

Medoid-based Clustering pada Kecamatan di Kabupaten Lebak dan Pandeglang Provinsi Banten Berdasarkan Trilogi Ketahanan Pangan

Weksi Budiaji^(1*,2), Juwarin Pancawati⁽¹⁾

¹Fakultas Pertanian, Universitas Sultan Ageng Tirtayasa

Jl. Palka Km 3, Sindang Sari, Pabuaran, Serang, Banten 42163

²Pusat Unggulan Ipteks Inovasi Pangan Lokal, Universitas Sultan Ageng Tirtayasa

Jl. Palka Km 3, Sindang Sari, Pabuaran, Serang, Banten 42163

e-mail: budiaji@untirta.ac.id (*penulis korespondensi), juwarinpancawati@untirta.ac.id

ABSTRAK

Kabupaten Lebak dan Pandeglang Provinsi Banten memiliki prevalensi status gizi buruk pada balita yang tinggi dan berada pada ketahanan pangan yang paling rendah diantara seluruh kabupaten di Provinsi Banten. Analisis kluster untuk mengelompokkan kecamatan di kedua kabupaten sangat penting untuk mengeksplorasi karakteristik kecamatan di Kabupaten Lebak dan Pandeglang. Data indikator trilogi ketahanan pangan yaitu ketersediaan bahan pangan, akses ke bahan pangan, dan pemanfaatan bahan pangan diambil dari data sekunder Badan Pusat Statistik Kabupaten Lebak dan Pandeglang tahun 2019. Ketiga variabel tersebut digunakan untuk menghitung jarak pada algoritma *simple k-medoids clustering*. Perhitungan jarak divariasikan dengan konsep jarak Euclidean, squared Euclidean, dan Manhattan. Hasil kluster kemudian divalidasi dengan consensus clustering, dan validasi internal. Jumlah kluster yang paling sesuai dengan mempertimbangkan hasil validasi adalah empat kluster yaitu kluster ketersediaan dan akses (kluster 1), kluster akses (kluster 2), kluster rawan (kluster 3), dan kluster ketersediaan (kluster 4). Kluster 3 sebagai kluster rawan pangan perlu diperhatikan karena tersusun sebanyak 38% dari total kecamatan di Kabupaten Lebak dan Pandeglang.

Kata kunci : Ketahanan pangan, Kluster, Jarak, Medoid, Consensus, Validasi.

ABSTRACT

Lebak and Pandeglang Regions in Banten Province have a high stunting prevalence of children under 5 years old and have the lowest value of food security index among regions in Banten Province. Cluster analysis to group districts in Lebak and Pandeglang Regions is indispensable to characterize the district members in those two regions. The trilogy data of food security indicator, i.e., the availability, access, and utility of the food, were obtained from Bureau of Statistics of Lebak and Pandeglang Regions 2019 secondary data. The trilogy data were supplied for distance calculation in the simple k-medoids clustering. The distances were varied among Euclidean, squared Euclidean, and Manhattan distances. The clustering result was then validated via consensus clustering and internal validation. The suitable number of clusters w.r.t validation results was four clusters defined as the available and access cluster (cluster 1), the access cluster (cluster 2), the vulnerable cluster (cluster 3), and the available cluster (cluster 4). The cluster 3 as the vulnerable cluster should be focused on because it consists of 38% from overall districts in Lebak and Banten Regions.

Keywords : Food security, Clustering, Distance, Medoid, Consensus, Validation.

1. PENDAHULUAN

Analisis kluster yang bertujuan mengelompokkan objek sejenis pada satu kelompok yang sama merupakan analisis eksploratif. Analisis eksploratif sangat bermanfaat untuk mengidentifikasi struktur data. Salah satu analisis kluster yang populer dan juga salah satu dari top 10 algoritma pada *data mining* adalah k-means (Wu et

al., 2008). Berbeda dengan k-means yang menggunakan rata-rata sebagai pusat dari kluster, k-medoids menggunakan salah satu objek sebagai pusat kluster. Simple k-medoids (SKM) telah dikembangkan dengan mengadopsi k-means dan mempertimbangkan kelemahan yang dimiliki k-means yaitu adanya kluster kosong dan *local optima* (Budiaji & Leisch, 2019).

