

Implementasi Algoritma Data Mining untuk Klustering Data Hasil Panen Kelapa Sawit

Implementation of Data Mining Algorithm for Clustering of Palm Oil Harvested Data

Widya Juli Mawaddah¹, Indra Gunawan², Ika Purnama Sari³

^{1,2,3} STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Article Info

Genesis Artikel:

Diterima, 11 Februari 2022

Direvisi, 20 Februari 2022

Disetujui, 3 Maret 2022

Kata Kunci:

Klustering
Kelapa Sawit
K-Means
Hasil Panen
Data Mining

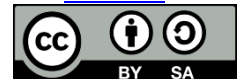
ABSTRAK

Kelapa sawit merupakan salah satu komoditas perkebunan yang memiliki peran strategis dalam pembangunan ekonomi Indonesia. Pada penelitian ini akan membahas hasil panen kelapa sawit pada PPKS Marihat yang merupakan salah satu cabang Pusat Penelitian Kelapa Sawit yang terletak di Kabupaten Simalungun, Medan, Sumatera Utara. Oleh karena itu penelitian ini bertujuan untuk mengklustering hasil panen kelapa sawit di PPKS Marihat, agar diketahui bagaimana pertumbuhannya. Algoritma klustering yang digunakan adalah K-Means. Dengan menggunakan metode ini data akan dikelompokkan menjadi 3 (tiga) *Cluster*, dimana penerapan proses K-Means Klustering menggunakan tools *Rapid Miner*. Data yang digunakan adalah data hasil panen kelapa sawit di PPKS Marihat tahun 2020 yang terdiri dari 100 item data. Hasil yang diperoleh adalah hasil panen dengan nilai yang amat baik sebanyak 66 item, data panen dengan nilai baik sebanyak 32 item, dan data panen dengan nilai cukup baik sebanyak 2 item, berdasarkan Jlh Netto dan Jlh Bruto terhadap setiap wilayah. Berdasarkan hal tersebut dapat disimpulkan bahwa Algoritma K-Means dapat digunakan untuk melakukan Klustering hasil panen kelapa sawit di PPKS Marihat.

ABSTRACT

Palm oil is one of the plantation commodities that has a strategic role in Indonesia's economic development. In this study, we will discuss oil palm yields at PPKS Marihat, one of the Oil Palm Research Center branches located in Simalungun Regency, Medan, North Sumatra. Know how it grows. The Clustering algorithm is used in K-Means. Using this method, the data will be grouped into 3 (three) Clusters, where the application of the K-Means Clustering process uses the Rapid Miner tools. The data used is data on oil palm harvests at PPKS Marihat in 2020, consisting of 100 data items. The results obtained are crop yields with an excellent value of 66 items, harvest data with a good deal of 32 items, and harvest data with a reasonably good value of 2 items, based on net total and gross amount for each region. Based on this, it can be concluded that the K-Means Algorithm can be used to Cluster oil palm yields at PPKS Marihat.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Korespondensi:

Widya Juli Mawaddah,
Program Studi Teknik Informatika,
STIKOM Tunas Bangsa, Pematangsiantar, indonesia
Email: widyajulimawaddah@gmail.com

1. PENDAHULUAN

Kelapa sawit salah satu tumbuhan yang dapat di tumbuh subur di wilayah Indonesia merupakan tanaman yang biji buahnya dapat menghasilkan minyak yang biasa digunakan dalam keseharian kita. Salah satu hasil dari biji buah kelapa sawit yaitu minyak sayur atau yang biasa digunakan untuk memasak [1]. Indonesia merupakan negara agraris yang diberi berkah kekayaan alam yang berlimpah baik sumber daya alam hayati maupun non hayati. Berbagai hasil alam tumbuh subur di Indonesia, karena tanah yang subur tanaman kelapa sawit dapat tumbuh subur di negara Indonesia.

Kelapa sawit terkenal terdiri dari empat macam golongan yaitu Varietas, Nigrescens, Virescens, Albescens, dan terbagi menjadi tiga tipe yaitu tipe Dura, Tenera, Pasifera. Masing-masing tipe dibedakan berdasarkan tebal tempurung. Daerah pertama penghasil kelapa sawit di Indonesia adalah daerah Jawa Barat (Lebak dan Tangerang), Lampung, Riau, Sumatera Barat, Sumatera Utara, dan Aceh. Negara penghasil kelapa sawit selain Indonesia adalah Malaysia, Amerika Tengah dan Nigeria. Di Sumatera Utara memiliki Pusat Penelitian Kelapa Sawit (PPKS) yang bertempat di Medan dan memiliki cabang di daerah Kabupaten Simalungun yaitu PPKS Marihat.

Kelapa sawit merupakan tanaman komoditas utama perkebunan Indonesia, di karena kan nilai ekonomi yang tinggi dan kelapa sawit merupakan tanaman penghasil minyak nabati terbanyak diantara tanaman penghasil minyak nabati yang lainnya. Sekarang ini prospek dari kelapa sawit sangat menguntungkan hal ini disebabkan karena hasil akhir dari pengolahan kelapa sawit seperti minyak goreng memiliki nilai ekonomi yang sangat tinggi [2]. Minyak goreng merupakan salah satu kebutuhan pokok bagi manusia, karena minyak goreng digunakan dalam keseharian kita. Selain diolah menjadi minyak goreng biji hasil panen kelapa sawit juga diperjualbelikan untuk mendapat kan biji benih yang berkualitas. Untuk mengetahui jenis benih yang berkualitas maka harus mempunyai kriteria tertentu.

