Neptunus : Jurnal Ilmu Komputer Dan Teknologi Informasi
Vol. 2, No.3 Agustus 2024
e-ISSN : 3031-898X, dan p-ISSN : 3031-8998 Hal 408-418
DOI: https://doi.org/10.61132/neptunus.v2i3.320
Available online at: https://journal.arteii.or.id/index.php/Neptunus

Penerapan Algoritma *Self Orginizing Mapping* untuk Analisis *Clustering* Indeks Pembangunan Manusia di Jawa Barat

Ahmad Fauzi[1*], Eko Harli[2], Tria Hadi Kusmanto[3]
[123]Program Studi Teknik Informatika, Universitas Indraprasta PGRI Jakarta, Indonesia

Alamat: Jl. Raya Tengah No.80, RT.6/RW.1, Gedong, Kec. Ps. Rebo, Kota Jakarta Timur
Korespondensi penulis: ahmadfauzi.udzi@gmail.com

Abstract. *The importance of the Human Development Index (HDI) as a key indicator of development success, encompassing health, education, and standard of living dimensions, is crucial for economic growth. Jawa Barat Province, as one of the provinces with a high HDI value, still needs to understand the characteristics of its cities/regencies based on the HDI components for more targeted development planning. The SOM algorithm is capable of clustering data effectively. The results of the study indicate that Jawa Barat Province has 2 clusters based on the HDI components.*
Keywords: *Human Development Index; Self-Orginizing Map, Cluster*

Abstrak. Pentingnya IPM sebagai indikator utama keberhasilan pembangunan yang mencakup dimensi kesehatan, pendidikan, dan standar hidup menjadi salah satu hal yang penting dalam pertumbuhan ekonomi. Propinsi Jawa Barat sebagai salah satu propinsi dengan nilai IPM tinggi tetap perlu untuk mengetahui karakteristik Kota/Kabupaten berdasarkan komponen pembentuk IPM sehingga dapat diidentifikasi untuk perencanaan pembangunan yang lebih tepat sasaran. Algoritma SOM mampu dalam melakukan klasterisasi data dengan baik. Hasil penelitian menunjukan bahwa di Propinsi Jawa Barat memiliki 2 cluster berdasarkan komponen-komponen IPM.
Kata kunci: Indeks Pembangunan Manusia, Self-Organizing Map, Klaster

1. LATAR BELAKANG

Indeks Pembangunan Manusia (IPM) merupakan parameter keberhasilan upaya peningkatan keterampilan penduduk yang dikembangkan dengan memperhatikan tiga aspek yaitu sehat dan panjang umur, kualitas pendidikan, dan hidup dengan layak. Selain itu juga IPM mendefinisikan apakah suatu negara termasuk kategori negara yang maju, negara sedang berkembang atau negara yang masih terbelakang (Azfirmawarman et al., 2023).

Meningkatnya perhatian terhadap IPM mengindikasikan perlunya strategi pembangunan yang lebih terarah dan efisien. Hal ini dikarenakan IPM merupakan salah satu variabel yang signifikan dalam pertumbuhan ekonomi (Tesalonika, 2022).

Jawa Barat sendiri sebagai salah satu propinsi dengan nilai IPM tinggi di Indonesia yaitu sebesar 74,24 pada tahun 2023 naik 0,83% dari tahun sebelumnya dengan rataan pertumbuhan IPM pertahun dari tahun 2022 sampai dengan 2023 sebesar 0,74%. Dari data tersebut dapat dilihat karakteristik dari setiap Kota/Kabupaten yang ada di Jawa

Received: Juni 20, 2024; Revised: Juli 25, 2024; Accepted: Agustus 03, 2024; Online Available: Agustus 07, 2024
* Ahmad Fauzi, *ahmadfauzi.udzi@gmail.com*

Barat untuk melihat keterkaitan antar Kota/Kabupaten. Salah satu metode analisis yang dapat digunakan untuk memahami struktur data IPM adalah pengelompokan (*clustering*) menggunakan *Self-Organizing Map* (SOM) (Fatihin & Andayani, 2023). Metode ini dapat membantu mengidentifikasi pola dan hubungan tersembunyi dalam data IPM yang mungkin tidak terlihat dengan metode analisis tradisional.

