

K-Medoids Cluster Analysis for Grouping Provinces in Indonesia Based on Agricultural Households ST2023

Riska, Zamahsary Martha*, Dony Permana, Fadhilah Fitri

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: zamahsarymartha@fmipa.unp.ac.id

Submitted : 10 Juli 2024
Revised : 03 Agustus 2024
Accepted : 13 Agustus 2024

ABSTRACT

Agriculture plays a crucial role in Indonesia's national development, providing essential resources such as raw materials, household income, and contributing significantly to Gross Domestic Product (GDP). According to the 2023 Agricultural Census (ST2023), there has been an increase in the number of Agricultural Household Enterprises (RTUP) across various agricultural subsectors. However, the welfare of agricultural entrepreneurs remains low, with 48.68% of poor household heads working in this sector. Therefore, an analysis is needed to understand the patterns and characteristics of RTUPs in each province. This study aims to cluster the provinces in Indonesia based on the number of Agricultural Household Enterprises (RTUP) using K-Medoids cluster analysis. K-Medoids, an extension of K-Means, was chosen for its ability to handle outliers by using medoids as cluster centers instead of means. The research utilized data from the 2023 Agricultural Census, covering 38 provinces and eight variables representing different agricultural subsectors. The optimal number of clusters was determined using the Elbow method, resulting in four distinct clusters. The findings revealed that Cluster 1 consists of 12 provinces with moderate RTUP numbers, Cluster 2 includes 23 provinces with low RTUP numbers, Cluster 3 comprises one province with high RTUP numbers, and Cluster 4 contains two provinces with very high RTUP numbers. The cluster validation using the Davies-Bouldin Index (DBI) yielded a value of 0.722, indicating that the clustering results are optimal.

Keywords: *Agricultural Households, Davies Bouldin Index, K-Medoids.*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Indonesia merupakan negara agraris yang mana sebagian besar penduduknya bekerja di sektor pertanian dan kebanyakan lahannya diarahkan untuk kegiatan pertanian (Faqih, 2021). Pertanian memiliki peran penting dalam pembangunan nasional Indonesia. Selain itu, sektor pertanian juga memiliki peran yang besar, diantaranya sebagai penyedia bahan industri, pendapatan rumah tangga pedesaan, bahan pokok, dan penyumbang Produk Domestik Bruto (PDB) (Prasetyaningrum & Susanti, 2023). Data statistik dasar yang lengkap, luas, dan menyeluruh dalam sektor pertanian dapat diperoleh melalui pelaksanaan Sensus Pertanian (Badan Pusat Statistik, 2023). Sensus Pertanian tahun 2023 (ST2023) mengumpulkan data dari tujuh subsektor pada sektor pertanian, yaitu subsektor tanaman pangan, hortikultura, perkebunan, peternakan, perikanan (budi daya ikan dan penangkapan ikan), kehutanan, dan jasa pertanian. Adapun unit usaha yang dicakup dalam ST2023 meliputi Usaha Pertanian Perorangan (UTP) yang termasuk ke dalam Rumah Tangga Usaha Pertanian (RTUP), Usaha Perusahaan Pertanian Berbadan Hukum (UPB), dan Usaha Pertanian Lainnya (UTL). RTUP adalah rumah tangga di mana satu atau lebih anggota keluarganya mengelola usaha pertanian dengan tujuan menjual sebagian atau seluruh hasilnya, baik itu usaha pertanian milik sendiri, usaha bagi hasil, atau milik orang lain dengan imbalan upah, termasuk usaha jasa di bidang pertanian.