K-means dan SKM merupakan analisis kluster dengan metode algoritma partisi yang menggunakan konsep jarak sebagai ukuran kemiripan objek. Proses partisi memungkinkan setiap objek berpindah dari kluster satu ke kluster yang lain. Hasil analisis kluster dengan metode algoritma partisi sangat besar dipengaruhi oleh pemilihan konsep jarak yang digunakan (Leisch, 2006), sehingga penggunaan jarak yang beragam sangat penting, seperti Pamulang et al. (2021) dan Aditya et al. (2021) yang menggunakan variasi jarak pada algoritma partisi.

Artikel ini membahas penggunaan berbagai variasi jarak dengan algoritma SKM pada variabel trilogi ketahanan pangan. Trilogi ketahanan pangan yang terdiri atas ketersediaan bahan pangan, akses ke bahan pangan, dan pemanfaatan bahan pangan (Simatupang, 2007). Ketiga variabel ini digunakan untuk mengukur ketahanan pangan di Jawa Tengah (Pujiati et al. 2020) dan bahkan kabupaten-kabupaten di Indonesia (Mun'im, 2012; FSC & WFP, 2015).

Studi FSC & WFP (2015) menyebutkan bahwa Kabupaten Lebak dan Pandeglang memiliki ketahanan pangan yang paling rendah diantara seluruh kabupaten di Provinsi Banten, sehingga kedua kabupaten ini menjadi fokus perhatian di artikel ini. Kedua kabupaten memiliki persentase prevalensi status gizi (berat badan/ umur) buruk pada balita yang tinggi (4.8 % dan 6.98%) melebihi rata-rata Provinsi Banten (3.65%) (Balitbangkes, 2019). Belum ada analisis ketahanan pangan pada level kecamatan di kedua kabupaten juga menjadi alasan pentingnya penyusunan gerombol pada tingkat kecamatan.

2. METODE PENELITIAN

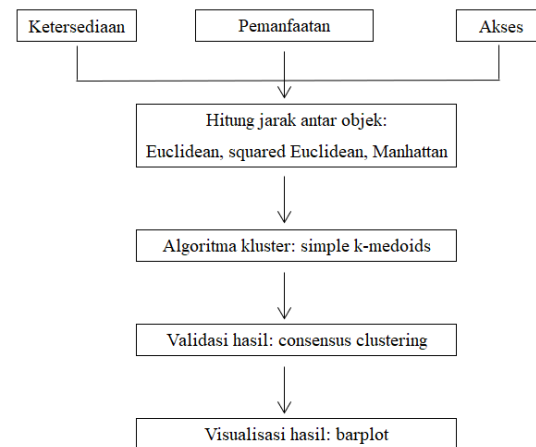
Medoid merupakan objek yang menjadi pusat kluster. Pengembangan simple k-medoid (SKM) yang mengadopsi k-means dengan mempertimbangkan kluster kosong dan local optima telah dilakukan (Budiaji & Leisch, 2019). Algoritma tersebut juga diimplementasikan pada paket R dengan nama paket *kmed* (Budiaji, 2021).

Algoritma SKM diterapkan pada data ketersediaan, akses, dan pemanfaatan bahan pangan. Data yang digunakan adalah data sekunder dari BPS Lebak (2019) dan BPS Pandeglang (2019). Data ketersediaan pangan menggunakan indikator rasio produksi sereal/ia tiap kecamatan dengan konsumsi normatif (FSC & WFP, 2015). Akses bahan pangan merupakan jumlah rumah tangga pengguna/pelanggan air bersih per 10000 penduduk, sedangkan pemanfaatan bahan pangan merasiokan jumlah fasilitas kesehatan tiap kecamatan per kapita.

Variabel trilogi ketahanan pangan yaitu ketersediaan, akses, dan pemanfaatan digunakan sebagai dasar perhitungan jarak tiap kecamatan. SKM sebagai metode algoritma partisi sangat dipengaruhi oleh jarak yang digunakan untuk mengukur kemiripan (Leisch, 2006). Jarak yang digunakan adalah *Euclidean*, *squared Euclidean*, dan *Manhattan*.

