

# Analisis Cluster Menggunakan Algoritma K-Means

*by* Prizka Rismawati Arum

---

**Submission date:** 07-Sep-2023 03:27PM (UTC+0700)

**Submission ID:** 2159725452

**File name:** tera\_Barat\_Berdasarkan\_Indeks\_Pembangunan\_Manusia\_Tahun\_2021.pdf (388.4K)

**Word count:** 3173

**Character count:** 18624



## 1 Analisis Cluster Menggunakan Algoritma K-Means Pada Provinsi Sumatera Barat Berdasarkan Indeks Pembangunan Manusia Tahun 2021

### Cluster Analysis Using the K-Means Algorithm in West Sumatra Province Based on the 2021 Human Development Index

Li<sup>1</sup> Anggraini<sup>1</sup>, Prizka Rismawati Arum, M.Stat<sup>2</sup>

<sup>1</sup> Universitas Muhammadiyah Semarang, Semarang

<sup>2</sup> Universitas Muhammadiyah Semarang, Semarang

Corresponding author : [prizka.rismawatiarum@unimus.ac.id](mailto:prizka.rismawatiarum@unimus.ac.id)

#### Abstrak

27 Indeks Pembangunan Manusia (IPM) merupakan tolak ukur angka kesejahteraan suatu daerah atau negara. IPM Provinsi Sumatera Barat berada diatas rata-rata nasional. Akan tetapi, masih terdapat Kabupaten/Kota di Provinsi Sumatera Barat memiliki pencapaian dibawah rata-rata angka nasional. Kondisi pembangunan manusia yang belum merata akan menyulitkan pemerintah untuk meningkatkan sumber daya manusia. Penelitian ini menggunakan algoritma *k-means* untuk mengelompokkan Kabupaten/Kota Provinsi Sumatera Barat berdasarkan kemiripan karakteristik yang dilihat dari indikator indeks pembangunan manusia yakni Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Perkapita (PRP). Terdapat tiga cluster berdasarkan IPM yaitu Cluster 1 : Kabupaten Kepulauan Mentawai, Kabupaten Sijunjung, Kabupaten Pasaman, Kabupaten Solok Selatan, dan Kabupaten Pasaman Barat. Cluster 2 : Kota Padang, Kota Solok, Kota Padang Panjang, Kota Bukittinggi, Kota Payakumbuh, dan Kota Pariaman. Cluster 3 : Kabupaten Pesisir Selatan, Kabupaten Solok, Kabupaten Tanah Datar, Kabupaten Padang Pariaman, Kabupaten Agam, Kabupaten Lima Puluh Kota, Kabupaten Dharmasraya, dan Kota Sawahlunto. Berdasarkan Angka Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Pengeluaran Perkapita dapat disimpulkan terdapat perbedaan yang signifikan pada tahun 2021.

**Kata Kunci** : IPM, K-Means, Pengelompokan, Sumatera Barat

#### Abstract

37 The Human Development Index (HDI) is a benchmark for the welfare rate of a region or country. The HDI of West Sumatra Province is above the national average. However, there are still regencies/cities in West Sumatra Province that have achievements below the national average. Uneven human development conditions will make it difficult for the government to improve human resources. This study uses the *k-means* algorithm to group regencies/cities in West Sumatra Province based on similar characteristics seen from the human development index indicators, namely Life Expectancy (AHH), School Length Expectations (HLS), Average Length of Schooling (RLS), and Per capita Expenditure (PRP). There are three clusters based on HDI, namely Cluster 1: Mentawai Islands, Sijunjung, Pasaman, South Solok, and West Pasaman. Cluster 2: Padang City, Solok, Padang Panjang, Bukittinggi, Payakumbuh, and Pariaman. Cluster 3: Pesisir Selatan, Solok Regency, Tanah Datar, Padang Pariaman, Agam, Fifty Cities Regency, Dharmasraya, and Sawahlunto City. Based on Life Expectancy, School Length Expectations, Average Length of Schooling, and Per capita Expenditure, it can be concluded that there is a significant difference in 2021.

**Keywords** : Cluster, HDI, K-Means, West Sumatera



## PENDAHULUAN

Keberhasilan suatu negara dalam pembangunan nasional tidak hanya dilihat dari pertumbuhan ekonomi akan tetapi juga dilihat dari peningkatan kualitas sumber daya manusia dalam pembangunan manusia. Keberhasilan dalam upaya membangun kualitas hidup manusia dapat diukur dengan Indeks Pembangunan Manusia (IPM). IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam bentuk pendapatan, kesehatan, pendidikan, dan aspek lain dalam kehidupan (BPS, 2021). Menurut Setiawan (2013) aspek terpenting kehidupan ini dilihat dari usia yang panjang dan hidup sehat, tingkat pendidikan yang memadai, dan standar hidup yang layak.

