

Pengelompokan Tingkat Kesehatan Masyarakat Menggunakan Shelf Organizing Maps Dengan Cluster Validation Idb dan I-Dunn

Budi Dwi Satoto
Manajemen Informatika
Fakultas Teknik Universitas Trunojoyo Madura
Jl. Telang PO Box 2 Kamal, Bangkalan
Email : budids@yahoo.com

Bain Khusnul Khotimah, Adam Muhammad
Teknik Informatika
Fakultas Teknik Universitas Trunojoyo Madura
Jl. Telang PO Box 2 Kamal, Bangkalan
Email : Bainkk@yahoo.com

Abstrak—Data hasil RISKESDAS Jawa Timur merupakan data mentah yang belum diolah. Hal ini disebabkan karena data tersebut merupakan data riil atau data nyata yang didapatkan dari hasil survei lapangan secara langsung. Adapun Tujuan penelitian ini adalah Adanya gambaran daerah yang telah dikelompokkan mendapatkan perlakuan yang sama dari sisi infrastruktur dan pelayanan kesehatan. Pada proses awal penelitian ini dilakukan preprocessing dilanjutkan *Clustering* menggunakan SOM untuk menentukan *Cluster* data diikuti dengan proses *Validation* untuk meningkatkan akurasi hasil *cluster* menggunakan IDB dan I-DUNN. Hasil yang dicapai pada penelitian ini menggunakan skenario uji coba pada modul data penyakit mata dengan *record* kabupaten di Jawa Timur, dimana dihasilkan nilai IDB dengan hasil terbesar ada pada tiga *Cluster* dengan nilai 32,8657, dan nilai terkecil ada pada sepuluh *cluster* dengan Nilai 9,8597, sedangkan nilai I-DUNN nya dengan hasil terkecil pada tiga *Cluster* dengan nilai 0,6667 dan nilai terbesar terdapat pada sepuluh *Cluster* dengan nilai 0,9. Dapat disimpulkan bahwa pada metode *Cluster* SOM menggunakan *Cluster validation* IDB, semakin kecil nilai indeksnya maka hasil semakin bagus sedangkan jika menggunakan *Cluster validation* I-DUNN semakin besar nilai indeksnya maka semakin baik hasil yang di dapatkan.

Keywords; *Clustering, Self Organizing Map, IDB, Index-dunn, Cluster Validation*

I. PENDAHULUAN

RISKESDAS (Riset Kesehatan Dasar) merupakan sebuah penelitian kuantitatif yang dilakukan oleh badan litbang Kementerian Kesehatan secara berkala dalam kurun waktu 3 tahunan. Data tersebut memiliki dimensi jamak sehingga dapat digolongkan sebagai data multidimensional. Dengan fakta diatas dapat diasumsikan, bila data mengalami *Granularity* dalam OLAP terkait dengan dimensi tertentu dan hirarki sebagai sebuah kisi, maka fokusnya adalah *roll-up* dan *drill-down* menggunakan algoritma dengan kompleksitas waktu yang optimal. algoritma baru diperlukan menggunakan struktur data dinamis sehingga didapatkan pemanfaatan ruang yang lebih baik dan mengurangi waktu komputasi[1].

Penelitian ini bertujuan untuk memudahkan pengelompokan data RISKESDAS di propinsi Jawa Timur. Adapun Penelitian sebelumnya yang mendasari penelitian ini adalah penelitian mengenai pengolahan data dimensi tinggi (multidimensional data) yang dapat dianggap sebagai sebuah *sample function*. Proses fungsional ini terbukti berhasil untuk menganalisa data berbentuk *spectrometric*. Penanganan data yang memiliki kompleksitas tinggi dan memiliki tingkat kemiripan yang rendah tidak dapat dilakukan dengan hanya berbasis pada perhitungan *centroid* saja. Hal ini dikarenakan *input* yang diterima bukanlah sebuah vektor. Bentuk dari data yang juga dipertimbangkan apakah data tersebut terstruktur yaitu data dengan *internal struktur* seperti *intervals data, distributions, functional data*, dsb ataukah *semi-struktural data* berbentuk *trees, XML doc, SQL queries*, dsb.

