

Penerapan Metode *K-Means Clustering* untuk Pemetaan Pengelompokan Lahan Produksi Tandan Buah Segar

Abdussalam Al Masykur¹, Siska Kurnia Gusti², Suwanto Sanjaya³, Febi Yanto⁴, Fadhilah Syafria⁵

^{1,2,3,4,5} Fakultas Sains dan Teknologi UIN Sultan Syarif Kasim Riau/Teknik Informatika
Jl. HR. Soebrantas No. 155 Simpang Baru Panam Pekanbaru 28293, Indonesia

e-mail: ¹11750115166@students.uin-suska.ac.id, ²siskakurniagusti@uin-suska.ac.id,
³suwantosanjaya@uin-suska.ac.id, ⁴febiyanto@uin-suska.ac.id, ⁵fadhilah.syafria@uin-suska.ac.id

Informasi Artikel Diterima: 16-03-2023 Direvisi: 05-04-2023 Disetujui: 15-04-2023

Abstrak

Kelapa sawit merupakan tanaman pertanian penting yang dibudidayakan secara luas di Indonesia, menyediakan sumber pendapatan dan lapangan kerja yang signifikan bagi masyarakat setempat. Salah satunya di perkebunan Sei Lukut, Desa Maredan Barat, Kecamatan Tualang, Kabupaten Siak, Provinsi Riau, PT Surya Intisari Raya, yang merupakan perusahaan swasta mengelola perkebunan kelapa sawit. Memiliki 4 bagian lahan kelapa sawit yang terdiri dari 216 blok dengan total sekitar 4.000 Ha. Pemetaan perkebunan kelapa sawit yang akurat sangat penting untuk pengelolaan lahan, pembangunan berkelanjutan dan membantu perusahaan dalam memutuskan kebijakan apa yang akan diterapkan untuk meningkatkan akurasi dan produktivitas produksi kelapa sawit. Dalam penelitian ini penulis menggunakan metode *K-Means Clustering* untuk memetakan perkebunan kelapa sawit pada PT Surya Intisari Raya yang menerapkan *Davies Bouldin Indeks* dengan alat *RapidMiner* sebagai evaluasi jumlah *cluster* yang optimal. Hasil penelitian menunjukkan bahwa metode *K-means clustering* efektif dalam pemetaan lahan produksi tandan buah segar dengan tingkat kecocokan sistem dengan *RapidMiner* 99,54% dengan nilai *Davies Bouldin Index* terkecil sebesar 0,921 dengan jumlah *cluster* 3 yang terdiri dari *Cluster C1* (Produktivitas Sedang) 96 blok lahan, *Cluster C2* (Produktivitas Rendah) 41 blok lahan, dan *Cluster C3* (Produktivitas Tinggi) 79 blok lahan. Penelitian ini menunjukkan potensi metode *K-Means clustering* untuk pemetaan lahan produksi tandan buah segar yang dapat memberikan informasi berharga bagi perusahaan.

Kata Kunci: *K-Means Clustering*; Lahan Produksi Tandan Buah Segar; Pengelompokan

Abstract

Oil palm is an important agricultural crop that is widely cultivated in Indonesia, providing a significant source of income and employment for local people. One of them is in the Sei Lukut plantation, Maredan Barat Village, Tualang District, Siak Regency, Riau Province, PT Surya Intisari Raya, which is a private company that manages oil palm plantations. Has 4 parts of oil palm land consisting of 216 blocks with a total of around 4,000 Ha. Accurate mapping of oil palm plantations is essential for land management, sustainable development, and assisting companies in deciding what policies to implement to increase the accuracy and productivity of palm oil production. In this study, the authors used the K-Means Clustering method for oil palm plantations at PT Surya Intisari Raya which applied the Davies Bouldin Index with the RapidMiner tool as an evaluation of the optimal number of clusters. The results showed that the K-means clustering method was effective in the recession of fresh fruit bunch production areas with a system compatibility rate with RapidMiner of 99.54% with the smallest Davies Bouldin Index value of 0.921 with the number of clusters 3 consisting of C1 Cluster (Medium Productivity) 96 blocks land, Cluster C2 (Low Productivity) 41 land blocks, and Cluster C3 (High Productivity) 79 land blocks. This research shows the potential of the K-Means clustering method to address fresh fruit bunch production areas which can provide valuable information for companies.

Keywords: *K-Means Clustering, Fresh Fruit Bunches Production Area, Grouping*



1. Pendahuluan

Indonesia merupakan negara agraris yang diberi berkah kekayaan alam yang berlimpah baik sumber daya alam hayati maupun non hayati (Erlangga et al., 2019). Indonesia juga merupakan salah satu negara yang mempunyai banyak sekali perkebunan kelapa sawit (Ismail, 2017). Menurut (Hendra Effendi, Ahmad Syahrial, Sefran Prayoga, 2021) baik pemerintah maupun swasta di Indonesia mengelola perkebunan kelapa sawit. Salah satu bisnis tersebut adalah PT. Surya Intisari Raya yang memproduksi minyak sawit. PT. Surya Intisari Raya didirikan pada tahun 1992 dan memiliki perkebunan di Kebun Sei Lukut, Desa Maredan Barat, Kecamatan Tualang, Kabupaten Siak, Provinsi Riau. Perkebunan tersebut terdiri dari 216 blok dengan total luas sekitar 4.000 Ha dan 4 bagian properti kelapa sawit. Dengan 28.000 pohon sawit dan luas rata-rata 20 hektar, setiap blok dapat menghasilkan rata-rata 57 ton minyak sawit per bulan. Berdasarkan wawancara dengan salah satu narasumber terkait, PT Surya Intisari Raya memiliki proses evaluasi lahan yang dilakukan dengan melihat data tingkat pencapaian produksi per bulan berdasarkan target yang telah ditentukan, dimana data tersebut memiliki kelemahan pada saat visualisasi yang masih menggunakan data numerik sehingga tidak dapat optimal untuk pencatatan dan pemantauan produksi tandan buah segar yang diperoleh perusahaan. Dengan tidak adanya pemetaan blok produksi buah kelapa sawit maka sistem pemetaan pengelompokan hasil tandan buah segar kelapa sawit *K-Means Clustering* ini dibuat untuk membantu perusahaan dalam menentukan kebijakan yang diambil sehingga diharapkan dapat meningkatkan akurasi dan produktifitas hasil produksi kelapa sawit bagi perusahaan.