Provinsi Sumatera Barat pada tahun 2021 memiliki pencapaian IPM dengan peringkat 9 tertinggi di Indonesia yakni sebesar 72,65 dengan angka IPM Sumatera Barat ini berhasil melewati rata-rata nasional. Akan tetapi, pencapaian angka IPM yang tinggi tidak menjadikan bahwa Kabupaten/Kota di Provinsi Sumatera Barat tersebut juga memiliki IPM yang tinggi. Perlu dilakukan analisis untuk mengetahui indikator-indikator penyusun IPM di Kabupaten/Kota sebagai penyumbang nilai IPM Provinsi Sumatera Barat dan mengelompokkan wilayah Kabupaten/Kota di Sumatera Barat sebagai perencanaan dan evaluasi program pemerintah agar tepat sasaran mencapai tujuan meningkatkan angka pembangunan manusia. Analisis Statistik yang dapat digunakan adalah analisis *cluster* dengan metode K-Means.

K-Means salah satu algoritma data mining yang banyak digunakan baik dalam dunia bisnis, akademik, ataupun industri. Algoritma ini bekerja dengan cara membagi data dalam sejumlah *cluster* untuk dianalisis faktor kesamaan (*similarity*) maupun ketidaksamaan (*dissimilarity*) yang melekat pada kumpulan data tersebut. K-Means mengelompokkan Kabupaten/Kota di Provinsi Sumatera Barat berdasarkan kemiripan karakteristik daerah yang ditinjau dari empat indikator yakni Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Perkapita (PRP).

## TINJAUAN PUSTAKA

### 1. Analisis Cluster

Pengelompokkan dimana anggota dari setiap partisi mempunyai persamaan berdasarkan matriks tertentu adalah dengan menggunakan *Cluster*. Analisis *Cluster* atau analisis kelompok merupakan teknik analisa data yang bertujuan untuk mengelompokkan individu atau objek ke dalam beberapa kelompok yang memiliki sifat berbeda antar kelompok, sehingga individu atau objek yang terletak di dalam satu kelompok akan mempunyai sifat relative homogen. Tujuan analisis *cluster* adalah mengelompokkan objek-objek tersebut.



Kelebihan analisis *Cluster* yakni (1) dapat mengelompokkan data observasi dalam jumlah besar dan variabel yang relative banyak. (2) dapat dipakai dalam skala data ordinal, interval dan rasio. Sedangkan kelemahan dari analisis *Cluster* ini adalah (1) pengelompokan bersifat subjektifitas peneliti karena hanya melihat dari gambar dendogram. (2) untuk data heterogen antara objek penelitian yang satu dengan yang lain akan sulit bagi peneliti untuk menentukan jumlah kelompok yang di bentuk. (3) metode-metode yang dipakai memberikan perbedaan yang signifikan, sehingga dalam perhitungan biasanya masing-masing metode dibandingkan. (4) semakin besar observasi, biasanya tingkat kesalahan akan semakin besar.

Proses inti *Clustering* adalah pengelompokan data, yang biasa dilakukan dengan dua metode yaitu :

- a. Metode Hierarki dimana metode ini mengelompokkan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk macam pohon dimana ada hierarki (tingkatan yang jelas) antara objek, dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut.
- b. Metode Nonhierarki dimana metode ini dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan. Pusat *cluster* yang dipilih adalah pusat *cluster* sementara yang terus diperbaharui untuk setiap iterasi sampai dengan kriteria tercapai, sehingga terdapat kemungkinan jika objek berpindah dari satu *cluster* ke *cluster* yang lain. Metode non hierarki yang terkenal disebut *K-Means Cluster*.

## 2. *K-Means Cluster*

Jhonson (2002) *K-Means Cluster* adalah metode yang dirancang untuk mengelompokkan item, bukan variabel ke dalam *cluster* yang telah ditentukan jumlahnya di awal. Menurut Witen (2012) *K-Means* merupakan salah satu algoritma dalam data mining untuk mengelompokkan data berdasarkan jarak dengan membagi data ke beberapa *cluster* serta algoritma hanya dapat bekerja pada data numerik. Ada banyak pendekatan untuk membuat *cluster*, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam kelompok yang sama berdasarkan tingkat persamaan di antara anggota-anggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah *clustering*.



