

Analisis *Clustering* K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020

Febiyanti Alfiah¹⁾, Almadayani¹⁾, Danial Al Farizi¹⁾, Edy Widodo¹⁾

¹⁾Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta, Indonesia

febiantialfiah1010@gmail.com; almadayani045@gmail.com; edywidodo@uii.ac.id

ABSTRAK

Keberadaan pandemi COVID-19 di Indonesia, mengakibatkan kemiskinan di Indonesia semakin tinggi terutama di Jawa Timur yang menjadi satu diantara provinsi lain dengan kasus COVID-19 tinggi di Indonesia. Tujuan penelitian ini yaitu mengetahui pengelompokan kabupaten/kota di Jawa Timur yang mempunyai kesamaan karakteristik berdasarkan indikator kemiskinan tahun 2020. Penelitian ini menggunakan data yang didapatkan dari Badan Pusat Statistik. Metode yang digunakan ialah metode *k-medoids clustering* yang merupakan metode partisi *clustering* guna pengelompokan n objek ke dalam k cluster. Berdasarkan hasil penelitian, diperoleh pengelompokan karakteristik masing-masing cluster yang dibentuk berdasarkan nilai indikator kemiskinan di Jawa Timur tahun 2020 sebanyak 2 cluster. Dimana 30 kabupaten/kota pada cluster 1 dan 8 kabupaten/kota pada cluster 2. Cluster 1 memiliki karakteristik Persentase Rumah Tangga yang Mempunyai Sanitasi Layak, Angka Harapan Hidup, dan Persentase Angka Melek Huruf Umur 15-55 Th tinggi. Sedangkan cluster 2 memiliki karakteristik Persentase Rumah Tangga Miskin Penerima Raskin, Persentase Penduduk Miskin, dan Persentase Pengeluaran Perkapita untuk Makanan dengan Status Miskin tinggi.

Kata kunci: *Clustering*; Jawa Timur; *K-medoids*; kemiskinan

K-Medoids Clustering Analysis Based on Poverty Indicators in East Java in 2020

ABSTRACT

The existence of the pandemic COVID-19 in Indonesia has resulted in higher poverty in Indonesia, especially in East Java, which is one of the other provinces with high cases in Indonesia. The purpose of this study is to find out the grouping of regencies/cities in East Java that have similar characteristics based on the poverty indicators in 2020. This study uses data obtained from the *Badan Pusat Statistik*. The method used is *k-medoids clustering* method which is a clustering partition method for grouping n objects into k clusters. Based on the results of the study, it was found that the grouping of the characteristics of each cluster formed based on the value of the poverty indicator in East Java in 2020 was 2 clusters. Where 30 regencies/cities in cluster 1 and 8 regencies/cities in cluster 2. Cluster 1 has the characteristics of the percentage of households that have proper sanitation, life expectancy, and a high percentage of literacy rates aged 15-55 years. While cluster 2 has the characteristics of the percentage of poor households receiving *Raskin*, the percentage of poor people, and the percentage of per capita expenditure on food with high poor status.

Keywords: Clustering; East Java; K-Medoids; poverty

(Article History: Received 26-09-2021; Accepted 29-11-2021; Published 05-12-2021)

PENDAHULUAN

Salah satu permasalahan yang hampir dialami oleh semua negara adalah kemiskinan. Secara umum, kemiskinan ialah suatu keadaan seseorang atau kelompok orang yang tidak dapat mencukupi kebutuhan dasarnya (Hendayanti & Nurhidayati, 2021).

Kemiskinan yang terus meningkat menjadi permasalahan negara Indonesia yang masih belum terselesaikan sampai saat ini (Wulandary, 2015). Berbagai bentuk program dan bantuan dari pemerintah telah dilakukan untuk mengentaskan kemiskinan. Namun hal itu masih belum dapat menyelesaikan permasalahan kemiskinan di Indonesia.

Pada Maret 2020, Indonesia dikejutkan dengan wabah COVID-19 yang menyebabkan tingkat kemiskinan di Indonesia terutama Jawa Timur mengalami peningkatan. Kemiskinan di Jawa Timur meningkat sebesar 4,58 juta dibanding dengan sebelum adanya pandemi (Wijayanto, 2021). Dengan adanya pandemi tersebut, pemerintah melakukan upaya penanggulangan, salah satunya adalah pembatasan aktivitas masyarakat yang hampir berdampak pada semua sektor. Salah satu pengaruh dari pembatasan kegiatan masyarakat yaitu pada sektor bisnis yang kemudian berdampak pada perekonomian. Dengan adanya pengelompokan wilayah, dapat mengetahui masalah kemiskinan di setiap kabupaten/kota di Jawa Timur. Sehingga dapat mempermudah pemerintah dalam menjalankan kebijakan yang akan ditargetkan di setiap kabupaten/kota di Jawa Timur.

