UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 3 Nomor 3 Tahun 2025 Halaman 271-278.
ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Comparison of Expectation-Maximization (EM) Algorithm and Kmeans for District/City Clustering in West Sumatera Province Based on Breadfruit Production

Mayrita Addila Putri dan Fadhilah Fitri*
Departemen Statistika , Universitas Negeri Padang, Padang, Indonesia
*Corresponding author: : fadhilahfitri@fmipa.unp.ac.id

Submitted : 09 Juli 2025
Revised : 07 Agustus 2025
Accepted : 20 Agustus 2025

ABSTRACT

Breadfruit (Artocarpus altilis) is an important food source that is highly nutritious and plays a strategic role in West Sumatra Province. However, challenges such as pests, diseases and marketing constraints affect its cultivation and productivity. This study employed K-means and expectation-maximisation (EM) clustering methods to categorise regions according to their breadfruit cultivation characteristics. The elbow method identified three optimal clusters for K-means and seven for EM. Evaluating the quality of the clusters using the silhouette coefficient produced values of 0.47 and 0.37 for EM and K-Means respectively, indicating that EM produced tighter, more distinct clusters. These results suggest that EM is a more effective method for describing the variation in breadfruit production in West Sumatra. With this in mind, the research is expected to inform strategic decision-making aimed at increasing the productivity and added value of breadfruit crops in the area..

Keywords: *Breadfruit production, Expectation-Maximization (EM), K-Means*

I. PENDAHULUAN

Tanaman sukun (*Artocarpus altilis*) merupakan salah satu tanaman pangan penting yang banyak ditemukan di daerah beriklim tropis, termasuk Indonesia (Sombamori Janggat et al., 2022). Tanaman ini telah lama dibudidayakan sebagai sumber pangan alternatif yang memiliki nilai ekonomi tinggi (Noviasari et al., 2023). Buah sukun mengandung nilai gizi yang signifikan, terutama terdapat karbohidrat sekitar 28,1%, protein 1,4 %, dan lemak 0,2% (Kementerian Kesehatan, 2017). Karena mengandung getah yang pahit dan zat tertentu yang dapat mengiritasi saluran pencernaan, buah sukun harus diolah terlebih dahulu, seperti direbus, dikukus, atau dipanggang agar aman dikonsumsi. Selain itu, sukun dapat diolah menjadi berbagai produk seperti keripik, kue, dan bubur yang meningkatkan nilai tambahnya (Sombamori Janggat et al., 2022).Tanaman ini juga berperan strategis dalam mendukung ketahanan pangan nasional dan memberikan kontribusi ekonomi bagi petani kecil di berbagai wilayah Indonesia (Noviasari et al., 2023).

Menurut Badan Pusat Statistik Provinsi Sumatera Barat, tanaman sukun memiliki peranan penting dalam diversifikasi pangan dan mendukung ekonomi lokal. Pada tahun 2023, budidaya sukun di Sumatera Barat menghadapi tantangan yang memengaruhi penurunan produktivitas dan kualitas hasil panen, seperti serangan hama dan penyakit tanaman (BPS, 2024). Salah satu contoh nyata dari penurunan produksi tersebut dapat ditemukan di Kabupaten Kepulauan Mentawai, yang tercatat sebagai daerah dengan produksi sukun terendah di Provinsi Sumatera Barat pada tahun 2023. Produksi sukun di wilayah ini hanya mencapai sekitar 50,3 kuintal dari 224 pohon dengan produktivitas yang sangat rendah, yaitu sebesar 0,22 kuintal per pohon. Kondisi ini mengindikasikan adanya tantangan serius dalam budidaya sukun di daerah tersebut, yang disebabkan oleh faktor-faktor seperti serangan hama dan penyakit, serta keterbatasan fasilitas pengolahan dan pemasaran yang menghambat pengembangan nilai tambah produk (BPS, 2024).

Penurunan produktivitas di wilayah seperti Kabupaten Kepulauan Mentawai maupun Kabupaten /kota di Provinsi Sumatera Barat mengharuskan dilakukannya analisis yang komprehensif dalam mengidentifikasi pola tantangan dalam budidaya sukun serta mengelompokkan wilayah berdasarkan karakteristik yang serupa, sehingga strategi pengembangan yang tepat dapat dirumuskan dan diimplementasikan. Sehubungan dengan hal tersebut, metode pengelompokan data (*clustering*) seperti *k-means* dan *Expectation-Maximization* (EM) dapat diterapkan (Ha

DOI: https://doi.org/10.24036/ujsds/vol3-iss3/403

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

et al., 2011). Metode k-means melakukan pengelompokan data berdasarkan jarak antara setiap data dengan pusat cluster, sedangkan metode EM menggunakan pendekatan probabilistik yang mempertimbangkan distribusi data serta memberikan probabilitas keanggotaan setiap data pada cluster tertentu (Ha et al., 2011).