Kondisi RTUP di Indonesia saat ini masih dalam kondisi yang masih tergolong kurang baik. Meskipun sektor pertanian mempunyai potensi yang besar dan terjadi penyerapan tenaga kerja yang tinggi, realitasnya menunjukkan bahwa sebagian besar pelaku petani masih termasuk dalam golongan masyarakat menengah ke bawah (Abdurrahman, 2016). Hal ini ditunjukkan dengan jumlah penduduk miskin usia 15 tahun keatas yang memiliki lapangan pekerjaan utama di sektor pertanian mencapai 5.216.951 (29,05%) dari 17.957.667 total penduduk miskin di Indonesia berdasarkan data Susenas Maret 2022. Sebanyak 2.682.886 (48,68%) dari 5.511.407 total kepala rumah tangga miskin di Indonesia bekerja di sektor pertanian (Badan Pusat Statistik, 2022). Keadaan ini tidak baik bagi Indonesia karena tingginya jumlah penduduk miskin yang bergantung pada sektor pertanian menunjukkan bahwa sektor ini belum mampu

mengangkat kesejahteraan masyarakat secara signifikan, padahal sektor ini memiliki potensi yang besar. Selain itu, tingginya jumlah penduduk miskin dapat berpengaruh terhadap pertumbuhan ekonomi suatu negara (Padang & Murtala, 2020). Oleh karena itu, penting dilakukan analisis kluster untuk membantu pemerintah dan pemangku kepentingan merancang program dan kebijakan yang lebih tepat sasaran dengan mengidentifikasi pola dan karakteristik yang berbeda di antara masing-masing provinsi di Indonesia berdasarkan jumlah RTUP.

Penelitian ini bertujuan untuk mengetahui pengelompokan provinsi di Indonesia menggunakan analisis kluster. Analisis kluster yang digunakan dalam penelitian ini adalah analisis kluster *K-Medoids*. Analisis kluster *K-Medoids* merupakan pengembangan dari analisis kluster *K-Means* yang dirancang untuk mengatasi kelemahan *K-Means* dalam menghadapi *outlier*. Perbedaan utama antara *K-Medoids* dan *K-Means* terletak pada penggunaan pusat kluster; *K-Medoids* menggunakan objek perwakilan (*medoid*) sebagai pusat kluster, sedangkan *K-Means* menggunakan nilai rata-rata (*mean*) sebagai pusat kluster. (Kamila, dkk., 2019).

Penelitian terdahulu oleh Meiriza (2023) mengenai perbandingan analisis kluster *K-Means* dan *K-Medoids* dalam mengelompokkan program BPJS Ketenagakerjaan, diperoleh bahwa *K-Medoids* memiliki performa yang lebih bagus dalam penelitian tersebut.

II. METODE PENELITIAN

Penelitian ini merupakan penelitian terapan dengan menggunakan data yang bersumber dari publikasi Sensus Pertanian tahun 2023 (ST2023) oleh Badan Pusat Statistik (BPS) yang mencakup 38 provinsi di Indonesia. Dalam penelitian ini terdapat 8 variabel yang merupakan jumlah RTUP pada masing-masing subsektor yang terdapat dalam sektor pertanian berdasarkan ST2023. Adapun 8 variabel tersebut yaitu jumlah Rumah Tangga Usaha Tanaman Pangan (X_1), jumlah Rumah Tangga Usaha Hortikultura (X_2), jumlah Rumah Tangga Usaha Perkebunan (X_3), jumlah Rumah Tangga Usaha Peternakan (X_4), jumlah Rumah Tangga Usaha Budi Daya Ikan (X_5), jumlah Rumah Tangga Usaha Penangkapan Ikan (X_6), jumlah Rumah Tangga Usaha Kehutanan (X_7), dan jumlah Rumah Tangga Usaha Jasa Pertanian (X_8).

Pada penelitian ini proses pengolahan data dengan bantuan *Software RStudio*. Adapun tahap-tahap analisis yang dilakukan adalah sebagai berikut:

1. Menyiapkan data yang akan diteliti yaitu jumlah RTUP subsektor Sensus Pertanian 2023 sebanyak 8 variabel dan 38 provinsi di Indonesia. Kemudian melakukan *cleaning data* untuk memastikan data yang digunakan akurat dan lengkap.
2. Melakukan pengujian asumsi dalam analisis kluster. Menurut Hair, dkk., (2019:211) terdapat uji asumsi yaitu sampel yang digunakan sudah mewakili populasi yang ada (sampel representatif) dan tidak terdapat hubungan antar variabel (multikolinearitas).
3. Penentuan jumlah k optimum menggunakan metode *Elbow*. Metode *Elbow* adalah metode yang dapat digunakan untuk penentuan berapa banyak kluster yang terbaik dengan membandingkan nilai *Sum of Square Error (SSE)* (Putri & Maori, 2021). Nilai SSE dapat diperoleh dengan rumus:

$$SSE = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \varphi_k)^2 \quad (1)$$