K-Medoids adalah metode partisi dari *clustering* yang menggabungkan n objek ke dalam k cluster. *Medoids* dari setiap cluster menjadi objek dari algoritma *k-medoids*. Setiap objek yang berdekatan dengan pusat cluster akan digabungkan untuk membangun cluster baru (Aurora *et al.*, 2016). Algoritma *K-Medoids* digunakan untuk menangani kekurangan dari algoritma *k-means* atas kesensitifannya terhadap *outlier*. Keberadaan *outlier* pada *k-means* dapat mendistorsi nilai rata-rata cluster karena objek tersebut memiliki karakteristik sangat jauh terhadap sebagian besar data lainnya (Wira *et al.*, 2019). Dengan analisis ini peneliti dapat mengetahui pengelompokan kabupaten/kota yang memiliki karakteristik sama berdasarkan indikator kemiskinan.

Berdasarkan permasalahan kemiskinan di Indonesia, telah dilakukan penelitian dari beberapa peneliti terdahulu diantaranya yaitu penelitian dari (Wahyudi *et al.*, 2019) yang bertujuan untuk melihat tinggi rendahnya tingkat kemiskinan di Indonesia dan diperoleh tingkat kemiskinan tertinggi di Indonesia beranggotakan 23 provinsi dan tingkat kemiskinan terendah beranggotakan 11 provinsi. Dari uraian di atas, membuat peneliti tertarik untuk mengelompokkan wilayah di Jawa Timur yang memiliki karakteristik sama berdasarkan indikator-indikator kemiskinan menggunakan metode *K-Medoids Clustering*, sehingga dapat dengan mudah menemukan solusi atau

mengambil kebijakan untuk mengurangi tingkat kemiskinan di Jawa Timur. Terdapat perbedaan antara penelitian ini dengan kajian pustaka tersebut, diantaranya yaitu penelitian ini melakukan uji asumsi terlebih dahulu dan menggunakan visualisasi peta yang bertujuan untuk mempermudah memahami informasi *clustering*.

Tujuan penelitian ini yaitu untuk mengetahui pengelompokan Kabupaten/Kota di Jawa Timur yang mempunyai kesamaan karakteristik berdasarkan indikator kemiskinan Tahun 2020.

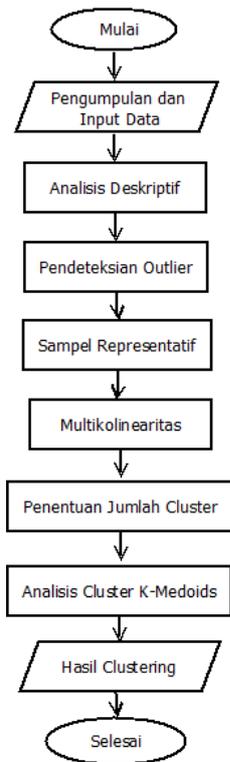
METODE PENELITIAN

Penelitian ini menggunakan data indikator kemiskinan Provinsi Jawa Timur tahun 2020 yang didapatkan dari Badan Pusat Statistik (BPS). Indikator kemiskinan yang digunakan ada 6 indikator yaitu yaitu Persentase Penduduk Miskin, Persentase Angka Melek Huruf Umur 15-55 Th, Persentase Pengeluaran Perkapita untuk Makanan dengan Status Miskin, Persentase Rumah Tangga Penerima Raskin (Beras Miskin), Angka Harapan Hidup/AHH, dan Persentase Rumah Tangga yang Mempunyai Sanitasi Layak (Febianto & Palasara, 2019). Metode yang digunakan ialah metode *K-Medoids Clustering Analysis*. Gambar 1 merupakan diagram alir proses penelitian.

