



\*Corresponding author: Zulkifli Rais,  
Department of Statistics, Faculty of  
Mathematical and Natural Science,  
University Negeri Makassar, Makassar,  
Indonesia

E-mail: [zulkifli.rais89@unm.ac.id](mailto:zulkifli.rais89@unm.ac.id)

## RESEARCH ARTICLE

# Applied the Self-Organizing Maps (SOM) Method for Clustering Educational Equity in South Sulawesi

Andi Restu Gunawan, Sudarmin, & Zulkifli Rais\*

Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia

**Abstract:** This research aims to group regencies/cities based on education indicators and identify the characteristics of each group formed based on education indicators. The method used in this research is Self self-organizing map (SOM). SOM is an artificial neural network that requires no assumptions and a method that produces a representation of the input space from low-dimensional training samples. The data used in this research are 9 variables regarding pure enrollment rates, gross enrollment rates, and student-to-teacher ratios at each level of education in 24 districts/cities in South Sulawesi in 2020-2021 which come from BPS publications. Based on the results obtained, 4 clusters were formed, each of which had its characteristics. The clusters formed include Cluster 1 consisting of 7 regencies/cities, cluster 2 consisting of 10 regencies/cities, cluster 3 consisting of 4 regencies/cities, and Cluster 4 consisting of 2 regencies. Based on the results of cluster validation using the Dunn index, 4 optimal clusters were obtained with a value of 0.42.

**Keywords:** Clusters, Self Organizing Maps, Dunn Index, Education.

## 1. Pendahuluan

Jaringan saraf tiruan (JST) merupakan suatu konsep rekayasa pengetahuan dalam bidang kecerdasan buatan yang didesain dengan mengadopsi sistem saraf manusia, yang pemrosesan utamanya ada pada otak dan bagian terkecil pada otak adalah sel saraf yang disebut pemroses informasi atau neuron. Menurut (Kapita dkk, 2020) JST adalah mekanisme komputasi yang mampu memperoleh, mewakili, menghitung pemetaan dari satu ruang multivariat ke informasi yang lain, dan diberi satu himpunan data yang mewakili pemetaan tersebut. Dalam penelitian ini akan dilakukan *clustering* pada wilayah kabupaten/kota di Sulawesi Selatan.

*Clustering* adalah proses pengelompokan data ke dalam beberapa *cluster* sehingga data-data di suatu *cluster* memiliki kemiripan maksimum. Menurut (Annisa dkk., 2022) *clustering* merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster* serta Objek yang di dalam *cluster* memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan *cluster* yang lain.

JST terdapat dua jenis metode pembelajaran yaitu *Supervised learning* dan *unsupervised learning*. Menurut (Retnoningsih & Pramudita, 2020) *Supervised learning* merupakan salah satu teknik *machine learning* yang menggunakan *dataset (data training)* yang sudah berlabel (*labeled data*) untuk melakukan pembelajaran pada mesin, sehingga mesin mampu mengidentifikasi label input dengan menggunakan fitur yang dimiliki untuk selanjutnya melakukan prediksi maupun klasifikasi, sedangkan *unsupervised learning* adalah teknik dengan menarik kesimpulan berdasarkan *dataset* yang merupakan input data *labeled response*.



Ada beberapa metode pembelajaran *Supervised learning* yang di kembangkan oleh beberapa peneliti diantaranya adalah *Decision Tree*, *K-Nearest Neighbor (KNN)*, *Naive Bayes*, *Regresi*, dan *Super Vector Machine* serta beberapa metode pembelajaran *Unsupervised learning* yang telah diciptakan oleh para peneliti, diantaranya adalah *DBSCAN*, *Fuzzy C-Means*, *K-Means*, dan *Self Organizing Maps* (Retnoningsih & Pramudita, 2020).

Penelitian ini akan di gunakan metode *Self-Organizing Map (SOM)*. *SOM* sendiri adalah JST yang dilatih menggunakan pembelajaran tanpa pengawasan atau *Unsupervised learning* untuk menghasilkan representasi berdimensi rendah (dua dimensi) yang dimana ruang input dari contoh-contoh pelatihan yang disebut dengan peta. Pada proses penerapan algoritma pembelajaran *SOM*, perlu dilakukan validasi cluster untuk menentukan apakah suatu model cluster tersebut baik sehingga dapat digunakan sebagai kesimpulan (Alkhalidi dkk., 2020).

Adanya pengelompokan dan klasifikasi yang terbentuk maka akan di peroleh informasi kabupaten/kota yang masih rendah berdasarkan indikator pendidikan. Salah satu persoalan dalam bidang pendidikan di Indonesia yaitu kurang meratanya pendidikan yang dimana hal tersebut di sebabkan oleh kurang memadainya informasi dan fasilitas yang sama antara kabupaten/kota di Sulawesi selatan.

## 2. Kajian Literatur

### 2.1. Analisis Cluster

Analisis *cluster* merupakan salah satu teknik analisis statistik dengan tujuan untuk memilah objek ke dalam *cluster* berdasarkan kesamaan karakteristik objek (Afrimayani dkk., 2019). Analisis *cluster* mengelompokkan objek (responden, produk, atau entitas lainnya) sehingga masing-masing objek mempunyai kemiripan dengan yang lain dalam suatu *cluster* (Rahayu & Hakim, 2019). Hasil *cluster* suatu objek harus memiliki internal (*within cluster*) homogenitas yang tinggi dan memiliki eksternal (*between cluster*) heterogenitas yang tinggi (Halim & Widodo, 2017).

Menurut (Halim & Widodo, 2017) konsep dasar pengukuran analisis *cluster* merupakan konsep pengukuran jarak (*distance*) serta kesamaan (*similarity*). Konsep pengukuran jarak merupakan ukuran tentang jarak pisah antar obyek, sebaliknya konsep pengukuran kesamaan merupakan berbicara tentang sesuatu ukuran kedekatan. Salah satu cara untuk mengukur jarak antar dua objek yang akan di gunakan dalam penelitian ini adalah *Euclidean Distance*.

*Euclidean distance* merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari 2 (dua) buah titik dalam *Euclidean space* (meliputi bidang euclidean dua dimensi, tiga dimensi, atau bahkan lebih) (Nishom, 2019). Untuk mengukur tingkat kemiripan data dengan rumus euclidean distance digunakan rumus berikut.

