

Received: 12 June 2023

Revised: 16 December 2023

Accepted: 30 December 2023

Published: 31 December 2023

Perbandingan Metode *K-Means Clustering* dengan *Self-Organizing Maps (SOM)* untuk Pengelompokan Provinsi di Indonesia Berdasarkan Data Potensi Desa

Lisa Rianti Iyohu^{1, a)}, Ismail Djakaria^{1, b)}, La Ode Nashar^{1, c)}

¹*Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Negeri Gorontalo, Bone Bolango, 96554, Indonesia*

Email: ^{a)}lisariantiiyohu@gmail.com, ^{b)}iskar@ung.ac.id, ^{c)}laode.nashar@ung.ac.id

Abstract

K-Means is a method of grouping data into several different groups so that data that has similar characteristics is made into one group while data that has different characteristics is made into a different group, where this method works by minimizing the distance between the data and the cluster center. In addition to K-Means clustering, there is also the Self Organizing Maps (SOM) method which is an undirected method, meaning that layers consisting of neurons are arranged into groups based on input values, where each data grouping process is based on the characteristics or features of the data. Clustering is carried out in Provinces in Indonesia based on village potential data in 2021 with the aim of knowing the performance comparison of K-Means clustering and Self Organizing Maps (SOM). Determination of the optimal number of clusters is carried out using the Elbow method, the results in the study obtained 3 clusters for both K-Means clustering and Self Organizing Maps (SOM). The clustering results are evaluated using the Davies Bouldin Index (DBI) value and show that clustering using the Self Organizing Maps (SOM) method provides better results than using the K-Means clustering method where the DBI value is 0.1829366. The clustering results using the Self Organizing Maps (SOM) method for cluster 1 consist of 31 province members, cluster 2 consists of 1 province member and cluster 3 consists of 2 province members.

Keywords: K-Means clustering, Self-Organizing Maps (SOM), Elbow Method, Davies Bouldin Index, Potential Village.

Abstrak

K-Means adalah metode pengelompokan data kedalam beberapa kelompok berbeda sehingga data yang mempunyai kesamaan karakteristik dijadikan satu kelompok sementara data yang mempunyai perbedaan karakteristik dijadikan satu kelompok yang berbeda, dimana metode ini bekerja dengan meminimalkan jarak antara data dan pusat *cluster*. Selain *K-Means clustering*, terdapat juga metode *Self Organizing Maps (SOM)* yang merupakan metode tanpa pengarahan,

artinya lapisan yang terdiri atas neuron-neuron disusun menjadi kelompok berdasarkan nilai input, dimana setiap proses pengelompokan data didasarkan atas karakteristik atau fitur-fitur datanya. Pengelompokan dilakukan pada Provinsi yang ada di Indonesia berdasarkan data potensi desa tahun 2021 dengan tujuan untuk mengetahui perbandingan kinerja *K-Means clustering* dan *Self Organizing Maps* (SOM). Penentuan jumlah *cluster* optimal dilakukan menggunakan metode *Elbow*, hasil pada penelitian diperoleh 3 *cluster* baik untuk *K-Means clustering* maupun *Self Organizing Maps* (SOM). Hasil pengelompokan dievaluasi menggunakan nilai *Davies Bouldin Index* (DBI) dan menunjukkan bahwa pengelompokan menggunakan metode *Self Organizing Maps* (SOM) memberikan hasil yang lebih baik dibandingkan menggunakan metode *K-Means clustering* dimana diperoleh nilai DBI sebesar 0,1829366. Hasil pengelompokan menggunakan metode *Self-Organizing Maps* (SOM) untuk *cluster* 1 terdiri atas 31 anggota Provinsi, *cluster* 2 terdiri atas 1 anggota Provinsi dan *cluster* 3 terdiri atas 2 anggota provinsi.

Kata-kata kunci: *K-Means Clustering*, *Self Organizing Maps* (SOM), Metode *Elbow*, *Davies Bouldin Index*, Potensi Desa.

PENDAHULUAN

Analisis *Cluster* ialah suatu analisis yang berfungsi mengidentifikasi kelompok atas dasar pengukuran kemiripan. *Clustering* merupakan pengelompokan objek dalam beberapa *cluster* (kelompok), yang mempunyai tingkat kemiripan sifat-sifat terhadap objek lain dibentuk menjadi satu *cluster*, sedangkan yang tidak mempunyai tingkat kesamaan karakteristik, dikelompokkan pada *cluster* yang lain (Zubedi, 2020). Secara umum, pengelompokan pada analisis dalam *cluster* terdiri dari dua metode, yakni metode hierarki dan *partitional clustering* (Matdoan and Delsen, 2020). Metode Hierarki terdiri atas lima metode, yang terdiri atas *Single Linkage*, *Complete Linkage*, *Centroid Linkage*, *Average Linkage* dan Metode *Ward* Sedangkan yang termasuk dalam metode *partitional clustering* (non-hierarki) dan yang paling banyak digunakan adalah *K-Means Clustering* (Andiani, Septiani and Riana, 2022).

Dalam *K-Means Clustering*, dimana K merupakan nilai *cluster* yang telah ditentukan lebih awal, sedangkan rata-rata dari grup data yang akan dijadikan *cluster* (kelompok) disebut *Means*. jadi *K-Means Clustering* merupakan metode yang dalam proses pengelompokannya tidak dengan pengawasan (Setiadi and Sikumbang, 2020). *K-Means* ialah metode pengelompokan data kedalam beberapa kelompok berbeda sehingga data yang mempunyai kesamaan karakteristik dijadikan satu kelompok sementara data yang mempunyai perbedaan karakteristik dijadikan satu kelompok yang berbeda, dimana metode ini bekerja dengan meminimalkan jarak antara data dan pusat *cluster* (Dacwanda and Nataliani, 2021).