Kebutuhan Tandan Buah Segar (TBS) akan meningkat seiring dengan meningkatnya kebutuhan, hasil produksi kelapa sawit tiap bulannya akan dilakukan peng-inputan data hasil produksi untuk dijadikan laporan bulanan oleh perusahaan yang bekerja di bagian produksi kelapa sawit (PKS). Banyak cabang ilmu Komputer dapat memecahkan masalah yang kompleks, salah satunya data mining [3]–[12]. Data mining merupakan teknik pengelolaan data dalam jumlah besar yang dapat digunakan untuk klasifikasi, asosiasi, estimasi / prediksi hingga klustering. Banyak penelitian-penelitian terdahulu terkait yang membahas tentang pengelompokan menggunakan algoritma K-Means, diantaranya: Penelitian untuk pengelompokan daerah rawan bencana berdasarkan provinsi yang ada di Indonesia. Hasil dari penelitian ini berupa pengelompokan data daerah rawan bencana yang dibagi menjadi 3 *Cluster*, yakni *Cluster* tinggi terdiri dari 4 provinsi, *Cluster* normal terdiri dari 14 Provinsi dan *Cluster* rendah terdiri 16 Provinsi [13]. Berikutnya penelitian yang dilakukan untuk pengelompokan kepadatan penduduk, indeks pembangunan manusia, angka pengangguran terbuka dan Rata-rata Lama Sekolah berdasarkan provinsi di Indonesia. Hasil dari penelitian ini berupa *Cluster* 1 yang terdiri dari 12 provinsi, *Cluster* 2 terdiri dari 6 provinsi dan *Cluster* 3 terdiri dari 1 provinsi, *Cluster* 4 terdiri dari 6 provinsi dan *Cluster* 5 terdiri dari 9 provinsi [14]. Selanjutnya penelitian tentang klustering laju pertumbuhan PDRB menurut lapangan usaha di Kota Surabaya dengan Algoritma K-Means. Hasil dari penelitian ini terdapat 9 kategori/sector dengan *Cluster* tinggi, 5 kategori/sector dengan *Cluster* sedang dan 3 kategori/sector dengan *Cluster* rendah [6]. Berikutnya penelitian yang dilakukan untuk meng kluster distribusi kasus rabies di kota Palembang menggunakan data mining K-Means. Pengolahan data pada penelitian ini menggunakan software *Rapid Miner* dengan hasil bahwa dari 16 kecamatan di Palembang, tujuh kecamatan termasuk kluster wilayah sangat rawan rabies (C0), sedangkan empat kecamatan termasuk di *Cluster* daerah rawan rabies (C1), dan lima kecamatan yang termasuk dalam *Cluster* regional tidak rawan rabies (C2) [15]. Selanjutnya penelitian dengan menggunakan algoritma K-Means untuk mengelompokkan tingkat keterlibatan siswa program sarjana sains universitas Amerika Utara dalam lingkungan e-learning [16]. Penelitian-penelitian terkait ini lah yang melatarbelakangi dilakukannya penelitian untuk mengelompokkan hasil panen kelapa sawit di PPKS Marihat menggunakan Algoritma K-Means Klustering.

2. METODE PENELITIAN

2.1. Pengumpulan Data

Sumber data penelitian diperoleh dari kantor PPKS Unit Marihat. Data yang digunakan yaitu data hasil panen kelapa sawit pada tahun 2020. Pengumpulan data penelitian dengan cara berikut :

1. Pengamatan Langsung (Observasi)

Pengamatan dilakukan secara langsung dengan datang ke kantor PPKS yang ada di Marihat guna memperoleh data yang diperlukan.

2. Wawancara

Wawancara dilakukan dengan memberikan pertanyaan kepada pegawai yang bekerja di bagian yang mengolah hasil panen agar mendapatkan informasi tentang hasil panen kelapa sawit

Tabel 1. Data Hasil Panen Kelapa Sawit

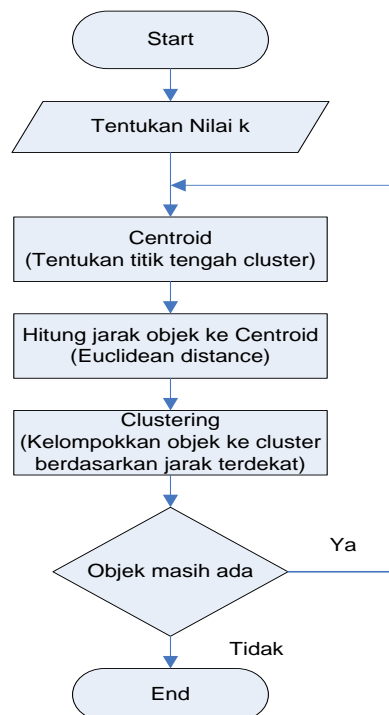
No	Barang	Jlh Netto	Jlh Bruto
1	DxP Simalungun	1000	330
2	DxP Simalungun	1535	110
3	DxP Yangambi	1870	110
4	DxP PPKS 540	1670	201