Metode *K-Medoids Clustering*

K-Medoids Clustering merupakan variasi metode *K-Means*. Perbedaan metode ini dibandingkan dengan *k-means* yaitu pada penggunaan *medoids* bukan berdasarkan rata-rata setiap objek cluster (Farissa *et al.*, 2021). *K-Medoids* digunakan untuk mengurangi sensitivitas partisi terhadap nilai *outlier* dalam kumpulan data (Vercillis, 2009). Tujuan dari *K-Medoids Clustering* yaitu menanggulangi kekurangan *K-Means Clustering* dari kesensitifannya terhadap data pencilan (*outlier*) yang dapat berpengaruh pada distribusi data (Han & Kamber, 2017).

K-Medoids Clustering adalah metode partisi dari *clustering* yang menggabungkan n objek ke dalam k cluster. Objek pada sekelompok objek yang mewakili suatu cluster disebut dengan *medoids* (Setyawati, 2017).



Gambar 1. Flowchart penelitian

Tahapan-tahapan menggunakan algoritma K-Medoids sebagai berikut (Wira et al., 2019):

1. Menentukan jumlah *cluster* sebanyak *k* dengan pertimbangan teoritis dan konseptual.
2. Mencari jarak tiap objek terhadap *cluster* terdekat dengan menggunakan ukuran *Euclidian distance*.

$$d_{ab} = \sqrt{\sum_{k=1}^p (x_{ak} - x_{bk})^2} \quad (1)$$

dengan:

- p : Banyaknya variabel
 - d_{ab} : Jarak antara objek a serta objek b
 - x_{ak} : Nilai objek a dari variabel k
 - x_{bk} : Nilai objek b dari variabel k
3. Kemudian menentukan pusat *cluster* baru secara random untuk dijadikan calon *non medoids* pada masing-masing objek.
 4. Menghitung jarak antara setiap objek pada tiap-tiap *cluster* terhadap calon *non medoids*.
 5. Menghitung antara jumlah simpangan (S) dengan jumlah jarak baru – jumlah jarak lama. Apabila $S < 0$, maka mengganti objek dengan data *cluster non medoids* agar terbentuk kumpulan *k* objek baru yang disebut *medoids*.

6. Selanjutnya, mengulangi cara 3 – 5 hingga *medoids* tidak berubah. Kemudian diperoleh *cluster* dan masing-masing anggota *clusternya*.

HASIL DAN PEMBAHASAN

Analisis Deskriptif

Analisis deskriptif diperlukan untuk melihat gambaran/rangkuman umum dari data yang didapatkan dengan melihat nilai minimum (*min*), nilai rata-rata (*mean*), dan nilai maksimal (*max*).

Tabel 1. Deskriptif indikator kemiskinan

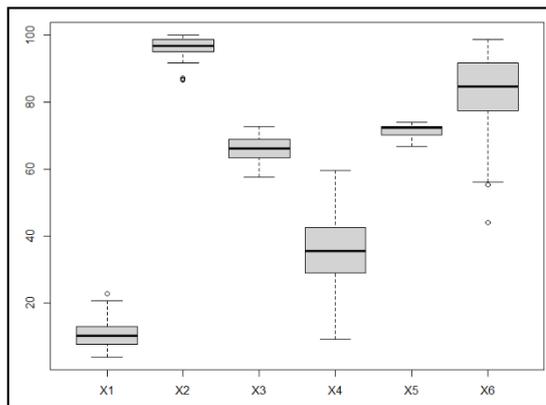
Indikator Kemiskinan		Min	Mean	Max
X_1	Persentase Penduduk Miskin (%)	3,890	11,021	22,780
X_2	Persentase Angka Melek Huruf Umur 15-55 Th (%)	85,56	96,11	100
X_3	Persentase Pengeluaran Perkapita untuk Makanan dengan Status Miskin (%)	57,77	65,97	72,84
X_4	Persentase Rumah Tangga Penerima Raskin (%)	9,14	35,57	59,53
X_5	Angka Harapan Hidup/AHH (%)	66,74	71,64	74,18
X_6	Persentase Rumah Tangga yang Mempunyai Sanitasi Layak (%)	44,07	82,07	98,71

Berdasarkan Tabel 1 didapatkan bahwa Persentase Angka Melek Huruf Umur 15-55 Th tinggi yaitu dengan nilai minimum sebesar 85,56%, nilai rata-rata sebesar 96,11%, dan nilai maksimum sebesar 100%. Kabupaten/kota memiliki tingkat kemiskinan rendah apabila angka Persentase Penduduk Miskin, Persentase Rumah Tangga Miskin Penerima

Raskin, dan Persentase Pengeluaran Perkapita untuk Makanan dengan Status Miskin rendah. Sedangkan angka Persentase Angka Melek Huruf Umur 15-55 Th, AHH, dan Persentase Rumah Tangga yang Mempunyai Sanitasi Layak tinggi.