Dimana C_k adalah K kluster yang terbentuk, k adalah banyak kluster, x_i adalah nilai objek ke- i , φ_k adalah rata-rata K kluster pada nilai k ($k=1,2,3,\dots,K$).

4. Bangkitkan pusat kluster (*medoid*) secara acak sebanyak k optimum yang sudah didapatkan.
5. Kelompokkan objek ke *medoid* terdekat berdasarkan jarak terkecil objek ke *medoid*. Menurut Johnson (2014), jarak dapat dihitung dengan menggunakan jarak *Euclidean* dengan rumus:

$$d_{i,C} = \sqrt{\sum_{k=1}^p (x_{ik} - C_{jk})^2} \quad , i = 1,2,3, \dots, n \text{ dan } C = 1,2,3, \dots, m \quad (2)$$

Dimana $d_{i,C}$ adalah jarak antara non-medoid ke- i dengan medoid ke- C , x_{ik} adalah nilai objek non-medoid ke- i pada variabel ke- k , C_{jk} adalah nilai objek medoid ke- j pada variabel ke- k , p adalah banyaknya variabel yang diamati, n adalah banyaknya objek yang diamati, m adalah banyaknya medoid yang terbentuk.

6. Setelah jarak dari setiap objek ke *medoid* didapatkan dan dilakukan penempatan objek berdasarkan jarak terdekatnya, langkah selanjutnya adalah menghitung jumlah/total jarak (*Total Cost*). Berikut Persamaan untuk menghitung nilai *Total Cost*.

$$Total\ Cost = \sum_{c=1}^m \sum_{i=1}^n d_{i,c} \tag{3}$$

7. Sebagai *medoid* baru, lakukan kembali pemilihan *medoid* dari setiap kluster yang ada. Kemudian, tentukan kembali jarak objek ke *medoid* dan lakukan kembali penempatan objek. Lalu, hitung kembali nilai *Total Cost*.
8. Selanjutnya lakukan pengurangan/selisih dari *Total Cost* yang baru dengan *Total Cost* sebelumnya. Kemudian lanjutkan iterasi hingga diperoleh $S_{total\ cost} > 0$ atau tidak terjadi perubahan anggota *medoid*.
9. Interpretasi hasil pengelompokan dari analisis kluster *K-Medoids*
10. Selanjutnya melakukan validasi hasil analisis kluster menggunakan *Davies Bouldin Index* (DBI). Validasi analisis kluster adalah proses untuk menilai/mengevaluasi hasil analisis kluster yang dilakukan secara objektif dan dihitung secara kuantitatif. *Davies Bouldin Index* (DBI) merupakan metode yang dapat digunakan untuk mengevaluasi hasil analisis karena memiliki dua kelebihan utama yaitu, memaksimalkan jarak antar kluster dan meminimalkan jarak antar titik dalam satu kluster (Kapita, dkk., 2022). dengan rumus:

$$DBI = \frac{1}{k} \sum_{j=1}^k \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \tag{5}$$

Nilai R_{ij} diperoleh melalui persamaan berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Dengan SSW_i , SSW_j dan $SSB_{i,j}$ dapat diperoleh melalui persamaan berikut:

$$SSW_i = \frac{1}{m_i} \sum_{n=1}^{m_i} d(x_i, c_j)$$

$$SSW_j = \frac{1}{m_j} \sum_{n=1}^{m_j} d(x_i, c_j)$$

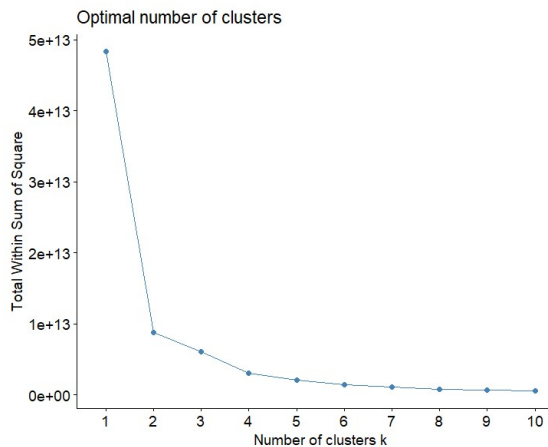
$$SSB_{i,j} = d(c_i, c_j)$$

Dimana R_{ij} adalah rasio antara kluster i dan kluster j , SSW_i , SSW_j adalah *sum of square within* kluster i dan j , $SSB_{i,j}$ adalah *sum of square between* kluster i dan j , $d(x_i, c_j)$ adalah jarak data ke- i ke *centroid* ke- j , $d(c_i, c_j)$ adalah jarak *centroid* ke- i ke *centroid* ke- j , dan k adalah jumlah kluster.

III. HASIL DAN PEMBAHASAN

A. Penentuan Jumlah Kluster

Penentuan jumlah kluster optimum dalam penelitian menggunakan metode *Elbow*. Jumlah kluster optimum ditunjukkan seperti Gambar 1.



Gambar 1. Plot Metode *Elbow*.

Berdasarkan Gambar 1, hasil dari metode *elbow* menunjukkan bahwa jumlah kluster optimum pada pengelompokan provinsi di Indonesia berdasarkan jumlah RTUP subsektor ST2023 menggunakan analisis kluster *K-Medoids* adalah sebanyak 4 kluster. Hal ini dikarenakan setelah penurunan yang signifikan pada $k=2$ dan 3, sedangkan setelah $k=4$ terlihat bahwa grafik telah landai dan tidak terjadi penurunan yang signifikan.

B. Analisis Kluster *K-Medoids*

Langkah selanjutnya setelah jumlah kluster optimum didapatkan adalah pilih empat titik data secara acak yang akan digunakan sebagai pusat kluster. Pusat kluster dalam analisis kluster *K-Medoids* disebut sebagai *medoid*. Kemudian hitung jarak antara setiap objek *non-medoid* dan *medoid* menggunakan ukuran jarak *Euclidean*. Proses analisis kluster *K-Medoids* dilakukan secara berulang hingga posisi *medoid* tidak mengalami perubahan. Hasil *Medoid* akhir dapat dilihat pada Tabel 2.

Tabel 2. *Medoid* Akhir Analisis Kluster *K-Medoids*

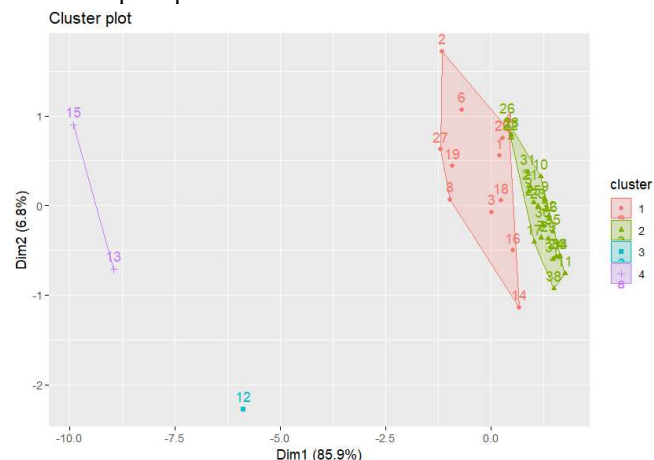
Klaster	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	449.091	177.658	367.071	179.295	15.934	30.936	8.760	6.128
2	44.331	49.267	108.199	57.188	12.843	23.179	3.548	1.568
3	2.210.536	1.281.005	556.753	1.061.094	216.599	35.745	589.182	55.064
4	2.614.586	1.941.929	1.037.462	2.250.765	181.323	71.411	1.035.652	60.835

