

KLASTERISASI FAKTOR-FAKTOR KEMISKINAN DI PROVINSI JAWA BARAT MENGGUNAKAN *K-MEDOIDS CLUSTERING*

Pradini Nurul Safitri¹, Rafika Aristawidya², Shafa Bunga Faradilla³

Corresponding author : S. B. Faradilla

Universitas Islam Indonesia, 18611049@students.uui.ac.id¹

Universitas Islam Indonesia, 18611050@students.uui.ac.id²

Universitas Islam Indonesia, 18611064@students.uui.ac.id³

Received : 17 September 2021, Revised : 23 Oktober 2021, Accepted : 25 Oktober 2021

© Mathematics Education Unugiri 2021

Abstract

Difficulty in meeting basic food and non-food need measured on the expenditure side is referred to as poverty. Thus, the purpose of this study is to find out the poverty group by district/city in West Java Province. This method used is the K-Medoids Clustering algorithm because K-Medoids can handle data that is very sensitive to outliers. The results of the study showed that three groups were obtained, with group one having as many as eight regions, group two having members as many as 15 regions, and group three having members as many as four regions.

Keywords: *Poverty, K-Medoids Clustering, West Java Province*

Abstrak

Kesulitan untuk memenuhi kebutuhan dasar makanan dan selain makanan diukur berdasarkan sisi pengeluaran disebut sebagai kemiskinan. Sehingga, tujuan penelitian ini yaitu mengetahui kelompok kemiskinan menurut kabupaten/kota di Provinsi Jawa Barat. Adapun metode yang digunakan yaitu algoritma *K-Medoids Clustering*, dikarenakan *K-Medoids* mampu mengatasi data yang sangat sensitif terhadap *outlier*. Hasil dari penelitian menunjukkan bahwa diperoleh tiga kelompok, dengan kelompok satu memiliki anggota sebanyak delapan wilayah, kelompok dua memiliki anggota sebanyak yaitu 15 wilayah, dan kelompok tiga memiliki anggota sebanyak empat wilayah.

Kata kunci: *Kemiskinan, K-Medoids Clustering, Provinsi Jawa Barat*

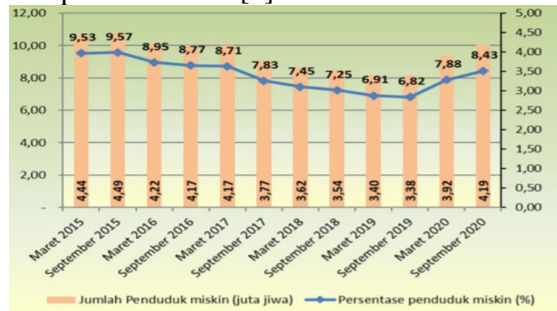
1. Pendahuluan

Kemiskinan adalah masalah global yang sering dihadapi dan menjadi perhatian dunia. Kemudian, arti lain dari kemiskinan juga dapat dikatakan seperti masyarakat yang menganggur dan kurangnya lahan produktif sebagai asset pendapatan untuk masyarakat miskin atau masyarakat yang kurang mampu untuk memperoleh suatu kebutuhan paling dasar yaitu makan, air dan tempat tinggal [1].

Negara miskin sering kali dihadapkan pada masalah pertumbuhan dan pendapatan yang tidak merata, disisi lain banyak negara

berkembang mengalami pertumbuhan ekonomi yang tinggi, namun kurang memberi manfaat bagi penduduk yang tergolong miskin. Indonesia sendiri tergolong negara berkembang, masalah kemiskinan selalu menjadi perhatian. Berbagai permasalahan yang muncul akan menyebabkan terjadinya pengangguran, kriminalitas, sulit untuk mendapatkan tempat tinggal yang baik, dan kemacetan. Dengan demikian, hal tersebut disebabkan karena jumlah penduduk yang semakin meningkat disuatu kabupaten/kota [1]. Provinsi Jawa Barat merupakan salah satu wilayah yang memiliki permasalahan kemiskinan yang

meningkat, hal tersebut dapat dilihat berdasarkan grafik pada Gambar 1[2].



Gambar 1 Presentase Penduduk Miskin Jawa Barat tahun 2015-2017

Berdasarkan Gambar 1 periode Maret 2015 - September 2019 tingkat kemiskinan di daerah Jawa Barat mengalami tren cenderung menurun. Sedangkan pada periode Maret - September 2020 mengalami kenaikan.

