

IMPLEMENTASI ALGORITMA K-MEANS UNTUK PENGELOMPOKAN KELURAHAN DKI JAKARTA BERDASARKAN BANTUAN PANGAN BULOG DI PT YASA ARTHA TRIMANUNGGAL

Raehan Ramadhan^{1*}, Arief Wibowo²

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*} 2011500036@student.budiluhur.ac.id, ² arief.wibowo@budiluhur.ac.id

(* : corresponding author)

Abstrak- Penelitian ini mengembangkan sistem *Data Mining* klusterisasi dengan Algoritma KMeans untuk mengelompokkan kelurahan di Provinsi DKI Jakarta berdasarkan Distribusi Bantuan Pangan dari Bulog yang dilakukan oleh PT Yasa Artha Trimanunggal. Provinsi DKI Jakarta memiliki banyak kelurahan dengan karakteristik yang beragam, sehingga sering kali terjadi ketidakakuratan dalam penyaluran bantuan pangan. Tujuan penelitian ini adalah untuk meningkatkan efisiensi dan efektivitas penyaluran bantuan pangan dengan mengidentifikasi wilayah yang benar-benar membutuhkan bantuan secara lebih akurat. Metode K-Means digunakan untuk mengelompokkan wilayah berdasarkan total penerima tiap kelurahan, total penduduk tiap kelurahan, dan rata-rata usia penerima tiap kelurahan. Dari hasil penelitian ini, jumlah kluster yang optimal ditemukan untuk mengelompokkan data kelurahan di DKI Jakarta adalah 3 kluster yang terdiri dari wilayah dengan tingkat kebutuhan bantuan pangan yang tinggi, kedua dengan tingkat kebutuhan sedang, dan ketiga dengan tingkat kebutuhan rendah. Evaluasi hasil klusterisasi menggunakan metode *Davies-Bouldin Index* menunjukkan nilai 0,5442 yang menandakan bahwa pengelompokan yang dilakukan cukup baik dalam memisahkan wilayah-wilayah berdasarkan kebutuhan bantuan pangan. Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam meningkatkan pemahaman terhadap pola penyaluran bantuan pangan serta memberikan informasi yang berharga bagi PT Yasa Artha Trimanunggal dalam merencanakan dan mengambil keputusan terkait Distribusi Bantuan Pangan di DKI.

Kata Kunci: Algoritma K-Means, Distribusi Bantuan Pangan, klusterisasi .

IMPLEMENTATION OF K-MEANS ALGORITHM FOR CLUSTERING JAKARTA URBAN VILLAGES BASED ON BULOG FOOD ASSISTANCE AT PT YASA ARTHA TRIMANUNGGAL

Abstract- This research develops a clustering *Data Mining* system with the KMeans Algorithm to group urban villages in DKI Jakarta Province based on the Distribution of Food Aid from Bulog carried out by PT Yasa Artha Trimanunggal. DKI Jakarta Province has many urban villages with diverse characteristics, so inaccuracies often occur in the distribution of food aid. The purpose of this research is to improve the efficiency and effectiveness of food aid distribution by identifying areas that really need assistance more accurately. The K-Means method is used to cluster areas based on the total recipients of each kelurahan, the total population of each kelurahan, and the average age of recipients of each kelurahan. From the results of this study, the optimal number of clusters found to group data on urban villages in DKI Jakarta is 3 clusters consisting of areas with high levels of need for food assistance, second with moderate levels of need, and third with low levels of need. Evaluation of the clustering results using the *Davies-Bouldin Index* method shows a value of 0.5442 which indicates that the clustering is good enough to separate areas based on food assistance needs. This research is expected to make a significant contribution in improving the understanding of food aid distribution patterns as well as providing valuable information for PT Yasa Artha Trimanunggal in managing food aid distribution.