Menurut hasil penelitian terdahulu, metode *Expectation-Maximization* (EM) clustering memiliki keunggulan dalam mengolah data yang memiliki distribusi kompleks serta *cluster* yang saling tumpang tindih secara parsial, sehingga mampu menghasilkan pengelompokan yang lebih akurat (Ha et al., 2011).Sebaliknya, metode k-means clustering lebih sederhana dan efektif digunakan pada data dengan *cluster* yang berbentuk bulat dan terpisah secara jelas (Zainudin et al., 2018). Untuk mengevaluasi kualitas hasil pengelompokan dari kedua metode tersebut, biasanya digunakan koefisien silhouette, yang berfungsi untuk menentukan metode mana yang paling sesuai dengan karakteristik data yang dianalisis(Mardiani, 2015). Pemilihan metode clustering yang tepat sangat penting agar hasil penelitian dapat memberikan gambaran yang valid dan dapat digunakan sebagai dasar pengambilan keputusan dalam pengembangan budidaya sukun (Ha et al., 2011)

Penelitian ini bertujuan untuk membandingkan efektivitas metode k-means dan *Expectation-Maximization* (EM) dalam mengelompokkan data produksi sukun berdasarkan Kabupaten/Kota di Provinsi Sumatera Barat dengan menggunakan koefisien *silhouette* sebagai indikator evaluasi kualitas *cluster*. Melalui perbandingan tersebut, diharapkan dapat diperoleh metode *clustering* yang paling tepat dalam mengelompokkan Kabupaten/Kota di Provinsi Sumatera Barat berdasarkan karakteristik budidaya sukun, sehingga pola tantangan yang dihadapi dapat diidentifikasi secara lebih akurat dan komprehensif. Selanjutnya, hasil penelitian ini diharapkan memberikan kontribusi yang signifikan dalam mendukung pengambilan keputusan strategis oleh pemerintah daerah maupun pelaku usaha pertanian dalam upaya meningkatkan produktivitas serta nilai tambah tanaman sukun di tingkat Kabupaten/Kota di Provinsi Sumatera Barat..

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari lembaga resmi, yaitu Badan Pusat Statistika Provinsi Sumatera Barat. Data yang digunakan berupa data produksi sukun di 19 Kabupaten/Kota Provinsi Sumatera Barat pada tahun 2023. Variabel – Variabel yang digunakan dalam penelitian ini sebagai berikut :

Tabel 1. Variabel Penelitian

Variabel	Penjelasan
X_1	Tanaman menghasilkan (pohon)
X_2	Produksi (Ton)
X_3	Produktivitas (Kuintal/Pohon)

B. Teknik Analisis Data

Analisis ini dilakukan dengan menggunakan perangkat lunak R-Studio sebagai alat bantu dalam proses pengolahan data. Penelitian ini menerapkan perbandingan metode *Expectation-Maximization* (EM) clustering deengan metode *K-Means* dengan menggunakan koefisien silhouette sebagai ukuran validasi cluster. Adapun teknik analisis yang digunakan sebagai berikut :

a. *Expectation-Maximization* (EM)

Expectation-Maximization (EM) merupakan algoritma partisi yang berbasis model dengan menggunakan konsep probabilitas untuk melakukan estimasi parameter pada model statistik. Algoritma ini dapa bekerja secara iteratif melalui dua langkah utama yaitu ekspektasi (E-step) dan langkah maksimal (M-step). Pada langkah ini algoritma akan dihitung nilai harapan dari variabel tersembunyi berdasarkan parameter saat ini, sedangkan pada langkah maksimasi , parameter akan diperbarui dengan cara memaksimalkan fungsi *likelihood* yang diharapan. Proses ini dilakukan secara berulang hingga konvergensi tercapai dengan menghasilkan estimasi parameter yang optimal dan meningkatkan akurasi model dalam mereprentasikan data. *Expectation-Maximization* (EM) dapat dihitung dengan tahapan berikut (Atika Nurani Ambarwati, 2019) :

