

Pengelompokkan Kabupaten/Kota di Provinsi Sulawesi Utara Berdasarkan Indikator Pendidikan Menggunakan Analisis Cluster

Citra Filadelfia Manoppo, Charles Mongi*, Christie Montolalu

Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sam Ratulangi, Manado, 95115

*Corresponding author: mongicharles@unsrat.ac.id

Abstrak

Indikator pendidikan merupakan salah satu tolok ukur bagi pemerintah untuk meningkatkan mutu pendidikan di Indonesia. Penelitian ini bertujuan untuk mengetahui hasil *cluster* yang terbentuk dan pengelompokan yang terbaik dengan menggunakan metode *K-Means* jarak *Euclidean* dan *Manhattan* untuk pengelompokan Kabupaten/Kota di Provinsi Sulawesi Utara berdasarkan Indikator Pendidikan. Pada penelitian ini objek yang digunakan adalah 15 Kabupaten/Kota di Provinsi Sulawesi Utara dan variabelnya adalah indikator pendidikan khususnya jumlah sekolah, jumlah guru dan jumlah siswa dari tingkat SD-sederajat sampai SMA/SMK-sederajat. Sumber data yaitu Badan Pusat Statistik Provinsi Sulawesi Utara tahun 2020. Berdasarkan penelitian yang dilakukan dengan metode *K-Means* jarak *Euclidean* dan *Manhattan* jumlah yang dibentuk untuk pengelompokan Kabupaten/Kota yaitu sebanyak 3 cluster. Perbandingan dari penelitian ini menggunakan simpangan baku. Hasil dari perbandingan kedua jarak tersebut diperoleh bahwa jarak *Euclidean* merupakan jarak yang baik untuk digunakan dalam penelitian ini, karena memiliki nilai simpangan baku dalam kelompok yaitu 0,2216 dan nilai simpangan baku antar kelompok yaitu 1,61251 sehingga nilai rasio simpangan baku yang didapat yaitu sebesar 0,13743.

Kata kunci : Indikator Pendidikan, *K-Means*, *Euclidean*, *Manhattan*

Grouping Districts/Cities in North Sulawesi Province Based on Education Indicators Using Cluster Analysis

Abstrak

Education indicators are one of the benchmarks for the government to improve the quality of education in Indonesia. This study aims to determine the results of the clusters formed and the best grouping using the K-Means method of Euclidean and Manhattan distances for the grouping of Regencies/Cities in North Sulawesi Province based on Education Indicators. In this study, the objects used were 15 regencies/cities in North Sulawesi Province and the variables were education indicators, especially the number of schools, the number of teachers, and the number of students from SD-equivalent to SMA/SMK-equivalent levels. The data source is the Central Bureau of Statistics of North Sulawesi Province in 2020. Based on research conducted using the K-Means method, the Euclidean and Manhattan distances are formed for the grouping of Regencies/Cities, namely 3 clusters. The comparison of this study uses the standard deviation. The results of the comparison of the two distances are obtained that the Euclidean distance is a good distance to be used in this study because it has a standard deviation value within the group that is 0.2216 and the standard deviation value between groups is 1.61251 so the standard deviation ratio value obtained is equal to 0.13743.

Keywords: Education Indicators, K-Means, Euclidean, Manhattan

PENDAHULUAN

Pendidikan merupakan hal yang sangat penting dalam kehidupan. Pada dasarnya pendidikan merupakan suatu kebutuhan yang harus dipenuhi dalam kehidupan bermasyarakat. Pendidikan mencakup 2 konsep yaitu belajar dan mengajar. Maju atau tidaknya suatu bangsa tergantung pada kualitas pendidikannya. Indikator pendidikan merupakan salah satu tolak ukur bagi pemerintah untuk meningkatkan mutu pendidikan di Indonesia.

