

Pengelompokan Provinsi Di Indonesia Berdasarkan Luas Panen, Produksi, Dan Produktivitas Padi Menggunakan Algoritma K-Medoid

Elisa Syafaqoh¹

Program Studi Matematika, Universitas Islam Negeri Sunan Ampel Surabaya¹

e-mail: elisa.syagaqoh26@gmail.com

* Corresponding Author

Received: 20 Agustus 2023; Accepted: 1 September 2023; Published: 30 September 2023

Abstrak. Sektor pertanian merupakan salah satu dari kebutuhan pokok pangan yang menjadi komponen penting dalam mendorong pertumbuhan perekonomian. Salah satu komoditi terbesar dalam sektor pertanian yaitu beras yang merupakan hasil olahan padi yang menjadi makanan pokok mayoritas warga Indonesia. Tujuan dilakukannya penelitian ini untuk mengelompokkan Provinsi di Indonesia berdasarkan luas panen, produksi, dan produktivitas padi menggunakan metode Analisis K-Medoid Clustering (PAM). Cluster yang terbentuk sebanyak 2 cluster dengan Average Silhouette sebesar 0,86 dengan spesifikasi struktur yang terbentuk kuat serta seluruh data telah terletak pada cluster yang tepat. Cluster 1 sebagai kelompok provinsi dengan tingkat produktivitas rendah yang terdiri dari 30 provinsi, dan cluster 2 sebagai kelompok provinsi dengan tingkat produktivitas tinggi yang terdiri dari 4 provinsi. Nilai silhouette dari masing-masing cluster tersebut sebesar 0,89 dengan spesifikasi struktur yang terbentuk kuat dan 0,66 dengan spesifikasi struktur yang terbentuk baik.

Kata Kunci: K-Medoid, koefisien silhouette, clustering, padi

Pendahuluan

Tiga syarat pokok untuk kehidupan yang layak adalah sandang, pangan, dan papan. setiap manusia membutuhkan pangan untuk bertahan hidup, menjadikannya salah satu kebutuhan utama manusia (Marisa et al., 2021). Salah satu kebutuhan pangan yang mendasar dan menjadi faktor kunci dalam mendorong pertumbuhan ekonomi adalah sektor pertanian. Mayoritas penduduk Indonesia mengonsumsi nasi sebagai sumber pangan utama sehingga menjadikannya salah satu hasil pertanian yang sangat penting (Septiadi & Joka, 2019). Berdasarkan data yang dihimpun oleh FAOSTAT tahun 2021, Indonesia menjadi negara yang menghasilkan beras terbanyak keempat di dunia dengan total produksi sebesar 54,415,294 ton (FAOSTAT, n.d.). Berdasarkan data yang dihimpun oleh Badan Pusat Statistik (BPS), luas panen padi di Indonesia mencapai 10,45 juta ha pada tahun 2022. Luasnya meningkat 40 ribu ha dari tahun sebelumnya yang sebesar 10,41 juta ha (BPS, n.d.).

Kebutuhan pangan semakin meningkat sebagai akibat dari pertumbuhan penduduk Indonesia. Kementerian Pertanian telah menerapkan sejumlah skema untuk mendorong produktivitas padi dalam upaya memenuhi kebutuhan beras penduduk. Untuk mendorong hasil padi, Kementerian Pertanian telah membuat serta menjalankan P3KP atau Tujuh Strategi Utama Penguatan Pembangunan Pertanian untuk Kedaulatan Pangan (Akbar et al., 2018). Untuk

mengetahui prioritas daerah produktivitas padi yang masih terbelang rendah di Indonesia dapat ditangani menggunakan teknik data mining.

Data mining merupakan teknik pengolahan data untuk mengungkap pola tersembunyi dalam data sehingga dapat menciptakan sebuah pengetahuan (Nurlaela et al., 2020). Salah satu metode data mining yang digunakan pada penelitian ini yaitu metode clustering. Clustering merupakan teknik menyusun kumpulan data kedalam beberapa kelompok yang memungkinkan objek pada satu kelompok memiliki kemiripan serta perbedaan dengan kelompok lain (Rahmayani, 2018). Analisis clustering ini menggunakan algoritma K-Medoid yaitu metode clustering yang menghimpun n objek kedalam k cluster. Objek yang berada pada kumpulan objek yang mewakili suatu cluster disebut dengan medoids (Alfiah et al., 2021).