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

- d = jarak antara x dan y
- x = data pusat klaster
- y = data pada atribut
- i = setiap data
- n = jumlah data,
- $x_i$  = data pada pusat klaster ke  $i$
- $y_i$  = data pada setiap data ke  $i$

### 2.2. Artificial Neural Network

*Artificial Neural Network* atau Jaringan Saraf Tiruan adalah suatu paradigma pengolahan informasi yang termotivasi dari sistem kerja syaraf biologis, semacam kinerja otak, yang

memproses suatu informasi (Cynthia & Ismanto, 2017). Menurut (Raharjo dkk, 2020) jaringan syaraf tiruan dibentuk sebagai suatu generalisasi model matematika dari jaringan syaraf biologi, dengan asumsi bahwa :

- a. Pemrosesan informasi terjadi pada banyak elemen sederhana yang biasa disebut *neuron*
- b. Sinyal mengalir diantara *neuron-neuron* melalui suatu sambungan penghubung
- c. Setiap sambungan penghubung memiliki penghubung antar neuron memiliki bobot yang akan memperkuat atau memperlemah sinyal
- d. Untuk menentukan output, setiap neuron menggunakan fungsi aktivitas (biasanya bukan fungsi linear) yang dikenakan pada input yang diterima. Menurut (Cynthia & Ismanto, 2017) jaringan saraf tiruan, seperti manusia, belajar dari contoh ataupun pola. Sebuah jaringan syaraf tiruan merupakan hasil konfigurasi untuk aplikasi tertentu, seperti pengenalan pola atau klasifikasi data, melalui suatu proses pembelajaran. Proses pembelajaran dalam sistem biologis menambahkan penyesuaian koneksi sinaptik yang ada diantar neuron. Jaringan syaraf tiruan memiliki kemampuan untuk mengekstrak hubungan antara input atau output dari sebuah proses, tanpa adanya keberadaan fisik.

### 2.3. Self Organizing Maps

Teuvon Kohonen adalah seorang ilmuwan Finlandia yang memperkenalkan jaringan kohonen pada tahun 1982 (Amalia & Yunhasnawa, 2019). *SOM* adalah jenis jaringan saraf tiruan yang dipertahankan menggunakan pembelajaran tanpa pengawasan untuk menghasilkan representasi ruang input dari sampel pelatihan berdimensi rendah (biasanya dua dimensi), yang disebut map (Oral & Oral, 2010).

*SOM* merupakan salah satu algoritma pembentukan cluster berdasarkan ciri ataupun fitur-fitur data. *SOM* ialah salah satu bentuk topologi yang dalam proses pelatihannya tidak membutuhkan pengawasan (target output). Kelebihan penggunaan *SOM* dalam pengelompokan objek karena bisa memetakan vektor input yang berukuran tinggi ke dalam ruang berdimensi dua. Tidak hanya itu, *SOM* bisa mengelompokkan data kategori serta data input yang tidak lengkap (Annas dkk, 2007).

Pada jaringan *SOM* masing-masing *neuron* akan dibentuk berdasarkan nilai tertentu dalam suatu *cluster* (Umar dkk, 2018). Dalam *SOM*, informasi yang ada setiap neuron tersebut diperbaharui berdasarkan proses pembelajaran yang dilakukan. *SOM* menggabungkan dua aspek penting, yaitu mengurangi jumlah data dengan mengkluster, dan mengatur data pada tampilan berdimensi rendah (Septiara, 2018). Menurut Haykin (1999), dalam (Rahayu & Hakim, 2019) terdapat empat komponen penting dalam *SOM*:

- a. *Initialization*. Menentukan bobot acak awal secara *random*.
- b. *Competition*. Pada tahap ini untuk setiap pola input, neuron menghitung nilai masing-masing fungsi diskriminan yang memberi dasar untuk kompetisi. Neuron tertentu dengan nilai terkecil dari fungsi diskriminan dinyatakan sebagai pemenang (*winning neuron*).
- c. *Cooperation* : *Winning neuron* menentukan suatu lokasi spasial dari lingkungan topologi *excited neuron* untuk memberi dasar kerja sama dalam suatu lingkungan neuron.
- d. *Adaption* : *Excited neuron* menurunkan nilai fungsi diskriminan yang berkaitan dengan pola input melalui penyesuaian bobot terkait, sehingga respon dari *winning neuron* ke aplikasi berikutnya dengan pola input yang sama akan meningkat.

Adapun tahapan dalam pola jaringan Kohonen *SOM*, menurut (Halim & Widodo, 2017) adalah dengan melakukan langkah berikut.:

- a. Inisialisasi berupa bobot ( $w_{ij}$ ) yang diperoleh secara acak untuk tiap *node*. Setelah bobot ( $w_{ij}$ ) diberikan maka jaringan diberikan *input* ( $x_i$ ).

- b. Setelah *input* diterima jaringan akan melakukan perhitungan jarak *vector*  $d_{ij}$  yang didapat dengan menjumlahkan selisih antara vektor bobot ( $w_{ij}$ ) dengan vektor *input* ( $x_i$ ).

$$d_j = \sum_{i=1}^n (w_{ij} - x_i)^2 \quad (2.2)$$

- c. Setelah jarak antara *node* diketahui maka ditentukan nilai minimum dari perhitungan jarak vektor  $d_{ij}$ , maka tahap selanjutnya melakukan perubahan bobot.

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha [x_i - w_{ij}] \quad (2.3)$$

- d. Pada proses untuk mendapatkan bobot baru memerlukan nilai *learning rate* ( $\alpha$ ) yaitu  $0 \leq \alpha \leq 1$ . Nilai *learning rate* pada setiap *epoch* akan berkurang menjadi  $\alpha (i+1) = 0,5 \alpha$ .  $\alpha$  yang digunakan yaitu 0,05.
- e. Kondisi penghentian pengujian dilakukan dengan menghitung selisih antara bobot  $w_{ij}(\text{new})$  dengan  $w_{ij}(\text{old})$ , apabila pengujian sudah mencapai konvergensi sehingga pengujian dapat dihentikan.

#### 2.4. Validasi Cluster

*Cluster* validasi adalah hasil prosedur evaluasi analisa *cluster* secara kuantitatif dan objektif. Menurut Jain & Dubes 1988 dalam (Hermayadi dkk, 2013) ada tiga pendekatan untuk melakukan evaluasi terhadap validasi *cluster*.

- Eksternal : mengevaluasi hasil dari metode *clustering* berdasarkan praspesifikasi struktur yang diterima dari sebuah data yang mencerminkan intuisi pengguna tentang struktur *clustering* dari data,
- Internal : mengevaluasi hasil *clustering* dalam konsep kuantitatif yang didapat dari data, dan 3 kriteria relatif, membandingkan sebuah struktur *clustering* dengan struktur *clustering* yang lain yang didapatkan dari metode *clustering* yang sama tetapi nilai-nilai parameternya dimodifikasi,
- Relative : mengevaluasi beberapa solusi *cluster* yang berbeda dari data dibandingkan dengan menggunakan algoritma yang sama dengan parameter yang berbeda.

