

Perbandingan *K-Means* Dan *K-Medoids Clustering* Dalam Pengelompokan Kabupaten/Kota Berdasarkan Produksi Tanaman Pangan Di Provinsi Sumatera Barat

Syahraini¹, Yusmet Rizal²

¹²Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Negeri Padang
e-mail: syahrainii7@gmail.com

Abstrak

Jumlah produksi tanaman pangan dalam lima tahun terakhir di Sumatra Barat selalu menurun yang dapat mempengaruhi perekonomian dan ketahanan pangan di Provinsi Sumatera Barat. Dalam meningkatkan ketahanan pangan dilakukan clustering untuk mengidentifikasi kemiripan potensi antar Kabupaten/Kota berdasarkan produksi tanaman pangan di Sumatera Barat menggunakan *K-Means* dan *K-Medoids Clustering*. Penelitian ini bertujuan untuk membandingkan dan mengetahui algoritma mana yang lebih optimal antara *K-Means* dan *K-Medoids Clustering* dalam pengelompokan Kabupaten/Kota berdasarkan produksi tanaman pangan di Provinsi Sumatera Barat serta menginterpretasikan hasil pengelompokannya. Penelitian ini merupakan penelitian terapan dan jenis data yang digunakan dalam penelitian ini adalah data sekunder tentang jumlah produksi tanaman pangan di Provinsi Sumatera Barat tahun 2022. Sumber data diperoleh dari website resmi Badan Pusat Statistika Sumatera Barat. Dengan menggunakan 7 variabel yaitu padi, jagung, kedelai, kacang hijau, kacang tanah, ubi jalar dan ubi kayu. Pada penelitian ini dilakukan proses clustering dengan menggunakan jarak Euclidean dan evaluasi cluster dengan menggunakan DBI (Davies Bouldin Index). Hasil penelitian menunjukkan bahwa cluster yang terbentuk yaitu sebanyak 3 cluster dilihat dari $K = 3$ dengan menggunakan metode elbow. Pada *K-Means Clustering* diperoleh hasil cluster 1 sebanyak 11 Kabupaten/Kota, cluster 2 sebanyak 6 Kabupaten/Kota, cluster 3 sebanyak 2 Kabupaten/Kota dan hasil evaluasi cluster DBI sebesar 0.64. Pada *K-Medoids Clustering* diperoleh hasil cluster 1 sebanyak 8 Kabupaten/Kota, cluster 2 sebanyak 4 Kabupaten/Kota, cluster 3 sebanyak 4 Kabupaten/Kota dan hasil evaluasi cluster DBI sebesar 0.92. Sehingga, algoritma yang lebih optimal dalam pengelompokan Kabupaten/Kota berdasarkan Produksi Tanaman Pangan di Provinsi Sumatera Barat adalah menggunakan *K-Means Clustering*.

Kata Kunci: *Tanaman Pangan, Clustering, K-Means K-Medoids*

Abstract

The amount of food crop production in the last five years in West Sumatra has always decreased, which can affect the economy and food security in West Sumatra Province. To improve food security, clustering is carried out to identify potential similarities between districts/cities based on food crop production in West Sumatra using *K-Means* and *K-Medoids Clustering*. This research aims to compare and find out which algorithm is more optimal between *K-Means* and *K-Medoids Clustering* in grouping districts/cities based on

food crop production in West Sumatra Province and interpreting the grouping results. This research is applied research and the type of data used in this research is secondary data regarding the amount of food crop production in West Sumatra Province in 2022. The data source was obtained from the official website of the West Sumatra Central Statistics Agency. Using 7 variables, namely rice, corn, soybeans, green beans, peanuts, sweet potatoes and cassava. In this research, a clustering process was carried out using Euclidean distance and cluster evaluation using DBI (Davies Bouldin Index). The research results showed that the clusters formed were 3 clusters seen from $K = 3$ using the elbow method. In K-Means Clustering, the results for cluster 1 were 11 districts/cities, cluster 2 was 6 districts/cities, cluster 3 was 2 districts/cities and the DBI cluster evaluation result was 0.64. In K-Medoids Clustering, the results obtained for cluster 1 were 8 districts/cities, cluster 2 was 4 districts/cities, cluster 3 was 4 districts/cities and the DBI cluster evaluation result was 0.92. So, a more optimal algorithm for grouping districts/cities based on food crop production in West Sumatra Province is to use K-Means Clustering.