Keywords: K-Means Algorithm, Food aid distribution, Clustering

1. PENDAHULUAN

Rendahnya pendapatan berdampak langsung pada keterbatasan akses terhadap kebutuhan pangan bagi sebagian masyarakat Indonesia. Pangan merupakan kebutuhan vital yang tidak dapat diabaikan, terutama di negara dengan populasi besar seperti Indonesia. Penyaluran bantuan pangan, khususnya beras, memainkan peran penting dalam memastikan kesejahteraan masyarakat. PT Yasa Artha Trimanunggal, yang berlokasi di Jakarta Barat, adalah salah satu perusahaan yang terlibat dalam penyaluran bantuan pangan dari Badan Urusan Logistik (Bulog) di Provinsi DKI Jakarta. Namun, data penerima bantuan yang dikelola oleh PT Yasa Artha Trimanunggal seringkali tidak akurat, terutama dalam mengidentifikasi penerima yang benar-benar membutuhkan. Ketidakakuratan ini dapat menyebabkan bantuan tidak tepat sasaran dan menurunkan efektivitas program. Oleh

karena itu, sangat penting bagi pemerintah dan lembaga terkait untuk memastikan bahwa bantuan tersebut disalurkan secara efisien dan efektif, terutama di wilayah dengan tingkat kebutuhan yang paling mendesak di Provinsi DKI Jakarta. Di sinilah teknologi *Data Mining* menjadi penting untuk membantu mengolah dan mengelompokkan data secara cepat dan akurat, sehingga memudahkan penentuan wilayah atau daerah yang paling membutuhkan bantuan pangan. *Data Mining* tidak hanya bertujuan untuk mengidentifikasi pola historis dalam penyaluran bantuan, tetapi juga memberikan kemampuan untuk memprediksi kebutuhan masa depan dengan lebih akurat dan efektif. Dalam hal ini, PT Yasa Artha Trimanunggal berkomitmen untuk menggunakan metode K-Means *Clustering* dalam mengelompokkan penerima bantuan pangan berdasarkan wilayah yang paling membutuhkan. Pendekatan ini diharapkan dapat meningkatkan akurasi peramalan dan memungkinkan penyesuaian kebijakan penyaluran yang lebih optimal dan efisien. Dalam menghadapi dinamika sosial dan ekonomi yang terus berubah, perusahaan seperti PT Yasa Artha Trimanunggal harus mampu beradaptasi dengan cepat. Penerapan metode K-Means *Clustering* menjadi penting untuk mengidentifikasi perubahan pola penyaluran bantuan yang terjadi. Ini memberikan peluang bagi PT Yasa Artha Trimanunggal untuk mengambil tindakan preventif dan menyesuaikan kebijakan dengan lebih cermat dan efisien guna menghadapi tantangan yang muncul. Metode K-Means dapat digunakan untuk mengelompokkan wilayah berdasarkan berbagai indikator ekonomi dan sosial, seperti tingkat pendapatan, tingkat pengangguran, dan indikator kesehatan. Ini memungkinkan PT Yasa Artha Trimanunggal untuk mengidentifikasi daerah yang benar-benar membutuhkan bantuan pangan dengan lebih mendesak. Dengan menggunakan *Clustering*, wilayah dengan karakteristik serupa akan dikelompokkan bersama, memudahkan identifikasi daerah prioritas untuk penyaluran bantuan. Penelitian sebelumnya yang dilakukan oleh dengan judul "Implementasi *Data Mining* dengan Metode *Clustering* Algoritma K-Means untuk Pengelompokan Data Tilang di Instansi Pemerintah"[1] serta penelitian oleh Frisma Handayanna (2023) dengan judul "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin di Provinsi Banten" diharapkan dapat memberikan peningkatan yang lebih mendalam dan kompleks pada penelitian ini, sehingga mampu meningkatkan pemahaman serta relevansinya dalam konteks yang lebih luas. Penelitian ini mengembangkan sistem *Data Mining* Klasterisasi dengan Algoritma K-Means untuk mengelompokkan kelurahan di provinsi DKI Jakarta berdasarkan penyaluran Bantuan Pangan BULOG di PT Yasa Artha Trimanunggal. Sebagai perbandingan, penelitian sebelumnya yang dilakukan oleh (Radana Sembiring et al., 2021) Implementasi *Data Mining* Dalam Mengelompokkan Jumlah Penduduk Miskin berdasarkan Provinsi Menggunakan Algoritma K-Means. Perbedaan utama antara penelitian ini dan penelitian sebelumnya terletak pada fokus analisisnya. Penelitian sebelumnya memusatkan perhatian pada jumlah penduduk miskin berdasarkan provinsi, penelitian ini lebih menitikberatkan pada analisis pola penyaluran bantuan pangan dari BULOG. Dengan perspektif ini, diharapkan penelitian ini dapat memberikan wawasan baru terkait distribusi bantuan pangan di tingkat kelurahan di DKI Jakarta, serta menyediakan informasi yang berharga bagi PT Yasa Artha Trimanunggal dalam merencanakan dan mengambil keputusan terkait penyaluran bantuan pangan di seluruh Indonesia. Setelah menerapkan algoritma K-Means untuk mengelompokkan kelurahan di Provinsi DKI Jakarta berdasarkan distribusi bantuan pangan dari Bulog, dilakukan evaluasi terhadap hasil *clustering* yang diperoleh. Evaluasi ini menggunakan persentase keberhasilan yang dihitung dengan membandingkan hasil *clustering* dengan data acuan yang diharapkan. Persentase keberhasilan yang diperoleh sebesar 93.07% ini menunjukkan bahwa algoritma K-Means berhasil mengelompokkan kelurahan dengan tingkat akurasi yang tinggi. Hasil ini mencerminkan efektivitas algoritma dalam mengelompokkan data sesuai dengan tujuan penelitian. Dengan menguraikan latar belakang ini secara terperinci, diharapkan penelitian ini akan menjadi tonggak penting dalam meningkatkan efektivitas dan efisiensi penyaluran bantuan pangan beras dari Bulog melalui PT Yasa Artha Trimanunggal. Selain itu, penelitian ini juga diharapkan dapat memberikan pemahaman yang lebih mendalam terhadap pola penyaluran bantuan, sehingga dapat memberikan kontribusi yang lebih akurat dan efektif bagi semua pihak yang terlibat dalam manajemen distribusi bantuan pangan di tingkat provinsi.