Validasi *cluster* yang digunakan peneliti adalah validasi internal, karena validasi ini didasarkan pada evaluasi hasil *clustering* dimana dalam konsep kuantitatif yang mana pada data pendidikan ini berupa data kuantitatif. Adapun beberapa metode menurut (Susilowati dkk, 2021) pada validasi internal, diantaranya *connectivity*, nilai *silhouette*, dan indeks *Dunn*. Pada penelitian kali ini akan di gunakan salah satu validasi *internal* yaitu validasi *cluster* indeks *Dunn*. Indeks *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. *Dunn*, menghasilkan *cluster* terbaik apabila nilai *Dunn* yang diperoleh semakin besar dan untuk nilai Indeks *Dunn* yang besar atau tinggi tersebut menerangkan bahwa *cluster* yang terbentuk telah terpecah antara *cluster* yang satu dengan yang lain dengan teratur dan penuh atau padat.

$$Dunn = \frac{d_{min}}{d_{max}} \quad (2.4)$$

Keterangan:

$d_{min}$  = jarak terkecil antara observasi pada *cluster* yang berbeda

$d_{max}$  = jarak terbesar pada masing-masing *cluster* data.



## 2.5. Pendidikan

Pendidikan merupakan suatu proses yang sangat penting untuk meningkatkan kecerdasan, keterampilan, mempertinggi budi pekerti, memperkuat kepribadian, dan mempertebal semangat kebersamaan agar dapat membangun diri sendiri dan Bersama sama membangun bangsa (Nurhuda, 2022). Dalam kondisi apapun, manusia tidak dapat menolak efek dari penerapan pendidikan dalam sehari-hari. Pendidikan dibagi menjadi tiga, yaitu pendidikan formal, pendidikan informal, dan pendidikan non formal. Pendidikan formal terdiri dari SD hingga ke perguruan tinggi. Pendidikan informal adalah jenis pendidikan atau pelatihan yang terdapat di dalam keluarga atau masyarakat yang diselenggarakan tanpa ada organisasi tertentu. Pendidikan non formal adalah segala bentuk pendidikan yang diberikan secara terorganisasi tetapi diluar wadah pendidikan formal. Dan pada kali ini akan dibahas mengenai hal-hal yang berkaitan dengan pendidikan formal.

## 3. Research Method and Materials

### 3.1. Data dan Variabel

Data yang digunakan dalam penelitian ini berupa data sekunder atau data informasi yang telah ada, data ini berupa data indikator Pendidikan. Adapun variabel yang digunakan yaitu Angka Partisipasi Kasar SD, Angka Partisipasi Kasar SMP, Angka Partisipasi Kasar SMA, Angka Partisipasi Murni SD, Angka Partisipasi Murni SMP, Angka Partisipasi Murni SMA, Rasio Siswa perGuru SD, Rasio Siswa perGuru SMP, dan Rasio Siswa perGuru SMA. Jumlah data yang digunakan adalah 24 data selaras dengan jumlah kabupaten di Sulawesi Selatan yang berjumlah 24. Deskripsi variabel yang digunakan dapat dilihat sebagai berikut.:

- a. Angka Partisipasi Kasar SD ( $X_1$ ) : Proporsi anak sekolah pada suatu jenjang pendidikan tertentu dalam kelompok umur yang sesuai dengan jenjang pendidikan tersebut berdasarkan jenjang pendidikan SD.
- b. Angka Partisipasi Kasar SMP ( $X_2$ ) : Proporsi anak sekolah pada suatu jenjang pendidikan tertentu dalam kelompok umur yang sesuai dengan jenjang pendidikan tersebut berdasarkan jenjang pendidikan SMP.
- c. Angka Partisipasi Kasar SMA ( $X_3$ ) : Proporsi anak sekolah pada suatu jenjang pendidikan tertentu dalam kelompok umur yang sesuai dengan jenjang pendidikan tersebut berdasarkan jenjang pendidikan SMA.
- d. Angka Partisipasi Murni SD ( $X_4$ ) : Proporsi anak sekolah pada suatu kelompok tertentu yang bersekolah pada tingkat yang sesuai dengan kelompok umurnya berdasarkan jenjang Pendidikan SD.
- e. Angka Partisipasi Murni SMP ( $X_5$ ) : Proporsi anak sekolah pada suatu kelompok tertentu yang bersekolah pada tingkat yang sesuai dengan kelompok umurnya berdasarkan jenjang Pendidikan SMP.
- f. Angka Partisipasi Murni SMA ( $X_6$ ) : Proporsi anak sekolah pada suatu kelompok tertentu yang bersekolah pada tingkat yang sesuai dengan kelompok umurnya berdasarkan jenjang Pendidikan SMA.
- g. Rasio Siswa per Guru SD ( $X_7$ ) : Jumlah perbandingan antara guru dan murid berdasarkan jenjang Pendidikan SD.
- h. Rasio Siswa per Guru SMP ( $X_8$ ) : Jumlah perbandingan antara guru dan murid berdasarkan jenjang Pendidikan SMP.
- i. Rasio Siswa per Guru SMA ( $X_9$ ) : Jumlah perbandingan antara guru dan murid berdasarkan jenjang Pendidikan SMA.

### 3.2. Teknik Analisis Data

Adapun teknik analisis data pada penelitian ini yaitu:



- a) Melakukan analisis deskriptif terhadap variable - variabel yang digunakan.
- b) Melakukan cleaning data
- c) Menentukan *cluster* terbaik dengan menggunakan validasi *Dunn*.
- d) Melakukan pengelompokan dengan algoritma SOM.