Provinsi Sulawesi Utara memiliki 15 Kabupaten/Kota dengan jumlah sekolah, guru dan siswa yang berbeda-beda. Untuk mengetahui dan mengevaluasi sejauh mana pemerataan pendidikan atau karakteristik tingkat pendidikan 2 Kabupaten/Kota di Provinsi Sulawesi Utara, diperlukan indikator pendidikan. Berdasarkan data pada Badan Pusat Statistik Provinsi Sulawesi Utara tahun 2020, untuk indikator pendidikan di Provinsi Sulawesi Utara masih belum maksimal apabila ditinjau berdasarkan jumlah sekolah, jumlah guru dan siswa, khususnya jumlah sekolah tingkat SD sampai SMA/SMK, jumlah guru tingkat SD sampai SMA/SMK dan jumlah siswa tingkat SD sampai SMA/SMK. Oleh karena itu, penting untuk mengkaji indikator pendidikan. Indikator pendidikan pada penelitian ini adalah indikator untuk jumlah sekolah, siswa dan guru di Kabupaten/Kota Provinsi Sulawesi Utara. Oleh karena itu, dengan 15 Kabupaten/Kota di Provinsi Sulawesi Utara sebagai objek dan indikator pendidikan sebagai peubah, maka penelitian ini bertujuan mengelompokkan Kabupaten/Kota di Provinsi Sulawesi Utara berdasarkan indikator pendidikan. Model analisis yang digunakan untuk mencapai tujuan tersebut adalah dengan menggunakan analisis kuantitatif, dimana data dianalisis menggunakan metode statistika yaitu analisis cluster.

Analisis *cluster* merupakan suatu metode untuk mengelompokkan variabel atau objek ke dalam beberapa kelompok. *K-Means cluster* adalah salah satu metode yang membutuhkan jumlah *cluster* yang ditentukan terlebih dahulu sebagai k , dan kemudian membagi n objek kedalam k *cluster*. Kemiripan antara anggota dalam satu *cluster* sangat tinggi sedangkan kemiripan antar anggota dengan *cluster* yang lain sangat rendah (Muzakir,2014).

Jarak yang biasa digunakan dalam analisis *cluster* diantaranya adalah jarak *Euclidean*, jarak *Mahalanobis*, jarak *Manhattan*, jarak *Log-likelihood* (Mongi, 2015).

Berdasarkan penelitian terdahulu, perbandingan hasil analisis *K-Means* dan *K-Medoids* juga dilakukan oleh Sangga (2018) yaitu Perbandingan Algoritma *K-Means* dan *K-Medoids* dalam Pengelompokan Komoditas Peternakan di Provinsi Jawa Tengah tahun 2015.

Penelitian lain tentang perbandingan hasil analisis *K-Means* dan *K-Medoids* juga dilakukan oleh Fuady dan Nugraha (2017) yang bertujuan untuk Mengelompokkan 82 Kota Di Indonesia Berdasarkan Indeks Harga Konsumen.

METODOLOGI PENELITIAN

Waktu Penelitian

Penelitian ini dilaksanakan dari bulan Desember 2021, mulai dari penyusunan proposal, pengambilan data serta pengolahan data.

Prosedur Penelitian

Pada penelitian ini digunakan analisis cluster dengan metode *K-means* dengan jarak *Euclidean* dan *Manhattan*. Analisis data diolah dengan bantuan *microsoft excel* dan *software RStudio*.

Prosedur 1

Mendeskripsikan data menggunakan analisis deskripif.

Prosedur 2

Melakukan standarisasi/pembakuan data. Menurut Hair,et al. (2006),pembakuan data adalah proses mengkonversi nilai masing-masing variabel awal menjadi nilai standar dengan rata-rata 0 dan standar deviasi 1 untuk menghilangkan bias yang disebabkan 14 karena perbedaan skala dari beberapa variabel yang digunakan dalam analisis. Nilai standar untuk x_{ij} adalah:

$$Z_{ij} = \frac{x_{ij} - m_j}{S_j} \quad (1)$$

Keterangan:

- x_{ij} = nilai objek ke- i pada variabel ke- j
- i = 1,2,3, ... ,n
- Z_{ij} = data x_{ij} yang sudah terstandarkan
- S_j = simpangan baku dari variabel ke- j
- m_j = rata-rata dari variabel ke- j

Prosedur 3

Pengelompokan dengan metode *K-Means* jarak *Euclidean*

Adapun persamaan untuk menghitung jarak *Euclidean* adalah sebagai berikut (Prasetyo, 2012):

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2)$$

Prosedur 4

Pengelompokan dengan metode *K-Means* jarak *Manhattan*

Adapun persamaan untuk jarak *Manhattan* adalah sebagai berikut (Prasetyo, 2012):

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (3)$$

Prosedur 5

Membandingkan hasil *K-Means* jarak *Euclidean* dan *Manhattan* dengan melihat nilai simpangan baku. Untuk mengetahui metode mana yang mempunyai kinerja terbaik, dapat digunakan rata-rata simpangan baku dalam *Cluster* (σ_W) dan simpangan baku antar *Cluster* (σ_B) (Bunkers, dkk.1996)

