

Implementasi Metode K-Means *Clustering* Untuk Memetakan Daerah Potensial Penghasil Padi di Provinsi Sumatera Utara

Christian Jodhy Silalahi¹, Alfonsus Situmorang², Jimmy F. Naibaho³
^{1,2,3}Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Jan 9, 2022

Revised, Feb 11, 2022

Accepted, Feb 20, 2022

Keywords:

K-Means,

GIS,

Data Mining,

Cluster,

Potensial Penghasil Padi.

ABSTRAK

Dalam rangka memenuhi kebutuhan beras, dinas pertanian berupaya untuk mengoptimalkan hasil pertanian dengan memetakan atau mengelompokkan daerah yang menghasilkan tanaman padi di daerah Sumatera Utara dengan metode *KMeans clustering*. Berdasarkan data hasil pertanian padi di Dinas Pertanian Provinsi Sumatera Utara, menampilkan beberapa daerah dengan hasil panen padi yang bervariasi jumlahnya. Untuk itu diperlukan pengelompokan daerah potensial penghasil padi dengan membuat aplikasi pemetaan (*GIS*) untuk mengetahui daerah mana saja yang menghasilkan padi dengan jumlah banyak ataupun sedikit. Pembagian hasil panen biasanya dilakukan berdasarkan nama kecamatan penghasil padi. Oleh karena itu, dibutuhkan metode untuk memudahkan dalam pengelompokan daerah penghasil padi. Dengan pendekatan pengklasteran *K-means*, pembagian kelompok daerah dapat dilakukan berdasarkan luas panen (Ha), produksi (ton), tahun panen dan lokasi panen padi (Kabupaten) di Sumatera Utara. Pada penelitian ini dilakukan pengklasteran daerah potensial penghasil padi menggunakan algoritma *K-Means*. Dengan menggunakan metode *K-Means*, dapat memudahkan pengelompokan suatu daerah dengan hasil panen padi terbesar, sedang dan rendah. Hasilnya adalah sebuah gambaran yang menunjukkan pengelompokan daerah berdasarkan hasil pertanian padi. Dari hasil pengujian terhadap 10 data data kasus uji didapatkan nilai dan telah dihitung rata – rata C1 : (255,1323) C2 : (110,6107) dan C3 : (165,0971) hasil rata rata dari C1,C2,C3 : (530,8401). Pada masing masing 10 data tersebut sehingga dapat dikatakan bahwa sistem bekerja dengan cukup baik dan dapat diterapkan pada Dinas Pertanian Sumatera Utara.

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



Christian Jodhy Silalahi:

Christian Jodhy Silalahi,
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: jodhys.silalahi5@gmail.com

1. PENDAHULUAN

Produksi beras dalam negeri diharapkan dapat memenuhi semua kebutuhan masyarakat Indonesia karenadengan berhasilnya pemenuhanberas dalam negeri berarti pemerintah tidak memerlukan tindakan untuk mengimpor berasdari negaralain. Produksi beras di Indonesia padatahun ke tahun terus meningkat karena harusmemenuhi target yang telah dicapai pada tahun sebelumnya. Tetapi bukan berarti dapat mencukupi ketersediaan beras karena setiaptahun pula jumlah penduduk meningkat sehinggapeningkatan jumlah produksi beras dilakukan untuk mengimbangi tingginya jumlah penduduk Indonesia yang mengkonsumsi beras. Untukitu dalam rangka memenuhi kebutuhan beras, dinas pertanian berupaya untuk mengoptimalkan hasil pertanian dengan memetakan atau mengelompokkan daerah yang menghasilkan tanaman padi di daerah Sumatera Utara dengan metode *K-Mean clustering*.

Berdasarkan data hasil pertanian padi di Dinas Pertanian Provinsi Sumatera Utara, menampilkan beberapa daerah dengan hasilpanen padi yang bervariasi jumlahnya. Untuk itudiperlukan pengelompokan daerah potensial penghasil padi dengan membuat aplikasi pemetaan (*GIS*) untuk mengetahui daerah mana saja yang menghasilkan padi dengan jumlah banyak ataupun sedikit. Pembagian hasil panen biasanya dilakukan

berdasarkan namakecamatanpenghasil padi. Oleh karena itu, dibutuhkan metode untuk memudahkan dalam pengelompokan daerah penghasilpadi.