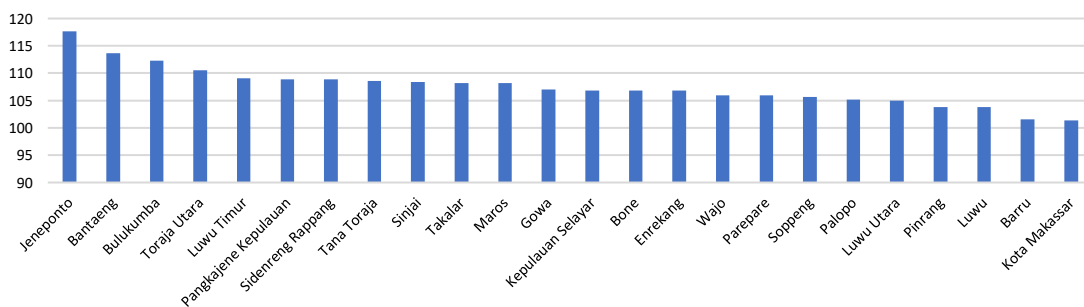
Adapun langkah-langkahnya yaitu:

- a) Melakukan inialisasi secara acak
- b) Menggunakan perhitungan jarak vektor
- c) Menghitung perubahan bobot
- d) Pengujian dihentikan. Pengujian dihentikan apabila selisih antara pembobot baru dan lama sudah mencapai konvergensi sehingga dapat dihentikan.
- e) Melakukan interpretasi hasil dari setiap cluster yang terbentuk. Setelah hasil pengelompokan diketahui, maka dilakukan interpretasi mengenai kelompok yang terbentuk.

## 4. Results and Discussion

### 4.1. Analisis Deskripsi

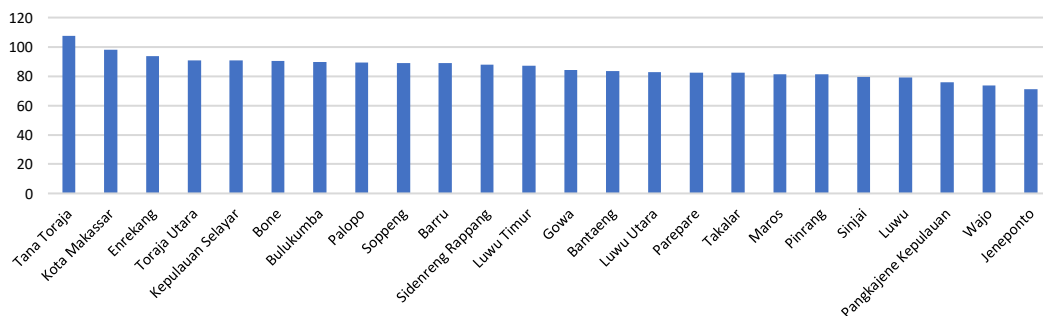
#### a. Angka Partisipasi Kasar SD



**Gambar 1.** Nilai Angka Partisipasi Kasar SD

Berdasarkan Gambar 1 Nilai Angka Partisipasi Kasar(APK) SD tertinggi jika diurutkan dari yang tertinggi, ada di kabupaten/kota Jeneponto, Bantaeng, dan Bulukumba. Sementara nilai terendahnya berasal dari kabupaten/kota Pinrang, Barru, dan Kota Makassar.

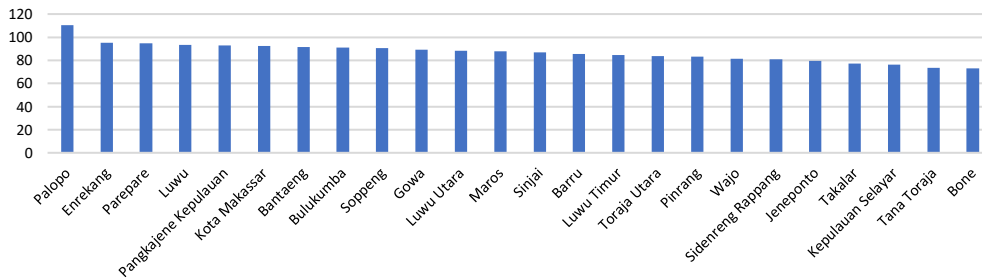
#### b. Angka Partisipasi Kasar SMP



**Gambar 2.** Nilai Angka Partisipasi Kasar SMP

Berdasarkan Gambar 2 Nilai APK SMP jika diurutkan dari yang tertinggi, ada di kabupaten/kota Tana Toraja, Kota Makassar, dan Enrekang. Sementara nilai terendahnya berasal dari Pangkajene Kepulauan, Wajo dan Jeneponto.

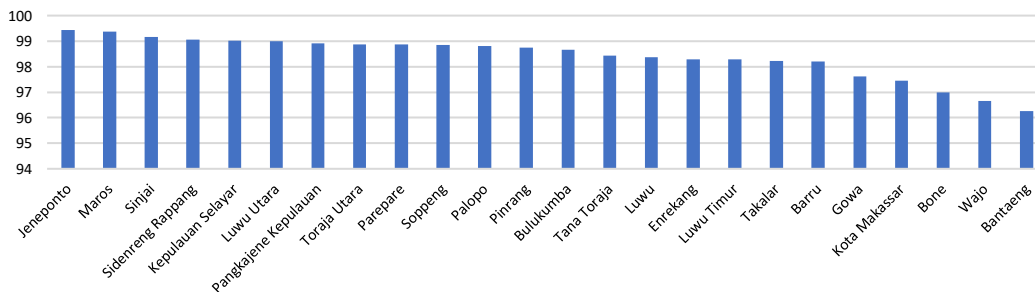
*c. Angka Partisipasi Kasar SMA*



**Gambar 3.** Nilai Angka Partisipasi Kasar SMA

Berdasarkan Gambar 4 Nilai Angka Partisipasi Murni (APM) SD jika diurutkan dari yang tertinggi, ada di kabupaten/kota Jeneponto, Maros, dan Sinjai. Sementara nilai terendahnya berasal dari Kepulauan Bone, Wajo, dan Bantaeng.

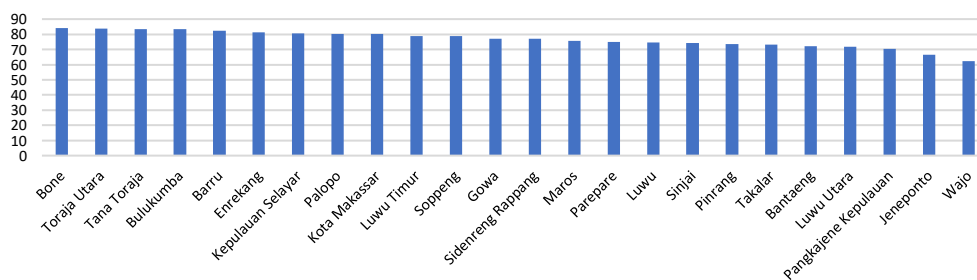
*d. Angka Partisipasi Murni SD*



**Gambar 4.** Nilai Angka Partisipasi Murni SD

Berdasarkan Gambar 4 Nilai Angka Partisipasi Murni (APM) SD jika diurutkan dari yang tertinggi, ada di kabupaten/kota Jeneponto, Maros, dan Sinjai. Sementara nilai terendahnya berasal dari Kepulauan Bone, Wajo, dan Bantaeng.