Saat ini, terdapat banyak data yang perlu dianalisa menggunakan metode pengelompokan unik yang disebut *cluster*. Pengelompokan ini didasarkan pada tidak adanya pola atau deskripsi kelas yang jelas. Jadi, hasil yang diperoleh bersifat intern data. Salah satu cara untuk menentukan kualitas partisi dari pengelompokan adalah dengan menggunakan validasi indeks *cluster*. Tujuan dari indeks adalah memastikan mana data yang merupakan *outlier* dan data mana yang dipadatkan dalam suatu *cluster* [2].

Pemilihan fitur merupakan hal penting dalam *automatic classification*. Penggunaan SOM dalam hal ini tidak hanya mengurangi dimensi dari fitur yang ada, namun juga dapat meningkatkan ketepatan hasil klasifikasi. Dalam prakteknya, informasi eksternal seperti kelas label sering kali tidak tersedia di banyak skenario aplikasi. Oleh karena itu, dalam situasi tidak ada informasi eksternal yang tersedia, validasi internal yang adalah satu-satunya pilihan untuk memvalidasi *cluster*[3]. Dengan penggabungan metode tersebut jarak antar *cluster* yang didapat akan lebih besar dan lebih jelas. Penggunaan data yang besar dan algoritma iteratif maka analisis *cluster* memiliki kepekaan akan kebutuhan yang tinggi dalam komputasi.

II. KARYA TERKAIT

A. Preprocessing

Normalisasi adalah sebuah proses *preprocessing* yang dilakukan untuk membentuk ulang data-data yang ada dengan rentang 0 (Nol) hingga 1 (satu). Hal ini dilakukan dengan maksud agar data yang diproses memiliki rentang yang lebih sempit, namun masih tetap mengakomodir perbedaan dan nilai dari data tersebut.

$$\sum_{i=1}^n f(x) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Dengan :

$f(x)$ = Fungsi $f(x)$ dengan variable x untuk normalisasi

x_i = Nilai x dengan iterasi dari $i=1..n$

x_{\min} = Nilai x minimum

x_{\max} = Nilai x maksimum

Dengan demikian maka data yang memiliki nilai diantara maksimum dan minimum akan memiliki nilai lebih besar dari 0 ($X > 0$) dan lebih kecil dari 1 ($X < 1$) atau dengan notasi matematis kondisi tersebut bisa dinyatakan dengan $\{X | 0 < X < 1, X \in R\}$. Dengan demikian maka terdapat persamaan yang dapat digunakan dalam proses normalisasi sebagaimana dirumuskan persamaan 1.

B. Clustering

Clustering merupakan *unsupervised Classification* pola tertentu dalam kelompok (*Cluster*) tertentu. Secara umum, *Clustering* adalah metode untuk membagi data menjadi kelompok berdasarkan kemiripannya. Salah satu area penelitian yang signifikan dalam data *mining* adalah mengembangkan metode untuk memodernisasi pengetahuan menggunakan pengetahuan yang ada, karena pada umumnya dapat meningkatkan efisiensi penggalian data, terutama untuk database dengan ukuran yang sangat besar[4]

C. Self Organizing Maps (SOM)

SOM *Clustering* dapat diimplementasikan untuk menunjang berbagai penelitian di bidang yang terkait dengan pengolahan data antara lain PL-G-SOM (*Parameterless-Growing-SOM*) dan T-SOM (*Transient-SOM*) yang seringkali digunakan dalam bidang robotika. PL-G-SOM dan T-SOM digunakan untuk memetakan data resonansi suara yang kemudian divisualkan. Ini digunakan untuk memilih respon apa yang tepat untuk menanggapi suara yang ditangkap.

Dalam bidang *information retrieval*, SOM juga diimplementasikan untuk menganalisa data yang didapat, SOM digunakan sebagai metode untuk meng-*cluster* data tersebut yang kemudian dilanjutkan dengan proses DF (*Document Frequency*), TF-IDF (*Term Frequency-Inverse Document Frequency*), MI (*Mutual Information*) dan proses-proses yang lainnya[5].

Adapun langkah Algoritma *Shelf Organizing Maps* adalah sebagai berikut :

1) *Inisialisasi bobot*. Yaitu menentukan secara acak bobot awal sebagai W_{ij} (banyaknya kriteria = jumlah data).

2) *Repeat* (keputusan untuk mengulang kembali iterasi).

a) Menentukan data yang akan digunakan.

b) Menentukan *centroid* dari obyek tersebut.