Penelitian terkait penggunaan SOM dalam analisis data IPM telah dilakukan dengan berbagai fokus, termasuk optimalisasi algoritma dan interpretasi hasil *cluster*. Misalnya, penelitian SOM untuk mengelompokkan data IPM di Propinsi Kalimantan Barat dimana dari penelitian tersebut terdapat 4 *cluster*, dimana nilai Rata-rata Lama Sekolah (RLS) dan jumlah murid mempengaruhi jumlah anggota setiap cluster (Midyanti & Bahri, 2023).

Terdapat juga penelitian yang mengkaji penggunaan SOM untuk menganalisa data indikator sosial Propinsi Nusa Tenggara Timur, dari hasil penelitian didapat algoritma SOM dapat dengan baik mengelompokkan setiap Kota/Kabupaten yang memiliki karaktersistik yang sama kedalah 4 buah cluster (Imani et al., 2023).

Meskipun banyak penelitian telah mengeksplorasi penggunaan SOM untuk analisis data IPM, terdapat beberapa celah yang belum banyak dieksplorasi. Pertama, penelitian sebelumnya cenderung fokus pada analisis di tingkat makro tanpa memperhatikan variasi pada komponen-komponen spesifik IPM. Kedua, masih sedikit studi yang mengintegrasikan analisis SOM dengan kebijakan pembangunan lokal yang spesifik. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan menganalisis IPM menurut komponennya dan mengidentifikasi kelompok-kelompok daerah berdasarkan kesamaan dalam setiap komponen IPM. Urgensi penelitian ini terletak pada kebutuhan untuk mengembangkan strategi pembangunan yang lebih tepat sasaran berdasarkan karakteristik unik dari setiap kelompok daerah.

Penelitian ini bertujuan untuk mengelompokkan daerah berdasarkan komponen-komponen IPM menggunakan metode SOM dan mengidentifikasi karakteristik unik dari setiap kelompok daerah yang terbentuk.

2. KAJIAN TEORITIS

Clustering

Clustering merupakan proses mempartisi sekumpulan objek data menjadi subset. Subset adalah *cluster*, sehingga objek dalam *cluster* yang sama, memiliki karakteristik yang mirip antara satu sama lain, namun berbeda dengan objek di *cluster* lain (Han & Kamber, 2006). *Clustering* berbeda dari klasifikasi, jika klasifikasi memiliki target dalam pengolahan data, lain hal dalam *clustering*, karena tidak ada variabel target untuk *clustering* (Larose & Larose, 2015).

Self Orginizing Map (SOM)

Self-Organizing Map (SOM) adalah salah satu jenis jaringan saraf tiruan yang diperkenalkan oleh Teuvo Kohonen pada tahun 1982. SOM bekerja dengan memetakan data berdimensi tinggi ke dalam peta dua dimensi yang memudahkan visualisasi dan interpretasi klaster (biasanya berupa *grid* 2 dimensi) (Kusumadewi & Purnomo, 2004). Maka dapat disimpulkan bahwa SOM mengurangi dimensi data dan menampilkan kesamaan antar data (Anis & Isnanto, 2014).

SOM merupakan salah satu metode data *mining* yang termasuk kedalam tipe *Unsupervised Learning* (Frigui, 2008), dimana Algoritma *Unsupervised Learning* bersifat deskriptif, yang akan berguna untuk mengelompokkan atau mengkategorikan data (Abijono et al., 2021).