Selain *K-Means clustering*, terdapat juga metode *clustering* lainnya yaitu *Self Organizing Maps* (SOM). Metode (SOM) adalah metode tanpa pengarahan, artinya lapisan yang terdiri atas neuron-neuron disusun menjadi kelompok berdasarkan nilai input, dimana setiap proses pengelompokan (*clustering*) data didasarkan atas karakteristik atau fitur-fitur datanya (Kapita and Abdullah, 2020). *Self Organizing Maps* (SOM) adalah *clustering* yang harus menggunakan data berdasarkan karakteristik yang sama, kemudian data-data yang memiliki kesamaan karakteristik tersebut ditempatkan kedalam satu *cluster*. Namun, jika data mempunyai perbedaan karakteristik, maka data tersebut akan ditempatkan ke 2 dalam *cluster* lainnya yang berbeda (Fawaz, Fitriyani and Rosalia, 2022).

Potensi desa merupakan kemampuan suatu desa yang dapat dikembangkan guna untuk meningkatkan taraf kesejahteraan masyarakat dalam desa tersebut (Soleh, 2017). Sejak tahun 1980 Badan Pusat Statistik (BPS) telah melaksanakan pendataan potensi desa (Podes). Sejak saat itu, dalam kurun waktu sepuluh tahun pengumpulan data potensi desa secara rutin dilaksanakan sebanyak tiga kali untuk mendukung pelaksanaan sensus pertanian, sensus ekonomi dan sensus penduduk. Oleh karenanya, ketersediaan prasarana dan potensi yang dimiliki oleh setiap daerah dapat ditinjau

perkembangannya secara sistematis. Potensi desa yang dimaksud dalam penelitian kali ini ialah usaha industri. Usaha industri merupakan badan ekonomi yang melakukan kegiatan ekonomi yang ditujukan untuk produksi barang atau jasa, terletak di gedung atau tempat tertentu dan memiliki catatan administrasi sendiri tentang struktur produksi dan biaya serta ada orang yang bertanggung jawab atas usaha tersebut (www.bps.go.id).

Pendataan potensi desa tahun 2021 menunjukkan potensi industri yang dimiliki desa atau kelurahan. Terdapat 3.196 desa/kelurahan yang memiliki industri kulit, 44.338 desa/kelurahan yang memiliki industri furnitur, 21.336 desa/kelurahan yang memiliki industri barang logam, 12.589 desa/kelurahan yang memiliki industri tekstil, 28.819 desa/kelurahan yang memiliki industri pakaian jadi, 18.037 desa/kelurahan yang memiliki industri gerabah, 29.097 desa/kelurahan yang memiliki industri kayu, 45.169 desa/kelurahan yang memiliki industri makanan, 39.812 desa/kelurahan yang memiliki industri minuman, 2.452 desa/kelurahan yang memiliki industri pengolahan tembakau, 785 desa/kelurahan yang memiliki industri kertas, 11.420 desa/kelurahan yang memiliki industri percetakan dan reproduksi media rekaman, 2.896 desa/kelurahan yang memiliki industri alat angkutan, 6.047 desa/kelurahan yang memiliki industri kerajinan, 14.198 desa/kelurahan yang memiliki reparasi dan pemasangan mesin (Badan, 2022). Dalam mendukung bergeraknya perekonomian, salah satu penunjangnya adalah usaha industri. Oleh karena itu perlu untuk menilai tingkat kapasitas desa berdasarkan potensi desa sehingga kebijakan dan strategi pembangunan dapat dengan baik dan tepat sasaran. Sehingga, sangat penting untuk mempertimbangkan pengelompokan dan karakteristik Provinsi berdasarkan data potensi desa.

Pada penelitian sebelumnya oleh (Muningsih, Maryani and Handayani, 2021) mengenai Penerapan Metode *K-Means* dan Optimasi Jumlah *Cluster* dengan *Index Davies Bouldin* untuk *Clustering* Provinsi Berdasarkan Potensi Desa, dimana dalam penelitian tersebut terbentuk 3 *cluster*. Pada *cluster* pertama terdiri atas 31 anggota Provinsi dengan kategori Provinsi yang memiliki potensi desa (industri) sedikit, *cluster* kedua terdiri atas 2 anggota Provinsi dengan kategori Provinsi yang memiliki potensi desa (industri) banyak, dan *cluster* ke tiga terdiri atas 1 anggota Provinsi dengan kategori Provinsi yang memiliki potensi desa (industri) sedang. Selanjutnya penelitian yang dilakukan oleh (Imani, Alfassa and Yolanda, 2022) dengan judul *Self Organizing Maps (SOM) Clustering* untuk Analisis Data Indikator Sosial di Provinsi Nusa Tenggara Timur. dimana pada penelitian ini terbentuk 4 *cluster*. Pada *Cluster* 1 beranggota 16 Kabupaten merupakan *cluster* yang menggambarkan wilayah secara umum, dengan terdapat beberapa aspek yang menonjol seperti umur harapan hidup, jumlah penduduk, tingkat pengangguran terbuka, dan angka buta huruf. *Cluster* 2 beranggota 3 kabupaten merupakan *cluster* yang hampir menyerupai *cluster* 1 namun tidak memiliki nilai umur harapan hidup, jumlah penduduk, dan tingkat pengangguran terbuka setinggi dan menonjol *cluster* 1. *cluster* 3 merupakan *cluster* yang menggambarkan kabupaten yang berpotensi akan dapat berkembang dan bertransformasi menjadi kota dikemudian hari. *Cluster* 4 beranggota 1 kota yaitu Kota Kupang dengan hampir segala aspek variabel penelitian memiliki nilai dan indeks yang tinggi. Selanjutnya yaitu penelitian yang dilakukan oleh (Hidayatin, Adinugroho and Dewi, 2019) dengan judul Pengelompokan wilayah berdasarkan penyandang masalah kesejahteraan sosial (PMKS) dengan optimasi algoritma *K-Means* menggunakan *Self Organizing Maps (SOM)*. Pada penelitian tersebut terbentuk 2 *cluster*. dimana *cluster* pertama terdiri atas 12 kabupaten/kota sedangkan *cluster* kedua terdiri atas 26 kabupaten/kota.