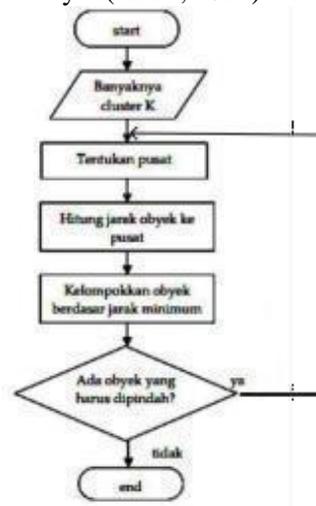
Dengan pendekatan pengklasteran K- means, pembagian kelompok daerah dapat dilakukan berdasarkan luas panen (Ha), produksi(ton), tahun panen dan lokasipanen padi(Kabupaten) di Sumatera Utara. Pada penelitian ini dilakukan pengklasteran daerah potensial penghasil padi menggunakan algoritma K- Means. Dengan menggunakan metode K-Means,dapat memudahkan pengelompokan suatudaerahdengan hasil panen padi terbesar, sedang danrendah. Hasilnya adalah sebuah gambaran menunjukkan pengelompokan daerah berdasarkan hasil pertanian padi. Pengelompokan ini akan diterapkan menggunakan system informasi berbasis webdengan memanfaatkan Google Map API. Dengan data yang sudah dikelompokkan menggunakan algoritma K-Means diharapkan dapat mempermudahdinas pertanian dalam menghitung hasil pertanian ditiap daerahnya agar mengetahui daerah mana yang menghasilkan padi terbanyak, sedang, dan sedikit. Berdasarkan latar belakang masalah yang sudah diuraikan diatas maka penelitian ini berjudul “IMPLEMENTASI METODE K- MEANS *CLUSTERING* UNTUK MEMETAKAN DAERAH POTENSIAL PENGHASIL PADIDI PROVINSI SUMATERA UTARA”.

2. METODE PENELITIAN

Analisa metode merupakan langkah- langkah sistematis dalam melakukan proses perhitungan yang direpresentasikan pada kesesuaian data yang didapatkan dari tempat penelitian

2.1.Penerapan Metode *K-Means Clustering*

K-Means merupakan salah satu metode data *clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan kedalam kelompok yang lainnya. (Aulia, 2021).



Gambar 1. Flowchart *K-Means Clustering*

1. Menentukan Banyak Cluster

Untuk melakukan *clustering* dengan algoritma *k-means* langkah yang pertamakali yaitu menentukan banyak *cluster* yang akan dibentuk. Pada penelitian ini, terdapat tiga *cluster* yaitu *cluster* untuk kabupaten dengan produksi padi banyak (C1), produksi sedang (C2) dan produksisedikit (C3).

2. Menentuka titik pusat (*centroid*)

Setelah menentukan banyak *cluster* yang akan dibentuk, langkah selanjutnya yaitu denganmenentukan titik pusat (*centroid*) dari tiap-tiap *cluster*. Pengambilan titik pusat secara random(acak). Titik pusat yang diambil pada penelitian ini adalah C1 adalah titik pusatuntuk daerah dengan produksi padi tinggi dengan titik pusat (107,161; 466,944), C2 adalah titikpusat untuk daerah produksi padi sedikit dengan titik pusat (7,016; 36,616).

3. Menentukan banyak (*cluster*)

Untuk melakukan clustering dengan algoritma k-means langkah yang pertama kali yaitu menentukan banyak cluster yang akan dibentuk. Pada penelitian ini, terdapat tiga cluster yaitu cluster untuk kabupaten dengan produksi padi banyak (C1), produksi sedang (C2) dan produksi sedikit (C3).

4. Menghitung jarak setiap objek ke titik pusat (*centroid*)

Langkah selanjutnya untuk melakukan cluster setelah menentukan titik pusat adalah dengan menghitung jarak setiap data dengan titik pusat yang sudah ditentukan sebelumnya.