Keywords: *Food Crop, Clustering, K-Means, K-Medoids*

PENDAHULUAN

Sektor Pertanian merupakan salah satu sektor terpenting di Indonesia dalam pembangunan nasional. Sumatera Barat merupakan salah satu provinsi yang sedang berusaha untuk meningkatkan perekonomian daerah terutama dibidang pertanian, dikarenakan pertanian masih menjadi sektor strategis bagi perekonomian Sumatera Barat. Hal ini terdapat dalam Rencana Pembangunan Jangka Menengah Daerah (RPJMD) Provinsi Sumatera Barat 2021-2026 (Bappeda Sumatera Barat, 2021). Sektor pertanian merupakan sektor unggulan yang berkontribusi besar dalam meningkatkan dan mengembangkan perekonomian daerah (BPS Provinsi Sumatera Barat, 2023). Salah satu subsektor yang mempengaruhi pertumbuhan sektor pertanian yaitu subsektor tanaman pangan.

Menurut Widodo & Setijorini (2019), Tanaman pangan merupakan kelompok tanaman budi daya yang mengandung karbohidrat, protein dan lemak yang dimanfaatkan oleh manusia sebagai sumber energi. Komoditas tanaman pangan yang dihasilkan di Sumatera Barat adalah padi, jagung, kedelai, kacang tanah, kacang hijau, ubi jalar dan ubi kayu. Berikut adalah data hasil produksi tanaman pangan tahun 2018-2022:

Tabel 1. Data Hasil Produksi Tanaman Pangan (Ton) di Sumatera Barat Tahun 2018 - 2022

Tanaman Pangan	Hasil Produksi (Ton)				
	2018	2019	2020	2021	2022
Padi	1.485.094,48	1.485.015,01	1.389.289,29	1.317.209,38	1.373.532,19
Jagung	995.179,20	922.149,46	941.485,95	948.063,16	853.023,56
Kedelai	2.225,54	2.259,00	46,58	6,27	19,08
Kacang hijau	311,99	340,96	296,88	241,00	391,6
Kacang tanah	4.810,06	4.498,52	4.439,04	4.703,12	2.996,37
Ubi jalar	138.066,00	121.518,42	133.930,30	125.200,65	122.958,53
Ubi kayu	203.171,00	170.941,32	154.728,76	153.412,02	143.330,27
Jumlah	2.828.858,27	2.706.722,69	2.624.216,80	2.548.835,60	2.496.251,60

Sumber: BPS Sumbar

Jumlah hasil produksi tanaman pangan di Provinsi Sumatera Barat selama lima tahun terakhir mengalami penurunan. Jika jumlah pangan berkurang terus menerus maka tidak akan mampu memenuhi kebutuhan konsumsi masyarakat di Provinsi Sumatera Barat dikarenakan jumlah penduduk di Sumatera Barat yang setiap tahunnya selalu meningkat. Hal ini merupakan ancaman yang serius sehingga perlu dilakukan peningkatan produksi tanaman pangan untuk memenuhi kebutuhan pangan masyarakat Sumatera Barat.

Potensi produksi tanaman pangan di Sumatera Barat dapat diketahui dengan melihat kemiripan potensi antar wilayah, sehingga kebijakan pembangunan pertanian khususnya tanaman pangan dapat difokuskan pada wilayah yang potensial. Untuk mengidentifikasi kemiripan potensi antar Kabupaten/Kota berdasarkan produksi tanaman pangan di Sumatera Barat dapat dilakukan pengelompokan. Untuk mengetahui kesamaan karakteristik setiap kelompok antar Kabupaten/Kota dapat menggunakan *clustering*.