Dengan studi literatur penelitian pertama tentang klasifikasi blok perkebunan kelapa sawit menjadi dua kelompok, yaitu blok berbuah dan blok tidak produktif, dengan menggunakan empat variabel (TBS, produksi/ha, jumlah tandan, dan bobot tandan). menggunakan *Micosoft Exel* dan *Rapidminer* (Pulungan et al., 2019). Kemudian pada penelitian lain membagi area perkebunan kelapa sawit menjadi dua kelompok, tinggi dan rendah, menggunakan algoritma clustering k-means dengan perangkat *Microsoft Exel* dan *Rapidminer* (Pasaribu et al., 2021). Terlepas dari kenyataan bahwa keduanya menggunakan algoritma *k-means clustering*.

Pengambilan *K-Means Clustering* digunakan berdasarkan analisis perbandingan antara *K-Means* dan *Fuzzy C-Means (FCM)* yang dilakukan oleh Suomi G. dan Sanjay Kumar D., didapatkan hasil yang membuktikan

bahwa algoritme *K-Means* lebih cepat dengan elapsed time 0.433755 detik dibandingkan dengan algoritme *FCM* yang memiliki elapsed time sebesar 0.781679 detik (Jipkate & Gohokar, 2012).

Penelitian ini berbeda dengan beberapa penelitian yang disebutkan di atas karena fokusnya pada PT perkebunan kelapa sawit Surya Intisari Raya, dan faktor-faktornya berjumlah delapan, (luas, total pokok, janjang panen, brondolan, curah hujan, pupuk, target, persentase capaian) yang mengelompokkan produksi tandan buah segar kedalam 3 cluster (produksi tinggi, sedang dan rendah). Pemetaan disini merupakan pengelompokkan suatu kumpulan wilayah yang berkaitan dengan beberapa letak geografis wilayah yang meliputi dataran tinggi, pegunungan, sumber daya alam, serta penduduk yang berpengaruh terhadap sosial dan kebudayaan yang memiliki ciri khas khusus dalam penggunaan skala yang tepat (Kharistiani, 2013).

Evaluasi *clustering Davies Bouldin Index* digunakan untuk mengevaluasi hasil *clustering* guna menentukan jumlah kelompok yang paling menguntungkan, dengan tujuan untuk mengetahui seberapa tinggi kualitas hasil *clustering* (Muningsih et al., 2021). David L. Davies dan Donald W. Bouldin menciptakan *Davies Bouldin Index (DBI)* pada tahun 1979. Dalam teknik pengelompokan dimana kohesi digambarkan sebagai total kedekatan data ke titik pusat cluster dari cluster, satu metode digunakan untuk menilai validitas atau jumlah cluster yang paling efektif adalah *Indeks Davies-Bouldin* (Bates & Kalita, 2016). Kuantitas dan kedekatan data hasil klaster digunakan dalam metode penilaian *Indeks Davies Bouldin* untuk menentukan apakah klaster menghasilkan hasil yang sangat baik atau buruk. Sebuah sistem yang dapat mengkategorikan lahan kelapa sawit menurut tingkat produktivitas menggunakan teknik *k-means clustering*, penilaian model menggunakan *Davies Boundin Index*, dan mengevaluasi akurasi sistem dibandingkan dengan aplikasi *RapidMiner* merupakan temuan dari penelitian ini.

2. Metode Penelitian

Alur tahapan penelitian dapat diuraikan sebagai berikut:

1) Identifikasi Masalah

Salah satu langkah pertama dalam memecahkan suatu masalah adalah mengidentifikasinya (Pasaribu et al., 2021). Tidak adanya pemetaan blok produksi buah kelapa sawit sebagai bagian dari upaya yang lebih besar untuk meningkatkan akurasi dan produktivitas estimasi hasil merupakan kendala utama tujuan studi ini.

2) Pengumpulan Data

Dengan melihat informasi produksi minyak sawit per bulan, statistik dikumpulkan. Untuk kumpulan statistik pertama, penulis menggunakan informasi dari PT. Surya Intisari Raya Sei Lukut mengenai hasil produksi minyak sawit bulan Agustus 2021.

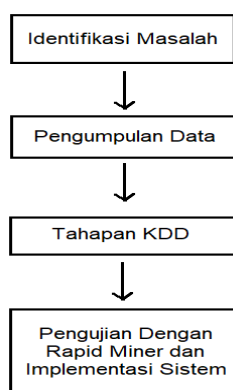
3) Tahapan KDD

Langkah pertama dalam proses ini adalah pemilihan data, atau memilih data dari sekelompok data, dan yang kedua adalah pembersihan data agar dapat menjadi bahan penemuan pengetahuan dalam basis data. Kemudian tahap *Transformation* tidak dilakukan karena jenis data sudah sama. Ketiga dilakukanlah proses *Data Mining* dengan metode *k-means*. Selanjutnya yang terakhir adalah proses *Interpretation Data* berupa variabel penting pendukung dan pemetaan yang bertujuan agar dapat dimengerti oleh pihak yang berkepentingan (Yuli Mardi, 2019).

