

OPTIMALISASI PENGELOMPOKAN KECAMATAN BERDASARKAN INDIKATOR PENDIDIKAN MENGGUNAKAN METODE CLUSTERING DAN DAVIES BOULDIN INDEX

Firli Irhamni¹
firli@if.trunojoyo.ac.id
Universitas Trunojoyo Madura

Fitri Damayanti²
fitri2708@yahoo.com
Universitas Trunojoyo Madura

Bain Khusnul K³
bainkk@gmail.com
Universitas Trunojoyo Madura

Mifftachul A⁴
mitha.affatur@gmail.com
Universitas Trunojoyo Madura

ABSTRAK

Bidang pendidikan saat ini merupakan salah satu bidang prioritas yang diutamakan oleh pemerintah. Diperlukan informasi pendukung mengenai kondisi pendidikan tersebut agar perencanaan dan kebijakan pemerintah dapat tepat sasaran. Salah satu perencanaan yang diharapkan pemerintah adalah pemerataan pendidikan. Indikator pendidikan terdapat pada kecamatan sebagai salah satu organisasi pemerintah. Pengetahuan mengenai informasi pemerataan pendidikan diperlukan untuk pengelompokan kecamatan berdasarkan indikator pendidikan. Pengelompokan (*clustering*) merupakan salah satu metode *data mining* yang membagi data ke dalam kelompok – kelompok yang mempunyai objek dengan kemiripan karakteristiknya. Salah satu metode *clustering* adalah *Self organizing Maps (SOM)*. Pada penelitian ini, akan mengoptimalkan kelompok kecamatan untuk mendapatkan tingkat kemiripan yang terbaik menggunakan metode *Davies Bouildin Index (DBI)* dari hasil pengelompokan kecamatan menggunakan *SOM*. Pada uji coba yang telah dilakukan menggunakan jumlah kelompok dari tiga sampai dengan sepuluh kelompok dengan menggunakan *learning rate* sebesar 0.9 dan nilai *centroid* acak. Sehingga diperoleh kelompok kecamatan yang paling optimal berdasarkan kemiripan karakteristiknya.

Kata Kunci: *clustering*, indikator pendidikan, *Self Organizing Map (SOM)*, *Davies Bouldin Index (DBI)*.

I. Pendahuluan

Bidang pendidikan saat ini merupakan salah satu bidang prioritas yang diutamakan oleh pemerintah. Berdasarkan UU 20 / 2003 tentang Sisdiknas menyatakan bahwa anggaran pendidikan sekurang-kurangnya adalah 20% dari APBN. Tujuan dari Undang-Undang tersebut supaya adanya peningkatan dan pemerataan pendidikan di Indonesia. Pemerataan pendidikan diharapkan dapat mengurangi kesenjangan pendidikan antara suatu daerah dengan daerah lainnya. Diperlukan informasi pendukung mengenai kondisi pendidikan tersebut agar perencanaan dan kebijakan pemerintah dapat tepat sasaran.

Salah satu perencanaan yang diharapkan pemerintah adalah pemerataan pendidikan. Untuk melakukan pemerataan pendidikan menggunakan indikator pendidikan yang terdapat pada kecamatan sebagai salah satu organisasi pemerintah. Salah satu parameter

penyebab keberhasilan pendidikan dapat dilihat dari indikator pendidikan di suatu daerah. Kondisi pemerataan pendidikan disuatu daerah dapat diukur dari rendahnya nilai Angka Partisipasi Murni (APM) dan nilai Angka Partisipasi Kasar (APK). Indikator lainnya yang memengaruhi kondisi pemerataan pendidikan adalah factor yang berhubungan dengan sarana dan prasarana pendidikan yaitu jumlah sekolah, ruang kelas, dan tenaga pengajar [1].

Pengelompokan kecamatan berdasarkan tingkat pendidikan SMA/SMK/MA tersebut menggunakan metode *clustering* yaitu *Self Organizing Map (SOM)* yang dapat menangani data dengan jumlah fitur yang cukup banyak. *Clustering* data digunakan untuk membagi data menjadi kelompok berdasar kemiripan pola yaitu kesamaan indikator pendidikan. Hasil *clustering* data tersebut kemudian diolah dengan metode *Davies Bouldin Index (DBI)* untuk menghasilkan jumlah *cluster* yang paling optimal. Penelitian ini menggunakan data

indikator pendidikan setiap kecamatan di kabupaten Lamongan -Jawa Timur sebanyak 27 kecamatan dan 10 indikator pendidikan.