### 3. Indeks Pembangunan Manusia

Pembangunan manusia didefinisikan oleh *United Nation Development Program* (UNDP) sebagai suatu proses untuk memperluas pilihan-pilihan bagi penduduk, dalam artian bahwa manusia diberikan kebebasan untuk memilih pilihan lebih banyak dalam hal memenuhi kebutuhan hidup. Tak ukur untuk mengetahui keberhasilan pembangunan manusia adalah dengan indeks pembangunan manusia (IPM). Indeks Pembangunan Manusia (IPM) merupakan indikator yang dapat menggambarkan perkembangan pembangunan manusia secara terukur dan representative. IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan dan sebagainya. Pada dasarnya IPM diharapkan mampu mewakili kinerja pembangunan manusia sehingga dapat dibandingkan antar wilayah atau bahkan antar waktu.

IPM didasari oleh 3 dimensi diantaranya adalah: 1) Umur panjang dan hidup sehat (*a long and healthy life*). 2) Pengetahuan (*knowledge*). 3) Standar hidup layak (*decent standard of living*). Kemudian BPS menggabungkan 3 dimensi tersebut dengan 4 komponen yaitu, angka harapan hidup (AHH), harapan lama sekolah (HLS), rata-rata lama sekolah (RLS), dan pengeluaran per kapita.

#### 1 METODE

##### 1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh melalui publikasi lembaga pemerintahan resmi yaitu Badan Pusat Statistik periode tahun 2021. Unit observasi dalam penelitian ini adalah Sumatera Barat.

##### 2. Variabel Penelitian

Dalam penelitian ini, variabel yang digunakan adalah angka harapan hidup (X1), harapan lama sekolah (X2), rata-rata lama sekolah (X3), dan pengeluaran perkapita (X4).

##### 3. Analisis Data

Wahidiah (2010) menjelaskan tahapan dalam algoritma K-Means clustering sebagai berikut :

- Menentukan jumlah cluster
- Melakukan standarisasi data yaitu transformasi ke bentuk Z-Score jika terdapat perbedaan ukuran satuan yang besar pada data antar variabel. Standarisasi menggunakan rumus Z-Score yaitu :

$$Z_i = \frac{x_i - \bar{x}}{s}$$

dimana  $Z_i$  = nilai Z-Score ke-i;  $x_i$  = Nilai data ke-i;  $\bar{x}$  = Nilai rata-rata; S = Nilai Standarisasi,  $i = 1, 2, 3, \dots, n$ .



- c. Menentukan nilai *centroid* untuk awal iterasi secara acak. Selanjutnya jika menentukan nilai *centroid* yang merupakan tahap dari iterasi, dapat digunakan rumus.

$$\bar{V}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj}$$

dimana  $\bar{V}_{ij}$  = *centroid*/ rata-rata *cluster* ke-*i* untuk variabel ke-*j*;  $i, k$  = indeks dari *cluster* *j* adalah indeks dari variabel;  $N_i$  = jumlah data yang menjadi anggota *cluster* ke-*i*;  $x_{kj}$  = nilai data ke-*k* yang ada di dalam *cluster* tersebut untuk variabel ke-*j*

- d. Menghitung *Euclidean Distance* untuk mengetahui jarak antara titik *centroid* dengan titik tiap objek menggunakan rumus

$$D_e = \sqrt{(x_i - S_i)^2 + (y_i - t_i)^2}$$

dimana  $D_e$  = *Euclidean Distance*;  $i$  = banyaknya objek (0,1,2, ..., n);  $(s,t)$  = koordinat *centroid*;  $(x,y)$  = koordinat objek.

- e. Pengelompokkan objek dalam menentukan anggota *cluster* dengan menghitung jarak minimum objek. Nilai yang diperoleh adalah 0 atau 1, nilai 1 berarti data yang dialokasikan ke *cluster* dan nilai 0 data yang alokasikan ke *cluster* yang lain.
- f. Kembali ke tahap 2, melakukan repetisi isampai diperoleh nilai *centroid* tetap dan anggota *cluster* tidak berpindah ke *cluster* yang lain.