1. Ekspektasi (E-step) digunakan untuk menghitung nilai ekspektasi dari fungsi *log likelihood* dengan rumus sebagai berikut :

$$Q = E\left[log[L(\theta)]\,|x_n, \hat{\theta}^{r-1}\right]$$

DOI: https://doi.org/10.24036/ujsds/vol3-iss3/403

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

2. Maksimal (M-step) digunakan untuk mencari nilai taksiran parameter berdasarkan hasil perhitungan pada langkah E-step yang bertujuan untuk memaksimumkan fungsi *log likelihood*. Proses ini dapat dinyatakan dengan persamaan berikut :

$$\hat{\eta}_j = \sum_{\eta=1}^{n} \frac{p(j|x_h\hat{\eta}_j{}^{(r-1)}, \hat{\mu}_{ij}{}^{(r-1)}, \hat{\sigma}_{ij}{}^{(r-1)}}{n} \quad (1)$$

$$g\hat{\mu}_{ij} = \frac{\sum_{\eta=1}^{n} x_{ih} p(j|x_h\hat{\eta}_j{}^{(r-1)}, \hat{\mu}_{ij}{}^{(r-1)}, \hat{\sigma}_{ij}{}^{(r-1)}}{n\hat{\eta}_j{}^{r}} \quad (2)$$

$$\hat{\mu}_{ij}{}^{2(r)} = \frac{\sum_{\eta=1}^{n}\sum_{j=1}^{c}(x_{ih} - \hat{\mu}_{ij}{}^{(r-1)})^2 p(j|x_h\hat{\eta}_j{}^{(r-1)}, \hat{\mu}_{ij}{}^{(r-1)}, \hat{\sigma}_{ij}{}^{(r-1)}}{\sum_{\eta=1}^{n}\sum_{j=1}^{c} p(j|x_h\hat{\eta}_j{}^{(r-1)}, \hat{\mu}_{ij}{}^{(r-1)}, \hat{\sigma}_{ij}{}^{(r-1)}} \quad (3)$$

Langkah M-step dilakukan secara terus menerus hingga mendapatkan nilai yang kovergen

3. Melakukan pemilihan kelompok terbaik melalui nilai BIC terkecil dari model berdasarkan persamaan berikut :

$$BIC_{LL} = -2LL + \ln(N)M$$

Dimana:
N : Banyakanya objek pengamatan
M : Jumlah parameter
LL : Nilai likelihood yang telah optimum

b. *K-Means*

K- means merupakan metode pengelompokkan data yang bekerja dengan cara menentukan pusat *cluster* atau *centroid* berdasarkan kemiripan antar data. Metode ini mengelompokkan data ke dalam beberapa kelompok sehingga data memiliki karakteristik yang serupa dalam *cluster* yang sama. Proses pengelompokkan dilakukan dengan mengukur jarak antara data dan *centroid*, kemudian melakukan pembaharuan posisi *centroid* hingga konvergen , sehingga menghasilkan pembagian cluster yang optimal (Mardiani, 2015).

Penentuan jumlah cluster yang tepat dalam metode ini dilakukan dengan menggunakan metode *elbow*. Metode *elbow* yaitu teknik yang menghitung nilai *Sum of Squares Errors* (SSE) untuk berbagai nilai cluster K. Nilai SSE tersebut kemudian diplot terhadap K, dan titik "*elbow*" pada grafik dimana penurunan SSE mulai melambat secara signifikan dengan memilih cluster terbaik. Metode *elbow* dapat dihitung dengan persamaan berikut(Rahmattullah et al., 2023) :

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

Dimana :
K : Jumlah *cluster*
x_i : Data ke – i dalam *cluster* ke- k
C_k: Pusat cluster ke- k
$\|x_i - C_k\|$: Jarak *euclidean* antara data dan pusat *cluster*

c. Koefisien *silhouette*

Koefisien *silhouette* merupakan metode yang diganakan untuk menghitung hasil pengelompokkan (*clustering*). Metode ini dilakukan dengan menggabungkan dua aspek penting yaitu kohesi dan saparasi. Kohesi digunakan untuk mengukur seberapa dekt objek-objek dalam satu cluster, sedangkan saparasi digunakan untuk mengukur seberapa

