

ANALISIS CLUSTERING PRESTASI PEMANEN KELAPA SAWIT MENGUNAKAN METODE HARD C-MEANS PADA PTPN IV PALMCO KEBUN SAWIT HULU

CLUSTERING ANALYSIS OF OIL PALM HARVESTER PERFORMANCE USING THE HARD C-MEANS METHOD AT PTPN IV PALMCO SAWIT HULU ESTATE

Putri Serianti¹, Juanda Nargaza², Rizki Julia Utama³, Jansen Fernando Bangun⁴

¹Program Studi Informatika, Fakultas Sains & Teknologi, Universitas Ubudiyah Indonesia

²Program Studi Sistem Informasi, Fakultas Sains & Teknologi, Universitas Ubudiyah Indonesia

Corresponding Author: Putriserianti@uui.ac.id

Abstrak— Evaluasi prestasi pemanen kelapa sawit di PTPN IV PalmCo Kebun Sawit Hulu selama ini belum didukung oleh pengelompokan yang objektif, sehingga keputusan pembinaan dan pemberian penghargaan belum berbasis data. Penelitian ini bertujuan mengelompokkan prestasi pemanen menggunakan metode Hard C-Means (HCM) dan mengevaluasi kualitas pengelompokan yang dihasilkan. Data yang digunakan adalah rerata capaian tahun 2024–2025 dari 268 pemanen pada 13 afdeling dengan tiga variabel, yaitu Tandan Buah Segar (TBS), brondolan, dan hari kerja (HK). Data dinormalisasi dengan Min–Max, kemudian dikelompokkan ke dalam tiga cluster menggunakan jarak Euclidean. Kualitas cluster diukur menggunakan Silhouette Coefficient dan Davies–Bouldin Index (DBI). Algoritma konvergen pada iterasi ke-9 dan menghasilkan cluster prestasi tinggi sebanyak 57 pemanen (21,3%), prestasi sedang 139 pemanen (51,9%), dan prestasi rendah 72 pemanen (26,9%). Nilai DBI sebesar 0,771 merupakan yang terendah untuk $c = 2-7$, sedangkan nilai Silhouette sebesar 0,429 menunjukkan struktur cluster yang tergolong lemah–moderat; meskipun demikian, partisi yang sama diperoleh dari 100 inisialisasi acak dan dari normalisasi z-score. Analisis lanjutan menunjukkan bahwa produktivitas harian cluster rendah dan sedang tidak berbeda signifikan (971 dan 982 kg/HK; $p = 0,848$), sehingga rendahnya prestasi lebih disebabkan oleh jumlah hari kerja yang kecil. Hasil ini dapat dijadikan dasar evaluasi kinerja, pembinaan, dan pemberian penghargaan pemanen.

Kata kunci: *Hard C-Means, clustering, prestasi pemanen, kelapa sawit, Davies–Bouldin Index, Python.*

Abstract— The evaluation of oil palm harvester performance at PTPN IV PalmCo Sawit Hulu Estate has not been supported by an objective grouping, so decisions on coaching and rewards are not yet data-driven. This study aims to cluster harvester performance using the Hard C-Means (HCM) method and to evaluate the quality of the resulting clusters. The data consist of the 2024–2025 average achievements of 268 harvesters across 13 divisions (afdeling), described by three variables: Fresh Fruit Bunches (FFB), loose fruits, and working days. The data were normalized using Min–Max scaling and then grouped into three clusters based on the Euclidean distance. Cluster quality was measured using the Silhouette Coefficient and the Davies–Bouldin Index (DBI). The algorithm converged at the ninth iteration, producing a high-performance cluster of 57 harvesters (21.3%), a moderate-performance cluster of 139 harvesters (51.9%), and a low-performance cluster of 72 harvesters (26.9%). The DBI value of 0.771 was the lowest for $c = 2-7$, while the Silhouette value of 0.429 indicates a weak-to-moderate cluster structure. Nevertheless, the same partition was obtained from 100 random initializations and from z-score normalization. Further analysis showed that the daily productivity of the low- and moderate-performance clusters did not differ significantly (971 and 982 kg per working day; $p = 0.848$), indicating that low performance is mainly attributable to fewer working days. These results can serve as a basis for performance evaluation, coaching, and reward decisions for harvesters.

Keywords: *Hard C-Means, clustering, harvester performance, oil palm, Davies–Bouldin Index, Python.*

I. PENDAHULUAN

Indonesia merupakan produsen dan eksportir minyak kelapa sawit (*Crude Palm Oil/CPO*) terbesar di dunia. Badan Pusat Statistik mencatat luas areal perkebunan kelapa sawit nasional pada tahun 2024 sebesar 16,01 juta hektare dengan produksi sekitar 47,47 juta ton [1], sedangkan Kementerian Pertanian melaporkan produksi CPO tahun 2025 sebesar 46,55 juta ton [2]; perbedaan angka antarsumber tersebut dipengaruhi oleh perbedaan cakupan dan metode pendataan. Besarnya kontribusi komoditas ini menjadikan produktivitas kebun sebagai isu strategis, baik bagi perusahaan swasta maupun badan usaha milik negara. Salah satu pengelola terbesar adalah PalmCo, entitas hasil konsolidasi sejumlah

PTPN yang berfokus pada pengelolaan komoditas kelapa sawit.

