

Implementasi Metode K-Means Untuk Mengelompokkan Kawasan Potensi Pertanian Karet Produktif

Seberlina Laia¹, Deski Helsa Pane², Egi Affandi³

^{1,2,3}Program Studi Sistem Informasi, STMIK Triguna Dharma

Email: ¹seberlina@gmail.com, ²deski_pn2@gmail.com, ³egiaffandi@gmail.com

Email Penulis Korespondensi: sberlinn4@gmail.com

Abstrak

Nias Selatan yang mempengaruhi kelangsungan kegiatan pemberdayaan lahan potensial lahan pertanian karet dan harga karet yang cenderung terus menurun dikarenakan kualitas karet. Adapun masalah yang muncul dalam menentukan daerah pertanian produktif ialah biaya observasi yang mahal dan pembebasan secara administrasi dengan tanah adat yang sangat sulit. Permasalahan tersebut untuk mengetahui pengelompokan kawasan yang berpotensi pada tanaman karet masih belum berbentuk sistem ataupun masih begitu lambat, maka adapun bidang keilmuan yang digunakan adalah data mining dimana data-data tersebut akan diproses dengan menggunakan algoritma K-Means. Penerapan metode K-Means telah digunakan dalam berbagai kasus menganalisa data seperti yang digunakannya metode K-Means untuk mengetahui mengelompokkan kawasan potensi petani karet produktif. Sehingga Metode K-Means sangat sesuai dan dapat digunakan sebagai alat bantu dalam menganalisa mengelompokkan kawasan potensi petani karet produktif. Hasil penelitian ini bermanfaat bagi pemerintah desa dalam mendapatkan hasil yang maksimal dalam pengelompokkan data dalam menganalisa mengelompokkan kawasan potensi petani karet produktif dengan menggunakan metode K-Means yang lebih cepat dan akurat.

Kata Kunci: Data Mining, K-Means, Pengelompokan, Pertanian, Produksi Karet

1. PENDAHULUAN

Salah satu sebab banyaknya kerugian dalam pemberdayaan lahan pertanian karet dan petani perkebunan rakyat yang merugi akibat ketidak stabilan harga karet yakni mengenai tingginya biaya pembukaan lahan baru bagi pemerintah Nias Selatan yang mempengaruhi kelangsungan kegiatan pemberdayaan lahan potensial lahan pertanian karet dan harga karet yang cenderung terus menurun dikarenakan kualitas karet. Adapun masalah yang muncul dalam menentukan daerah pertanian produktif ialah biaya observasi yang mahal dan pembebasan secara administrasi dengan tanah adat yang sangat sulit. Oleh sebab itu, untuk mengetahui pengelompokan kawasan yang berpotensi pada tanaman karet masih belum berbentuk sistem ataupun masih begitu lambat, maka adapun bidang keilmuan yang digunakan adalah data mining dimana data-data tersebut akan diproses dengan menggunakan algoritma K-Means.

Data mining merupakan suatu proses yang digunakan untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan mendapatkan beberapa informasi penting dari suatu data [1]. Istilah pengenalan pola juga cocok digunakan karena pengetahuan yang akan digali memang berupa pola-pola yang mungkin juga masih perlu digali dari dalam potongan data yang sedang dihadapi [2]. Algoritma K-Means merupakan algoritma yang sederhana untuk diimplementasikan, memiliki kinerja yang relatif cepat, mudah beradaptasi, dan umum digunakan. K-Means merupakan salah satu algoritma clustering yang paling penting dalam bidang Data mining secara historis [3].

Dalam menganalisa mengelompokkan kawasan potensi petani karet produktif metode yang digunakan untuk pengelolaan *Data mining* pada kasus menganalisa kawasan berpotensi dan digunakan untuk pembentukan *cluster* adalah metode *K-Means*. Penerapan metode *K-Means* telah digunakan dalam berbagai kasus menganalisa data seperti yang digunakannya metode *K-Means* untuk mengetahui mengelompokkan kawasan potensi petani karet produktif. Sehingga Metode *K-Means* sangat sesuai dan dapat digunakan sebagai alat bantu dalam menganalisa mengelompokkan kawasan potensi petani karet produktif. Maka hasil proses dibuat dapat dikembangkan dengan menggunakan sistem berbasis komputer. Oleh sebab itu sistem yang dirancang dan dibangun menggunakan keilmuan *data mining* dengan metode *K-Means*. Dari sistem tersebut akan mendapatkan hasil yang maksimal dalam pengelompokkan data dalam menganalisa mengelompokkan kawasan potensi petani karet produktif dengan menggunakan metode *K-Means* yang lebih cepat dan akurat.

Metode *K-Means* digunakan untuk pengelompokan buku di perpustakaan Yayasan Nurul Islam Indonesia Baru [4], Penerapan algoritma k-means, mampu mempercepat dalam menentukan pengelompokan siswa kelas unggulan dengan menginputkan nilai centroid dan mencari nilai terdekat[5]. Hasil yang maksimal dalam pengelompokkan data kredit bermasalah dengan menggunakan algoritma k-means yang lebih efisien maupun efektif [6].

2. METODOLOGI PENELITIAN

2.1 Data mining

Berawal dari beberapa disiplin ilmu dan tujuan utama data mining adalah untuk menemukan, atau menambang pengetahuan dari data atau informasi yang kita miliki [7]. *Data mining*, sering juga disebut sebagai bagian dari

knowledge discovery in database (KDD). KDD merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam *set data* berukuran besar. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan [8].