2. METODE PENELITIAN

2.1 Klasterisasi

Klasterisasi atau *clustering* adalah metode untuk menemukan dan mengelompokkan data berdasarkan kemiripan karakteristik. Metode ini mengidentifikasi kelompok atau *cluster* dari data yang mirip, sehingga pola dalam data lebih mudah dikenali[2]. *Clustering* adalah teknik dalam *Data Mining* yang bersifat tanpa arahan (*unsupervised*), diterapkan tanpa pelatihan atau target output tertentu. Ada dua jenis utama metode *Clustering*: *hierarchical* dan *non-hierarchical*. *Hierarchical Clustering* mengorganisir data dalam struktur bertingkat untuk mengelompokkan pada berbagai tingkat kedalaman, sementara *non-hierarchical Clustering*, seperti K-Means, mengelompokkan data ke dalam sejumlah *cluster* yang ditentukan berdasarkan kemiripan karakteristik. Kedua pendekatan ini penting dalam analisis data, memungkinkan penemuan wawasan berharga dari dataset yang

kompleks [3]. *Clustering* membagi sekelompok objek data ke dalam kelompok yang disebut *cluster*, di mana setiap *cluster* terdiri dari objek yang memiliki kesamaan karakteristik, berbeda signifikan dengan objek dalam *cluster* lain. Proses ini menggunakan algoritma yang telah diatur sebelumnya, memungkinkan analisis lebih efisien dan terstruktur, mengidentifikasi pola dalam data yang mungkin tidak terlihat, dan memberikan wawasan baru serta pemahaman yang lebih dalam tentang struktur data yang kompleks [4].