## HASIL DAN PEMBAHASAN

### 1. Data Indikator Indeks Pembangunan Manusia

Berikut disajikan data yang akan digunakan untuk mengelompokkan Kabupaten/Kota di Provinsi Sumatera Barat berdasarkan indeks pembangunan manusia.

**Tabel 1. Data Indikator IPM**

Wilayah	AHH	HLS	RLS	PRP
Kepulauan Mentawai	64,73	12,89	7,20	6321,00
Pesisir Selatan	70,96	13,33	8,27	9270,00
Kab.Solok	68,79	13,05	7,87	10215,00
Sijunjung	66,36	12,38	8,12	10389,00
Tanah Datar	70,12	14,34	8,62	10616,00
Padang Pariaman	68,97	13,68	7,88	11050,00
Agam	72,53	13,88	8,97	9662,00
Lima Puluh Kota	69,84	13,30	8,07	9668,00
Pasaman	67,59	12,81	8,10	8440,00
Solok Selatan	68,01	12,72	8,32	10367,00
Dharmasraya	71,53	12,44	8,55	11324,00
Pasaman Barat	67,94	13,68	8,27	9089,00



Wilayah	AHH	HLS	RLS	PRP
Padang	73,69	16,53	11,59	14540,00
Kota Solok	73,73	14,33	11,04	12168,00
Sawahlunto	70,10	13,18	10,32	10195,00
Padang Panjang	72,82	15,07	11,63	10754,00
Bukittinggi	74,50	14,98	11,34	13331,00
Payakumbuh	73,84	14,27	10,81	13317,00
Pariaman	70,38	14,55	10,67	12818,00

Data yang diperoleh, selanjutnya diproses menggunakan bantuan software Rstudio 4.2.1 untuk membentuk *cluster* berdasarkan indikator IPM di Provinsi Sumatera Barat menggunakan metode non hierarki.

## 2. Proses pengelompokkan dengan metode Non Hierarki

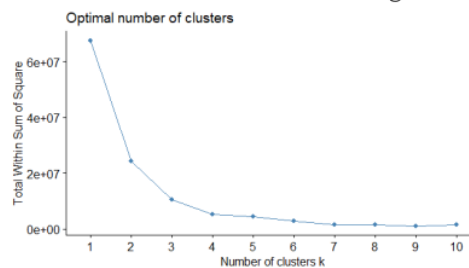
Pada statistik deskriptif dibawah ini menunjukkan bahwa semua data lengkap yaitu AHH, HLS, RLS, dan PRP pada 19 Kabupaten/Kota di Provinsi Sumatera Barat.

31

**Tabel 2. Statistik Deskriptif**

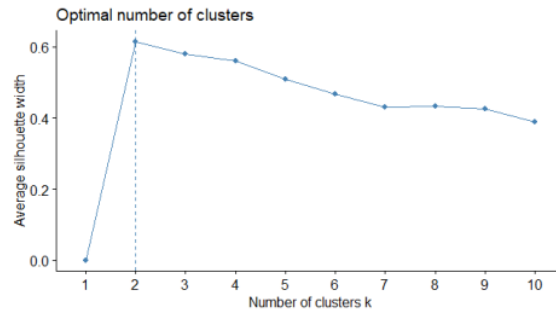
	N	Min	Max	Mean	Std. Dev
AHH	19	64.73	74.50	70.3384	2.74589
HLS	19	12.38	16.53	13.7584	1.05606
RLS	19	7.20	11.63	9.2442	1.49133
PRP	19	6321	14540	10712.3158	1935.63
Valid (listwise)	N 19				

Selanjutnya menentukan cluster, karena analisis K-Means merupakan analisis non hierarki. Maka nilai k ditentukan sendiri oleh peneliti. Namun untuk membuktikan nilai k yang digunakan paling optimal bisa membandingkan dengan beberapa metode yaitu metode WSS, dan *Sillhouette* sebagai berikut.



**Gambar 1. Within Sum of Square**

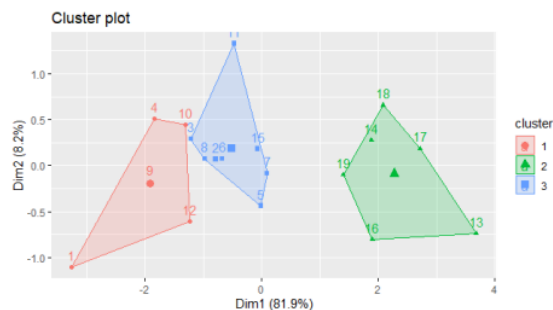
dengan menggunakan metode *Within Sum of Square* (WSS) dapat dilihat pada Gambar 1. Grafik yang pergerakannya landai pertama yaitu pada angka 3-4, berarti nilai K optimum metode WSS adalah 3.