2.2 Metode K-Means dengan teknik Clustering

Algoritma *K-means* merupakan salah satu algoritma dengan partitional, karena *K-Means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya [9]. *Algoritma K-means* menggunakan proses secara berulang-ulang untuk mendapatkan basis data *cluster*. Metode *K-Means* dikenal sebagai suatu metode pengelompokan yang paling terkenal diberbagai bidang karena sangat mudah diimplementasi dan mampu klasterisasi penjualan data dalam jumlah yang banyak. Algoritme *K-Means* juga sangat akurat dalam pengelompokan data [10].

Tahapan dalam perhitungan algoritme *K-Means* yaitu tentukan banyaknya *cluster* berdasarkan *cluster* data yang akan dibentuk. Selanjutnya tentukan *centroid* atau pusat *cluster* awal secara acak. Kemudian hitung setiap data dan kelompokkan ke *cluster* terdekat berdasarkan jarak antara data dengan pusat *cluster* awal. Lalu Hitung kembali *centroid* dengan keanggotaan *cluster* yang sekarang. Selanjutnya pada setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses perhitungan selesai [11]:

Langkah- langkah melakukan *clustering* dengan metode *K-means* adalah sebagai berikut [12]:

1. Menentukan jumlah *Cluster* misalkan sebanyak $k = 3$
2. Alokasikan data ke dalam kelompok secara acak.
3. Hitung pusat kelompok (*centroid* / rata-rata) dari data yang ada di masing-masing kelompok. Lokasi *centroid* setiap kelompok diambil dari rata-rata (mean) semua nilai data pada setiap fiturnya. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke- i dalam sebuah kelompok, dan p menyatakan dimensi data, maka persamaan untuk menghitung *centroid* fitur ke- i digunakan persamaan 1.

$$d(x_i, \mu_j) = \sqrt{\sum (x_i, \mu_j)^2}$$

4. Alokasikan masing-masing data ke *centroid* / rata-rata terdekat. Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok, diantaranya adalah *Euclidean*. Pengukuran jarak pada ruang jarak (*distance space*) *Euclidean* dapat dicari menggunakan persamaan 2.

$$d = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

5. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok atau apabila ada perubahan nilai *centroid* di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data Penilaian

Berikut ini adalah deskripsi data yang diuji setelah dilakukan pengumpulan data melalui wawancara pihak pengolah Kawasan Karet Selatan pada Nias Selatan.

Tabel 1. Daerah / Desa Nias Selatan

No	Kode Daerah	Nama Daerah	Masa Panen	Sudah Panen
1	A001	Hilimazihono	330	315
2	A002	Hiligumandu	58	53
3	A003	Lolomaya	48	45
4	A004	Lolowau	202	202
5	A005	Amandraya	172	166
6	A006	Gomo	90	81
7	A007	Hilimegai	102	92
8	A008	Simandaolo	81	81
9	A009	Hilibadalu	63	57
10	A010	Hililwuo	95	81
11	A011	Lolomatau	64	58
12	A012	Lahusa	151	146
13	A013	Susua	170	162
14	A014	Goo	133	128
15	A015	Hilinduria	140	132
16	A016	Onohazumba	330	325
17	A017	Toma	121	115

18	A018	Hilimbuan	40	36
...
78	A077	Talio	72	61

3.1.1 Menentukan Inisialisasi Jumlah Cluster Dan Titik Awal Cluster

Sebelum melakukan perhitungan terlebih dahulu harus menentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini data-data yang ada akan dikelompokkan menjadi tiga *cluster*. Selanjutnya adalah menentukan titik pusat awal dari setiap *cluster*, maka untuk titik pusat *cluster* yang dipilih dapat dilihat pada tabel 2. sebagai berikut:

Tabel 2. Contoh Data Titik Pusat *Cluster* pada iterasi-1

Titik Pusat Awal	Nama Daerah	Masa Panen	Sudah Panen
<i>Cluster 1</i>	Hiliamaetaniha	460	402
<i>Cluster 2</i>	Fuge	92	81
<i>Cluster 3</i>	Fanedanu	20	18

3.1.2 Inisialisasi Ke Pusat Cluster

Inisialisasi ke pusat *cluster* dengan menghitung jarak data ke *Centroid* menggunakan rumus *Euclidean*, data tersebut dari *Cluster* terdekatnya.

- a. Menghitung data pertama dengan titik *cluster 1*

$$C1 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$C1 = \sqrt{(330-460)^2 + (315-402)^2}$$

$$C1 = 156,426$$

- b. Menghitung data pertama dengan titik *cluster 2*

$$C2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$C2 = \sqrt{(330-92)^2 + (315-81)^2}$$

$$C2 = 333,766$$

- c. Menghitung data pertama dengan titik *cluster 3*

$$C3 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$C3 = \sqrt{(330-20)^2 + (315-18)^2}$$

$$C3 = 429,312$$

3.1.3 Klasterisasi Kawasan Setiap Data Centroid

Berikut tabel dibawah ini dapat dilihat untuk hasil lengkap dari perhitungan pada tahap iterasi pertama:

Tabel 3. Hasil Perhitungan Iterasi ke-1

No	Kode Daerah	C1	C2	C3	JARAK TERDEKAT	Berada Di Cluster
1	A001	156,426	333,766	429,312	156,426	1
2	A002	532,358	44,045	51,662	44,045	2
3	A003	545,154	56,851	38,897	38,897	3
4	A004	326,441	163,527	258,805	163,527	2
5	A005	372,344	116,726	212,151	116,726	2
6	A006	489,838	2,000	94,175	2,000	2
7	A007	473,565	14,866	110,454	14,866	2
8	A008	496,671	11,000	87,693	11,000	2
9	A009	525,960	37,643	58,052	37,643	2
10	A010	486,072	3,000	97,949	3,000	2
11	A011	524,549	36,235	59,464	36,235	2
12	A012	401,269	87,784	183,153	87,784	2
13	A013	376,431	112,450	207,933	112,450	2
14	A014	426,620	62,370	157,699	62,370	2
15	A015	418,688	70,036	165,517	70,036	2
16	A016	151,093	340,852	436,290	151,093	1
17	A017	444,173	44,688	140,036	44,688	2

18	A018	557,096	68,768	26,907	26,907	3
...
78	A078	516,551	28,284	67,476	28,284	2

Dari tabel 3. diatas dapat hasil iteras ke-1 sebagai berikut:

- C1= { A001,A016,A045}
 - C2={ A002, A004, A005, A006, A007, A008, A009, A010, A011, A012, A013, A014, A015, A017, A020, A023, A026, A027, A034, A035, A038, A040, A046, A049, A053, A056, A063, A072, A073, A074, A075, A078}
 - C3={ A003, A018, A019, A021, A022, A024, A025, A028, A029, A030, A031, A032, A033, A036, A037, A039, A041, A042, A043, A044, A047, A048, A050, A051, A052, A054, A055, A057, A058, A059, A060, A061, A062, A064, A065, A066, A067, A068, A069, A070, A071, A076}
1. Hitung nilai WCV (*Within Cluster Variation*) dengan cara memangkatkan jarak terdekat *cluster* dan menjumlahkan setiap nilai WCV.

$$WCV = 156,426 + 44,045 + 38,897 + 163,527 + \dots + 28,284$$

$$WCV = 3171,189$$
 2. Hitung nilai BCV (*Between Cluster Variation*) dengan cara menjumlahkan hasil dari jarak diantara setiap *centroid*.
 - a.
$$D(m1,m2) = \sqrt{(m1-m2)^2}$$

$$= \sqrt{(460-92)^2+(402-81)^2}$$

$$= 488,329$$
 - b.
$$D(m1,m3) = \sqrt{(m1-m3)^2}$$

$$= \sqrt{(460-20)^2+(402-18)^2}$$

$$= 95,671$$
 - c.
$$D(m2-m3) = \sqrt{(m2-m3)^2}$$

$$= \sqrt{(92-20)^2+(81-18)^2}$$

$$= 584,000$$

Nilai BCV = d(m1,m2) + d(m1,m3)+d(m2,m3)
 = 488,329 + 95,671 + 584,000
 = 1168,000
 3. Menghitung nilai besar rasio dengan membandingkan nilai BCV dan WCV.

$$BCV/WCV = 1168,000 / 3171,189$$

$$= 0,368$$

3.1.4 Memperbarui Nilai Centroid Baru

Kemudian langkah berikutnya pada iterasi ke-2 adalah Menghitung kembali pusat-pusat baru dari *cluster* yang baru terbentuk dengan menghitung nilai rata-rata pada *cluster*.

Diketahui :

s = Masa panen

t = Sudah panen

1. Pada *cluster* 1 terdapat 3 data maka nilai rata-rata pada *cluster*-1 adalah

$$s = (330 + 330 + 460)/3 = 373,333$$

$$t = (315 + 325 + 402)/3 = 347,333$$
2. Pada *cluster* 2 terdapat 9 data maka nilai rata-rata pada *cluster*-2 adalah

$$s = (58 + 202 + 172 + 90 + 102 + 81 + 63 + 95 + 64 + 151 + 170 + 133 + 140 + 121 + 170 + 151 + 90 + 70 + 92 + 80 + 80 + 55 + 65 + 62 + 173 + 160 + 58 + 177 + 152 + 153 + 201 + 72) / 32 = 175,333$$

$$t = (53 + 202 + 166 + 81 + 92 + 81 + 57 + 81 + 58 + 146 + 162 + 128 + 132 + 115 + 156 + 140 + 88 + 62 + 81 + 77 + 73 + 52 + 61 + 55 + 160 + 156 + 51 + 177 + 140 + 150 + 201 + 61) / 32 = 170,000$$
3. Pada *cluster* 3 terdapat 67 data maka nilai rata-rata pada *cluster*-3 adalah

$$s = (48 + 40 + 26 + 31 + 20 + 30 + 22 + 25 + 31 + 21 + 25 + 20 + 52 + 36 + 43 + 35 + 36 + 30 + 20 + 35 + 50 + 45 + 57 + 35 + 45 + 30 + 45 + 31 + 40 + 33 + 40 + 43 + 50 + 44 + 40 + 34 + 56 + 40 + 35 + 22 + 36 + 35 + 40) / 43 = 36,093$$

$$t = (45 + 36 + 25 + 22 + 19 + 27 + 20 + 20 + 29 + 18 + 21 + 18 + 50 + 34 + 41 + 36 + 31 + 25 + 19 + 34 + 42 + 39 + 44 + 32 + 37 + 24 + 41 + 27 + 35 + 28 + 35 + 41 + 47 + 40 + 35 + 27 + 48 + 39 + 33 + 20 + 34 + 32 + 35) / 43 = 32,209$$