Penelitian ini menggunakan metode non hirarki yakni *K-Medoids*. Pusat klaster dari metode *K-Medoids* ditentukan dari perwakilan *medoid* atau objek. Sehingga, peneliti menggunakan metode *K-Medoids* untuk mengelompokkan kemiskinan di Provinsi Jawa Barat. Dengan demikian, peneliti ingin mengetahui kelompok kemiskinan berdasarkan faktor-faktor kemiskinan di Provinsi Jawa Barat menggunakan *k-medoids clustering*.

2. Penelitian Terdahulu

Berdasarkan faktor-faktor mengenai kemiskinan bahwa pada tahun 2017, yang meliputi rata-rata lama sekolah, presentase rumah tangga yang bekerja di pertanian, tingkat pengangguran terbuka, dan pengeluaran rumah tangga per kapita menggunakan metode *average linkage hierarchical clustering* menunjukkan bahwa diperoleh dua kelompok kabupaten/kota di Pulau Jawa dengan tingkat kemiskinan rendah dan tinggi. Adapun kelompok tingkat kemiskinan rendah terdiri dari kota besar di Indonesia, sedangkan kelompok dengan tingkat kemiskinan tinggi meliputi wilayah pedesaan pada kabupaten/kota [1].

Adapun penelitian [3] menggunakan metode *cluster average linkage* meliputi angka melek huruf, tingkat pengangguran terbuka, angka partisipasi sekolah, dan pendidikan yang merupakan faktor-faktor kemiskinan pada tahun 2017 maka diperoleh sebanyak tiga kelompok dengan tingkat kemiskinan rendah, sedang, dan tinggi.

Mengacu pada *website Open Knowledge Maps* menunjukkan bahwa belum dilakukan adanya penelitian terhadap faktor-faktor atau komponen-komponen kemiskinan di Provinsi Jawa Barat menggunakan *K-Medoids Clustering*. *K-Medoids Clustering* mirip dengan Algoritma *K-Means* [4], dimana *K-Means* sangat sensitif terhadap *outlier* sehingga mampu diatasi oleh *K-Medoids* [5]. Persamaan penelitian ini dengan beberapa penelitian sebelumnya yaitu serupa dalam melakukan pengelompokan berdasarkan faktor kemiskinan. Adapun faktor-faktor tersebut dalam penelitian ini meliputi jumlah penduduk miskin, tingkat pengangguran terbuka, rata-rata lama sekolah, dan produk domestik regional bruto. Sehingga, peneliti tertarik untuk melakukan pengelompokan di Provinsi Jawa Barat berdasarkan faktor kemiskinan pada tahun 2020 di suatu kabupaten/kota.

Algoritma *K-Medoids Clustering* digunakan dalam penelitian, yang menunjukkan bahwa masyarakat yang berasal dari wilayah Indonesia dengan ciri suhu badan diatas 36,9 derajat *celcius* dan disertai dengan batuk serta demam [6].

Penelitian [7] menggunakan algoritma *K-Medoids Clustering* menunjukkan bahwa diperoleh tiga kelompok, dengan kelompok tunagrahita ringan berjumlah 5 siswa, kelompok tunagrahita sedang berjumlah 16 siswa dan kelompok tunagrahita berat berjumlah 6 siswa.

Adapun dalam penelitian [8] menggunakan *K-Medoids Clustering* yang bertujuan untuk mengetahui potensi daerah rawan gempa bumi, yaitu diperoleh tiga kelompok yang terbentuk. Kelompok pertama dicirikan dengan kekuatan gempa antara 3 SR hingga 3,4 SR, kelompok kedua dicirikan dengan kekuatan gempa pada daerah rawan gempa bumi antara 3,5 SR hingga 4,5 SR, dan kelompok ketiga dicirikan dengan kekuatan gempa pada daerah rawan gempa bumi yang berkisar antara 4,5 SR sampai dengan 5,7 SR.

3. Landasan Teori

3.1. Kemiskinan

Dalam [9] menurut *World Bank*, salah satu penyebab kemiskinan adalah pendapatan dan aset yang kurang untuk memenuhi kebutuhan dasar. Seseorang dengan keterbatasan lapangan pekerjaan berkaian dengan kemiskinan. Kemiskinan disebabkan oleh tingkat pendidikan dan kesehatan yang tidak layak, serta tidak memiliki pekerjaan disebut pengangguran.