Kebun Sawit Hulu merupakan salah satu unit PTPN IV PalmCo yang berlokasi di Kecamatan Sawit Seberang, Kabupaten Langkat, Sumatera Utara, dengan luas sekitar 9.671 hektare yang terbagi ke dalam 13 afdeling. Pada kebun ini, capaian panen setiap pemanen telah direkap menggunakan Microsoft Excel, tetapi belum terdapat mekanisme pengelompokan prestasi yang objektif. Akibatnya, manajemen belum memiliki dasar kuantitatif yang konsisten untuk memberikan penghargaan kepada pemanen berprestasi, melakukan pembinaan kepada pemanen dengan capaian rendah, maupun merencanakan strategi peningkatan produktivitas.

Data mining menyediakan pendekatan untuk mengekstraksi pola bernilai dari data berukuran besar dengan menggabungkan teknik statistik, matematika, dan *machine learning* [3]. Pendekatan ini berperan dalam mengubah data operasional yang kompleks menjadi informasi yang mendukung pengambilan keputusan [4], [5]. Salah satu teknik yang paling banyak digunakan adalah *clustering*, yaitu pengelompokan objek berdasarkan kemiripan karakteristiknya tanpa memerlukan label kelas sebelumnya.

Sejumlah penelitian telah menerapkan *clustering* pada sektor kelapa sawit. K-Means digunakan untuk segmentasi perkebunan berdasarkan luas areal, produksi, dan produktivitas [6], untuk mengelompokkan produktivitas blok panen [7], kualitas minyak kelapa sawit [8], daerah produksi [9], [10], serta kualitas produk kelapa sawit [13]. Fuzzy C-Means (FCM) diterapkan untuk mengelompokkan hasil perkebunan nasional [11], dan perbandingan K-Means dengan FCM juga telah dilakukan pada data produksi hortikultura [12]. Ula dkk. [14] menunjukkan bahwa C-Means menghasilkan proses *clustering* yang lebih baik dengan jumlah iterasi lebih sedikit dibandingkan K-Means pada data produksi kelapa sawit di Aceh Utara. Namun, penelitian-penelitian tersebut umumnya berfokus pada unit wilayah, lahan, atau produk, bukan pada kinerja individu tenaga kerja panen, dan sebagian besar belum melaporkan validasi implementasi maupun ukuran kualitas cluster yang dihasilkan.

Berdasarkan celah tersebut, penelitian ini menerapkan metode Hard C-Means (HCM) untuk mengelompokkan prestasi pemanen kelapa sawit di PTPN IV PalmCo Kebun Sawit Hulu. HCM dipilih karena setiap pemanen harus ditempatkan secara tegas pada satu kategori prestasi sehingga hasilnya mudah dioperasionalkan dalam kebijakan sumber daya manusia. Kontribusi penelitian ini adalah: (1) penerapan HCM pada data kinerja individu 268 pemanen menggunakan tiga variabel, yaitu TBS, brondolan, dan hari kerja; (2) validasi silang antara perhitungan manual pada Microsoft Excel dan implementasi Python, dilengkapi evaluasi kualitas cluster menggunakan *Silhouette Coefficient*, *Davies–Bouldin Index* (DBI), dan uji stabilitas inisialisasi; serta (3) interpretasi karakteristik cluster dan peringkat prestasi pemanen sebagai dasar evaluasi kinerja oleh manajemen kebun.

II. STUDI PUSTAKA

A. Clustering dan Konsep Keanggotaan

Clustering merupakan teknik *unsupervised learning* yang bertujuan meminimalkan jarak antardata dalam satu cluster dan memaksimalkan jarak antarcluster [15]. Berdasarkan mekanisme keanggotaannya, *clustering* dibedakan menjadi *hard clustering* dan *fuzzy (soft) clustering*. Pada *hard clustering*, setiap data hanya menjadi anggota satu cluster dengan derajat keanggotaan bernilai 0 atau 1, sedangkan pada *fuzzy clustering* setiap data dapat memiliki derajat keanggotaan antara 0 dan 1 terhadap beberapa cluster [16], [17]. Hard C-Means merupakan representasi *hard clustering*, sedangkan Fuzzy C-Means merepresentasikan *fuzzy clustering* [16], [18]. Keanggotaan tegas pada HCM sesuai untuk kasus penilaian kinerja karena setiap pegawai perlu ditetapkan ke satu kategori keputusan.

B. Algoritma Hard C-Means

HCM mempartisi n data ke dalam c cluster dengan meminimalkan fungsi objektif jumlah kuadrat jarak setiap data terhadap pusat clusternya [16]. Sebelum proses *clustering*, setiap variabel dinormalisasi menggunakan metode Min–Max agar variabel dengan rentang nilai besar, seperti TBS, tidak mendominasi perhitungan jarak:

$$x'_{kj} = \frac{x_{kj} - x_{j,\min}}{x_{j,\max} - x_{j,\min}} \quad (1)$$

dengan x_{kj} adalah nilai variabel ke- j pada data ke- k . Fungsi objektif HCM dinyatakan sebagai berikut:

$$J = \sum_{k=1}^n \sum_{i=1}^c u_{ik} \|x'_k - v_i\|^2 \quad (2)$$

dengan $u_{ik} \in \{0, 1\}$ adalah keanggotaan data ke- k pada cluster ke- i dan v_i adalah centroid cluster ke- i . Jarak setiap data terhadap centroid dihitung menggunakan jarak Euclidean pada m variabel:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x'_{kj} - v_{ij})^2} \quad (3)$$

