

Penerapan Algoritma K-Means *Clustering* Pada Produksi Perkebunan Kelapa Sawit Menurut Provinsi

Deny Haryadi

Teknologi Informasi, Institut Teknologi Telkom Jakarta
Jln. Daan Mogot KM 11 Cengkareng Jakarta Barat Indonesia

denyharyadi@ittelkom-jkt.ac.id

Abstract

Kelapa sawit merupakan salah satu perkebunan yang produksinya berpengaruh menjadi salah satu komoditas ekspor utama Indonesia. Tingginya produksi perkebunan kelapa sawit dipengaruhi oleh adanya permintaan akan kebutuhan konsumsi minyak nabati dunia. Selain itu, tingginya produksi perkebunan kelapa sawit memiliki peranan besar terhadap perekonomian nasional terutama dari sisi penerimaan pajak dan pundi-pundi devisa negara. Tujuan dari penelitian ini untuk pengelompokan provinsi-provinsi yang menjadi prioritas terhadap produksi perkebunan kelapa sawit di Indonesia. Berdasarkan hasil pengujian yang telah dilakukan dalam penelitian ini menggunakan algoritma K-Means *Clustering* yaitu *Cluster 1* merupakan kategori provinsi dengan produksi perkebunan kelapa sawit rendah atau *Low* yaitu 14 (aceh, sumatera barat, Bengkulu, Lampung, kep.bangka belitung, kep.riau, Jawa Barat, Banten, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Barat, Papua Barat dan Papua) dari 21 kategori provinsi yang diuji, kemudian *cluster 2* adalah kategori provinsi dengan produksi perkebunan kelapa sawit sedang atau *Medium* yaitu 4 (Jambi, Sumatera Selatan, Kalimantan Barat, dan Kalimantan Timur) dari 21 kategori provinsi yang diuji, dan terakhir adalah *cluster 3* merupakan kategori provinsi dengan produksi perkebunan kelapa sawit tinggi atau *High* yaitu 3 (Sumatera Utara, Riau, dan Kalimantan Tengah) dari 21 kategori provinsi yang diuji.

Keywords: Algoritma K-Means *Clustering*, Perkebunan Kelapa Sawit, Produksi, Provinsi

I. INTRODUCTION

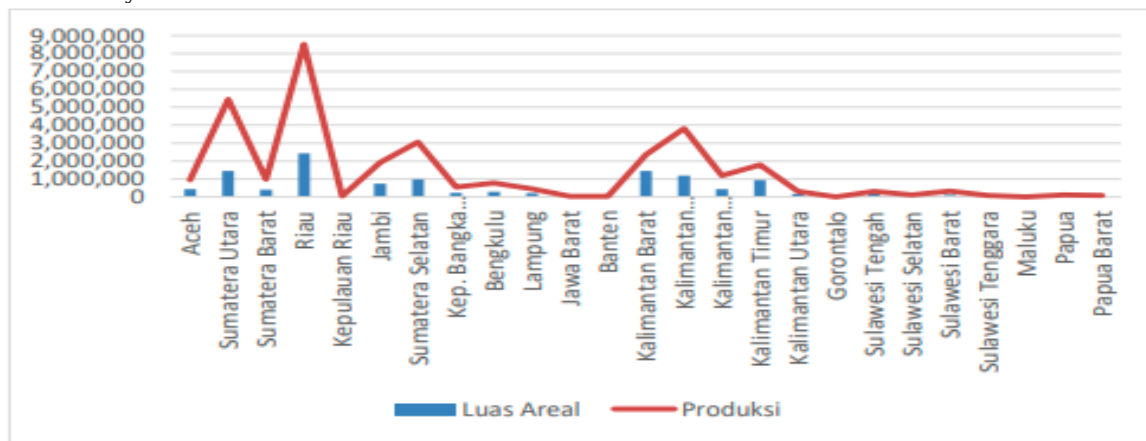
Indonesia merupakan negara penghasil kelapa sawit terbesar dunia dengan produksi di atas 40,56 juta ton pada tahun 2018. Kelapa sawit merupakan salah satu perkebunan yang produksinya berpengaruh menjadi salah satu komoditas ekspor utama Indonesia. Produk kelapa sawit mulai dari minyak kelapa sawit, mentega, cokelat kemasan, sabun mandi, sampo, deterjen, kosmetik, hingga bahan bakar biodiesel. Tingginya produksi perkebunan kelapa sawit dipengaruhi oleh adanya permintaan akan kebutuhan konsumsi minyak nabati dunia. Selain itu, tingginya produksi perkebunan kelapa sawit memiliki peranan besar terhadap perekonomian nasional terutama dari sisi penerimaan pajak dan pundi-pundi devisa negara. Berdasarkan data Badan Pusat Statistik (BPS) bulan Desember 2019 luas perkebunan kelapa sawit di Indonesia mencapai 14,32 juta hektar. Perkebunan kelapa sawit di Indonesia berada pada pulau Sumatera, Kalimantan, Jawa, Sulawesi, Papua dan berada pada pulau tertentu di Indonesia. Hal ini dikarenakan kondisi geografis wilayah Indonesia yang memiliki suhu hangat, sinar matahari, dan curah hujan tinggi sangat cocok untuk pengembangan perkebunan kelapa sawit [1]. Berdasarkan

uraian perlu dilakukan suatu metode pengelompokan provinsi-provinsi yang menjadi prioritas terhadap produksi perkebunan kelapa sawit di Indonesia. Agar dikemudian hari, produksi perkebunan kelapa sawit di Indonesia tidak menurun dan dapat berpengaruh terhadap perekonomian nasional terutama dari sisi penerimaan pajak dan pundi-pundi devisa negara.

II. LITERATURE REVIEW

A. Produksi Perkebunan Kelapa Sawit

Perkebunan merupakan bagian dari sektor pertanian. Produksi perkebunan diartikan sebagai usaha menciptakan nilai untuk masa kini dan masa depan melalui sektor perkebunan. Hasil dari perkebunan berupa buah, getah, rotan, dedaunan, dan batang pohon[2]. Kelapa sawit merupakan salah satu komoditas yang memiliki andil besar dalam menghasilkan pendapatan asli daerah, produk domestik bruto, dan kesejahteraan masyarakat. Kegiatan perkebunan kelapa sawit telah memberikan pengaruh eksternal yang bersifat positif atau bermanfaat bagi wilayah sekitarnya[3]. Perkebunan kelapa sawit merupakan komoditi strategis pada perekonomian nasional dan kesejahteraan masyarakat[4]. Luas areal produksi kelapa sawit telah tersebar di seluruh wilayah Indonesia.



