

Peningkatan Distribusi Bantuan Sosial di Pangkalpinang dengan Pengelompokan Berbantuan Algoritma *K-Means*

DWITRA GUSTI ALRISCKI¹, ACHMAD FAUZAN²

^{1, 2}) Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Jl. Kaliurang KM. 14,5, Sleman, 55584, Indonesia

e-mail: achmadfauzan@uii.ac.id

ABSTRAK

Program bantuan sosial (Bansos) merupakan kebijakan penting yang diimplementasikan untuk mengatasi kemiskinan dan meningkatkan kesejahteraan masyarakat. Penelitian ini bertujuan untuk meningkatkan efektivitas program bantuan sosial di Kota Pangkalpinang melalui penerapan metode *K-Means Clustering* dalam pengelompokan kelurahan dengan tingkat kesejahteraan rendah. Data yang digunakan adalah Data Terpadu Kesejahteraan Sosial (DTKS) yang mencakup penerima bantuan dari berbagai program seperti PBI, BST, dan lainnya. Metode *K-Means Clustering* diterapkan untuk mengelompokkan kelurahan berdasarkan indikator kemiskinan, yang kemudian dianalisis untuk profilisasi *cluster*. Analisis *Principal Component Analysis* (PCA) dilakukan untuk mengatasi multikolinearitas antar variabel. *Silhouette coefficient* digunakan untuk menentukan jumlah *cluster* yang ideal untuk memastikan validitas pengelompokan. Hasil penelitian menunjukkan tiga *cluster* utama dengan nilai *silhouette coefficient* sebesar 0.458. *Cluster* pertama memiliki penerima bantuan terbanyak, sedangkan *cluster* ketiga memiliki penerima bantuan terendah. Penggunaan metode ini diharapkan dapat meningkatkan efektivitas dan efisiensi distribusi bantuan sosial dengan memastikan bantuan tepat sasaran sesuai dengan tingkat kemiskinan masing-masing kelurahan. Oleh karena itu, penelitian ini diharapkan dapat membantu kebijakan bantuan sosial Kota Pangkalpinang.

Kata Kunci: Bantuan Sosial, Kemiskinan, *K-Means Clustering*, Koefisien *Silhouette*, Pangkalpinang.

ABSTRACT

The social assistance program (Bansos) is an important policy implemented to address poverty and improve community welfare. This research aims to improve the effectiveness of social assistance programs in Pangkalpinang City through the application of the *K-Means clustering* method in grouping sub-districts with low welfare levels. The data used is the Integrated Social Welfare Data (SWD), which includes recipients of assistance from various programs such as PBI, BST, and others. The *K-Means clustering* method is applied to group villages based on poverty indicators, which are then analyzed for cluster profiling. *Principal Component Analysis* (PCA) is conducted to address multicollinearity among variables. The *silhouette coefficient* is used to determine the ideal number of clusters to ensure the validity of the clustering. The research results show three main clusters with a *silhouette coefficient* value of 0.458. The first cluster has the most aid recipients, while the third cluster has the fewest aid recipients. The use of this method is expected to improve the effectiveness and efficiency of social assistance distribution by ensuring that aid is targeted according to the poverty levels of each sub-district. Therefore, this research is expected to assist the social assistance policies of Pangkalpinang City.

Keywords: Social Assistance, Poverty, *K-Means Clustering*, *Silhouette Coefficient*, Pangkalpinang

1. PENDAHULUAN

Program bantuan sosial (Bansos) merupakan kebijakan penting yang diimplementasikan oleh pemerintah untuk mengatasi kemiskinan dan meningkatkan kesejahteraan masyarakat. Di berbagai kota di Indonesia termasuk Pangkalpinang, berbagai program seperti Program Bantuan Iuran (PBI), Jaminan Kesehatan Nasional dan Bantuan Sosial Tunai (BST) telah dijalankan untuk mendukung masyarakat yang kurang mampu (Kementerian Sosial, 2021). Namun, pelaksanaan

program-program ini sering menghadapi kendala seperti ketidaktepatan sasaran dan ketimpangan distribusi bantuan, yang dapat disebabkan oleh kurangnya data yang akurat dan metode distribusi yang tidak efektif (Sumarto & Suryahadi, 2004). Beberapa penelitian menunjukkan bahwa salah satu masalah utama adalah data penerima yang tidak valid, yang mengakibatkan banyak bantuan tidak tepat sasaran (Sumarto et al., 2005). Selain itu, birokrasi yang kompleks dan kurangnya koordinasi antar lembaga juga sering menghambat distribusi bantuan secara merata (Skoufias et al., 2011).