Setiap data kemudian dialokasikan secara tegas ke cluster dengan jarak terkecil:

$$u_{ik} = \begin{cases} 1, & \text{jika } d_{ik} = \min_{1 \leq l \leq c} d_{lk} \\ 0, & \text{lainnya} \end{cases} \quad (4)$$

Setelah seluruh data dialokasikan, centroid diperbarui sebagai rata-rata anggota cluster:

$$v_{ij} = \frac{\sum_{k=1}^n u_{ik} x'_{kj}}{\sum_{k=1}^n u_{ik}} \quad (5)$$

Langkah (3)–(5) diulang hingga perubahan centroid antariterasi tidak melebihi toleransi ε . Pada penelitian ini digunakan $\varepsilon = 0$, yaitu iterasi dihentikan ketika centroid tidak lagi berubah:

$$\max_i \|v_i^{(t)} - v_i^{(t-1)}\| \leq \varepsilon \quad (6)$$

C. Evaluasi Kualitas Cluster

Kualitas hasil *clustering* dinilai menggunakan indeks validitas internal [19]. *Silhouette Coefficient* mengukur seberapa sesuai suatu data dengan clusternya dibandingkan dengan cluster terdekat lainnya [20]:

$$s(k) = \frac{b(k) - a(k)}{\max\{a(k), b(k)\}} \quad (7)$$

dengan $a(k)$ adalah rata-rata jarak data ke- k terhadap anggota cluster yang sama dan $b(k)$ adalah rata-rata jarak terkecil terhadap anggota cluster lain. Nilai s berada pada rentang -1 sampai 1 ; semakin besar nilainya, semakin baik pemisahan cluster. *Davies–Bouldin Index* mengukur rasio sebaran di dalam cluster terhadap jarak antarcentroid [21]:

$$DBI = \frac{1}{c} \sum_{i=1}^c \max_{j \neq i} \frac{\sigma_i + \sigma_j}{d(v_i, v_j)} \quad (8)$$

dengan σ_i adalah rata-rata jarak anggota terhadap centroid cluster ke- i . Semakin kecil nilai DBI, semakin kompak dan terpisah cluster yang terbentuk. Kesamaan dua partisi, misalnya antara hasil dua inisialisasi yang berbeda, diukur menggunakan *Adjusted Rand Index* (ARI) dengan nilai 1 menunjukkan partisi yang identik [22].

D. Python untuk Analisis Data

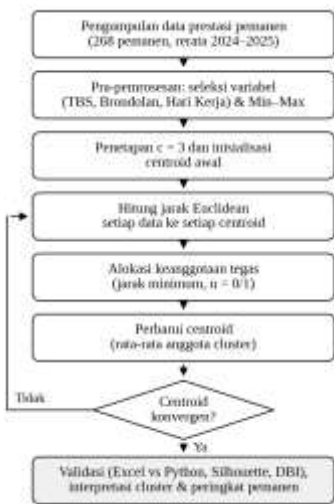
Python merupakan bahasa pemrograman tingkat tinggi dengan sintaks yang mudah dibaca dan ekosistem pustaka

yang luas [23]. Pustaka NumPy menyediakan komputasi array numerik yang efisien [24], sedangkan scikit-learn menyediakan implementasi algoritma *machine learning* dan metrik evaluasi *clustering* [25]. Pada penelitian ini Python digunakan untuk mengimplementasikan HCM, memverifikasi hasil perhitungan manual, menghitung indeks validitas, serta menyajikan visualisasi hasil.

III. METODE

A. Tahapan Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan tahapan seperti ditunjukkan pada Gambar 1. Tahapan dimulai dari pengumpulan data, pra-pemrosesan, penerapan HCM secara iteratif hingga konvergen, kemudian validasi, interpretasi cluster, dan penyusunan peringkat prestasi pemanen.



Gambar 1. Tahapan penelitian

B. Lokasi, Waktu, dan Data Penelitian

Penelitian dilaksanakan di PTPN IV PalmCo Kebun Sawit Hulu, Desa Sawit Hulu dan Desa Alur Melati, Kecamatan Sawit Seberang, Kabupaten Langkat, Sumatera Utara, pada Maret–Juli 2026. Data diperoleh melalui observasi partisipatif dan studi dokumentasi atas rekap capaian produksi pemanen per afdeling. Data berupa rerata capaian tahun 2024 dan 2025 dari 268 pemanen pada 13 afdeling dengan tiga variabel, yaitu TBS (kg), brondolan (kg), dan hari kerja (HK). Statistik deskriptif data disajikan pada Tabel 1.

Tabel 1. Statistik deskriptif data pemanen (n = 268)

Variabel	Min	Maks	Rerata	Std. dev.
TBS (kg)	48.348	485.338	213.444	75.451
Brondolan (kg)	3.745	37.407	14.375	6.254
Hari kerja (HK)	41	267	198,1	40,0

Tabel 1 menunjukkan bahwa rentang nilai ketiga variabel sangat berbeda; TBS bernilai ratusan ribu kilogram, sedangkan HK hanya ratusan hari. Kondisi ini menegaskan perlunya normalisasi sebelum perhitungan jarak.