No	Barang	Jlh_Netto	Jlh_Bruto
5	DxP PPKS 540	1600	99
6	DxP PPKS 540	975	110
7	DxP PPKS 540	1430	113
8	DxP PPKS 540	1480	148
9	DxP Simalungun	1900	540
10	DxP Simalungun	1310	720
11	DxP PPKS 540	700	60
12	DxP Yangambi	1220	55
13	DxP PPKS 540	1160	165
14	DxP PPKS 540	1540	55
15	DxP Simalungun	1350	110
16	DxP Simalungun	1873	110
17	DxP Simalungun	1400	378
18	DxP Yangambi	1510	113
19	DxP Simalungun	920	110
20	DxP Simalungun	1465	223
21	DxP Simalungun	870	60
22	DxP Simalungun	1730	56
23	DxP Yangambi	1540	115
24	DxP PPKS 540	800	110
25	DxP Simalungun	2150	6
26	DxP PPKS 540	1070	220
27	DxP PPKS 540	1400	111
28	DxP PPKS 540	1042	57
29	DxP PPKS 540	2500	250
30	DxP PPKS 540	1310	55
31	DxP PPKS 540	2206	110
32	DxP Yangambi	405	60
33	DxP Yangambi	360	45
34	DxP PPKS 540	1520	220
35	DxP Yangambi	2110	70
36	DxP PPKS 540	1773	80
37	DxP Simalungun	1610	110
38	DxP Simalungun	1725	108
39	DxP Simalungun	1040	10
40	DxP Simalungun	3350	550
41	DxP Yangambi	1005	100
42	DxP Yangambi	1940	167
43	DxP PPKS 540	1380	110
44	DxP PPKS 540	1540	222
45	DxP Yangambi	810	58
46	DxP Yangambi	1116	60
47	DxP Yangambi	1380	113
48	DxP PPKS 540	1895	110
49	DxP Simalungun	1210	110
50	DxP PPKS 540	927	70
51	DxP PPKS 540	1020	110
52	DxP Yangambi	1110	60
53	DxP Simalungun	945	220
54	DxP PPKS 540	1400	40
55	DxP Simalungun	1650	110
56	DxP Simalungun	2200	50
57	DxP PPKS 540	2970	278
58	DxP PPKS 540	1395	670
59	DxP Yangambi	1185	60
60	DxP Simalungun	2050	55
61	DxP Simalungun	1840	330
62	DxP Simalungun	510	220
63	DxP Simalungun	2130	175
64	DxP Simalungun	2260	23
65	DxP Simalungun	345	5
66	DxP Simalungun	980	65
67	DxP Simalungun	1610	110
68	DxP Simalungun	2200	73
69	DxP PPKS 540	1050	166
70	DxP PPKS 540	1890	223
71	DxP Yangambi	905	165
72	DxP PPKS 540	1380	261
73	DxP PPKS 540	1600	110
74	DxP PPKS 540	1850	110
75	DxP PPKS 540	1540	333
76	DxP PPKS 540	950	55
77	DxP Yangambi	1570	222
78	DxP Yangambi	4520	424

No	Barang	Jlh_Netto	Jlh_Bruto
79	DxP PPKS 540	1205	110
80	DxP PPKS 540	1678	168
81	DxP Simalungun	870	570
82	DxP Simalungun	1880	169
83	DxP Simalungun	1510	140
84	DxP Simalungun	1860	221
85	DxP Simalungun	1540	50
86	DxP Simalungun	2700	292
87	DxP Simalungun	3000	100
88	DxP Simalungun	1820	60
89	DxP PPKS 540	1430	170
90	DxP PPKS 540	1435	110
91	DxP PPKS 540	1990	110
92	DxP PPKS 540	1175	55
93	DxP PPKS 540	1240	17
94	DxP PPKS 540	1350	330
95	DxP Yangambi	2030	110
96	DxP Simalungun	920	40
97	DxP Simalungun	1040	110
98	DxP Simalungun	1010	55
99	DxP Simalungun	2410	400
100	DxP PPKS 540	1050	110

2.2. Flowchart K-Means

Berikut ini disajikan Diagram alir dari metode *Clustering K-Means*.



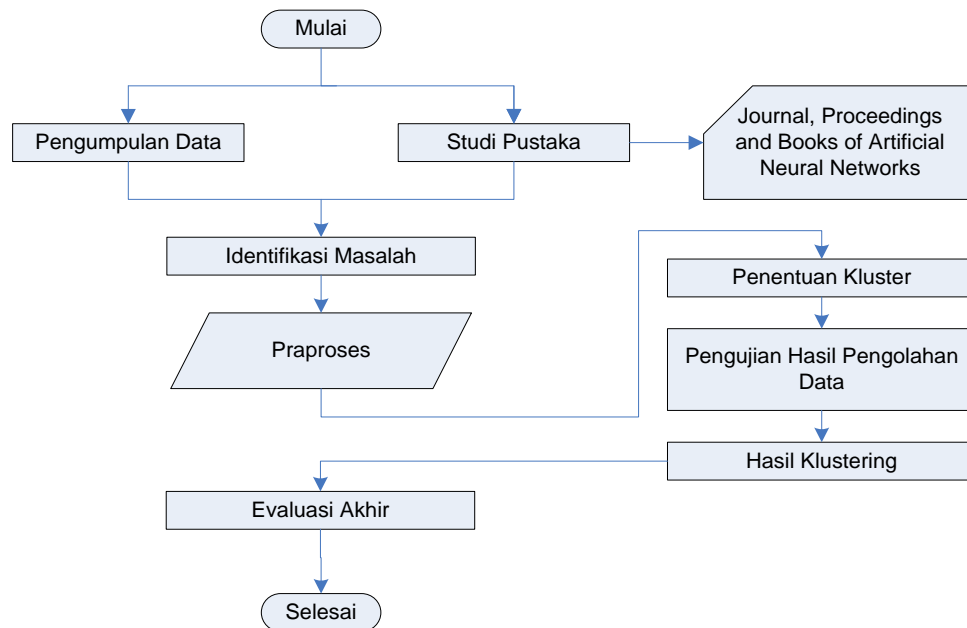
Gambar 1. Diagram alir metode *K-Means*

Langkah-langkah metode *K-Means* dapat dijelaskan sebagai berikut :

1. Tentukan jumlah *Cluster* (*k*) pada data set.
2. Tentukan nilai pusat (*Centroid*).
Penentuan nilai *Centroid* pada tahap awal dilakukan secara random
3. Pada masing-masing *record*, hitung jarak terdekat dengan *Centroid*.
Ada beberapa cara yang dapat digunakan untuk mengukur jarak data ke pusat kelompok, diantaranya *Euclidean*, *Manhattan/City Block*, dan *Minkowsky*. Setiap cara memiliki kelebihan dan kekurangan masing-masing. Pada penelitian ini, jarak *Centroid* yang digunakan adalah *Euclidean Distance*,
4. Kelompokkan objek berdasarkan jarak ke *Centroid* terdekat
5. Ulangi langkah ke-3 hingga langkah ke-4, lakukan *iterasi* hingga *Centroid* bernilai optimal.