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

$$W_{i+1} = W_{ij} + \delta_i * (x_i + W_{ij}) \quad (2)$$

dengan

W_{i+1} = *Centroid* yang baru, $t+1$

X_i = Data pada iterasi ke- t

W_{ij} = Bobot pada iterasi ke- t

δ_i = *Learning rate* pada iterasi ke- i , tiap kenaikan (iterasi) *learning rate* = *learning rate* awal *0.5

3) sampai tidak ada perubahan *centroid* atau *threshold* sudah terpenuhi.

4) Iterasi pada langkah ke-2 akan berhenti apabila *threshold* terpenuhi, untuk mencapai nilai *threshold* terpenuhi dilakukan dengan menghitung nilai MSE.

5) Menetapkan setiap obyek terhadap *centroid* dan menentukan letak *Cluster* tersebut.

D. Cluster Validation (IDB dan I-DUNN)

1) *Index Davies Bouldin*. *Index Davies-Bouldin* merupakan *Cluster validity* yang dibuat oleh D.L. Davies. Pendekatan pengukuran ini untuk memaksimalkan jarak inter-*Cluster*. Yang dalam penelitian ini, IDB digunakan untuk memvalidasi data pada masing-masing *Cluster* sehingga outlier ditempatkan pada *Cluster* yang benar-benar sesuai.

$$\text{var}(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (3)$$

$$R_i = \max_{j=1..k, i \neq j} R_{ij} \quad (4)$$

$$R_{ij} = \frac{\text{var}(C_i) + \text{var}(C_j)}{\|c_i - c_j\|} \quad (5)$$

$$DB = \frac{1}{k} \sum_{i=1}^k R_i \quad (6)$$

Keterangan :

DB : validasi Davies Bouldin

var : *variance* dari data

N : Banyaknya data dalam *Cluster* yang sudah Terbentuk

x : data ke- i

\bar{x} : rata-rata dari tiap *Cluster*

R : jarak antar *Cluster*

Skema *Clustering* yang optimal menurut Indeks Davies Bouldin adalah yang memiliki Indeks Davies Bouldin minimal.

2) *Index Dunn*. *Index Dunn* merupakan salah satu pengukuran *Cluster validity* yang diajukan oleh J.C. Dunn. *Cluster validity* ini berlandaskan pada fakta bahwa *Cluster* yang terpisah itu biasanya memiliki jarak antar *Cluster* yang besar dan diameter intra *Cluster* yang kecil.

$$D = \min_{j=i+1..n_c} \left(\min_{j=i+1..n_c} \left(\frac{d(c_i, c_j)}{\max_{k=1..n_c} (diam(c_k))} \right) \right) \quad (7)$$

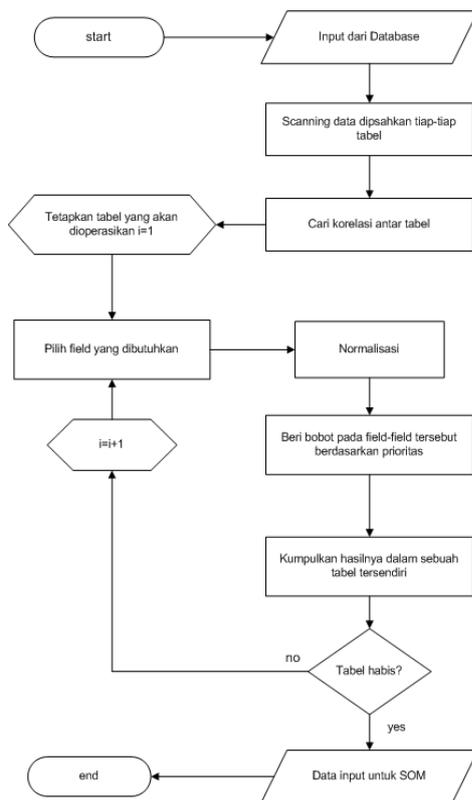
Dimana nilai $d(c_i, c_j)$ dan $diam(c_i)$ ini didefinisikan sebagai berikut:

$$d(c_i, c_j) = \min_{x \in c_i, y \in c_j} (d(x, y)) \quad (8)$$

$$diam(c_i) = \max_{x, y \in c_i} (d(x, y)) \quad (9)$$

Nilai pada *Index dunn* ini jika nilainya semakin besar, maka hasil *Clustering* akan semakin bagus.