**Gambar 2. Silhouette**

untuk penentuan nilai K pada metode *silhouette*, dilihat dari garis tertinggi atau melihatnya dengan melihat garis yang paling optimum adalah 2, namun ada opsional lain yaitu menentukan nilai K dengan menggunakan grafik paling tinggi berikutnya setelah grafik paling tinggi pertama yakni grafik tertinggi kedua, yaitu di angka 3, berarti nilai K optimum metode *silhouette* adalah 3.

Setelah menentukan jumlah *cluster*, dengan menggunakan data yang telah distandarisasi dilakukan pembentukan *cluster*. Berdasarkan hasil output didapatkan hasil sebagai berikut.



**Gambar 3. Cluster Plot**

23 Pada Gambar 3 dengan menggunakan  $k=3$ , diketahui *cluster* 1 terdiri dari 5 Kabupaten/Kota, *cluster* 2 terdiri 6 Kabupaten/Kota, *cluster* 3 terdiri 8 Kabupaten/Kota. Berikut anggota *cluster* yang telah di bentuk sesuai dengan Gambar 3.





Setelah diketahui 19 Kabupaten/Kota di Provinsi Sumatera Barat masuk ke cluster mana, selanjutnya didapatkan nilai *Cluster Means* atau hasil pusat *cluster* sebagai berikut.

**Tabel 3. Cluster Means**

	Cluster		
	1	2	3
AHH	-1,24274	1,02757	0,00604
HLS	-0,81664	1,13306	-0,33939
RLS	-0,83296	1,29803	-0,45293
PRP	-0,92534	1,08958	-0,23884

Tabel 3 diketahui bahwa data tersebut masih terkait dengan proses standarisasi yang mengacu pada *z-score* dengan ketentuan sebagai berikut.

- 1) Angka negatif (-) berarti data dibawah rata-rata total.
- 2) Angka positif (+) berarti diatas rata-rata total.

Berdasarkan Tabel 3 dengan ketentuan yang telah dijelaskan, dapat didefinisikan sebagai berikut.

*Cluster 1* : Pada *cluster 1* diperoleh bahwa pada tahun 2021 angka AHH, HLS, RLS, dan PRP berada dibawah rata-rata total. Dari ciri-ciri tersebut dapat dikatakan bahwa angka HLS dan RLS sangat rendah terjadi pada tahun 2021.

*Cluster 2* : Pada *cluster 2* diperoleh bahwa pada tahun 2021 angka AHH, HLS, RLS, dan PRP berada di atas rata-rata total. Dari ciri-ciri tersebut dapat dikatakan bahwa angka AHH sangat rendah terjadi pada tahun 2021.

*Cluster 3* : Pada *cluster 3* diperoleh bahwa pada tahun 2021 angka AHH berada diatas rata-rata total. Sedangkan angka HLS, RLS, dan PRP berada dibawah rata-rata total. Dari ciri-ciri tersebut dapat dikatakan bahwa angka PRP sangat rendah terjadi pada tahun 2021.

Setelah membentuk 3 *cluster*, langkah selanjutnya adalah melihat apakah variabel-variabel yang telah terbentuk *cluster* mempunyai perbedaan pada setiap *cluster*. Dalam hal ini dapat dilihat dari *F* dan nilai probabilitas (*sig*) masing-masing variabel. Hal ini dilakukan dengan melihat *output* Anova berikut.

**Tabel 4. Anova**

	ANOVA					
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	Df		
Z_AHH	9,259	1	-0,514	17	18,009	0,001
Z_HLS	11,258	1	-0,397	17	28,388	0,000



Z_RLS	14,775	1	-0,19	17	77,888	0,000
Z_PRP	10,411	1	-0,446	17	23,319	0,000

Berdasarkan Tabel 4 jika semakin besar angka F suatu variabel dan angka *sig* adalah 0.05 maka semakin besar pula perbedaan variabel tersebut pada variabel lainnya. Sebagai contoh, angka F terbesar (77,888) ada pada Z RLS, dengan angka *sig* 0,000 yang berarti signifikansinya adalah nyata. Hal ini berarti faktor angka RLS sangat membedakan karakteristik ketiga *cluster* tersebut. atau dapat dikatakan angka RLS pada ketiga *cluster* yang sangat berbeda antara *cluster* 1 dengan *cluster* lainnya. Begitu juga dengan variabel lainnya.