2.3. Tahapan Penelitian

Adapun tahapan yang dilakukan untuk menyelesaikan masalah penelitian disusun dalam kerangka kerja berikut:



Gambar 2. Tahapan Penelitian

Keterangan:

1. Pengumpulan Data
Dataset penelitian berupa data hasil panen kelapa sawit di PPKS Marihat tahun 2020.
2. Studi Pustaka
Dilakukan untuk melengkapi pengetahuan dasar dan teori-teori yang digunakan dalam penelitian (artikel ilmiah, buku, prosiding dll)
3. Identifikasi Masalah
Dilakukan setelah semua data-data terpenuhi kemudian didapatkan *dataset* yang sesuai untuk dilakukan proses yang ditentukan
4. Praproses
Tahapan yang dikerjakan dengan melakukan perubahan terhadap beberapa tipe data pada atribut *dataset* dengan tujuan untuk mempermudah pemahaman terhadap isi *record*, juga melakukan seleksi dengan memperhatikan konsistensi data, *missing value* dan *redundant* pada data.
5. Penentuan Kluster
Hasil dari tahap ini adalah menentukan jumlah klustering yang ditetapkan dengan menggunakan datamining K-Means
6. Penguji Hasil Pengolahan Data
Setelah proses penentuan kluster selesai, maka dilakukan tahapan uji coba terhadap hasil pengolahan data dengan menggunakan *Software RapidMiner*.
7. Hasil Klustering
Hasil pengelompokan yang diperoleh berdasarkan jumlah kluster yang telah ditetapkan.
8. Evaluasi Akhir
Dilakukan untuk mengetahui apakah testing hasil pengolahan data sesuai

3. HASIL DAN PEMBAHASAN

3.1. Klustering Hasil Panen Sawit Menggunakan K-Means

Berikut langkah-langkah penyelesaian yang dilakukan penulis dalam *clustering* hasil panen sawit menggunakan *K-Means*:

1. Menentukan jumlah data yang ingin di *cluster*, dimana sampel data jumlah panen yang akan digunakan dalam proses *clustering* adalah data netto dan bruto. Jumlah data dapat kita lihat pada tabel 1 yaitu berjumlah 100 data.
2. Menetapkan nilai K jumlah *cluster* benih sawit sebanyak 3 *cluster* (k-3). *Cluster* yang dibentuk yaitu *cluster* amat baik, *cluster* baik, dan *cluster* cukup baik.

3. Menentukan nilai centroid (pusat *cluster*) awal yang telah di tentukan secara random berdasarkan nilai variabel data yang akan di *cluster* sebanyak yang ditentukan sebelumnya. Nilai centroid data awal untuk iterasi 1 adalah :

Tabel 2. Penentuan Pusat Awal

Penentuan Pusat Awal		
Data ke-34 pusat <i>cluster</i> 0	1520	220
Data ke-78 pusat <i>cluster</i> 1	4520	242
Data ke-65 pusat <i>cluster</i> 2	345	5

- a. Menghitung jarak setiap data panen terhadap pusat *cluster*.

Setelah data nilai pusat *cluster* awal ditentukan, maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat *cluster* dengan menggunakan rumus yang perhitungannya dicari menggunakan excel, dapat dilihat pada tabel 3.

Tabel 3. Perhitungan jarak data dengan titik pusat *cluster* iterasi 1

No	Barang	Jlh_Netto	Jlh_Bruto	C0	C1	C2	CLUS
1	DxP Simalungun	1000	330	531,5073	3521,1	731,1976	0
2	DxP Simalungun	1535	110	111,018	2987,917	1194,623	0
3	DxP Yangambi	1870	110	366,8787	2653,286	1528,61	0
4	DxP PPKS 540	1670	201	151,1985	2850,295	1339,418	0
5	DxP PPKS 540	1600	99	145,0552	2923,499	1258,515	0
6	DxP PPKS 540	975	110	555,9901	3547,457	638,6901	0
7	DxP PPKS 540	1430	113	139,8177	3092,692	1090,362	0
8	DxP PPKS 540	1480	148	82,36504	3041,453	1143,973	0
9	DxP Simalungun	1900	540	496,7897	2636,893	1644,46	0
10	DxP Simalungun	1310	720	542,3099	3245,394	1201,02	0
11	DxP PPKS 540	700	60	835,4639	3824,333	359,2353	2
12	DxP Yangambi	1220	55	342,3814	3305,294	876,4274	0
13	DxP PPKS 540	1160	165	364,1772	3360,882	830,557	0
14	DxP PPKS 540	1540	55	166,2077	2985,862	1196,046	0
15	DxP Simalungun	1350	110	202,4846	3172,747	1010,47	0
16	DxP Simalungun	1873	110	369,7418	2650,289	1531,603	0
17	DxP Simalungun	1400	378	198,4036	3122,963	1118,997	0
18	DxP Yangambi	1510	113	107,4663	3012,763	1169,995	0
19	DxP Simalungun	920	110	610	3602,419	584,5083	2
20	DxP Simalungun	1465	223	55,08176	3055,059	1141,019	0
21	DxP Simalungun	870	60	669,4027	3654,535	527,8731	2
22	DxP Simalungun	1730	56	266,4507	2796,193	1385,939	0
23	DxP Yangambi	1540	115	106,8878	2982,705	1200,052	0
24	DxP PPKS 540	800	110	728,3543	3722,341	466,9582	2
25	DxP Simalungun	2150	6	665,354	2381,721	1805	0
26	DxP PPKS 540	1070	220	450	3450,07	756,2076	0
27	DxP PPKS 540	1400	111	162,1142	3122,749	1060,312	0
28	DxP PPKS 540	1042	57	505,0277	3482,917	698,9371	0
29	DxP PPKS 540	2500	250	980,4591	2020,016	2168,882	0
30	DxP PPKS 540	1310	55	267,0674	3215,442	966,2945	0
31	DxP PPKS 540	2206	110	694,7633	2317,762	1863,96	0
32	DxP Yangambi	405	60	1126,421	4119,023	81,3941	2
33	DxP Yangambi	360	45	1173,126	4164,662	42,72002	2
34	DxP PPKS 540	1520	220	0	3000,081	1194,508	0
35	DxP Yangambi	2110	70	608,7693	2416,13	1766,196	0
36	DxP PPKS 540	1773	80	289,1522	2751,773	1429,968	0
37	DxP Simalungun	1610	110	142,1267	2912,992	1269,35	0
38	DxP Simalungun	1725	108	233,6001	2798,21	1383,839	0
39	DxP Simalungun	1040	10	523,9275	3487,725	695,018	0
40	DxP Simalungun	3350	550	1859,516	1209,861	3054,022	1
41	DxP Yangambi	1005	100	528,7958	3517,867	666,8021	0
42	DxP Yangambi	1940	167	423,3308	2581,09	1603,206	0
43	DxP PPKS 540	1380	110	178,0449	3142,773	1040,312	0
44	DxP PPKS 540	1540	222	20,09975	2980,067	1214,543	0
45	DxP Yangambi	810	58	728,2472	3714,56	468,0107	2
46	DxP Yangambi	1116	60	434,5296	3408,862	772,9592	0
47	DxP Yangambi	1380	113	176,2073	3142,649	1040,62	0
48	DxP PPKS 540	1895	110	390,8005	2628,317	1553,552	0
49	DxP Simalungun	1210	110	328,9377	3312,631	871,3495	0
50	DxP PPKS 540	927	70	611,6772	3597,115	585,6185	2
51	DxP PPKS 540	1020	110	511,957	3502,488	683,1179	0
52	DxP Yangambi	1110	60	440,1136	3414,853	766,9746	0
53	DxP Simalungun	945	220	575	3575,068	637,3578	0
54	DxP PPKS 540	1400	40	216,3331	3126,532	1055,58	0
55	DxP Simalungun	1650	110	170,2939	2873,034	1309,217	0
56	DxP Simalungun	2200	50	700,928	2327,931	1855,546	0