Gambar 1 Luas Areal dan Produksi Kelapa Sawit Menurut Provinsi Tahun 2016
Sumber: Direktorat Jenderal Perkebunan, 2017

B. Algoritma K-Means Clustering

Data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting[5]. Berdasarkan fungsionalitas, *data mining* terbagi dalam 5 metode yaitu estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi. Salah satu metode pada data mining, yaitu *Clustering* yang merupakan metode pengelompokan data. Tujuan dari pengelompokan *cluster* ini adalah untuk menemukan pengelompokan dari serangkaian pola, titik, objek maupun dokumen. Objek yang berada didalam pengelompokan *cluster* yang sama memiliki kemiripan antar satu kelompok dan memiliki perbedaan dengan objek oleh kelompok *cluster* lain[6]. K-Means adalah sebuah metode *Clustering* yang termasuk dalam pendekatan *partitioning*. Algoritma K-Means merupakan model *Centroid*. *Mode Centroid* adalah model yang menggunakan *Centroid* untuk membuat *Cluster*. *Centroid* adalah titik tengah suatu *Cluster*. *Centroid* berupa *Centroid* nilai. *Centroid* digunakan untuk menghitung jarak suatu objek data terhadap *Centroid*. Suatu objek data termasuk dalam *Cluster* jika memiliki jarak terpendek terhadap *Centroid Cluster* tersebut. Algoritma K-Means dapat diartikan sebagai algoritma pembelajaran yang sederhana untuk memecahkan suatu permasalahan pengelompokan yang bertujuan untuk meminimalkan kesalahan ganda [7]. Dalam menentukan nilai *Centroid* untuk awal iterasi, nilai awal *Centroid* dilakukan secara acak. Sedangkan jika menentukan nilai *Centroid* yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Keterangan:

V_{ij} = *Centroid* rata-rata *cluster* ke-*i* untuk variabel ke-*j*

N_i = jumlah *cluster* ke -i
 i, k = indeks dari *cluster*
 j = indeks dari variabel
 X_{kj} = nilai data ke -k variabel ke -j dalam cluster

Menghitung jarak antara titik Centroid dengan titik tiap objek:

$$D = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Keterangan:

D = Euclidean Distance
 i = banyaknya objek
 (x,y) = koordinat objek
 (s,t) = koordinat *centroid*

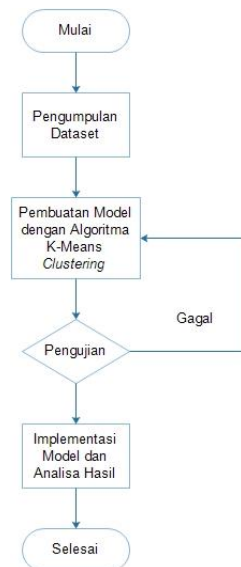
Metode *cross-validation* digunakan untuk menghindari *overlapping* pada data *testing*, adapun tahapan *cross-validation* adalah: a. Bagi data menjadi *subset* yg berukuran sama. b. Gunakan setiap *subset* untuk data *testing* dan sisanya untuk data *training*. *Cross-validation* biasa disebut juga dengan *K-fold cross-validation*, seringkali *subset* dibuat *stratified* (bertingkat) sebelum *cross-validation* dilakukan, karena stratifikasi akan mengurangi variansi dari estimasi[8].

C. RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. *RapidMiner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. *RapidMiner* sebelumnya bernama YALE (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*. *RapidMiner* didistribusikan di bawah lisensi AGPL (*GNU Affero General Public License*) versi 3. *RapidMiner* menyediakan GUI (*Graphic User Interface*) untuk merancang sebuah *pipeline* analitis. GUI ini akan menghasilkan file XML (*Extensible Markup Language*) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data. File ini kemudian dibaca oleh *RapidMiner* untuk menjalankan analisis secara otomatis[9].

III. RESEARCH METHOD

Dalam pembangunan sistem ini dimulai dari pengumpulan data produksi perkebunan kelapa sawit menurut provinsi (ribu ton) yang berasal dari Badan Pusat Statistik (BPS) pada tahun 2008-2020. Selanjutnya dilakukan pembuatan model dengan metode *clustering* menggunakan algoritma *k-means* di *RapidMiner*. Tujuan dilakukannya penelitian ini yaitu untuk mengelompokkan produksi perkebunan kelapa sawit menurut provinsi.



Gambar 2 Metode Penelitian

Teknik pengumpulan data pada penelitian ini ialah dokumentasi dan studi pustaka. Data diolah dengan proses pemilihan atau seleksi data, proses pembersihan data (membuang *missing value*, duplikasi data, dan memeriksa inkonsistensi data dan memperbaiki kesalahan pada data), dan proses transformasi (mengubah format data awal menjadi sebuah format data standar untuk proses pembacaan data pada program maupun *tool* yang digunakan). Pemodelan pada penelitian ini dilakukan dengan metode *clustering* memakai algoritma k-means. Penelitian ini diimplementasikan dengan *tools RapidMiner*. Validasi model pada penelitian ini menggunakan *K-fold cross-validation*.

IV. RESULTS AND DISCUSSION

Penelitian ini menggunakan algoritma k-means untuk mengelompokkan produksi perkebunan kelapa sawit menurut provinsi. Sumber data sebagai objek pada penelitian ini adalah data yang diambil dari Badan Pusat Statistik (BPS) pada tahun 2008-2020. Data yang digunakan dalam penelitian ini terdiri dari atribut atau variabel seperti provinsi dan tahun.

Tabel I Dataset Produksi Perkebunan Kelapa Sawit (Ribu Ton)

PROVINSI	TAHUN												
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
ACEH	564,7	693	616,5	592,2	654,8	817,53	945,6	896,3	732,7	867,3	1037,4	1133,3	1134,6
SUMATERA UTARA	3882,4	3862,4	3899,6	4010,7	3975,43	4549,2	4870,2	5193,1	3983,7	4852	5737,3	5647,3	5776,8
SUMATERA BARAT	961,5	896,3	985,9	927,4	930,12	1022,33	924,8	926,6	1183,1	1209,2	1248,3	1253,4	1312,3
RIAU	4812,9	5311,4	5496	5895,5	6384,54	6647	6993,2	8059,8	7668,1	7591,2	8496	9512,9	9984,3
JAMBI	1626,5	1499,9	1644,1	1773,1	1718,29	1749,62	1773,7	1794,9	1435,1	1783	2691,3	2884,4	3022,6
SUMATERA SELATAN	1891,4	2313,5	2542,8	2417,7	2492,9	2690,62	2791,8	2821,9	2929,5	2987	3793,6	4049,2	4267
BENGKULU	560,3	736	796	827,1	802,02	787,05	798,8	747,5	750,2	849,7	1047,7	1032,1	1063,4
LAMPUNG	416,3	389,3	405,7	424	433,82	424,05	455,9	434,3	425,9	456	487,2	414,2	384,9
KEP. BANGKA BELITUNG	412,9	446,6	490,2	504,6	504,6	508,13	516,6	523,1	726,6	756,1	900,3	815,7	843
KEP. RIAU	10,6	11,3	14,1	15,2	37,2	36,77	45	45,1	21,4	25,1	28,9	22,8	20
JAWA BARAT	12,6	20,7	16,4	22,1	22,72	32,64	33	35	32,8	42,4	46	32,2	33,1
BANTEN	25	25,1	26	21,8	27,67	27,08	24,3	25,5	27,5	32,1	38,4	31,3	27,4
KALIMANTAN BARAT	1124,4	1331,7	1426,9	1443,9	1830,86	1794,47	1965,5	2168,1	2192,6	2529	3086,9	5235,3	5471,4
KALIMANTAN TENGAH	1295,7	1798,1	1724,7	2499,3	2958,28	3127,14	3158,2	3573	4260,1	5209,8	7230,1	7664,8	7685,8
KALIMANTAN SELATAN	891,1	1041,4	1049,2	1045,2	1233,08	1244,04	1460,6	1049,5	1750,4	1560,1	1464,2	1665,4	1561,1
KALIMANTAN TIMUR	338,4	456,4	700	905,5	1298,14	1349,78	1407,3	1586,6	2358,4	2538,7	3786,5	3988,9	3823,2
SULAWESI TENGAH	126,6	144,3	145,8	185,1	234,08	244,07	254,4	275,3	316,8	417,1	383,6	381,7	371,7
SULAWESI SELATAN	21,5	28,2	34,9	42,7	47,18	49,82	78,9	111,5	105,1	113,1	105,7	91	100,3
SULAWESI BARAT	325,8	260,5	264,4	249,3	253,73	282,74	285,5	294,6	434,1	527,3	386,2	348,4	348
PAPUA BARAT	40	57,7	66,2	54,1	53,72	53,72	74	74	135,9	131,2	98,1	103,5	106,4
PAPUA	49,5	66,7	136,4	103,1	93,48	93,48	94	93,5	20,6	158,9	345,1	437,7	557,6