Rumus untuk menghitung jarak setiap objek ke titik pusat adalah

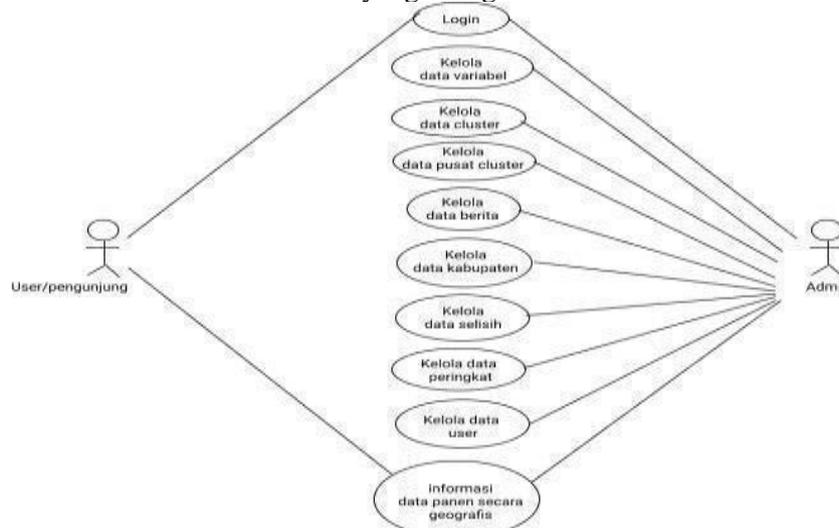
$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

3. HASIL DAN PEMBAHASAN

Implementasi Sistem peramalanpersediaan vaksin berbasis web pada dinas kesehatan kabupaten Toba ini dilakukan dengan menggunakan bahasa dan pemograman PHP, dengan Basis data yangdigunakan ialah MySQL, dengan versi 5.0. Aplikasi PHP tersebut dapat dijalankan pada berbagai platform sistem operasi dan perangkat keras, tetapi implementasi dan pengujian sepenuhnya

3.1. Use Case Diagram

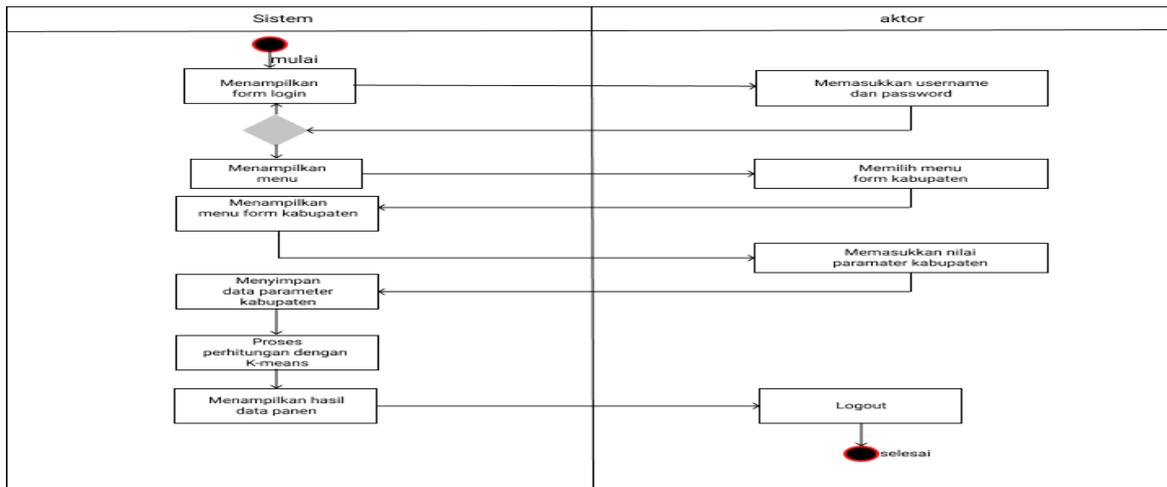
Use Case diagram adalah suatu urutaninteraksi yang saling berkaitan antara sistem dan actor.