Tabel 4. Titik Pusat *Cluster* pada iterasi-2

Titik Pusat Awal	Masa Panen	Sudah Panen
<i>Cluster 1</i>	373,333	347,333
<i>Cluster 2</i>	175,333	170,000
<i>Cluster 3</i>	36,093	32,209

3.1.5 Perulangan Iterasi

Apabila data setiap *cluster* belum berhenti, lakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah. Nilai *Centroid* berubah dari nilai *Centroid* sebelumnya, maka algoritma dilanjutkan ke langkah berikutnya. Setelah diketahui nilai *k* dan pusat *cluster* awal selanjutnya mengukur jarak antara pusat *cluster* menggunakan *Euclidian distance*, kemudian akan didapatkan matriks jarak yaitu C1, C2 dan C3 sebagai berikut:

- a. Menghitung data pertama dengan titik *cluster 1*

$$C1 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$C1 = \sqrt{(330-373,333)^2 + (315-347,333)^2}$$

$$C1 = 54,067$$

- b. Menghitung data pertama dengan titik *cluster 2*

$$C2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$C2 = \sqrt{(330-175,333)^2 + (315-170,000)^2}$$

$$C2 = 212,007$$

- c. Menghitung data pertama dengan titik *cluster 3*

$$C3 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$C3 = \sqrt{(330 - 36,093)^2 + (315 - 32,209)^2}$$

$$C3 = 407,863$$

Berikut tabel dibawah ini dapat dilihat untuk hasil lengkap dari perhitungan *Euclidean Distance* pada tahap iterasi ke-2 :

Tabel 5. Hasil Perhitungan Iterasi ke-2

No	Kode Daerah	C1	C2	C3	JARAK TERDEKAT	Berada Di Cluster
1	A001	54,067	212,007	407,863	54,067	1
2	A002	431,355	165,699	30,202	30,202	3
3	A003	444,125	178,434	17,475	17,475	3
4	A004	224,671	41,655	237,390	41,655	2
5	A005	270,956	5,207	190,711	5,207	2
6	A006	388,859	123,300	72,708	72,708	3
7	A007	372,581	107,060	88,987	88,987	3
8	A008	395,465	129,691	66,311	66,311	3
9	A009	424,971	159,335	36,586	36,586	3
10	A010	385,231	119,893	76,489	76,489	3
11	A011	423,557	157,921	37,999	37,999	3
12	A012	299,945	34,178	161,716	34,178	2
13	A013	275,123	9,615	186,485	9,615	2
14	A014	325,372	59,633	136,260	59,633	2
15	A015	317,510	51,889	144,065	51,889	2
16	A016	48,750	218,968	414,859	48,750	1
17	A017	343,003	77,312	118,590	77,312	2
18	A018	456,114	190,450	5,444	5,444	3
19	A019	473,856	208,148	12,403	12,403	3
...
78	A078	415,678	150,196	46,024	46,024	3

Dari tabel 5. diatas dapat hasil iteras ke-2 sebagai berikut:

- C1= { A001, A016, A045 }

- C2={ A004, A005, A012, A013, A014, A015, A017, A020, A023, A053, A056, A072, A073, A074, A075}
- C3={ A002, A003, A006, A007, A008, A009, A010, A011, A018, A019, A021, A022, A024, A025, A026, A027, A028, A029, A030, A031, A032, A033, A034, A035, A036, A037, A038, A039, A040, A041, A042, A043, A044, A046, A047, A048, A049, A050, A051, A052, A054, A055, A057, A058, A059, A060, A061, A062, A063, A064, A065, A066, A067, A068, A069, A070, A071, A076, A077, A078}

1. Hitung nilai WCV (*Within Cluster Variation*) dengan cara memangkatkan jarak terdekat *cluster* dan menjumlahkan setiap nilai WCV.

$$WCV = 54,067 + 30,202 + 17,475 + 41,655 \dots + 46,024$$

$$WCV = 2062,534$$

2. Hitung nilai BCV (*Between Cluster Variation*) dengan cara menjumlahkan hasil dari jarak diantara setiap *centroid*.

$$\begin{aligned} \text{a. } D(m1,m2) &= \sqrt{(m1-m2)^2} \\ &= \sqrt{(373,333-175,333)^2+(347,333-170,000)^2} \\ &= 265,803 \end{aligned}$$

$$\begin{aligned} \text{b. } D(m1,m3) &= \sqrt{(m1-m3)^2} \\ &= \sqrt{(373,333- 36,093)^2+(347,333- 32,209)^2} \\ &= 195,893 \end{aligned}$$

$$\begin{aligned} \text{c. } D(m2-m3) &= \sqrt{(m2-m3)^2} \\ &= \sqrt{(175,333- 36,093)^2+(170,000- 32,209)^2} \\ &= 461,556 \end{aligned}$$

$$\begin{aligned} \text{Nilai BCV} &= d(m1,m2) + d(m1,m3)+d(m2,m3) \\ &= 265,803 + 195,893 + 461,556 \\ &= 923,252 \end{aligned}$$

3. Menghitung nilai besar rasio dengan membandingkan nilai BCV dan WCV.

$$\begin{aligned} BCV/WCV &= 923,252 / 2062,534 \\ &= 0,448 \end{aligned}$$

Kemudian langkah berikutnya pada iterasi ke-3 adalah Menghitung kembali pusat-pusat baru dari *cluster* yang baru terbentuk dengan menghitung nilai rata-rata pada *cluster*.