No	Barang	Jlh_Netto	Jlh_Bruto	C0	C1	C2	CLUS
57	DxP PPKS 540	2970	278	1451,16	1550,418	2639,158	0
58	DxP PPKS 540	1395	670	467,0385	3154,173	1242,87	0
59	DxP Yangambi	1185	60	371,2479	3339,962	841,7987	0
60	DxP Simalungun	2050	55	555,0901	2477,069	1705,733	0
61	DxP Simalungun	1840	330	338,3785	2681,444	1529,918	0
62	DxP Simalungun	510	220	1010	4010,06	271,0166	2
63	DxP Simalungun	2130	175	611,6576	2390,939	1793,077	0
64	DxP Simalungun	2260	23	765,7735	2270,586	1915,085	0
65	DxP Simalungun	345	5	1194,508	4181,721	0	2
66	DxP Simalungun	980	65	561,8051	3544,422	637,8283	0
67	DxP Simalungun	1610	110	142,1267	2912,992	1269,35	0
68	DxP Simalungun	2200	73	695,7076	2326,147	1856,246	0
69	DxP PPKS 540	1050	166	473,092	3470,832	723,1501	0
70	DxP PPKS 540	1890	223	370,0122	2630,069	1560,304	0
71	DxP Yangambi	905	165	617,4545	3615,82	582,4088	2
72	DxP PPKS 540	1380	261	145,8801	3140,057	1066,19	0
73	DxP PPKS 540	1600	110	136,0147	2922,982	1259,385	0
74	DxP PPKS 540	1850	110	347,8505	2673,261	1508,658	0
75	DxP PPKS 540	1540	333	114,7563	2981,389	1239,197	0
76	DxP PPKS 540	950	55	593,4012	3574,894	607,0626	0
77	DxP Yangambi	1570	222	50,03998	2950,068	1244,072	0
78	DxP Yangambi	4520	424	3006,928	182	4195,973	1
79	DxP PPKS 540	1205	110	333,654	3317,627	866,3862	0
80	DxP PPKS 540	1678	168	166,337	2842,963	1342,929	0
81	DxP Simalungun	870	570	738,2412	3664,708	771,2652	0
82	DxP Simalungun	1880	169	363,5946	2641,009	1543,736	0
83	DxP Simalungun	1510	140	80,62258	3011,728	1172,796	0
84	DxP Simalungun	1860	221	340,0015	2660,083	1530,321	0
85	DxP Simalungun	1540	50	171,1724	2986,179	1195,847	0
86	DxP Simalungun	2700	292	1182,195	1820,687	2372,424	0
87	DxP Simalungun	3000	100	1484,857	1526,618	2656,699	0
88	DxP Simalungun	1820	60	340	2706,127	1476,025	0
89	DxP PPKS 540	1430	170	102,9563	3090,839	1097,474	0
90	DxP PPKS 540	1435	110	139,0144	3087,823	1095,046	0
91	DxP PPKS 540	1990	110	482,7007	2533,441	1648,348	0
92	DxP PPKS 540	1175	55	382,4265	3350,223	831,5047	0
93	DxP PPKS 540	1240	17	345,8453	3287,708	895,0804	0
94	DxP PPKS 540	1350	330	202,4846	3171,221	1056,243	0
95	DxP Yangambi	2030	110	521,7279	2493,496	1688,268	0
96	DxP Simalungun	920	40	626,4184	3605,663	576,0642	2
97	DxP Simalungun	1040	110	492,4429	3482,503	702,8869	0
98	DxP Simalungun	1010	55	536,0271	3514,978	666,8771	0
99	DxP Simalungun	2410	400	908,0198	2115,907	2102,439	0
100	DxP PPKS 540	1050	110	482,7007	3472,51	712,7763	0

Mengelompokkan masing-masing data obyek ke dalam centroid yang paling terdekat, bandingkan nilai C0,C1,C2 kemudian lihat nilai masing-masing *cluster* sesuai urutan *cluster* dengan menggunakan bantuan tanda (1) yang mengatakan data menjadi anggota pada *cluster*, seperti tabel-tabel berikut.