Tahapan SOM

Langkah-langkah algoritma SOM dapat dijabarkan sebagai berikut (Klir & Yuan, 2015):

1. Inisialisasi vektor input x_1, x_2, x_3, ..., x_n
2. Inisialisasi *neuron* output sebanyak y_1, y_2, y_3, ..., y_n
3. Menentukan *weight* (bobot) neuron output dengan nilai antara x_{min} dan x_{max}
4. Mengulangi langkah 5 sampai 8 hingga tidak ada *update weight* (bobot) atau telah mencapai kondisi stop (*error* terkecil).
5. Pemilihan acak salah satu data dari vektor input sebagai data training.
6. Mencari jarak terdekat dari masing-masing neuron output ke data input menggunakan rumus *euclidian distance*.

$$jarak = \sqrt{\sum_{i=1}^{10} (data[i] - bobot[i])^2}$$

Dari seluruh bobot (D_i) dicari yang paling kecil jaraknya, indeks dari bobot (D_i) ini disebut winning neuron.

7. Untuk setiap bobot w_{ij} diperbaharui bobot tetangga menggunakan rumus dengan persamaan sebagai berikut :

$$w_{ij}(t+1) = w_{ij}(t) + a(t)[x_i - w_{ij}(t)]$$

8. Mengupdate bobot bias (*error*).
9. Simpan bobot yang telah konvergen.

3. METODE PENELITIAN

Penelitian ini menggunakan metode deskriptif-kuantitatif dengan pendekatan *clustering* menggunakan SOM. Metode deskriptif-kuantitatif adalah pendekatan penelitian yang digunakan untuk menggambarkan karakteristik dari populasi atau fenomena yang sedang diteliti dengan menggunakan data kuantitatif (Jayusman & Shavab, 2020). Pendekatan ini menggabungkan deskripsi mendetail dari data yang dikumpulkan dengan analisis statistik untuk memberikan gambaran yang komprehensif tentang objek penelitian. Tujuan utama dari metode ini adalah untuk mendapatkan gambaran yang akurat dan sistematis mengenai fenomena atau populasi yang diteliti tanpa membuat kesimpulan kausalitas (Akbar et al., 2023). Pendekatan ini dipilih karena mampu mengelompokkan data yang kompleks dan mengidentifikasi pola-pola tersembunyi dalam data IPM.

Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari laporan tahunan BPS dan publikasi resmi lainnya yang memuat data IPM dan komponennya. Instrumen pengumpulan data adalah lembar pengumpulan data yang disusun berdasarkan komponen-komponen IPM, yaitu:

1. Angka harapan hidup untuk dimensi kesehatan.
2. Rata-rata lama sekolah dan harapan lama sekolah untuk dimensi pendidikan.
3. Pengeluaran per kapita untuk dimensi standar hidup.

Populasi dalam penelitian ini adalah seluruh Kota/Kabupaten di Propinsi Jawa Barat yang memiliki data IPM. Data IPM diambil dari publikasi resmi Badan Pusat Statistik (BPS) Propinsi Jawa Barat pada tahun 2023.

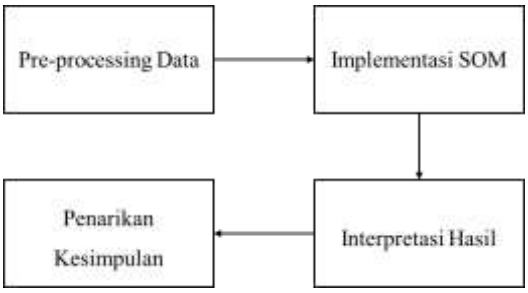
Teknik Analisis Data

Analisis data dilakukan menggunakan metode SOM, yang merupakan jenis jaringan syaraf tiruan (*Artificial Neural Network*) yang digunakan untuk *clustering* dan visualisasi data. Langkah-langkah analisis meliputi:

1. Normalisasi data komponen IPM untuk memastikan skala yang sama.
2. Implementasi algoritma SOM menggunakan Bahasa Pemrograman R.
3. Validasi hasil *clustering* menggunakan *silhouette score* untuk memastikan kualitas *clustering* yang dihasilkan.
4. Interpretasi hasil *clustering* untuk mengidentifikasi karakteristik unik dari setiap kelompok daerah.