4) Pengujian Dengan RapidMiner dan Implementasi Sistem

Data yang sudah melewati tahapan KDD kemudian dilakukan pengujian dengan *RapidMiner*. Kemudian di Implementasikan pada sistem yang akan di buat. Setelah itu dilakukan perbandingan kecocokan antara sistem dengan *Rapidminer*.

Dari tahapan penelitian tersebut dapat ditunjukkan pada gambar 1:



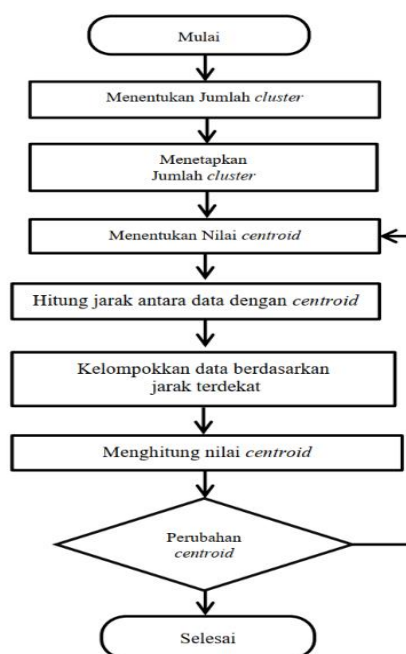
Gambar 1. Tahapan Penelitian

2.1 Data Mining

Data Mining adalah proses mengumpulkan banyak data untuk mempelajari sesuatu dan kemudian mengubahnya menjadi informasi yang bermanfaat (Adiya & Desnelita, 2019). *Clustering* tujuan dari metode penambangan data, yang seringkali mencakup analisis data, adalah untuk mengumpulkan data dengan kualitas yang sebanding. *Cluster* tidak digunakan untuk mengkategorikan, mengukur, atau meramalkan nilai kuantitas objektif (Erlangga et al., 2019). Untuk mempermudah

dalam memahami alur komputasi sebagai alur untuk mengetahui hasil penerapan *cluster* pada data yang akan diproses, maka diperlukan *flowchart* saat mencari *cluster* berdasarkan data yang saat ini dapat diakses (Maulida, 2018).

Pengumpulan data dibagi menjadi K kategori menggunakan algoritma *k-means* (lin Parlina, Agus Perdana Windarto, Anjar Wanto, 2018).



Sumber: (Pasaribu et al., 2021)

Gambar 2. Flowchart Metode K-Means

Penulis menggunakan algoritme *k-means* untuk mengkategorikan hasil buah sawit yang produktif, dan langkah-langkah berikut dilakukan untuk menyelesaikan proses tersebut (Pasaribu et al., 2021):

- 1) Memutuskan berapa banyak data yang perlu dicluster, dan memilih hasil panen yang paling banyak untuk sampel data keluaran yang akan digunakan dalam proses *clustering*.
- 2) Pilih 3 kelompok untuk jumlah K dari jumlah *cluster* ($k-3$).
- 3) Menetapkan *centroid* (pusat *cluster*) asli yang dihitung secara acak menggunakan nilai faktor data dalam *cluster* serta nilai K yang telah ditetapkan sebelumnya.
- 4) Tentukan pemisahan antara setiap titik data yang berguna dan pusat *cluster*.
- 5) Berdasarkan jarak data terpendek ke pusat *cluster*, tentukan lokasi *cluster* dari setiap data keluaran produktif.
- 6) Menghitung nilai *centroid* hasil iterasi dengan memeriksa kondisi penghentian proses iterasi, jika nilai *centroid* hasil iterasi sama dengan nilai *centroid* sebelumnya atau nilai

centroid optimal dan posisi *cluster* sawit produktif data produksi tidak berubah lagi, maka proses iterasi berhenti. Jika setelah satu putaran iterasi nilai *centroid* berbeda atau tidak optimum dan posisi data keluaran buah sawit produktif masih berfluktuasi, maka proses akan dilanjutkan pada putaran iterasi berikutnya.

2.2 RapidMiner

Pada tahap ini dilakukan *modelling* dan evaluasi menggunakan tools *RapidMiner* dengan metode K-Means.

2.3 Implementasi Sistem

Implementasi sistem disini adalah *Knowledge Discovery in Databases* (KDD) mencakup langkah-langkah yang terlibat dalam penguraian data untuk membuat skema sistem.

3. Hasil dan Pembahasan

3.1 Pengolahan Data

Setelah dilakukan seleksi dan pembersihan data, penulis memperoleh informasi produksi minyak sawit langsung dari PT Surya Intisari Raya Sei Lukut pada bulan Agustus sebagai berikut:

Tabel 1. Data Hasil Produksi Periode Agustus 2021

No	Blok	Luas (ha)	Total Pokok	Janjang Panen (kg)	Brondolan (kg)	Cukuran	Pupuk	Target	Persentase Pencapaian %
1	A17	13,6	191,0	181,4	166,0	22,0	105,0	218,3	90,70
2	A18	18,9	276,4	292,8	311,0	5,0	24,0	10,3	77,41
3	A19	18,2	253,6	188,7	159,0	22,0	129,0	393,1	51,97
4	A20	19,2	286,3	381,0	423,0	22,0	114,0	334,9	126,67
5	A21	24,1	348,3	447,8	436,0	22,0	148,0	452,3	108,54
6	A22	16,0	223,6	325,9	340,0	22,0	123,0	273,0	131,43
7	A23	21,5	302,0	278,0	271,0	22,0	128,0	413,2	73,83
8	A23	5,30	689,0	877,846	470,00	22,0	103,400	892,224	103,66
9	A24	13,3	178,1	255,18,2	244,00	22,0	267,200	297,04,74	94,12
10	A24	1,80	226,0	210,9,81	90,00	22,0	339,00	335,2,03	65,63
11	J16	10,4	112,8	111,59,1	137,00	19,0	282,00	129,38,2	96,84
12	J17	12,0	129,3	108,24,0	116,00	19,0	323,60	878,2,15	136,46