C. Penerapan Hard C-Means

Seluruh data dinormalisasi ke rentang [0, 1] menggunakan (1). Jumlah cluster ditetapkan $c = 3$ sesuai kebutuhan manajemen untuk membedakan prestasi tinggi, sedang, dan rendah; pemilihan ini selanjutnya diuji secara empiris

menggunakan metode *Elbow*, *Silhouette*, dan DBI untuk $c = 2$ sampai 7. Centroid awal dipilih secara acak dari data ternormalisasi, yaitu data ke-31, ke-120, dan ke-233. Iterasi dilakukan menggunakan (3)–(5) hingga memenuhi kriteria (6).

D. Validasi dan Evaluasi

Validasi dilakukan dalam empat bentuk. Pertama, hasil perhitungan manual pada Microsoft Excel dibandingkan dengan keluaran implementasi Python untuk setiap data. Kedua, kualitas cluster diukur menggunakan *Silhouette* (7) dan DBI (8). Ketiga, stabilitas solusi diuji dengan menjalankan HCM menggunakan 100 kombinasi centroid awal acak dan membandingkan hasilnya dengan partisi utama menggunakan ARI. Keempat, perbedaan karakteristik antarcluster diuji menggunakan uji Kruskal–Wallis, dan perbandingan dua cluster menggunakan uji Mann–Whitney pada taraf signifikansi 5%.

IV. HASIL DAN PEMBAHASAN

A. Normalisasi dan Inisialisasi Centroid

Normalisasi Min–Max menghasilkan nilai setiap variabel pada rentang [0, 1]. Sebagai contoh, data ke-1 dengan TBS 258.615 kg, brondolan 18.811 kg, dan 221 HK memiliki nilai ternormalisasi (0,4812; 0,4476; 0,7965). Centroid awal yang digunakan ditunjukkan pada Tabel 2.

Tabel 2. Centroid awal (data ternormalisasi)

Cluster	Data ke-	TBS	Brondolan	HK
C1	31	0,1330	0,1004	0,4513
C2	120	0,3870	0,2176	0,8142
C3	233	1,0000	0,8385	0,8319

B. Perhitungan Jarak dan Alokasi Keanggotaan

Jarak setiap data terhadap ketiga centroid dihitung menggunakan (3). Untuk data ke-1, jarak terhadap C1 adalah $\sqrt{[(0,4812 - 0,1330)^2 + (0,4476 - 0,1004)^2 + (0,7965 - 0,4513)^2]} = 0,6007$. Dengan cara yang sama diperoleh jarak ke C2 sebesar 0,2491 dan ke C3 sebesar 0,6506, sehingga data ke-1 dialokasikan ke C2 sesuai (4). Contoh hasil perhitungan pada iterasi pertama disajikan pada Tabel 3.

Tabel 3. Contoh jarak Euclidean dan alokasi cluster pada iterasi pertama

Data ke-	d_{1k}	d_{2k}	d_{3k}	Cluster
1	0,6007	0,2491	0,6506	C2
2	0,9686	0,6486	0,3284	C3
3	0,3263	0,2098	0,8985	C2

C. Proses Iterasi dan Konvergensi

Tabel 4 merangkum jumlah anggota setiap cluster dan perubahan maksimum koordinat centroid pada setiap iterasi. Perubahan terbesar terjadi pada iterasi pertama (0,2982) karena centroid awal C3 berada pada data ekstrem, yaitu pemanen dengan TBS tertinggi. Perubahan kemudian menurun tajam dan keanggotaan stabil sejak iterasi ke-8. Pada iterasi ke-9 centroid tidak lagi berubah sehingga algoritma dinyatakan konvergen.

Tabel 4. Jumlah anggota cluster dan perubahan centroid per iterasi

Iterasi	C1	C2	C3	Δ centroid maks.
1	71	173	24	0,2982
2	76	152	40	0,0722
3	76	148	44	0,0133
4	76	145	47	0,0097
5	74	144	50	0,0102

Iterasi	C1	C2	C3	Δ centroid maks.
6	73	142	53	0,0121
7	73	139	56	0,0086
8	72	139	57	0,0037
9	72	139	57	0,0000

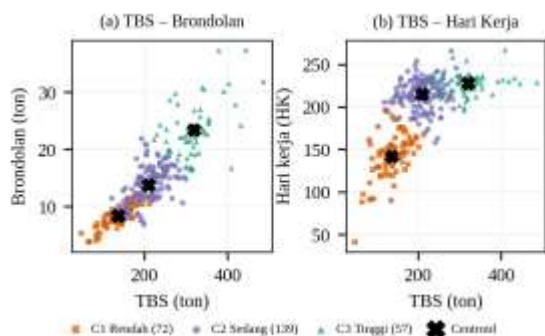
D. Karakteristik Cluster

Centroid akhir dan karakteristik setiap cluster ditunjukkan pada Tabel 5. Label prestasi ditetapkan berdasarkan urutan nilai centroid: C3 memiliki nilai tertinggi pada ketiga variabel sehingga dikategorikan sebagai prestasi tinggi, C2 sebagai prestasi sedang, dan C1 sebagai prestasi rendah. Perlu dicatat bahwa nomor cluster pada HCM hanya merupakan penanda urutan inisialisasi dan tidak menunjukkan peringkat, sehingga interpretasi harus didasarkan pada nilai centroid.