Penelitian ini relevansi dengan penelitian relevan pertama (Muningsih dkk., 2021) dari segi penggunaan metode *clustering K-Means* untuk mengelompokkan potensi desa, bedanya pada penelitian ini, peneliti menggunakan metode *Elbow* untuk menentukan jumlah *cluster* optimal. Pada penelitian relevan kedua (Imani dkk., 2022) memiliki relevansi dengan penelitian ini dari segi penggunaan metode *Clustering Self Organizing Maps (SOM)*. Kemudian untuk penggunaan metode pengklasteran dengan menggabungkan metode *K-Means* dengan *Self Organizing Maps (SOM)* didasarkan pada penelitian relevan ketiga (Iskarimah dkk., 2019), namun dalam penelitian tersebut, peneliti hanya menggunakan *Silhouette coeficient* sebagai validasi, sedangkan untuk penelitian kali ini peneliti menggunakan metode *Davies Bouldin Index (DBI)* sebagai indeks validasi.

Berdasarkan uraian di atas, maka pada artikel ini akan diuraikan mengenai manakah diantara *K-Means clustering* dengan *Self Organizing Maps* (SOM) yang menjadi metode terbaik Untuk mengelompokkan Provinsi di Indonesia berdasarkan data Potensi Desa.

METODOLOGI

Bahan dan Data

Pada penelitian ini, menggunakan data sekunder berupa data potensi desa tahun 2021 yang diperoleh melalui publikasi yang bersumber dari Badan Pusat Statistik (BPS) republik Indonesia tahun 2022. Sampel yang digunakan adalah seluruh Provinsi yang ada di Indonesia yang terdiri dari 34 Provinsi. Variabel yang digunakan dalam penelitian ini yaitu: Industri Kulit (X_1), Industri *Furniture* (X_2), Industri Barang Logam (X_3), Industri Tekstil (X_4), Industri Pakaian Jadi (X_5), Industri Gerabah (X_6), Industri Kayu (X_7), Industri Makanan (X_8), Industri Minuman (X_9), Industri Pengolahan Tembakau (X_{10}), Industri Kertas (X_{11}), Industri Percetakan dan Reproduksi Media Rekaman (X_{12}), Industri Alat Angkutan (X_{13}), Industri Kerajinan (X_{14}), Reparasi dan Pemasangan Mesin (X_{15}).

Metode Penelitian

Langkah-langkah analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Menginput data potensi desa yang terdiri dari 15 variabel untuk 34 Provinsi.
2. Melakukan uji asumsi Multikolinearitas.
3. Melakukan Analisis Komponen Utama (AKU).
 - a. Melakukan uji *Kaiser Mayer Olkin* (KMO).
 - b. Melakukan uji *Barlett Sphericity*.
4. Menentukan jumlah *k cluster* optimal menggunakan metode *Elbow*.
5. Melakukan pengelompokkan dengan metode *K-Means clustering*.
 - a. Menentukan jumlah *cluster* atau nilai *k*.
 - b. Menetapkan nilai *centroid* awal.
 - c. Menghitung jarak objek ke masing-masing *centroid* dari masing-masing *cluster* menggunakan persamaan jarak *Euclidean* Sebagai berikut (Solichin and Khairunnisa, 2020):

$$D_{(ij)} = \sqrt{\sum_{k=1}^n (X_{ik} - C_{jk})^2} \quad (1)$$

- d. Mengalokasikan objek ke *centroid* terdekat.
- e. Melakukan iterasi, dan tentukan nilai *centroid* baru menggunakan persamaan sebagai berikut (Solichin and Khairunnisa, 2020):

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} X_q \quad (2)$$

- f. Ulangi langkah c sampai dengan langkah e apabila terdapat data yang berpindah *cluster*.
6. Melakukan pengelompokkan dengan metode *Self Organizing Maps* (SOM).
 - a. Menginisialisasi bobot.
 - b. Menginput Data.