Metode yang lebih sistematis dan didukung oleh data, seperti metode *cluster*, diperlukan untuk menyelesaikan masalah ini. Analisis kluster (*cluster analysis*) adalah teknik statistik yang digunakan untuk mengelompokkan objek berdasarkan kesamaan karakteristik. Teknik ini digunakan untuk mengidentifikasi pola dalam data yang kompleks, terutama dalam bidang-bidang seperti pemasaran, biologi, dan ilmu sosial. *K-Means* adalah salah satu algoritma yang paling umum digunakan dalam analisis *cluster*, yang membagi data ke dalam banyak *cluster* berdasarkan jarak terdekat dari *centroid* atau pusat *cluster*. Proses optimasi *K-Means* bertujuan untuk meminimalkan variansi dalam *cluster* yang terbentuk (Heraldi et al., 2019).

Penelitian terdahulu menunjukkan berbagai aplikasi metode *K-Means* dalam pengelompokan data kemiskinan dan penerima bantuan sosial. Misalnya, penelitian oleh (Putra & Anggrawan, 2021) yang mengelompokkan penerima bantuan sosial di Lombok menggunakan metode *K-Means* menunjukkan bahwa mayoritas penerima bantuan sudah tepat sasaran. Penelitian oleh Manurung et al. (2020) membandingkan algoritma *K-Means* dan *K-Medoids* untuk pengelompokan data masyarakat miskin di Kecamatan Hatonduhan dan menemukan bahwa kedua metode tersebut efektif dalam mengidentifikasi kelompok masyarakat yang membutuhkan bantuan. Selain itu, penelitian oleh Ramadani et al. (2019) menggunakan pendekatan *K-Means* untuk mengelompokkan masyarakat miskin di Binjai. Metode ini memperhitungkan jarak *Euclidean* dan *Manhattan*. Studi ini memberikan wawasan tambahan tentang cara terbaik untuk mengatasi kemiskinan.

Metode ini dapat membagi kelurahan berdasarkan tingkat kemiskinan, sehingga bantuan dapat didistribusikan secara lebih proporsional dan tepat sasaran (Filki, 2022). Penggunaan *K-Means Clustering* juga dapat meningkatkan transparansi dan akuntabilitas dalam distribusi bantuan, membantu pemerintah memantau dan melaporkan penggunaan dana dengan lebih efektif (Sumarto & Suryahadi, 2004). Data dalam situasi ini dikelompokkan ke dalam berbagai kelompok (*cluster*) berdasarkan karakteristik yang mirip satu sama lain. Setiap kelompok terdiri dari data yang memiliki karakteristik yang mirip satu sama lain, sementara kelompok lain dikelompokkan ke dalam kelompok yang berbeda. Dengan demikian, penerapan metode ini diharapkan dapat membantu dalam mengidentifikasi kelurahan-kelurahan yang memiliki tingkat kemiskinan yang tinggi dan membutuhkan bantuan lebih banyak (Nana et al., 2022). Selain itu, pengelompokan yang lebih tepat dan berbasis data ini dapat mendukung perencanaan program bantuan sosial yang lebih strategis, sehingga manfaatnya dapat dirasakan secara merata oleh seluruh masyarakat yang membutuhkan,

Berdasarkan data dari Kota Pangkalpinang meskipun Provinsi Kepulauan Bangka Belitung secara umum berada di peringkat ke-5 dengan tingkat kemiskinan terendah di Indonesia, Kota Pangkalpinang justru menunjukkan garis kemiskinan yang lebih tinggi dibandingkan wilayah lainnya dalam provinsi tersebut. Garis kemiskinan di Kota Pangkalpinang pada tahun 2021 tercatat sebesar Rp. 808.839, yang lebih tinggi dibandingkan rata-rata wilayah lain di provinsi tersebut (Marlina et al., 2022). Faktor utama yang menyebabkan tingginya garis kemiskinan di wilayah ini berasal dari komponen makanan yang memberikan kontribusi yang besar terhadap kenaikan garis kemiskinan. Dengan demikian, tujuan dari penelitian ini adalah untuk menerapkan pendekatan kelompokan *K-Means* dalam pembagian kelurahan yang kurang mampu. Selain itu, tujuan dari penelitian ini adalah untuk memberikan rekomendasi tentang cara terbaik untuk mengoptimalkan program bantuan sosial di Kota Pangkalpinang.