Pada perhitungan algoritma K-Means ini menggunakan dataset yang akan diolah yaitu sebanyak 21 provinsi yang akan dikelompokkan kedalam tiga *cluster* yaitu *low*, *medium*, dan *high*. Dengan pemodelan yang sudah

ditetapkan sebelumnya, maka berikut adalah contoh untuk perhitungan *Euclidean Distance* pada record ke 1 dari proses iterasi pertama:

$$D1(c1) =$$

$$\sqrt{(564,7-256)^2 + (693-294)^2 + (616,5-342)^2 + (592,2-369)^2 + (654,8-459)^2 + (817,53-474)^2 + (945,6-522)^2 + (896,3-498)^2 + (732,7-732)^2 + (867,3-778)^2 + (1037,4-938)^2 + (1133,3-1002)^2 + (1134,6-981)^2} = 966$$

$$D1(c2) =$$

$$\sqrt{(564,7-471)^2 + (693-575)^2 + (616,5-586)^2 + (592,2-704)^2 + (654,8-831)^2 + (817,53-850)^2 + (945,6-886)^2 + (896,3-972)^2 + (732,7-1098)^2 + (867,3-1293)^2 + (1037,4-1688)^2 + (1133,3-2031)^2 + (1134,6-2067)^2} = 1.578$$

$$D1(c3) =$$

$$\sqrt{(564,7-2043)^2 + (693-2188)^2 + (616,5-2283)^2 + (592,2-2349)^2 + (654,8-2423)^2 + (817,53-2609)^2 + (945,6-2728)^2 + (896,3-2920)^2 + (732,7-2669)^2 + (867,3-2877)^2 + (1037,4-3436)^2 + (1133,3-3645)^2 + (1134,6-3794)^2} = 7.128$$

Tabel II Euclidean Distance dan Matriks Kelompok Data Iterasi 1

C1	C2	C3	Jarak Terpendek	C1	C2	C3
966	1.578	7.128	966	T		
14.714	12.879	6.809	6.809			T
1.764	1.406	6.338	1.406		T	
24.093	22.185	16.202	16.202			T
5.074	3.243	3.041	3.041			T
8.564	6.677	855	855			T
1.041	1.699	7.141	1.041	T		
1.075	3.000	8.650	1.075	T		
385	2.157	7.902	385	T		
2.226	4.234	10.067	2.226	T		
2.206	4.214	10.051	2.206	T		
2.218	4.226	10.060	2.218	T		
7.782	5.769	3.332	3.332			T
14.160	12.153	7.476	7.476			T
2.645	1.521	5.442	1.521		T	
5.894	3.958	4.161	3.958		T	
1.305	3.324	9.157	1.305	T		
2.039	4.050	9.889	2.039	T		
1.220	3.227	8.982	1.220	T		
2.013	4.026	9.862	2.013	T		
1.604	3.547	9.462	1.604	T		
Index				12	3	6

Langkah berikutnya perlu ditentukan kembali titik kluster baru yang dihitung dengan mencari nilai rata-rata berdasarkan dari data anggota masing-masing kelompok kluster. Berikut adalah perhitungan untuk penentuan nilai titik kluster baru dari proses iterasi pertama:

$$C1 \text{ baru ke 2 (2008)} = \frac{2.566}{12} = 214$$

$$C1 \text{ baru ke 2 (2009)} = \frac{2.879}{12} = 240$$

C2 baru ke 2 (2008)	$= \frac{2.191}{3} = 730$	C2 baru ke 2 (2009)	$= \frac{2.394}{3} = 798$
C3 baru ke 2 (2008)	$= \frac{14.633}{6} = 2439$	C3 baru ke 2 (2009)	$= \frac{16.117}{6} = 2686$
C1 baru ke 2 (2010)	$= \frac{3.013}{12} = 251$	C1 baru ke 2 (2011)	$= \frac{3.041}{12} = 253$
C2 baru ke 2 (2010)	$= \frac{2.735}{3} = 912$	C2 baru ke 2 (2011)	$= \frac{2.878}{3} = 959$
C3 baru ke 2 (2010)	$= \frac{16.734}{6} = 2789$	C3 baru ke 2 (2011)	$= \frac{18.040}{6} = 3007$
C1 baru ke 2 (2012)	$= \frac{3.165}{12} = 264$	C1 baru ke 2 (2013)	$= \frac{3.357}{12} = 280$
C2 baru ke 2 (2012)	$= \frac{3.461}{3} = 1154$	C2 baru ke 2 (2013)	$= \frac{3.616}{3} = 1205$
C3 baru ke 2 (2012)	$= \frac{19.360}{6} = 3227$	C3 baru ke 2 (2013)	$= \frac{20.558}{6} = 3426$
C1 baru ke 2 (2014)	$= \frac{3.606}{12} = 301$	C1 baru ke 2 (2015)	$= \frac{3.556}{12} = 296$
C2 baru ke 2 (2014)	$= \frac{3.793}{3} = 1264$	C2 baru ke 2 (2015)	$= \frac{3.563}{3} = 1188$
C3 baru ke 2 (2014)	$= \frac{21.553}{6} = 3592$	C3 baru ke 2 (2015)	$= \frac{23.611}{6} = 3935$
C1 baru ke 2 (2016)	$= \frac{3.730}{12} = 311$	C1 baru ke 2 (2017)	$= \frac{4.376}{12} = 365$
C2 baru ke 2 (2016)	$= \frac{5.292}{3} = 1764$	C2 baru ke 2 (2017)	$= \frac{5.308}{3} = 1769$
C3 baru ke 2 (2016)	$= \frac{22.469}{6} = 3745$	C3 baru ke 2 (2017)	$= \frac{24.952}{6} = 4159$
C1 baru ke 2 (2018)	$= \frac{4.905}{12} = 409$	C1 baru ke 2 (2019)	$= \frac{4.844}{12} = 404$
C2 baru ke 2 (2018)	$= \frac{6.499}{3} = 2166$	C2 baru ke 2 (2019)	$= \frac{6.908}{3} = 2303$
C3 baru ke 2 (2018)	$= \frac{31.035}{6} = 5173$	C3 baru ke 2 (2019)	$= \frac{34.994}{6} = 5832$
C1 baru ke 2 (2020)	$= \frac{4.990}{12} = 416$	C1 baru ke 2 (2020)	$= \frac{36.208}{6} = 6035$
C2 baru ke 2 (2020)	$= \frac{6.697}{3} = 2232$		