a. Pada *cluster 1* terdapat 3 data maka nilai rata-rata pada *cluster-1* adalah

$$S = (330 + 330 + 460) / 3 = 373,333$$

$$T = (315 + 325 + 402) / 3 = 347,333$$

b. Pada *cluster 2* terdapat 15 data maka nilai rata-rata pada *cluster-2* adalah

$$S = (202 + 172 + 151 + 170 + 133 + 140 + 121 + 170 + 151 + 173 + 160 + 177 + 152 + 153 + 201) / 15 = 161,73$$

$$T = (202 + 166 + 146 + 162 + 128 + 132 + 115 + 156 + 140 + 160 + 156 + 177 + 140 + 153 + 201) / 15 = 155,40$$

c. Pada *cluster 3* terdapat 61 data maka nilai rata-rata pada *cluster-3* adalah

$$S = 58 + 48 + 90 + 102 + 81 + 63 + 95 + 64 + 40 + 26 + 31 + 20 + 30 + 22 + \dots + 72) / 61 = 47,15$$

$$T = 53 + 45 + 81 + 92 + 81 + 57 + 81 + 58 + 36 + 25 + 22 + 19 + 27 + 20 + \dots + 61) / 61 = 42,48$$

Tabel 6. Titik Pusat *Cluster* pada iterasi-3

TITIK PUSAT AWAL	DAERAH / DESA	KAWASAN
<i>Cluster 1</i>	373,333	347,333
<i>Cluster 2</i>	161,733	155,400
<i>Cluster 3</i>	47,150	42,483

Setelah diketahui nilai k dan pusat *cluster* awal selanjutnya mengukur jarak antara pusat *cluster* menggunakan *Euclidian distance*, kemudian akan didapatkan matriks jarak yaitu C1, C2 dan C3 sebagai berikut:

Menghitung jarak antara data dan titik *cluster* yang sudah dipilih

a. Menghitung data pertama dengan titik *cluster 1*

$$C1 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$C1 = \sqrt{(330-373,333)^2+(315-347,333)^2}$$

$$C1 = 54,067$$

b. Menghitung data pertama dengan titik *cluster 2*

$$C2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$C2 = \sqrt{(330-161,733)^2 + (315-155,400)^2}$$

$$C2 = 231,918$$

c. Menghitung data pertama dengan titik *cluster* 3

$$C3 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$C3 = \sqrt{(330-47,150)^2 + (315-42,483)^2}$$

$$C3 = 392,772$$

Berikut tabel dibawah ini dapat dilihat untuk hasil lengkap dari perhitungan *Euclidean Distance* pada tahap iterasi ke-3:

Tabel 7. Hasil Perhitungan Iterasi ke-3

No	C1	C2	C3	Jarak Terdekat	Berada Di Cluster
1	54,067	231,918	392,772	54,067	1
2	431,355	145,761	15,110	15,110	3
3	444,125	158,504	2,656	2,656	3
4	224,671	61,587	222,315	61,587	2
5	270,956	14,757	175,624	14,757	2
6	388,859	103,349	57,616	57,616	3
7	372,581	87,107	73,895	73,895	3
8	395,465	109,787	51,277	51,277	3
9	424,971	139,395	21,493	21,493	3
10	385,231	99,943	61,426	61,426	3
11	423,557	137,980	22,906	22,906	3
12	299,945	14,268	146,631	14,268	2
13	275,123	10,578	171,395	10,578	2
14	325,372	39,703	121,175	39,703	2
15	317,510	31,936	128,974	31,936	2
16	48,750	238,910	399,775	48,750	1
17	343,003	57,370	103,501	57,370	2
18	456,114	170,515	9,652	9,652	3
19	473,856	188,222	27,441	27,441	3
...
78	415,678	130,244	30,990	30,990	3

Dari tabel 7. diatas dapat hasil iterasi ke-3 sebagai berikut:

- C1= {A001,A016 dan A045}
 - C2={ A004, A005, A012, A013, A014, A015, A017, A020, A023, A053, A056, A072, A073, A074, A075}
 - C3={ A002, A003, A006, A007, A008, A009, A010, A011, A018, A019, A021, A022, A024, A025, A026, A027, A028, A029, A030, A031, A032, A033, A034, A035, A036, A037, A038, A039, A040, A041, A042, A043, A044, A046, A047, A048, A049, A050, A051, A052, A054, A055, A057, A058, A059, A060, A061, A062, A063, A064, A065, A066, A067, A068, A069, A070, A071, A076, A077, A078}
1. Hitung nilai WCV (*Within Cluster Variation*) dengan cara memangkatkan jarak terdekat *cluster* dan menjumlahkan setiap nilai WCV.
 $WCV = 59,544^2 + 15,152^2 + 2,711^2 + 61,617^2 + 14,802^2 + 57,656^2 + \dots + 31,026^2$
 $WCV = 1352,243$
 2. Hitung nilai BCV (*Between Cluster Variation*) dengan cara menjumlahkan hasil dari jarak diantara setiap *centroid*.
 - a. $D(m1,m2) = \sqrt{(m1-m2)^2}$
 $= \sqrt{(373,333-161,733)^2 + (347,333-155,400)^2}$
 $= 285,680$
 - b. $D(m1,m3) = \sqrt{(m1-m3)^2}$
 $= \sqrt{(373,333-47,15)^2 + (347,333-42,115)^2}$
 $= 160,871$
 - c. $D(m2-m3) = \sqrt{(m2-m3)^2}$
 $= \sqrt{(161,733-47,150)^2 + (155,400-42,483)^2}$
 $= 446,463$

$$\begin{aligned} \text{Nilai } BCV &= d(m1,m2) + d(m1,m3)+d(m2,m3) \\ &= 285,680 + 160,871 + 446,463 \\ &= 893,014 \end{aligned}$$