Rumus rata-rata simpangan baku dalam *Cluster* :

$$(\sigma_W) = K^{-1} \sum_{k=1}^K \sigma_k \quad (4)$$

Rumus simpangan baku *Cluster* ke- k σ_k :

$$\sigma_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu_k)^2} \quad (5)$$

Rumus simpangan baku antar *Cluster* σ_B :

$$\sigma_B = [(K)^{-1} \sum_{k=1}^K (\mu_k - \mu)^2]^{1/2} \quad (6)$$

Rumus Rasio simpangan baku *Cluster* σ :

$$\sigma = \frac{\sigma_W}{\sigma_B} \times 100\% \quad (7)$$

Dengan :

- K : Banyaknya *Cluster* yang terbentuk
- σ_k : Simpangan baku *Cluster* ke- k
- N : Jumlah Anggota dari setiap *Cluster*
- μ_k : Rata-rata *Cluster* ke- k
- X_i : Anggota *Cluster*, dari $I=1,2,\dots,N$
- μ : Rata-rata keseluruhan *Cluster*

HASIL DAN PEMBAHASAN

Hasil 1

Gambaran umum data penelitian disajikan pada Tabel 1.

Tabel 1. Analisis Deskriptif 12 Variabel

Var	Min	Mean	Maks	St. Dev
X1	60	155.2	350	89.67178
X2	19	53.2	107	30.37198
X3	5	17.8	54	12.88521
X4	4	12.46667	37	8.314073
X5	603	1274.333	2673	669.7187
X6	361	656.8667	1420	318.0067
X7	85	347.3333	1038	244.4979
X8	73	273.2667	829	185.3342
X9	5870	15730.8	42527	10111.11
X10	3000	8415.867	24273	5550.403
X11	1173	4706.4	16231	3717.736
X12	850	3539.733	11920	2881.724

Hasil 2

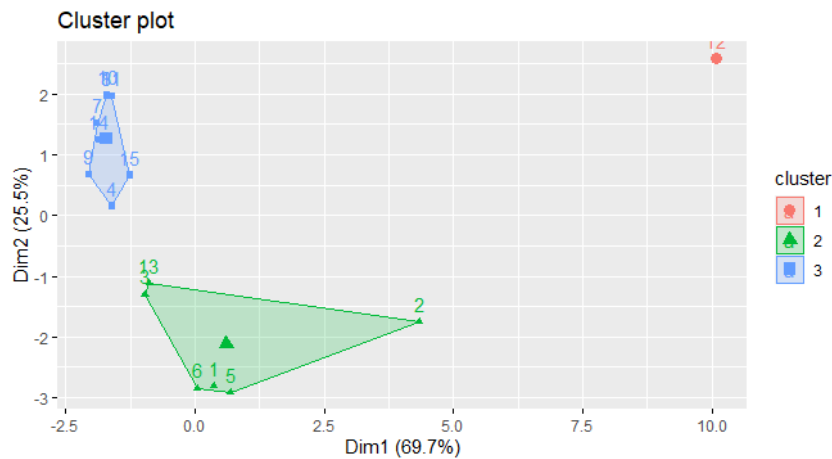
Hasil standarisasi data dapat dilihat dalam Tabel 2.

Tabel 2. Hasil Standarisasi Data

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
0,834	0,882	0,248	0,425	0,861	0,595	-0,03	-0,1	0,76	0,62	-0,06	-0,22
2,172	1,771	1,18	0,786	1,818	1,554	1,4	0,81	1,22	0,99	1,05	0,504
0,723	0,487	0,404	-0,42	0,312	-0,18	0,05	-0,3	-0,4	-0,5	-0,49	-0,45
-0,43	-0,34	0,093	-0,3	-0,286	-0,41	0,4	-0,6	-0,7	-0,7	-0,43	-0,93
0,935	1,113	0,171	0,545	0,574	0,818	0,24	0,52	0,34	0,46	0,291	0,38
0,488	0,718	0,481	0,425	0,585	0,683	-0,03	0,01	0,55	0,3	0,119	0,242
-0,59	-0,76	-0,76	-1,02	-0,678	-0,66	-0,64	-0,9	-0,7	-0,7	-0,55	-0,85
-0,57	-0,9	-0,99	-0,78	-0,853	-0,88	-0,92	-0,8	0,98	-1	-0,78	-0,8
-0,64	-0,34	-0,45	-0,78	-0,683	-0,53	-0,59	-0,6	-0,5	-0,4	-0,39	-0,51
-0,85	-0,9	-0,84	-0,9	-0,79	-0,74	-0,96	-1,1	-0,8	-0,8	-0,85	-0,87
-1,06	-0,86	-0,99	-0,3	-1,002	-0,81	-1,07	-0,5	-0,8	-0,9	-0,95	-0,71
1,347	1,607	2,809	2,951	2,088	2,4	2,82	3	2,65	2,86	3,1	2,908
-0,46	-0,4	-0,14	0,305	-0,302	-0,16	-0,15	0,38	0,46	0,38	0,076	0,724
-0,99	-0,96	-0,53	-0,66	-0,908	-0,93	-0,31	-0,3	-0,7	-0,5	-0,22	-0,14
-0,91	-1,13	-0,68	-0,3	-0,735	-0,74	-0,21	0,39	-0,5	-0,2	0,09	0,729