Langkah-langkah perhitungan metode *k-means clustering*:

- Menentukan Jumlah *Cluster* dan *Centroid*.
Menggunakan *RapidMiner*, algoritma *k-means* akan digunakan untuk menerapkan data ke dalam langkah pengelompokan (Pasaribu et al., 2021). Informasi tersebut kemudian akan dimasukkan ke dalam *RapidMiner*. Jika pengelompokan yang dimaksud adalah 3, algoritma *k-means* kemudian diterapkan, dan nilai titik tengah atau pusat dihasilkan dari data yang diperoleh. Penentuan *cluster* dibagi menjadi tiga bagian, yaitu C1, C2, dan C3, dan dijelaskan pada Tabel 7. Pemilihan *centroid* awal merupakan salah satu faktor yang dapat mempengaruhi hasil *cluster* (Febrita & Amaniyah, 2022). Menggunakan *Rapidminer*, tentukan lokasi *cluster* ini. Kita harus mengambil nilai tergantung pada atribut untuk mendapatkan nilai untuk setiap *cluster*. Sehingga nilai C1, C2, dan C3 dapat dilihat pada tabel terlampir.

Tabel 2. Nilai *Centroid* Awal

Cluster	Blok	Luas (ha)	Total Pokok	Janjang Panen (kg)	Brondolan (kg)	Cukuran	Pupuk	Target	Persentase Pencapaian %
C1	D17	22,07	22,0	18,04	159,00	22,00	5,78	52,91	54,68
C2	C25	14,85	14,0	43,63	111,00	22,00	5,62	76,62	47,50
C3	H22	25,12	25,0	50,64	347,00	11,00	13,7	81,3,0	106,05

- Menemukan jumlah minimum di antara tiga *centroid* dengan menghitung jarak dari *centroid*, dari mana kita memperoleh tabel jarak. Berikut ini adalah grafik jarak dari pusat *cluster*.

Tabel 3. Nilai Iterasi ke-1

No	Blok	C1	C2	C3	Cluster
1	A17	15011,34	10623,65	39685,49	C2
2	A18	16179,35	20546,02	19520,78	C1
3	A19	8491,18	11227,36	30155,86	C1
4	A20	21340,59	25140,69	17704,96	C3
..
214	J15	32290,24	26545,80	61150,19	C2
215	J16	24183,60	18914,73	51996,58	C2
216	J17	28204,75	22869,93	55044,54	C2

- Menentukan *Cluster* atau Pengelompokan
Pada iterasi 1, *cluster* yang cocok dengan nilai yang sama dipilih ketika menemukan *cluster* dengan nilai *cluster* berdasarkan nilai klaster yang sama. Selain itu, jika grup pada

iterasi berikutnya cocok dengan grup pada iterasi sebelumnya, perhitungan diakhiri dengan menggunakan teknik *k-means*. Konsekuensinya, sampai nomor iterasinya sama, cari grup pada iterasi berikutnya. Menjumlahkan nilai yang disebutkan dalam klaster pada bagan di atas akan membantu Anda menentukan nilai *centroid* berikutnya menggunakan *centroid* baru dari proses pertama. Pusat baru untuk menemukan klaster berikut diperoleh dengan menjumlahkan nilai-nilai *cluster* yang dipilih bersama-sama, kemudian mendistribusikan hasilnya ke seluruh jumlah nilai. Dan dari iterasi pertama, nilai *centroid* baru diperoleh (Pasaribu et al., 2021). Komputasi *cluster* kemudian ditangguhkan sampai nomor *cluster* iterasi tetap konstan dan tidak berubah.

Tabel 4. Centroid Baru

C 1	21,6 7	282 6,26	3454 9,90	3371 ,45	196 ,59	9325 ,51	3558 1,97	115, 65
C 2	8,04	102 5,17	1009 1,19	926, 09	188 ,36	2758 ,24	1267 9,95	101, 02
C 3	24,6 5	329 2,95	5044 5,33	4798 ,73	208 ,98	1633 8,08	5382 2,07	108, 43

Dari table 4 diatas adalah *centroid* baru setelah beberapa kali iterasi dilakukan dan berhenti pada iterasi ke-8.

Tabel 5. Nilai Iterasi ke-8

No	Blok	C1	C2	C3	Cluster
1	A17	21452,11	14562,95	45884,86	C2
2	A18	10083,17	37136,49	24445,66	C1
3	A19	16.628,10	29959,09	35026,20	C1
4	A20	4768,84	36181,17	24293,30	C1
5	A21	15137,77	49248,85	10414,44	C3
6	A22	9003,29	28608,28	32254,31	C1
7	A23	9555,31	35264,33	26175,08	C1
8	A23A	38166,01	4374,82	63339,38	C2
9	A24	12741,42	23037,10	37384,89	C1
10	A24A	46791,22	12566,05	71910,52	C2
..
..
214	J15	42460,80	81970,00	67556,45	C2
215	J16	33302,83	1191,12	58430,23	C2
216	J17	36.407,69	4010,25	61542,41	C2

Perhitungan *cluster* dihentikan karena iterasi ke-7 dan ke-8 sama atau data tidak berubah dan hasil yang diperoleh yaitu:

Cluster 1 (C1) sebanyak 96 data blok sawit, *Cluster 2* (C2) sebanyak 41 data blok sawit, *Cluster 3* (C3) sebanyak 79 data blok sawit.