- c. Menghitung jarak vektor menggunakan persamaan jarak *Euclidean* sebagai berikut (Kapita *et al.*, 2022):

$$d_j = \sum (W_{ij} - X_i)^2 \tag{3}$$

- d. Setelah menghitung jarak vektor, menentukan nilai minimum dari perhitungan jarak vektor.
- e. Melakukan perubahan bobot menggunakan persamaan sebagai berikut (Kapita *et al.*, 2022):

$$W_{ij}(\text{baru}) = W_{ij}(\text{lama}) + \alpha [X_i - W_{ij}(\text{lama})] \tag{4}$$

- f. Menghentikan iterasi jika sudah mencapai iterasi maksimum dimana selisih antara bobot baru dengan bobot lama hanya berubah sedikit saja.
7. Membandingkan hasil *cluster* dari kedua metode menggunakan nilai *Davies Bouldin Index* untuk menentukan metode terbaik.
 8. Menginterpretasi hasil pengklasteran dari metode terbaik.

HASIL DAN PEMBAHASAN

Sebelum melakukan proses pengolahan data, terlebih dahulu dilakukan pengumpulan data, dimana proses pengumpulan data dalam penelitian ini diperoleh dari publikasi Badan Pusat Statistik (BPS). Data yang digunakan merupakan data potensi desa yaitu usaha industri tahun 2021 dengan sampel yang digunakan adalah seluruh Provinsi yang ada di Indonesia yang terdiri dari 15 variabel penelitian yaitu Industri Kulit (X₁), Industri *Furniture* (X₂), Industri Barang Logam (X₃), Industri Tekstil (X₄), Industri Pakaian Jadi (X₅), Industri Gerabah (X₆), Industri Kayu (X₇), Industri Makanan (X₈), Industri Minuman (X₉), Industri Pengolahan Tembakau (X₁₀), Industri Kertas (X₁₁), Industri Percetakan dan Reproduksi Media Rekaman (X₁₂), Industri Alat Angkutan (X₁₃), Industri Kerajinan (X₁₄), Reparasi dan Pemasangan Mesin (X₁₅).

Selanjutnya pada data yang digunakan akan dilihat apakah memenuhi asumsi multikolinearitas dengan melihat koefisien korelasi.

TABEL 1. Koefisien Korelasi

Var	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅
X ₁	1														
X ₂	0,89	1													
X ₃	0,90	0,98	1												
X ₄	0,74	0,88	0,82	1											
X ₅	0,86	0,97	0,97	0,86	1										
X ₆	0,70	0,79	0,78	0,67	0,75	1									
X ₇	0,90	0,99	0,97	0,85	0,97	0,77	1								
X ₈	0,85	0,98	0,97	0,84	0,98	0,76	0,98	1							
X ₉	0,88	0,98	0,97	0,85	0,98	0,77	0,98	0,99	1						
X ₁₀	0,85	0,91	0,92	0,78	0,86	0,74	0,88	0,88	0,85	1					
X ₁₁	0,98	0,91	0,93	0,74	0,86	0,73	0,91	0,87	0,89	0,91	1				
X ₁₂	0,94	0,97	0,99	0,80	0,95	0,77	0,97	0,96	0,96	0,94	0,96	1			
X ₁₃	-	0,14	0,14	-	0,08	0,18	0,18	0,17	0,17	0,08	0,05	0,11	1		
	0,02			0,02											

X ₁₄	0,94	0,96	0,97	0,81	0,97	0,76	0,97	0,95	0,95	0,90	0,94	0,98	0,07	1	
X ₁₅	0,92	0,99	0,99	0,84	0,97	0,77	0,98	0,98	0,98	0,93	0,94	0,98	0,13	0,98	1

Berdasarkan TABEL 1 dapat dilihat bahwa nilai korelasi lebih besar dari 0,5 artinya terdapat korelasi antar variabel dari yang cukup tinggi bahkan sampai yang sangat tinggi. Sehingga perlu dilakukan transformasi variabel independen menjadi variabel baru atau yang dikenal dengan nama *Principal Component Analysis (PCA)*.

Pada uji asumsi multikolinearitas, didapatkan nilai korelasi yang lebih besar dari 0,5 sehingganya perlu dilakukan penyederhanaan variabel dengan cara mengurangi dimensinya melalui transformasi variabel independen menjadi variabel baru yang dapat menghilangkan korelasi antar variabel sehingga variabel sudah tidak saling berkorelasi. Untuk penerapan analisis komponen utama memerlukan asumsi kelayakan sampel yang dilakukan dengan uji *Kaiser Mayer Olkin (KMO)* dan asumsi keberadaan korelasi yang dilakukan dengan uji *Bartlett Sphericity*.

TABEL 2. Nilai KMO dan *Barlett Sphericity*

KMO	0,8960805
<i>Barlett Sphericity (Khi-Squared)</i>	1313,858
<i>Degree of Freedom (df)</i>	105
<i>sig</i>	6,16 x 10 ⁻²⁰⁶

Pada TABEL 2, terlihat bahwa nilai KMO sebesar 0,8960805. Seperti kriteria keputusan uji KMO yang ada pada tabel 2.2, maka nilai KMO yang didapatkan artinya data layak digunakan atau telah memenuhi asumsi Analisis Komponen Utama. p-value lebih kecil dari α (0,05) sehingga memberikan keputusan tolak H₀. Artinya terdapat korelasi yang signifikan antar variabel sehingga perlu untuk dilakukan Analisis Komponen Utama (AKU).