Hasil clustering kemudian divalidasi dengan salah satu metode kriteria relatif yaitu *consensus clustering* (Monti et al., 2003). Metode ini menggunakan dasar teknik *re-sampling* yang hasilnya dapat menggambarkan kestabilan anggota kluster. Validasi dengan metode kriteria relatif dapat divisualkan dalam *heatmap* dengan gambar blok diagonal mengindikasikan kluster yang dihasilkan stabil.

Hasil kluster yang paling sesuai dengan kondisi kecamatan di Kabupaten Lebak dan Pandeglang dideskripsikan dengan barplot. Barplot ditandai nilai tengah per kluster untuk menginterpretasikan hasil kluster (Dolnicar & Leisch, 2014). Gambar 1 menunjukkan proses analisis data trilogi ketahanan pangan sampai dengan visualisasi. Proses ini dibantu dengan software R (R Core Team, 2020), paket *kmed* (Budiaji, 2021), dan *rgl* (Murdoch & Adler, 2021).



Gambar 1. Langkah-langkah analisis data sampai hasil akhir

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi variabel trilogi ketahanan pangan

Ketersediaan bahan pangan diukur dengan produksi sereal/ia tiap kecamatan tahun 2018 (BPS Lebak, 2019; BPS Pandeglang, 2019) yang terdiri atas produksi padi dan jagung dan konsumsi normatif. Konsumsi normatif sereal/ia per hari per kapita tiap kecamatan diperoleh dari perkalian

jumlah pendudukan dengan 0.3 (FSC & WFP, 2015). Indeks ketersediaan bahan pangan kemudian merupakan rasio produksi sereal dengan konsumsi normatif. Berbeda dengan indeks ketersediaan FSC & WFP (2015), indeks ketersediaan yang dihitung ini jika nilai lebih dari 1 mengindikasikan kelebihan bahan pangan, sedangkan nilai kurang dari 1 menunjukkan kekurangan bahan pangan.

Tabel 1 menunjukkan bahwa kecamatan di Kabupaten Lebak mayoritas kekurangan bahan pangan. Kondisi ini bertolak belakang dengan kecamatan di Kabupaten Pandeglang. Rataan nilai indeks kelebihan pangan di Kabupaten Pandeglang walaupun lebih kecil yaitu sebesar 2.79 dibandingkan Kabupaten Lebak (3.52), rentang/jangkauan nilai kelebihan pangan di Kabupaten Lebak lebih tinggi.

Tabel 1. Ringkasan ketersediaan bahan pangan

| Kabupaten | Status Ketersediaan Bahan Pangan * | Jumlah Kecamatan | Nilai Indeks | | |
|------------|------------------------------------|------------------|--------------|--------|------|
| | | | Min | Rataan | Mak |
| Lebak | Kekurangan (indeks < 1) | 24 | 0 | 0.20 | 0.77 |
| | Kelebihan (indeks > 1) | 4 | 1.44 | 3.52 | 6.05 |
| Pandeglang | Kekurangan (indeks < 1) | 3 | 0.50 | 0.68 | 0.82 |
| | Kelebihan (indeks > 1) | 32 | 1.18 | 2.79 | 5.34 |
| Total | | 63 | | | |

* Kelebihan bahan pangan ditandai dengan nilai indeks yang lebih dari 1

Jumlah rumah tangga pengguna/ pelanggan air bersih per 10000 jumlah penduduk merupakan indikator dari akses bahan pangan. Semakin tinggi indeks rumah tangga pengguna/ pelanggan air bersih mengindikasikan semakin mudah akses ke bahan pangan. Tabel 2 menunjukkan bahwa kedua kabupaten memiliki karakter yang mirip dalam hal akses ke bahan pangan. Setengah dari keseluruhan kecamatan di kedua kabupaten bukan merupakan pengguna/ pelanggan air bersih.