Tabel 5. Centroid akhir dan karakteristik cluster

Keterangan	C1	C2	C3
Kategori prestasi	Rendah	Sedang	Tinggi
Jumlah pemanen	72 (26,9%)	139 (51,9%)	57 (21,3%)
Centroid TBS (norm.)	0,2018	0,3690	0,6215
Centroid brondolan (norm.)	0,1376	0,2979	0,5845
Centroid HK (norm.)	0,4470	0,7693	0,8268
Rerata TBS (kg)	136.525	209.616	319.942
Rerata brondolan (kg)	8.376	13.773	23.420
Rerata HK	142,0	214,9	227,9
Rerata TBS/HK (kg)	970,7	981,9	1.407,3

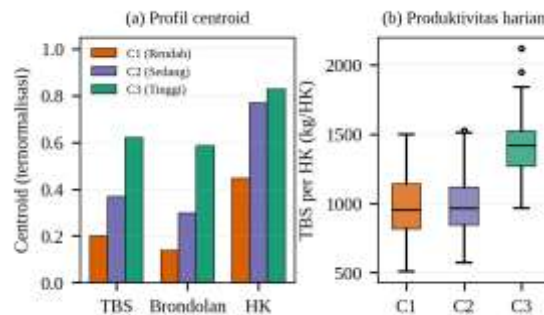
Cluster prestasi tinggi (C3) beranggotakan 57 pemanen dengan rerata TBS 319.942 kg, sekitar 2,3 kali lipat rerata cluster prestasi rendah. Cluster prestasi sedang (C2) merupakan kelompok terbesar dengan 139 pemanen (51,9%), sedangkan cluster prestasi rendah (C1) beranggotakan 72 pemanen. Uji Kruskal–Wallis menunjukkan perbedaan yang signifikan antarcluster pada TBS ($H = 188,0$), brondolan ($H = 185,3$), dan HK ($H = 166,0$), seluruhnya dengan $p < 0,001$. Karena ketiga variabel tersebut juga digunakan untuk membentuk cluster, hasil uji ini bersifat konfirmatif-deskriptif, yaitu menegaskan besarnya perbedaan antarcluster, dan tidak dimaksudkan sebagai bukti independen atas keberadaan kelompok. Sebaran data dan posisi centroid ditunjukkan pada Gambar 2.



Gambar 2. Sebaran hasil clustering: (a) TBS–brondolan, (b) TBS–hari kerja

Gambar 2(a) menunjukkan hubungan positif yang kuat antara TBS dan brondolan, sehingga ketiga cluster tersusun berurutan sepanjang diagonal. Gambar 2(b) memperlihatkan bahwa pemisahan C1 terhadap cluster lainnya terutama terjadi pada sumbu hari kerja, sedangkan pemisahan C2 dan C3 lebih ditentukan oleh besaran TBS. Temuan ini

dipertegas oleh profil centroid dan produktivitas harian pada Gambar 3.



Gambar 3. (a) Profil centroid ternormalisasi, (b) produktivitas harian (TBS per HK)

Gambar 3(a) menunjukkan bahwa centroid HK C1 (0,447) jauh lebih rendah dibandingkan C2 (0,769) dan C3 (0,827). Sebaliknya, Gambar 3(b) menunjukkan bahwa produktivitas harian C1 (970,7 kg/HK) hampir sama dengan C2 (981,9 kg/HK). Uji Mann–Whitney mengonfirmasi bahwa produktivitas harian C1 dan C2 tidak berbeda signifikan ($p = 0,848$), sedangkan C3 memiliki produktivitas harian yang jauh lebih tinggi daripada C2 (1.407,3 kg/HK; $p < 0,001$). Dengan demikian, rendahnya capaian tahunan C1 terutama disebabkan oleh jumlah hari kerja yang sedikit, bukan oleh rendahnya kemampuan memanen per hari. Sementara itu, keunggulan C3 bersumber dari produktivitas harian yang tinggi. Implikasinya, intervensi manajemen perlu dibedakan: pemanen C1 memerlukan penelusuran penyebab rendahnya kehadiran, seperti status kerja, kesehatan, atau penugasan, sedangkan peningkatan pemanen C2 menuju C3 lebih relevan melalui pelatihan teknik panen dan dukungan peralatan.