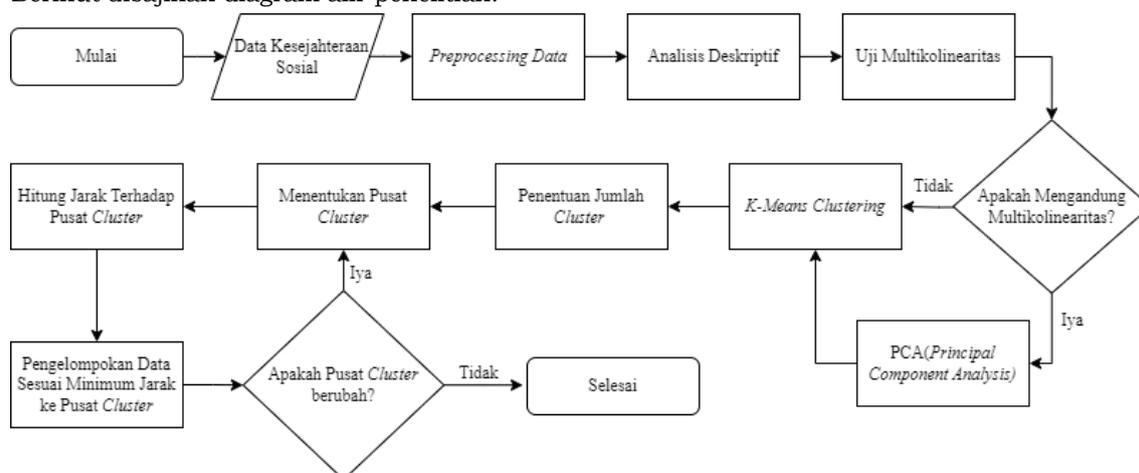
2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari *website* resmi Kementerian Sosial yaitu data Data Terpadu Kesejahteraan Sosial (DTKS) Kota Pangkalpinang yang berisikan penerima bansos di kota pangkalpinang 2020 hingga 2023. Pada penelitian ini, data yang digunakan adalah kelurahan di kota pangkalpinang berjumlah 42 kelurahan. Variabel penelitian yang digunakan berjumlah 6 dimana merupakan jenis – jenis bantuan dengan definisi yang disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Definisi Operasional
Program Bantuan Iuran (PBI)	Jumlah individu yang menerima bantuan iuran untuk Program Jaminan Kesehatan Nasional (JKN) yang diberikan oleh pemerintah untuk masyarakat kurang mampu.
Yatim Piatu	Jumlah anak-anak yang tidak memiliki kedua orang tua (yatim piatu) yang menerima bantuan sosial.
Bantuan Sosial Tunai (BST)	Jumlah individu yang menerima bantuan sosial dalam bentuk tunai dari pemerintah.
Sembilan Bahan Pokok (Sembako)	Jumlah individu yang menerima bantuan berupa paket sembilan bahan pokok (beras, gula, minyak goreng, daging, telur, susu, jagung, minyak tanah, dan garam) dari pemerintah.
Bantuan Langsung Tunai Bahan Bakar Minyak (BLT BBM)	Jumlah individu yang menerima bantuan tunai yang diberikan oleh pemerintah kepada masyarakat miskin dan rentan yang terdampak kenaikan harga bahan bakar minyak.
Bantuan Langsung Tunai Minyak Goreng (BLT MIGOR)	Jumlah individu yang menerima bantuan tunai yang diberikan oleh pemerintah kepada masyarakat miskin dan rentan untuk membantu mereka dalam membeli minyak goreng yang mengalami kenaikan harga

Berikut disajikan diagram alir penelitian:



Gambar 1. Diagram Alir Penelitian

Dalam statistik deskriptif, penyajian data menjadi aspek penting untuk memudahkan pemahaman dan interpretasi. Data dapat disajikan dalam bentuk tabel yang memuat informasi berupa baris dan kolom, serta grafik yang memberikan gambaran visual mengenai distribusi data (Rahma et al., 2022). Penyajian data dalam bentuk grafik meliputi tabel, diagram batang, diagram garis, diagram lingkaran, dan diagram titik (Wahyuni, 2020). Multikolinearitas terjadi ketika dua atau lebih variabel independen dalam model memiliki hubungan linear yang kuat. Ini menyebabkan estimasi parameter yang tidak stabil dan hasil yang bias. *Variance Inflation Factor* (VIF) digunakan untuk mendeteksi multikolinearitas, dengan nilai VIF lebih dari 10 menunjukkan multikolinearitas serius (Kim, 2019). VIF digunakan untuk menentukan seberapa besar variasi koefisien regresi yang diestimasi meningkat dalam kasus di mana variabel independen saling berkorelasi. VIF dihitung menggunakan (Shrestha, 2020).