Tabel 2. Ringkasan akses ke bahan pangan

| Kabupaten | Status Akses Bahan Pangan * | Jumlah Kecamatan | Nilai Indeks | | |
|------------|-----------------------------|------------------|--------------|--------|------|
| | | | Min | Rataan | Mak |
| Lebak | Kurang (indeks = 0) | 14 | 0 | 0 | 0 |
| | Baik (indeks > 0) | 14 | 0.01 | 2.07 | 7.23 |
| Pandeglang | Kurang (indeks = 0) | 17 | 0 | 0 | 0 |
| | Baik (indeks > 0) | 18 | 0.02 | 2.52 | 8.87 |
| Total | | 63 | | | |

* Akses ke bahan pangan yang baik diindikasikan dengan nilai indeks yang tidak 0

Pemanfaatan bahan pangan menggunakan jumlah fasilitas Kesehatan per kecamatan per kapita sebagai indikator. Semakin tinggi indeks pemanfaatan bahan pangan mengindikasikan semakin mudah dalam pemanfaatan bahan pangan. Tabel 3 menunjukkan bahwa kedua kabupaten memiliki karakter yang berbeda pada pemanfaatan bahan pangan. Kabupaten Pandeglang mayoritas dengan pemanfaatan bahan pangan yang baik,

sedangkan Kabupaten Lebak sebaliknya, lebih dari 50% kecamatan di Kabupaten Lebak memiliki indeks pemanfaatan bahan pangan yang kurang.

Tabel 3. Ringkasan pemanfaatan bahan pangan

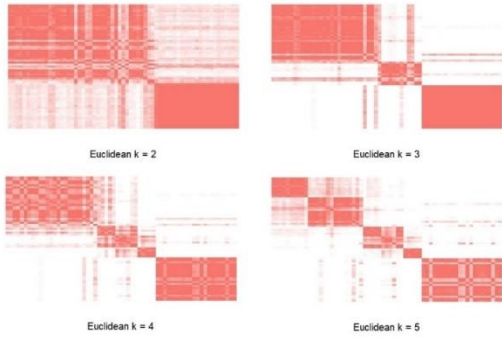
| Kabupaten | Status Pemanfaatan Bahan Pangan * | Jumlah Kecamatan | Nilai Indeks | | |
|------------|-----------------------------------|------------------|--------------|--------|------|
| | | | Min | Rataan | Mak |
| Lebak | Kurang (indeks < 1) | 18 | 0.52 | 0.74 | 0.97 |
| | Baik (indeks > 1) | 10 | 1.01 | 1.28 | 1.66 |
| Pandeglang | Kurang (indeks < 1) | 13 | 0.39 | 0.7 | 0.98 |
| | Baik (indeks > 1) | 22 | 1.03 | 1.38 | 2.84 |
| Total | | 63 | | | |

* Pemanfaatan bahan pangan yang baik diindikasikan dengan nilai indeks yang lebih dari 1

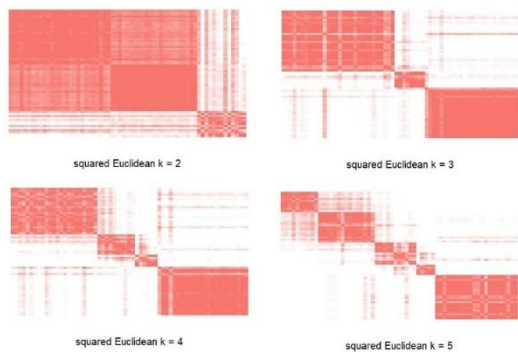
3.2 Analisis kluster

Analisis kluster diterapkan pada data kecamatan di Kabupaten Lebak dan Pandeglang. Data yang digunakan adalah data asli (bukan data terstandarisasi) karena semua variabel yang digunakan berupa indeks hasil dari proporsi, sehingga setiap variabel memiliki kontribusi yang sama. Penggunaan data tanpa distandarisasi juga memudahkan dalam interpretasi. Data asli ini kemudian dihitung jarak antar objeknya. Jarak yang digunakan adalah *Euclidean*, *squared Euclidean*, dan *Manhattan*.