DOI: https://doi.org/10.24036/ujsds/vol3-iss3/403

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

jauh objek-objek tersebut dari *cluster* lain yang paling dekat. Proses perhitungan nilai *silhouette* dapat dilakukan dengan menggunakan persamaan sebagai berikut (Aselnino & Wijayanto, 2024) :

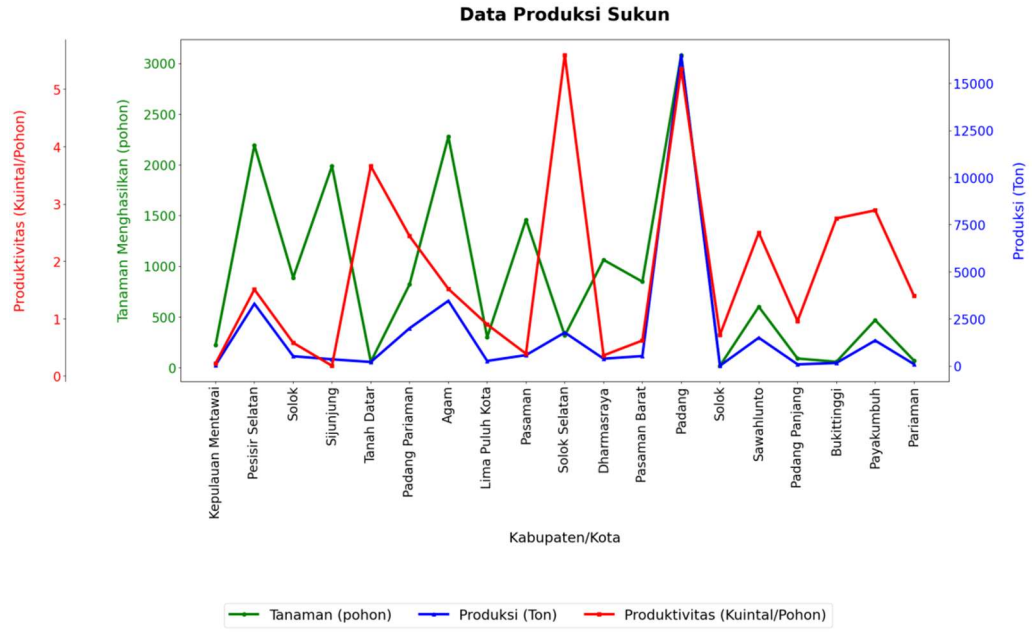
$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

Dimana :
$s(i)$: nilai *silhouette*
$a(i)$: rata –rata jarak objek i dengan objek yang berada di *cluster* berbeda
$b(i)$: rata –rata jarak objek i dengan seluruh objek yang berada di *cluster* yang sama

III. HASIL DAN PEMBAHASAN

a. Data analisis

Data yang disajikan mencakup tiga variabel utama, yaitu jumlah tanaman yang menghasilkan (pohon), produksi (ton), dan produktivitas (kuintal / pohon) yang terdiri atas 19 Kabupaten/Kota di Provinsi Sumatera Barat. Data analisis ini dapat dilihat pada **Gambar 1**.



Gambar 1. Data Analisis

Berdasarkan **Gambar 1** terdapat variasi yang signifikan antar Kabupaten/Kota di Provinsi Sumatera Barat dengan jumlah tanaman yang menghasilkan paling sedikit terdapat di Solok sebanyak 19 pohon, sedangkan tertinggi berada di Padang dengan jumlah mencapai 3.081 pohon. Produksi terendah dicatat di Solok sebesar 13,65 ton, sementara produksi tertinggi mencapai 16.506,4 ton di Padang. Produktivitas tanaman juga beragam, mulai dari 0,18 kuintal per pohon di Sijunjung hingga 5,6 kuintal per pohon di Solok Selatan. Perbedaan yang cukup besar antar kabupaten dan kota ini menunjukkan perlunya perhatian khusus dalam perencanaan pengembangan dan pengelolaan sumber daya pertanian di masing-masing wilayah.