2.2 Algoritma K-Means Clustering

Algoritma K-Means adalah salah satu algoritma *Clustering* yang sangat populer dan sering digunakan dalam analisis data untuk mengelompokkan data berdasarkan karakteristik yang serupa. Algoritma ini efektif dalam mengidentifikasi kelompok data yang memiliki kemiripan tertentu, yang kemudian disebut sebagai *Cluster* atau klaster. *Cluster* tersebut dirancang untuk mengklasifikasikan objek berdasarkan kesamaan karakteristik, sehingga memungkinkan analisis yang lebih mendalam dan terstruktur[5]. K-Means merupakan algoritma *Clustering* yang khusus bekerja dengan atribut numerik. Hal ini disebabkan oleh metode dasarnya yang menggunakan perhitungan jarak untuk membagi data ke dalam sejumlah *Cluster* tertentu. Setiap objek dalam data dievaluasi berdasarkan kedekatannya dengan pusat *Cluster* yang telah ditentukan, dan kemudian dikelompokkan ke dalam *Cluster* yang paling dekat. Proses ini terus berulang hingga pembagian *Cluster* mencapai stabilitas[6]. Kemampuan K-Means untuk mengelompokkan data dengan cara ini membuatnya sangat berguna dalam berbagai aplikasi, mulai dari segmentasi pasar, analisis perilaku pelanggan, hingga pengelompokkan dalam bidang biologi dan Kesehatan. Dengan demikian, K-Means menjadi alat yang sangat berharga dalam penemuan pola tersembunyi dan penggalian wawasan dari data numerik yang besar dan kompleks [7].

2.3 Data Penelitian

Penelitian ini menjelaskan pendekatan metodologi terstruktur untuk menerapkan Algoritma K-Means *Clustering* dalam analisis klasterisasi wilayah berdasarkan data penyaluran bantuan pangan oleh PT Yasa Artha Trimanunggal. Metodologi ini dirancang secara rinci untuk memastikan bahwa setiap tahap penelitian dilakukan secara efektif dan relevan. Penelitian ini mencakup langkah-langkah seperti identifikasi data, pembersihan, pemrosesan data, penerapan algoritma klasterisasi, dan interpretasi hasil. Tujuan akhir dari penelitian ini adalah memberikan wawasan mendalam tentang pola penyaluran bantuan pangan dan memberikan rekomendasi yang dapat diandalkan untuk perbaikan di masa depan. Pendekatan metodologis ini penting untuk memastikan konsistensi, validitas, dan reliabilitas analisis klasterisasi. Dengan pemahaman mendalam terhadap setiap langkah penelitian, peneliti dapat memastikan integrasi data yang baik, penerapan algoritma K-Means yang tepat, dan hasil klasterisasi yang bermakna. Peneliti akan bekerja sama dengan PT Yasa Artha Trimanunggal untuk mendapatkan dataset yang komprehensif. Tahap pra-pemrosesan mencakup pembersihan, penanganan nilai hilang, dan normalisasi data. Implementasi algoritma K-Means melibatkan penentuan jumlah klaster optimal melalui pendekatan iteratif. Hasil klasterisasi akan dievaluasi menggunakan metrik seperti indeks *Davies-Bouldin* dan *Silhouette* untuk mengukur kualitas klaster [8]. Analisis lebih lanjut akan mengidentifikasi karakteristik klaster dan implikasinya terhadap penyaluran bantuan pangan. Dengan demikian, penelitian ini tidak hanya menghasilkan temuan klasterisasi yang akurat tetapi juga memberikan kontribusi penting dalam penggunaan *Data Mining* untuk analisis distribusi bantuan pangan di tingkat provinsi. Pendekatan ini menjadi landasan yang kuat untuk memahami dan memberikan implikasi dari hasil analisis klasterisasi serta menjadi referensi bagi penelitian selanjutnya.

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, penerapan metode klasterisasi dengan algoritma K-Means mencakup beberapa tahapan penting. Tahapan pertama adalah menentukan jumlah klaster yang diinginkan, diikuti oleh inisialisasi pusat klaster. Kemudian, dilakukan iterasi untuk menemukan klaster yang optimal. Proses ini melibatkan pembagian data ke dalam klaster berdasarkan kedekatan dengan pusat klaster terdekat, pembaruan pusat klaster berdasarkan rata-rata data dalam klaster, dan pengulangan langkah-langkah tersebut sampai tercapai konvergensi. Hasil akhir dari proses ini adalah pembentukan kelompok klaster yang mencerminkan pola atau karakteristik dalam data, yang kemudian diinterpretasikan. Untuk membantu memahami hasil klasterisasi, visualisasi digunakan.

3.1 Proses Perhitungan Algoritma K-Means

Sebelum memulai iterasi algoritma klasterisasi K-Means, langkah pertama yang dilakukan adalah inisialisasi pusat klaster. Pada tahap ini, posisi awal pusat klaster ditetapkan untuk digunakan selama proses iteratif. Inisialisasi pusat klaster ini sangat mempengaruhi konvergensi algoritma dan hasil akhir dari proses klasterisasi.