Tabel III Titik Pusat Awal Klaster Baru ke-2

Titik pusat awal cluster	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Cluster Baru ke -1	214	240	251	253	264	280	301	296	311	365	409	404	416
Cluster Baru ke-2	730	798	912	959	1.154	1.205	1.264	1.188	1.764	1.769	2.166	2.303	2.232
Cluster Baru ke-3	2.439	2.686	2.789	3.007	3.227	3.426	3.592	3.935	3.745	4.159	5.173	5.832	6.035

Nilai titik klaster baru tersebut digunakan kembali untuk perhitungan iterasi ke-2, dengan langkah yang sama seperti sebelumnya, maka dibawah ini adalah contoh untuk perhitungan *Euclidean Distance* terhadap data ke-1 pada proses iterasi kedua:

D1(c1) =

$$\sqrt{(564,7-214)^2 + (693-240)^2 + (616,5-251)^2 + (592,2-253)^2 + (654,8-264)^2 + (817,53-280)^2 + (945,6-301)^2 + (896,3-296)^2 + (732,7-311)^2 + (867,3-365)^2 + (1037,4-409)^2 + (1133,3-404)^2 + (1134,6-416)^2} = 1.916$$

D1(c2) =

$$\sqrt{(564,7-730)^2 + (693-798)^2 + (616,5-912)^2 + (592,2-959)^2 + (654,8-1154)^2 + (817,53-1205)^2 + (945,6-1264)^2 + (896,3-1188)^2 + (732,7-1764)^2 + (867,3-1769)^2 + (1037,4-2166)^2 + (1133,3-2303)^2 + (1134,6-2232)^2} = 2.563$$

$D1(c3) =$

$$\sqrt{(564,7-2439)^2 + (693-2686)^2 + (616,5-2789)^2 + (592,2-3007)^2 + (654,8-3227)^2 + (817,53-3426)^2 + (945,6-3592)^2 + (896,3-3935)^2 + (732,7-3745)^2 + (867,3-4159)^2 + (1037,4-5173)^2 + (1133,3-5832)^2 + (1134,6-6035)^2} = 11.436$$

Tabel IV Euclidean Distance dan Matriks Kelompok Data Iterasi 2

C1	C2	C3	Jarak Terpendek	C1	C2	C3
1.916	2.563	11.436	1.916	T		
15.784	11.691	3.421	3.421			T
2.733	1.944	10.682	1.944		T	
25.211	20.984	12.020	12.020			T
6.161	2.255	7.264	2.255		T	
9.681	5.488	3.748	3.748			T
1.908	2.623	11.470	1.908	T		
488	4.058	12.984	488	T		
1.154	3.219	12.196	1.154	T		
1.044	5.383	14.367	1.044	T		
1.026	5.363	14.350	1.026	T		
1.036	5.374	14.360	1.036	T		
8.953	4.889	5.322	4.889		T	
15.371	11.081	3.957	3.957			T
3.697	1.253	9.769	1.253		T	
7.112	3.080	7.158	3.080		T	
208	4.456	13.448	208	T		
866	5.197	14.185	866	T		
252	4.318	13.293	252	T		
837	5.169	14.159	837	T		
626	4.743	13.720	626	T		
Index				12	5	4

Langkah berikutnya perlu ditentukan kembali titik klaster baru yang dihitung dengan mencari nilai rata-rata berdasarkan dari data anggota masing-masing kelompok klaster. Berikut adalah perhitungan untuk penentuan nilai titik klaster baru untuk proses iterasi ketiga:

$$\begin{aligned} \text{C1 baru ke 2 (2008)} &= \frac{2.566}{12} = 214 \\ \text{C2 baru ke 2 (2008)} &= \frac{4.942}{5} = 988 \\ \text{C3 baru ke 2 (2008)} &= \frac{11.882}{4} = 2971 \\ \text{C1 baru ke 2 (2010)} &= \frac{3.013}{12} = 251 \\ \text{C2 baru ke 2 (2010)} &= \frac{5.806}{5} = 1161 \\ \text{C3 baru ke 2 (2010)} &= \frac{13.663}{4} = 3416 \\ \text{C1 baru ke 2 (2012)} &= \frac{3.165}{12} = 264 \\ \text{C2 baru ke 2 (2012)} &= \frac{7.010}{5} = 1402 \end{aligned}$$

$$\begin{aligned} \text{C1 baru ke 2 (2009)} &= \frac{2.879}{12} = 240 \\ \text{C2 baru ke 2 (2009)} &= \frac{5.226}{5} = 1045 \\ \text{C3 baru ke 2 (2009)} &= \frac{13.285}{4} = 3321 \\ \text{C1 baru ke 2 (2011)} &= \frac{3.041}{12} = 253 \\ \text{C2 baru ke 2 (2011)} &= \frac{6.095}{5} = 1219 \\ \text{C3 baru ke 2 (2011)} &= \frac{14.823}{4} = 3706 \\ \text{C1 baru ke 2 (2013)} &= \frac{3.357}{12} = 280 \\ \text{C2 baru ke 2 (2013)} &= \frac{7.160}{5} = 1432 \end{aligned}$$

$$\begin{aligned}
 \text{C3 baru ke 2 (2012)} &= \frac{15.811}{4} = 3953 & \text{C3 baru ke 2 (2013)} &= \frac{17.014}{4} = 4253 \\
 \text{C1 baru ke 2 (2014)} &= \frac{3.606}{12} = 301 & \text{C1 baru ke 2 (2015)} &= \frac{3.556}{12} = 296 \\
 \text{C2 baru ke 2 (2014)} &= \frac{7.532}{12} = 1506 & \text{C2 baru ke 2 (2015)} &= \frac{7.526}{12} = 1505 \\
 \text{C3 baru ke 2 (2014)} &= \frac{17.813}{4} = 4453 & \text{C3 baru ke 2 (2015)} &= \frac{19.648}{4} = 4912 \\
 \text{C1 baru ke 2 (2016)} &= \frac{3.730}{12} = 311 & \text{C1 baru ke 2 (2017)} &= \frac{4.376}{12} = 365 \\
 \text{C2 baru ke 2 (2016)} &= \frac{8.920}{12} = 1784 & \text{C2 baru ke 2 (2017)} &= \frac{9.620}{12} = 1924 \\
 \text{C3 baru ke 2 (2016)} &= \frac{18.841}{4} = 4710 & \text{C3 baru ke 2 (2017)} &= \frac{20.640}{4} = 5160 \\
 \text{C1 baru ke 2 (2018)} &= \frac{4.905}{12} = 409 & \text{C1 baru ke 2 (2019)} &= \frac{4.844}{12} = 404 \\
 \text{C2 baru ke 2 (2018)} &= \frac{12.277}{12} = 2455 & \text{C2 baru ke 2 (2019)} &= \frac{15.027}{12} = 3005 \\
 \text{C3 baru ke 2 (2018)} &= \frac{25.257}{4} = 6314 & \text{C3 baru ke 2 (2019)} &= \frac{26.874}{4} = 6719 \\
 \text{C1 baru ke 2 (2020)} &= \frac{4.990}{12} = 416 & \text{C3 baru ke 2 (2020)} &= \frac{27.714}{4} = 6928 \\
 \text{C2 baru ke 2 (2020)} &= \frac{15.191}{5} = 3038 & &
 \end{aligned}$$