3. Menghitung nilai besar rasio dengan membandingkan nilai *BCV* dan *WCV*.

$$\begin{aligned} BCV/WCV &= 893,014 / 1352,243 \\ &= 0,683 \end{aligned}$$

Dari perhitungan telah dilakukan dengan menerapkan *algoritma K-Means*, karena jumlah objek pada iterasi kedua dan ketiga memiliki kesamaan maka perhitungan *iterasi* berhenti di iterasi ketiga. Pada tahap ini dapat diketahui hasil klusterisasi menggunakan *algoritma K-Means Clustering* untuk klusterisasi Kawasan karet yang paling Baik Kawasan Produksi Karet, cukup Baik Kawasan Produksi Karet dan kurang Baik Kawasan Produksi Karet di kawasan nias selatan.

Tabel 8. Daerah Yang Paling Baik Kawasan Produksi Karet

No	Kode Daerah	Nama Daerah	JARAK TERDEKAT
1	A001	Hilimazihono	Paling Baik
16	A016	Onohazumba	Paling Baik
45	A045	Hiliamaetaniha	Paling Baik
4	A004	Lolowau	Baik
5	A005	Amandraya	Baik
12	A012	Lahusa	Baik
13	A013	Susua	Baik
14	A014	Goo	Baik
15	A015	Hilinduria	Baik
17	A017	Toma	Baik
20	A020	Telukdalam	Baik
23	A023	Bawolahusa	Baik
53	A053	Hilifalago	Baik
56	A056	Koendrafo	Baik
72	A072	Ramba-ramba	Baik
73	A073	Rapa-rapa Melayu	Baik
74	A074	Saeru Melayu	Baik
75	A075	Sambulu	Baik

Tabel 9. Daerah Yang Kurang Baik Produksi Karet

No	Kode Daerah	Nama Daerah	JARAK TERDEKAT
2	A002	Hiligumandu	Kurang Baik
3	A003	Lolomaya	Kurang Baik
6	A006	Gomo	Kurang Baik
7	A007	Hilimegai	Kurang Baik
8	A008	Simandaolo	Kurang Baik
9	A009	Hilibadalu	Kurang Baik
10	A010	Hililwuo	Kurang Baik
11	A011	Lolomatau	Kurang Baik
18	A018	Hilimbuan	Kurang Baik
19	A019	Jowo	Kurang Baik
21	A021	Bawogosali	Kurang Baik
22	A022	Bawohosi	Kurang Baik
24	A024	Bawolato	Kurang Baik
25	A025	Bawolowalani	Kurang Baik
26	A026	Bawonahono	Kurang Baik
27	A027	Doli-doli Idanotae	Kurang Baik
28	A028	Duru	Kurang Baik
29	A029	Eho	Kurang Baik
30	A030	Ehosakhozi	Kurang Baik
31	A031	Fadoro Ewo	Kurang Baik
32	A032	Fanedanu	Kurang Baik
33	A033	Fondrakoraya	Kurang Baik

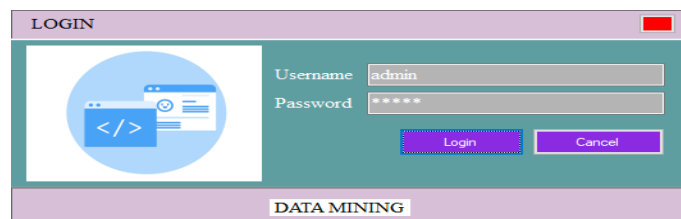
34	A034	Fuge	Kurang Baik
35	A035	Gobo	Kurang Baik
36	A036	Golambanua	Kurang Baik
37	A037	Golambanua II	Kurang Baik
38	A038	Gondia	Kurang Baik
39	A039	Hayo	Kurang Baik
40	A040	Helefanikha	Kurang Baik
41	A041	Hili Amo Dula	Kurang Baik
42	A042	Hili'alawa	Kurang Baik
43	A043	Hiliadulo	Kurang Baik
44	A044	Hiliametaluo	Kurang Baik
46	A046	Hiliamauzula	Kurang Baik
47	A047	Hiliana'a Gomo	Kurang Baik
48	A048	Hiliana'a Susua	Kurang Baik
49	A049	Hilianombasela	Kurang Baik
50	A050	Hilidohona	Kurang Baik
51	A051	Hilifadolo	Kurang Baik
52	A052	Hilifadolo	Kurang Baik
54	A054	Hilifalawu	Kurang Baik
55	A055	Hilifondege Hilizoroilawa	Kurang Baik
57	A057	Koto	Kurang Baik
58	A058	Labara	Kurang Baik
59	A059	Labuan Bajau	Kurang Baik
60	A060	Labuan Hiu	Kurang Baik
61	A061	Labuan Rima Baru	Kurang Baik
62	A062	Labuan Rima	Kurang Baik
63	A063	Lagundri	Kurang Baik
64	A064	Lahusa Fau	Kurang Baik
65	A065	Lahusa Idanotae	Kurang Baik
66	A066	Lahusa	Kurang Baik
67	A067	Lalimanawa	Kurang Baik
68	A068	Mahang Labara	Kurang Baik
69	A069	Makole	Kurang Baik
70	A070	Olayama	Kurang Baik
71	A071	Onohondro	Kurang Baik
76	A076	Sialema	Kurang Baik
77	A077	Sibaranun	Kurang Baik
78	A078	Talio	Kurang Baik