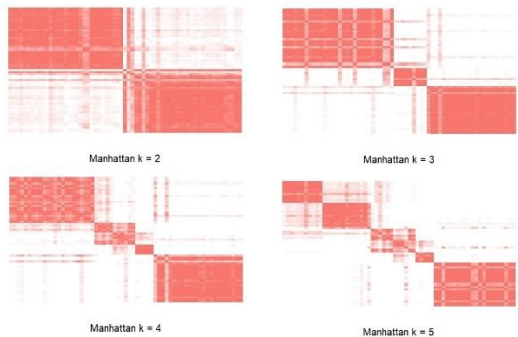
Hasil kluster dengan jarak *Euclidean*, *squared Euclidean*, dan *Manhattan* divalidasi dengan consensus clustering (Monti et al., 2003; Rupji et al., 2019) yang mendasarkan pada teknik bootstrapping. Heatmap dengan blok warna diagonal yang bersih dan penuh (tanpa titik-titik putih) menggambarkan hasil yang paling stabil, sedangkan Heatmap dengan warna yang tidak penuh (memiliki titik-titik putih yang banyak) menunjukkan separasi didalam cluster kurang baik. Gambar 2, 3, dan 4 menunjukkan *heatmap* hasil kluster dengan jumlah kluster 2, 3, 4, dan 5 pada jarak *Euclidean*, *squared Euclidean*, dan *Manhattan*. Gambar 2 memperlihatkan jumlah kluster 3 menghasilkan *heatmap* yang paling stabil karena memiliki blok warna diagonal paling penuh, sedangkan Gambar 3 dan 4 jumlah kluster 3 dan 4 menghasilkan *heatmap* yang paling baik.



Gambar 2. Heatmap hasil validasi kluster pada jumlah kluster (k) = 2, 3, 4, dan 5 dengan jarak *Euclidean*



Gambar 3. Heatmap hasil validasi kluster pada jumlah kluster (k) = 2, 3, 4, dan 5 dengan jarak *squared Euclidean*



Gambar 4. Heatmap hasil validasi kluster pada jumlah kluster (k) = 2, 3, 4, dan 5 dengan jarak *Manhattan*

Teknik validasi kriteria internal juga diaplikasikan selain teknik kriteria relatif dengan *consensus clustering* untuk memperkuat hasil validasi. Validasi internal digunakan untuk mengetahui kualitas hasil dari partisi. Ukuran yang digunakan pada validasi internal adalah kepadatan dan pemisahan (Arbelaitz et al., 2013). Kriteria internal yang digunakan adalah silhouette

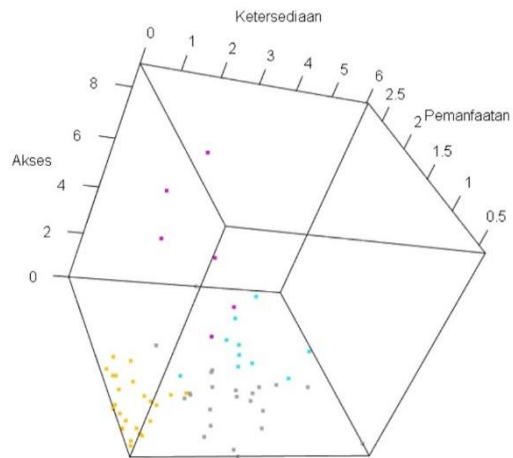
(Rousseeuw, 1987) dan *medoid-based shadow value* (Budiaji, 2019). Nilai silhouette berkisar [-1,1] sedangkan nilai *medoid-based shadow value* berada pada [0,1]. Semakin tinggi nilai silhouette/ *medoid-based shadow value*, menunjukkan cluster terpartisi dengan baik (nilai 1 artinya terpisah sempurna). Tabel 4 menunjukkan bahwa jumlah kluster 2, 3, atau 4 dengan jarak *squared Euclidean* menghasilkan validasi internal yang paling baik karena nilainya paling tinggi. Berdasarkan pertimbangan kriteria relatif, internal, dan kemudahan analisis lanjutan terkait ketahanan pangan di tingkat kecamatan, jumlah kluster 4 dengan jarak *squared Euclidean* dipilih sebagai jumlah kluster yang paling cocok/ sesuai.