Clustering merupakan proses pengelompokan data dalam beberapa *cluster* berdasarkan karakteristik yang dimiliki, dimana dalam satu *cluster* yang sama memiliki tingkat kemiripan yang tinggi sedangkan antar *cluster* lainnya memiliki tingkat kemiripan yang rendah. Tujuannya untuk mencari hasil kelompok yang optimal dan mengetahui objek pada *cluster* serupa (Han, Kamber & Pei, 2012). Salah satu metode analisis *cluster* yang dapat digunakan yaitu metode non hierarki.

Metode non hierarki merupakan metode pengelompokan data yang langsung membagi data menjadi beberapa kelompok yang ditentukan tanpa menunjukkan hubungan antar kelompoknya karena metode ini menentukan jumlah *cluster* terlebih dahulu dan menentukan nilai *centroid* pada tiap *cluster* (Han & Kamber, 2006). Metode non hierarki yang paling terkenal dan umum digunakan adalah *K-Means* dan *K-Medoids Clustering*. Keuntungan dari kedua metode tersebut yaitu proses pengerjaan yang lebih cepat dibanding metode hierarki dan mudah digunakan pada objek yang berjumlah besar.

K-Means Clustering adalah salah satu metode *partitional clustering* yang berguna untuk mengolah dan mengelompokkan data menjadi satu atau lebih *cluster*. *K-Means Clustering* akan berjalan jika jumlah *cluster* diketahui terlebih dahulu, kemudian sebanyak *k* titik dipilih secara acak sebagai *centroid* (pusat *cluster*) (Witten & Frank, 2005). *K-Medoids Clustering* mirip dengan *K-Means Clustering* yang juga merupakan *clustering* dari *partitional clustering* (Kaufman & Rousseeuw, 1990). Namun, perbedaan dari kedua *clustering* tersebut terletak pada titik pusat *cluster*. Titik pusat *cluster* pada *K-Medoids Clustering* adalah titik data yang sebenarnya. Titik ini disebut *Medoid* (Murphy, 2012).

Pada proses *cluster nonhierarki* variabel cenderung memiliki *range* yang sangat bervariasi antara satu dan yang lainnya. Hal ini menyebabkan suatu data mempunyai pengaruh yang besar terhadap hasil. Oleh karena itu, perlu dilakukan normalisasi data. Namun, jika skala pada data sudah memiliki satuan data yang sama tidak perlu dilakukan normalisasi data. Adapun normalisasi data dapat dilakukan dengan normalisasi *min-max* dengan menggunakan rumus berikut (Han & Kamber, 2006):

$$v' = \frac{v - \min_A}{\max_A - \min_A}$$

(1)

Proses saat melakukan *K-Means* dan *K-Medoids Clustering*, terlebih dahulu menentukan jumlah *cluster* yang akan dibentuk dengan menggunakan metode *elbow* dengan cara menghitung nilai *SSE* (*Sum of Squares Error*) (Hastie, Tibshirani & Friedman, 2008). Perhitungan nilai *SSE* dapat dilakukan dengan persamaan:

$$SSE = \sum_{k=1}^K \|x_i - c_k\|^2$$

(2)