Menurut [10], kesulitan untuk memenuhi kebutuhan dasar makanan dan selain makanan yang diukur berdasarkan aspek pengeluaran disebut sebagai kemiskinan. BPS mengukur kemiskinan menggunakan Garis Kemiskinan (GK). GK yang dihasilkan hanya menggambarkan kondisi kemiskinan penduduk makro yang dapat digunakan untuk menggambarkan situasi kemiskinan di suatu daerah.

3.2. Multikolinieritas

Multikolinieritas yaitu dimana terjadi korelasi yang kuat antar variabel sehingga variabel-variabel yang digunakan harus terhindar dari multikolinearitas. Multikolinearitas bisa dideteksi dengan salah satu cara yaitu menghitung nilai *tolerance/VIF*. Nilai $VIF < 0,1$ atau nilai $VIF > 10$ maka terjadi multikolinearitas antar variabel [11].

3.3. K-Medoids Clustering

K-Medoids merupakan metode partisi *clustering* untuk mengelompokkan kumpulan n objek ke dalam k kluster. *K-Medoids* menggunakan *medoid* (objek) dengan menggunakan perwakilan pusat kluster untuk tiap-tiap kluster. Setiap objek yang berdekatan dengan pusat kluster akan dikumpulkan untuk membentuk sebuah kluster yang baru. *K-Medoids* mempunyai keunggulan dalam mengatasi kekurangan pada *K-Means* yang sensitif pada *outlier*. Meskipun dilakukan pengacakan terhadap data, hasil pengelompokan pada *K-Medoids* tetap sama [5]. Berikut tahapan-tahapan dari algoritma *K-Medoids*.

1. Memilih acak objek sebanyak k dari n objek yang ada sebagai representatif (*medoids*) banyaknya kluster yang ingin dibentuk.
2. Menghitung *distance* setiap objek *non-medoid* terhadap setiap *medoid* dan menempatkan setiap objek ke kluster terdekat dengan *medoid*-nya. Hitung *distance* setiap objek terhadap masing-masing *medoids* dapat dihitung dengan jarak *Euclidean* seperti berikut.

$$(x_{ij}, O_{im}) = \sqrt{(x_{i1} - O_{m1})^2 + (x_{i2} - O_{m2})^2 + \dots + (x_{iq} - O_{mq})^2}$$

(Error! No text of specified style in document..1)

- dengan $d(x_{ij}, O_{mj})$ adalah *distance* dari data ke- i pada variabel ke- j terhadap *medoids* ke- m dimana $m = 1, 2, \dots, k$ serta $j = 1, 2, \dots, q$.
3. Seleksi dengan acak pada objek tiap kluster selaku calon *medoid* baru
4. Kalkulasikan suatu *distance* pada masing-masing objek yang terdapat di tiap-tiap kluster melalui kandidat *medoid* baru
5. Kalkulasikan jumlah seluruh simpangan atau S dengan mengkalkulasikan nilai jumlah *distance* baru dengan jumlah *distance* lama. Apabila $S < 0$, lakukanlah penukaran suatu objek pada data kluster agar terbentuk serangkaian k objek baru untuk *medoid*
6. Berikutnya lakukan pengulangan pada tahap ketiga sampai pada tahap kelima, sampai *medoid* tidak lagi mengalami perubahan, sehingga diperoleh kluster dan juga anggota masing-masing kluster.

4. Metode Penelitian

Menggunakan data sekunder yang diambil pada situs BPS Jawa Barat. Populasinya adalah seluruh kabupaten/kota di Provinsi Jawa Barat, sedangkan masing-masing kabupaten/kota di Provinsi Jawa Barat merupakan sampel. Metode yang digunakan yaitu *K-Medoids Clustering*.

Variabel-variabel penelitian yang digunakan antara lain empat variabel yakni Kemiskinan (X1), Tingkat Pengangguran Terbuka (X2), Rata-rata Lama Sekolah (X3), dan Produk Domestik Regional Bruto (PDRB) (X4).

Adapun langkah-langkah analisis dalam penelitian ini adalah sebagai berikut.

- a. Mendeskripsikan setiap variabel indikator untuk mengetahui karakteristiknya.
- b. Melakukan pengumpulan data dan input data
- c. Melakukan uji multikolinieritas
- d. Melakukan penentuan jumlah kluster
- e. Jika jumlah kluster sudah ditentukan, maka dilanjutkan dengan melakukan analisis *K-Medoids Clustering*
- f. Menarik kesimpulan

5. Pembahasan

5.1. Statistika Deskriptif

Statistika deskriptif berfungsi untuk memberikan sebuah gambaran umum mengenai variabel yang digunakan yaitu Kemiskinan (X1),

Tingkat Pengangguran Terbuka (X2), Rata-rata Lama Sekolah (X3), dan Produk Domestik Regional Bruto (PDRB) (X4). Analisis deskriptif yang diperoleh dapat dilihat pada Tabel 1 berikut ini.