Setelah *cluster* C1, C2, dan C3 didapatkan tahap selanjutnya adalah menentukan kategori produksi tinggi, sedang, dan rendah untuk masing masing blok. Untuk mendapatkannya digunakan 2 variabel penting pendukung yaitu target dan persentase pencapaian dari 8 banyak variabel yang digunakan.

Tabel 6. Penentuan *Cluster* Berdasarkan Variabel Penting Pendukung

Cluster	Blok	Target	Persentase Pencapaian%
C1	D7	35912,58	54,68
C2	C25	31029,89	47,50
C3	H22	48194,43	106,05

Kemudian disini ada proses *Interpretation data* yang merupakan bagian dari proses *Knowledge Discovery in Database (KDD)* untuk mendapatkan kategori *cluster* produksi tinggi, sedang dan rendah. Dari 2 variabel yaitu target dan persentase pencapaian didapat:

C3 dengan data target 48194.44, persentase pencapaian 106.05 mempunyai data yang paling tinggi diantara ketiga data, sehingga C3 dapat dikategorikan menjadi produksi tinggi.

C1 dengan data target 35912.58, persentase pencapaian 54.68 mempunyai data tidak tinggi dan tidak rendah diantara ketiga data, sehingga C1 dapat dikategorikan menjadi produksi sedang.

C2 dengan data target 31029.89, persentase pencapaian 47,50 mempunyai data yang paling rendah diantara ketiga data, sehingga C2 dapat dikategorikan menjadi produksi rendah.

3.2 Pengujian *RapidMiner*

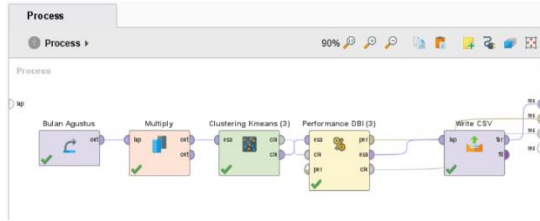
Teknik *k-means RapidMiner* digunakan untuk menganalisis data, dan jumlah grup dioptimalkan dengan mencari nilai *DBI* terkecil. Sebuah teknik pengukuran yang disebut *Indeks Davies Bouldin* memaksimalkan jarak dalam sebuah *cluster* sekaligus berusaha meminimalkan jarak antar lokasi dalam sebuah cluster (Tukiyat & Djohan, 2022). Tabel 7 di bawah mencantumkan jumlah grup dan nilai *DBI*-nya:

Tabel 7. Nilai *DBI* Tiap *Cluster*

Cluster	Nilai <i>DBI</i>
3	0,921
4	0,928
5	1,078
6	1,000

Berdasarkan hasil analisis, terdapat 3 kelompok untuk optimasi, masing-masing dengan nilai *DBI* 0,921. Hasilnya, tiga

kelompok dipilih sebagai jumlah optimal untuk kumpulan pengelompokan data ini.



Gambar 3. Modeling Clustering Metode K-Means di RapidMiner

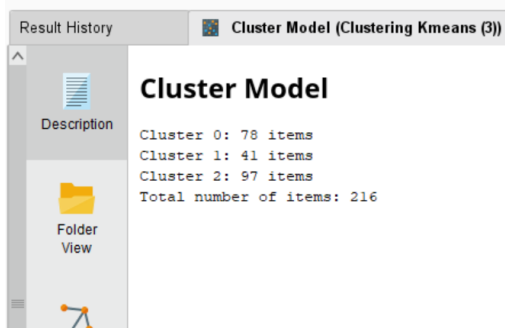
Pada Gambar 3 dilakukan modeling cluster dengan metode k-means pada tools RapidMiner dimana jumlah cluster yang dimodel adalah 3.

R.	cluster	Luas Blok	Total Pokok	Janjang Pa...	Bronddan...	Carah Rujan...	Pepuk	Target/Budget	Persentase...
1	A17	13.660	1913	18193.540	1650	225	10522	21890.130	90.896
2	A18	18.490	2768	29257.260	3110	225	15224	41810.310	77.415
3	A19	18.620	2538	18881.770	1590	225	12966	39388.410	51.974
4	A20	19.320	2862	38153.490	4230	225	11449	33458.890	126.673
5	A21	24.130	3488	44751.950	4300	225	14824	45247.430	108.541
6	A22	16.600	2239	32546.230	3490	225	12317	27351.000	131.426
7	A23	21.250	3020	27808	2710	225	12830	41334.520	73.832
8	A23A	5.300	889	8778.460	470	225	1034	8922.240	103.656
9	A24	13.320	1781	25518.220	2440	225	2672	28704.740	94.120
10	A24A	1.800	226	2109.810	90	225	339	3352.030	65.626
11	B12	2.650	320	2812.490	230	225	800	4118.880	73.867
12	B13	11.010	1412	7766.060	820	225	3530	16513.350	51.995
13	B14	13.200	1738	34790.870	3470	225	4346	16017.710	238.866

Sumber: Penulis (2023)

Gambar 4. Data Hasil Clustering Metode K-Means di RapidMiner

Pada gambar 4 adalah hasil saat tool run diklik pada RapidMiner dan memunculkan hasil cluster pada setiap blok.