Beberapa kajian mengenai kualitas K-Medoid clustering terhadap beberapa permasalahan yang telah diusulkan diantaranya yaitu analisis clustering terhadap pengelompokan kondisi jalan di Semarang yang menghasilkan 4 cluster dengan nilai silhouette sebesar 0,57432 (Asmiatun, 2019). Klasterisasi data penyakit pasien di RSUD Bandung menghasilkan nilai silhouette sebesar 0,409373 dengan sebanyak 3 cluster (Andini & Arifin, 2020). Pengelompokan wilayah sebaran cacat pada anak dengan menerapkan metode K-Medoid dan K-Means mendapatkan hasil bahwa K-Medoid dengan nilai silhouette sebesar 0,5009 menjadi metode yang lebih baik dibanding dengan metode K-Means dengan nilai silhouette sebesar 0,1443 (Marlina et al., 2018).

Dari beberapa penelitian menunjukkan bahwa metode K-Medoid Clustering memiliki hasil yang baik. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan Provinsi di Indonesia berdasarkan luas panen, hasil produksi dan produktivitas padi menggunakan metode K-Medoid Clustering.

Metode Penelitian

Clustering

Pada data mining, clustering berfungsi untuk mengelompokkan objek kedalam cluster. Cluster merupakan data yang saling terikat satu sama lain dalam satu kelompok dengan adanya kesamaan karakter diantara sekelompok data atau objek (Dwilestari et al., 2021). Clustering dan klasifikasi seringkali dianggap sama, namun nyatanya terdapat perbedaan yaitu pada proses clustering tidak terdapat variabel target untuk kelompok. Terdapat banyak algoritma clustering dengan keunggulan dan kelemahan tersendiri, namun prinsipnya tetap sama yaitu clustering data berdasarkan karakteristik dan memperhitungkan jarak kesamaan antara data cluster (Pramesti et al., 2017).

Algoritma K-Medoid

Peter J. Rousseeuw dan Leonard Kaufman menciptakan algoritma K-Medoid atau Partitioning Around Medoid (PAM) pada tahun 1987. Metode ini bertujuan untuk menghimpun sekumpulan n objek ke dalam k cluster dan menggunakan medoid sebagai pusat cluster (Wira et al., 2019). Keunggulan metode k-medoid yaitu dapat mengatasi data yang sensitif terhadap noise dan outlier pada dataset yang dapat mempengaruhi distribusi data. Berikut langkah-langkah analisis menggunakan algoritma k-medoid (Dwilestari et al., 2021):

1. Menentukan k cluster
2. Alokasikan setiap data (objek) pada cluster terdekat menggunakan persamaan ukuran jarak Euclidean Distance:

$$d_{euc}(X_{ij}, C_{kj}) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (X_{ij} - C_{kj})^2} \quad (1)$$

Dimana:

$d_{euc}(X_{ij}, C_{kj})$ = Jarak Euclidean antara pengamatan ke-i variabel ke-j ke pusat cluster ke-k pada variabel ke-j

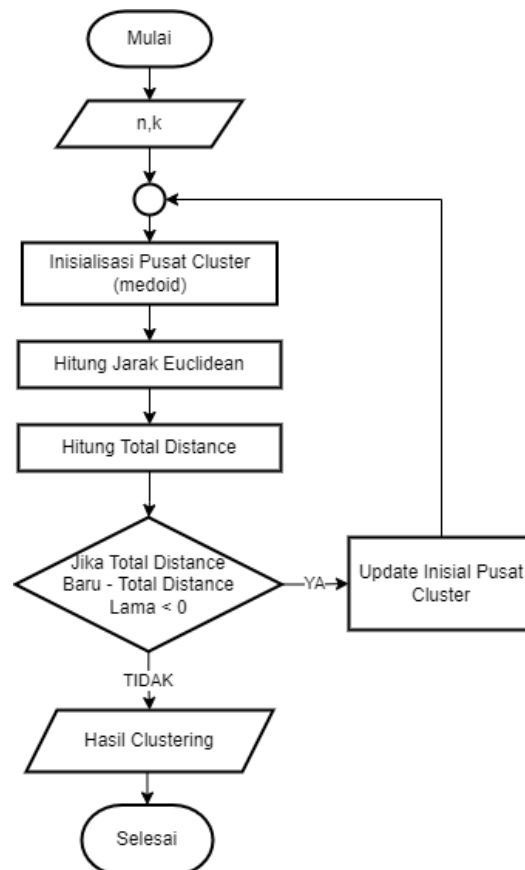
X_{ij} = objek pengamatan ke-i pada variabel ke-j