Berdasarkan Tabel 2, diperoleh nilai *medoid* akhir dari proses analisis kluster *K-Medoids*. Hasil pengelompokan provinsi di Indonesia berdasarkan jumlah RTUP subsektor ST2023 menggunakan analisis kluster *K-Medoids* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengelompokan Analisis Kluster *K-Medoids*

Klaster	Banyak Anggota (Provinsi)
1	12
2	23
3	1
4	2

Dari Tabel 3, kluster 1 mempunyai anggota sebanyak 12 provinsi, kluster 2 mempunyai anggota sebanyak 23 provinsi, kluster 3 hanya memiliki anggota 1 provinsi, dan kluster 4 mempunyai anggota sebanyak 2 provinsi. Hasil pengelompokan tersebut divisualisasikan seperti pada Gambar 2.



Gambar 2. Plot Hasil Analisis Kluster *K-Medoids*.

Dari Gambar 2, dapat diketahui anggota dari setiap kluster yang terbentuk. Kluster 1 terdapat 12 provinsi yaitu Provinsi Aceh, Sumatera Barat, Sumatera Utara, Sumatera Selatan, Riau, Lampung, Banten, DI Yogyakarta, Nusa

Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, dan Sulawesi Selatan. Klaster 2 memiliki anggota sebanyak 23 provinsi yaitu Provinsi Bengkulu, Jambi, Kepulauan Riau, Kepulauan Bangka Belitung, DKI Kakarta, Bali, Kalimantan Tengah, Kalimantan Timur, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat, Papua Barat Daya, Papua Tengah, Papua Selatan, dan Papua Pegunungan. Sedangkan klaster 3 hanya memiliki 1 anggota yaitu Provinsi Jawa Barat. Sedangkan Klaster 4 memiliki anggota sebanyak 2 provinsi yaitu Provinsi Jawa Tengah dan Jawa Timur. Setelah hasil klaster diperoleh, selanjutnya dapat dilihat penciri/karakteristik dari masing-masing klaster dari nilai rata-ratanya, yang mana tersaji pada Tabel 4.

Tabel 4. Karakteristik/Penciri Klaster Berdasarkan Nilai Rata-rata

Klaster	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Kategori
1	465.629	230.105	448.567	317.183	24.745	26.962	62.628	11.506	Sedang
2	81.497	63.533	117.667	67.077	5.408	18.868	11.597	1.719	Rendah
3	2.210.536	1.281.005	556.753	1.061.094	216.599	35.745	589.182	55.064	Tinggi
4	2.939.140	1.996.071	1.115.722	2.818.038	151.267	81.295	930.243	65.052	Sangat Tinggi

Dari Tabel 4, bisa dilihat bahwa provinsi-provinsi pada klaster 1 merupakan provinsi dengan rata-rata jumlah RTUP sedang dengan subsektor yang unggul pada klaster ini adalah subsektor perkebunan dan tanaman pangan. Provinsi-provinsi pada klaster 2 merupakan provinsi dengan rata-rata jumlah RTUP rendah dengan subsektor yang unggul pada klaster ini adalah subsektor perkebunan. Provinsi-provinsi pada klaster 3 merupakan provinsi dengan rata-rata jumlah RTUP tinggi dengan subsektor yang unggul pada klaster ini adalah subsektor tanaman pangan. Sedangkan provinsi-provinsi pada klaster 4 merupakan provinsi dengan rata-rata jumlah RTUP sangat tinggi dengan subsektor yang unggul pada klaster ini adalah subsektor tanaman pangan dan peternakan.

C. Validasi Hasil Klaster

Uji Validasi berguna dalam menilai kualitas hasil klaster. Hasil validasi menggunakan nilai DBI untuk analisis klaster *K-Medoids* yaitu sebesar 0,722. Suatu analisis klaster bila memiliki nilai DBI yang kecil atau mendekati nol, dapat dikatakan klaster tersebut sudah optimal. Maka, dengan nilai DBI sebesar 0,722 tersebut dapat dikatakan bahwa klaster yang dihasilkan menggunakan analisis klaster *K-Medoids* sudah optimal.