*e. Angka Partisipasi Murni SMP*

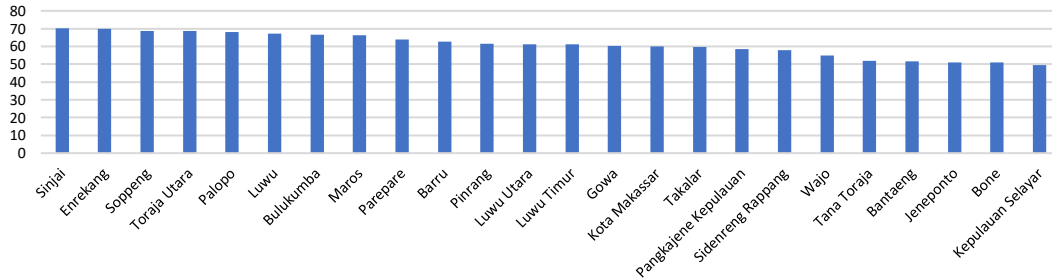


**Gambar 5.** Nilai Angka Partisipasi Murni SMP



Berdasarkan Gambar 5 Nilai APM SMP jika diurutkan dari yang tertinggi, ada di kabupaten/kota Bone, Toraja Utara, dan Tana Toraja. Sementara nilai terendahnya berasal dari Pangkajene Kepulauan, Jeneponto, dan Wajo.

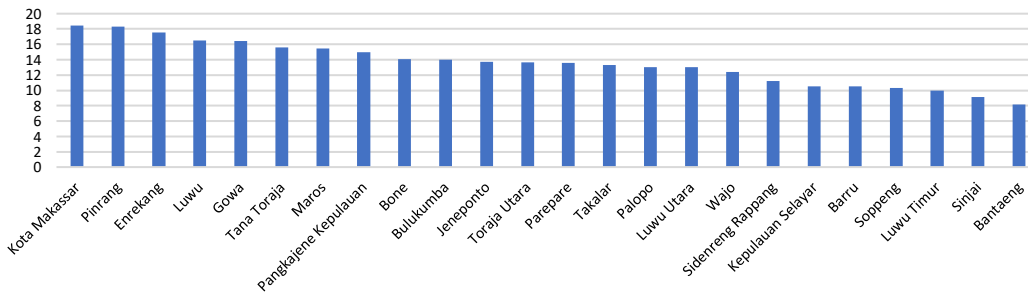
*f. Angka Partisipasi Murni SMA*



**Gambar 6.** Nilai Angka Partisipasi Murni SMA

Berdasarkan Gambar 6 Nilai Angka Partisipasi Murni SMA jika diurutkan dari yang tertinggi, ada di kabupaten/kota Sinjai, Enrekang, dan Soppeng. Sementara nilai terendahnya berasal dari Kepulauan Jeneponto, Bone, dan Kepulauan Selayar.

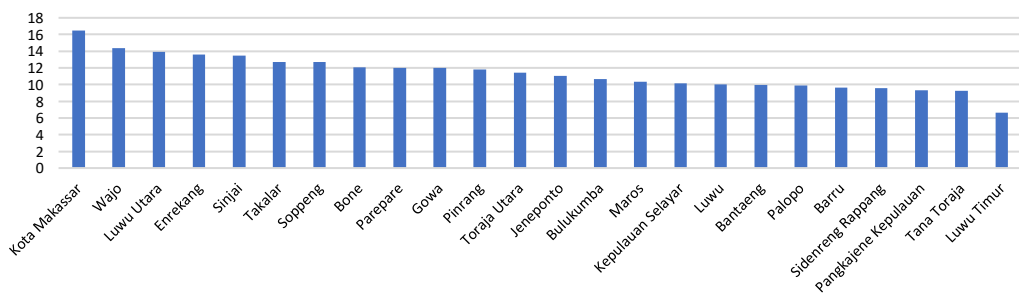
*g. Rasio Siswa per Guru SD*



**Gambar 7.** Nilai Rasio Siswa per Guru SD

Berdasarkan Gambar 7 Nilai Rasio Siswa per Guru(RSG) SD jika diurutkan dari yang tertinggi, ada di kabupaten/kota Makassar, Pinrang, dan Enrekang . Sementara nilai terendahnya berasal dari Kabupaten Luwu timur, Sinjai, dan Bantaeng.

*h. Rasio Siswa per Gowa SMP*

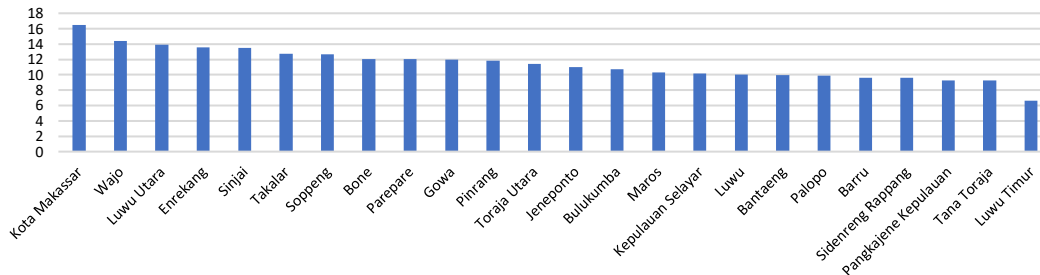


**Gambar 8.** Nilai Rasio Siswa per Guru SMP



Berdasarkan Gambar 8 Nilai RSG SMP jika diurutkan dari yang tertinggi, ada di kabupaten/kota Kota Makassar, Wajo, dan Luwu Utara. Sementara nilai terendahnya berasal dari Pangkajene Kepulauan, Tana Toraja, dan Luwu Timur.

*i. Rasio Siswa per Guru SMA*



**Gambar 9.** Nilai Rasio Siswa per Guru SMA

Berdasarkan Gambar 9 Nilai RSG SMA jika diurutkan dari yang tertinggi, ada di kabupaten/kota Pinrang, Soppeng, dan Kepulauan Selayar. Sementara nilai terendahnya berasal dari Luwu Timur, Luwu Utara, dan Toraja Utara.