Hasil 3

Metode *k-means* mengelompokkan data didasarkan pada jarak terdekat dengan pusatnya. Analisis *cluster* metode *k-means* diawali dengan menentukan jumlah *cluster*, pada penelitian ini objek akan dikelompokkan kedalam 3 *cluster*.



Gambar 1. Cluster Plot Jarak Euclidean

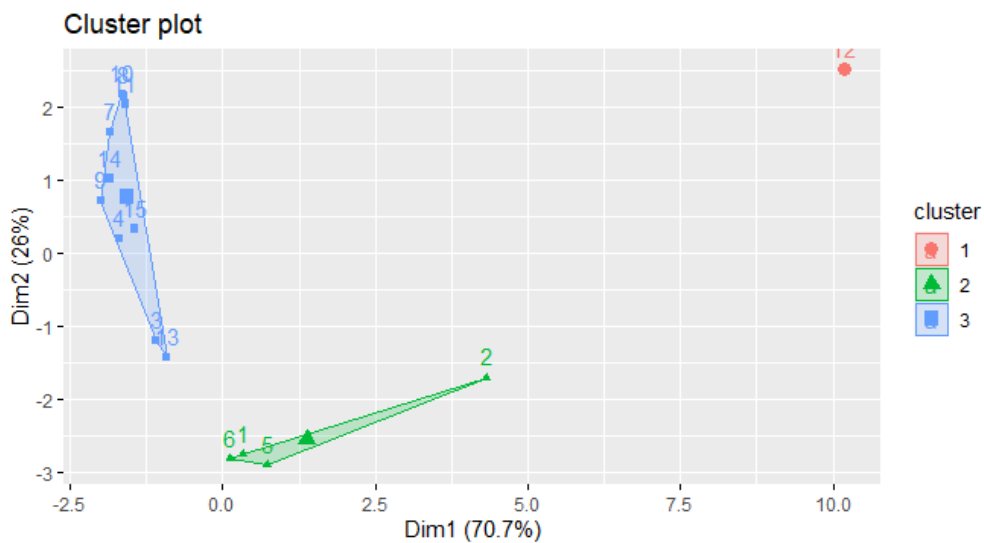
Tabel 3. Hasil Cluster Menggunakan Jarak Euclidean

C	Jumlah Anggota	Kabupaten/Kota
1	1	Kota Manado
2	6	Kab. Bolaang Mongondow, Kab. Minahasa, Kab. Kep. Sangihe, Kab. Minahasa Selatan, Kab. Minahasa Utara, Kota Bitung
3	8	Kab. Kepulauan Talaud, Kab. Bolaang Mongondow Utara, Kab. Kepulauan Siau Tagulandang Biaro, Kab. Minahasa Tenggara, Kab. Bolaang Mongondow Selatan, Kab. Bolaang Mongondow Timur, Kota Tomohon, Kota Kotamobagu

Pada Tabel 3 dapat diketahui masing-masing anggota *cluster* dari ketiga *cluster* yang terbentuk. *Cluster* 1 terdiri dari 1 Kabupaten/Kota, *cluster* 2 terdiri dari 6 Kabupaten/Kota, *cluster* 3 terdiri dari 8 Kabupaten/Kota.

Hasil 4

Metode *k-medoids* merupakan metode pengklasteran untuk mengelompokkan sekumpulan *n* objek ke *k cluster*, pengelompokan ini menggunakan *medoids* sebagai pusat *clusternya*.



