 0,013 + 0,05(0,000 - 0,013) = 0,012$$

$$w_2 D_2 = 0,211 + 0,05(0,222 - 0,211) = 0,211$$

$$w_2 D_3 = 0,160 + 0,05(0,111 - 0,160) = 0,157$$

$$w_2 D_4 = 0,651 + 0,05(0,500 - 0,651) = 0,643$$

$$w_2 D_5 = 0,328 + 0,05(0,143 - 0,328) = 0,319$$

$$w_2 D_6 = 0,313 + 0,05(0,875 - 0,313) = 0,341$$

$$w_2 D_7 = 0,036 + 0,05(0,000 - 0,036) = 0,034$$

$$w_2 D_8 = 0,263 + 0,05(0,250 - 0,263) = 0,262$$

$$w_2 D_9 = 0,996 + 0,05(1,000 - 0,996) = 0,996$$

$$w_2 D_{10} = 0,991 + 0,05(1,000 - 0,991) = 0,992$$

$$w_2 D_{11} = 0,815 + 0,05(1,000 - 0,815) = 0,824$$

$$w_2 D_{12} = 1,000 + 0,05(1,000 - 1,000) = 1,000$$

$$w_2 D_{13} = 0,798 + 0,05(1,000 - 0,798) = 0,808$$

Sehingga didapat *weight vector* baru yang akan digunakan untuk perhitungan jarak dengan data observasi ke dua dapat dilihat pada Tabel 8.

Tabel 8. Hasil update weight vector w₂O₁

	D ₀	D ₁	D ₂	D ₃	D ₄	D ₅	D ₆	...	D ₁₃
w ₁	0,390	0,076	0,893	0,570	0,893	0,798	0,238	...	0,247
w ₂	0,001	0,012	0,211	0,157	0,643	0,319	0,341	...	0,808
w ₃	0,243	0,951	0,537	0,621	0,581	0,803	0,829	...	0,806

Hasil *update weight vector* pada observasi ke-1 digunakan untuk menghitung jarak dengan data observasi ke-2 dan seterusnya. sehingga hasil minimum dari perhitungan jarak tersebut digunakan untuk *update bobot*. Hasil *update bobot* akan di gunakan untuk perhitungan jarak data observasi selanjutnya, jika *weight vector* pemenang berasal dari w_2 , maka bobot w_2 akan di perbarui, begitu seterusnya sampai maksimal batas iterasi, jika iterasi 1 sudah terpenuhi selanjutnya dilakukan pembaruan *learning rate* yang akan digunakan pada tahapan iterasi 2, dengan menggunakan Persamaan (4) = $0,05 * 0,01 = 0,0005$ dan Persamaan (5) = $0,05 - 0,0005 = 0,0495$. Sehingga didapatkan hasil *cluster* yang ada pada Tabel 9.

Tabel 9. Letak Cluster

Nama	JC	PK	PS	IT	BMB	...	W1	W2	W3	Min	Cluster
N-1	0,000	0,000	0,222	0,111	0,500	...	2,462	0,552	2,065	0,552	1
N-1	0,000	0,000	0,222	0,111	0,500	...	2,462	0,552	2,065	0,552	1
...
N-22047	0,000	0,000	0,222	0,000	0,500	...	2,354	0,387	2,065	0,387	1