Dalam penelitian ini, penulis menentukan pusat kluster Tabel 1 berdasarkan nilai dari total penerima terbesar dan total penduduk terbesar, serta rata usia penerima tertinggi di setiap wilayah.

Tabel 1. Inisialisasi Pusat Kluster

<i>Centroid</i>	Total Penerima	Total Penduduk
1	60	1,213
2	1,154	95,676
3	8,269	29,841

Selanjutnya proses iterasi. Iterasi dalam algoritma *K-Means* merupakan proses berulang yang dimulai setelah pusat awal kluster ditetapkan. Proses ini terdiri dari dua tahap utama. Pertama, setiap titik data ditempatkan ke dalam kluster dengan pusat terdekat, langkah ini disebut "*Assign Cluster*". Selanjutnya, pusat kluster diperbarui dengan menghitung rata-rata posisi semua titik data dalam kluster tersebut, yang disebut "*Update Pusat Cluster*". Kedua langkah ini diulang terus menerus sampai tidak ada perubahan signifikan dalam posisi pusat kluster atau sampai jumlah iterasi yang telah ditentukan tercapai. Tujuan dari proses ini adalah meminimalkan jumlah kuadrat jarak antara titik data dan pusat kluster, sehingga terbentuk kluster yang mencerminkan pola atau struktur dalam dataset. Proses ini menggunakan rumus (1) berikut.

$$\sqrt{((x_2 - x_1)^2 + (y_2 - y_1))^2 + (z_2 - z_1)^2} \quad (1)$$

Dengan data yang ada di tabel 1 dan dari rumus (1) tersebut didapatkan hasil iterasi pertama dan iterasi selanjutnya sampai data menjadi hasil yang konsisten sebagai Tabel 2 berikut.

Tabel 2. Iterasi Pertama

<i>Centroid</i>	Total penerima	Total Penduduk	RataUsiaPenerima
1	8.269	29.841	51,295
2	1.154	95.676	54,268
3	3	2.593	73,666
	Jumlah C1		200
	Jumlah C2		10
	Jumlah C3		51

Berikutnya, nilai centroid yang baru harus ditentukan untuk iterasi selanjutnya dengan menggunakan persamaan (2) berikut:

$$C = \frac{\sum m}{n} \quad (2)$$

Pada setiap iterasi, dilakukan penghitungan ulang centroid berdasarkan data yang telah diperbarui, dimulai dari iterasi kedua dan seterusnya. Nilai centroid yang baru beserta hasil iterasi kedua dapat dilihat pada tabel 3 berikut ini:

Tabel 3. Iterasi Kedua

<i>Centroid</i>	Total Penerima	Total Penduduk	RataUsiaPenerima
1	1.124	25.720	52,150
2	1.316	76.687	44,725
3	565	8.242	58,137
	Jumlah C1		136
	Jumlah C2		20
	Jumlah C3		105

Apabila hasil *clustering* masih berubah, maka perlu menetapkan nilai centroid baru untuk iterasi selanjutnya. Proses ini diulang hingga mencapai centroid akhir. Hasil centroid terbaru dan hasil iterasi ketiga dapat dilihat pada tabel 4 di bawah ini:

Tabel 4. Iterasi Ketiga

<i>Centroid</i>	Total Penerima	Total Penduduk	RataUsiaPenerima
1	1.162	27.793	50,449

<i>Centroid</i>	Total Penerima	Total Penduduk	RataUsiaPenerima
2	1.306	65.596	48,136
3	787	11.805	57,318
Jumlah C1			100
Jumlah C2			29
Jumlah C3			132

Proses iterasi dihentikan pada iterasi kesebelas Tabel 5 karena sudah mencapai centroid akhir. Proses iterasi dihentikan di iterasi ke sebelas karena data telah dikelompokkan secara konsisten/tidak berubah.