IV. KESIMPULAN

Pengelompokan provinsi-provinsi di Indonesia berdasarkan jumlah Rumah Tangga Usaha Pertanian (RTUP) subsektor Sensus Pertanian tahun 2023 dengan analisis klaster *K-Medoids* menghasilkan 4 klaster optimum. Klaster 1 beranggotakan 12 provinsi yang merupakan provinsi dengan rata-rata jumlah RTUP sedang dan subsektor yang unggul pada klaster ini adalah subsektor perkebunan dan tanaman pangan. Klaster 2 beranggotakan 23 provinsi yang merupakan provinsi dengan rata-rata jumlah RTUP rendah dan subsektor yang unggul pada klaster ini adalah subsektor perkebunan. Klaster 3 hanya beranggotakan 1 provinsi yang merupakan provinsi dengan rata-rata jumlah RTUP tinggi dan subsektor yang unggul pada klaster ini adalah subsektor tanaman pangan. Sedangkan Klaster 4 beranggotakan 2 provinsi yang merupakan provinsi dengan rata-rata jumlah RTUP sangat tinggi dan subsektor yang unggul pada klaster ini adalah subsektor tanaman pangan dan peternakan.

DAFTAR PUSTAKA

- Abdurrahman, M. (2016). "Analisis Cluster Untuk Mengklasifikasikan Sektor Pertanian Berdasarkan Data Jumlah Rumah Tangga Usaha Pertanian di Indonesia", *Tugas Akhir*, 52 Hal., Universitas Islam Indonesia, Yogyakarta, Indonesia, 16 Maret 2016.
- Badan Pusat Statistik. (2022). *Statistik Penduduk dan Kemiskinan Sektor Pertanian Tahun 2023*. Jakarta: Badan Pusat Statistik.

- . (2023). *Hasil Pencacahan Lengkap Sensus Pertanian 2023*. Jakarta: Badan Pusat Statistik.
- . (2023). *Produk Domestik Bruto Indonesia Triwulanan*. Jakarta: Badan Pusat Statistik.
- Faqih, A. (2021). “Pengaruh Pembangunan Sektor Pertanian terhadap Kesempatan Kerja dan Distribusi Pendapatan”, *Jurnal Ilmu Pertanian*, Vol. 1, No. 2, hal. 30-35.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Black, W. C., & Anderson, R. E. (2019). *Multivariate data analysis (Eighth)*. Andover: Cengage Learning EMEA.
- Kamila, I., Khairunnisa, U., & Mustakim. (2019). “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau”, *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, Vol. 5, No. 1, hal 119–125.
- Kapita, S. N., Mubarak, A., Do Abdullah, S., & Fhadli, M. (2022). “Penerapan Algoritma Clustering Khonen-Som Dengan Validasi Davies Bouldin Index Pada Pengelompokan Potensi Uang Di Indonesia”, *IJIS-Indonesian Journal On Information System*, Vol. 7, No. 2, hal 134-143.
- Maori, N. A., & Evanita, E. (2023). “Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering”, *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, Vol. 14, No. 2, hal 277-288.
- Meiriza, A., Ali, E., Rahmiati, & Agustin. (2023). “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan”, *Indonesian Journal of Computer Science*, Vol. 12, No. 2, hal. 714-728.
- Padang, L., & Murtala, M. (2020). “Pengaruh Jumlah Penduduk Miskin Dan Tingkat Pengangguran Terbuka Terhadap Pertumbuhan Ekonomi Di Indonesia”, *Jurnal Ekonomika Indonesia*, Vol. 9, No. 1, hal. 9-16.
- Prasetyaningrum, E., & Susanti, P. (2023). “Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan”, *Jurnal Media Informatika Budidarma*, Vol. 7, No. 4, hal. 1775-1783.
- Putri, A. L. R., & Dwidayati, N. (2021). Analisa perbandingan k-means dan fuzzy c-means dalam pengelompokan daerah penyebaran COVID-19 Indonesia. *Unnes Journal of Mathematics*, 10(2), 50-55.