Model Penelitian

Model penelitian yang digunakan dalam penelitian ini adalah model *clustering* dengan SOM, yang terdiri dari beberapa tahap. Gambar 1 manampilkan tahapan dari penelitian:



Gambar 1 Model Penelitian

4. HASIL DAN PEMBAHASAN

Pre Processing Data

Proses ini menghasilkan data bersih, dimana data-data yang tidak diperlukan pada penelitian ini dihilangkan, selain itu juga atribut-atribut yang tidak diperlukan juga dihilangkan sehingga data yang tersedia menjadi bersih. Atribut yang digunakan

merupakan komponen pembentuk IPM seluruh Kota/Kabupaten di Propinsi Jawa Barat tahun 2023. Tabel 1 menunjukan sebagian data hasil *pre-processing*.

Tabel 1. Sebagian hasil *pre-processing*

Kota/Kabupaten	UHH	HLS	RLS	PPK
Bogor	71,92	12,64	8,37	11153
Sukabumi	71,83	12,38	7,33	9482
Cianjur	70,79	12,03	7,22	8626
Bandung	74,27	12,73	9,1	11018

Implementasi SOM

Penelitian ini menggunakan bahasa pemrograman R dalam melakukan analisis data dengan *clustering* SOM, berikut adalah hasil dari setiap langkah-langkah implementasi:

1. Ringkasan Data

Langkah ini digunakan untuk mengetahui secara deskriptif bentuk data yang digunakan, gambar menunjukan hasil ringkasan data yang dimaksud.

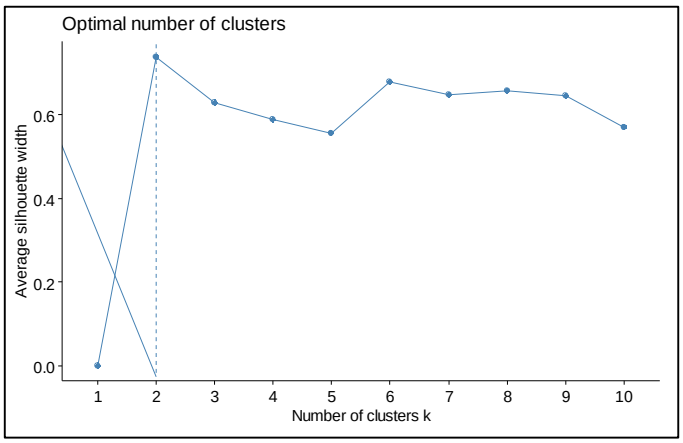
```
     Kota                UHH              HLS              RLS              PPK
 Length:27          Min.   :70.19   Min.   :11.91   Min.   : 6.940   Min.   : 8562
 Class :character   1st Qu.:72.00   1st Qu.:12.26   1st Qu.: 7.865   1st Qu.: 9880
 Mode  :character   Median :72.92   Median :12.64   Median : 8.230   Median :11136
                    Mean   :73.00   Mean   :12.88   Mean   : 8.870   Mean   :11526
                    3rd Qu.:74.28   3rd Qu.:13.47   3rd Qu.: 9.970   3rd Qu.:12449
                    Max.   :75.79   Max.   :14.29   Max.   :11.660   Max.   :18236
```

Gambar 2. Ringkasan Data

Dari hasil tersebut didapat jumlah data yang digunakan sebayak 27 data dengan nilai usia harapan hidup (UHH) terbesar adalah 75.79, nilai harapan lama sekolah (HLS) terbesar adalah 14.29, nilai rata-rata lama sekolah (RLS) adalah 11.66 dan pengeluaran perkapita (PPK) terbesar adalah Rp. 18.236.