II. Metode

Clustering

Clustering merupakan salah satu metode dalam data mining yaitu teknik pengelompokan data, pengamatan atau memperhatikan dan membentuk kelas obyek yang memiliki kemiripan. *Clustering* tidak mempunyai target output. Pada metode ini tidak dapat ditentukan hasil output selama proses pembelajaran. Selama proses pembelajaran, nilai bobot disusun dalam suatu range tertentu tergantung pada nilai input yang diberikan. Tujuan pembelajaran ini adalah mengelompokkan unit-unit yang hampir sama dalam suatu area tertentu. *Clustering* berbeda dengan klasifikasi, dalam hal tidak ada variabel target untuk *clustering*. *Clustering* tidak mengklasifikasikan, meramalkan, atau memprediksi nilai dari sebuah variabel target dan digunakan ketika kita tidak mengetahui bagaimana data harus dikelompokkan.

Self organizing maps (SOM)

SOM merupakan metode berdasarkan model dari pendekatan jaringan syaraf tiruan. SOM dikembangkan oleh Prof. Teuvo Kohonen ilmuwan Finlandia pada tahun 1982. Jaringan SOM merupakan suatu jaringan yang banyak digunakan untuk membagi pola masukan ke dalam beberapa kelompok *cluster*. SOM menyediakan suatu teknik visualisasi data yang membantu memahami data yang memiliki dimensi yang kompleks dengan mengurangi dimensi data kedalam peta. Dalam nilai bobot terdapat pola nilai input yang dikumpulkan dalam *cluster*. Selama proses SOM, unit *cluster* yang mempunyai nilai bobot akan dicocokkan dengan pola input yang terdekat dan dipilih sebagai pemenang [2].

Euclidean Distance

Dari beberapa penelitian untuk clustering, pengukuran jarak yang pada umumnya sering digunakan adalah jarak *Euclidean*. *Euclidean Distance* dianggap sebagai distance matrix yang mengadopsi prinsip Pythagoras. Hal ini dikarenakan pola perhitungannya yang menggunakan aturan pangkat dan akar kuadrat. *Euclidean* akan memberikan hasil jarak yang relatif kecil karena menggunakan aturan akar kuadrat. Jarak antara Nilai Random/ Bobot dan

data dihitung dengan menggunakan rumus persamaan 1 berikut [2]:

$$D_i = \sqrt{\sum_{j=1}^n (w_{ij} - x_i)^2} \quad (1)$$

dimana :

D_i : Jarak *Euclidean*
 w_{ij} : Bobot Neuron ke- i (bobot akhir)
 x_i : input vector ke x_i

Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 adalah metrik untuk mengevaluasi hasil algoritma clustering [3].

Rancangan Sistem

Tahapan proses pada pada penelitian ini sebagai berikut :

1. Menyiapkan data dari kecamatan dan variabel indikator pendidikan.
2. Data indikator pendidikan akan diproses menjadi data hasil normalisasi untuk input indikator pendidikan.
3. Proses data kecamatan akan menggunakan metode *clustering* SOM dan disimpan dengan nama bobot termasuk melakukan perhitungan jarak dengan menggunakan *Euclidean distance* antar bobot tersebut.
4. Bobot yang sudah dihitung jaraknya menggunakan *Euclidean distance* akan ditampilkan dan berkelompok sesuai dengan kesamaan karakteristik dari data tersebut.

Dari proses *Clustering* tersebut kemudian dilakukan penilaian menggunakan *Davies-Bouldin Index* (DBI) untuk menentukan jumlah cluster paling optimal dalam proses *Clustering* tersebut.

Sebelum menghitung *Davies-Bouldin Index* (DBI) [3], dihitung terlebih dahulu *variance* dari masing-masing *cluster*.

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

dimana :

\bar{x} : rata-rata dari *cluster* x dan N adalah jumlah anggota *cluster*.

Kemudian menghitung *Davies-Bouldin Index* (DBI) dengan persamaan :

$$DBI = \frac{1}{k} \cdot \sum_{i=1}^k R_i \quad (3)$$

dengan

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

dan

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

dimana :

C_i : cluster i dan c_i adalah centroid dari cluster i.