Tabel 5. Iterasi Kesebelas

<i>Centroid</i>	Total Penerima	Total Penduduk	RataUsiaPenerima
1	881	24.127	5,189
2	1.153	50.901	56,334
3	1.004	17.645	57,173
Jumlah C1			20
Jumlah C2			48
Jumlah C3			193

Proses iterasi pada algoritma K-Means telah dihentikan setelah mencapai kondisi konvergensi, yaitu ketika pusat kluster tidak lagi mengalami perubahan signifikan atau ketika jumlah iterasi yang telah ditentukan sebelumnya tercapai. Hasil akhir dari klusterisasi diperoleh dari konfigurasi pusat kluster pada iterasi terakhir. Setelah proses konvergensi tercapai, kluster yang terbentuk berada dalam kondisi stabil, memungkinkan kita untuk menginterpretasikan dan menganalisis hasil klusterisasi dengan lebih akurat. Pada titik ini, hasil klusterisasi mencerminkan struktur dan pola yang ada dalam dataset, memberikan wawasan berharga untuk langkah-langkah analisis selanjutnya. Dengan demikian, setelah mencapai konvergensi, dapat dengan percaya diri mengevaluasi dan menggunakan hasil klusterisasi yang stabil untuk berbagai aplikasi dan penelitian lebih lanjut.

Tabel 6. Hasil Iterasi

<i>Centroid</i>	Rata – rata Nilai		
	Total Penerima	Total Penduduk	RataUsiaPenerima
1	881	24.127	5,189
2	1.153	50.901	56,334
3	1.004	17.645	57,173
Jumlah C1			20
Jumlah C2			48
Jumlah C3			193

Dari data tabel 6 yang dihasilkan melalui proses iterasi terakhir pada algoritma K-Means, di mana pusat kluster atau Centroid telah diperbarui, kita dapat memperoleh informasi yang sangat penting. Dengan membagi jumlah anggota di setiap kluster, kita bisa menghitung rata-rata posisi atau nilai tengah dari setiap kluster tersebut. Perhitungan ini memberikan gambaran yang lebih mendetail tentang distribusi dan karakteristik setiap kluster. Dengan kata lain, melalui analisis ini, kita dapat mengevaluasi sejauh mana setiap kluster mencerminkan pola atau kelompok data yang seragam di dalamnya.

3.2 *Sum of Square Within Cluster (SSWC)*

Langkah berikutnya adalah menghitung SSWC (*Sum of Squares Within Cluster*) dengan persamaan (3), yang mengukur seberapa dekat titik-titik data dalam kluster terhadap pusat kluster. Nilai SSWC yang rendah menunjukkan bahwa data dalam kluster berdekatan, menciptakan kluster yang padat dan seragam[9]. Untuk menghitung SSWC, jarak setiap titik data dari pusat klusternya diukur, kemudian dikuadratkan dan dijumlahkan. Semakin kecil nilai SSWC, semakin baik kualitas klusterisasi karena menunjukkan bahwa titik-titik dalam kluster sangat dekat dengan pusatnya. SSWC yang rendah menandakan kluster dengan homogenitas tinggi, artinya kluster tersebut lebih representatif terhadap pola data. Oleh karena itu, SSWC adalah metrik penting untuk menilai efektivitas algoritma K-Means, di mana nilai SSWC yang lebih kecil mengindikasikan kualitas klusterisasi yang lebih tinggi.

$$SSWC = \sum_{i=1}^K \sum_{j=1}^{n_i} ||x_{ij} - c_i||^2 \quad (3)$$

Dengan rincian sebagai berikut:

- K adalah jumlah Klaster.
- n_i adalah jumlah titik dalam Klaster ke – i.
- x_{ij} adalah titik ke – j dalam klaster ke – i
- c_i adalah pusat klaster ke – i

Berikut Tabel 7 adalah nilai SSWC yang diperoleh dari hasil perhitungan di atas, yang akan memberikan gambaran lebih lanjut mengenai kualitas klasterisasi yang telah dilakukan.