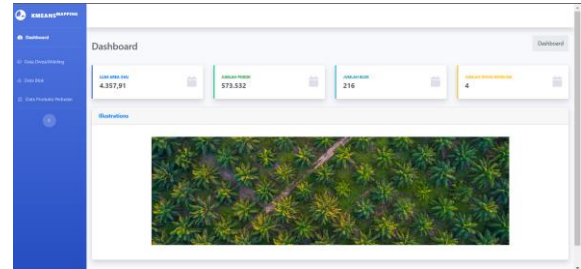


Gambar 5. Jumlah Data Hasil Clustering Metode K-Means di RapidMiner

Pada gambar 5 adalah jumlah data saat tool run diklik pada RapidMiner memunculkan hasil Cluster C0 (Produktifitas Sedang) terdiri atas 97 blok lahan, Cluster C1 (Produktifitas Rendah) terdiri atas 41 blok lahan, dan Cluster C2 (Produktifitas Tinggi) terdiri atas 78 blok lahan.

3.3 Implementasi Sistem Pemetaan Pengelompokan Lahan Produksi TBS Menggunakan K-Means

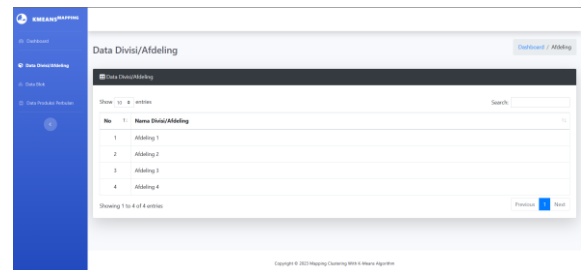
1) Halaman Dashboard



Gambar 6. Halaman Dashboard di Sistem Pemetaan Pengelompokan Lahan Produksi TBS Menggunakan K-Means

Tampilan utama program ditunjukkan pada Gambar 6. Pengguna memiliki akses ke situs web lengkap di halaman ini dan dapat melihat luas area, jumlah pokok, jumlah blok, dan jumlah afdeling.

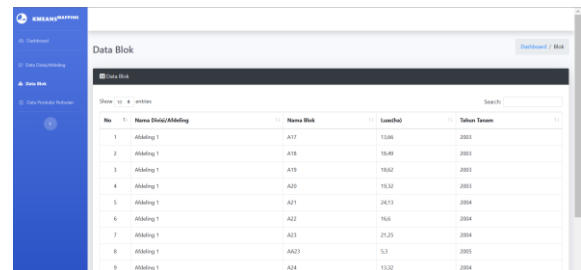
2) Halaman Data Afdeling



Gambar 7. Halaman Data Afdeling di Sistem Pemetaan Pengelompokan Lahan Produksi TBS Menggunakan K-Means

Gambar 7 adalah tampilan halaman data afdeling. Pengguna dapat melihat nama afdeling.

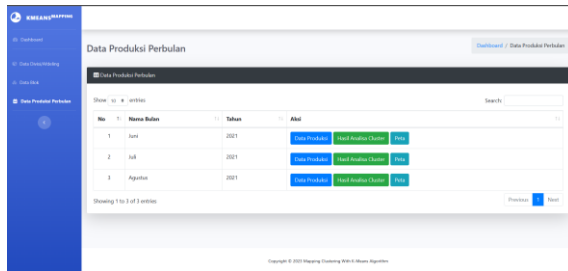
3) Halaman Data Blok



Gambar 8. Halaman Data Blok Afdeling di Sistem Pemetaan Pengelompokan Lahan Produksi TBS Menggunakan K-Means

Gambar 8 adalah tampilan halaman data blok. Pengguna dapat melihat nama afdeling, nama blok, luas, dan tahun tanam.

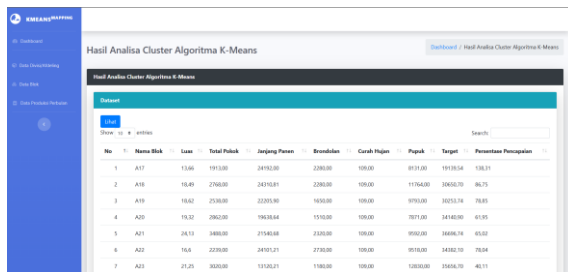
4) Halaman Data Produksi



Gambar 9. Halaman Data Produksi Afdeling Sistem Pemetaan Pengelompokan Lahan Produksi TBS

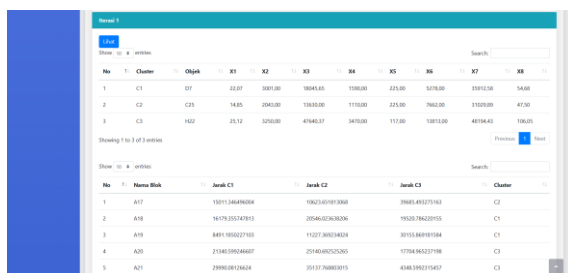
Tampilan lembar data manufaktur ditunjukkan pada Gambar 9. Pengguna dapat melihat statistik bulanan dan tahunan produksi minyak sawit di sistem ini. Kemudian pengguna dapat memilih tombol aksi untuk melihat data produksi, hasil analisa *cluster*, dan peta.

5) Halaman Hasil Analisa Cluster



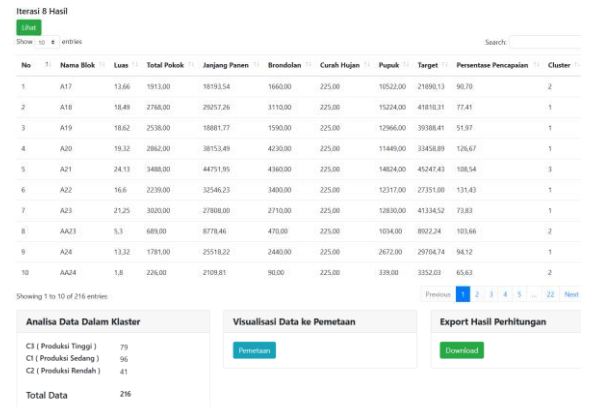
Gambar 10. Halaman Hasil Analisa Cluster di Sistem Pemetaan Pengelompokan Lahan Produksi TBS Menggunakan K-Means

Gambar 10 merupakan tampilan halaman analisa *cluster*. Pada halaman ini pengguna dapat melihat hasil analisa *cluster* setiap bulan yang dipilih dari halaman data produksi. Data tersebut berupa nama blok, luas, total pokok, janjang panen, brondolan, curah hujan, pupuk, target dan persentase pencapaian.