III. Hasil dan Pembahasan

Data Uji Coba

Tabel 1. Tabel Data Indikator Pendidikan

No.	Kecamatan	APM	APK	AT	APS	AU
1	Sukorame	14.5	15.125	107.692	0	0
2	Bluluk	47.587	75.737	89.899	0.354	0.354
3	ngimbang	48.349	63.365	112.4	0.124	0.124
4	sambeng	13.105	19.008	104.167	0	0
5	Mantup	52.398	66.244	107.203	0	0
...
21	Laren	20.855	32.757	169.697	0	0
22	solokuro	4.611	6.614	121.739	0	0
23	Paciran	27.033	52.057	127.5	0.121	0.121
24	brondong	10.078	18.936	74.074	0	0
25	Sarirejo	5.686	6.549	104.651	0	0

Perhitungan dengan SOM Euclidian Distance

Tahapan SOM menggunakan *Euclidian Distance* adalah sebagai berikut:

1. Menentukan *Learning rate* secara manual : 0.6 setiap kenaikan *epoch* (iterasi)
 $learning\ rate = learning\ rate\ awal * 0.5$
2. Inisialisasi Bobot awal secara random (Banyaknya Kriteria = Jumlah Data).
 Pada tabel 2 menunjukkan bobot awal dengan 3 Cluster.

Tabel 2. Bobot Awal

AP M	AP K	AT	AP S	AU	AL	R-S/RB	R-K/RB	R-M/G	R-M/S
0.42	0.6	0.79	0.26	0.57	0.69	0.94	0.17	0.52	0.12
0.18	0.79	0.76	0	0.77	0.97	0.23	0.93	0.39	0.03
0.43	0.99	0.99	0.84	0.31	0.76	0.14	0.73	0.35	0.35

Untuk setiap data dihitung menggunakan dengan *Euclidean Distance*. Berikut contoh perhitungan setiap data terhadap bobot menggunakan persamaan 1 :

$$D = ((0.274-0.42)^2)+((0.188-0.6)^2)+((0.635-0.79)^2)+((0-0.26)^2)+((0-0.57)^2)+((1-0.69)^2)+((0.475-0.94)^2)+((0.522-0.17)^2)+((0.494-0.52)^2)+((0.26-12)^2) = 1.065$$

$$D = ((0.274-0.18)^2)+((0.188-0.79)^2)+((0.635-0.76)^2)+((0-0)^2)+((0-0.77)^2)+((1-0.97)^2)+((0.475-0.23)^2)+((0.522-0.93)^2)+((0.494-0.39)^2)+((0.26-0.03)^2) = 1.271$$

$$D = ((0.274-0.43)^2)+((0.188-0.99)^2)+((0.635-0.27)^2)+((0-0.84)^2)+((0-0.31)^2)+((1-0.76)^2)+((0.475-0.14)^2)+((0.522-0.73)^2)+((0.494-0.35)^2)+((0.26-0.35)^2) = 1.844$$

3. Setelah didapat *winner*, maka dilakukan *update* bobot dengan menggunakan persamaan 6 :

$$w_{ij}(baru) = w_{ij}(lama) + \alpha [x_i - w_{ij}(lama)] \quad (6)$$

dimana :

- $W_{ij}(baru)$: Update bobot
- $W_{ij}(lama)$: Bobot lama
- X_i : Input data ke- X_i
- A : *Learning rate*

Berikut contoh perhitungan update bobot dengan menggunakan persamaan 6 (perhitungan setiap data menggunakan rumus *Euclidean Distance*):

$$Update\ bobot = [(0.42\ 0.6\ 0.79\ 0.26\ 0.57\ 0.69\ 0.94\ 0.17\ 0.52\ 0.12) + 0.6 [(0.274\ 0.188\ 0.635\ 0\ 0\ 1\ 0.475\ 0.522\ 0.494\ 0.26) - (0.42\ 0.6\ 0.79\ 0.26\ 0.57\ 0.69\ 0.94\ 0.17\ 0.52\ 0.12)]] = [0.332\ 0.353\ 0.697\ 0.104\ 0.228\ 0.876\ 0.661\ 0.381\ 0.504\ 0.204]$$

Jadi, bobot baru iterasi ke-1 untuk bobot data selanjutnya seperti pada tabel 3 berikut:

Tabel 3. Update Bobot

APM	APK	AT	APS	AU	AL	R-S/RB	R-K/RB	R-M/G	R-M/S
0.332	0.353	0.697	0.1	0.228	0.876	0.661	0.381	0.504	0.204
0.18	0.79	0.76	0	0.77	0.97	0.23	0.93	0.39	0.03
0.43	0.99	0.27	0.84	0.31	0.76	0.14	0.73	0.35	0.35

Selanjutnya hitung update bobot dan jarak *Euclidean* untuk setiap bobot sampai data ke X_n . Bobot *epoch* 1 terakhir digunakan untuk bobot awal *epoch* 2.