*4.2. Cleaning Data*

*Cleaning* data (pembersihan data) adalah proses untuk memperbaiki atau menghapus data yang tidak akurat, tidak relevan, atau tidak lengkap sehingga data dapat digunakan secara efektif. Penelitian ini akan menggunakan *cleaning* data berupa Standarisasi data. Standarisasi data pada metode *SOM* berguna untuk menghindari bias yang dapat muncul akibat perbedaan skala pada setiap variabel atau fitur dalam dataset serta dapat mempercepat proses iterasi pada *SOM*. Tabel 1 menyajikan hasil dari *cleaning* data dengan metode standarisasi data.

*4.3. Validasi Indeks Dunn*

Penelitian ini menggunakan salah satu jenis dari validasi internal, yaitu validasi Indeks *Dunn* dengan Jumlah *cluster* yang akan dipilih adalah jumlah *cluster* dengan nilai indeks *Dunn* mendekati 1. Hasil validasi indeks *Dunn* disajikan pada Tabel 2.

Tabel 2 menunjukkan hasil validasi *cluster* dimana nilai dari Indeks *Dunn* paling besar yaitu sebesar 0,42 pada *cluster* 4. Sehingga dari hasil validasi *cluster* diperoleh nilai *cluster* terbaik adalah 4, maka dalam penelitian ini menggunakan jumlah *cluster* 4 untuk mengelompokan pemerataan Pendidikan di Sulawesi Selatan.

**Tabel 1.** Hasil *Cleaning* data

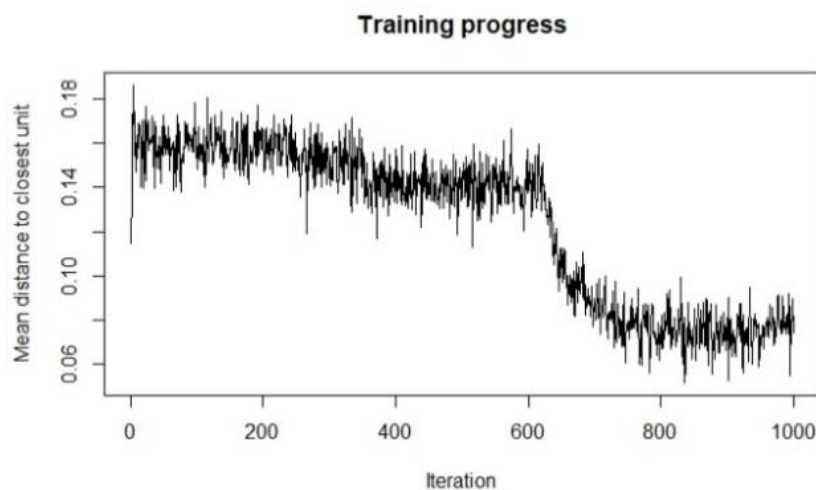
Kabupaten/Kota	X1	X2	X3	X4	X5	X6	X7	X8	X9
Selayar	-0,19	0,61	-1,25	0,73	0,73	-1,72	-1,04	-0,57	0,94
Bulukumba	1,33	0,49	0,50	0,31	1,18	0,84	0,18	-0,33	0,09
Bantaeng	1,69	-0,32	0,55	-2,55	-0,79	-1,41	-1,87	-0,68	-0,21
<b>Jeneponto</b>	<b>2,80</b>	<b>-1,87</b>	<b>-0,90</b>	<b>1,24</b>	<b>-1,80</b>	<b>-1,50</b>	<b>0,07</b>	<b>-0,17</b>	<b>0,03</b>
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<b>Toraja Utara</b>	<b>0,82</b>	<b>0,63</b>	<b>-0,39</b>	<b>0,57</b>	<b>1,24</b>	<b>1,18</b>	<b>0,06</b>	<b>0,02</b>	<b>-2,62</b>
Kota Makassar	-1,68	1,55	0,66	-1,13	0,65	-0,14	1,74	2,41	0,48
Parepare	-0,44	-0,44	0,97	0,57	-0,33	0,44	0,02	0,31	0,16
Palopo	-0,63	0,46	2,84	0,49	0,66	1,09	-0,16	-0,71	-1,01

**Tabel 2.** Hasil Uji Validasi Cluster

Banyaknya Cluster	Dunn
2	0.35
3	0.37
<b>4</b>	<b>0.42</b>
5	0.36
6	0.31

#### 4.4. Pengelompokan dengan Self Organizing Maps

Setelah melakukan analisis deskriptif dan *cleaning* data, selanjutnya melakukan *clustering* dengan menggunakan algoritma *self organizing maps*. Sebelum melakukan algoritma SOM, terlebih dahulu melakukan *training progress* untuk meminimalisir rata-rata jarak suatu objek ke unit terdekat.

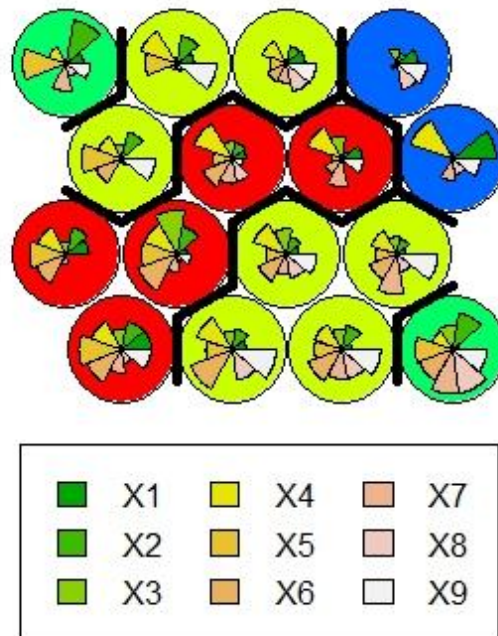


**Gambar 10.** Training Progress

Hasil Gambar 10 menerangkan banyaknya *training progress* yang membuktikan banyaknya iterasi yang dilakukan terhadap jarak rata-rata ke unit terdekat. Pada penelitian ini proses iterasi dihentikan setelah dilakukan 1000 iterasi, dimana terlihat pada iterasi 600 ke atas, menampilkan proses iterasi yang konvergen. Setelah melakukan *training progress*, selanjutnya melakukan *clustering* dengan menggunakan algoritma *self organizing maps*. Berikut adalah hasil diagram *fan* dengan tampilan *hexagonal* dengan *grid* 4 x 4 yang dihasilkan dari proses algoritma SOM dengan menggunakan 4 *cluster*.

Berdasarkan Gambar 11 terdapat 4 warna terbentuk yang menunjukkan bahwa hasil *cluster* setiap Kabupaten/Kota di Sulawesi Selatan. Untuk lebih jelasnya dapat dilihat pada Tabel 3 tentang *cluster* yang terbentuk dan karakteristiknya dari masing-masing *cluster*. Berikut ini adalah hasil dari pengelompokan berdasarkan indikator Pendidikan di Sulawesi Selatan.