Gambar 2. *Use Case Diagram* pengelompokan data panen

3.2. Activity Diagram

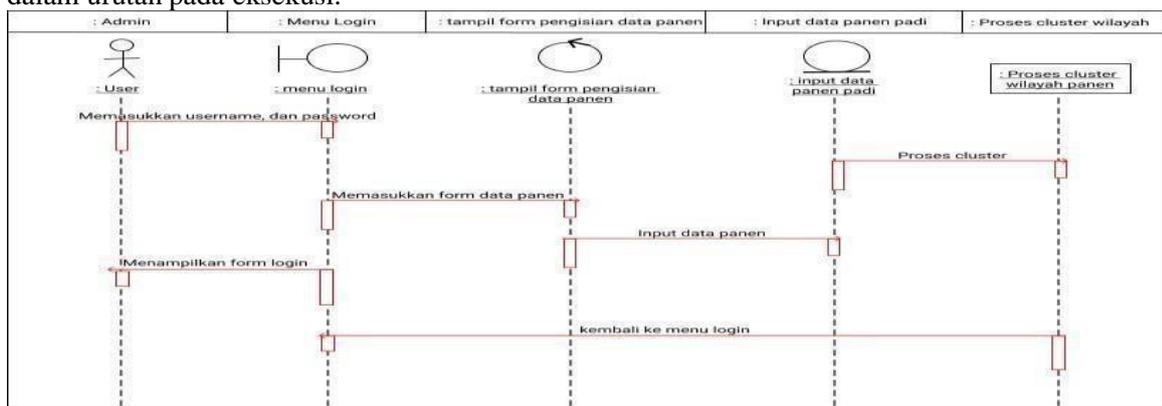
Activity diagram ialah sesuatu yang menjelaskan tentang alur kegiatan dalam program yang sedang dirancang, bagaimanaproses alir berawal, keputusan yang mungkin terjadi, dan bagaimana sistem akan berakhir. Activity diagram juga dapat menjelaskan metode paralel yang mungkin terjadi pada beberapa eksekusi. Gambar 3.2. adalah activity diagram dari aplikasi yang akan dibangun.



Gambar 3. Activity Diagram yang akan di bangun

3.3. Sequence Diagram

Sequence Diagram adalah suatu diagram yang menjelaskan interaksi objek dan menunjukkan (memberi tanda atau petunjuk) komunikasi diantara objek-objek tersebut. Sequence diagram digunakan untuk menjelaskan perilaku pada sebuah skenario dan menggambarkan bagaimana entitas dan sistem berinteraksi, termasuk pesan yang dipakai saat interaksi. Semua pesan digambarkan dalam urutan pada eksekusi.



Gambar 4 Sequence Diagram dari aplikasi yang akan di bangun

3.4. Proses Perhitungan Dengan Menggunakan Metode K-Means Clustering

Algoritma K-Means Clustering adalah suatu metode penganalisaan data atau metode data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi.

Metode K-Means Clustering berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya. Dan mempunyai karakteristik yang berbedadengan data yang ada di dalam kelompok yang lain.

Data clustering menggunakan metode K-Means clustering ini dilakukan dengan algoritma dasar sebagai berikut :

1. Menentukan jumlah cluster. Pada penelitian ini, terdapat tiga cluster yaitu cluster untuk kabupaten dengan produksi padi banyak (C1), produksi sedang (C2) dan produksi sedikit (C3).
2. Menentukan nilai centroid. Dalam menentukan nilai centroid untuk awal iterasi, nilai awal centroid dilakukan secara acak. Titik pusat yang diambil pada penelitian ini adalah C1 adalah titik pusat untuk daerah dengan produksi padi tinggi dengan titik pusat (107,161; 466,944), C2 adalah titik pusat untuk daerah dengan produksi sedang dengan titik pusat (33,489; 138,2), dan C3 adalah titik pusat untuk daerah produksi padi sedikit dengan titik pusat (7,016; 36,616).
3. Menghitung jarak antara titik centroid dengan titik tiap objek, Untuk menghitung jarak tersebut

dapat menggunakan Euclidean Distance, yaitu :

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

Dimana :

De adalah Euclidean Distance (x,y) merupakan koordinat object dan (s,t) merupakan koordinat centroid Pada proses perhitungan di bab ini, penulis mengambil beberapa data 10 Kabupaten pada tahun 2020 yang ada di Provinsi Sumatera Utara