Gambar 11. Halaman Hasil Analisa Cluster bagian 2 di Sistem Pemetaan Pengelompokan Lahan Produksi TBS Menggunakan K-Means
 Pada Gambar 11 merupakan perhitungan *k-means* pada iterasi 1 yang

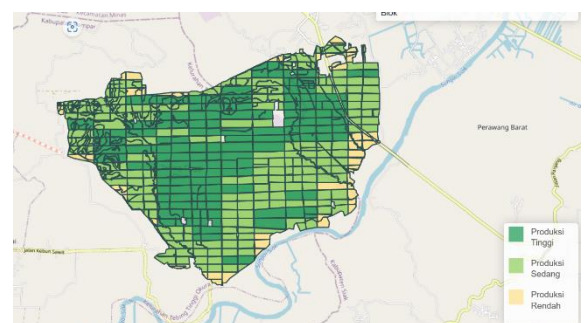
merupakan menentukan *centroid* awal yang kemudian dihitung jarak kedekatan datanya.



Gambar 12. Halaman Hasil Analisa Cluster bagian 3 di Sistem Pemetaan Pengelompokan Lahan Produksi TBS Menggunakan K-Means

Pada Gambar 12 merupakan hasil Analisa perhitungan K-Means dan berhenti pada iterasi 8 karena data sama dan tidak ada yang berubah. Setelah perhitungan *k-means* selesai maka didapatkan lah hasil analisa data *cluster* C1 (Produksi Sedang) sebanyak 96, C2 (Produksi Rendah) sebanyak 41 dan C3 (Produksi Tinggi) sebanyak 79 blok sawit. Kemudian pengguna dapat memilih tombol aksi pemetaan yang akan menampilkan peta, dan tombol aksi *download* untuk hasil perhitungan.

6) Halaman Peta Hasil Analisa Cluster



Sumber: Penulis (2023)

Gambar 13. Halaman Peta Hasil Analisa Cluster di Sistem Pemetaan Pengelompokan Lahan Produksi TBS Menggunakan K-Means

Gambar 13 merupakan tampilan halaman peta hasil analisa *cluster*. Pada halaman ini pengguna dapat melihat peta dari hasil analisa *cluster* setiap bulan yang dipilih dari halaman data produksi dan halaman analisa *cluster*. Pewarnaan disini menunjukkan perbedaan *cluster* pada setiap bloknya. Warna hijau tua adalah *cluster* dengan produksi tinggi, hijau

muda adalah produksi sedang, sedangkan jingga muda adalah produksi rendah. Dengan adanya pemetaan ini membuat inovasi baru dari setiap penelitian yang ada. Pemetaan disini mengimplementasikan data yang berasal dari numerik.

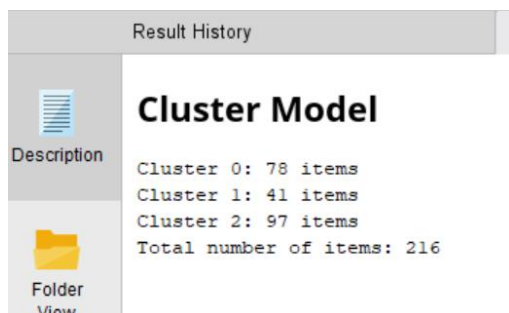
3.4 Perbandingan Kecocokan Sistem RapidMiner

Dilakukan dengan cara melakukan perbandingan kecocokan antara hasil perhitungan sistem dengan hasil perhitungan *clustering RapidMiner*.

Analisa Data Dalam Klaster	
C3 (Produksi Tinggi)	79
C1 (Produksi Sedang)	96
C2 (Produksi Rendah)	41
Total Data	216

Gambar 14. Hasil *Cluster* di Sistem Pemetaan Pengelompokan Lahan Produksi TBS Menggunakan *K-Means*

Sistem agregasi lahan sawit subur diklaster menggunakan teknik *k-means*, dan hasilnya ditunjukkan pada Gambar 14. Diketahui bahwa C1 (Produktivitas Sedang) terdiri dari 96 blok lahan, C2 (Produktivitas Rendah) terdiri dari 41 blok lahan, dan C3 (Produktivitas Tinggi) terdiri dari lebih dari 79 blok tanah.



Gambar 15. Hasil *Clustering* dengan *RapidMiner*

Dengan menggunakan *Rapidminer*, *Cluster* C0 (Produktivitas Sedang), C1 (Produktivitas Rendah), dan C2 (Produktivitas Tinggi) masing-masing terdiri dari 97, 41, dan 78 data.