Pengecekan *Outlier*

Mendeteksi *outlier* menggunakan *boxplot* untuk mengetahui apakah data yang digunakan terdapat data *outlier*. Jika terdapat data atau titik yang keluar dari *boxplot* maka data tersebut memiliki data *outlier* (Gambar 2).



Gambar 2. Output pengecekan *outlier*

Berdasarkan Gambar 2 terdapat titik yang keluar dari *boxplot* yaitu pada variabel X_1 , X_2 , dan X_6 yang menunjukkan bahwa data yang digunakan tersebut mengandung *outlier*. Sehingga metode *cluster* yang cocok digunakan pada analisis ini yaitu metode *K-Medoids Clustering*. Pada *K-Medoids* yang menjadi pusat *cluster* yakni objek representatif (*medoids*) yang kemudian menempatkan setiap objek terhadap *cluster* terdekat dengan *medoids*, sehingga cocok dalam data yang mengandung *outlier*.

Asumsi *Cluster*

Pada penelitian ini terdapat asumsi yang harus dipenuhi yaitu *sample representative* dan tidak terdapat multikolinearitas.

Sample Representative

Berdasarkan hasil dari pengujian KMO (*Kaiser Mayer Olkin*) diperoleh nilai KMO sebesar 0,8046603. Nilai KMO yang diperoleh bernilai antara 0,5-1, sehingga sampel mewakili populasi atau sampel representatif yang artinya uji asumsi untuk sampel representatif terpenuhi (Ningrat *et al.*, 2016).

Uji Multikolinearitas

Pengujian multikolinearitas memperhatikan nilai VIF, jika nilai VIF yang diperoleh ≤ 10 , maka tidak terjadi multikolinearitas (asumsi terpenuhi) (Mahmudan, 2020). Berikut merupakan uji hipotesis dari uji multikolinearitas:

Hipotesis

H_0 : Tidak ada multikolinearitas

H_1 : Ada multikolinearitas

Tingkat Signifikansi, $\alpha = 5\%$

Tabel 2. Output uji multikolinearitas

Indikator	VIF
X_1	1,941098
X_2	2,82901
X_3	1,867921
X_4	1,508485
X_5	3,653026
X_6	3,014712

Berdasarkan Tabel 2 diperoleh semua variabel memiliki nilai VIF < 10 , maka disimpulkan gagal tolak H_0 (tidak ada multikolinieritas). Pada tingkat kepercayaan 95%, data menunjukkan gagal tolak H_0 yang artinya semua indikator kemiskinan di Jawa Timur tahun 2020 berdasarkan Kabupaten/Kota tidak terdapat korelasi antar semua variabelnya.

Penentuan Jumlah *Cluster*

Dalam menentukan jumlah *cluster* atau k menggunakan metode *silhouette*, metode *silhouette* digunakan pendekatan nilai rata-rata guna memprediksi kualitas *cluster* yang terbentuk terbentuk (Anggara, *et al.*, 2016). Sehingga nilai rata-rata yang semakin tinggi akan semakin baik.

Jumlah $s(i)$ diperoleh dengan menggabungkan $a(i)$ dan $b(i)$ (Dewi & Pramita, 2019):

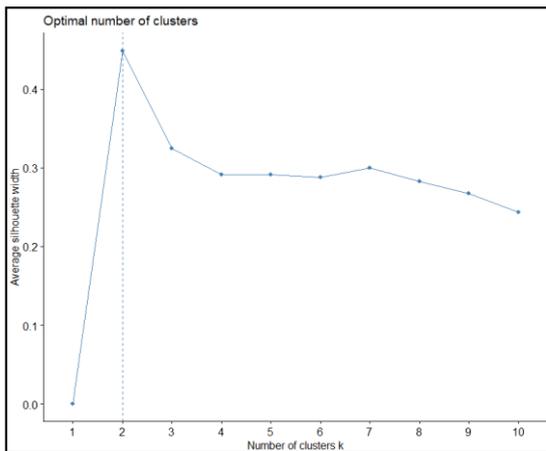
$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)} & \text{if } a(i) < b(i) \\ 0 & \text{if } a = b(i) \\ \frac{b(i)}{a(i)} - 1 & \text{if } a > b(i) \end{cases} \quad (2)$$

Sehingga dapat dirumuskan:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Berdasarkan Gambar 3, diperoleh banyaknya *cluster* optimum yaitu sebanyak 2

atau $k = 2$, Hal ini dilihat dari garis putus-putus pada tampilan grafik metode *silhouette*.