Dari Tabel 3 diketahui jumlah serta masing-masing anggota dari setiap *cluster* yang terbentuk. Pada pembentukan *cluster* tersebut di dapatkan hasil *clustering* menggunakan algoritma SOM. Hasil yang terbentuk yaitu *cluster* 1 beranggotakan 7 Kabupaten, *cluster* 2 beranggotakan 10 Kabupaten/Kota, *cluster* 3 beranggotakan 4 Kabupaten dan *cluster* 4 beranggotakan 3 Kabupaten.



**Gambar 11.** Diagram *fan* untuk 4 *Cluster*

**Tabel 3.** Hasil Pengelompokan 4 *Cluster*

Cluster	Jumlah Anggota	Anggota
1	7	Bulukumba, Maros, Pangkajene Kepulauan, Luwu Utara, Luwu Timur, Toraja Utara, dan Palopo
2	10	Kepulauan Selayar, Takalar, Gowa, Sinjai, Barru, Soppeng, Sidenreng Rappang, Pinrang, Luwu, dan Parepare
3	4	Bone, Enrekang, Tana Toraja, dan Kota Makassar
4	3	Bantaeng, Jenepono, dan Wajo

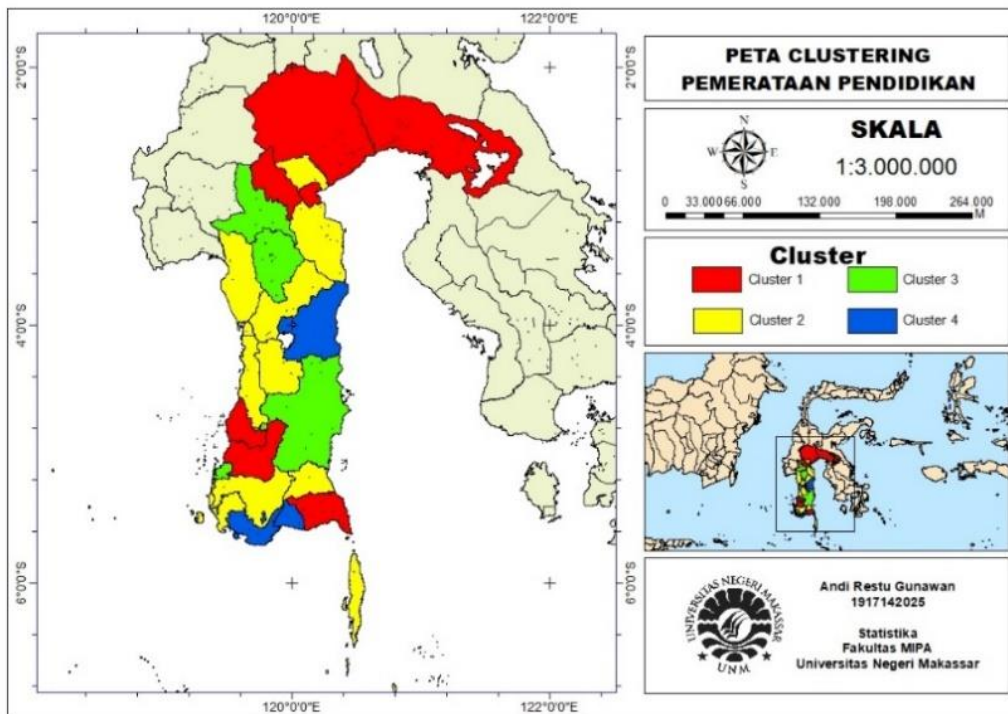
#### 4.5. Pemetaan dengan ARCGis

Setelah dilakukan *cluster* kemudian dilakukan Pemetaan menggunakan *Arcgis* dengan 4 *cluster*. Pemetaan ini berguna untuk melihat visualisasi hasil *Clustering SOM* ke dalam bentuk peta, sehingga dapat mempermudah pemahaman dalam melihat *cluster* yang terbentuk dari segi visualisasi. Berikut pemetaan dengan menggunakan *Software Arcgis*.

Dari hasil pemetaan pada gambar 12 dapat di simpulkan bahwa tidak setiap cluster yang terbentuk bersinggungan wilayah dengan cluster lainnya. Hal ini di sebabkan karena adanya pengaruh kedekatan jarak yang berbeda antra data satu dengan data yang lainnya, yang dapat di ukur dengan pengukuran jarak *Euclidean distance*.

#### 4.6. Deskripsi Hasil Cluster

Setelah dilakukan pengelompokan dan pemetaan, selanjutnya dilakukan perhitungan rata-rata dari tiap *cluster* untuk mengenali karakteristik dari tiap *cluster* (profilisasi) berdasarkan indikator Pendidikan di Sulawesi Selatan.



Gambar 12. Pemetaan dengan *Argis*

Tabel 4. Deskripsi hasil *Cluster*

Variabel	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Angka Partisipasi Kasar SD ( $X_1$ )	108	106	106,01	112
Angka Partisipasi Kasar SMP ( $X_2$ )	85,3	84,5	97,4	76
Angka Partisipasi Kasar SMA ( $X_3$ )	91,1	85,8	83,5	84
Angka Partisipasi Murni SD ( $X_4$ )	98,8	98,6	97,8	97,4
Angka Partisipasi Murni SMP ( $X_5$ )	77,7	76,6	82,3	67,1
Angka Partisipasi Murni SMA ( $X_6$ )	64,4	62,2	58,2	52,5
Rasio Siswa per Guru SD ( $X_7$ )	13,4	13	16,4	11,4
Rasio Siswa per Guru SMP ( $X_8$ )	10,3	11,4	12,8	11,8
Rasio Siswa per Guru SMA ( $X_9$ )	8,48	17,4	14,4	14,3

Tabel 4 diperoleh deskripsi hasil *cluster* untuk membuat tingkat kemiripan atau kesamaan karakteristik suatu objek dalam hal ini objeknya adalah indikator Pendidikan di Sulawesi selatan dan didapatkan rataan APK, APM, dan RSG dengan 4 *cluster* yang sudah bisa diinterpretasikan.

## 5. Kesimpulan

Dari hasil penelitian dan pembahasan pada bab sebelumnya, dapat disimpulkan beberapa hal berikut ini.