$$VIF = \frac{1}{1 - R^2} \tag{1}$$

R^2 adalah nilai koefisien determinasi yang diperoleh dari regresi antara satu variabel independen terhadap variabel independen lainnya. PCA digunakan untuk mereduksi dimensi dengan menghitung *initial eigenvalue*. Langkah-langkahnya melibatkan standarisasi data, perhitungan

matriks kovariansi dari data yang telah distandarisasi, perhitungan *eigenvalue* dan *eigenvector*, pengurutan *eigenvector* berdasarkan *eigenvalue*, pemilihan komponen utama berdasarkan *eigenvector* teratas, dan penggunaan matriks transformasi untuk menghasilkan *principal components* (Fadilah & Rosha, 2022). Menurut Hastie et al. (2009), salah satu kriteria untuk memilih jumlah komponen dalam PCA adalah berdasarkan jumlah variansi yang dijelaskan oleh komponen-komponen tersebut. Hasilnya adalah representasi data yang lebih sederhana dan dimensi yang lebih rendah, di mana varian terbesar ditempatkan pada komponen utama pertama, diikuti oleh varian terbesar kedua, dan seterusnya.

K-Means adalah algoritma *clustering* yang populer dalam pembelajaran mesin yang digunakan untuk mengelompokkan data ke dalam sejumlah kelompok (k) yang telah ditentukan sebelumnya. Setiap kelompok atau *cluster* ditandai oleh pusat (*centroid*) yang merupakan rata-rata dari semua titik data dalam *cluster* tersebut. Proses iteratif ini melibatkan pengoptimalan posisi *centroid* dan mengelompokkan titik data berdasarkan jarak terdekat ke *centroid*, sehingga menghasilkan *cluster* yang kohesif secara internal dan berbeda satu sama lain (Hasibuan et al., 2023).

Algoritma *K-Means* menggunakan rumus yang menghitung jarak *Euclidean* antara setiap titik data dengan *centroid*-nya. Langkah-langkah utama dalam algoritma *K-Means* adalah sebagai berikut (Tendean & Purba, 2020). Proses *clustering* diawali dengan pemilihan jumlah *cluster* k yang diinginkan. Langkah berikutnya adalah menentukan secara acak k titik awal yang berfungsi sebagai *centroid*. Setiap titik data kemudian dipindahkan ke *centroid* terdekat berdasarkan jarak *Euclidean*, yang memungkinkan pembentukan awal *cluster*. Setelah itu, posisi *centroid* dihitung ulang sebagai rata-rata dari semua titik data yang berada dalam *cluster* tersebut. Langkah memindahkan titik data dan menghitung ulang *centroid* diulangi hingga posisi *centroid* tidak lagi berubah secara signifikan atau mencapai batas maksimum iterasi yang telah ditentukan. Rumus jarak *Euclidean* disajikan pada Persamaan (2).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

x_i = Nilai variabel ke- i dari observasi x , y_i = Nilai variabel ke- i dari observasi y dan n = Jumlah dimensi atau variabel yang diamati dalam ruang tersebut.

Dalam analisis *cluster* non-hirarkis, ada banyak metode yang dapat digunakan untuk mengoptimalkan jumlah *cluster*. Salah satu metode yang dapat digunakan untuk menentukan jumlah *cluster* yang ideal adalah metode *silhouette*, yang menilai seberapa mirip sebuah objek dengan *clusternya* sendiri dibandingkan dengan *cluster* lain. Nilai *silhouette* berkisar dari -1 hingga 1; nilai yang lebih tinggi menunjukkan bahwa objek berada dalam *clusternya* dengan benar. Metode ini membantu menentukan jumlah *cluster* optimal dengan memaksimalkan nilai rata-rata *Silhouette* (Handoyo et al., 2014).

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

$a(i)$ adalah rata-rata jarak antara elemen i dan semua elemen lain dalam *cluster* yang sama dan $b(i)$ adalah jarak rata-rata elemen i dengan elemen-elemen di *cluster* terdekat lainnya (*cluster* yang bukan milik elemen i).

3. HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

Pada penelitian ini, akan dilakukan analisis deskriptif sebagai gambaran umum Data Terpadu Kesejahteraan Sosial (DTKS) Kota Pangkalpinang. Penelitian ini terdiri dari 42 jumlah data. Gambaran umum dari data yang digunakan disajikan pada Tabel 2.