Tabel 1. Dataset Produksi Padi 2020

No.	Kota / Kabupaten	Luas Lahan (Ha)	Produksi (Ton)
1.	Nias	11,952	42,076
2.	Mandailing Natal	107,161	466,944
3.	Tapanuli Selatan	42,317	228,915
4.	Tapanuli Tengah	33,489	138,2
5.	Tapanuli Utara	49,379	212,185
6.	Toba Samosir	21,77	145,812
7.	Labuhan Batu	7,016	36,616
8.	Asahan	16,34	103,291
9.	Simalungun	72,249	377,726
10.	Dairi	13,684	72,468

Pada bagian dibawah ini terdapat contoh perhitungan jarak data dengan titik pusatnya. Contoh untuk Kabupaten Nias tahun 2020 dengan luas panen 11,952 Ha dan Produksi 42,076 ton, maka :

Perhitungan jarak Kabupaten Nias dengan pusat cluster pertama adalah :

$$d11 = \sqrt{(11,952 - 107,161)^2 + (42,076 - 466,944)^2}$$

$$d11 = 435,4050655$$

Perhitungan jarak Kabupaten Nias dengan pusat cluster kedua adalah :

$$d12 = \sqrt{(11,952 - 33,489)^2 + (42,076 - 138,2)^2}$$

$$d12 = 98,50718626$$

Perhitungan jarak Kabupaten Nias dengan pusat cluster ketiga adalah :

$$d13 = \sqrt{(11,952 - 7,016)^2 + (42,076 - 36,616)^2}$$

$$d13 = 7,360414119$$

Jika kita terapkan rumus Euclidean distance pada setiap data Kabupaten, maka akan menghasilkan data seperti pada tabel berikut.

Tabel 2. Perhitungan *Euclidean Distance*

No.	Kot Kabupaten	a / Luas Lahan (Ha)	Produksi (Ton)	C1	C2	C3
1.	Nias	11,952	42,076	435,4050655	98,50718626	7,360414119
2.	Mandailing	107,161	466,944	0	336,8978794	441,8271253
3.	Tapanuli Tengah	42,317	228,915	246,7033627	91,14353959	195,5123168
4.	Tapanuli Selatan	33,489	138,2	336,8978794	0	104,9768012
5.	Tapanuli Utara	49,379	212,185	261,2296071	75,67213705	180,6075788
6.	Toba Samosir	21,77	145,812	332,2911138	13,97417278	110,1882341
7.	Labuhan Batu	7,016	36,616	441,8271253	104,9768012	0
8.	Asahan	16,34	103,291	374,8225693	38,89378462	67,32378927
9.	Simalungun	72,249	377,726	95,80552838	242,6417983	347,2914862
10.	Dairi	13,684	72,468	405,4001284	68,65081099	36,46680858

Pengelompokan data berdasarkan cluster. Untuk menentukan anggota cluster baru adalah dengan memperhitungkan jarak minimum objek. Pada tabel dibawah ini terdapat jarak minimum dari masing masing cluster.

Tabel 3. Jarak minimum objek

No.	Jarak
	Terpendek
1.	7,360414119
2.	0
3.	91,14353959
4.	0
5.	75,67213705
6.	13,97417278
7.	0
8.	38,89378462
9.	95,80552838
10.	36,46680858

Jika data cluster tersebut dikelompokkan maka akan menghasilkan skema tabel pada bagian dibawah ini. Angka 1 pada tabel adalah jarak terpendek (minimum) dari masing masing cluster.

Tabel 4. Skema Kelompok data

No.	C1	C2	C3
1.			1
2.	1		
3.		1	
4.		1	
5.		1	
6.		1	
7.			1
8.		1	
9.	1		
10.			1

Menentukan atau memperbaharui centroid baru

- A. Pada Cluster 1 terdapat 2 data, maka nilai rata-rata pada cluster 1 adalah Average Luaslahan (data2 + data9) = Average(107,161 + 72,249) = 89,705 Average Produksi (data2 + data 9) = Average(466,944 + 377,726) = 422,335
- B. Pada Cluster 2 terdapat 5 data, maka nilai rata-rata pada cluster 2 adalah Average Luaslahan (data3 + data4 + data5 + data6) = Average(42,317 + 33,489 + 49,379 + 21,77 + 16,34) = 32,659 Average Produksi (data3 + data4 + data5 + data6) = Average(228,915 + 138,2 + 212,185 + 145,812 + 103,291) = 165,6806
- C. Pada Cluster 3 terdapat 3 data, maka nilai rata-rata pada cluster 3 adalah : Average Luas lahan(data1 + data7 + data10) = Average(11,952 + 7,016 + 13,684) = 50,38666667 Average Produksi(data1 +data7 + data10) = Average(42,076 + 36,616+ 72,468) = 26,23066667