Tabel 4. Kelompok Data 1

Pengelompokkan Data		Kelompok Data 1	
No	C0	C1	C2
1	1		
2	1		
3	1		
4	1		
5	1		
6	1		
7	1		
8	1		
9	1		
10	1		
11			1
12	1		
13	1		
14	1		
15	1		
16	1		
17	1		
18	1		
19			1
...
100	1		

Yang termasuk dalam *clustering* 0 sebanyak 87 dan yang termasuk *clustering* 1 sebanyak 2 dan yang termasuk *clustering* 2 sebanyak 11. Selanjutnya kembali ke langkah 3 yaitu menentukan pusat *cluster* baru sampai hasil pengelompokan hasil *cluster* sama dengan kelompok sebelumnya, berikut tabel-tabel pengelompokannya.

Tabel 5. Centroid Baru Data 1

C0	C1	C2
1561,183908	3935	738,8182
156,6781609	487	100,2727

Dengan menggunakan pusat *cluster* sebanyak 3 pusat *cluster* yang diambil dengan rumus berhenti di iterasi ke 2 dengan hasil akhir sebagai berikut.

Tabel 6. Kelompok Data 2

Pengelompokan Data		Kelompok Data 2	
No	C0	C1	C2
1			1
2	1		
3	1		
4	1		
5			1
6	1		
7	1		
8	1		
9	1		
10	1		
11			1
12	1		
13	1		
14	1		
15	1		
16	1		
17	1		
18	1		
19			1
...
100	1		

Yang termasuk dalam *clustering* 0 sebanyak 68 dan yang termasuk *clustering* 1 sebanyak 4 dan yang termasuk *clustering* 2 sebanyak 28. Iterasi berhenti jika hasil *cluster* sudah sama dengan hasil *cluster* sebelumnya. Karena hasil *cluster* iterasi 1 dan iterasi ke 2 belum sama, maka dilanjutkan pencarian sampai mencapai hasil yang sama untuk melihat *cluster* yang terbaik. Berikut adalah hasil *cluster* dari iterasi selanjutnya, pencarian dilakukan dengan rumus *clustering* K-means di Excel.

Tabel 7. Centroid Baru Data 2

C0	C1	C2
1656,765	3460	904,2857
161,1912	338	117,5

Tabel 8. Kelompok Data 3

Pengelompokan Data		Kelompok Data 3	
No	C0	C1	C2
1			1
2	1		
3	1		
4	1		
5	1		
6			1
7	1		
8	1		
9	1		
10	1		
11			1
12			1
13			1
14	1		
15	1		
16	1		
17	1		
18	1		
19			1
...
100			1

Yang termasuk dalam *clustering* 0 sebanyak 59 dan yang termasuk *clustering* 1 sebanyak 5 dan yang termasuk *clustering* 2 sebanyak 36.

Tabel 9. Centroid Baru Data 3

C0	C1	C2
1714,237	3308	948,3333
169,0847	328,8	110,6389

Tabel 10. Kelompok Data 4

Pengelompokkan Data		Kelompok Data 4	
No	C0	C1	C2
1	1		
2	1		
3			1
4			1
5	1		
6	1		
7	1		
8	1		
9			1
10	1		
11	1		
12	1		
13	1		
14	1		
15	1		
16			1
17	1		
18	1		
19	1		
...
100	1		

Yang termasuk dalam *clustering* 0 sebanyak 66 dan yang termasuk *clustering* 1 sebanyak 2 dan yang termasuk *clustering* 2 sebanyak 32.

Tabel 11. Centroid Baru Data 4

C0	C1	C2
1721,207	3308	958,1081
171,0517	328,8	109,1351

Tabel 12. Kelompok Data 5

Pengelompokkan Data		Kelompok Data 5	
No	C0	C1	C2
1	1		
2	1		
3			1
4			1
5	1		
6	1		
7	1		
8	1		
9			1
10	1		
11	1		
12	1		
13	1		
14	1		
15	1		
16			1
17	1		
18	1		
19	1		
...
100	1		

Yang termasuk dalam *clustering* 0 sebanyak 66 dan yang termasuk *clustering* 1 sebanyak 2 dan yang termasuk *clustering* 2 sebanyak 32. Karena hasil *cluster* dari kelompok data 5 dengan sebelumnya yaitu data kelompok 4 sama, maka iterasi sudah dapat dihentikan, maka didapatkan hasil *cluster* terbaik yaitu :