Untuk mengukur kemiripan atau ketidakmiripan dari titik data ke *centroid* menggunakan jarak *euclidean*. Semakin dekat jarak antar objek satu dengan objek-objek lainnya, maka semakin mirip karakteristik suatu objek data tersebut (Bishop, 2006). Perhitungan jarak *euclidean* dapat dilakukan dengan persamaan:

$$d(x, y) = \|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (3)$$

Algoritma *K-Means* menentukan nilai *centroid* dengan menggunakan nilai rata-rata (*means*), sedangkan algoritma *K-Medoids* menentukan nilai *centroid* dengan menggunakan nilai data sebenarnya (Han & Kamber, 2006). Perhitungan *centroid* baru pada *K-Means Clustering* dapat dilakukan dengan persamaan:

$$c_i = \frac{1}{n} \sum_{k=1}^n x_i \quad (4)$$

Algoritma *K-Medoids* akan berhenti jika nilai simpangan total (*S*) positif ($S > 0$). Perhitungan *centroid* baru pada *K-Means Clustering* dapat dilakukan dengan persamaan:

$$S = \text{Total Jarak Baru} - \text{Total Jarak Lama} \quad (5)$$

Interpretasi hasil suatu *cluster* merupakan cara mendeskripsikan karakteristik masing-masing *cluster* yang digunakan untuk menjelaskan perbedaan dari setiap *cluster*. Sehingga, dapat diketahui tingkatan masing-masing *cluster*. Untuk mengetahui karakteristik suatu *cluster* dengan cara menghitung rata-rata dari nilai setiap variabel (Hair JR, dkk, 2010).

Untuk mengetahui tingkat kualitas dari hasil *clustering* dilakukan evaluasi *cluster*. Salah satu metode yang dapat digunakan untuk mengevaluasi hasil *clustering* adalah *Davies Bouldin Index* (DBI). Hasil *cluster* yang terbaik ditunjukkan dengan nilai *Davies Bouldin Index* (DBI) yang terkecil. Jika nilai DBI semakin mendekati nilai 0 maka *cluster* yang diperoleh semakin baik. DBI diperoleh dengan menghitung nilai SSW, SSB dan Rasio. Perhitungan Nilai DBI dapat dilakukan dengan persamaan:

$$SSW_i = \frac{1}{n} \sum_{i=1}^n d(x, c) \quad (6)$$

$$SSB_{i,j} = d(c_i, c_j) \quad (7)$$

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (8)$$

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \quad (9)$$

METODE

Penelitian ini merupakan penelitian terapan dengan jenis data yang digunakan adalah data sekunder tentang jumlah produksi tanaman pangan di Provinsi Sumatera Barat tahun 2022. Sumber data diperoleh dari website resmi Badan Pusat Statistika Sumatera Barat yaitu <https://sumbar.bps.go.id/>. Objek penelitian yang digunakan pada penelitian ini adalah seluruh Kabupaten/Kota yang ada di Provinsi Sumatera Barat dengan total keseluruhan 19 wilayah yang terdiri dari 12 Kabupaten dan 7 Kota. variabel yang digunakan dalam penelitian ini sebagai berikut:

X_1 : Produksi padi

X_2 : Produksi jagung

X_3 : Produksi kedelai

X_4 : Produksi kacang hijau
 X_5 : Produksi kacang tanah
 X_6 : Produksi ubi jalar
 X_7 : Produksi ubi kayu

Untuk mempermudah dalam melakukan *clustering* maka berdasarkan data yang akan digunakan dapat disusun seperti pada Tabel 2.

Tabel 2. Struktur Data

Data ke- i	X_1	X_2	X_3	...	X_p
1	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$...	$x_{1,7}$
2	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$...	$x_{2,7}$
3	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$...	$x_{3,7}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$...	$x_{i,p}$

Dengan $x_{i,j}$ adalah nilai variabel yang menyatakan hasil produksi tanaman pangan di Sumatera Barat tahun 2022, dimana $i = 1, 2, 3, \dots, n$ dan $j = 1, 2, 3, \dots, p$. Pada penelitian ini banyaknya data yang digunakan yaitu $n = 19$ dan variabel yang digunakan yaitu $p = 7$. Sehingga, membentuk matriks 19×7 .