C_{kj} = pusat kelompok ke-k pada variabel ke-j

p = banyaknya variabel yang diamati

n = banyaknya pengamatan yang diamati

3. Pilih secara acak satu objek pada setiap cluster sebagai calon medoid baru
4. Hitung jarak setiap objek pada cluster menggunakan calon medoid baru dengan persamaan (1)
5. Hitung simpangan total (S) dengan menghitung nilai total distance baru – total distance lama. Apabila $S < 0$, ubah objek dengan data cluster untuk membentuk kumpulan k objek baru sebagai medoid.
6. Ulangi langkah 3,4,5 hingga tidak terjadi perubahan pada medoid, sehingga didapatkan cluster beserta anggota masing-masing.



Gambar 1. Flowchart K-Medoid

Koefisien Silhouette

Koefisien silhouette merupakan gabungan dari metode cohesion yang berfungsi untuk menghitung kedekatan antar sekelompok objek, serta metode separation untuk menghitung jarak antar kelompok dengan kelompok lainnya. Hasil cluster metode K-Medoid dievaluasi menggunakan koefisien silhouette untuk menentukan kualitas hasil cluster. Nilai silhouette dapat dihitung menggunakan persamaan berikut (Handoyo et al., 2014):

$$s_i = \frac{b[i]-a[i]}{\text{Max}(a[i],b[i])} \tag{2}$$

Dimana:

$s(i)$ = silhouette pada titik ke-i

$a(i)$ = rata-rata perbedaan objek ke-i dengan objek lain dalam satu cluster

$b(i)$ = minimal perbedaan objek ke-i dengan objek lain dari cluster lain

Kualitas cluster dikatakan baik apabila nilai silhouette yang dihasilkan semakin mendekati angka 1. Berikut tabel penafsiran nilai koefisien silhouette (Hidayati et al., 2021):

Tabel 1. Interpretasi nilai silhouette

Koefisien Silhouette	Interpretasi
0,71 – 1,00	Struktur yang dihasilkan kuat
0,51 – 0,70	Struktur yang dihasilkan baik
0,26 – 0,50	Struktur yang dihasilkan lemah
≤ 0,25	Tidak terstruktur

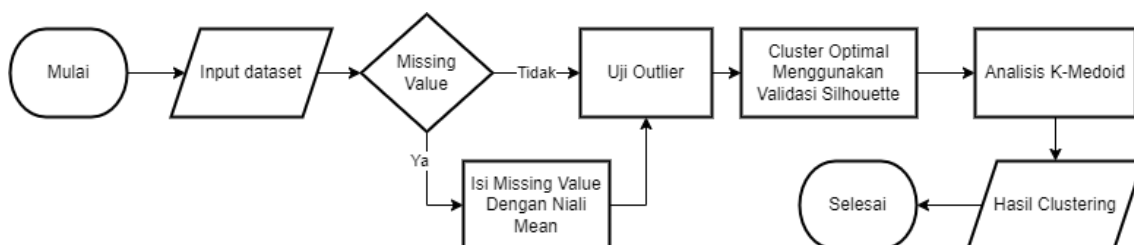
Pengumpulan Data dan langkah Penelitian

Dalam penelitian ini secara keseluruhan dilakukan secara kuantitatif dengan menerapkan algoritma K-Medoid Clustering berdasarkan luas panen (LP), produksi (P), dan produktivitas padi (PP) di Indonesia menggunakan software Rstudio. Data penelitian diperoleh melalui laman web BPS tahun 2022.