Tabel V Titik Pusat Awal Klaster Baru ke-3

Titik pusat awal cluster	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Cluster Baru ke -1	214	240	251	253	264	280	301	296	311	365	409	404	416
Cluster Baru ke-2	988	1.045	1.161	1.219	1.402	1.432	1.506	1.505	1.784	1.924	2.455	3.005	3.038
Cluster Baru ke-3	2.971	3.321	3.416	3.706	3.953	4.253	4.453	4.912	4.710	5.160	6.314	6.719	6.928

Nilai titik klaster baru tersebut digunakan kembali untuk perhitungan iterasi ke-3, dengan langkah yang sama seperti sebelumnya, maka dibawah ini adalah contoh untuk perhitungan *Euclidean Distance* terhadap data ke-1 pada proses iterasi ketiga:

D1(c1) =

$$\sqrt{(564,7-214)^2 + (693-240)^2 + (616,5-251)^2 + (592,2-253)^2 + (654,8-264)^2 + (817,53-280)^2 + (945,6-301)^2 + (896,3-296)^2 + (732,7-311)^2 + (867,3-365)^2 + (1037,4-409)^2 + (1133,3-404)^2 + (1134,6-416)^2} = 1.916$$

D1(c2) =

$$\sqrt{(564,7-988)^2 + (693-1045)^2 + (616,5-1161)^2 + (592,2-1219)^2 + (654,8-1402)^2 + (817,53-1432)^2 + (945,6-1506)^2 + (896,3-1505)^2 + (732,7-1784)^2 + (867,3-1924)^2 + (1037,4-2455)^2 + (1133,3-3005)^2 + (1134,6-3038)^2} = 3.738$$

D1(c3) =

$$\sqrt{(564,7-2971)^2 + (693-3321)^2 + (616,5-3416)^2 + (592,2-3706)^2 + (654,8-3953)^2 + (817,53-4253)^2 + (945,6-4453)^2 + (896,3-4912)^2 + (732,7-4710)^2 + (867,3-5160)^2 + (1037,4-6314)^2 + (1133,3-6719)^2 + (1134,6-6928)^2} = 14.429$$

Tabel VI Euclidean Distance dan Matriks Kelompok Data Iterasi 3

C1	C2	C3	Jarak Terpendek	C1	C2	C3
1.916	3.738	14.429	1.916	T		
15.784	10.560	2.285	2.285			T

C1	C2	C3	Jarak Terpendek	C1	C2	C3
2.733	3.095	13.659	2.733	T		
25.211	19.809	9.002	9.002			T
6.161	1.312	10.265	1.312		T	
9.681	4.332	6.690	4.332		T	
1.908	3.797	14.457	1.908	T		
488	5.272	15.970	488	T		
1.154	4.446	15.190	1.154	T		
1.044	6.593	17.366	1.044	T		
1.026	6.574	17.348	1.026	T		
1.036	6.585	17.359	1.036	T		
8.953	3.607	8.289	3.607		T	
15.371	9.897	4.205	4.205			T
3.697	2.328	12.733	2.328		T	
7.112	2.275	10.102	2.275		T	
208	5.676	16.445	208	T		
866	6.410	17.183	866	T		
252	5.545	16.284	252	T		
837	6.384	17.157	837	T		
626	5.920	16.732	626	T		
Index				13	5	3

Langkah berikutnya perlu ditentukan kembali titik klaster baru yang dihitung dengan mencari nilai rata-rata berdasarkan dari data anggota masing-masing kelompok klaster menggunakan rumus berikut dibawah ini:

C1 baru ke 2 (2008) = $\frac{3.527}{13} = 271$	C1 baru ke 2 (2009) = $\frac{3.776}{13} = 290$
C2 baru ke 2 (2008) = $\frac{5.872}{5} = 1174$	C2 baru ke 2 (2009) = $\frac{6.643}{5} = 1329$
C3 baru ke 2 (2008) = $\frac{9.991}{3} = 3330$	C3 baru ke 2 (2009) = $\frac{10.972}{3} = 3657$
C1 baru ke 2 (2010) = $\frac{3.999}{13} = 308$	C1 baru ke 2 (2011) = $\frac{3.969}{13} = 305$
C2 baru ke 2 (2010) = $\frac{7.363}{5} = 1473$	C2 baru ke 2 (2011) = $\frac{7.585}{5} = 1517$
C3 baru ke 2 (2010) = $\frac{11.120}{3} = 3707$	C3 baru ke 2 (2011) = $\frac{12.406}{3} = 4135$
C1 baru ke 2 (2012) = $\frac{4.095}{13} = 315$	C1 baru ke 2 (2013) = $\frac{4.379}{13} = 337$
C2 baru ke 2 (2012) = $\frac{8.573}{5} = 1715$	C2 baru ke 2 (2013) = $\frac{8.829}{5} = 1766$
C3 baru ke 2 (2012) = $\frac{13.318}{3} = 4439$	C3 baru ke 2 (2013) = $\frac{14.323}{3} = 4774$
C1 baru ke 2 (2014) = $\frac{4.531}{13} = 349$	C1 baru ke 2 (2015) = $\frac{4.482}{12} = 345$
C2 baru ke 2 (2014) = $\frac{9.399}{5} = 1880$	C2 baru ke 2 (2015) = $\frac{9.421}{5} = 1884$
C3 baru ke 2 (2014) = $\frac{15.022}{3} = 5007$	C3 baru ke 2 (2015) = $\frac{16.826}{3} = 5609$
C1 baru ke 2 (2016) = $\frac{4.913}{13} = 378$	C1 baru ke 2 (2017) = $\frac{5.586}{13} = 430$
C2 baru ke 2 (2016) = $\frac{10.666}{5} = 2133$	C2 baru ke 2 (2017) = $\frac{11.398}{5} = 2280$
C3 baru ke 2 (2016) = $\frac{15.912}{3} = 5304$	C3 baru ke 2 (2017) = $\frac{17.653}{3} = 5884$

$$\begin{aligned}
 \text{C1 baru ke 2 (2018)} &= \frac{6.153}{13} = 473 \\
 \text{C2 baru ke 2 (2018)} &= \frac{14.823}{5} = 2965 \\
 \text{C3 baru ke 2 (2018)} &= \frac{21.463}{3} = 7154 \\
 \text{C1 baru ke 2 (2020)} &= \frac{6.303}{12} = 485 \\
 \text{C2 baru ke 2 (2020)} &= \frac{18.145}{5} = 3629 \\
 \text{C1 baru ke 2 (2019)} &= \frac{6.097}{13} = 469 \\
 \text{C2 baru ke 2 (2019)} &= \frac{17.823}{5} = 3565 \\
 \text{C3 baru ke 2 (2019)} &= \frac{22.825}{3} = 7608 \\
 \text{C3 baru ke 2 (2020)} &= \frac{23.447}{3} = 7816
 \end{aligned}$$