Gambar 3. Output penentuan jumlah k metode *silhouette*

Clustering

Hasil *clustering* menggunakan jarak *Euclidean Distance* dengan *cluster* yang terbentuk sebanyak 2 *cluster* dan jumlah data yang digunakan sebanyak 38. Ada 30 kabupaten/kota pada *cluster 1* serta 8 kabupaten/kota pada *cluster 2* seperti pada Tabel 3.

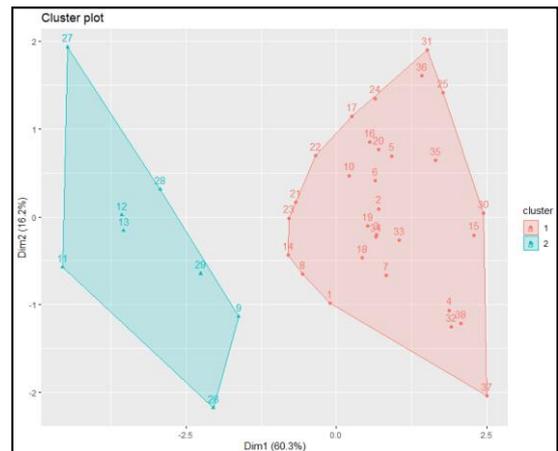
Tabel 3. Hasil *clustering*

Cluster	Kabupaten/Kota	Jumlah
1	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Banyuwangi, Pasuruan, Sidoarjo, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Kota Surabaya, Ngawi, Kota Pasuruan, Bojonegoro, Tuban, Kota Madiun, Lamongan, Kota Probolinggo, Gresik, Kota Kediri, Kota Malang, Kota Blitar, Kota Mojokerto, dan Kota Batu	30
2	Jember, Bondowoso, Situbondo, Probolinggo, Bangkalan, Sampang, Sumenep, dan Pamekasan	8

Dari *output* Gambar 4 didapatkan hasil *clustering* menggunakan *cluster plot*, diketahui bahwa *cluster 1* area warna merah dan *cluster 2* area warna hijau, setiap warna memiliki karakteristik yang berbeda-beda.

Kemudian, untuk melihat karakteristik dari setiap *cluster* dengan melihat nilai rata-

rata dari masing-masing indikator disetiap *clusternya*.



Gambar 4. Output *k-medoids clustering*

Tabel 4. Profilisasi data

Indikator	Cluster 1	Cluster 2
X_1	9,52	16,7
X_2	97,4	91,2
X_3	64,8	70,4
X_4	34,2	40,6
X_5	72,4	68,7
X_6	87,4	62

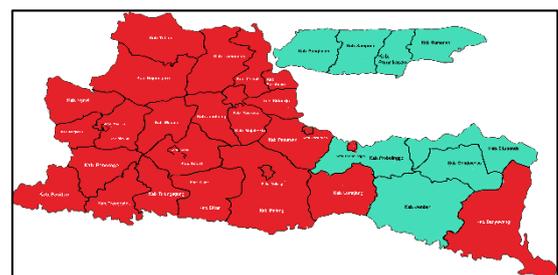
Berdasarkan Tabel 4 karakteristik setiap indikator kemiskinan, didapatkan hasil *clustering* yang diinterpretasikan sebagai berikut:

Tabel 5. Tabel indikator

Cluster	Indikator Tinggi	Indikator Rendah
Cluster 1	X_2, X_5, X_6	X_1, X_3, X_4
Cluster 2	X_1, X_3, X_4	X_2, X_5, X_6

Visualisasi K-Medoids Clustering

Visualisasi ini untuk mempermudah melihat hasil *clustering* berupa sebuah peta Provinsi Jawa Timur.



Gambar 5. Visualisasi *clustering*

Berdasarkan Gambar 5 diketahui bahwa wilayah berwarna merah menunjukkan karakteristik Persentase Angka Melek Huruf Umur 15-55 Th, AHH, dan Persentase Rumah Tangga yang Mempunyai Sanitasi Layak tinggi yang terdiri dari 30 kabupaten/kota. Sedangkan wilayah berwarna hijau menunjukkan karakteristik Persentase Penduduk Miskin, Persentase Rumah Tangga Miskin Penerima Raskin, dan Persentase Pengeluaran Perkapita untuk Makanan dengan Status Miskin tinggi yang terdiri dari 8 kabupaten/kota.