- Gambaran umum indikator Pendidikan di setiap Kabupaten/Kota di Sulawesi Selatan Tahun 2021 tercatat bahwa Kabupaten Jeneponto menunjukkan nilai APK dan APM SD tertinggi, Tana Toraja menunjukkan nilai APK SMP tertinggi, Palopo menunjukkan nilai APK SMA tertinggi, Bone menunjukkan nilai APM SMP tertinggi, Sinjai menunjukkan nilai APM SMA tertinggi, Bantaeng menunjukkan nilai RSG SD terbaik, Luwu timur menunjukkan nilai RSG SMP terbaik, dan Toraja utara menunjukkan nilai RSG SMA terbaik.
- Berdasarkan dari hasil validasi Indeks *Dunn* diperoleh *cluster 4* yang memiliki nilai validasi Indeks *Dunn* yang optimal yaitu dengan nilai 0,42.

- c. Berdasarkan pemetaan dengan *ArGIS* dapat disimpulkan bahwa tidak setiap cluster yang terbentuk berada dalam satu wilayah yang bersih. karena adanya perbedaan jarak *Euclidean distance*.
- d. Karakteristik yang terdapat pada masing-masing *cluster* yaitu ada pada *cluster 1* menunjukkan peningkatan APK tertinggi dan nilai RSG terbaik di banding *cluster* lainnya. *Cluster 2* menunjukkan nilai RSG SMA yang lebih tinggi dibandingkan *cluster* yang lainnya. *Cluster 3* menunjukkan nilai APK yang semakin tinggi jenjang Pendidikannya maka semakin rendah nilai APKnya bahkan pada *cluster 3* ini memiliki nilai rata-rata APK SMA yang terendah di banding 3 *cluster* lainnya. *Cluster 4* menunjukkan nilai rata-rata APM yang paling rendah di banding 3 *cluster* lainnya dan menunjukkan nilai RSG SD terbaik di banding 3 *cluster* lainnya.

## References

- Afrimayani, A., Yozza, H., & Devianto, D. (2019). Pengelompokan Negara Di Dunia Berdasarkan Data Runtun Waktu Realisasi Penanaman Modal Asing Di Indonesia Menggunakan Analisis Cluster. *Jurnal Matematika UNAND*, 8(2), 157. <https://doi.org/10.25077/jmu.8.2.157-164.2019>
- Alkhalidi, M. W., Nadeak, B., & Sayuthi, M. (2020). *Sistem Informasi Geografis Pemetaan Wilayah Penyalahgunaan Narkoba Menggunakan Metode SOM ( Self-Organizing Map ) Studi Kasus : Kabupaten Aceh Tenggara*. 2(1), 1–9.
- Amalia, E. L., & Yunhasnawa, Y. (2019). Klinik Laptop Menggunakan Metode Self Organizing Maps ( SOM ). *Seminar Informatika Aplikatif Polinema*, 18–23.
- Annas, S., Kanai, T., & Koyama, S. (2007). Principal Component Analysis and Self-Organizing Map for Visualizing and Classifying Fire Risks in Forest Regions. *Agricultural Information Research*, 16(2), 44–51. <https://doi.org/10.3173/air.16.44>
- Annisa, K., Ginting, B. S., & Syari, M. A. (2022). Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat. *Jurnal Ilmu Komputer Dan Informatika Volume*, 6(1), 113–129.
- Cynthia, E. P., & Ismanto, E. (2017). Eka Pandu Cynthia, 2) Edi Ismanto. *RABIT : Jurnal Teknologi Dan Sistem Informasi Univrab*, 2(2), 83–98.
- Halim, N. N., & Widodo, E. (2017). Clustering dampak gempa bumi di indonesia menggunakan kohonen self organizing maps. *Prosiding SI MaNIS (Seminar Nasional Integrasi Matematika Dan Nilai Islami)*, 1(1), 188–194. <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/62>
- Hermayadi, R., Purnomo, H. M., & Purnama, I. K. E. (2013). *Pengelompokan Data Koordinat BTS Menggunakan K-Means dan Visualisasi Berbasis Goodle Map*. (2), 1–8.
- Idrus, M. (2012). Mutu pendidikan dan pemerataan pendidikan di daerah [Quality of Education and Equitable Education in Regions]. *Psikopedagogia*, 1(2), 1–10.
- Kapita, S. N., Abdullah, S. Do, Studi, P., Informatika, T., Teknik, F., Khairun, U., Pendidikan, N., & Sekolah, M. (2020). Pengelompokan data mutu sekolah menggunakan jaringan syaraf tiruan Kohonen som grouping of school quality data using ann of kohonen som. 3(1), 56–61. <https://doi.org/10.33387/jiko>
- Kartikasari, M. D. (2021). *Self-Organizing Map Menggunakan Davies-Bouldin Index dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan*. 3(2), 187–196.
- Miftahuddin, Y., Umaroh, S., & Karim, F. R. (2020). Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan. *Jurnal*

- Tekno Insentif*, 14(2), 69–77. <https://doi.org/10.36787/jti.v14i2.270>
- Nanggala, A. (2020). Jurnal Pendidikan Kewarganegaraan Undiksha Vol. 8 No. 2 (Mei, 2020). *Jurnal Pendidikan Kewarganegaraan Undika*, 8(2), 157–167.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>
- Nurhuda, H. (2022). Faktor Dan Solusi Yang Ditawarkan National Education Problems ; Factors and Solutions. *Jurnal Pemikiran Dan Pendidikan Dasar Islam*, 127–137.
- Oral, E. L., & Oral, M. (2010). Predicting construction crew productivity by using Self Organizing Maps. *Automation in Construction*, 19(6), 791–797. <https://doi.org/10.1016/j.autcon.2010.05.001>
- Raharjo, M., Napih, M., Putra, J. L., & Mustofa, M. (2020). Prediksi Pengaruh Matakuliah Terhadap Peminatan Outline Tugas Akhir Mahasiswa Dengan Jaringan Syaraf Tiruan. *Jurnal Infortech*, 2(1), 78–83. <https://doi.org/10.31294/infortech.v2i1.7965>
- Rahayu, B., & Hakim, R. B. F. (2019). *Longsor Di Indonesia Menggunakan Kohonen Self Organizing Maps (Som)*. 1–17.
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>
- Susilowati, T., Sugiarto, D., & Mardianto, I. (2021). *JURNAL RESTI*. 1(10).
- Syahril, & Qashlim, A. (2017). *pemerataan pendidikan menggunakan self organizing maps (som)*. 3(2), 27–34.
- Umar, R., Fadlil, A., & Az Zahra, R. R. (2018). *kebazanah informatika Jurusan di SMK*. 4(2), 131–137.