Tabel 2. Statistika Deskriptif

Jenis Bansos	Minimal	Rerata	Maksimal
PBI	148	734.9	1773
Yatim Piatu	0	8.024	20
BST	32	108	208
Sembako	47	181.9	390
BLT BBM	58	180.9	354
BLT Migor	56	171.2	328

Mengenai berbagai jenis bantuan sosial (bansos), terlihat bahwa nilai minimal berada pada bantuan yatim piatu dengan nilai 0, yang menunjukkan bahwa ada kelurahan di mana tidak ada penerima bantuan. Sedangkan nilai maksimal tertinggi berada pada bantuan PBI (Penerima Bantuan Iuran) dengan nilai 1773, yang menunjukkan jumlah penerima bantuan tertinggi dalam data yang tersedia. Nilai rata-rata tertinggi juga terdapat pada PBI dengan 734.9, menunjukkan bahwa sebagian besar penerima berada di sekitar nilai ini. Rentang penerima bantuan PBI adalah yang terluas, dari 148 hingga 1773, menunjukkan variasi yang signifikan dalam jumlah penerima. Secara keseluruhan, data ini menunjukkan bahwa bantuan yatim piatu memiliki jumlah penerima terendah dan rentang yang paling kecil dibandingkan dengan jenis bantuan lainnya, sementara PBI memiliki jumlah penerima dan rentang tertinggi. BST dan bantuan sembako juga menunjukkan variasi yang cukup besar dalam jumlah penerima, namun tidak sebesar PBI. BLT BBM menunjukkan jumlah penerima yang cukup konsisten dengan nilai rata-rata yang mendekati bantuan sembako.

3.2 Uji Multikolinearitas

Tabel 2. Nilai uji Multikolinearitas

PBI	Yatim Piatu	BST	Sembako	BLT BBM	BLT Migor
7.278	1.806	4.836	45.781	253.750	291.079

Terdapat beberapa pasangan variabel menunjukkan nilai VIF di atas 10, yang menunjukkan adanya korelasi yang kuat antara variabel-variabel tersebut. Hal ini selaras dengan penjelasan Kim dalam *Multicollinearity and misleading statistical results* serta Noora Shrestha dalam *Detecting Multicollinearity in Regression Analysis*. Terutama pada variabel BLT BBM dan BLT MIGOR yang memiliki nilai VIF sangat tinggi. Dengan demikian, dapat disimpulkan bahwa terjadi multikolinearitas, oleh karena itu dilakukan *Principal Component Analysis (PCA)* untuk mengurangi multikolinearitas.

3.3 Principal Component Analysis (PCA)

Karena terindikasi memiliki multikolinearitas yang kuat maka akan dilakukan analisis PCA dengan hasil sebagai berikut:

Tabel 3. Hasil PCA

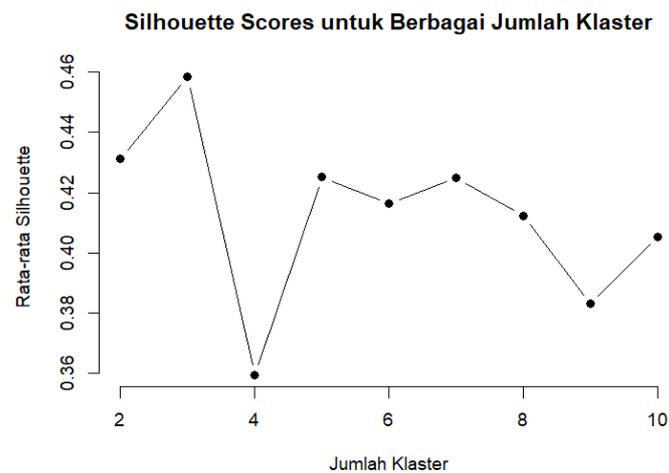
Komponen	Nilai Eigen	Percentage of Variance	Cummulative of Variance
Komponen 1	4.724	78.743	78.743
Komponen 2	0.805	13.426	92.170
Komponen 3	0.322	5.376	97.546
Komponen 4	0.130	2.179	99.726
Komponen 5	0.014	0.241	99.968
Komponen 6	0.001	0.031	100

Hasil analisis PCA menunjukkan bahwa dua komponen utama sudah cukup untuk menjelaskan sebagian besar variansi dalam data. Komponen pertama menjelaskan 78.74% variansi, sedangkan komponen kedua menambahkan 13.43%, sehingga totalnya menjadi 92.17%. Penggunaan dua komponen ini diusulkan karena mereka menangkap informasi yang signifikan sambil mempertahankan kesederhanaan dan efisiensi analisis. Hal ini selaras dengan prinsip

yang dijelaskan oleh Hastie et al. (2009) dalam *The Elements of Statistical Learning*, dimana pemilihan jumlah komponen didasarkan pada variansi yang dijelaskan.