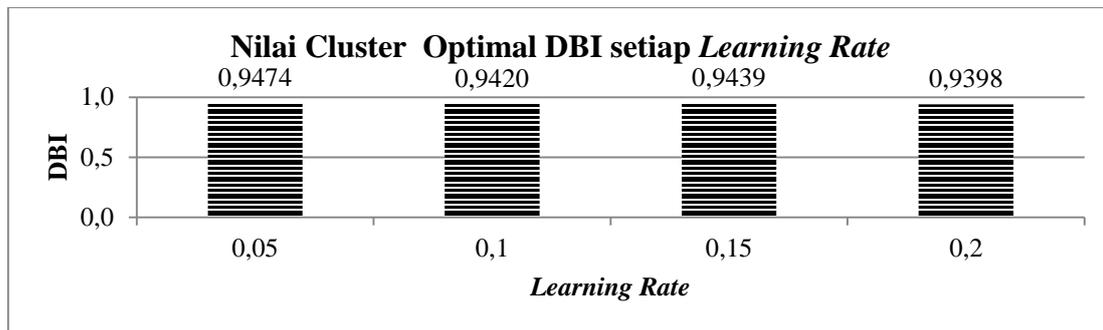
3.8 Hasil

Berdasarkan data yang diolah didapatkan hasil perhitungan nilai DBI pada Percobaan menggunakan aplikasi matlab. Berikut ini adalah hasil perhitungan nilai DBI pada Tabel 10.

Tabel 10. Hasil validasi DBI

Kombinasi	Cluster	iterasi	Learning Rate	Jumlah Anggota Cluster	DBI
1	3	500	0,05	Cluster 0: 1 Cluster 1: 16217 Cluster 2: 5829	0,9474
2		750	0,05	Cluster 0: 16215 Cluster 1: 1 Cluster 2: 5831	0,9499
3		1000	0,05	Cluster 0: 1 Cluster 1: 22045 Cluster 2: 1	0,9528
4		500	0,10	Cluster 0: 5845 Cluster 1: 16201 Cluster 2: 1	0,9420
5		750	0,10	Cluster 0: 1 Cluster 1: 16215 Cluster 2: 5831	0,9478
6		1000	0,10	Cluster 0: 16215 Cluster 1: 5831 Cluster 2: 1	0,9485
7		500	0,15	Cluster 0: 16195 Cluster 1: 5851 Cluster 2: 1	0,9439
8		750	0,15	Cluster 0: 5830 Cluster 1: 1 Cluster 2: 16216	0,9474
9		1000	0,15	Cluster 0: 16215 Cluster 1: 1 Cluster 2: 5831	0,9484
10		500	0,20	Cluster 0: 5827 Cluster 1: 1 Cluster 2: 16219	0,9398
...
36		1000	0,20	Cluster 0: 1 Cluster 1: 0 Cluster 2: 0 Cluster 3: 5831 Cluster 4: 16215	0,9477

Perbandingan nilai DBI yang mempunyai nilai optimal pada setiap percobaan kombinasi dengan *learning rate* 0,05; 0,10; 0,15; dan 0,20 dapat dilihat pada Gambar 4.



Gambar 4. Nilai *Cluster* Optimal pada Setiap Percobaan

Dari Gambar grafik di atas diperoleh *cluster* optimal berdasarkan setiap percobaan, pada kombinasi 10 terdapat pada *learning rate* 0,2; *iterasi* 500 dan *cluster* 3 dengan nilai DBI 0,9398.

3.9 Implementasi Sistem

Adapun Implementasi sistem pengelompokan dengan menggunakan algoritma SOM dapat dilihat pada Gambar 5 dan Gambar 6.



Gambar 5. *Input* Parameter SOM



Gambar 6. Hasil Analisis Perhitungan SOM

Gambar 5 adalah halaman proses untuk *input* parameter dari metode *Self Organizing Map*. Selanjutnya, Gambar 6 merupakan tampilan hasil analisis perhitungan SOM dari percobaan yang telah dilakukan, berisi *weigh vector*, lihat *cluster* terbaik dan grafik *cluster*.