Sehingga didapatkan hasil perhitungan centroid baru seperti dibawah ini :

$$C1 = (89,705; 422,335)$$

$$C2 = (32,659; 165,6806)$$

$$C3 = (50,3866; 26,23066667)$$

Kembali ke step 3 sampai menghasilkan cluster yang konversi (sama). Menghitung jarak antara titik centroid dengan titik tiap objek dengan menggunakan rumus *Euclidean Distance*.

Pada bagian dibawah ini terdapat contoh perhitungan jarak data dengan titik pusat terhadap cluster baru. Contoh untuk Kabupaten Nias tahun 2020 dengan luas panen 11,952 Ha dan Produksi 42,076 ton, maka :

Perhitungan jarak Kabupaten Nias dengan pusat cluster pertama adalah :

$$d_{21} = \sqrt{(11,952 - 89,705)^2 + (42,076 - 422,335)^2}$$

$$d_{21} = 388,1268299$$

Perhitungan jarak Kabupaten Nias dengan pusat cluster kedua adalah :

$$d_{22} = \sqrt{(11,952 - 32,659)^2 + (42,076 - 165,6806)^2}$$

$$d_{22} = 125,32708$$

Perhitungan jarak Kabupaten Nias dengan pusat cluster ketiga adalah :

$$d_{23} = \sqrt{(11,952 - 50,3866)^2 + (42,076 - 26,23066667)^2} d_{23} = 41,5728059$$

Jika kita terapkan rumus Euclidean distance terhadap cluster baru pada setiap data Kabupaten, maka akan menghasilkan data seperti pada tabel berikut :

Tabel 5. Hasil Perhitungan Euclidean distance

No.	Kabupaten	Kota / (Ha)	Luas Lahan	Produksi (Ton)	C1	C2	C3
1.	Nias	11,952		42,076	388,1268299	125,32708	41,5728059
2.	Mandailing	107,161	Natal	466,944	47,90276419	310,3388216	444,3552263
3.	Tapanuli	42,317		228,915	199,1404503	63,96769737	202,8449124
4.	Tapanuli	33,489	Tengah	138,2	289,6427746	27,49313144	113,2371968
5.	Tapanuli	49,379	Utara	212,185	213,9841321	49,41879824	185,9570635
6.	Toba	21,77		145,812	284,7457353	22,65682208	122,9577525
7.	Labuhan	7,016		36,616	394,4827217	131,5873642	44,59674737
8.	Asahan	16,34		103,291	327,3705808	64,48854122	84,24648648
9.	Simalungun	72,249		377,726	47,90276419	215,7095727	352,1745746
10.	Dairi	13,684		72,468	358,0308787	95,12433665	59,03369152

Kemudian data akan dikelompokkan kembali berdasarkan cluster. Untuk menentukan anggota cluster baru adalah dengan memperhitungkan jarak minimum objek. Pada tabel dibawah ini terdapat jarak minimum dari masing masing cluster yang baru.

Tabel 6. Jarak minimum objek terhadap *cluster* baru

No.	Jarak Terpendek
1.	41,5728059
2.	47,90276419
3.	63,96769737
4.	27,49313144
5.	49,41879824
6.	22,65682208
7.	44,59674737
8.	64,48854122
9.	47,90276419
10.	59,03369152

Jika data tersebut dikelompokkan maka akan menghasilkan skema tabel pada bagian dibawah ini. Angka 1 pada tabel adalah jarak terpendek (minimum) dari masing masing *cluster*

Tabel 7. Skema Kelompok data ke 2

No.	C1	C2	C3
1.			1
2.	1		
3.		1	
4.		1	
5.		1	
6.		1	
7.			1
8.		1	
9.	1		
10.			1