III. METODOLOGI PERCOBAAN



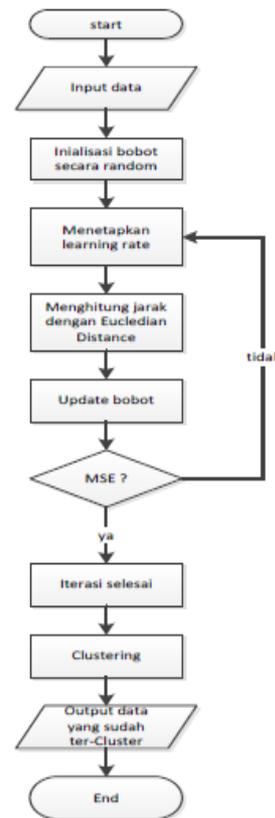
Gambar 1. Flowchart Preprocessing

Proses *Preprocessing* Seperti ditunjukkan pada Gambar 1 diawali dengan memilih kolom yang akan digunakan. Kolom mewakili variabel yang diperlukan, dalam penelitian ini yaitu parameter yang digunakan dalam tingkat kesehatan masyarakat. Pemilihan tabel juga memperhatikan korelasi antar tabel. Kolom mana saja yang dibutuhkan dipilih dan proses selanjutnya adalah melakukan normalisasi terhadap data. Kemudian berlanjut dengan pembobotan dari *content*

tiap-tiap kolom berdasarkan prioritasnya. Proses diulangi hingga semua tabel terselesaikan.

Setelah mencari nilai rata-rata kelasnya, proses normalisasi data dilakukan terhadap semua anggota di dalam kelas sehingga seluruh data akan berada pada rentang nilai antara 0 s/d 1. Setelah itu, data akan diberi bobot menurut variabel masing-masing di awal proses. Hasilnya akan disimpan pada tabel tersendiri. *Output* dari proses ini adalah sebuah tabel baru yang berisi *content* data yang telah memiliki bobot. *Output* inilah yang kemudian menjadi input untuk proses selanjutnya yaitu masuk ke proses SOM.

Dalam hal ini penentuan bobot secara tidak langsung akan mempengaruhi proses selanjutnya, dikarenakan bobot akan memberi prioritas kepada data *output* yang dikehendaki. Sehingga penentuan algoritma pembobotan yang tepat diperlukan sehingga *cluster* yang terbentuk dapat diuji dan dilakukan uji validasi. Langkahnya ditunjukkan pada Gambar 2.



Gambar 2. Flowchart Clustering SOM

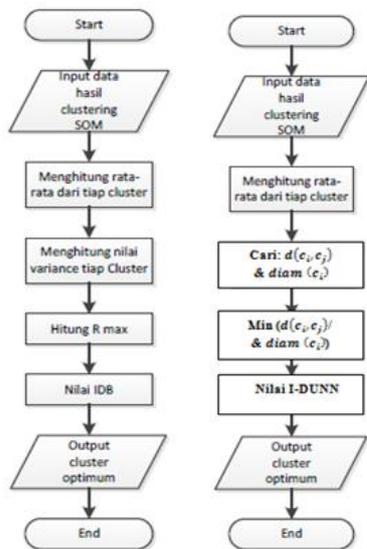
Input dari proses ini adalah data yang ada dalam *tabel* baru hasil *preprocessing*, merupakan *output* dari proses data *warehousing*. Data tersebut dikonversi ke dalam sebuah *matrix* berordo $I \times J$. *Matrix* ini yang akan diproses lagi menggunakan SOM dan *cluster validation*.

Tahap pertama yang dilakukan adalah penentuan *winner* dari data set yang sudah ada dalam *matrix*. Penetapan *winner* ini menggunakan metode statistik dimana *winner* dari masing-masing *cluster* adalah nilai maksimum, minimum dan nilai tengah (*median*) dari data. Nilai maksimum digunakan

sebagai *winner* dari *cluster* dengan tingkat kesehatan baik, nilai minimum digunakan sebagai *winner* dari *cluster* dengan tingkat kesehatan buruk dan nilai median digunakan sebagai *winner* dari *cluster* dengan tingkat kesehatan kurang baik.

Setelah diketahui masing *winner*nya, tahapan berikutnya adalah *cluster validation*. *Euclidian distance* akan menghitung jarak kedekatan antar data dimulai dengan data yang ada disekitar *winner* kemudian data yang ada disekitar data yang sudah diproses. Tahap ini dilakukan berulang-ulang hingga semua data diproses secara keseluruhan. Tahap inilah yang menyebabkan hasil visualisasi dari *cluster* tidak selalu berbentuk lingkaran atau yang biasa disebut dengan *encircled areas*. Tahap ini juga yang meningkatkan tingkat akurasi dari *clustering* menggunakan SOM. Untuk SOM yang digunakan tidak menggunakan epoch diasumsikan apabila proses telah mencapai jumlah maksimum *epoch*, maka proses akan berhenti meskipun belum semua data terproses. Sebagai gantinya maka digunakan MSE (*Mean Square Error*) karena lebih bisa mengakomodir data yang lebih bersifat dinamis.