Untuk mengetahui lebih detail Kabupaten/Kota masing-masing *cluster* dapat dilihat pada Tabel 5.

**Tabel 5. Anggota Cluster**

Anggota Clustering							
1	Jarak	2	Jarak	3	Jarak		
Kepulauan Mentawai	1.653	Padang	1.768	Pesisir Selatan	0.591		
Sijunjung	0.929	Kota Solok	0.718	Kab. Solok	0.809		
Pasaman	0.362	Padang Panjang	1.122	Tanah Datar	0.915		
Solok Selatan	0.887	Bukittinggi	0.565	Padang Pariaman	0.842		
Pasaman Barat	0.853	Payakumbuh	0.781	Agam	0.999		
				Pariaman	1.135	Lima Puluh Kota	0.496
						Dhamasraya	1.148
				Sawahlunto	1.197		

Pada Tabel 5 menunjukkan nomor *cluster* dari Kabupaten/Kota, dan kolom jarak ke pusat *cluster* menunjukkan jarak terdekat antara data dengan pusat *cluster*, dimana semakin kecil berarti semakin mirip dengan *cluster* terkait. Dengan bantuan *software* SPSS dapat diinterpretasikan sebagai berikut.

*Cluster* 1 : Kabupaten Kepulauan Mentawai, Kabupaten Sijunjung, Kabupaten Pasaman, Kabupaten Solok Selatan, dan Kabupaten Pasaman Barat. Jarak ke pusat *cluster* 1 adalah 1.653; 0.929; 0.362; 0.887; 0.852. Jarak ke pusat *cluster* 1 Kabupaten Sijunjung, Kabupaten Solok Selatan dan Kabupaten Pasaman Barat tidak berbeda jauh sehingga dapat disimpulkan karakteristik di *cluster* 1 yang mirip adalah Kabupaten Sijunjung, Kabupaten Solok Selatan dan Kabupaten Pasaman Barat.



*Cluster 2* : Kota Padang, Kota Solok, Kota Padang Panjang, Kota Bukittinggi, Kota Payakumbuh, dan Kota Pariaman. Jarak ke pusat *cluster 2* adalah 1.768; 0.718; 1.122; 0.565; 0.781; 1.135. Jarak ke pusat *cluster 2* Kota Solok dan Kota Payakumbuh tidak berbeda jauh sehingga dapat disimpulkan karakteristik di *cluster 2* yang mirip adalah Kota Solok dan Kota Payakumbuh.

*Cluster 3* : Kabupaten Pesisir Selatan, Kabupaten Solok, Kabupaten Tanah Datar, Kabupaten Padang Pariaman, Kabupaten Agam, Kabupaten Lima Puluh Kota, Kabupaten Dhamasraya, dan Kota Sawahlunto. Jarak ke pusat *cluster 3* adalah 0.591; 0.809; 0.915; 0.842; 0.999; 0.496; 1.148; 1.197. Jarak ke pusat *cluster 3* Kabupaten Solok, Kabupaten Padang Pariaman, Kabupaten Tanah Datar dan Kabupaten Agam tidak berbeda jauh sehingga dapat disimpulkan karakteristik di *cluster 3* yang mirip adalah Kabupaten Padang Pariaman, Kabupaten Tanah Datar dan Kabupaten Agam.

## KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dengan *K-Means Cluster* pada data indikator IPM Kabupaten/Kota di Provinsi Sumatera Barat Tahun 2021, diperoleh kesimpulan pengelompokan Kabupaten/Kota berdasarkan indikator Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Perkapita (PRP) diperoleh 3 *Cluster* sebagai berikut :