Tabel 2. Sampel Data

No	Provinsi	LP	P	PP
1	Aceh	271750,20	55,55	1509456,00
2	Sumatera Utara	411462,10	50,76	2088584,00
3	Suamtera Barat	271883,10	50,52	1373532,00
4	Riau	51054,04	41,83	213557,20
5	Jambi	60539,59	45,88	277743,80
6	Suamtera Selatan	513378,20	54,06	2775069,00

Data pada tabel 2 akan diproses menggunakan algoritma K-Medoid. Proses tersebut dapat dilihat melalui flowchart berikut:



Gambar 2. Flowchart penelitian

Berdasarkan gambar 2 didapat tahap penelitian sebagai berikut:

1. Input data penelitian.
2. Lakukan pengecekan terhadap missing value, apabila terdapat missing value maka dapat diisi dengan nilai mean.

3. Lakukan pengecekan terhadap outlier, apabila terdapat outlier maka dapat diselesaikan menggunakan K-Medoid.
4. Menentukan jumlah cluster optimal menggunakan koefisien silhouette.
5. Melakukan analisis K-Medoid Clustering menggunakan jumlah cluster optimal.
6. Setelah dilakukan analisis, akan didapatkan hasil clustering.

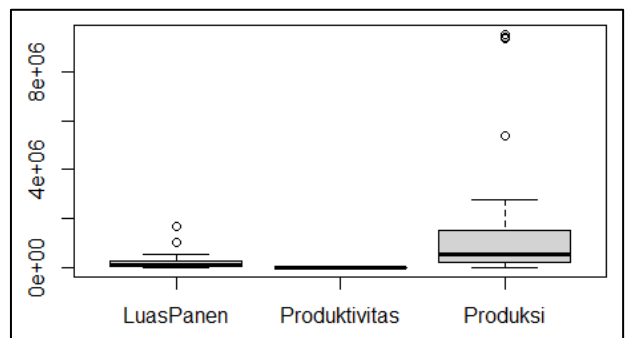
Hasil dan Pembahasan

Dalam melakukan klasterisasi, langkah pertama yang dilakukan yaitu pengecekan terhadap missing value yang menghasilkan seperti gambar berikut.

```
> summary(is.na(dt))
LuasPanen      Produktivitas      Produksi
Mode :logical  Mode :logical  Mode :logical
FALSE:34      FALSE:34      FALSE:34
```

Gambar 3. Output missing value

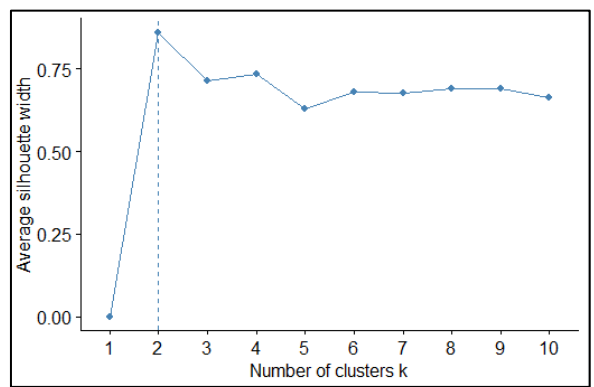
Gambar 3 menunjukkan bahwa tidak terdapat data yang terindikasi missing value, sehingga dapat dilanjutkan pada tahap berikutnya yaitu memeriksa outlier yang menghasilkan seperti gambar berikut.



Gambar 4. Output outlier

Melalui gambar 4 menunjukkan data luas panen dan produksi terdapat outlier, yaitu pencilan yang tidak mengikuti mayoritas pola serta berada jauh dari pusat data. Data yang terdapat outlier dapat mempengaruhi hasil analisis serta tidak mewakili keadaan populasi, namun algoritma K-Medoid memiliki keunggulan dalam mengolah data outlier (Nahdliyah et al., 2019).

Setelah dilakukan pemeriksaan terhadap missing value dan outlier, langkah berikutnya yaitu menentukan jumlah cluster optimal menggunakan koefisien silhouette (SI). Berikut grafik hasil pengujian beberapa cluster menggunakan SI.