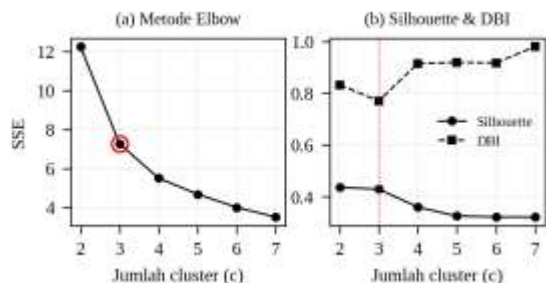
E. Validasi Implementasi dan Kualitas Cluster

Perbandingan hasil perhitungan manual pada Microsoft Excel dan implementasi Python menunjukkan bahwa seluruh 268 data memperoleh cluster yang sama (kesesuaian 100%). Centroid hasil iterasi pertama pada Excel, yaitu C1 (0,1993; 0,1367; 0,4454), C2 (0,4061; 0,3345; 0,7783), dan C3 (0,7018; 0,7105; 0,8320), identik dengan keluaran Python, dan kedua implementasi sama-sama konvergen pada iterasi ke-9. Hal ini menunjukkan bahwa perhitungan manual telah dilakukan dengan benar dan implementasi Python dapat digunakan untuk proses evaluasi berikutnya secara lebih cepat dan konsisten.

Uji stabilitas menggunakan 100 kombinasi centroid awal acak menghasilkan partisi yang identik dengan partisi utama pada seluruh percobaan ($ARI = 1,000$) dengan rata-rata 10,5 iterasi hingga konvergen. Solusi yang sama juga diperoleh menggunakan K-Means dengan inisialisasi k-means++ pada pustaka scikit-learn [25] serta ketika normalisasi Min–Max diganti dengan z-score ($ARI = 1,000$). Artinya, hasil pengelompokan tidak bergantung pada pemilihan centroid awal maupun metode normalisasi. Namun, normalisasi Min–Max sensitif terhadap nilai ekstrem; ketika satu data dengan HK sangat rendah (41 HK) dikeluarkan, kesesuaian partisi turun menjadi $ARI = 0,917$, sehingga penanganan pencilaan perlu diperhatikan pada penerapan berikutnya. Kesesuaian jumlah cluster diuji pada rentang $c = 2$ sampai 7 sebagaimana ditunjukkan pada Tabel 6 dan Gambar 4.

Tabel 6. Indeks validitas untuk jumlah cluster $c = 2-7$

c	SSE	Silhouette	DBI
2	12,243	0,436	0,833
3	7,262	0,429	0,771
4	5,498	0,359	0,915
5	4,678	0,325	0,920
6	4,000	0,322	0,918
7	3,500	0,321	0,981

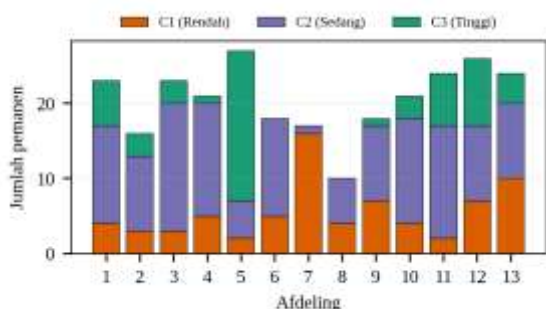


Gambar 4. (a) Kurva Elbow, (b) nilai Silhouette dan DBI terhadap jumlah cluster

Kurva *Elbow* pada Gambar 4(a) menunjukkan titik siku pada $c = 3$; penurunan SSE dari $c = 2$ ke $c = 3$ sebesar 40,7%, sedangkan dari $c = 3$ ke $c = 4$ hanya 24,3%. Nilai DBI terendah juga diperoleh pada $c = 3$ (0,771). Nilai *Silhouette* pada $c = 3$ (0,429) hampir sama dengan $c = 2$ (0,436) dan lebih tinggi dibandingkan $c \geq 4$. Nilai *Silhouette* per cluster adalah 0,504 (C1), 0,409 (C2), dan 0,382 (C3), dan hanya 2 dari 268 data yang memiliki nilai negatif. Berdasarkan kriteria Kaufman dan Rousseeuw [26], nilai *Silhouette* pada rentang 0,26–0,50 menunjukkan struktur cluster yang lemah, sedangkan struktur yang wajar (*reasonable*) memerlukan nilai di atas 0,50. Dengan demikian, cluster yang terbentuk lebih tepat dipandang sebagai segmentasi dari data kinerja yang bersifat kontinu, bukan kelompok yang terpisah secara alami; hal ini tampak pada Gambar 2, ketika batas antara C2 dan C3 tidak tajam. Keterbatasan ini tidak mengurangi kegunaan praktis hasil *clustering* sebagai dasar kategorisasi, karena partisi yang dihasilkan stabil, DBI-nya terendah, dan $c = 3$ sesuai dengan kebutuhan manajerial.

F. Sebaran Cluster per Afdeling

Sebaran anggota cluster pada setiap afdeling disajikan pada Gambar 5. Afdeling 5 mendominasi cluster prestasi tinggi dengan 20 dari 27 pemanen (74,1%) berada pada C3, disusul Afdeling 12 dengan 9 pemanen. Sebaliknya, 16 dari 17 pemanen Afdeling 7 (94,1%) berada pada C1, sementara Afdeling 6 dan Afdeling 8 tidak memiliki pemanen pada C3.



Gambar 5. Sebaran cluster prestasi pemanen pada setiap afdeling

Konsentrasi cluster pada afdeling tertentu mengindikasikan bahwa prestasi pemanen tidak hanya dipengaruhi faktor individu, tetapi juga faktor kebun seperti

umur dan komposisi tanaman, kerapatan buah, topografi, maupun kebijakan pengaturan hari kerja di tingkat afdeling. Oleh karena itu, hasil *clustering* sebaiknya digunakan bersama informasi kondisi afdeling agar evaluasi kinerja pemanen berlangsung secara adil.