Pada iterasi pertama, tidak ada data cluster yang berpindah, maka iterasi dihentikan dan proses clustering menghasilkan data sebagai berikut :

Tabel 8. Data hasil iterasi pertama

No	Kota / Kabupaten	Luas Lahan (Ha)	Produksi (Ton)	c1	c2	c3	Jumlah c1,c2,c3
1	Nias	11,952	42,076	388,126	125,327	41,572	555,025
2	Mandailing Natal	107,161	466,944	47,902	310,338	444,355	802,595
3	Tapanuli Selatan	42,317	228,915	199,14	63,967	202,844	465,951
4	Tapanuli Tengah	33,489	138,2	289,642	27,493	113,237	430,372
5	Tapanuli Utara	49,379	212,185	213,984	49,418	185,957	449,359
6	Toba Samosir	21,77	145,812	284,745	22,656	122,957	430,358
7	Labuhan Batu	7,016	36,616	394,482	131,587	44,596	570,665
8	Asahan	16,34	103,291	327,37	64,488	84,246	476,104
9	Simalungun	72,249	377,726	47,902	215,709	352,174	615,785
10	Dairi	13,684	72,468	358,03	95,124	59,033	512,187
Rata-rata				255,1323	110,6107	165,0971	530,8401

Jadi dapat disimpulkan bahwa sistem perhitungan pengelompokan suatu daerah dengan hasil panen padi terbesar, sedang dan rendah menggunakan metode K-Means Clustering berdasarkan 10 data yang telah dihitung mendapatkan hasil rata - rata sebagai berikut :

$$= \frac{\text{hasil penjumlahan pada tiap masing masing cluster}}{\text{jumlah banyaknya data}}$$

C1 : 255,1323

C2 : 110,6107

C3 : 165,0971

Hasil rata rata dari C1,C2 dan C3 : 530,8401

dari 10 data kabupaten tersebut yang dimana C1 produksi terbanyak, C2 produksi sedang, dan C3 produksi sedikit bahwa sistem dapat berfungsi dengan baik sesuai dengan berjalan nya sistem.

4. KESIMPULAN

1. Dari permasalahan diatas, data produksi tanaman padi di Provinsi Sumatera utara dapat diselesaikan dengan teknik data mining menggunakan aturan *clustering* untuk mengelompokkan daerah potensial tanaman padi di Provinsi Sumatera utara.
2. Software data mining yang penulis rancang dapat memudahkan penyelesaian tugas data mining dengan menggunakan algoritma K-Means. Hasil analisis kelayakan software yang telah dijelaskan diatas dapat diimplementasikan pada Dinas Pertanian Provinsi Sumatera untuk mengelompokkan data panen berdasarkan wilayah pada masing-masing kabupaten.
3. Dari hasil pengujian terhadap 10 data kasus uji didapatkan nilai dan rata rata dari 3 cluster adalah sebagai berikut

C1 : 255,1323

C2 : 110,6107

C3 : 165,0971

Hasil rata rata dari C1,C2 dan C3 :530,8401

pada masing masing 10 data tersebut sehingga dapat dikatakan bahwa sistem bekerja dengan cukup baik dan dapat diterapkan di Dinas Pertanian Sumatera Utara.

REFERENSI

- [1] Abdi, H. (2020, October 14). *Google Maps adalah Aplikasi Peta Online Gratis, Kenali cara penggunaannya*
- [2] Ariata. (2021, March 3). *Apa Itu JavaScript? Pemahaman Dasar tentang JavaScript untuk Pemula.*
- [3] Aulia, S. (2021). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K- Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi, 1(1)*, 1–5.
- [4] Nisa, A. C. (2020, September 3). *Mengenal Lebih Dalam Apa Itu PHP dan Kelebihanannya.*
- [5] Setiawan, S. (2021, March 3). *Pengertian Database Dan Perangkat Lunak.*
- [6] Sora. (2015, September 24). *Pengertian UML Dan Jenis- Jenisnya Serta Contoh Diagramnya.*
- [7] Wardhani, A. K. (2016). K-Means Algorithm Implementation for Clustering of Patients Disease in Kagen Clinic of Pekalongan. *Jurnal Transformatika, 14(1)*, 30.