DOI: https://doi.org/10.24036/ujsds/vol3-iss3/403

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

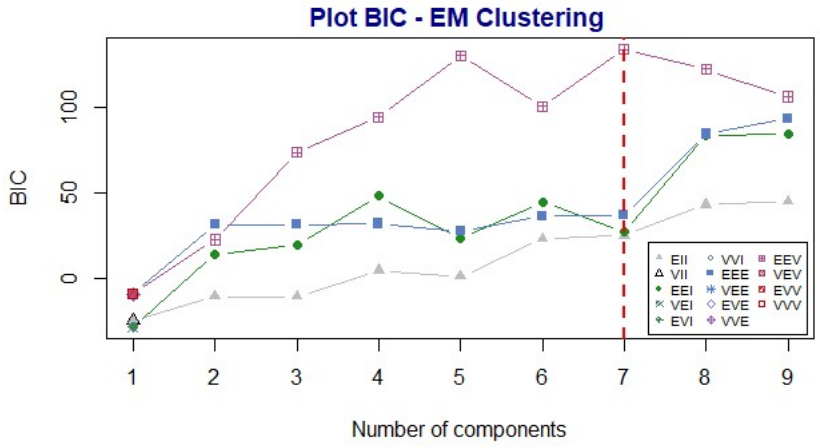
b. *Algoritma Expectation-Maximization* (EM)

Algoritma EM digunakan untuk mengelompokkan kabupaten/kota berdasarkan data produksi dan produktivitas. Proses penerapan metode ini diawali dengan menentukan nilai *log-likelihood*, dearajat kebebasan (df) dan nilai *Bayesian Information Criterion* (BIC) yang hasinya disajikan pada **Tabel 2.**

Tabel 2. *Algoritma Expectation-Maximization*

log-likelihood	n	df	BIC
142,065	19	51	133,9636

Berdasarkan Tabel 3. Terdapat nilai *log-likelihood* sebesar 142,065 dan *Bayesian Information Criterion* (BIC) sebesar 133,9636 menunjukkan bahwa model yang digunakan memiliki kemampuan yang baik dalam menjelaskan data. Hasil *Algoritma* EM Tabel 3 digunakan untuk menentukan jumlah *cluster* pada algoritma ini. Penentuan jumlah cluster dilakuan dengan membuat plot berdasarkan hasil yang telah diperoleh. Plot BIC- EM cluster dapat dilihat pada **Gambar 2.**



Gambar 2. Plot BIC – EM Cluster

Berdasarkan **Gambar 2** dapat diketahui bahwa kabupaten/kota terbagi menjadi tujuh cluster dengan karakteristik yang serupa. Pengelompokkan ini dapat mempermudah identifikasi wilayah berdasarkan potensi produksi tanaman sukun. yang membantu dalam pengelompokan wilayah berdasarkan potensi produksi sukun. *Cluster* ini dapat digunakan sebagai dasar pengambilan keputusan dalam pengembangan pertanian wilayah. Hasil dari penentuan cluster dapat digunakan untuk mengelompokkan Kabupaten/Kota ke dalam masing-masing *cluster* . Hasil cluster tersebut dapat disajikan pada **Tabel 3.**

Tabel 3. Hasil EM *Cluster* Menurut Kabupaten/Kota

CLuster	Kabupaten/Kota
1	Kepulauan Mentawai, Lima Puluh Kota, Solok, Padang Panjang, Pariaman
2	Pesisir Selatan, Sijunjung, Agam
3	Solok, Pasaman,Dharmasraya, Pasaman Barat
4	Tanah Datar,Bukittinggi
5	Padang Pariaman, Sawahlunto, Payakumbuh
6	Solok Selatan
7	Padang

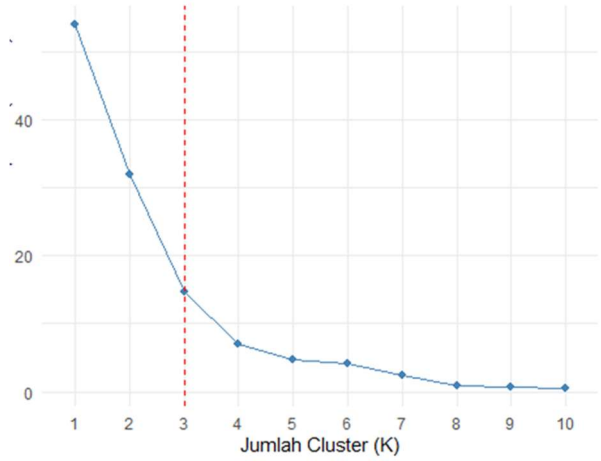
DOI: https://doi.org/10.24036/ujsds/vol3-iss3/403