Hasis analisis komponen utama menggunakan software R, diperoleh nilai seperti yang ada pada tabel berikut:

TABEL 3. Eigen Value

Komponen	Eigen Value
1	12,730
2	1,056
3	0,405
4	0,362
5	0,192
6	0,149
7	0,041
8	0,022
9	0,016
10	0,007
11	0,005
12	0,004
13	0,003
14	0,001
15	0,001

TABEL 3, menunjukkan nilai eigen value dari masing-masing komponen yang terbentuk. Nilai eigen dapat menjelaskan varians dari setiap komponen utama atau besarnya kontribusi varians dalam menjelaskan keragaman data asal. Penentuan banyaknya komponen yang digunakan didasarkan pada nilai eigen value > 1. Berdasarkan tabel 4.2 terlihat bahwa nilai eigen value > 1 yaitu pada komponen 1 dan komponen 2. yang artinya komponen 1 dan komponen 2 secara bersama-sama telah dapat menjelaskan varians dari ke-15 variabel. Sehingga, dari dua komponen tersebut dianggap sudah mewakili 15 variabel potensi desa dalam hal ini adalah usaha industri di setiap Provinsi yang ada di

Indonesia tahun 2021. Sehingga didapatkan *score* komponen utama untuk 34 Provinsi yaitu sebagai berikut:

TABEL 4. *Score* Komponen Utama

Provinsi	PC1	PC2
Aceh	844,12471	325,220583
Sumatera Utara	1722,56248	612,013482
Sumatera Barat	-543,65222	-22,727956
Riau	-181,18757	-2771,526060
Jambi	-1207,60144	137,770601
Sumatera Selatan	-53,29642	102,426525
Bengkulu	-1558,00341	97,025853
Lampung	986,83728	-31,183212
Kepulauan Bangka Belitung	-2031,85986	8,048082
Kepulauan Riau	-2154,29488	28,357865
DKI Jakarta	-2429,17666	131,535376
Jawa Barat	7094,23159	260,020860
Jawa Tengah	12311,19653	279,642138
DI Yogyakarta	-1791,80238	24,132106
Jawa Timur	10798,97890	-506,846796
Banten	-488,21356	172,463465
Bali	-1340,93202	0,973745
Nusa Tenggara Barat	-985,59968	-263,701932
Nusa Tenggara Timur	589,02688	-71,760977
Kalimantan Barat	-1177,31392	-19,858657
Kalimantan Tengah	-1600,69539	59,175360
Kalimantan Selatan	-721,36750	299,112545
Kalimantan Timur	-1697,01868	114,811289
Kalimantan Utara	-2475,13304	56,698599
Sulawesi Utara	-1147,52586	142,466169
Sulawesi Tengah	-682,68367	127,802769
Sulawesi Selatan	1347,04940	338,816795
Sulawesi Tenggara	43,74264	61,958970
Gorontalo	-1674,20222	-9,784163
Sulawesi Barat	-2015,52153	62,232055
Maluku	-1677,21021	137,518744
Maluku Utara	-1687,43955	51,643128
Papua Barat	-2352,90962	67,139517
Papua	-2063,10911	-1,616869

Berdasarkan TABEL 4, diperoleh nilai atau *score* Komponen Utama (KU) untuk seluruh Provinsi di Indonesia. Nilai inilah yang selanjutnya akan digunakan untuk mengelompokkan Provinsi di Indonesia.

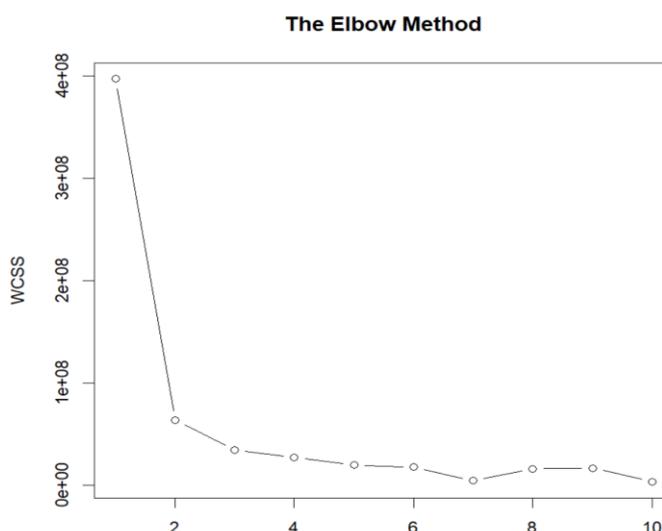
Proses selanjutnya yaitu menentukan jumlah cluster optimal pada penelitian ini menggunakan metode *elbow*, dimana dengan menghitung nilai *Sum of Squared Error* (SSE). Hasil perhitungan SSE dapat dilihat pada tabel berikut:

TABEL 5. Pengujian Metode Elbow

Cluster	SSE	Selisih
1	397.033.405	-
2	63.501.712	333.531.693
3	34.286.091	29.215.621

4	27.228.723	7.057.368
5	24.924.158,6	2.304.564, 4
6	17.717.643,5	7.206.515,1
7	17.439.229,67	278.413,83
8	16.761.799,82	677.729,85
9	16.079.748,84	682.050,98
10	15.654.720,05	425.028,79