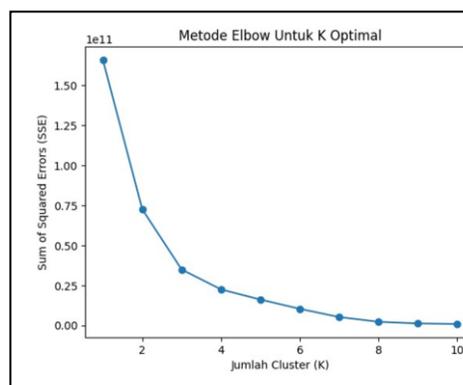
Adapun langkah-langkah pada teknik analisis data dalam penelitian ini sebagai berikut:

1. Mengidentifikasi masalah.
2. Mengambil data yang akan digunakan di *website* resmi BPS Sumatera Barat.
3. Melakukan normalisasi data menggunakan persamaan (1) (Jika diperlukan, jika tidak langsung ke langkah selanjutnya).
4. Menentukan jumlah *cluster* optimal yang akan dibentuk dengan menggunakan metode *Elbow* menggunakan persamaan (2).
5. Melakukan *Clustering* dengan *K-Means* dan *K-Medoids Clustering*, dengan tahapan sebagai berikut:
 - a. *K-Means Clustering*
 - 1) Menentukan titik *centroid* awal sebanyak k buah secara acak.
 - 2) Menghitung jarak dari masing-masing objek ke *centroid* awal dengan jarak *Euclidean* menggunakan persamaan (3).
 - 3) Kelompokkan setiap objek data berdasarkan jarak terdekat dari masing-masing objek ke *centroid* awal yang telah diperoleh.
 - 4) Tentukan *centroid* baru dengan menghitung rata-rata dari data untuk setiap *cluster* menggunakan persamaan (4)
 - 5) Jika masih ada data yang pindah *cluster*, ulangi langkah 3 hingga 5. Jika sudah tidak ada yang berpindah, maka proses *clustering* berhenti.
 - b. *K-Medoids Clustering*
 - 1) Menentukan titik *centroid* awal sebanyak k buah secara acak.
 - 2) Menghitung jarak dari masing-masing objek ke *medoid* sementara.
 - 3) Kelompokkan setiap objek data berdasarkan jarak terdekat dari masing-masing objek ke *medoid*, lalu hitung total jarak terdekat.
 - 4) Menentukan *medoid* baru (*non-medoid*) secara acak.
 - 5) Kelompokkan setiap objek data berdasarkan jarak terdekat dari masing-masing objek ke *medoid* baru, lalu hitung total jarak terdekat baru.
 - 6) Menghitung simpangan total (S) dengan persamaan (5).
 - a. Jika $S > 0$, proses *clustering* berhenti.

- b. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai *medoid*.
- 7) Lakukan iterasi sampai diperoleh nilai $S > 0$.
6. Melakukan interpretasi hasil yang telah diperoleh dengan *profiling cluster*.
7. Visualisasi hasil *cluster* ke peta.
8. Mengevaluasi hasil *cluster* dengan uji validasi *Davies Bouldin Index* (DBI) untuk mengetahui algoritma mana yang lebih optimal. Adapun tahapan dari uji validasi DBI sebagai berikut:
 - a. Menentukan nilai *SSW* menggunakan persamaan (6).
 - b. Menentukan nilai *SSB* menggunakan persamaan (7).
 - c. Menentukan nilai *Rasio* menggunakan persamaan (8).
 - d. Menentukan nilai *DBI* menggunakan persamaan (9).
9. Menentukan algoritma mana yang lebih optimal.
10. Menarik Kesimpulan.