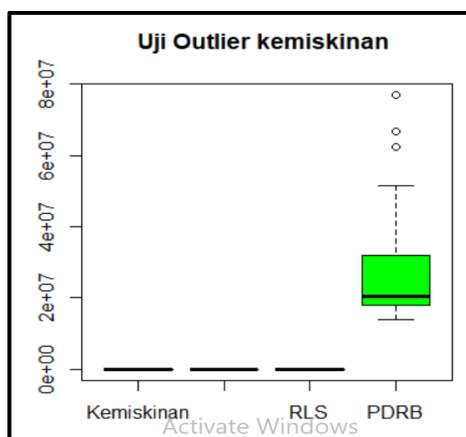
Tabel 1. Analisis Deskriptif

Hasil	Variabel			
	X1 (%)	X2 (%)	X3(Thn)	X4 (Rp)
Min.	2.450	5.080	6.30	13866961
Mean	8.425	9.980	8.54	28323931
Max.	12.970	14.290	11.28	76947023

Berdasarkan Tabel 1. Gambaran umum pada variabel Kemiskinan, Pengangguran, RLS, dan PDRB berdasarkan kabupaten/kota di Provinsi Jawa Barat tertinggi yaitu variabel Kemiskinan berada pada wilayah kota Tasikmalaya sebesar 12.970, variabel Pengangguran di Bogor sebesar 14.290, variabel RLS di kota Depok sebesar 11.280, dan variabel PDRB sebesar 76947023 di kota Sukabumi. Kemudian untuk nilai terendah pada variabel Kemiskinan berada di wilayah kota Depok sebesar 2.450, variabel Pengangguran di Pangandaran sebesar 5.080, variabel RLS di Indramayu sebesar 6.300, dan variabel PDRB sebesar 13866961 di Garut. Sedangkan untuk rata-rata Kemiskinan adalah 8.425, untuk Pengangguran sebesar 9.980, RLS sebesar 8.541, dan PDRB sebesar 28323931.

5.2. Uji Data Outlier

Data *outlier* adalah data yang sering disebut sebagai data pencilan. Penelitian ini dilakukan uji data outlier menggunakan grafik boxplot. Berikut tampilan *output*-nya.



Gambar 2 Uji Deteksi Outlier

Dari Gambar 2 didapatkan hasil bahwa terdapat adanya *outlier* pada salah satu variabel. Oleh karena

itu, untuk data dengan *outlier* disarankan menggunakan analisis kluster non hirarki yaitu *k-medoids clustering*. Selanjutnya dilakukan standarisasi data seperti Gambar 2.

5.3. Uji Multikolinieritas

Pengecekan multikolinearitas merupakan asumsi dalam analisis kluster non hirarki yaitu diharapkan no multikolinearitas atau sebaiknya tidak ada korelasi antar variabel. Pada pengecekan no multikolinearitas ini menggunakan nilai VIF atau nilai dari matriks korelasi dari semua variabel. Tabel 2 berikut merupakan hasil yang diperoleh dari pengujian multikolinieritas.

Tabel 2. Uji Multikolinieritas

Variabel	VIF
X1	2.098643
X2	1.179057
X3	2.211367
X4	1.099583

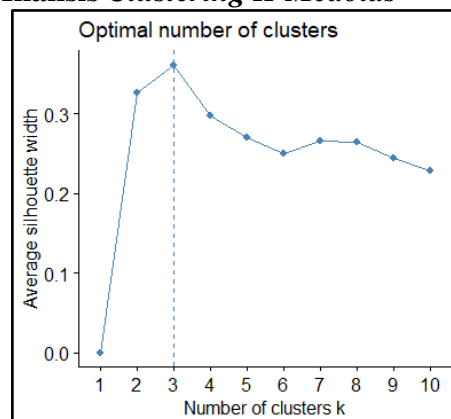
Pada Tabel 2 terlihat bahwa nilai VIF yang diberikan antar variabel Kemiskinan, Pengangguran, RLS, dan PDRB memiliki nilai VIF kurang dari 10 ($VIF < 10$). Dilakukan uji hipotesis berikut.