KESIMPULAN

Dengan menggunakan analisis *k-medoids clustering* diperoleh *cluster* optimal sebanyak 2 *cluster*. Terdapat 30 kabupaten/kota pada *cluster* 1 dan 8 kabupaten/kota pada *cluster* 2. *Cluster* 1 memiliki karakteristik Persentase Rumah Tangga yang Mempunyai Sanitasi Layak, AHH, dan Persentase Angka Melek Huruf Umur 15-55 Th tinggi. Sedangkan *cluster* 2 memiliki karakteristik Persentase Rumah Tangga Miskin Penerima Raskin, Persentase Penduduk Miskin, Persentase Pengeluaran Perkapita untuk Makanan dengan Status Miskin tinggi.

DAFTAR PUSTAKA

Anggara, M., Sujiani, H. & Nasution, H. 2016. Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member Di Alvaro Fitness. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, **1(1)**: 1-6.

Aurora, P., Deepali & Varshney, S., 2016. Analysis of K-means an K-Medoids Algorithm for Big Data. *Procedia Computer Science*, **78**: 507-512.

Dewi, D.A.I.C. & Pramita, D.A.K. 2019. Analisis Perbandingan Metode Elbow Dan Silhouette Pada Algoritma Clustering K-Medoids Dalam Pengelompokan Produksi Kerajinan Bali. *Jurnal Matrix*, **9(3)**: 102-109.

Farissa, R.A., Mayasari, R. & Umidah, Y. 2021. Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient. *Journal of Applied Informatics and Computing (JAIC)*, **5(2)**: 109-116.

Febianto, N.I. & Palasara, D.N. 2019. Analisis Clustering K-Means Pada Data Informasi Kemiskinan di Jawa barat Tahun 2018. *SISFOKOM*, **8(2)**: 130-139.

Han, J. & Kamber, M. 2017. Data Mining, Concept and Techniques. Morgan Kauffman Publisher, Waltham.

Hendayanti, N.P.N. & Nurhidayati, M. 2021. Klasifikasi Tingkat Keparahan Kemiskinan Provinsi di Indonesia. *Jurnal Matematika dan Pendidikan Matematika*, **5(1)**: 14-21.

Mahmudan, A. 2020. Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering. *Jurnal Matematika, Statistika, & Komputasi (JMSK)*, **17(1)**: 1-13.

Ningrat, D.R., Maruddani, D.A.I. & Wuryandari, T. 2016. Analisis Cluster dengan Algoritma K-Means dan Fuzzy C-Means Clustering untuk Pengelompokan Data Obligasi Korporasi. *Jurnal Gaussian*, **5(4)**: 641-650.

Setyawati, A.W. 2017. Implementasi Algoritma Partitioning Around Medoid (PAM) untuk Pengelompokan Sekolah Menengah Atas di DIY Berdasarkan Nilai Daya Serap Ujian Nasional [Skripsi]. Fakultas Sains dan Teknologi, Universitas Sanata Darma, Yogyakarta.

Vercillis, C. 2009. Business Intelligence, Data Mining and Optimization for Decision Making. Wiley, Milan.

Wahyudi, D., Handrizal, H & Wanto, A. 2019. Mengelompokkan Garis Kemiskinan Menurut Provinsi Menggunakan Algoritma K-Medoids. *Prosiding Seminar Nasional Riset Informasi Sains (SENARIS)*: 452-461.

- Wijayanto. 2021. Dampak Pandemi Covid-19, Penduduk Miskin di Jatim Meningkat. [Online] Availabl at: <https://radarsurabaya.jawapos.com/read/2021/03/30/250918/dampak-pandemi-covid-19-penduduk-miskin-di-jatim-meningkat>
- Wira, B., Budianto, A.E. & Wiguna, A.S. 2019. Implementasi Metode K-Medoids Clustering Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang. *Jurnal Terapan Sains & Teknologi*, **1(3)**: 54-69.
- Wulandary, F.H. 2015. Pengaruh Pertumbuhan Ekonomi Inflasi, Pengangguran, dan Pendidikan Terhadap Kemiskinan Provinsi di Indonesia Tahun 2008-2012 [Skripsi]. Universitas Atma Jaya, Yogyakarta.