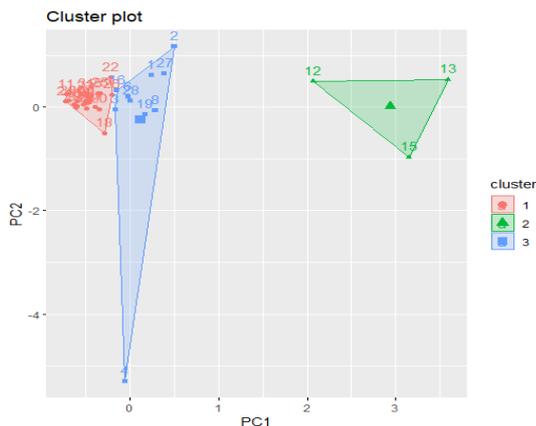
Berdasarkan hasil pengujian metode *Elbow* seperti yang ada pada TABEL 5, menunjukkan bahwa nilai penurunan paling besar yang kemudian diikuti penurunan nilai yang kecil pada *cluster* selanjutnya yaitu pada *cluster* 3, dimana memiliki nilai selisih yang besar dengan *cluster* ke-2 dan memiliki selisih nilai yang kecil dengan *cluster* selanjutnya. Untuk menentukan *cluster* yang optimal dapat dilihat pada titik yang membentuk siku pada GAMBAR 1 di bawah ini.



GAMBAR 1. Grafik Metode *elbow*

Berdasarkan GAMBAR 1 dapat dilihat bahwa titik yang membentuk siku yaitu di antara *cluster* 2 dan *cluster* 3. Maksud dari titik yang membentuk siku adalah penurunan nilai yang besar antara dua titik yang kemudian diikuti oleh penurunan nilai yang kecil. Sehingga, jumlah *cluster* optimal untuk pengelompokkan menggunakan metode *K-Means clustering* dan *Self Organizing Maps (SOM)* yang terbentuk adalah 3 *cluster* karena terdapat penurunan nilai yang besar antara *cluster* 2 dan *cluster* 3 dan diikuti oleh penurunan nilai yang kecil pada *cluster* selanjutnya.

Berikut ini adalah tampilan hasil pengelompokkan *K-Means clustering* dengan nilai *centroid* dari hasil pengelompokkan yang diolah menggunakan *software R*:



GAMBAR 2. Visualisasi K-Means Clustering

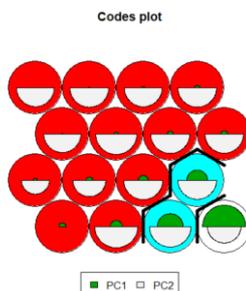
Dari gambar 2 dapat dilihat bahwa terbentuk 3 cluster dengan cluster pertama diasosiasikan dengan warna merah, cluster kedua diasosiasikan dengan warna hijau dan cluster ketiga diasosiasikan dengan warna biru. Untuk lebih jelasnya, anggota masing-masing cluster dapat dilihat pada TABEL 6 di bawah ini:

TABEL 6. Hasil K-Means Clustering

Cluster	Jumlah	Anggota
1	21	Provinsi Jambi, Provinsi Bengkulu, Provinsi Kepulauan Bangka Belitung, Provinsi Kepulauan Riau, Provinsi DKI Jakarta, Provinsi DI Yogyakarta, Provinsi Bali, Provinsi Nusa Tenggara Barat, Provinsi Kalimantan Barat, Provinsi Kalimantan Tengah, Provinsi Kalimantan Selatan, Provinsi Kalimantan Timur, Provinsi Kalimantan Utara, Provinsi Sulawesi Utara, Provinsi Sulawesi Tengah, Provinsi Gorontalo, Provinsi Sulawesi Barat, Provinsi Maluku, Provinsi Maluku Utara, Provinsi Papua Barat dan Provinsi Papua
2	3	Provinsi Jawa Barat, Provinsi Jawa Tengah dan Provinsi Jawa Timur
3	10	Provinsi Aceh, Provinsi Sumatera Utara, Provinsi Sumatera Barat, Provinsi Riau, Provinsi Sumatera Selatan, Provinsi Lampung, Provinsi Banten, Provinsi Nusa Tenggara Timur, Provinsi Sulawesi Selatan dan Provinsi Sulawesi Tenggara

TABEL 6, menunjukkan cluster yang terbentuk serta jumlah anggota masing-masing cluster setelah dikelompokkan dengan K-Means clustering. Anggota Provinsi yang masuk ke dalam satu cluster merupakan wilayah yang memiliki sifat-sifat yang homogen berdasarkan data yang ada. Cluster 1 terdiri atas 21 anggota Provinsi, Cluster 2 terdiri atas 3 anggota Provinsi, dan Cluster 3 terdiri atas 10 anggota Provinsi.

Berikut ini adalah tampilan hasil pengelompokkan Self Organizing Maps (SOM) dari hasil pengelompokkan yang diolah menggunakan software R:



GAMBAR 3. Visualisasi *Self Organizing Maps* (SOM)

Berdasarkan GAMBAR 3 dapat dilihat bahwa terbentuk 3 *cluster* dengan *cluster* pertama diasosiasikan dengan lingkaran yang berwarna merah, *cluster* kedua diasosiasikan dengan lingkaran yang berwarna putih dan *cluster* ketiga diasosiasikan dengan lingkaran yang berwarna biru. Untuk lebih jelasnya, anggota masing-masing *cluster* dapat dilihat pada TABEL 7 di bawah ini:

TABEL 7. Hasil *Self Organizing Maps* (SOM)