Gambar 2. Cluster Plot Jarak Manhattan

Tabel 3. Hasil *Cluster* Menggunakan Jarak *Manhattan*

C	Jumlah Anggota	Kabupaten/Kota
1	1	Kota Manado
2	4	Kab. Bolaang Mongondow, Kab. Minahasa, Kab. Minahasa Selatan, Kab. Minahasa Utara
3	10	Kab. Kep. Sangihe, Kab. Kepulauan Talaud, Kab. Bolaang Mongondow Utara, Kab. Kepulauan Siau Tagulandang Biaro, Kab. Minahasa Tenggara, Kab. Bolaang Mongondow Selatan, Kab. Bolaang Mongondow Timur, Kota Bitung, Kota Tomohon, Kota Kotamobagu

Pada Tabel 4, dapat diketahui masing-masing anggota *cluster* dari ketiga *cluster* yang terbentuk. *Cluster* 1 terdiri dari 1 Kabupaten/Kota, *cluster* 2 terdiri dari 4 Kabupaten/Kota, *cluster* 3 terdiri dari 10 Kabupaten/Kota.

Hasil 5

Tabel 5. Perbandingan Hasil Simpangan Baku Jarak *Euclidean* dan *Manhattan*

Simpangan Baku	K-Means	
	<i>Euclidean</i>	<i>Manhattan</i>
Simpangan baku dalam <i>cluster</i> σ_W	0,2216	0,24534
Simpangan baku antar <i>cluster</i> σ_B	1,61251	1,5366
Rasio simpangan baku <i>cluster</i> σ	0,13743	0,15966

Pada Tabel 5 didapatkan nilai simpangan baku dalam *cluster* σ_W dan antar *cluster* σ_B untuk setiap jarak. Memilih metode yang terbaik dengan melihat nilai simpangan baku dalam *cluster* paling kecil dan nilai simpangan baku antar *cluster* yang terbesar. Berdasarkan Tabel 5 dapat dilihat bahwa metode *k-means* dengan jarak *Euclidean* memiliki nilai simpangan baku dalam *cluster* yang paling kecil dan nilai simpangan baku antar *cluster* yang paling besar sehingga memiliki nilai rasio simpangan baku yang terkecil yaitu 0,13743. Oleh karena itu, jarak *Euclidean* merupakan jarak yang baik untuk digunakan dalam melakukan pengelompokan pada penelitian ini.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan pada 15 Kabupaten/Kota di Provinsi Sulawesi Utara dikelompokkan dalam 3 *cluster*. Hasil pengelompokan menggunakan algoritma *K-means* jarak *Euclidean* yaitu *cluster* 1 terdiri dari 1 Kabupaten/Kota, *cluster* 2 terdiri dari 6 Kabupaten/Kota, *cluster* 3 terdiri dari 8 Kabupaten/Kota. Hasil pengelompokan menggunakan algoritma *K-means* jarak *Manhattan* yaitu *cluster* 1 terdiri dari 1 Kabupaten/Kota, *cluster* 2 terdiri dari 4 Kabupaten/Kota, *cluster* 3 terdiri dari 10 Kabupaten/Kota. Jarak *Euclidean* memiliki nilai rasio simpangan baku yang terkecil dengan nilai 0,13743. Sehingga jarak tersebut baik digunakan untuk melakukan pengelompokan dalam penelitian ini.

DAFTAR PUSTAKA

- Bunkers, M.J., & dkk. 1996. Definition of Climate Region In The Northern Plains Using an Objective Cluster Modification Technique. *Journal Climate*, Vol 9.
- Hair, et al. 2006. Multivariate data analysis. edisi ke-6. Pearson Education.
- Fuady, M.B.I. dan Nugraha, J. 2017. Implementasi Metode K-Means Dan K-Medoids Untuk Mengelompokkan 82 Kota Di Indonesia Berdasarkan Indeks Harga Konsumen. Prosiding Penelitian Semnas UII 2017. Hlm 327-337

- Mongi, C., 2015. Penggunaan Analisis Two Step Clustering Untuk Data Campuran. *Jurnal d'Cartesian*, **4(1)**:9-19
- Muzakir, A. 2014. Analisis dan Pemanfaatan Algoritma K-Means Clustering pada Data Nilai Siswa sebagai Penentuan Penerima. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST).
- Prasetyo, Lambang Agung. 2012. Aplikasi Sistem Informasi Geografis pada Kerusakan Bangunan Akibat Erupsi Merapi Tahun 2010 di Kabupaten Sleman Daerah Istimewa Yogyakarta Berbasis WEB. [skripsi]. FMIPA, UII.
- Sangga, V.A.P. 2018. Perbandingan Algoritma K-Means Dan Algoritma K-Medoids Dalam Pengelompokan Komoditas Peternakan Di Provinsi Jawa Tengah Tahun 2015 [skripsi]. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.