Tabel VII Titik Pusat Awal Klaster Baru ke-4

Titik pusat awal cluster	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Cluster Baru ke -1	271	290	308	305	315	337	349	345	378	430	473	469	485
Cluster Baru ke-2	1.174	1.329	1.473	1.517	1.715	1.766	1.880	1.884	2.133	2.280	2.965	3.565	3.629
Cluster Baru ke-3	3.330	3.657	3.707	4.135	4.439	4.774	5.007	5.609	5.304	5.884	7.154	7.608	7.816

Nilai titik klaster baru tersebut digunakan kembali untuk perhitungan iterasi ke-4, dengan langkah yang sama seperti sebelumnya, maka dibawah ini adalah contoh untuk perhitungan *Euclidean Distance* terhadap data ke-1 pada proses iterasi keempat:

D1(c1) =

$$\sqrt{(564,7-271)^2 + (693-290)^2 + (616,5-308)^2 + (592,2-305)^2 + (654,8-315)^2 + (817,53-337)^2 + (945,6-349)^2 + (896,3-345)^2 + (732,7-378)^2 + (867,3-430)^2 + (1037,4-473)^2 + (1133,3-469)^2 + (1134,6-485)^2} = 1.712$$

D1(c2) =

$$\sqrt{(564,7-1174)^2 + (693-1329)^2 + (616,5-1473)^2 + (592,2-1517)^2 + (654,8-1715)^2 + (817,53-1766)^2 + (945,6-1880)^2 + (896,3-1884)^2 + (732,7-2133)^2 + (867,3-2280)^2 + (1037,4-2965)^2 + (1133,3-3565)^2 + (1134,6-3629)^2} = 5.103$$

D1(c3) =

$$\sqrt{(564,7-3330)^2 + (693-3657)^2 + (616,5-3707)^2 + (592,2-4135)^2 + (654,8-4439)^2 + (817,53-4774)^2 + (945,6-5007)^2 + (896,3-5609)^2 + (732,7-5304)^2 + (867,3-5884)^2 + (1037,4-7154)^2 + (1133,3-7608)^2 + (1134,6-7816)^2} = 16.651$$

Tabel VIII Euclidean Distance dan Matriks Kelompok Data Iterasi 4

C1	C2	C3	Jarak Terpendek	C1	C2	C3
1.712	5.103	16.651	1.712	T		
15.576	9.235	3.698	3.698			T
2.523	4.418	15.884	2.523	T		
25.003	18.428	6.864	6.864			T
5.957	1.410	12.488	1.410		T	
9.474	3.013	8.920	3.013		T	
1.701	5.152	16.683	1.701	T		
335	6.647	18.193	335	T		
950	5.830	17.412	950	T		
1.253	7.985	19.587	1.253	T		

C1	C2	C3	Jarak Terpendek	C1	C2	C3
1.235	7.966	19.570	1.235	T		
1.244	7.976	19.580	1.244	T		
8.766	2.526	10.397	2.526		T	
15.180	8.586	5.314	5.314			T
3.490	3.569	14.955	3.490	T		
6.934	2.025	12.233	2.025		T	
380	7.067	18.665	380	T		
1.075	7.802	19.404	1.075	T		
275	6.931	18.505	275	T		
1.045	7.775	19.379	1.045	T		
793	7.314	18.952	793	T		
Index				14	4	3

Langkah berikutnya perlu ditentukan kembali titik kluster baru yang dihitung dengan mencari nilai rata-rata berdasarkan dari data anggota masing-masing kelompok kluster menggunakan rumus berikut dibawah ini:

C1 baru ke 2 (2008) = $\frac{4.418}{14} = 316$	C1 baru ke 2 (2009) = $\frac{4.817}{14} = 344$
C2 baru ke 2 (2008) = $\frac{4.981}{4} = 1245$	C2 baru ke 2 (2009) = $\frac{5.602}{4} = 1400$
C3 baru ke 2 (2008) = $\frac{9.991}{3} = 3330$	C3 baru ke 2 (2009) = $\frac{10.972}{3} = 3657$
C1 baru ke 2 (2010) = $\frac{5.048}{14} = 361$	C1 baru ke 2 (2011) = $\frac{5.014}{14} = 358$
C2 baru ke 2 (2010) = $\frac{6.314}{4} = 1578$	C2 baru ke 2 (2011) = $\frac{6.540}{4} = 1635$
C3 baru ke 2 (2010) = $\frac{11.120}{3} = 3707$	C3 baru ke 2 (2011) = $\frac{12.406}{3} = 4135$
C1 baru ke 2 (2012) = $\frac{5.328}{14} = 381$	C1 baru ke 2 (2013) = $\frac{5.623}{14} = 402$
C2 baru ke 2 (2012) = $\frac{7.340}{4} = 1835$	C2 baru ke 2 (2013) = $\frac{7.584}{4} = 1896$
C3 baru ke 2 (2012) = $\frac{13.318}{3} = 4439$	C3 baru ke 2 (2013) = $\frac{14.323}{3} = 4774$
C1 baru ke 2 (2014) = $\frac{5.991}{14} = 428$	C1 baru ke 2 (2015) = $\frac{5.532}{14} = 395$
C2 baru ke 2 (2014) = $\frac{7.938}{4} = 1985$	C2 baru ke 2 (2015) = $\frac{8.372}{4} = 2093$
C3 baru ke 2 (2014) = $\frac{15.022}{3} = 5007$	C3 baru ke 2 (2015) = $\frac{16.826}{3} = 5609$
C1 baru ke 2 (2016) = $\frac{6.663}{14} = 476$	C1 baru ke 2 (2017) = $\frac{7.146}{14} = 510$
C2 baru ke 2 (2016) = $\frac{8.963}{4} = 2229$	C2 baru ke 2 (2017) = $\frac{9.838}{4} = 2459$
C3 baru ke 2 (2016) = $\frac{15.912}{3} = 5304$	C3 baru ke 2 (2017) = $\frac{17.653}{3} = 5884$
C1 baru ke 2 (2018) = $\frac{7.617}{14} = 544$	C1 baru ke 2 (2019) = $\frac{7.763}{14} = 554$
C2 baru ke 2 (2018) = $\frac{13.358}{4} = 3340$	C2 baru ke 2 (2019) = $\frac{16.158}{4} = 4039$
C3 baru ke 2 (2018) = $\frac{21.463}{3} = 7154$	C3 baru ke 2 (2019) = $\frac{22.825}{3} = 7608$
C1 baru ke 2 (2020) = $\frac{7.864}{14} = 562$	C3 baru ke 2 (2020) = $\frac{23.447}{3} = 7816$
C2 baru ke 2 (2020) = $\frac{16.584}{4} = 4146$	