H_0 : No multikolinearitas

H_1 : Multikolinearitas

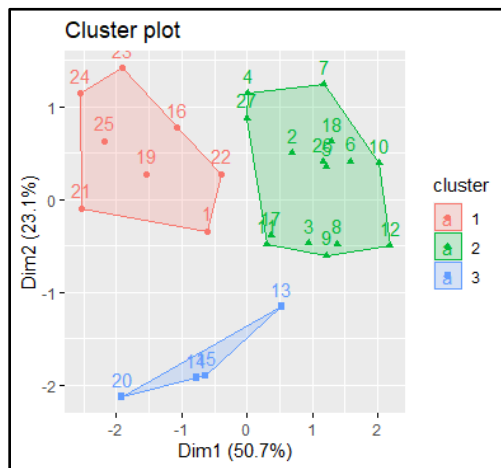
Diperoleh bahwa nilai *VIF* antar semua variabel kurang dari 10 ($VIF < 10$). Daerah kritis merupakan daerah penolakan, dimana H_0 ditolak jika $VIF > 10$. Karena semua nilai $VIF < 10$, maka keputusan yang diperoleh ialah gagal tolak H_0 . Maka disimpulkan bahwa pada tingkat kepercayaan 95%, diperoleh bahwa data tersebut tidak terdapat korelasi antar semua variabelnya.

5.4. Analisis Clustering K-Medoids



Gambar 3 Penentuan Kluster Optimal

Metode *silhouette* pada Gambar 3 menunjukkan hasil penentuan jumlah k kluster kemiskinan di provinsi Jawa Barat, dengan pendekatan nilai rata-rata yang digunakan untuk memprediksi kualitas dari kluster yang terbentuk. Dengan demikian, apabila nilai rata-ratanya semakin tinggi maka jumlah kluster yang terbentuk akan semakin bagus dan optimal, dimana jumlah k kluster optimum ketika $k = 3$.



Gambar 4 Plot Hasil Kluster

Tabel 3. Hasil Pengelompokan Kluster

Klaster	Pusat Klaster	Anggota	Jumlah Anggota
1	Bogor	Bogor, Bekasi, Kota Bogor, Kota Bandung, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi	8
2	Garut	Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Bandung Barat, Pangandaran, Kota Tasikmalaya, Kota Banjar.	15

Klaster	Pusat Klaster	Anggota	Jumlah Anggota
3	Purwakarta	Subang, Purwakarta, Karawang, Kota Sukabumi.	4

Setelah jumlah k telah ditetapkan, langkah selanjutnya yaitu melakukan proses kluster. Klusterisasi yang dilakukan bertujuan untuk mengelompokkan kabupaten/kota dengan mengelompokkan menjadi 3 kelompok (Tinggi, Sedang, dan Rendah). Berikut tampilan plot visualisasi *K-Medoids Clustering* yang diperoleh.

Pada Gambar 4 diperoleh 3 kluster kemiskinan dengan masing-masing kluster memiliki ciri khas warna yang berbeda-beda. Adapun warna merah menunjukkan kluster 1 yang beranggotakan 8 kota, warna hijau menunjukkan kluster 2 yang beranggotakan 15 kota dan warna biru menunjukkan kluster 3 yang beranggotakan 4 kota. Untuk mempermudah dalam membacanya, berikut pemaparan tabel pengelompokan di wilayah Jawa Barat.

Pada Tabel 3 merupakan hasil pengelompokan faktor kemiskinan menjadi 3 kelompok yang memiliki jumlah anggota berbeda-beda pada setiap kluster. Kelompok 1 memiliki anggota sebanyak 8 wilayah, kelompok 2 memiliki anggota sebanyak yaitu 15 wilayah, dan kelompok 3 memiliki anggota sebanyak 4 wilayah.

5.5. Profilisasi

Setelah mengetahui jumlah dan anggota kluster, selanjutnya melakukan profilisasi untuk mengetahui karakteristik dari setiap kelompok. Dalam melakukan profilisasi ini digunakan hasil rata-rata variabel pada setiap kluster

Tabel 4. Hasil Profilisasi

Klaster	X1	X2	X3	X4
1	5.58	11.81	10.22	26828650
2	9.95	8.71	7.74	19837542
3	8.35	11.06	8.13	63138455

Berdasarkan Tabel 4, maka dapat dilakukan interpretasi sebagai berikut.

- Klaster 1 merupakan kelompok kabupaten/kota dengan ciri jumlah Pengangguran dan RLS yang tinggi, Kemiskinan yang rendah, dan PDRB yang sedang.
- Klaster 2 merupakan kelompok kabupaten/kota dengan ciri jumlah Kemiskinan yang tinggi,

jumlah Pengangguran, RLS, dan PDRB yang rendah.