Tabel 4. Nilai validasi internal dengan silhouette dan *medoid-based shadow value*

| Kriteria | Jarak | Jumlah Kluster | | | |
|---------------------------|-------------------|----------------|---------------|---------------|-------|
| | | 2 | 3 | 4 | 5 |
| Silhouette | Euclidean | 0.369 | 0.470 | 0.480 | 0.438 |
| | squared Euclidean | 0.659* | 0.628* | 0.645* | 0.618 |
| | Manhattan | 0.386 | 0.461 | 0.464 | 0.427 |
| Medoid-based shadow value | Euclidean | 0.547 | 0.620 | 0.615 | 0.615 |
| | squared Euclidean | 0.790* | 0.788* | 0.797* | 0.771 |
| | Manhattan | 0.580 | 0.630 | 0.607 | 0.597 |

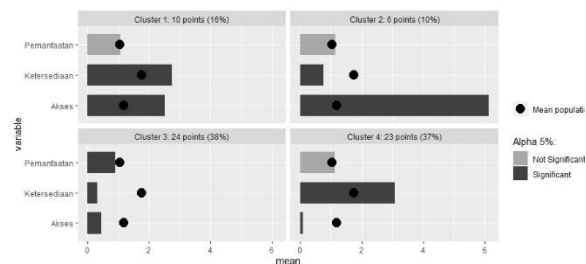
* Jumlah kluster yang cocok berdasarkan nilai yang paling tinggi

Plot 3 dimensi mudah dibuat karena variabel yang digunakan dalam analisis kluster ada 3. Gambar 5 menunjukkan plot objek dalam tiga dimensi dengan jumlah kluster sebanyak 4 berdasarkan hasil analisis kluster dengan jarak *squared Euclidean*. Untuk mempermudah interpretasi hasil analisis, barplot (Gambar 6) dibuat.



Gambar 5. Plot 3 dimensi hasil analisis kluster dengan $k = 4$ dengan jarak squared *Euclidean*

Gambar 6 menunjukkan bahwa analisis kluster menghasilkan kluster 3 (38%) dengan karakteristik bernilai rendah pada ketiga variabel (kluster rawan). Kluster 1 (16%) adalah kluster ketersediaan dan akses karena anggota kluster memiliki nilai yang tinggi pada kedua variabel tersebut. Kluster 2 (10%) merupakan kluster dengan akses ke bahan pangan yang sangat baik, sedangkan kluster 4 (37%) sangat baik pada ketersediaan bahan pangan. Analisis lanjutan perlu pada kluster 1 karena kluster tersebut merupakan kluster dengan kategori yang rawan pada ketiga aspek (ketersediaan, akses, dan pemanfaatan bahan pangan).



Gambar 6. Barplot 4 kluster berdasarkan variabel pemanfaatan, ketersediaan dan akses bahan pangan

Anggota kluster 3 berjumlah 24 kecamatan (38%) dari 63 kecamatan di Kabupaten Lebak dan Pandeglang. Hanya ada satu kecamatan di Kabupaten Pandeglang yang masuk dalam kluster 3 yaitu Kecamatan Cadasari, sedangkan 23 kecamatan lainnya berasal dari Kabupaten Lebak. Dengan jumlah kecamatan sebanyak 28 kecamatan di Kabupaten Lebak, berarti sebanyak 82% kecamatan di Kabupaten Lebak masih berkategori rawan.

4. KESIMPULAN

Kecamatan di Kabupaten Pandeglang memiliki ketersediaan dan pemanfaatan bahan pangan yang lebih baik daripada kecamatan di Kabupaten Lebak. Pada variabel akses ke bahan pangan kecamatan di kedua kabupaten berkarakter mirip yaitu setengahnya masih kurang akses. Analisis kluster yang telah divalidasi menghasilkan 4 kluster pada jarak squared *Euclidean* sebagai jumlah kluster yang paling sesuai. Kluster ketiga perlu diperhatikan karena merupakan kluster yang rawan pangan dengan anggota 38% dari total populasi kecamatan di Kabupaten Lebak dan Pandeglang.

UCAPAN TERIMA KASIH

Ucapan terima kasih kami sampaikan kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Sultan Ageng Tirtayasa (Untirta) yang telah mendanai penelitian ini melalui skema Hibah Penelitian Madya Fakultas Pertanian Untirta.

DAFTAR PUSTAKA

Aditya A, Sari BN, dan Padilah TN. 2021. Perbandingan pengukuran jarak Euclidean dan Gower pada kluster k-medoids. *Jurnal Teknologi dan Sistem Komputer* Vol. 9, No. 1, Hal. 1-7

Arbelaitz O, Gurrutxaga I, Muguerza J, Perez J, and Perona I. 2013. An extensive comparative study of cluster validity indices. *Pattern Recognition* 46, 243–256

Balitbangkes. 2019. *Laporan Provinsi Banten Riskesdas 2018/* Badan Penelitian dan Pengembangan Kesehatan. Jakarta : Lembaga Penerbit Badan Penelitian dan Pengembangan Kesehatan.