HASIL DAN PEMBAHASAN

Data yang digunakan yaitu data hasil produksi tanaman pangan di Provinsi Sumatera Barat pada tahun 2022 dengan 19 Kabupaten/Kota yang dijadikan sebagai objek dan produksi tanaman pangan yang menjadi variabel dalam penelitian ini yaitu padi, kedelai, jagung, kacang hijau, kacang tanah, ubi jalar, ubi kayu. Pada penelitian ini tidak dilakukan normalisasi data dikarenakan skala data sudah sama yaitu ton. Sehingga, proses pengolahan data lanjut menentukan jumlah *cluster*. Teknik dalam analisis *cluster nonhierarki* merupakan teknik yang menentukan jumlah k nya terlebih dahulu sebelum dilakukannya proses *clustering*. Pada penelitian ini untuk menentukan jumlah *cluster* yang optimal menggunakan metode *elbow* dengan cara mengkuadratkan jumlah kesalahan pada semua objek dalam *cluster*. Nilai *SSE* berguna untuk menunjukkan seberapa jauh titik-titik data dari pusat *cluster* yang diberikan pada titik data tersebut. Semakin besar jumlah *cluster* maka nilai *SSE* Semakin kecil.



Gambar 1. Jumlah Cluster Optimal

Hasil perhitungan *SSE* menggunakan *Google Colaboratory* menunjukkan bahwa jumlah *cluster* yang optimal ditunjukkan pada grafik *elbow* tersebut berada pada $K = 3$, karena pada titik $K = 3$ memiliki bentuk siku. Artinya, berdasarkan metode *elbow*, jumlah *cluster* optimal yang akan digunakan berjumlah 3 *cluster*.

1. Analisis *Cluster* Menggunakan *K-Means Clustering*

Pada proses *K-means Clustering* diperoleh hasil pengelompokan sebanyak 3 *cluster* dengan anggota tiap *clusternya* yang dapat dilihat pada **Tabel 3**.

Tabel 3. Hasil Cluster Menggunakan K-Means Clustering

Kabupaten/Kota	Cluster
Kab. Kepulauan Mentawai	1
Kab. Pesisir Selatan	3
Kab. Solok	2
Kab. Sijunjung	1
Kab. Tanah Datar	2
Kab. Padang Pariaman	2
Kab. Agam	2
Kab. Lima Puluh Kota	2
Kab. Pasaman	2
Kab. Solok Selatan	1
Kab. Dharmasraya	1
Kab. Pasaman Barat	3
Kota Padang	1
Kota Solok	1
Kota Sawahlunto	1
Kota Padang Panjang	1
Kota Bukittinggi	1
Kota Payakumbuh	1
Kota Pariaman	1

Berdasarkan **Tabel 3**, dapat dilihat bahwa hasil *cluster* dengan menggunakan *K-Means Clustering*, yaitu *cluster 1* terdapat 11 anggota, *cluster 2* terdapat 6 anggota dan *cluster 3* terdapat 2 anggota. Setelah diperoleh hasil *cluster* dengan *K-Means Clustering*, maka untuk mengetahui karakteristik dari masing-masing *cluster* dilakukan dengan *profiling cluster* dengan menghitung nilai rata-rata variabel setiap *clusternya*.

Tabel 4. Karakteristik Cluster

Variabel	Rata-rata			
	Cluster 1	Cluster 2	Cluster 3	Total
Produksi Padi (X1)	108.992,01	21.829,57	21.821,33	77.291,17
Produksi Jagung (X2)	57.532,65	36.176,84	1.551,69	44.895,98
Produksi Kedelai (X3)	0,14	2,92	0,00	1,00
Produksi Kacang Hijau (X4)	9,64	47,60	0,00	20,61
Produksi Kacang Tanah (X5)	214,39	106,07	0,84	157,70
Produksi Ubi Jalar (X6)	10.555,22	1.141,86	0,00	6.471,50
Produksi Ubi Kayu (X7)	11.466,35	2.173,81	2.078,77	7.543,70

Dari **Tabel 4**, dapat diketahui karakteristik dari setiap *cluster* yang terbentuk yaitu *Cluster 1* merupakan *cluster* dengan hasil produksi tanaman pangan yang tergolong tinggi karena memiliki rata-rata variabel yang tinggi, *Cluster 2* merupakan *cluster* dengan hasil produksi tanaman pangan yang tergolong sedang karena memiliki rata-rata variabel yang sedang dan *Cluster 3* merupakan *cluster* dengan hasil produksi tanaman pangan yang tergolong rendah karena memiliki rata-rata variabel yang rendah.