3.4 K-Means Clustering

Menggunakan data yang telah dilakukan analisis PCA akan dilakukan optimalisasi jumlah *cluster*, karena memastikan bahwa data dikelompokkan secara representatif dan bermakna sangat penting untuk algoritma *K-means*. Pemilihan jumlah *cluster* yang tepat memungkinkan untuk menangkap struktur alami dalam data, meningkatkan akurasi analisis, dan menghasilkan wawasan yang lebih tajam. Oleh karena itu digunakan metode *Silhouette* untuk menentukan jumlah *cluster* yang tepat.



Gambar 2. *Silhouette Plot*

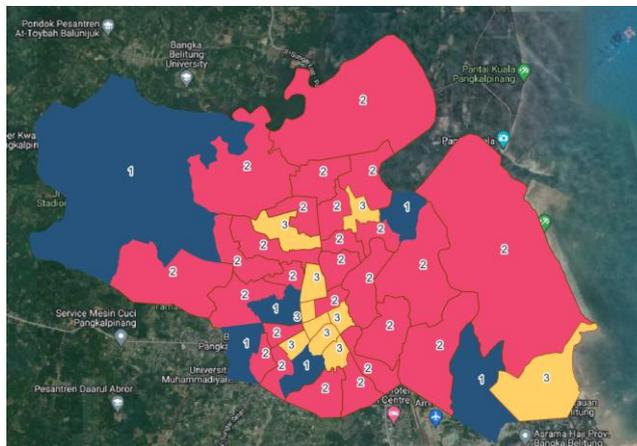
Pada Gambar 2 dapat dilihat bahwa garis putus-putus yang terbentuk pada metode *silhouette* terletak pada *cluster* 3 dengan nilai koefisien sebesar 0.458, sehingga dapat ditarik kesimpulan bahwa 3 merupakan jumlah *cluster* yang optimal. Setelah jumlah optimum *cluster* diketahui dapat dilanjutkan ke langkah yang berikutnya yaitu *K-Means Clustering*, untuk penelitian ini akan digunakan *software* R. Sehingga didapatkan hasil:



Gambar 3. Plot Pembagian *Cluster*

Pada Gambar 3 didapatkan bahwa terdapat 6 kelurahan yang dikelompokkan kedalam *cluster* 1 dan memiliki kelurahan paling sedikit, pada *cluster* 3 terdapat 10 kelurahan yang berada didalam kelompok, untuk *cluster* 2 tergabung atas 26 kelurahan serta *cluster* dengan jumlah kelurahan

yang paling banyak. Untuk penjelasan mengenai kelurahan mana saja yang terletak pada tiap *cluster* disajikan pada tabel dan visualisasi berikut.



Gambar 4. Visualisasi Cluster Pada Peta

Tabel 5. Pembagian Cluster

Cluster	Kelurahan
1	Air Itam, Tua Tunu, Ketapang, Keramat, Parit Lalang, Kejaksanaan.
2	Bukit Besar, Kacang Pedang, Gabek Satu, Sriwijaya, Temberan, Pintu Air, Bacang, Rejosari, Semabung Baru, Selindung, Gajah Mada, Bukit Merapin, Lontong Pancur, Gabek Dua, Bukit Sari, Jerambah Gantung, Pasir Putih, Selindung Baru, Ampui, Air Kepala Tujuh, Air Mawar, Opas Indah, Semabung Lama, Gedung Nasional, Taman Bunga, Asam.
3	Sinar Bulan, Air Salemba, Bukit Intan, Pasar Padi, Pasir Garam, Bintang, Masjid Jamik, Melintang, Batin Tikal, Rawa Bangun.

Setelah proses pengelompokan data (*clustering*) selesai dilakukan, langkah selanjutnya adalah melakukan *profiling* atau profilisasi pada setiap *cluster* yang terbentuk. Profilisasi *cluster* ini bertujuan untuk mengidentifikasi dan memahami karakteristik unik dari masing-masing *cluster*, seperti pola, kecenderungan, atau atribut-atribut yang dominan dalam setiap kelompok data.