Tabel 7. *Sum of Square Within Cluster (SSWC)*

SSW	Cluster 1	Cluster 2	Cluster 3
Nilai	16.588,2683	11.608,2809	7.422,5161

3.3 Sum of Square Between Cluster (SSBC)

SSBC (*Sum of Squares Between Clusters*) persamaan (4) adalah cara untuk mengukur seberapa jauh pusat klaster satu dengan lainnya dalam klasterisasi. Kalau nilai SSBC besar, itu berarti pusat-pusat klaster berada cukup jauh satu sama lain, sehingga kelompok-kelompok datanya benar-benar terpisah[10]. Untuk menghitung SSBC, kita melihat jarak antara pusat setiap klaster dan pusat keseluruhan data. Ini membantu kita memahami bagaimana pusat-pusat klaster tersebar di ruang data. Nilai SSBC yang besar berarti klaster-klasternya berbeda dengan jelas, jadi klasterisasi tersebut efektif dalam memisahkan data menjadi kelompok-kelompok yang berbeda. Oleh karena itu, SSBC adalah alat penting untuk menilai seberapa baik data dikelompokkan. Dengan menghitung SSBC, kita bisa mendapatkan gambaran lebih baik tentang struktur klaster yang terbentuk dan memastikan bahwa perbedaan antar klaster memang signifikan, yang membantu dalam memahami hasil klasterisasi.

$$SSBC = \sum_{i=1}^K n_i \cdot ||c_i - c_{global}||^2 \quad (4)$$

Dengan rincian sebagai berikut:

- K adalah jumlah Klaster.
- n_i adalah jumlah titik dalam klaster ke – i.
- c_i adalah pusat klaster ke – i.
- c_{global} adalah pusat klasternya secara keseluruhan

Berikut Tabel 8 adalah nilai SSBC yang diperoleh dari hasil perhitungan di atas, yang akan memberikan gambaran lebih lanjut mengenai kualitas klasterisasi yang telah dilakukan.

Tabel 8. *Sum of Square Between Cluster (SSBC)*

SSB	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0,0000	57.730,1482	52.387,0343
Cluster 2	57.730,1482	0,0000	33.266,1740
Cluster 3	52.387,0343	33.266,1740	0,0000

3.4 Davies-Bouldien Index

Langkah terakhir dalam proses *Knowledge Discovery in Databases* (KDD) adalah memberikan pemahaman dan makna pada setiap klaster yang telah terbentuk. Pada tahap ini, Centroid dari iterasi terakhir dibagi dengan jumlah anggota setiap klaster. Tujuannya adalah untuk memberikan deskripsi atau karakteristik rinci dari masing-masing klaster, sehingga dapat memahami lebih dalam pola atau atribut yang mewakili setiap kelompok data. Selain itu, evaluasi klaster dilakukan menggunakan metode *Davies-Bouldin Index* (DBI). Metode ini digunakan untuk menghitung rata-rata nilai setiap titik dalam himpunan data, yang membantu mengevaluasi seberapa baik klaster-klaster tersebut terpisah satu sama lain. Berikut untuk rumus (5) perhitungannya:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

Hasil evaluasi ini memberikan wawasan tambahan tentang kehomogenan dan pemisahan antar kluster, sehingga memperkuat pemahaman mengenai kualitas klusterisasi yang telah dilakukan. Berikut Tabel 9 hasil perhitungan DBI dengan rumus (5) di atas:

Tabel 9. Davies-Bouldien Index (DBI)

R	Cluster 1	Cluster 2	Cluster 3	R Max
Cluster 1	0,0000	0,4884	0,4583	0,4884
Cluster 2	0,4884	0,0000	0,5721	0,5721
Cluster 3	0,4583	0,5721	0,0000	0,5721

Dari nilai R Max kita dapat mencari untuk nilai DBI dengan mencari nilai rata – rata nya dengan cara sebagai berikut :

$$M = 4884 + 5721 + 5721 = 16326$$

$$DBI = \frac{0,4884 + 0,5721 + 0,5721}{3} = 0,5442$$

$$DB = \frac{1}{3} \times 0,5442 = 0,1814$$

Evaluasi kluster menggunakan perhitungan *Davies-Bouldin Index* (DBI) menghasilkan nilai sebesar 0.1814 dengan rincian perhitungan yang terlampir. Semakin kecil atau mendekati 0 nilai DBI, semakin baik kualitas kluster yang terbentuk. Hasil ini menunjukkan bahwa klusterisasi yang dilakukan sangat baik, mengindikasikan pemisahan yang jelas antara kluster data homogenitas yang tinggi di dalam masing-masing kluster.