Output dari proses *clustering* berbasis SOM ini adalah sebuah *maps* yang menggambarkan *cluster* dari data set yang diinputkan.



Gambar 3. Flowchart Diagram IDB dan I-DUNN

Gambar 3 menunjukkan *Flowchart* IDB untuk mendapatkan nilai *cluster* yang optimum yaitu dengan cara memasukkan hasil *clustering* SOM dan menghitung nilai rata-rata tiap *cluster*, menghitung nilai *variance*, menghitung nilai R (jarak antar *cluster*) maksimumnya dilanjutkan dengan menghitung nilai indeks IDB nya.

Hasil *clustering* SOM tadi juga diproses menggunakan I-DUNN dengan cara menghitung nilai rata-rata tiap *cluster*, mencari nilai jarak antar *cluster* minimum dan *cluster* diam serta menghitung nilai IDUNN nya.

IV. HASIL DAN PEMBAHASAN

A. Sumber Data

Sumber yang digunakan dalam penelitian kali ini adalah data kesehatan yang dilansir oleh KEMENKES RI dari hasil RISKESDAS JAWA TIMUR.

Tabel 1. Modul Data

Nama Modul	Jumlah Variabel		Jumlah Variabel
1. Normalitas Gizi	17	8. Cidera dan Disabilitas	38
2. Gizi Buruk	11	9. Perilaku Perokok	37
3. Ibu dan Anak	31	10. Perilaku Konsumsi Alkohol	17
4. Penyakit Menular	22	11. Perilaku dan Pengetahuan	23
5. Penyakit Tidak Menular	23	12. Pelayanan Kesehatan	76
6. Penyakit Mata	7	13. Ketanggapan Layanan Kesehatan	15
7. Kesehatan Gigi	25	14. Kesehatan Lingkungan	16

Seperti ditunjukkan Tabel 1, yang menjadi sample dalam uji coba sistem adalah data yang terdapat dalam modul penyakit mata. Sedangkan *record* yang digunakan adalah *record* Propinsi Jawa Timur dengan kabupaten sebagai elemen *record*-nya. Berikut adalah daftar variabel yang terdapat dalam modul penyakit mata: Low Vision, Kebutaan, Nakes Diagnosis, Berkabut dan Silau, Diagnosis Gejala, Operasi Katarak, Berkaca Mata setelah Operasi.

B. Uji Coba Dengan SOM

Setelah data di dapat, maka tahapan dalam SOM adalah sebagai berikut:

- 1) Menentukan bilangan terbesar dan terkecil dari tiap variabel.
- 2) Data melalui proses normalisasi menggunakan persamaan (1) akan didapat data yang sudah ter-normalisasi.
- 3) Berikutnya adalah tahap dimana *learning rate* ditentukan. Dalam percobaan kali ini, *learning rate* di-set secara manual dengan nilai 1 (satu).
- 4) Penetapan nilai MSE. Dalam percobaan kali ini nilai MSE yang ditetapkan adalah 0,1%. Jika nilai tersebut diubah kedalam pecahan maka $1/1000 = 0.001$. Percobaan kali ini tidak menggunakan *epoch* namun menggunakan MSE sebagai acuan kapan *learning* dihentikan.
- 5) Tahapan selanjutnya adalah penetapan bobot untuk tiap variabel. Untuk menentukannya dilakukan pemberian nilai secara acak (*Random*). Dalam percobaan kali ini dibatasi nilai *random* yang diberikan dalam rentang 1 dan 10, seperti ditunjukkan table 2.

Tabel 2. Bobot Awal

	Variabel						
	1	2	3	4	5	6	7
Bobot	7	4	1	3	5	9	7

- 6) Tahapan berikutnya adalah *training data* dengan bobot yang sudah, ditunjukkan table 3

Tabel 3. Bobot Awal dan *Record* Pertama

	Variabel						
	1	2	3	4	5	6	7
Bobot	7	4	1	3	5	9	7
Record	1	0.5	0.1538	0.6785	0.6433	0.230769	0.333

7) Record akan dikalikan dengan nilai bobot dan hal tersebut dilakukan pada setiap baris record data ternormalisasi yang diproses. Dari nilai diatas maka didapat nilai record baru untuk iterasi berikutnya.