Tabel 6. Profilisasi Cluster

Cluster	PBI	Yatim Piatu	BST	Sembako	BLT BBM	BLT Migor
Cluster 1	1367.5	12.5	139.5	340.5	323.66	305.5
Cluster 2	727.5	9.03	100.53	178.84	180.15	170.57
Cluster 3	374.7	2.7	64.3	94.7	97.2	92.2

Tabel 6 menyajikan data dari tiga *cluster* yang diukur berdasarkan enam variabel, yaitu PBI, Yatim Piatu, BST, Sembako, BLT BBM, dan BLT MIGOR, dengan penekanan pada perbedaan nilai di setiap *cluster*. *Cluster 1* menonjol dengan nilai tertinggi pada semua variabel: PBI sebesar 1367.5, Yatim Piatu 12.5, BST 139.5, Sembako 340.5, BLT BBM 323.667, dan BLT MIGOR 305.5. Sebaliknya, *Cluster 3* menunjukkan nilai terendah untuk semua variabel dengan PBI 374.7, Yatim Piatu 2.7, BST 64.3, Sembako 94.7, BLT BBM 97.2, dan BLT MIGOR 92.2. Analisis ini menunjukkan bahwa *Cluster 1* memiliki karakteristik paling unggul dalam semua variabel, sementara *Cluster 3* memiliki karakteristik paling rendah. *Cluster 2* memperlihatkan perbedaan yang mencolok di antara kedua *cluster* tersebut.

Berdasarkan analisis data jumlah penerima bansos yang terbagi dalam tiga *cluster*, direkomendasikan agar pemerintah memfokuskan alokasi dan pemantauan bantuan dengan lebih cermat. *Cluster 1*, yang menunjukkan jumlah penerima tertinggi, memerlukan pengawasan ekstra untuk memastikan distribusi bantuan yang efektif dan tepat sasaran. *Cluster 3*, dengan jumlah penerima terendah, membutuhkan evaluasi untuk mengidentifikasi kendala yang menghalangi penerima manfaat potensial. *Cluster 2*, yang memiliki variasi jumlah penerima, dapat dioptimalkan dengan program yang lebih spesifik dan penyesuaian kebijakan berdasarkan kebutuhan lokal.

4. SIMPULAN DAN SARAN

Data bantuan sosial di Kota Pangkalpinang menunjukkan variasi yang mencolok dalam jumlah penerima di setiap kelurahan. Jenis bantuan yang paling banyak diterima adalah Program Bantuan Iuran (PBI), dengan jumlah penerima berkisar antara 148 hingga 1773 orang. Rata-rata penerima PBI adalah yang tertinggi dibandingkan jenis bantuan lainnya, menunjukkan tingginya kebutuhan individu untuk bantuan pembayaran iuran Jaminan Kesehatan Nasional (JKN). Bantuan yatim piatu memiliki jumlah penerima paling sedikit, dengan variasi dari 0 hingga 20 penerima, menunjukkan bahwa beberapa kelurahan mungkin tidak memiliki anak yatim piatu yang menerima bantuan. Penelitian ini berhasil mengelompokkan kelurahan di Kota Pangkalpinang berdasarkan indikator kemiskinan menggunakan metode *K-Means Clustering*. Hasil analisis menunjukkan bahwa pengelompokan kelurahan ke dalam tiga *cluster* adalah yang paling optimal, dengan nilai koefisien *silhouette* sebesar 0.458, yang menunjukkan tingkat kesesuaian yang cukup baik antara data dan *cluster* yang terbentuk. *Cluster 1* terdiri dari kelurahan-kelurahan dengan jumlah penerima bantuan sosial tertinggi, *Cluster 2* terdiri dari kelurahan dengan jumlah penerima bantuan sedang, dan *Cluster 3* terdiri dari kelurahan dengan jumlah penerima bantuan terendah. Hasil menunjukkan bahwa *Cluster 1* memerlukan pengawasan ekstra untuk memastikan distribusi bantuan yang efektif, sementara *Cluster 3* memerlukan evaluasi untuk mengidentifikasi kendala yang menghalangi penerima manfaat potensial.