G. Peringkat Prestasi Pemanen

Peringkat prestasi disusun dengan mengurutkan pemanen berdasarkan kategori cluster (C3, C2, kemudian C1) dan selanjutnya berdasarkan capaian TBS di dalam setiap cluster. Sepuluh pemanen dengan peringkat tertinggi ditunjukkan pada Tabel 7; identitas pemanen disajikan dalam bentuk kode untuk menjaga kerahasiaan data karyawan. Seluruh pemanen pada sepuluh besar berada pada C3, dan lima di antaranya berasal dari Afdeling 5.

Karena keanggotaan cluster ditentukan oleh ketiga variabel, terdapat 17 anggota C3 yang capaian TBS-nya lebih rendah daripada pemanen dengan TBS tertinggi di C2 (299.322 kg). Pemanen tersebut tetap diperingkat lebih tinggi karena capaian brondolan dan hari kerjanya lebih baik. Sebagai uji kepekaan, peringkat juga disusun menggunakan skor komposit, yaitu rata-rata nilai ternormalisasi ketiga variabel di dalam setiap cluster. Kedua cara pemeringkatan menghasilkan korelasi Spearman sebesar 0,962 dan 8 dari 10 pemanen teratas yang sama, sehingga peringkat yang dihasilkan cukup konsisten. Meskipun demikian, bobot setiap variabel dalam penilaian prestasi sebaiknya ditetapkan bersama manajemen kebun.

Tabel 7. Sepuluh pemanen dengan peringkat prestasi tertinggi

No.	Kode	Afd.	TBS (kg)	Brondolan (kg)	HK
1	P-233	12	485.338	31.969	229
2	P-084	5	444.080	37.407	228
3	P-238	12	431.127	24.293	234
4	P-060	3	418.749	27.855	230
5	P-175	10	409.069	16.696	267
6	P-228	12	402.928	31.529	232
7	P-095	5	391.162	30.184	230
8	P-085	5	378.430	37.314	215
9	P-105	5	354.928	25.047	237
10	P-090	5	354.893	25.442	234

Dibandingkan penelitian sebelumnya yang mengelompokkan wilayah, blok, atau produk kelapa sawit [6]–[10], [13], penelitian ini menunjukkan bahwa pengelompokan pada tingkat individu tenaga kerja dapat menghasilkan informasi yang lebih operasional bagi manajemen kebun. Selain itu, pelaporan validasi implementasi, indeks validitas, dan uji stabilitas memperkuat keandalan hasil dibandingkan penelitian yang hanya menyajikan hasil pengelompokan [7], [13].

V. KESIMPULAN

A. Kesimpulan

Metode Hard C-Means berhasil diterapkan untuk mengelompokkan prestasi 268 pemanen kelapa sawit PTPN IV PalmCo Kebun Sawit Hulu berdasarkan variabel TBS, brondolan, dan hari kerja. Algoritma konvergen pada iterasi ke-9 dan menghasilkan cluster prestasi tinggi sebanyak 57 pemanen (21,3%), prestasi sedang 139 pemanen (51,9%), dan prestasi rendah 72 pemanen (26,9%). Hasil perhitungan manual pada Microsoft Excel identik dengan implementasi Python pada seluruh data, dan partisi yang sama diperoleh dari 100 inisialisasi acak. Jumlah tiga cluster didukung oleh metode *Elbow* dan nilai DBI terendah (0,771), sedangkan

nilai *Silhouette* sebesar 0,429 menunjukkan struktur cluster yang lemah-moderat sehingga hasilnya lebih tepat dimaknai sebagai segmentasi kinerja. Analisis karakteristik menunjukkan bahwa rendahnya prestasi cluster C1 terutama disebabkan oleh jumlah hari kerja yang sedikit, sedangkan keunggulan cluster C3 bersumber dari produktivitas harian yang tinggi. Hasil *clustering* dan peringkat pemanen dapat digunakan sebagai dasar evaluasi kinerja, pembinaan, dan pemberian penghargaan.

B. Saran

Penelitian selanjutnya disarankan menggunakan data dengan periode pengamatan yang lebih panjang atau data bulanan untuk menangkap dinamika kinerja, menambahkan variabel yang memengaruhi prestasi seperti usia, masa kerja, tingkat kehadiran, umur tanaman, dan kondisi lahan, serta menormalisasi capaian terhadap potensi produksi setiap afdeling agar perbandingan antarpemanen lebih adil. Penanganan pencilan dan penetapan bobot variabel bersama manajemen juga perlu dilakukan untuk meningkatkan ketepatan kategorisasi dan peringkat. Metode HCM juga dapat dibandingkan dengan Fuzzy C-Means, DBSCAN, atau *clustering* hierarkis, dan hasilnya dapat dikembangkan menjadi sistem informasi evaluasi kinerja pemanen yang terintegrasi dengan data produksi kebun.