3.1.6 Perulangan Iterasi

Fungsi dari *interface* (antarmuka) ini adalah untuk memberikan *input* dan menampilkan *output* dari aplikasi. Pada aplikasi ini memiliki *interface* yang terdiri dari *Menu login*, *Data Kawasan*, *Data Centroid* dan *Menu Proses K-Means*. Dalam halaman utama untuk menampilkan pada tampilan *Menu* pada awal sistem yaitu *Menu login* dan *Menu Utama*. Adapun *Menu* halaman utama sebagai berikut.

a. **Menu Login**

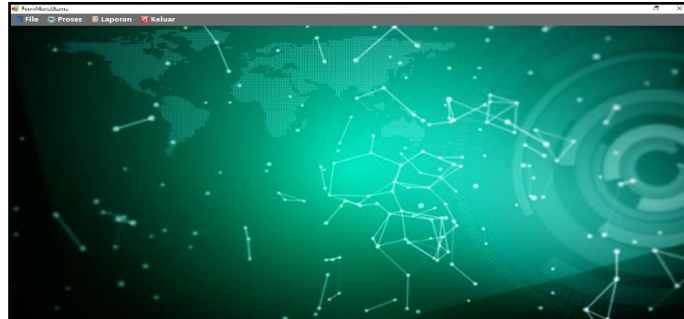
Menu Login digunakan untuk mengamankan sistem dari *user-user* yang tidak bertanggung jawab sebelum masuk ke *Menu Utama*. Berikut adalah tampilan *Menu Login* :



Gambar 1. *Menu Login*

b. Menu Utama

Menu Utama digunakan sebagai penghubung untuk *Menu Data Kawasan*, data *Centroid*, proses dan laporan. Berikut adalah tampilan *Menu Utama*.

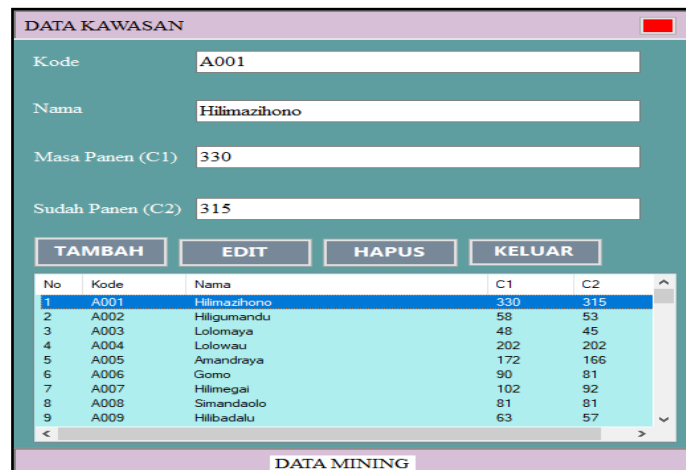


Gambar 2. Menu Utama

Administrator untuk menampilkan *Menu* pengolahan data pada penyimpanan data kedalam *database* yaitu *Menu Data Kawasan*, dan *Menu Centroid*. Adapun *Menu* halaman *adminstrator* utama sebagai berikut.

a. Menu Data Kawasan

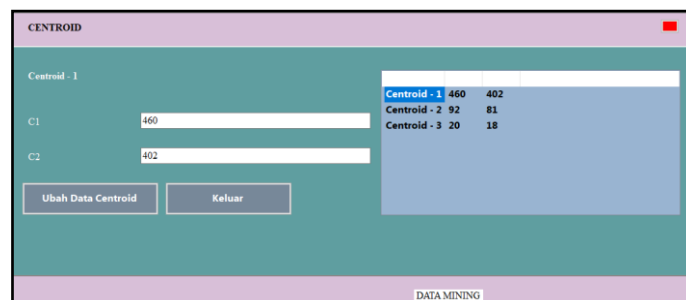
Menu data kawasan berfungsi untuk pengolahan dalam penginputan data, ubah data dan penghapusan data kawasan. Adapun *Menu* data kawasan sebagai berikut.



Gambar 3. Menu Data Kawasan

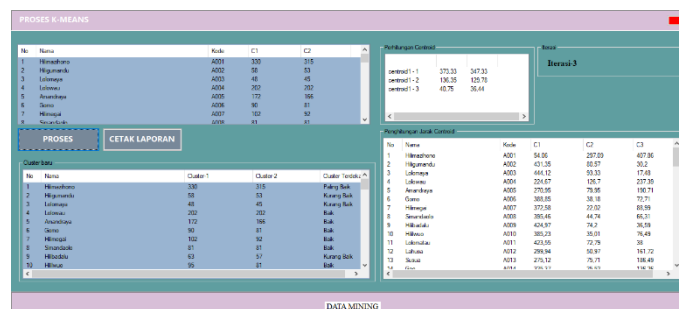
b. Menu Data Centroid

Menu Data Centroid untuk pengolahan data *Centroid* kawasan produksi karet. Adapun *Menu Data Centroid* adalah sebagai berikut.