2. Validasi *Cluster*

Validasi *cluster* dilakukan untuk menentukan jumlah *cluster* dengan menggunakan 3 metode validasi yaitu *Dunn*, *Silhoutte*, dan *Connectivity*. Dari ketiga metode tersebut jumlah *cluster* yang digunakan adalah dilihat dari nilai *Connectivity* yang paling rendah, jumlah *cluster* dengan Indeks *Dunn* paling besar, dan nilai *silhoutte* mendekati 1. Gambar merupakan hasil dari validasi *Silhoutte*.



Gambar 3. Hasil *Silhouette Coefficient*

Silhouette Coefficient digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik atau buruknya suatu obyek ditempatkan dalam suatu *cluster*. Metode ini merupakan gabungan dari metode separasi dan kohesi (Ayu et al., 2019) (Kodinariya & Makwana, 2013).

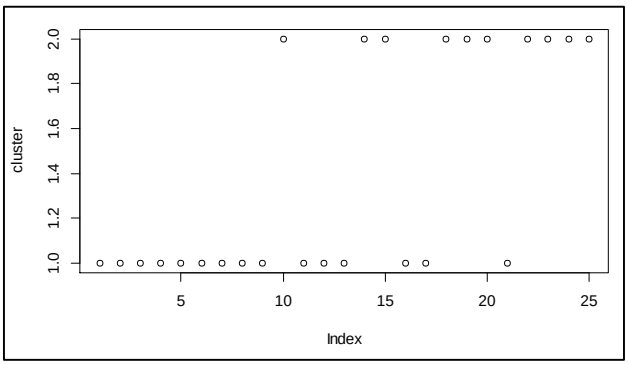
```
Validation Measures:
                          2       3       4       5

som Connectivity     4.2869 10.5948 16.3921 22.0877
    Dunn             0.7859  0.1042  0.1339  0.1369
    Silhouette       0.7378  0.6351  0.6376  0.6069

Optimal Scores:

             Score  Method Clusters
Connectivity 4.2869 som    2
Dunn         0.7859 som    2
Silhouette   0.7378 som    2
```

Gambar 4. Hasil Validasi *Cluster*

Berdasarkan output *summary*(*cluster*) diatas, diperoleh nilai indeks *connectivity* paling rendah pada *cluster* 2 yaitu 4.2869, dengan nilai indeks *dunn* paling besar yaitu 0.7859 pada *cluster* 2, dan nilai dari indeks *silhoutte* dengan nilai yang paling mendekati 1 yaitu 0.7378 pada cluster 2. Sehingga pada analisis kali ini menggunakan jumlah *cluster* 2 dalam teknik *clustering* kota berdasarkan komponen IPM.

3. *Clustering* SOM

Pemodelan SOM pada penelitian ini menggunakan grid som 5x5 dengan topologi heksagonal. Keuntungan menggunakan topologi ini adalah interaksi yang lebih halus dan visualisasi yang lebih efektif (Vesanto & Alhoniemi, 2000). Hasil pelatihan SOM kemudian dilakukan *cutree* untuk memotong pohon hierarki pada tingkat tertentu untuk mendapatkan *cluster*.

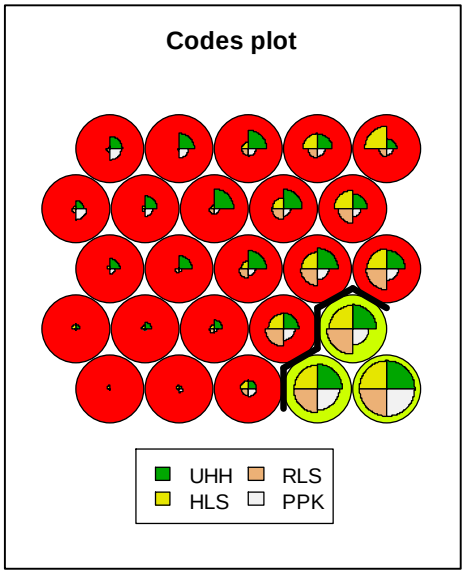


Gambar 5. Hasil *Cutree*

Gambar 5 menunjukan hasil *cutree* dimana pada hasil pelatihan terdapat 10 lingkaran pada cluster 2 dan 15 lingkaran pada cluster 1. Hasil ini sesuai dengan dimensi 5x5 dari pelatihan yang dilakukan yaitu sejumlah 25 lingkaran.