4. KESIMPULAN

Berdasarkan evaluasi dari hasil penelitian berjudul "Implementasi Data Mining Klusterisasi dengan Algoritma K-Means untuk Pengelompokan Kelurahan di Provinsi DKI Jakarta berdasarkan Total Penerima, Total Penduduk, dan Rata-rata Usia Penerima Bantuan Pangan Bulog di PT Yasa Artha Trimanggung," beberapa kesimpulan yang dapat ditarik adalah sebagai berikut: Algoritma K-Means terbukti efektif dalam mengelompokkan wilayah di DKI Jakarta berdasarkan data penerima bantuan pangan, di mana klusterisasi yang dihasilkan dapat mengidentifikasi area dengan kesamaan dalam total penerima, total penduduk, dan rata-rata usia penerima, sehingga dapat membantu dalam menentukan prioritas penyaluran bantuan. Selain itu, penggunaan *Davies-Bouldin Index* (DBI) sebagai alat evaluasi klusterisasi membantu menentukan jumlah kluster yang optimal; nilai DBI yang rendah, yakni 0,5442 dalam penelitian ini, menunjukkan bahwa kluster yang terbentuk memiliki jarak antar-kluster yang baik dan masing-masing kluster cukup kompak, yang berarti pembagian kluster sudah cukup baik dan terpisah dengan jelas. Penelitian ini juga memiliki kaitan langsung terhadap kebijakan penyaluran bantuan pangan di DKI Jakarta, di mana data dan klusterisasi yang dihasilkan memberikan dasar kuat untuk pengambilan keputusan yang lebih informatif dan strategis, diharapkan dapat meningkatkan kesejahteraan masyarakat secara keseluruhan.

DAFTAR PUSTAKA

- [1] R. Astuti and K. Ukar, "Implementasi Data Mining Dengan Metode *Clustering* Algoritma K-Means Untuk Pengelompokan Data Tilang di Instansi Pemerintah," *Media Informatika*, vol. 20, no. 2, pp. 109-121, 2021.
- [2] J. Homepage, D. Syaputri, P. Herwina Noprita, and S. Romelah, "Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 1–6, 2021.
- [3] Y. Radana Sembiring, R. Winanjaya, S. Tunas Bangsa, S. Utara, and I. A. Jln Sudirman Blok No, "Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma K-Means," *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 2, no. 2, pp.125-132, 2021. [Online]. Available: <https://www.bps.go.id>
- [4] N. Nurahman and J. Susanto, "Klusterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma K-Means," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 461, Apr. 2023, doi: 10.30865/jurikom.v10i2.5807.
- [5] F. Handayanna, "Penerapan Algoritma K-Means Untuk Klusterisasi Penduduk Miskin di Provinsi Banten," *INTI Nusa Mandiri*, vol. 18, no. 1, pp. 93–99, 2023, doi: 10.33480/inti.v18i1.4399.
- [6] A. Fatkhudin, A. Khambali, F. A. Artanto, N. A. Putra Zade, and U. Muhammadiyah Pekajangan Pekalongan, "Implementasi Algoritma *Clustering* K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika)," *Jurnal Minfo Polgan*, vol. 12, no. 2, 2023, doi: 10.33395/jmp.v12i2.12494.

- [7] I. Febriani, M. Safii, and O. Alfina, "Implementasi Data Mining Peningkatan Produksi Beras Menggunakan Metode K-Means *Clustering*," *Majalah Ilmiah Methoda*, vol. 12, no. 3, pp. 258-268, 2022.
- [8] T. B. Santos, "Aplikasi Data Mining untuk *Clustering* Daerah Penyebaran Penyakit Diare di DKI Jakarta Menggunakan Algoritma K-MEANS," *Jurnal Ilmiah FIFO*, vol. 11, no. 2, pp. 131-143, 2019, doi: 10.22441/fifo.2019.v11i2.003.
- [9] R. R. A. Aria, S. Susilowati, and I. R. Rahadjeng, "Data Mining Menentukan *Cluster* Penerima Program Bantuan dengan Metode K-Means," *Remik*, vol. 7, no. 1, pp. 291–300, 2023, doi: 10.33395/remik.v7i1.12030.
- [10] S. Indhira and B. Hendrik, "Penerapan Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit ISPA," *Journal of Information System and Education Development*, vol. 1, no. 3, pp. 31-35, 2023.