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Berdasarkan **Tabel 3** menunjukkan hasil pengelompokan kabupaten/kota di Provinsi Sumatera Barat berdasarkan metode *Expectation-Maximization* (EM) *clustering* menjadi tujuh *cluster*. Setiap *cluster* mengelompokkan kabupaten/kota yang memiliki karakteristik produksi sukun dan tantangan budidaya yang serupa. *Cluster* 1 terdiri dari Kepulauan Mentawai, Lima Puluh Kota, Solok, Padang Panjang, dan Pariaman, yang kemungkinan memiliki pola produksi dan kendala yang sejenis. *Cluster* 2 mengelompokkan Pesisir Selatan, Sijunjung, dan Agam, sedangkan *Cluster* 3 mencakup Solok, Pasaman, Dharmasraya, dan Pasaman Barat. *Cluster* 4 terdiri dari Tanah Datar dan Bukittinggi, sedangkan *Cluster* 5 meliputi Padang Pariaman, Sawahlunto, dan Payakumbuh. *Cluster* 6 hanya terdiri dari Solok Selatan, menunjukkan karakteristik yang cukup berbeda dari kabupaten lainnya. Terakhir, *Cluster* 7 hanya berisi Padang, yang menunjukkan bahwa karakteristik produksi sukun di Padang cukup unik dan berbeda dari kabupaten/kota lain di provinsi tersebut.

c. K- means

Metode *K-means* berkerja dengan cara membagi data kedalam sejumlah cluster yang telah ditentukan sebelumnya, dengan tujuan meminimalkan jarak antar objek dalam satu *cluster* sekaligus memaksimalkan jarak antar *cluster*. Dalam penelitan ini, metode *k-means* digunakan untuk mengelompokkn Kabupaten/Kota di Provinsi Sumatera Barat dengan menggunakan metode elbow untuk menentukan jumlah cluster. Hasil dari penentuan jumlah cluster tersebut disajikan pada **Gambar 3.**



Gambar 3. Penentuan jumlah cluster mengunakan metode *elbow*

Berdasarkan **Gambar 3,** hasil penentuan jumlah *cluster* menggunakan metode *elbow* menunjukkan bahwa kabupaten/kota di Provinsi Sumatera Barat dapat dikelompokkan menjadi tiga *cluster*. Pengelompokkan tiga *cluster* ini dianggap paling representative dalam menggambarkan pola produksi sukun di wilayah tersebut. Hasil dari penentuan cluster ini dapat digunakan untuk mengelompokkan Kabupaten/Kota ke dalam masing-masing *cluster* . Hasil *cluster* tersebut dapat disajikan pada **Tabel 4.**

Tabel 4. Hasil *Cluster K-Means* Menurut Kabupaten/Kota

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

Berdasarkan **Tabel 4**, hasil *clustering* dengan metode *K-Means* membagi kabupaten/kota di Provinsi Sumatera Barat menjadi tiga cluster utama. *Cluster* 1 hanya terdiri dari Kabupaten Padang, yang menunjukkan karakteristik produksi sukun yang unik dan berbeda dibandingkan dengan wilayah lainnya. *Cluster* 2 mengelompokkan kabupaten Pesisir Selatan, Solok, Sijunjung, Agam, Pasaman, Dharmasraya, dan Pasaman Barat, yang diduga memiliki pola

DOI: https://doi.org/10.24036/ujsds/vol3-iss3/403

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

produksi dan tantangan budidaya yang serupa. Sedangkan *Cluster* 3 merupakan kelompok terbesar yang meliputi Kepulauan Mentawai, Tanah Datar, Padang Pariaman, Lima Puluh Kota, Solok Selatan, Solok, Sawahlunto, Padang Panjang, Bukittinggi, Payakumbuh, dan Pariaman. Cluster ini menunjukkan wilayah dengan karakteristik produksi sukun yang relatif homogen dan berbeda dari dua cluster sebelumnya.

d. Koefisien *silhouette*

Koefisien *silhouette* digunakan untuk mengukur kualitas hasil clustering dengan menilai seberapa rapat dan terpisahnya klaster yang terbentuk. Nilai *silhouette* yang lebih tinggi menunjukkan bahwa objek-objek dalam suatu klaster memiliki kemiripan yang tinggi antar sesama anggota *cluster* dan berbeda secara signifikan dengan *cluster* lain. Nilai Koefisien *silhouette* dapat disajikan dalam **Tabel 5** sebagai berikut :

Tabel 5. Nilai Koefisien silhouette

Expectation-Maximization (EM)	K-Means
0,47	0,37

Berdasarkan **Tabel 5** nilai *silhouette* EM sebesar 0,47 lebih tinggi dibandingkan *K-Means* yang sebesar 0,37, Hal ini mengindikasikan bahwa pengelompokan dengan EM menghasilkan *cluster* yang lebih rapat dan terpisah dengan baik, sehingga lebih representatif untuk data produksi sukun di Provinsi Sumatera Barat.