- a. *Cluster 1* yaitu Kabupaten/Kota dengan indikator IPM rendah, yaitu Kabupaten Kepulauan Mentawai, Kabupaten Sijunjung, Kabupaten Pasaman, Kabupaten Solok Selatan, dan Kabupaten Pasaman Barat. Hal ini dikarenakan karakteristik Kabupaten/Kota di *cluster 1* memiliki kesamaan karakteristik yaitu nilai semua indikator IPM paling rendah jika dibandingkan dengan nilai indikator pada *cluster* lain.
- b. *Cluster 2* yaitu Kota Padang, Kota Solok, Kota Padang Panjang, Kota Bukittinggi, Kota Payakumbuh, dan Kota Pariaman. Hal ini dikarenakan karakteristik Kabupaten/Kota di *cluster 2* memiliki kesamaan karakteristik yaitu nilai semua indikator IPM paling tinggi jika dibandingkan dengan nilai indikator pada *cluster* lain.
- c. *Cluster 3* yaitu Kabupaten Pesisir Selatan, Kabupaten Solok, Kabupaten Tanah Datar, Kabupaten Padang Pariaman, Kabupaten Agam, Kabupaten Lima Puluh Kota, Kabupaten Dhamasraya, dan Kota Sawahlunto. Hal ini dikarenakan karakteristik Kabupaten/Kota di *cluster 3* memiliki kesamaan karakteristik yaitu nilai semua indikator IPM berada ditengah- tengah *cluster 1* dan *cluster 2*.



## DAFTAR PUSTAKA

- Badan Pusat Statistik. (2021). Indeks Pembangunan Manusia 2020. Jakarta: Badan Pusat Statistik.
- Ediyanto,*dkk.* (2013). Pengklasifikasian Karakteristik Dengan Metode *K-Means Cluster Analysis*. *Bulletin Ilmiah Mat.Stat. dan Terapannya (BIMASTER)*, vol. 02, No. 2, Hal. 133-136.
- Jhonson, R.A. & Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis, fifth Edition*. New Jersey: Pearson Prentice Inc.
- Puspita, Rindang Ndaru. (2021). Analisis K-Means Cluster Pada Kabupaten/Kota di Provinsi Banten Berdasarkan Indikator Indeks Pembangunan Manusia. *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika, Matematika dan Statistika*, Vol. 2, No. 3, Hal. 267-281.
- Setiawan, M. B., & Hakim, A. (2013). Indeks pembangunan manusia Indonesia. *Jurnal Economia*, 9(1), 18-26.
- Talakua, M.W., *dkk.* (2017). Analisis *Cluster* Dengan Menggunakan Metode *K-Means* Untuk Pengelompokkan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. Barekeng: Jurnal Ilmu Matematika dan Terapan, Vol. 11, No. 2, Hal. 119-128.
- Wakhidah, Nur. (2010). Clustering Menggunakan K-Means Algorithm. *Jurnal Transformatika*, 8(1): 45-52.
- Witten, H.I & Frank, E. (2012). *Data Mining Practical Machine Learning Tools and Technique 2<sup>nd</sup> Edition*. San Fransisco: Morgan Kaufmann

# Analisis Cluster Menggunakan Algoritma K-Means

## ORIGINALITY REPORT

23%

SIMILARITY INDEX

19%

INTERNET SOURCES

10%

PUBLICATIONS

4%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://prosiding.unimus.ac.id">prosiding.unimus.ac.id</a> Internet Source	3%
2	<a href="http://repository.unisba.ac.id">repository.unisba.ac.id</a> Internet Source	2%
3	<a href="http://jos.unsoed.ac.id">jos.unsoed.ac.id</a> Internet Source	1%
4	<a href="http://minangkabaunews.com">minangkabaunews.com</a> Internet Source	1%
5	<a href="http://www.journal.uim.ac.id">www.journal.uim.ac.id</a> Internet Source	1%
6	Rusdiansyah Rusdiansyah, Harun Al Rasyid, Suryanto Sosrowidigdo. "Implementation of address recording management using the K-Means clustering classification algorithm in Kebayoran District, DKI Jakarta", SinkrOn, 2021 Publication	1%
7	<a href="http://data.acehprov.go.id">data.acehprov.go.id</a> Internet Source	1%

8	<a href="http://ejournal.uin-suka.ac.id">ejournal.uin-suka.ac.id</a> Internet Source	1 %
9	<a href="http://eprints.uty.ac.id">eprints.uty.ac.id</a> Internet Source	1 %
10	<a href="http://fti.uajy.ac.id">fti.uajy.ac.id</a> Internet Source	1 %
11	<a href="http://lib.ibs.ac.id">lib.ibs.ac.id</a> Internet Source	1 %
12	<a href="http://jdih.blorakab.go.id">jdih.blorakab.go.id</a> Internet Source	1 %
13	Submitted to Universitas Sanata Dharma Student Paper	1 %
14	<a href="http://journal.unhas.ac.id">journal.unhas.ac.id</a> Internet Source	1 %
15	Submitted to Universitas Pertamina Student Paper	<1 %
16	<a href="http://www.stieumeltual.com">www.stieumeltual.com</a> Internet Source	<1 %
17	Febry Sandrian Sagala, Mugiarto Mugiarto, Wowon Priatna. "Algoritma K-Means Clustering Untuk Menentukan Rekomendasi Pemberian Beasiswa Bagi Siswa Berprestasi", Journal of Students' Research in Computer Science, 2021 Publication	<1 %