4. Visualisasi Data

Visualisasi data hasil pelatihan SOM dengan menggunakan diagram Kipas seperti yang terlihat pada gambar 5 dibawah ini.



Gambar 6. Diagram Kipas Hasil Implementasi SOM

5. Hasil Pemodelan

Dengan hasil validasi *clustering* yang berjumlah 2 *cluster*, terdapat 23 Kota/Kabupaten masuk termasuk kedalam *cluster* 1 dan 4 Kota/Kabupaten termasuk kedalam *cluster* 2. Tabel 2 menunjukan hasil lengkap dari pemodelan SOM.

Tabel 2. Hasil Clustering

Cluster	Kabupaten/Kota	Jumlah Anggota
1	Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Tasikmalaya, Kota Banjar	23
2	Kota Bandung, Kota Bekasi, Kota Depok, Kota Cimahi	4

Untuk melihat karakterisitik atau ciri khas dari setiap *cluster* perlu dilakukan profilisasi *cluster* dengan cara menghitung rataan dari setiap *cluster*. Tabel menunjukan rataan dari setiap *cluster*.

Tabel 3. Rataan Komponen IPM Setiap Cluster

Cluster	UHH	HLS	RLS	PPK
1	72,62	12,68	8,43	10.753,04
2	75,21	14,04	11,42	15.969,25

UHH = Usia Harapan Hidup
HLS = Harapan Lama Sekolah
RLS = Rata-rata Lama Sekolah
PPK = Pengeluaran Per Kapita

Dari tabel tersebut terlihat bahwa *cluster* 2 memiliki karakteristik setiap komponen IPM lebih tinggi dari *cluster* 1. Selain itu komponen rata-rata lama sekolah pada *cluster* 1 dapat menjadi perhatian khusus karena memiliki nilai cukup rendah. Hal ini dapat menjadi fokus penting agar perlunya perhatian khusus dalam bidang pendidikan di Propinsi Jawa Barat, terutama pada Kota/Kabupaten di *cluster* 1.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan, dapat disimpulkan bahwa algoritma SOM dapat melakukan *clustering* terhadap data IPM menurut komponen di Propinsi Jawa Barat dengan hasil *cluster* yaitu 2 *cluster* dengan nilai *silhouette* 0,7378. Selain itu dapat disimpulkan bahwa terdapat cukup besar perbedaan karakteristik antara 2 cluster tersebut, dimana jumlah anggota tiap cluster berbeda jauh cukup siginifikan. Hal ini menunjukan bahwa pemerataan baik dari sisi pendidikan, ekonomi dan kesehatan masih ada yang belum maksimal di beberapa daerah. Penelitian kedepannya dapat menambahkan

beberapa atribut lainnya diluar IPM untuk melihat lebih jauh lagi karakteristik tiap daerah di Jawa Barat.

DAFTAR REFERENSI

Abijono, H., Santoso, P., & Anggreini, N. L. (2021). ALGORITMA SUPERVISED LEARNING DAN UNSUPERVISED LEARNING DALAM PENGOLAHAN DATA. *G-Tech: Jurnal Teknologi Terapan*, *4*(2), 315–318. https://doi.org/10.33379/gtech.v4i2.635

Akbar, M. R., Magdalena, D., Septina, N., Sari, A. C., Susanti, O. I., Rohmah, Z., Septian, A., & Fajri, N. (2023). *Analisis Deskriptif Motivasi Belajar Siswa Melalui Model Belajar Kombinasi Pada Pandemi Covid-19*. *2*(5), 946–951.