8) Proses berikutnya adalah update learning rate. Ini dilakukan dengan mengalikan nilai learning rate awal dengan $\frac{1}{2}$ sehingga nilai learning rate akan berubah menjadi $\frac{1}{2}$ dari nilai awalnya pada tiap iterasi. Karena diawal learning rate di-set dengan nominal 1 maka $(1 \times \frac{1}{2}) = 0.5$

9) Update bobot dilakukan setelah proses update learning rate selesai. Update bobot menggunakan persamaan (2)

10) Langkah yang berikutnya adalah menghitung MSE dimana MSE merupakan hasil dari persamaan (4)

11) Langkah ke-7 hingga 10 dilakukan terus menerus hingga MSE yang didapat mendekati dengan yang ditetapkan pada langkah ke-5.

12) Dalam percobaan ini MSE terakhir yang didapatkan adalah 0.00086. Karena $0.000857631198821652 < 0.001$ (nilai MSE yang ditetapkan dalam langkah ke-5) maka iterasi dihentikan.

13) Karena iterasi sudah berakhir, dilakukan langkah ke-7 dengan data training pada iterasi terakhir lalu kemudian jumlahkan semua data yang berada dalam satu record menghasilkan data baru.

14) Data tersebut dipetakan kedalam 3, 4, 5, 6, 7, 8, 9, dan 10 kelompok dengan memanfaatkan nilai maksimum, minimum dan kaidah quartil dalam ilmu statistik.

Berikut adalah hasil dari Clustering SOM yang dipetakan ke dalam 3-10 Cluster ditunjukkan Tabel 4.

Tabel 4. Hasil Clustering SOM 3-10 Cluster

No	Daerah	Cluster						
		3	4	5	6	7	8	9
1	Pacitan	3	3	4	4	5	6	6
2	Ponorogo	1	3	3	4	4	5	5
3	Trenggalek	1	4	5	6	7	8	9
4	Tulungagung	1	4	4	5	6	7	8
5	Blitar	1	2	2	3	3	3	4
6	Kediri	2	3	3	4	4	5	5
7	Malang	2	3	3	4	4	5	5
8	Lumajang	2	3	4	4	5	5	6
9	Jember	2	2	2	2	3	3	3
10	Banyuwangi	2	1	1	1	1	1	1
11	Bondowoso	2	3	4	5	6	6	7
12	Situbondo	2	2	2	3	3	4	4
13	Probolinggo	2	2	3	3	3	4	4
14	Pasuruan	2	4	5	5	6	7	8
15	Sidoarjo	2	3	3	4	4	5	6
16	Mojokerto	2	2	2	3	3	3	4
17	Jombang	1	3	4	5	5	6	7
18	Nganjuk	1	3	3	4	4	5	5
19	Madiun	2	3	4	5	5	6	7
20	Magetan	1	3	3	4	4	5	5
21	Ngawi	1	1	1	1	1	1	1
22	Bojonegoro	2	3	3	4	4	5	5
23	Tuban	2	3	4	5	5	6	7
24	Lamongan	2	3	4	4	5	6	6
25	Gresik	1	4	5	6	7	7	8
26	Bangkalan	3	3	4	5	5	6	7
27	Sampang	3	4	5	6	7	7	8
28	Pamekasan	2	4	4	5	6	7	7
29	Sumenep	3	3	4	4	5	6	6
30	Kota Kediri	3	1	1	1	1	1	1
31	Kota Blitar	1	2	2	3	3	3	4
32	Kota Malang	2	3	4	5	5	6	7

33	Kota Probolinggo	2	4	5	6	7	8	9
34	Kota Pasuruan	3	4	5	6	7	8	9
35	Kota Mojokerto	2	3	3	4	4	5	5
36	Kota Madiun	2	2	3	3	4	4	5
37	Kota Surabaya	2	4	5	6	7	8	9
38	Kota Batu	1	3	3	4	5	5	6

Hasil dari proses Cluster SOM pada Tabel 4 digunakan pada proses selanjutnya yaitu pencarian nilai validasi IDB dan I-Dunn untuk 10 Cluster.