BPS Lebak. 2019. *Kabupaten Lebak dalam Angka*. Lebak: Badan Pusat Statistik Kabupaten Lebak.

BPS Pandeglang. 2019. *Kabupaten Pandeglang dalam Angka*. Pandeglang: Badan Pusat Statistik Kabupaten Lebak.

Budiaji W. 2019. Medoid-based shadow value validation and visualization. *International Journal of Advances in Intelligent Informatics* Vol. 5 (2), pp 76-88

Budiaji, Weksi, and Friedrich Leisch. 2019. Simple K-Medoids Partitioning Algorithm for Mixed Variable Data. *Algorithms* 12, no. 9: 177. <https://doi.org/10.3390/a12090177>

Budiaji, W. 2021. kmed: distance-based k-medoids. R package version 0.4.0. <https://cran.r-project.org/package=kmed>

Dolnicar, S. and F. Leisch. 2014. Using graphical statistics to better understand market segmentation solutions. *International Journal of Market Research* 56, 207–230.

FSC and WFP. 2015. *Food Security and Vulnerability Atlas of Indonesia*. Jakarta, Indonesia: FSC and WFP.

Leisch, F. 2006. A toolbox for K-centroids cluster analysis. *Comput. Stat. Data Anal.* 51, 526–544. [<http://dx.doi.org/10.1016/j.csda.2005.10.006>]

Monti S, Tamayo P, Mesirov J, and Golub T. 2003. Consensus Clustering: A Resampling-Based

- Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Mach. Learn.*, vol. 52, pp. 91–118, doi: 10.1023/A:1023949509487.
- Mun'im A. 2012. Analisis Pengaruh Faktor Ketersediaan, Akses, dan Penyerapan, Pangan terhadap Ketahanan Pangan di Kabupaten Surplus Pangan: Pendekatan Partial Least Square Path Modelling. *Jurnal Agro Ekonomi* Vol. 30 No. 1, Hal: 41-58
- Murdoch, D & Adler D. 2021. rgl: 3D Visualization Using OpenGL. R package version 0.108.3. <https://CRAN.R-project.org/package=rgl>
- Pamulang MNP, Aini MN, dan Enri U. 2021. Komparasi Distance Measure Pada K-Medoids Clustering untuk Pengelompokan Penyakit Ispa. *Edumatic: Jurnal Pendidikan Informatika* Vol. 5 No. 1, Hal. 99-107
- Pujiati S, Pertiwi A, Silfia CCS, Ibrahim DM, Hafida SHN. 2020. Analisis Ketersediaan, Keterjangkauan, dan Pemanfaatan Pangan dalam Mendukung Tercapainya Ketahanan Pangan Masyarakat di Provinsi Jawa Tengah. *Jurnal Sosial Ekonomi Pertanian* Vol. 16 No. 2, Hal: 123-133
- R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2015, available at: <https://www.r-project.org/>
- Rousseeuw, PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, vol. 20, pp. 53–65
- Rupji M, Dwivedi B, Kowalski J. 2019. NOJAH: NOt Just Another Heatmap for genome-wide cluster analysis. *PLoS ONE* 14(3): e0204542
- Simatupang P. 2007. Analisis Kritis Terhadap Paradigma dan Kerangka Dasar Kebijakan Ketahanan Pangan Nasional. *Forum Penelitian Agro Ekonomi* Vol. 25 No.1 Hal: 1–18
- Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S.; Steinbach M; Hand D; and Steinberg D. 2008. Top 10 algorithms in data mining. *Knowl. Inf. Syst*, 14, 1–37. [<http://dx.doi.org/10.1007/s10115-007-0114-2>]

LAMPIRAN

Data yang digunakan dalam artikel ini dapat diunduh pada DOI: 10.5281/zenodo.6525997. Kode software R juga tersedia untuk reproduksi ulang pada DOI: 10.5281/zenodo.6526018 (<https://zenodo.org/record/6526018/files/jstat.R?download=1>).