2. Analisis Cluster Menggunakan *K-Medoids Clustering*

Pada proses *K-means Clustering* diperoleh hasil pengelompokan sebanyak 3 cluster dengan anggota tiap cluster-nya yang dapat dilihat pada **Tabel 5**.

Tabel 5. Hasil Pengelompokan *K-Medoids Clustering*

Kabupaten/Kota	Cluster
Kab. Kepulauan Mentawai	3
Kab. Pesisir Selatan	1
Kab. Solok	1
Kab. Sijunjung	2
Kab. Tanah Datar	1
Kab. Padang Pariaman	1
Kab. Agam	1
Kab. Lima Puluh Kota	1
Kab. Pasaman	1
Kab. Solok Selatan	2
Kab. Dharmasraya	2
Kab. Pasaman Barat	1
Kota Padang	2
Kota Solok	3
Kota Sawahlunto	3
Kota Padang Panjang	3
Kota Bukittinggi	3
Kota Payakumbuh	3
Kota Pariaman	3

Berdasarkan **Tabel 5**, dapat dilihat bahwa hasil cluster dengan menggunakan *K-Means Clustering*, yaitu cluster 1 terdapat 8 anggota, cluster 2 terdapat 4 anggota dan cluster 3 terdapat 7 anggota. Setelah diperoleh hasil cluster dengan *K-Medoids Clustering*, maka untuk mengetahui karakteristik dari masing-masing cluster dilakukan dengan *profiling cluster* dengan menghitung nilai rata-rata variabel setiap cluster-nya. Adapun karakteristik dari masing-masing cluster dapat dilihat **Tabel 6**.

Tabel 6. Karakteristik Cluster

Variabel	Rata-rata			
	Cluster 1	Cluster 2	Cluster 3	Total
Produksi Padi (X_1)	119.685,81	72.935,77	17.757,52	77.291,17
Produksi Jagung (X_2)	54.040,84	104.045,72	644,85	44.895,98
Produksi Kedelai (X_3)	0,19	0,00	2,51	1,00
Produksi Kacang Hijau (X_4)	9,00	79,88	0,01	20,61
Produksi Kacang Tanah (X_5)	200,70	338,16	5,45	157,70
Produksi Ubi Jalar (X_6)	13.958,33	2.566,05	146,81	6.471,50
Produksi Ubi Kayu (X_7)	14.645,38	3.450,15	1.766,66	7.543,70

Dari **Tabel 6** dapat diketahui karakteristik dari setiap cluster yang terbentuk yaitu Cluster 1 merupakan cluster dengan hasil produksi tanaman pangan yang tergolong tinggi karena memiliki rata-rata variabel yang tinggi, Cluster 2 merupakan cluster dengan hasil produksi tanaman pangan yang tergolong sedang karena memiliki rata-

rata variabel yang sedang dan *Cluster 3* merupakan *cluster* dengan hasil produksi tanaman pangan yang tergolong rendah karena memiliki rata-rata variabel yang rendah.

3. Perbandingan Algoritma Terbaik

Tabel 7. Hasil Evaluasi Cluster

Clustering	DBI
<i>K-Means</i>	0,64
<i>K-Medoids</i>	0,92

Dari **Tabel 7** di atas, terlihat bahwa nilai DBI untuk *K-Means Clustering* sebesar 0,64. Sedangkan, *K-Medoids Clustering* sebesar 0,92. Jika nilai DBI semakin mendekati nilai 0 maka cluster yang diperoleh semakin baik (Davies & Bouldin, 1979). *K-Means Clustering* memperoleh nilai DBI yang mendekati 0 dibanding *K-Medoids Clustering*. Hal ini menunjukkan bahwa pengelompokan Kabupaten/Kota berdasarkan produksi tanaman pangan di Sumatera Barat Tahun 2022 dapat lebih optimal jika menggunakan *K-Means Clustering*.