4. Kesimpulan

Adapun kesimpulan dari penelitian ini adalah sebagai berikut:

1. Percobaan *clustering* yang dilakukan dengan kombinasi beberapa *cluster* (3, 4 dan 5), *Learning rate* (0,05; 0,10; 0,15 dan 0,20) dan *iterasi* (500, 750 dan 1000) menggunakan metode *Self Organizing Maps* (SOM), maka didapatkan *cluster* optimal berdasarkan validasi *Davies Bouldin Indeks* (DBI) diperoleh *cluster* optimal menggunakan 3 *cluster*, *Learning rate* 0,20 dan 500 *iterasi* dengan nilai DBI 0,9398.
2. Percobaan pengelompokan dengan beberapa *iterasi* Nilai rata-rata DBI paling tinggi berada pada *iterasi* 1000 dengan nilai DBI 0,9860. Kemudian dengan *iterasi* 500 dan nilai DBI paling rendah merupakan *iterasi* terbaik dengan nilai 0,9398.
3. Pengolahan dengan menggunakan percobaan dengan 36 kombinasi *cluster*, *learning* dan *iterasi*. *Cluster* yang terbentuk umumnya memiliki jumlah *cluster* dua dan tiga *cluster* selebihnya adalah data *outlier*. data *outlier* berada pada data ke 7.908 dan selebihnya terdapat pada data ke 9.729.
4. Persentase anggota *cluster* 0 dilihat dari aspek kesehatan didominasi tidak memiliki cacat. Selanjutnya dari aspek pendidikan banyak yang tidak bersekolah lagi dan dari aspek ekonomi didominasi dengan banyak tidak memiliki pekerjaan utama dan sebagian memiliki pekerjaan utama adalah berkebun.
5. Persentase anggota *cluster* 2 dilihat dari aspek kesehatan yang mendominasi tidak memiliki cacat sehingga banyak juga yang tidak memiliki penyakit kronis. Selanjutnya dari aspek pendidikan tidak banyak yang melanjutkan sekolah. Kemudian dari aspek ekonomi banyak yang bekerja, memiliki lapangan usaha dari pekerjaan utama terutama berkebun selanjutnya bekerja sebagai buruh/karyawan/pegawai swasta.

Referensi

- [1] G. Maeswara, *Biografi politik Susilo Bambang Yudhoyono*. Penerbit Narasi, 2009.
- [2] Astuti, S. Adyatma, and E. Normelani, "Pemetaan Tingkat Kesejahteraan Keluarga di Kecamatan Banjarmasin Selatan," *J. Pendidik. Geogr.*, vol. 4, no. 2, pp. 20–34, 2017, [Online]. Available: <http://ppjp.unlam.ac.id/journal/index.php/jpg>.
- [3] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. USA: Morgan Kaufmann, 2012.
- [4] Kursini and Emha T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi, 2009.
- [5] D. Setiani and R. F. Hakim, "CLUSTERING INDIKATOR PEMBANGUNAN BERKELANJUTAN DI INDONESIA MENGGUNAKAN ALGORITMA SELF-ORGANIZING MAPS (SOMs) KOHONEN DessySetiani dan RB.Fajriya Hakim," pp. 614–628.
- [6] Edward, "Clustering Menggunakan Self Organizing Maps (Studi Kasus: Data PPMB IPB)," 2006.
- [7] L. Rahmawati, A. D. Cahyani, and S. S. Putro, "Pemanfaatan Metode Cluster SOM – IDB sebagai Analisa Pengelompokan Penerima Beasiswa," *J. Sist. Inf. Indones.*, vol. 1, no. 1, pp. 11–17, 2015.
- [8] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.
- [9] E. Iswandy, "Sistem Penunjang Keputusan untuk Menentukan Penerimaan Dana Santunan Sosial Anak Nagari dan Penyalurannya Bagi Mahasiswa dan Pelajar Kurang Mampu di Kenagarian Barung – Barung Balantai Timur," *TEKNOIF*, vol. 10, 2015.
- [10] A. Cahyat, C. Gönner, and M. Haug, "Mngkaji Kemiskinan dan Kesejahteraa," Bogor: Center For International Forestry Research, 2007.
- [11] E. Prasetyo, "Data Mining Mengelola Data Menjadi Informasi Menggunakan Matlab," Yogyakarta: Andi, 2014.
- [12] S. D. Agustina, "Pengelompokan Jurnal Internasional Terindeks Scopus Menggunakan Self Organizing Maps Algorithm," 2017.