Cluster	Jumlah	Anggota
1	31	Provinsi Aceh, Provinsi Sumatera Utara, Provinsi Sumatera Barat, Provinsi Riau, Provinsi Jambi, Provinsi Sumatera Selatan, Provinsi Bengkulu, Provinsi Lampung, Provinsi Kepulauan Bangka Belitung, Provinsi Kepulauan Riau, Provinsi DKI Jakarta, Provinsi DI Yogyakarta, Provinsi Banten, Provinsi Bali, Provinsi Nusa Tenggara Barat, Provinsi Nusa Tenggara Timur, Provinsi kalimantan Barat, Provinsi kalimantan Tengah, Provinsi kalimantan Selatan, Provinsi kalimantan Timur, Provinsi kalimantan Utara, Provinsi Sulawesi Utara, Provinsi Sulawesi Tengah, Provinsi Sulawesi Selatan, Provinsi Sulawesi Tenggara, Provinsi Gorontalo, Provinsi Sulawesi Barat, Provinsi Maluku, Provinsi Maluku Utara, Provinsi Papua Barat dan Provinsi Papua.
2	1	Provinsi Jawa Barat
3	2	Provinsi Jawa Jawa Tengah dan Provinsi Jawa Timur

TABEL 7, menunjukkan *cluster* yang terbentuk serta jumlah anggota masing-masing *cluster* setelah dikelompokkan dengan *Self Organizing Maps* (SOM). *Cluster* 1 terdiri atas 31 anggota Provinsi, *cluster* 2 terdiri atas 1 anggota Provinsi dan *cluster* 3 terdiri atas 2 anggota Provinsi.

Selanjutnya menentukan metode terbaik menggunakan nilai *Davies Bouldin Index* (DBI). semakin kecil nilai DBI yang didapatkan dapat disimpulkan bahwa *cluster* yang diperoleh dari pengelompokkan menggunakan algoritma pengklasteran semakin baik.

TABEL 8. Nilai DBI

Metode	Nilai DBI
<i>K-Means Clustering</i>	0,6613664
<i>Self-Organizing Maps</i> (SOM)	0,1829366

Berdasarkan TABEL 8, terlihat bahwa pengelompokkan dengan *Self Organizing Maps* (SOM) memberikan nilai DBI yang lebih kecil dibandingkan dengan pengelompokkan menggunakan *K-Means Clustering*. Hal ini menunjukkan bahwa metode *Self Organizing Maps* (SOM) memberikan hasil pengelompokkan yang lebih baik dari metode *K-Means Clustering* dalam hal ini untuk pengelompokkan data potensi desa (Usaha Industri).

Selanjutnya dilakukan interpretasi hasil pengelompokkan menggunakan *Self Organizing Maps* (SOM).

TABEL 9. Nilai rata-rata tiap cluster

Variabel	Cluster 1	Cluster 2	Cluster 3
Industri Kulit	33,7	670	741

Industri Furnitur	832	4601	6965
Industri Barang Logam	377	2160	3737
Industri Tekstil	249	1264	1798
Industri Pakaian Jadi	492	3617	4968
Industri Gerabah	432	1772	2932
Industri Kayu	577	3180	4014
Industri Makanan	897	4247	6555
Industri Minuman	786	4387	5528
Industri Pengolahan Tembakau	19,7	171	834
Industri Kertas	6,90	148	212
Industri Percetakan dan reproduksi media rekaman	163	1486	2440
Industri Alat Angkutan	83,8	78	110
Industri Kerajinan	96,5	856	1100
Reparasi dan Pemasangan Mesin	232	1684	2656

TABEL 9, menunjukkan nilai rata-rata variabel pada tiap *cluster*. Berdasarkan tabel tersebut, hasil pengelompokkan diinterpretasikan sebagai berikut:

- Cluster* pertama terdiri dari daerah-daerah dengan nilai industri kulit, industri furnitur, industri Barang logam, industri tekstil, industri Pakaian jadi, industri gerabah, industri kayu, industri makanan, industri minuman, industri pengolahan tembakau, industri kertas, industri percetakan dan reproduksi media rekaman, industri alat angkutan, industri kerajinan, reparasi dan pemasangan mesin yang lebih rendah jika dibandingkan dengan *cluster 2* dan *cluster 3*. dimana, nilai potensi desa terendah yaitu usaha industri kulit. Nilai yang rendah tersebut dapat diartikan bahwa daerah pada *cluster 1* memiliki potensi desa (usaha industri) yang lebih sedikit dibandingkan *cluster* kedua dan *cluster* ketiga.
- Cluster* kedua terdiri dari daerah-daerah dengan nilai industri kulit, industri furnitur, industri Barang logam, industri tekstil, industri Pakaian jadi, industri gerabah, industri kayu, industri makanan, industri minuman, industri pengolahan tembakau, industri kertas, industri percetakan dan reproduksi media rekaman, industri kerajinan, reparasi dan pemasangan mesin yang lebih tinggi jika dibandingkan dengan *cluster 1* akan tetapi rendah jika dibandingkan dengan *cluster 3*. Namun, memiliki industri alat angkutan yang rendah jika dibandingkan dengan *cluster 1*. Nilai tersebut dapat diartikan bahwa daerah pada *cluster 2* memiliki potensi desa (usaha industri) yang sedang.
- Cluster* ketiga terdiri dari daerah-daerah dengan nilai industri kulit, industri furnitur, industri Barang logam, industri tekstil, industri Pakaian jadi, industri gerabah, industri kayu, industri makanan, industri minuman, industri pengolahan tembakau, industri kertas, industri percetakan dan reproduksi media rekaman, industri alat angkutan, industri kerajinan, reparasi dan pemasangan mesin yang lebih tinggi jika dibandingkan dengan *cluster 1* dan *cluster 2*. Nilai yang tinggi tersebut dapat diartikan bahwa daerah pada *cluster 3* memiliki potensi desa (usaha industri) yang lebih banyak dibandingkan *cluster* pertama dan *cluster* kedua.