- c. Klaster 3 merupakan kelompok kabupaten/kota dengan ciri jumlah Kemiskinan, Pengangguran dan RLS yang sedang, serta jumlah PDRB yang tinggi.

6. Penutup

Berdasarkan hasil penelitian yang diuraikan dapat ditarik kesimpulan bahwa Gambaran umum berdasarkan hasil analisis deskriptif yaitu Kota Tasikmalaya merupakan daerah yang memiliki jumlah Kemiskinan terbesar, dan Bogor merupakan wilayah yang memiliki jumlah Pengangguran terbesar. Sedangkan kota Depok merupakan wilayah yang memiliki jumlah RLS terbanyak, dan kota Sukabumi merupakan wilayah yang memiliki jumlah PDRB terbanyak.

Pengelompokan dengan menggunakan metode *K-medoids clustering* diperoleh hasil kluster 1 sebanyak 8 anggota. Klaster 2 sebanyak 15 anggota. Sedangkan klaster 3 sebanyak 4 anggota.

Berdasarkan karakteristik dari klaster 1, dengan jumlah Pengangguran tinggi dan Kemiskinan rendah, serta RLS tinggi dan PDRB sedang sehingga peneliti menamakannya sebagai wilayah dengan variabel kemiskinan yang cukup baik. Klaster 2 dengan jumlah Kemiskinan tinggi dan Pengangguran yang rendah, serta RLS dan PDRB yang rendah sehingga peneliti menamakannya sebagai wilayah dengan variabel kemiskinan yang kurang baik. Klaster 3 dengan jumlah Pengangguran dan Kemiskinan sedang, serta RLS sedang dan PDRB tinggi sehingga peneliti menamakannya sebagai wilayah dengan variabel kemiskinan yang cukup baik.

Referensi

- [1] A. W. Disty Wahyuli, Handrizal Handrizal, Iin Parlina, Agus Perdana Windarto, Dedi Suhendro, "Mengelompokan Garis Kemiskinan Menurut Provinsi Menggunakan Algoritma K-Medoids," *J. Ris. Sist. Inf. dan Tek. Inform.*, vol. 1, pp. 452–461, 2019, [Online]. Available: <https://tunasbangsa.ac.id/seminar/index.php/senaris/article/view/51%0A>.
- [2] Badan Pusat Statistik Jawa Barat, *Perkembangan Tingkat Kemiskinan Provinsi Jawa Barat*. Bandung: BPS Provinsi Jawa Barat, 2020.
- [3] Q. Nafisah and N. E. Chandra, "Analisis Cluster Average Linkage Berdasarkan Faktor-Faktor Kemiskinan di Provinsi Jawa Timur," *Zeta - Math J.*, vol. 3, no. 2, pp. 31–36, Nov. 2017, doi: 10.31102/zeta.2017.3.2.31-36.
- [4] W. Setiyawati, *Implementasi Algoritma Partitioning Around Medoids (PAM) untuk Pengelompokan Sekolah Menengah Atas di DIY berdasarkan Nilai Daya Serat Ujian Nasional*. Yogyakarta: (Skripsi) Jurusan Teknik Informatika Universitas Sanata Dharma, 2017.
- [5] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm For Big Data," *Procedia Comput. Sci.*, vol. 78, no. December 2015, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.
- [6] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.
- [7] R. R. Husin Sariangshah*, Wanayumini, "Penentuan Kelas Menggunakan Algoritma K Medoids Untuk Clustering Siswa Tunagrahita," *J. Media ...*, vol. 5, pp. 83–89, 2021, doi: 10.30865/mib.v5i1.2547.
- [8] D. Kurmiati, M. Z. Fauzi, Ripangi, A. Falegas, and Indria, "Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 45–57, 2021.
- [9] D. V. Ferezagia, "Analisis Tingkat Kemiskinan di Indonesia," *J. Sos. Hum. Terap.*, vol. 1, no. 1, pp. 1–6, Jul. 2018, doi: 10.7454/jsht.v1i1.6.
- [10] Badan Pusat Statistik, "Kemiskinan dan Ketimpangan," 2020. <https://www.bps.go.id/subject/23/kemiskinaan-dan-ketimpangan.html#subjekViewTab1>.
- [11] I. Ghozali, *Aplikasi Analisis Multivariat Edisi 4 dengan Program SPSS BP Universitas Diponegoro*. Semarang:

BPUniversitas Diponegoro, 2007.