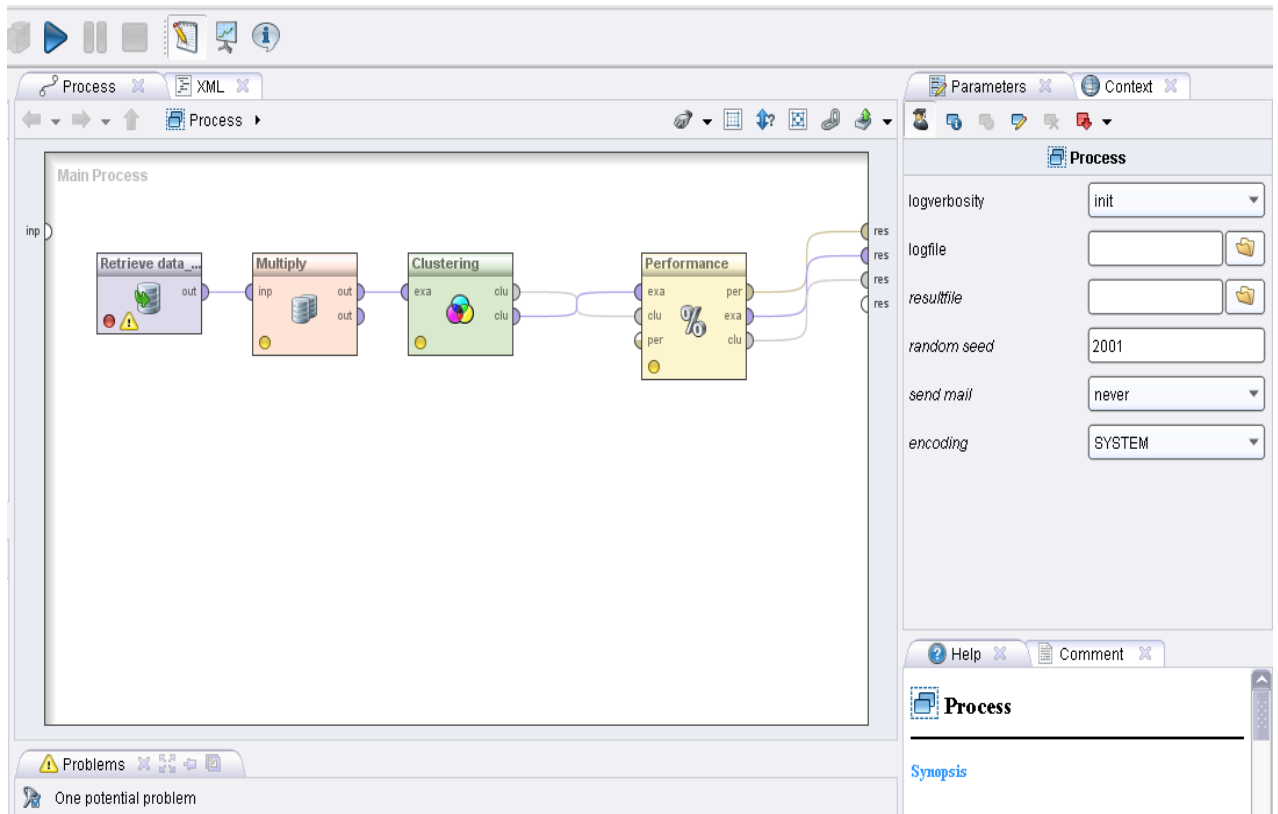
Tabel 13. Hasil *Cluster*

C0	C1	C2
66	2	32

3.2. Hasil Percobaan dan Pengujian

1. Pemrosesan Sistem

Pada tahap ini akan dijelaskan tahapan proses penggunaan K-Means di dalam *Rapid Miner* data yang telah di *Import*. Tahapan pertama drag atau klik dan tahan data yang telah disimpan pada Repositories ke lembar kerja Main Process, lalu klik *Clustering* and *Segmentation* dan pilih K-Means. Lalu hubungkan antara read excel dengan *Clustering* seperti gambar 3 berikut.

Gambar 3. Pemrosesan Data dengan K-Means Menggunakan *Rapid Miner*

2. Hasil Pengelompokkan (Klustering)

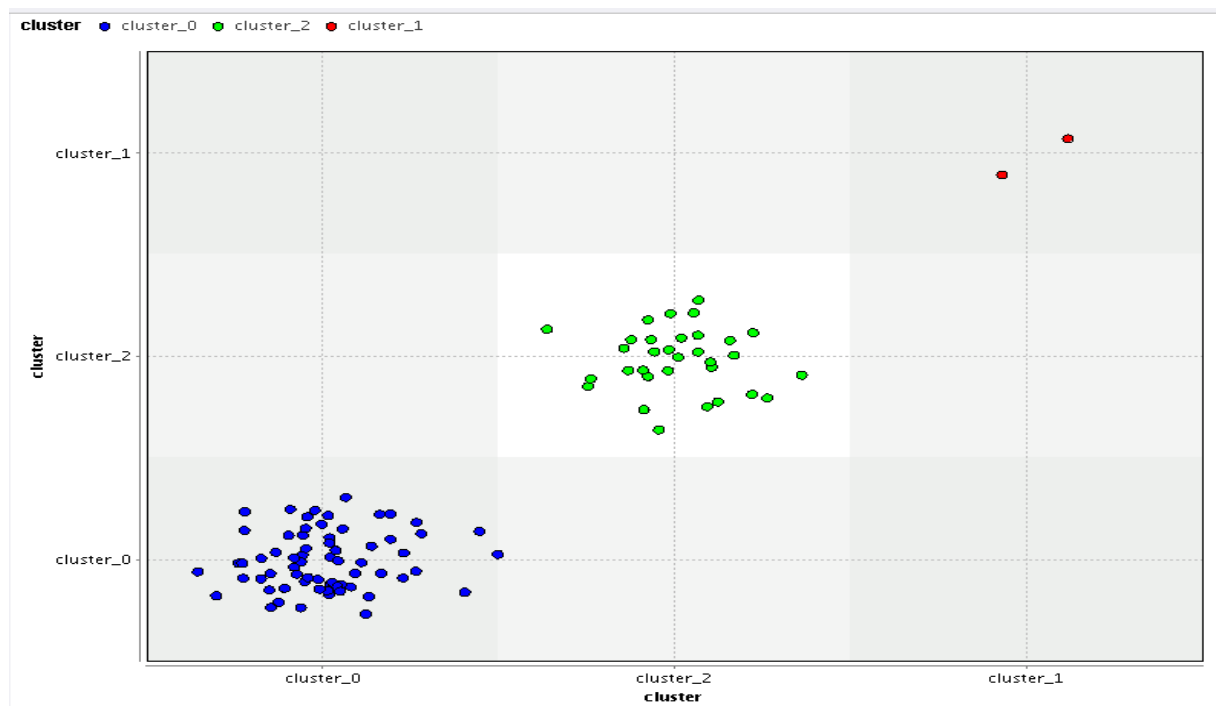
Untuk mendapatkan hasil pengelompokkan (klustering) maka pada tahap selanjutnya klik run (tanda panah biru) pada *toolbar*. Maka akan menampilkan hasil akhir serta langkah terakhir dalam penggunaan *tools rapidminer* ini. Dapat dilihat pada gambar 4.