C. Uji Coba SOM Menggunakan Validasi Indeks Davies-Bouldin

Setelah proses SOM dan menghasilkan nilai pengelompokan, dilanjutkan dengan perhitungan validasi menggunakan Index Davies-Bouldin. Dalam pengujian ini digunakan perbandingan hasil proses SOM menggunakan 3 cluster, 4 cluster, 5 cluster, 6 cluster, 7 cluster, 8 cluster, 9 cluster dan 10 cluster. Dari proses SOM menggunakan 8 Cluster di hasilkan :

- Jumlah Cluster 1 = 3 daerah
- Jumlah Cluster 2 = 0 daerah
- Jumlah Cluster 3 = 4 daerah
- Jumlah Cluster 4 = 3 daerah
- Jumlah Cluster 5 = 10 daerah
- Jumlah Cluster 6 = 9 daerah
- Jumlah Cluster 7 = 5 daerah
- Jumlah Cluster 8 = 4 daerah

Dengan menggunakan 8 cluster, maka pembagian daerah untuk masing masing cluster ditunjukkan tabel 5.

Tabel 5. Data SOM 8 Cluster

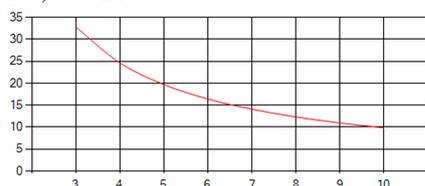
SOM	Cluster	No	Daerah	
8 Cluster	Cluster Pertama	1	Banyuwangi	
		2	Ngawi	
		3	Kota Kediri	
		Cluster Kedua	0	
	Cluster Ketiga	1	Blitar	
		2	Jember	
		3	Mojokerto	
		4	Kota Blitar	
	Cluster Keempat	1	Situbondo	
		2	Probolinggo	
3		Kota Madiun		
Cluster Kelima	1	Ponorogo		
	2	Kediri		
	3	Malang		
	4	Lumajang		
	5	Sidoarjo		
	6	Nganjuk		
	7	Magetan		
	8	Bojonegoro		
	9	Kota Mojokerto		
	10	Kota Batu		
Cluster Keenam	1	Pacitan		
	2	Bondowoso		
	3	Jombang		
	4	Madiun		
	5	Tuban		
	6	Lamongan		
	7	Bangkalan		
	8	Sumenep		
	9	Kota Malang		
Cluster Ketujuh	1	Tulungagung		
	2	Pasuruan		
	3	Gresik		
	4	Sampang		
	5	Pamekasan		
Cluster Kedelapan	1	Trenggalek		
	2	Kota Probolinggo		
	3	Kota Pasuruan		
	4	Kota Surabaya		

Adapun munculnya cluster dengan anggota kosong merupakan hasil perhitungan SOM dimana jarak terdekat objek ke pusat cluster tersebut tidak ditemukan seiring dengan kenaikan jumlah cluster. Hal ini ditunjukkan pula pada tabel 4, bahwa pada saat cluster 3 sampai dengan cluster 6 masih terdapat anggota cluster 2. Selanjutnya dari hasil clustering SOM tabel 5, dihitung nilai Index Bouldin dan didapatkan output IDB pada Tabel 6.

Tabel 6. Hasil *Clustering* SOM menggunakan Validasi Indeks Davies-Bouldin

No	Cluster	Hasil validasi I-DB
1	Tiga Cluster	32,86573681
2	Empat Cluster	24,64930261
3	Lima Cluster	19,71944209
4	Enam Cluster	16,43286841
5	Tujuh Cluster	14,08531578
6	Delapan Cluster	12,3246513
7	Sembilan Cluster	10,9552456
8	Sepuluh Cluster	9,859721043

Berdasarkan tabel 6 ditunjukkan bahwa hasil terbaik menggunakan validasi IDB terjadi pada *Cluster* sepuluh dengan nilai 9,85972104309579.



Gambar 4 : Grafik Index Davies Bouldin

Dari grafik Gambar 4 dapat diketahui bahwa semakin banyak *Cluster* yang dilakukan maka nilai IDB semakin optimal.