III. KESIMPULAN

Penentuan jumlah cluster optimal untuk pengelompokan data produksi sukun menurut Kabupaten/Kota di Provinsi Sumatera Barat dilakukan dengan metode elbow, yang menghasilkan tiga *cluster* untuk metode *K-Means* dan tujuh *cluster* untuk metode *Expectation-Maximization* (EM). Perbedaan jumlah *cluster* ini menunjukkan bahwa EM mampu menangkap variasi data secara lebih rinci dibandingkan *K-Means*.

Evaluasi kualitas *clustering* menggunakan nilai koefisien *silhouette* menunjukkan bahwa metode EM memperoleh nilai 0,47, lebih tinggi dibandingkan nilai 0,37 pada *K-Means*. Hal ini mengindikasikan bahwa cluster hasil EM lebih rapat dan terpisah dengan baik, sehingga pengelompokan menggunakan EM lebih representatif dan akurat dalam menggambarkan pola produksi sukun di wilayah tersebut. Oleh karena itu, metode EM lebih direkomendasikan untuk analisis data produksi sukun di Provinsi Sumatera Barat yang dapat mendukung pengambilan keputusan strategis.

DAFTAR PUSTAKA

Aselnino, P., & Wijayanto, A. W. (2024). Analisis Perbandingan Metode Hierarchical dan Non-Hierarchical dalam Pembentukan Cluster Provinsi di Indonesia Berdasarkan Indikator Women Empowerment. *Indonesian Journal of Applied Statistics*, *6*(1), 57.

Atika Nurani Ambarwati. (2019). LATENT CLASS CLUSTER ANALYSIS UNTUK PENGELOMPOKAN Latent Class Cluster Analysis for Grouping of Districts / Cities in Central Java province Based on Human Development Index Indicators 2017. *Variance*, *1*(2), 46–54.

BPS, P. S. B. (2024). Produksi Tanaman Hortikultura Provinsi Sumatera Barat 2023. *Produksi Tanaman Hortikultura Provinsi Sumatera Barat 2023*, *37*, 7–10.

Ha, J., Kambe, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*.

Kementerian Kesehatan. (2017). Food Composition Table—Indonesia (Daftar Komposisi Bahan Makanan). In *Tabel Komposisi Pangan Indonesia*.

Mardiani, M. (2015). Perbandingan Algoritma K-Means dan EM untuk Clusterisasi Nilai Mahasiswa Berdasarkan Asal Sekolah. *Creative Information Technology Journal*, *1*(4), 316. https://doi.org/10.24076/citec.2014v1i4.31

Noviasari, S., Rahma, Y. H., Nilda, C., & Safriani, N. (2023). PELUANG DAN POTENSI SUKUN (Artocarpus altilis) SEBAGAI INGREDIENT PANGAN. *Jurnal Ilmiah Mahasiswa Pertanian*, *8*(1), 221–229.

Rahmattullah, R., Indwiarti, I., & Rohmawati, A. A. (2023). Clustering Harga Rumah: Perbandingan Model K-

DOI: https://doi.org/10.24036/ujsds/vol3-iss3/403

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Means dan Gaussian Mixture Model. *E-Proceeding Of Engineering*, *10*(3), 3441–3449. https://openlibrary.telkomuniversity.ac.id/pustaka/files/185889/jurnal_eproc/clustering-harga-rumah-perbandingan-model-k-means-dan-gaussian-mixture-model.pdf

Sombamori Janggat, A., Nengah Kencana Putra, I., Made Sugitha, I., Studi Teknologi Pangan, P., Teknologi Pertanian, F., Udayana Kampus Bukit, U., & korespondensi, P. (2022). Itepa: Jurnal Ilmu dan Teknologi Pangan, Pengaruh Perbandingan Tepung Sukun (Artocarpus altilis) dan Tepung Kacang Merah (Phaseolus vulgaris L.) Terhadap Karakteristik Stik. *Alberto Sombamori Janggat, Dkk. /Itepa*, *11*(2), 2022–2177.

DOI: https://doi.org/10.24036/ujsds/vol3-iss3/403