Cluster Model

```
Cluster 0: 66 items
Cluster 1: 2 items
Cluster 2: 32 items
Total number of items: 100
```

Gambar 4. Nilai *Cluster Model Rapid Miner* (K-Means)

Keterangan: Jumlah *Cluster* 0 (panen amat baik) sebanyak 66 items. Jumlah *Cluster* 1 (panen baik) sebanyak 32 items, dan Jumlah *Cluster* 2 (panen cukup baik) sebanyak 2 items, berdasarkan Jlh Netto dan Jlh Bruto terhadap setiap wilayah. Sehingga dapat diketahui hasil pengelompokkan dari *Rapid Miner* berikut, yang dapat dilihat pada gambar 5.



Gambar 5. Hasil Pengelompokan dengan *Rapid Miner* (K-Means)

Pada data *import configuration wizard* kemudian pilih tempat kita menyimpan data yang akan digunakan. Selanjutnya pilih *file name* data yang akan digunakan. Lalu klik *next* pada bagian kanan bawah.

3. Pembahasan

Berdasarkan penjelasan sebelumnya serta hasil yang telah ditampilkan maka terdapat keterkaitan dari hasil yang didapat antara perhitungan manual algoritma dengan hasil yang ditampilkan oleh *Rapid Miner* dan dapat disimpulkan bahwa hasil perhitungan manual dari metode K-Means dengan bantuan Microsoft Excel 2010 hasilnya sama dengan pengujian menggunakan *Rapid Miner*.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dengan menggunakan algoritma K-Means, maka dapat disimpulkan bahwa hasil *Cluster 0* (panen amat baik) sebanyak 66 items. *Cluster 1* (panen baik) sebanyak 32 items, dan *Cluster 2* (panen cukup baik) sebanyak 2 items, berdasarkan Jlh Netto dan Jlh Bruto terhadap setiap wilayah. Pengujian data pada *Rapid Miner* dengan menggunakan algoritma K-Means dapat menampilkan 3 (tiga) kelas yang sama dengan perhitungan manual. Sehingga algoritma K-Means dapat digunakan untuk klustering hasil panen kelapa sawit di PPKS Marihat.

REFERENSI

- [1] Y. Andriani, A. Wanto, and H. Handrizal, "Jaringan Saraf Tiruan dalam Memprediksi Produksi Kelapa Sawit di PT. KRE Menggunakan Algoritma Levenberg Marquardt," *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, vol. 1, no. September, pp. 249–259, 2019.
- [2] C. Michael, P. Marpaung, and F. Siburian, "Analisis Hubungan Biaya Produksi Kelapa Sawit Terhadap Pendapatan Petani di Desa Pulo Bayu Kecamatan Hutabayu Raja, Kabupaten Simalungun Organik," *Jurnal Agroteknosains*, vol. 4, no. 1, pp. 8–16, 2020.
- [3] N. Arminarahmah, A. D. GS, G. W. Bhawika, M. P. Dewi, and A. Wanto, "Mapping the Spread of Covid-19 in Asia Using Data Mining X-Means Algorithms," *IOP Conf. Series: Materials Science and Engineering*, vol. 1071, no. 012018, pp. 1–7, 2021.
- [4] A. Pradipta, D. Hartama, A. Wanto, S. Saifullah, and J. Jalaluddin, "The Application of Data Mining in Determining Timely Graduation Using the C45 Algorithm," *IJISTECH (International Journal of Information System & Technology)*, vol. 3, no. 1, pp. 31–36, 2019.
- [5] T. H. Sinaga, A. Wanto, I. Gunawan, S. Sumarno, and Z. M. Nasution, "Implementation of Data Mining Using C4.5 Algorithm on Customer Satisfaction in Tirta Lihou PDAM," *Journal of Computer Networks, Architecture, and High-Performance Computing*, vol. 3, no. 1, pp. 9–20, 2021.
- [6] N. A. Febriyati, A. D. GS, and A. Wanto, "GRDP Growth Rate Clustering in Surabaya City uses the K- Means

- Algorithm,” *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 276–283, 2020.
- [7] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, “COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm,” *Journal of Physics: Conference Series*, vol. 1783, no. 1, p. 012027, 2021.
- [8] I. Parlina *et al.*, “Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category,” in *Journal of Physics: Conference Series*, 2019, vol. 1255, no. 1, p. 012031.
- [9] M. A. Hanafiah, A. Wanto, and P. B. Indonesia, “Implementation of Data Mining Algorithms for Grouping Poverty Lines by District/City in North Sumatra,” *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 315–322, 2020.
- [10] I. S. Damanik, A. P. Windarto, A. Wanto, Poningsih, S. R. Andani, and W. Saputra, “Decision Tree Optimization in C4.5 Algorithm Using Genetic Algorithm,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [11] A. Wanto *et al.*, *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis, 2020.
- [12] D. Hartama, A. Perdana Windarto, and A. Wanto, “The Application of Data Mining in Determining Patterns of Interest of High School Graduates,” *Journal of Physics: Conference Series*, vol. 1339, no. 1, p. 012042, Dec. 2019.
- [13] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, “Classification of Natural Disaster Prone Areas in Indonesia using K-Means,” *International Journal of Grid and Distributed Computing*, vol. 11, no. 8, pp. 87–98, 2018.
- [14] A. S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha, and M. Azmi, “Using K-Means Clustering to Cluster Provinces in Indonesia,” *Journal of Physics: Conference Series*, vol. 1028, no. 1, pp. 1–6, 2018.
- [15] K. Rahayu, L. Novianti, and M. Kusnandar, “Implementation Data Mining with K-Means Algorithm for Clustering Distribution Rabies Case Area in Palembang City,” *Journal of Physics: Conference Series*, vol. 1500, no. 1, pp. 1–9, 2020.
- [16] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, “Student Engagement Level in an e-Learning Environment: Clustering Using K-means,” *American Journal of Distance Education*, vol. 34, no. 2, pp. 137–156, 2020.