Tabel 3 : Hasil Pembahasan

Cluster Set	DBI	Cluster yang dihasilkan	Jumlah anggota cluster
3	0,182	3	Cluster 1 : 13
			Cluster 2 : 2
			Cluster 3 : 12
4	0,139	4	Cluster 1 : 4
			Cluster 2 : 4
			Cluster 3 : 2
			Cluster 4 : 17
5	0,526	5	Cluster 1 : 1
			Cluster 2 : 3
			Cluster 3 : 5
			Cluster 4 : 2
			Cluster 5 : 16
6	0,292	5	Cluster 1 : 1
			Cluster 2 : 2
			Cluster 3 : 6
			Cluster 4 : 2
			Cluster 5 : 16
7	1,081	5	Cluster 1 : 1
			Cluster 2 : 1
			Cluster 3 : 3
			Cluster 4 : 4
			Cluster 5 : 18
8	0,436	6	Cluster 1 : 1
			Cluster 2 : 2
			Cluster 3 : 3
			Cluster 4 : 4
			Cluster 5 : 5
			Cluster 6 : 12
9	2,056	8	Cluster 1 : 1
			Cluster 2 : 2
			Cluster 3 : 2
			Cluster 4 : 12
			Cluster 5 : 1
			Cluster 6 : 5
			Cluster 7 : 2
			Cluster 8 : 2
10	0,822	6	Cluster 1 : 1
			Cluster 2 : 1
			Cluster 3 : 5
			Cluster 4 : 4
			Cluster 5 : 10
			Cluster 6 : 2
			Cluster 7 : 1
			Cluster 8 : 1
			Cluster 9 : 2

4. Melakukan perhitungan dari data X_i sampai terakhir (dinamakan *epoch*1/iterasi1), kemudian setelah perhitungan data X_i -terakhir, melakukan pengecekan bobot (sebelum pindah ke *epoch*2).

Implementasi *Davies-Bouldin Index (DBI)*

Dalam penelitian ini menggunakan percobaan perhitungan 3 sampai 10 cluster dan menghasilkan hasil seperti terlihat dalam tabel 4.

Dari hasil perhitungan pada tabel 4 tersebut maka yang mempunyai DBI paling rendah 0,139 adalah perhitungan dengan cluster set 4 menghasilkan 4 *cluster*. Hasil kecamatan masing-masing *cluster* terdapat pada table 4.

Tabel 5. Data Kecamatan pada *Cluster* 4

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Karangbinangun	Kembangbahu	Sukorame	Ngimbang
Kalitengah	Pucuk	Bluluk	Sambeng
Solokuro	Lamongan		Mantup
Sarirejo	Brondong		Sugio
			Kedungpring
			Modo
			Babat
			Sukodadi
			Tikung
			Deket
			Glagah
			Turi
			Karanggeneng
			Sekaran
			Maduran
			Laren
			Paciran

IV. Simpulan dan Saran

Dari hasil penelitian Optimalisasi Pengelompokan Kecamatan Berdasarkan Indikator Pendidikan Menggunakan Metode Clustering dan *Davies-Bouldin Index* dapat diambil simpulan sebagai berikut :

- Penelitian ini memberikan kontribusi terhadap pengambilan kebijakan dari pejabat yang berwenang . Hal ini disebabkan metode *Clustering* dan *Davies-Bouldin Index* akan menentukan kelompok kecamatan yang paling optimal berdasarkan kemiripan kondisi kecamatan.
- Penggunaan *Davies-Bouldin Index (DBI)* menghasilkan cluster set yang paling optimal.

Untuk pengembangan lebih lanjut serta penyempurnaan perangkat lunak untuk Pengelompokan Kecamatan Berdasarkan Indikator Pendidikan menggunakan metode

Clustering, aplikasi ini tidak hanya untuk menentukan data pemerataan pendidikan saja, tetapi juga untuk mengelompokkan data yang lainnya. Disarankan untuk menggunakan metode yang lain dan dibandingkan agar kevalidannya lebih optimum.

V. Referensi

- [1] Silia Karti, Hanna dan Irhamah. 2013. Pengelompokkan kabupaten/kota di propinsi Jawa Timur Berdasarkan Indikator Pendidikan SMA/SMK/MA dengan Metode *C-Means* dan *Fuzzy C-Means*. Institut teknologi Sepuluh Nopember (ITS). Surabaya Vol. 2(2), 2337-3520 (2301-928X Print).
- [2] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, 2006. *Introduction to Data Mining*. Addison Wesley.
- [3] Davies, D. L.; Bouldin, D. W. "A Cluster Separation Measure", IEEE Transactions on Pattern Analysis and Machine Intelligence (2): 224, 1979.