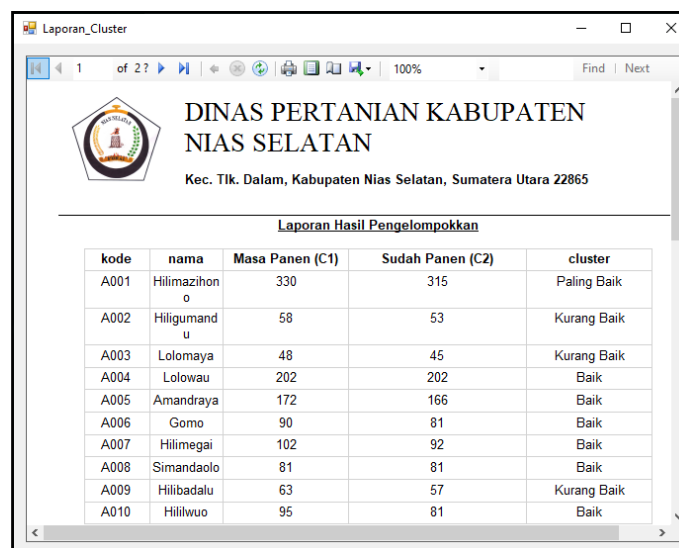


Gambar 4. Menu Data Centroid

Adapun hasil proses perhitungan kedalam program untuk pengelompokkan data kawasan produksi karet sebagai berikut.



Gambar 5. Hasil Analisa K-Means



Gambar 6. Laporan Hasil

4. KESIMPULAN

Berdasarkan hasil analisa dari permasalahan yang terjadi dengan kasus yang dibahas tentang mengelompokkan kawasan produksi karet dengan menerapkan metode *K-Means* terhadap sistem yang dirancang dan dibangun maka dapat dibuat kesimpulan sebagai berikut: Dengan menganalisa kawasan produksi karet dalam pengambilan data, dapat dikelompokkan data kawasan produksi karet dari tiga pengelompokkan (*Clustering*) dan proses analisa dapat di proses dengan metode *K-Means*. Pembuatan perancangan digunakan dengan bahasa pemodelan UML (*Unified Modeling Language*) yang terdiri dari *use case diagram*, dan *activity diagram* dan *class diagram*. Aplikasi yang dibangun dengan bahasa pemrograman visual dengan berbasis *desktop*. Dengan mengimplementasikan sistem data mining dengan berbasis *desktop* di bangun dengan *form login*, menu utama, pengolahan kawasan produksi karet dan proses metode *k-means* untuk pengelompokan kawasan produksi karet dengan cepat dan akurat.

UCAPAN TERIMAKASIH

Terima Kasih diucapkan kepada Bapak Deski Helsa Pane dan Bapak Egi Affandi serta pihak-pihak yang telah mendukung dalam proses penyelesaian penelitian ini yang tidak dapat disebutkan satu persatu. Kiranya memberi manfaat bagi pembaca.

DAFTAR PUSTAKA

- [1] D. Sunia And P. Alam Jusia, "Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means".
- [2] J. Hutagalung, *Kombinasi K-Means Clustering dan Metode MOORA*, 1st ed. Pp:105, Yogyakarta: Deepublish, 2021, ISBN: 978-623-02-3891-8, <https://penerbitbukudeepublish.com/shop/buku-kombinasi-k-means/>
- [3] S. Syahidatul Helma, R. R. Rustiyan, E. Normala, P. Studi Sistem Informasi Fakultas Sains Dan Teknologi, U. Islam Negeri Sultan Syarif Kasim Riau, J. Soebrantas No And S. Baru, "Clustering Pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K-Means," 2019.
- [4] D. Nofriansyah and I. Mariami, "Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul

- Islam Indonesia Baru Dengan Metode K-Means Clustering,” *J. CyberTech*, vol. 1, no. 1, pp. 1–12, 2021, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jct/index>
- [5] J. Hutagalung, Y. H. Syahputra and Z. P. Tanjung “Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022, doi: 10.35957/jatisi.v9i1.1516.
- [6] R. B. Hasibuan, Hafizah, and R. Mahyuni, “Penerapan Data Mining Clustering Dengan Menggunakan Algoritma K-Means Pada Data Nasabah Kredit Bermasalah,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 5, no. 1, pp. 7–15, 2022.
- [7] A. H. Nasyuha et al., “Frequent pattern growth algorithm for maximizing display items,” *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 19, no. 2, pp. 390–396, 2021, doi: 10.12928/TELKOMNIKA.v19i2.16192.
- [8] A. Fitri, Y. Syahra, and R. Kustini, “Penerapan Data Mining Dalam Mengklusterisasi Location Best Pb Tambahan Pada Regional IV PT Indomarco Prismatama Cab.Medan Dengan Menggunakan Metode K-Means,” *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 19, no. 2, p. 11, 2020, doi: 10.53513/jis.v19i2.2330.
- [9] M. H. Mhd. Gilang Suryanata, Deski Helsa Pane, “Implementasi Algoritma K-Means Untuk Mengukur Tingkat Kepuasan Siswa Terhadap Proses Pembelajaran,” *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 2, no. 2, pp. 118–125, 2019.
- [10] S. Natalia, B. Sembiring, H. Winata, and S. Kusnasari, “Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means,” *J. Sist. Inf. TGD*, vol. 1, pp. 31–40, 2022.
- [11] F. Yunita, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri),” 2018.
- [12] R. A. Indraputra And R. Fitriana, “K-Means Clustering Data Covid-19,” *Jurnal Teknik Industri*, Vol. X, No. 2622-5131, Pp. 275-282, 2020.