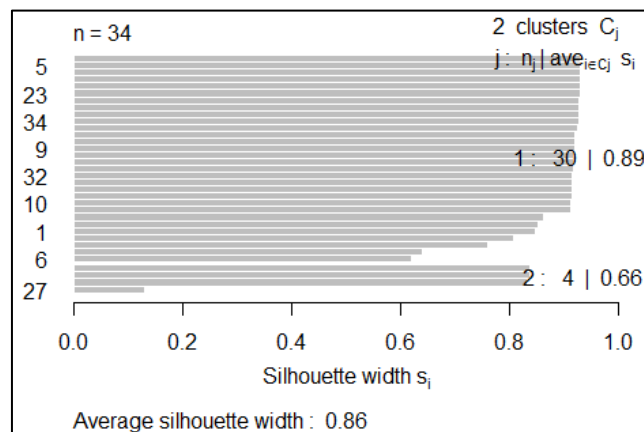
Gambar 5. Grafik cluster optimal

Berdasarkan gambar 5 didapat hasil k optimal 2 cluster dengan nilai average silhouette sebesar 0,86 yang ditandai dengan garis putus-putus. Tahap berikutnya yaitu melakukan klasterisasi menggunakan algoritma K-Medoid. Tahap awal dalam melakukan klasterisasi yaitu menentukan pusat medoid terlebih dahulu, karena k optimal sebanyak 2 cluster maka pusat medoid dapat dilihat pada tabel berikut.

Tabel 3. Pusat medoid

Provinsi	Cluster	LP	P	PP
Sulawesi Barat	1	69323,95	50,99	353513,3
Jawa Barat	2	1662404,00	56,75	9433723,0

Berdasarkan tabel 3 diperoleh pusat medoid pada cluster 1 yaitu Provinsi Sulawesi Barat dan cluster 2 terletak pada Provinsi Jawa Barat. Melalui pusat medoid dapat menentukan kategori setiap cluster yang terbentuk yaitu cluster 1 sebagai provinsi dengan tingkat rendah dan provinsi pada cluster 2 tergolong kategori tinggi. Selanjutnya dilakukan uji silhouette untuk mengetahui bahwa data disetiap cluster telah berada pada cluster yang tepat.



Gambar 6. Plot silhouette

Melalui gambar 6 dapat diketahui bahwa grafik data mengarah ke kanan yang menandakan bahwa semua objek data berada pada rentang positif sehingga data telah terletak pada cluster yang tepat. Apabila terdapat grafik data yang mengarah ke kiri atau berada pada rentang negatif, maka data tersebut terletak pada cluster yang kurang tepat. Selain itu juga dapat diketahui pada cluster 1 yang terdiri dari 30 provinsi dengan nilai silhouette sebesar 0,89 dan pada cluster 2 yang terdiri dari 4 provinsi diperoleh nilai silhouette sebesar 0,66. Selanjutnya hasil pengelompokkan dapat divisualisasikan berupa plot sebagai berikut.



Gambar 7. Output K-Medoid Clustering

Gambar 7 menunjukkan bahwa berdasarkan hasil analisis K-Medoid Clustering terbentuk 2 cluster, dengan cluster 1 sebagai provinsi dengan tingkat rendah yang ditandai dengan warna merah sedangkan cluster 2 sebagai provinsi dengan tingkat tinggi yang ditandai dengan warna biru. Untuk lebih jelasnya, hasil clustering akan ditunjukkan dalam tabel berikut.

Tabel 4. Hasil pengelompokan Provinsi

<i>Cluster</i>	<i>Provinsi</i>	<i>Kategori</i>	<i>Silhouette</i>
1	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI. Jakarta, DI. Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua	Rendah	0,89
2	Jawa Barat, Jawa Tengah, Jawa Timur, Sulawesi Selatan	Tinggi	0,66
<i>Average Silhouette Total</i>			0,86

Penutup

Berdasarkan metode K-Medoid didapat hasil cluster optimal yang dilihat melalui validasi silhouette yaitu dengan rata-rata sebesar 0,86 pada $k = 2$ cluster dengan seluruh data telah berada pada cluster yang tepat. Nilai silhouette tersebut termasuk dalam struktur kuat. Pada pengelompokannya terbagi menjadi 2 daerah yaitu daerah produksi tinggi yang terdiri dari Provinsi Jawa Barat, Jawa Tengah, Jawa Timur, dan Sulawesi Selatan. Sedangkan 30 provinsi lainnya tergolong rendah akan produktivitas padi. Bagian ini berisi simpulan sesuai dengan tujuan penelitian dan saran.