Tabel IX Titik Pusat Awal Klaster Baru ke-5

Titik pusat awal cluster	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Cluster Baru ke -1	316	344	361	358	381	402	428	395	476	510	544	554	562
Cluster Baru ke-2	1.245	1.400	1.578	1.635	1.835	1.896	1.985	2.093	2.229	2.459	3.340	4.039	4.146
Cluster Baru ke-3	3.330	3.657	3.707	4.135	4.439	4.774	5.007	5.609	5.304	5.884	7.154	7.608	7.816

Nilai titik klaster baru tersebut digunakan kembali untuk perhitungan iterasi ke-5, dengan langkah yang sama seperti sebelumnya, maka dibawah ini adalah contoh untuk perhitungan *Euclidean Distance* terhadap data ke-1 pada proses iterasi kelima:

D1(c1) =

$$\sqrt{(564,7-316)^2+(693-344)^2+(616,5-361)^2+(592,2-358)^2+(654,8-381)^2+(817,53-402)^2+(945,6-428)^2+(896,3-395)^2+(732,7-476)^2+(867,3-510)^2+(1037,4-544)^2+(1133,3-554)^2+(1134,6-562)^2} = 1.474$$

D1(c2) =

$$\sqrt{(564,7-1245)^2+(693-1400)^2+(616,5-1578)^2+(592,2-1635)^2+(654,8-1835)^2+(817,53-1896)^2+(945,6-1985)^2+(896,3-2093)^2+(732,7-2229)^2+(867,3-2459)^2+(1037,4-3340)^2+(1133,3-4039)^2+(1134,6-4146)^2} = 5.969$$

D1(c3) =

$$\sqrt{(564,7-3330)^2+(693-3657)^2+(616,5-3707)^2+(592,2-4135)^2+(654,8-4439)^2+(817,53-4774)^2+(945,6-5007)^2+(896,3-5609)^2+(732,7-5304)^2+(867,3-5884)^2+(1037,4-7154)^2+(1133,3-7608)^2+(1134,6-7816)^2} = 16.651$$

Tabel X Euclidean Distance dan Matriks Kelompok Data Iterasi 5

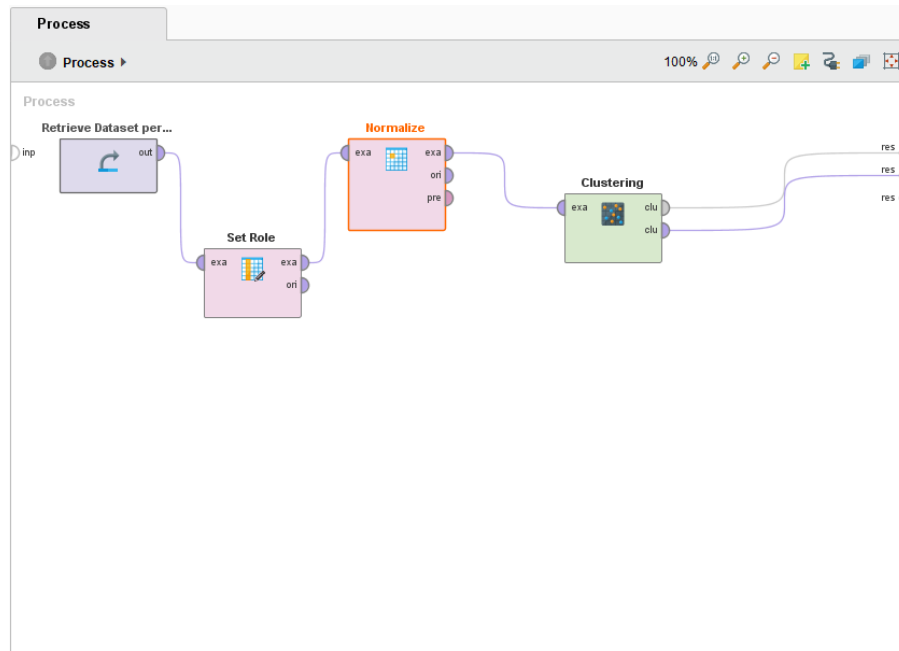
C1	C2	C3	Jarak Terpendek	C1	C2	C3
1.474	5.969	16.651	1.474	T		
15.333	8.568	3.698	3.698			T
2.278	5.299	15.884	2.278	T		
24.758	17.665	6.864	6.864			T
5.720	2.110	12.488	2.110		T	
9.230	2.456	8.920	2.456		T	
1.461	6.022	16.683	1.461	T		
290	7.514	18.193	290	T		
712	6.693	17.412	712	T		
1.498	8.837	19.587	1.498	T		
1.481	8.818	19.570	1.481	T		
1.490	8.828	19.580	1.490	T		
8.544	1.831	10.397	1.831		T	
14.952	7.731	5.314	5.314			T
3.240	4.462	14.955	3.240	T		
6.719	2.127	12.233	2.127		T	

C1	C2	C3	Jarak Terpendek	C1	C2	C3
607	7.922	18.665	607	T		
1.321	8.654	19.404	1.321	T		
452	7.794	18.505	452	T		
1.289	8.629	19.379	1.289	T		
1.013	8.153	18.952	1.013	T		
Index				14	4	3

Tabel XI Kelompok Kluster

PROVINSI	Kategori Kluster
ACEH	LOW
SUMATERA UTARA	HIGH
SUMATERA BARAT	LOW
RIAU	HIGH
JAMBI	MEDIUM
SUMATERA SELATAN	MEDIUM
BENGKULU	LOW
LAMPUNG	LOW
KEP. BANGKA BELITUNG	LOW
KEP. RIAU	LOW
JAWA BARAT	LOW
BANTEN	LOW
KALIMANTAN BARAT	MEDIUM
KALIMANTAN TENGAH	HIGH
KALIMANTAN SELATAN	LOW
KALIMANTAN TIMUR	MEDIUM
SULAWESI TENGAH	LOW
SULAWESI SELATAN	LOW
SULAWESI BARAT	LOW
PAPUA BARAT	LOW
PAPUA	LOW

Pada proses ini metode klasterisasi dengan algoritma K-Means diterapkan untuk pembentukan kelompok kluster dengan keakurasian yang tepat. Dalam penelitian ini menggunakan pengujian perhitungan dengan *tools* Rapid Miner, hasil pengujian yang didapat dengan menggunakan *tools* Rapid Miner adalah dengan tahapan langkah - langkah sebagai berikut:

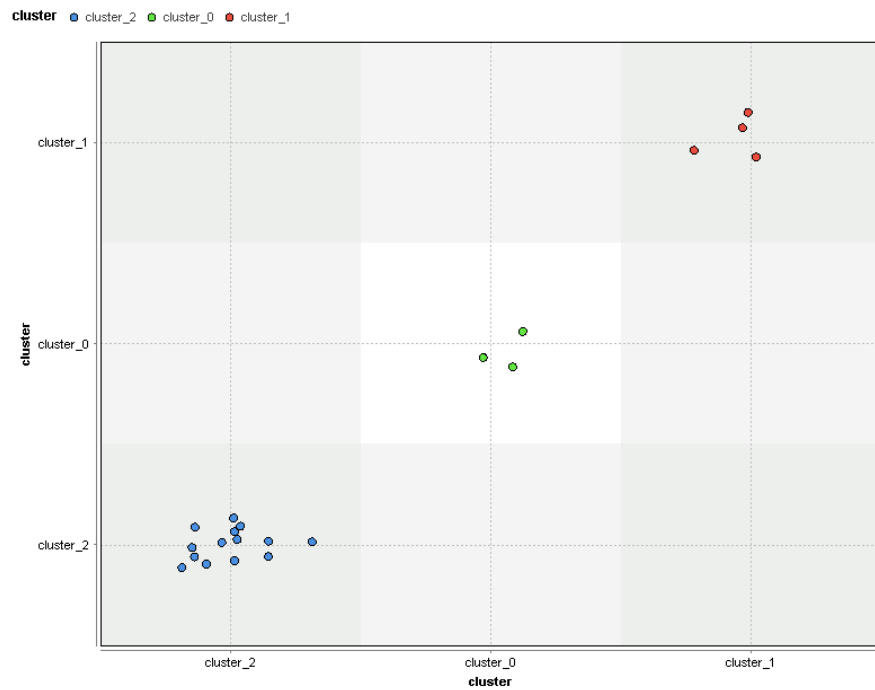


Gambar 3 Proses Rapid Miner Dengan Algoritma K-Means

Cluster Model

Cluster 0: 3 items
Cluster 1: 4 items
Cluster 2: 14 items
Total number of items: 21

Gambar 4 Cluster Model dari 217 record data



Gambar 5 Grafik Scatter Plot dari Kluster yang terbentuk

V. CONCLUSION

Dari hasil pengujian yang telah dilakukan dalam penelitian ini, maka dapat diambil suatu kesimpulan yaitu melalui beberapa tahapan didapatkan hasil bahwa proses *clustering* dengan algoritma K-Means berhenti pada iterasi ke-5, karena posisi objek dari masing – masing *cluster* sudah tidak berubah dan mendapatkan nilai yang optimal. Metode klasterisasi tersebut di proses dengan algoritma K-Means yang dimana hasilnya juga menunjukkan sebuah wawasan baru yaitu pengelompokan provinsi-provinsi yang menjadi prioritas terhadap produksi perkebunan kelapa sawit di indonesia berdasarkan 3 *cluster*. Cluster 1 merupakan kategori provinsi dengan produksi perkebunan kelapa sawit rendah atau *Low* yaitu 14 dari 21 kategori provinsi yang diuji, kemudian cluster 2 adalah kategori provinsi dengan produksi perkebunan kelapa sawit sedang atau *Medium* yaitu 4 dari 21 kategori provinsi yang diuji, dan terakhir adalah cluster 3 merupakan kategori provinsi dengan produksi perkebunan kelapa sawit tinggi atau *High* yaitu 3 dari 21 kategori provinsi yang diuji. Pengujian menggunakan *tools* RapidMiner juga dapat menghasilkan wawasan yang serupa yaitu masing-masing klaster memiliki anggota kelompok klaster sesuai dengan perhitungan manual seperti Cluster_0 pada Rapid Miner memiliki 3 anggota klaster yang merepresentasikan klaster *High*, Cluster_1 memiliki 4 anggota kelompok klaster sebagai representasi klaster *Medium*, dan Cluster_2 memiliki 14 anggota klaster yang sesuai dengan representasi klaster *Low*. Berdasarkan penelitian yang telah dilakukan bahwa provinsi sumatera utara, riau dan kalimantan tengah merupakan provinsi yang menjadi prioritas terhadap produksi perkebunan kelapa sawit di indonesia. Agar dikemudian hari, produksi perkebunan kelapa sawit di indonesia tidak menurun dan dapat berpengaruh terhadap perekonomian nasional terutama dari sisi penerimaan pajak dan pundi-pundi devisa negara.

REFERENCES

- [1] D. Haryadi and R. Mandala, "Prediksi Harga Minyak Kelapa Sawit Dalam Investasi Dengan Membandingkan Algoritma Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor," *IT Soc.*, vol. 4, no. 1, 2019, doi: 10.33021/itfs.v4i1.1181.
- [2] K. FIRMANSYAH, "DETERMINASI SUBSEKTOR PERKEBUNAN KARET DAN KELAPA SAWIT TERHADAP TINGKAT KEMISKINAN DI PROVINSI JAMBI TAHUN 2010-2018," *Orphanet J. Rare Dis.*, vol. 21, no. 1, pp. 1–9, 2020.
- [3] D. Anggraini, "ANALISIS PENGARUH PERKEBUNAN KELAPA SAWIT TERHADAP PEREKONOMIAN DI PROVINSI RIAU TAHUN 2002-2016," *Photosynthetica*, vol. 2, no. 1, pp. 1–13, 2018, [Online]. Available: <http://link.springer.com/10.1007/978-3-319-76887-8%0Ahttp://link.springer.com/10.1007/978-3-319-93594-2%0Ahttp://dx.doi.org/10.1016/B978-0-12-409517-5.00007-3%0Ahttp://dx.doi.org/10.1016/j.jff.2015.06.018%0Ahttp://dx.doi.org/10.1038/s41559-019-0877-3%0Aht>.
- [4] I. K. Saragih, D. Rachmina, and B. Krisnamurthi, "Analisis Status Keberlanjutan Perkebunan Kelapa Sawit Rakyat Provinsi Jambi," *J. Agribisnis Indones.*, vol. 8, no. 1, pp. 17–32, 2020, doi: 10.29244/jai.2020.8.1.17-32.
- [5] V. V. Galet Guntoro Setiaji, Khoirudin, "KOMPARASI METODE CLUSTERING K-MEANS DAN FUZZY C-MEANS UNTUK MEMPREDEKSI KETEPATAN WAKTULULUS," vol. 15, no. 1, pp. 38–42, 2019.
- [6] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [7] S. Hajar, A. A. Novany, A. P. Windarto, A. Wanto, and E. Irawan, "Penerapan K-Means Clustering Pada Ekspor Minyak Kelapa Sawit Menurut Negara Tujuan," pp. 314–318, 2020.
- [8] R. Mean *et al.*, "Evaluasi dan Validasi Evaluasi."
- [9] D. Aprilla, C. B. Donny, Aji, A. Lia, and W. I. Wayan, Simri, "DATA MINING dengan RAPIDMINER," *Produk-produk perangkat lunak gratis dan bersifat open source yang demikian banyak jumlahnya, telah memudahkan kita dalam melakukan proses Pengolah. dan Anal. data. Dalam melakukan Anal. terhadap data mining, RapidMiner merupakan salah sat*, pp. 1–128, 2013.