Anis, Y., & Isnanto, R. R. (2014). *Penerapan Metode Self-Organizing Map (SOM) Untuk Visualisasi Data Geospasial Pada Informasi Sebaran Data Pemilih Tetap (DPT)*. *01*, 48–57.

Ayu, D., Cahya, I., Ayu, D., & Pramita, K. (2019). *Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali*. *9*(3).

Azfirmawarman, D., Magriasti, L., & Yulhendri, Y. (2023). Indeks Pembangunan Manusia Di Indonesia. *Jurnal Pendidikan Dan Konseling (JPDK)*, *5*(5), 117–125. https://doi.org/10.31004/jpdk.v5i5.22864

Fatihin, C., & Andayani, S. (2023). *DISTRICT / CITY CLUSTERING IN EAST JAVA BASED ON HUMAN*. *9*(November), 134–147.

Frigui, H. (2008). Clustering: Algorithms and Applications. *2008 First Workshops on Image Processing Theory, Tools and Applications*, 1–11. https://doi.org/10.1109/IPTA.2008.4743793

Han, J., & Kamber, M. (2006). *Data Mining. Concepts and Techniques* (2nd ed.). Morgan Kaufmann. http://www.amazon.de/Mining-Concepts-Techniques-Kaufmann-Management/dp/1558609016%3FSubscriptionId%3D13CT5CVB80YFWJEPWS02%26tag%3Dws%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D1558609016

Imani, N., Alfassa, A., & Yolanda, A. (2023). ANALISIS CLUSTER TERHADAP INDIKATOR DATA SOSIAL DI PROVINSI NUSA TENGGARA TIMUR MENGGUNAKAN METODE SELF ORGANIZING MAP (SOM). *Jurnal Gaussian*, *11*(3), 458–467.

Jayusman, I., & Shavab, O. A. K. (2020). STUDI DESKRIPTIF KUANTITATIF TENTANG AKTIVITAS BELAJAR MAHASISWA DENGAN MENGGUNAKAN MEDIA PEMBELAJARAN EDMODO DALAM PEMBELAJARAN SEJARAH. *Jurnal Arte*, *7*(1), 13–20.

Klir, G. J., & Yuan, B. (2015). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall.

Kodinariya, T. M., & Makwana, P. R. (2013). *Review on determining number of Cluster in K-Means Clustering*. https://api.semanticscholar.org/CorpusID:10090179

Kusumadewi, S., & Purnomo, H. (2004). *Aplikasi logika fuzzy untuk pendukung keputusan*. Graha Ilmu.

Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics* (2nd ed.). Wiley Publishing.

Midyanti, D. M., & Bahri, S. (2023). *IMPLEMENTASI SELF ORGANIZING MAPS UNTUK PENGELOMPOKAN KABUPATEN / KOTA BERDASARKAN INDEKS PEMBANGUNAN MANUSIA IMPLEMENTATION OF SELF-ORGANIZING MAPS FOR DISTRICT / CITY GROUPING BASED ON THE HUMAN DEVELOPMENT INDEX*. *10*(6), 1265–1272. https://doi.org/10.25126/jtiik.2023107647

Tesalonika, A. (2022). *Analisis Pengaruh Tingkat Pengangguran dan Indeks Pembangunan Manusia terhadap Pertumbuhan Ekonomi di Provinsi Jawa Barat Periode Tahun 2017-2021*. *1*(4), 78–89.

Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, *11*(3), 586–600. https://doi.org/10.1109/72.846731

Laporan Instansi/Lembaga/Organisasi/Perusahaan

LPPSP. (2016). *Statistik Indonesia 2016*. Badan Pusat Statistik, 676. Jakarta. Diakses dari https://www.LPPSP.go.id/index.php/publikasi/326.