Daftar Pustaka

- Akbar, I., Budiraharjo, K., & Mukson, M. (2018). Analisis Faktor-Faktor Yang Mempengaruhi Produktivitas Padi Di Kecamatan Kesesi, Kabupaten Pekalongan. *Agrisocionomics: Jurnal Sosial Ekonomi Pertanian*, 1(2), 99. <https://doi.org/10.14710/agrisocionomics.v1i2.1820>
- Alfiah, F., Almadayani, Al Farizi, D., & Widodo, E. (2021). Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020. *Jurnal Ilmiah Sains*, 22(1), 1. <https://doi.org/10.35799/jis.v22i1.35911>
- Andini, A. D., & Arifin, T. (2020). Implementasi Algoritma K-Medoids Untuk Klasterisasi Data Penyakit Pasien Di Rsud Kota Bandung. *Jurnal Responsif: Riset Sains Dan Informatika*, 2(2), 128–138. <https://doi.org/10.51977/jti.v2i2.247>
- Asmiatun, S. (2019). Penerapan Metode K-Medoids Untuk Pengelompokan Kondisi Jalan Di Kota Semarang. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 6(2), 171–180. <https://doi.org/10.35957/jatisi.v6i2.193>
- BPS. (n.d.). *Luas Panen, Produksi, dan Produktivitas Padi Menurut Provinsi 2020-2022*. BPS (Badan Pusat Statistik). Retrieved March 31, 2023, from <https://www.bps.go.id/indicator/53/1498/1/luas-panen-produksi-dan-produktivitas-padi-menurut-provinsi.html>

- Dwilestari, G., Mulyawan, Martanto, & Ali, I. (2021). Analisis Clustering menggunakan K-Medoid pada Data Penduduk Miskin Indonesia. *JURSIMA: Jurnal Sistem Informasi Dan Manajemen*, 9(3), 282–290.
- FAOSTAT. (n.d.). *Crops and Livestock Products*. FAOSTAT (Food and Agriculture Organization of the United Nations). Retrieved April 12, 2023, from <https://www.fao.org/faostat/en/#data/QCL/visualize>
- Handoyo, R., Rumani, R., & Nasution, S. M. (2014). Perbandingan metode clustering menggunakan metode single linkage dan K-Means pada pengelompokan dokumen. *JSM STMIK Mikroskil*, 15(2), 73–82. <https://mikroskil.ac.id/ejurnal/index.php/jsm/article/view/161>
- Hidayati, R., Zubair, A., Pratama, A. H., & Indana, L. (2021). Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering. *Techno.Com*, 20(2), 186–197. <https://doi.org/10.33633/tc.v20i2.4556>
- Marisa, F., Zahma, A., Mui Bau, A., Noviansa, E., Neno, A. S., Lidya Maukar, A., Informatika, T., Malang, W., Borobudur, J., & 35, N. (2021). Digitasi Produktivitas Panen Padi Berbasis K-Means Clustering. *SMARTICS Journal*, 7(1), 21–26.
- Marlina, D., Putri, N. F., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 4(2), 64. <https://doi.org/10.24014/coreit.v4i2.4498>
- Nahdliyah, M. A., Widiharih, T., & Prahutama, A. (2019). Metode k-Medoids Clustering Dengan Validasi Silhouette Index dan C-Index. *JURNAL GAUSSIAN*, 8, 161–170.
- Nurlaela, S., Primajaya, A., & Padilah, T. N. (2020). Algoritma K-Medoids Untuk Clustering Penyakit Maag Di Kabupaten Karawang. *I N F O R M a T I K A*, 12(2), 56. <https://doi.org/10.36723/juri.v12i2.234>
- Pramesti, D. F., Tanzil Furqon, M., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(9), 723–732. <https://doi.org/10.1109/EUMC.2008.4751704>
- Rahmayani, M. T. I. (2018). Analisis Clustering Tingkat Keparahan Penyakit Pasien Menggunakan Algoritma K-Means. *Jurnal Inovasi Teknik Informatika*, 1(2), 40–44.
- Septiadi, D., & Joka, U. (2019). Analisis Respon dan Faktor-Faktor yang Mempengaruhi Permintaan Beras Indonesia. *Agrimor*, 4(3), 42–44. <https://doi.org/10.32938/ag.v4i3.843>
- Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. *RAINSTEK : Jurnal Terapan Sains & Teknologi*, 1(3), 53–68. <https://doi.org/10.21067/jtst.v1i3.3046>