Tabel 8. Perbandingan Kecocokan Hasil *Clustering*

No	Nama Blok	Hasil <i>Cluster</i> Sistem	Hasil <i>Cluster</i> <i>RapidMiner</i>
1	A17	C2	C2
2	A18	C1	C1
3	A19	C1	C1
4	A20	C1	C1
5	A21	C3	C3
6	A22	C1	C1
7	A23	C1	C1
8	A23A	C2	C2
9	A24	C1	C1
10	A24A	C2	C2
..
..
214	J15	C2	C2
215	J16	C2	C2
216	J17	C2	C2

Tabel 8 mencantumkan hasil perbandingan antara *RapidMiner* dan sistem. Berdasarkan hasil *clustering* yang sudah didapatkan, hasil perbandingan kecocokan pada sistem dan *Rapidminer* dapat dihitung dengan cara:

$$P = \frac{\text{Jumlah data yang benar disistem}}{\text{Jumlah keseluruhan data}} \times 100\%$$

$$P = \frac{215}{216} \times 100\% = 99,54\%$$

Sehingga dapat dinyatakan bahwa hasil perbandingan kecocokan sistem dengan aplikasi *RapidMiner* untuk pengelompokan lahan sawit produktif dengan menggunakan metode *k-means* adalah 99,54%.

4. Kesimpulan

Penerapan metode *K-Means* dengan penilaian nilai *DBI* 0,921 menghasilkan optimalisasi jumlah *cluster* terbaik yaitu 3. Kesimpulan ini dapat ditarik berdasarkan temuan penelitian yang dilakukan. Blok-blok kelapa sawit dikelompokkan berdasarkan hasil pengelompokan dengan menggunakan 8 variabel yaitu luas, total pokok, janjang panen, brondolan, curah hujan, pupuk, target, dan persentase pencapaian dengan kelompok *cluster* tersebut adalah produktivitas tinggi, produktivitas sedang, dan produktivitas rendah. Dari hasil pengujian dengan menggunakan data produksi periode Agustus 2021 diperoleh hasil dari *RapidMiner* dengan C0 (Produktivitas Sedang) terdiri dari 97 blok lahan, *cluster* C1 (Produktivitas Rendah) terdiri dari 41 blok lahan, dan *cluster* C2 (Produktivitas Tinggi) terdiri dari 78 blok lahan, dan hasil sistem dengan Cluster C1 (Produktivitas Sedang) terdiri dari 96 blok lahan, Cluster C2 (Produktivitas Rendah) terdiri dari 41 blok lahan, dan *cluster* C3 (Produktivitas

Tinggi) terdiri dari 79 blok lahan. Dengan menggunakan metode *k-means clustering*, sistem juga telah berhasil memetakan blok lahan kelapa sawit. Pewarnaan yang digunakan untuk menunjukkan berbagai *cluster* pada setiap blok adalah hijau tua untuk produksi tinggi, hijau muda untuk produksi sedang, dan jingga muda untuk produksi rendah. Dengan tingkat kecocokan 99,54% antara sistem dan *Rapidminer*, sistem ini juga dapat beroperasi dengan baik dan benar.

Referensi

- Adiya, M. H., & Desnelita, Y. (2019). *Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. 01*, 17–24.
- Bates, A., & Kalita, J. (2016). *Counting Clusters in Twitter Posts*.
<https://doi.org/10.1145/2905055.2905295>
- Erlangga, N., Solikhun, S., & Irawan, I. (2019). Penerapan Data Mining Dalam Mengelompokkan Produksi Jagung Menurut Provinsi Menggunakan Algoritma K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 702–709.
<https://doi.org/10.30865/komik.v3i1.1681>
- Febrita, R. E., & Amaniyah, M. (2022). Penentuan Kekekabatan Hewan Berdasarkan Struktur Protein IGF2 Menggunakan Metode K-Means dan N-Gram. *Jurnal Informatika*, 9(2), 140–147.
<https://doi.org/10.31294/inf.v9i2.13808>
- Hendra Effendi, Ahmad Syahrial, Sefran Prayoga, W. D. H. (2021). Penerapan Metode K-Means Clustering untuk Pengelompokan Lahan Sawit Produktif pada PT Kasih Agro Mandiri. *Teknomatika*, 11(02), 117–126.
- lin Parlina, Agus Perdana Windarto, Anjar Wanto, M. R. L. (2018). *Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Assessment Center Untuk Clustering Program Sdp*, 3(1), 87–93.
- Ismail. (2017). Perkebunan Kelapa Sawit Indonesia Dalam Perspektif Pembangunan Berkelanjutan. *Jurnal Ilmu-Ilmu Sosial Indonesia*, 43(1), 81–94.
<http://jmi.ipsk.lipi.go.id/index.php/jmiipks/article/view/717/521>
- Jipkate, B. R., & Gohokar, V. V. (2012). “ A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms .” *International Journal of Computational Engineering Research*, 2(3), 737–739.
- Kharistiani, E. (2013). Sistem Informasi Geografis Pemetaan Potensi Sma/Smk Berbasis Web (Studi Kasus : Kabupaten Kebumen). *Jurnal Sarjana Teknik Informatika*, 1, 712–720.
- Maulida, L. (2018). Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167.
<https://doi.org/10.14421/jiska.2018.23-06>
- Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, 9(1), 95–100.
www.bps.go.id
- Pasaribu, D. F., Damanik, I. S., Irawan, E., Suhada, & Tambunan, H. S. (2021). Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat. *BIOS : Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 2(1), 11–20.
<https://doi.org/10.37148/bios.v2i1.17>
- Pulungan, I. M., Saifullah, S., Fauzan, M., & Windarto, A. P. (2019). Implementasi Algoritma K-Means Clustering dalam Menentukan Blok Tanaman Sawit Paling Produktif. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 338.
<https://doi.org/10.30645/senaris.v1i0.39>
- Tukiyat, T., & Djohan, Y. (2022). Analisis Penyebaran Pandemi Covid-19 Di Kota Jakarta Menggunakan Metode Clustering K-Means Dan Density Based Spatial Clustering of Application With Noise (Dbscan). *Jurnal Informatika*, 9(1), 43–54.
<https://doi.org/10.31294/inf.v9i1.11226>
- Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2(2), 213–219.