KESIMPULAN DAN SARAN

Berdasarkan analisis dan pembahasan di atas didapatkan bahwa Pengelompokan Provinsi berdasarkan data potensi desa (usaha industri) menggunakan metode *Self Organizing Maps* (SOM) memberikan hasil yang lebih baik dibandingkan dengan pengelompokkan menggunakan metode *K-Means clustering* berdasarkan evaluasi menggunakan *Davies Bouldin Index* (DBI), dimana nilai DBI yang didapatkan yaitu 0,1829366. Hasil pengelompokkan provinsi di Indonesia berdasarkan potensi desa (usaha industri) terbentuk menjadi 3 *cluster*, dimana *cluster 1* terdiri atas 31 anggota provinsi, *cluster 2* terdiri atas 1 anggota Provinsi dan *cluster 3* terdiri atas 2 anggota Provinsi.

Adapun saran yang dapat penulis berikan adalah diharapkan pemerintah dapat memberikan perhatian serta mengembangkan potensi desa di daerah-daerah yang memiliki potensi desa yang sedikit agar usaha industri di Indonesia merata sehingganya dapat meningkatkan kualitas Sumber Daya

Manusia, Peneliti selanjutnya dapat menggunakan metode *clustering* lainnya seperti *Partitioning Around Medoids* (PAM) dan *Fuzzy C-Means*, peneliti selanjutnya dapat menggunakan penentuan jumlah *K cluster* optimal seperti metode *Silhouette* dan metode *gap statistic*, dan peneliti selanjutnya dapat menggunakan metode validasi lain untuk menentukan metode terbaik seperti *dunn index*.

REFERENSI

Andiani, D., Septiani, S.D.R. and Riana, A. (2022) 'Analisis Teknik non-Hierarki untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Berdasarkan Indikator Kesejahteraan Rakyat 2020', *Jurnal Riset Matematika dan Sains Terapan*, 2(1), pp. 21–28.

Badan, P.S. (2022) *Statistik Potensi Desa Indonesia*. Jakarta: Badan Pusat Statistik.

Dacwanda, D.O. and Nataliani, Y. (2021) 'Implementasi K-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan', *Jurnal Teknologi Informasi*, 18(2), pp. 125–138.

Fawaz, Fitriyani, N.S. and Rosalia, A.A. (2022) 'Perbandingan Algoritma Self Organizing Map dan Fuzzy C-Means dalam clustering hasil produksi ikan PPN Karangantu', *Jurnal Sistem Informasi dan Telematika (Telekomunikasi, Multimedia, dan Informatika)*, 13(2), pp. 102–109. Available at: <https://doi.org/10.36448/jsit.v13i2.2783>.

Hidayatin, I., Adinugroho, S. and Dewi, C. (2019) 'Pengelompokan Wilayah berdasarkan Penyandang Masalah Kesejahteraan Sosial (PMKS) dengan Optimasi Algoritme K-Means menggunakan Self Organizing Map (SOM)', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JPTIIK)*, 3(8), pp. 7524–7531. Available at: <http://j-ptiik.ub.ac.id>.

Imani, N., Alfassa, A.I. and Yolanda, A.M. (2022) 'Self Organizing Maps (SOM) Clustering untuk Analisis Data Indikator Sosial di Provinsi Nusa Tenggara Timur', *Jurnal Gaussian*, 11(3), pp. 458–467. Available at: <https://doi.org/10.14710/j.gauss.11.3.458-467>.

Kapita, S. *et al.* (2022) 'Penerapan Algoritma Clustering Khonen-SOM dengan Validasi Davies Bouldin Index pada Pengelompokan Potensi Udag di Indonesia', *Indonesian Journal on Information System*, 7(2), pp. 134–143.

Kapita, S.N. and Abdullah, S. Do (2020) 'Pengelompokan Data Mutu Sekolah Menggunakan Jaringan Syaraf Tiruan Kohonen-SOM', *Jurnal Informatika dan Komputer*, 3(1), pp. 56–61. Available at: <https://doi.org/10.33387/jiko>.

Matdoan, M. and Delsen, M.S.N. (2020) 'Penerapan Analisis Cluster dengan Metode Hierarki untuk Klasifikasi Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia', *STATMAT (Jurnal Statistika dan Matematika)*, 2(2), pp. 123–130.

Muningsih, E., Maryani, I. and Handayani, V.R. (2021) 'Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa', *Jurnal Sains dan Manajemen*, 9(1), pp. 95–100. Available at: <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839>.

Setiadi, A. and Sikumbang, E.D. (2020) 'K-Means Clustering dalam Penerimaan Karyawan Baru', *Informatics for Educators and Professionals*, 4(2), pp. 103–112.

Soleh, A. (2017) 'Strategi Pengembangan Potensi Desa', *Jurnal Sungkai*, 5(1), pp. 32–52.

Solichin, A. and Khairunnisa, K. (2020) 'Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta', *Fountain of Informatics Journal*, 5(2), pp. 52–59.

Zubedi, F. (2020) 'Analisis Cluster Hasil Try Out Siswa MTS Alhuda Gorontalo dengan Chi-Sim Cosimilarity dan K-Means Clustering', *Matematika dan Pendidikan Matematika*, 5(1), pp. 1–11.