DAFTAR PUSTAKA

- Fadilah, D., & Rosha, M. (2022). Analisis Faktor Pada Faktor-Faktor Yang Mempengaruhi Stres Guru SD Selama Sistem Pembelajaran Daring Era Covid-19 (Studi Kasus di SD Kecamatan Padang Timur). *Journal of Mathematics UNP*, 7(4), 75. <https://doi.org/10.24036/unpjomath.v7i4.13990>
- Filki, Y. (2022). Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa. *Jurnal Informatika Ekonomi Bisnis*, 4, 166–171. <https://doi.org/10.37034/infeb.v4i4.166>
- Handoyo, R., Mangkudjaja, R., & Nasution, S. M. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K - Means pada Pengelompokan Dokumen. *Jurnal SIFO Mikroskil*, 15(2), 73–82.
- Hasibuan, A., Kembuan, D. R. E., Manoppo, C. T. M., & Hermanto, M. T. (2023). Optimization of k-means algorithm in grouping data using the statistical gap method. *Journal of Intelligent Decision Support System (IDSS)*, 6(3), 112–120. <https://doi.org/https://doi.org/10.35335/idss.v6i3.149>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Vol. Second Edition*. Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Heraldi, H. Y., Aprilia, N. C., & Pratiwi, H. (2019). Analisis Cluster Intensitas Kebencanaan di Indonesia Menggunakan Metode K-Means. *Indonesian Journal of Applied Statistics*, 2(2), 137. <https://doi.org/10.13057/ijas.v2i2.34911>
- Kementerian Sosial. (2021, September 27). *Mensos: Data Penerima Bantuan Iuran Jaminan Kesehatan Sudah Terintegrasi dengan DTKS*. <https://Kemensos.Go.Id/Mensos-Data-Penerima-Bantuan-Iuran-Jaminan-Kesehatan-Sudah-Terintegrasi-Dengan-Dtks>.
- Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 72(6), 558–569. <https://doi.org/10.4097/kja.19087>
- Manurung, J., Ramadhan, P. S., & Suryanata, M. (2020). Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pengelompokan Data Masyarakat Miskin Pada Kantor Camat Hatonduhan STMIK Triguna Dharma. *Jurnal CyberTech*, 3(9), 1522–1531. <https://doi.org/https://doi.org/10.53513/jct.v3i9.3161>
- Marlina, M., Nurwasya, N., Valeriani, D., & Wulandari, A. (2022). Korelasi Umur, Penghasilan Tetap, Pendidikan Terakhir dan Rata-Rata Pengeluaran Pada Rumah Tangga Miskin di Kota Pangkalpinang. *Equity: Jurnal Ekonomi*, 10(2), 37–48. <https://doi.org/10.33019/equity.v10i2.112>

- Nana, S., Herman, Nining, R., Rini, A., & Wijaya, Y. A. (2022). Grouping of Sewing Tool Assistance Recipients Using K-Means Clustering Analysis. *International Journal of Social Science*, 2(2), 1513–1522. <https://doi.org/10.53625/ijss.v2i2.3085>
- Putra, L. G. R., & Anggrawan, A. (2021). Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(1), 205–214. <https://doi.org/10.30812/matrik.v21i1.1554>
- Rahma, R. W., Malau, N. A., Sova, M., Ngii, E., Sugiri, T., Ardhiarisca, O., Astuti, Y., & Saidah, H. (2022). *Statistik Deskriptif*. Widina. www.penerbitwidina.com
- Ramadani, S., I, A., & Pardede, A. M. H. (2019). Metode K-Means Untuk Pengelompokan Masyarakat Miskin Dengan Menggunakan Jarak Kedekatan Manhattan City Dan Euclidean (Studi Kasus Kota Binjai). *Journal Information System Development*, 4(2). <https://binjaikota.bps.go.id/>
- Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39–42. <https://doi.org/10.12691/ajams-8-2-1>
- Skoufias, E., Tiwari, S., & Zaman, H. (2011). Can We Rely on Cash Transfers to Protect Dietary Diversity During Food Crises? Estimates from Indonesia. In <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/278061468269665919/can-we-rely-on-cash-transfers-to-protect-dietary-diversity-during-food-crises-estimates-from-indonesia>. <https://www.researchgate.net/publication/228295599>
- Sumarto, S., & Suryahadi, A. (2004). *The Role of Agricultural Growth in Poverty Reduction in Indonesia* (Issue 60724). <https://ideas.repec.org/p/pramprapa/60724.html>
- Sumarto, S., Suryahadi, A., & Widyanti, W. (2005). Assessing the Impact of Indonesian Social Safety Net Programmes on Household Welfare and Poverty Dynamics. *The European Journal of Development Research*, 17(1), 155–177. <https://doi.org/10.1080/09578810500066746>
- Tendean, T., & Purba, W. (2020). Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means. *Jurnal Sains Dan Teknologi*, 1(2), 5–11. <https://doi.org/https://doi.org/10.34013/saintek.v1i2.31>
- Wahyuni, M. (2020). *Statistik Deskriptif untuk Penelitian Olah Data Manual dan SPSS Versi 25*. Bintang Pustaka Madani.