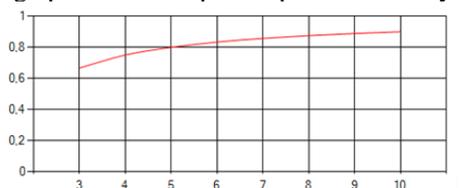
D. Uji Coba SOM menggunakan Validasi Indeks-Dunn

Hal yang sama dilakukan setelah proses SOM dan menghasilkan nilai dari beberapa pengelompokan. Selanjutnya perhitungan validasi menggunakan *Index Dunn*. Dalam pengujian kali ini digunakan perbandingan dari hasil proses SOM menggunakan 3 *cluster*, 4 *cluster*, 5 *cluster*, 6 *cluster*, 7 *cluster*, 8 *cluster*, 9 *cluster* dan 10 *cluster*. Dari pengukuran sampel modul penyakit mata, didapatkan *output* validasi I-Dunn sebagai berikut:

Tabel 7 : Hasil *Clustering* SOM menggunakan Validasi Indeks Dunn

No	Cluster	Hasil validasi I-Dunn
1	Tiga Cluster	0,666666667
2	Empat Cluster	0,75
3	Lima Cluster	0,8
4	Enam Cluster	0,833333333
5	Tujuh Cluster	0,857142857
6	Delapan Cluster	0,875
7	Sembilan Cluster	0,888888889
8	Sepuluh Cluster	0,9

Berdasarkan tabel 7 diketahui bahwa nilai Indeks Dunn paling optimal adalah pada sepuluh *Cluster* yaitu : 0,9.



Gambar 5. Grafik Index-dunn

Dari grafik Gambar 5 menunjukkan bahwa semakin banyak *Cluster* maka nilai Index-Dunn semakin optimal. Pada

penelitian ini, algoritma SOM dengan validasi IDB ini di implementasikan pada modul Gizi Buruk, Kesehatan Gigi, Kesehatan Ibu dan Anak, Kesehatan Lingkungan, Ketanggapan Layanan Kesehatan, Normalitas Gizi, Pelayanan Kesehatan, Penyakit Menular, Penyakit Tidak Menular, Perilaku dan Pengetahuan, Perilaku dan Pengetahuan, Perilaku Konsumsi Alkohol, Perilaku Perokok. Adapun hasilnya diinformasikan kepada dinas kesehatan propinsi jawa timur.

V. PENGAKUAN

Penelitian ini merupakan karya bersama dosen Manajemen Informatika dan Teknik Informatika Universitas serta mahasiswa Teknik Universitas Trunojoyo Madura.

VI. KESIMPULAN

Adapun kesimpulan penelitian ini adalah sebagai berikut:

a) *Data RISKESDAS* memiliki jumlah *variable* terkecil pada modul penyakit mata dengan 7 *variable*, dan terbesar modul kesehatan lingkungan 86 buah.

b) Hasil *Clustering SOM* dengan keanggotaan terlengkap didapatkan pada 6 *cluster* dikarenakan peningkatan jumlah *cluster* merapatkan jarak keanggotaan ke *centroid* di masing masing *cluster*, sehingga akan terdapat *cluster* yang tidak memiliki anggota.

c) Hasil uji coba data menggunakan validasi *Index Davies Bouldin* menunjukkan semakin kecil nilai IDB yang dihasilkan 9.857 maka semakin optimal nilai jarak yang dihasilkan.

d) Sedangkan hasil uji coba menggunakan validasi *Index Dunn* menunjukkan semakin besar nilai I-dunn 0.9 maka semakin optimal nilai jaraknya. Berkebalikan dengan IDB.

REFERENCES

- [1] Sen, S. and N. Chaki, "Efficient Traversal in Data Warehouse Based on Concept Hierarchy Using Galois Connections," in Emerging Applications of Information Technology (EAIT), 2011 Second International Conference on. 19-20 Feb. 2011.
- [2] Garay, A.B., G.P. Contreras, and R.P. Escarcina, "A GH-SOM optimization with SOM labelling and dunn index," in Hybrid Intelligent Systems (HIS), 2011 11th International Conference on. 5-8 Dec. 2011.
- [3] Yanchi, L., et al., "Understanding and Enhancement of Internal Clustering Validation Measures," *Cybernetics, IEEE Transactions on*, **43**, (3): p. 982-994, 2013.
- [4] Sarlin, P. and Z. Yao, "Clustering of the Self-Organizing Time Map," *Neurocomputing*, **121**, (0): p. 317-327, 12/9, 2013.
- [5] Juntunen, P., et al., "Cluster analysis by self-organizing maps: An application to the modelling of water quality in a treatment process," *Applied Soft Computing*, **13**, (7): p. 3191-3196, 7//, 2013.