REFERENSI

- [1] Badan Pusat Statistik, *Statistik Tanaman Perkebunan Tahunan Indonesia 2024 (Kelapa Sawit, Kopi, Kakao, Karet, Teh, dan Komoditas Perkebunan Unggulan)*. Jakarta: Badan Pusat Statistik, 2025.
- [2] Kementerian Pertanian Republik Indonesia, "CPO Indonesia 2025 meningkat, produksi capai 46,55 juta ton," Kementerian Pertanian Republik Indonesia, 2025.
- [3] A. Srirahayu dan L. S. Priadi, "Review paper data mining klasifikasi data mining," *Jurnal Ilmiah Informatika Global*, vol. 14, no. 1, 2023, doi: 10.36982/jiig.v14i1.2981.
- [4] E. Saputri, "Teknik dan aplikasi data mining di Indonesia: Tinjauan literatur satu dekade (2015–2024)," *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 4, no. 2, pp. 138–149, 2025, doi: 10.24246/itexplore.v4i2.2025.pp138-149.
- [5] J. Astuti dan T. Yuniarti, "Data mining modeling in clustering car products sales data in the automotive industry in Indonesia," *Jurnal Manajemen Industri dan Logistik*, vol. 7, no. 2, pp. 261–281, 2023, doi: 10.30988/jmil.v7i2.1258.
- [6] T. Yuniarti dan D. Hayati, "Segmentasi perkebunan kelapa sawit dengan data mining teknik K-Means clustering berdasarkan luas areal, produksi dan produktivitas," *INVENTORY: Industrial Vocational E-Journal on Agroindustry*, vol. 2, no. 2, p. 56, 2021, doi: 10.52759/inventory.v2i2.47.
- [7] I. Ramadhani dan M. Megawati, "Implementasi algoritma K-Means untuk klustering data produktivitas kelapa sawit," *Indonesian Journal of Informatic Research and Software Engineering*, vol. 3, no. 1, pp. 56–64, 2023, doi: 10.57152/ijirse.v3i1.488.
- [8] A. Priadi, "Klasterisasi kualitas minyak kelapa sawit menggunakan metode K-Means clustering (studi kasus PT Tasma Puja)," *Jurnal Sains dan Ilmu Terapan*, vol. 5, no. 1, pp. 40–45, 2022, doi: 10.59061/jsit.v5i1.76.
- [9] F. Nasari, "Pengelompokan daerah produksi kelapa sawit menggunakan algoritma K-Means clustering," *Jurnal Sains dan Ilmu Terapan*, vol. 4, no. 2, pp. 34–38, 2021, doi: 10.59061/jsit.v4i2.108.
- [10] D. Haryadi, "Penerapan algoritma K-Means clustering pada produksi perkebunan kelapa sawit menurut provinsi," *Journal of Informatics and Communication Technology (JICT)*, vol. 3, no. 1, pp. 50–64, 2021, doi: 10.52661/j_ict.v3i1.71.
- [11] N. Aisyah, A. I. Sukarni, dan D. C. R. N. Sari, "Pengelompokan hasil perkebunan di Indonesia menggunakan Fuzzy C-Means," *CESS (Journal of Computer Engineering, System and Science)*, vol. 9, no. 2, p. 361, 2024, doi: 10.24114/cess.v9i2.50682.
- [12] D. Ariel dan T. Handhayani, "Perbandingan efektivitas algoritma K-Means dan Fuzzy C-Means untuk clustering data produksi alpukat di Indonesia," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 13, no. 2, 2025, doi: 10.24912/jiksi.v13i2.35133.
- [13] S. E. N. Putra dan M. Furqan, "Penerapan data mining dalam pengelompokan kualitas produk kelapa sawit menggunakan algoritma K-Means clustering," *CESS (Journal of Computer Engineering, System and Science)*, vol. 9, no. 2, p. 766, 2024, doi: 10.24114/cess.v9i2.61682.
- [14] M. Ula, G. Perdinanta, R. Hidayat, dan I. Sahputra, "Analyze the clustering and predicting results of palm oil production in Aceh Utara," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 2, 2023, doi: 10.22146/ijccs.83195.
- [15] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010, doi: 10.1016/j.patrec.2009.09.011.
- [16] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981, doi: 10.1007/978-1-4757-0450-1.
- [17] D. Krasnov, D. Davis, K. Malott, Y. Chen, X. Shi, dan A. Wong, "Fuzzy C-Means clustering: A review of applications in breast cancer detection," *Entropy*, vol. 25, no. 7, p. 1021, 2023, doi: 10.3390/e25071021.
- [18] A. D. Kochuveetil dan R. Mathew, "A novel approach to fuzzy c-Means clustering using kernel function," *Intelligent Decision Technologies*, vol. 16, no. 4, pp. 643–651, 2022, doi: 10.3233/IDT-210091.
- [19] H.-Y. Wang, J.-S. Wang, dan G. Wang, "A survey of fuzzy clustering validity evaluation methods," *Information Sciences*, vol. 618, pp. 270–297, 2022, doi: 10.1016/j.ins.2022.11.010.
- [20] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [21] D. L. Davies dan D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.
- [22] L. Hubert dan P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985, doi: 10.1007/BF01908075.
- [23] E. Matthes, *Python Crash Course: A Hands-On, Project-Based Introduction to Programming*, 3rd ed. San Francisco: No Starch Press, 2023.
- [24] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020, doi: 10.1038/s41586-020-2649-2.
- [25] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [26] L. Kaufman dan P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons, 1990.