18

[sumbarpost.com](http://sumbarpost.com)

Internet Source

&lt;1 %

19

Submitted to Universitas Amikom

Student Paper

&lt;1 %

20

Submitted to Universitas Tadulako

Student Paper

&lt;1 %

21

[ejournal.ust.ac.id](http://ejournal.ust.ac.id)

Internet Source

&lt;1 %

22

Agus Bandiyono. "PENGARUH BELANJA PEMERINTAH DAERAH BERDASARKAN FUNGSI TERHADAP PENINGKATAN IPM DAN PENGENTASAN KEMISKINAN ( STUDI PADA KABUPATEN/KOTA DI PROVINSI ACEH)", INFO ARTHA, 2018

Publication

&lt;1 %

23

[journal.stikespemkabjombang.ac.id](http://journal.stikespemkabjombang.ac.id)

Internet Source

&lt;1 %

24

[jualjualskripsi.blogspot.com](http://jualjualskripsi.blogspot.com)

Internet Source

&lt;1 %

25

[openaccess.city.ac.uk](http://openaccess.city.ac.uk)

Internet Source

&lt;1 %

26

[dabar.srce.hr](http://dabar.srce.hr)

Internet Source

&lt;1 %

27

[jurnal.magelangkota.go.id](http://jurnal.magelangkota.go.id)

Internet Source

&lt;1 %

28

Samin Radjid, Nadia Istifarin, Meylani Tuasella. "PENGELOMPOKKAN JUMLAH PENDUDUK KABUPATEN BURU SELATAN BERDASARKAN JENIS KELAMIN PADA TAHUN 2018 DENGAN ALGORITMA K-MEANS", Pattimura Proceeding: Conference of Science and Technology, 2022

Publication

&lt;1 %

29

[m.moam.info](http://m.moam.info)

Internet Source

&lt;1 %

30

[prioritaspendidikan.org](http://prioritaspendidikan.org)

Internet Source

&lt;1 %

31

[pub.unj.ac.id](http://pub.unj.ac.id)

Internet Source

&lt;1 %

32

I T Utami, Rais Rais, W Seftiani. "ANALISIS KLAS TER PAUTAN LENGKAP UNTUK MENGELOMPOKKAN KABUPATEN/KOTA DI PROVINSI SULAWESI TENGAH BERDASARKAN INDIKATOR KRIMINALITAS", JURNAL ILMIAH MATEMATIKA DAN TERAPAN, 2019

Publication

&lt;1 %

33

Nur Asiska, Neva Satyahadewi, Hendra Perdana. "PENCARIAN CLUSTER OPTIMUM PADA SINGLE LINKAGE, COMPLETE LINKAGE DAN AVERAGE LINKAGE", Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya, 2019

&lt;1 %



34

[bankmakalah13.blogspot.com](http://bankmakalah13.blogspot.com)

Internet Source

<1 %

35

[ejurnal.litbang.pertanian.go.id](http://ejurnal.litbang.pertanian.go.id)

Internet Source

<1 %

36

[podatekzanglii.pl](http://podatekzanglii.pl)

Internet Source

<1 %

37

[ppjp.ulm.ac.id](http://ppjp.ulm.ac.id)

Internet Source

<1 %

38

[ecoplan.ulm.ac.id](http://ecoplan.ulm.ac.id)

Internet Source

<1 %

39

[ejournal.unib.ac.id](http://ejournal.unib.ac.id)

Internet Source

<1 %

40

[ejurnal.itats.ac.id](http://ejurnal.itats.ac.id)

Internet Source

<1 %

41

[www.aljazeera.com](http://www.aljazeera.com)

Internet Source

<1 %

42

Musfiani Musfiani. "ANALISIS CLUSTER DENGAN MENGGUNAKAN METODE PARTISI PADA PENGGUNA ALAT KONTRASEPSI DI KALIMANTAN BARAT", Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya, 2019

Publication

<1 %

43

[journal.ipb.ac.id](http://journal.ipb.ac.id)

Exclude quotes Off

Exclude matches Off

Exclude bibliography On