SIMPULAN

Berdasarkan hasil dan pembahasan yang diperoleh dapat disimpulkan bahwa dalam Pengelompokan Kabupaten/Kota berdasarkan Produksi Tanaman Pangan di Sumatera Barat Tahun 2022 menggunakan *K-Means* dan *K-Medoids Clustering* terbentuk menjadi 3 *Cluster* dengan anggota *cluster* yang berbeda. Tingkatan *cluster* yang terbentuk adalah tinggi, sedang dan rendah. Pengelompokan Kabupaten/Kota berdasarkan produksi tanaman pangan di Sumatera Barat Tahun 2022 menggunakan *K-Means* dan *K-Medoids Clustering* dapat diketahui bahwa *K-Means Clustering* lebih baik dibanding *K-Medoids Clustering*. Hal ini dapat diketahui dari evaluasi *cluster* menggunakan nilai DBI. Perbandingan evaluasi hasil *cluster* dengan nilai DBI menggunakan *K-Means Clustering* sebesar 0,64, sedangkan nilai DBI menggunakan *K-Medoids Clustering* sebesar 0,92. Oleh karena itu, pengelompokan Kabupaten/Kota berdasarkan akan memperoleh hasil terbaik apabila dilakukan dengan menggunakan *K-Means Clustering*.

DAFTAR PUSTAKA

- Badan Pusat Statistik (BPS) Provinsi Sumatera Barat. 2023. *Provinsi Sumatera Barat dalam Angka 2023*. Padang: BPS Provinsi Sumatera Barat.
- Bappeda Provinsi Sumatera Barat. *Rencana Pembangunan Jangka Menengah Daerah provinsi Sumatera Barat tahun 2016-2021*.
- Bishop, C.M. 2006. *Pattern Recognition and Machine Learning*. United Kingdom: Springer.
- Davies, D. L & Bouldin, D. W. 1979. A Cluster Separation Measure. *IEEE Transactions On Pattern Analysis And Machine Intelligene* Pami-1(2), 224-227. doi: <http://dx.doi.org/10.1109/TPAMI.1979.4766909>
- Hair JR, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. 2010. *Multivariate Data Analysis. 7th Edition*, New York: Perason.
- Han, J & Kamber, M. 2006. *Data Mining: Concept and Techniques. 2nd edition*, Amsterdam: Morgan Kaufmann-Elsevier.
- Han, J., Kamber, M., & Pei, J. 2012. *Data Mining: Concept and Techniques. 3rd edition*, Amsterdam: Morgan Kaufmann-Elsevier.

- Hastie, T., Tibshirani, R., & Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition*, Stanford: Springer.
- Kaufman, L., & Rousseeuw, P. 1990. *Finding Groups In Data: An Introduction to Cluster Analysis*. Hoboken: John Willey & Sons, Inc.
- Muningsih, Elly., Maryani, Ina., & Handayani, Vembria Rose. 2021. Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Evolusi: Jurnal Sains dan Manajemen*. 9(1), 95-100. doi: <https://doi.org/10.31294/evolusi.v9i1.10428>.
- Murphy, K.P. 2012. *Machine Learning :A Probabilistic Perspective*. London : Massachusetts Institute of Technology.
- Widodo, W.D., & Setijorini, L.E. 2019. *Budi Daya Tanaman Pangan Utama*. Banten: Universitas Terbuka.
- Witten, I.H., & Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques. 2nd edition*, San Francisco: Morgan Kaufmann-Elsevier