

PEMETAAN HASIL PRODUKSI BUAH-BUAHAN DENGAN TEKNIK DATA MINING K-MEDOIDS

Ira Audita

STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Korespondensi penulis: iraaudita917@gmail.com

Irfan Sudahri Damanik

STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Eka Irawan

STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

***Abstract.** Fruits are one of the horticultural commodities that play an important role for agricultural development in Indonesia. Broadly speaking, fruit production in North Sumatra Province during the 2018-2020 period has decreased. A decrease in the amount of fruit production can cause fruit prices to be expensive, and fruit stocks become a step. This study aims to determine the results of grouping fruit plants using the K-Medoids method which is part of Data Mining. The K-Medoids method is a clustering method that can break the dataset into several groups. In this study, the data used were sourced from the Central Statistics Agency in 2017-2021. The results of this study obtained as many as 21 commodities belonging to the low cluster and 2 commodities belonging to the high cluster. This research is expected to assist the Department of Agriculture of North Sumatra Province in seeking to increase the production of fruit crops in North Sumatra Province.*

***Keywords:** Fruits, Data Mining, Prediction Results, K-Medoids, Mapping.*

Abstrak. Buah-buahan merupakan salah satu komoditas hortikultura yang memegang peranan penting bagi pembangunan pertanian di Indonesia. Secara garis besar, produksi buah-buahan di Provinsi Sumatera Utara selama periode 2018-2020 mengalami penurunan. Penurunan jumlah produksi buah-buahan dapat mengakibatkan harga buah menjadi mahal, dan stok buah-buahan menjadi langka. Penelitian ini bertujuan untuk mengetahui hasil dari pengelompokan tanaman buah-buahan menggunakan metode K-Medoids yang merupakan bagian dari Data Mining. Metode K-Medoids ini merupakan metode *clustering* yang dapat memecahkan dataset menjadi beberapa kelompok. Pada penelitian ini data yang digunakan bersumber dari Badan Pusat Statistik pada tahun 2017-2021. Hasil dari penelitian ini diperoleh sebanyak 21 komoditas yang tergolong *cluster* rendah dan 2 komoditas yang tergolong dalam *cluster* tinggi. Penelitian ini diharapkan dapat Membantu Pihak Dinas Pertanian Provinsi Sumatera Utara dalam mengupayakan meningkatkan hasil produksi tanaman buah-buahan yang ada di Provinsi Sumatera Utara.

Kata kunci: Buah-buahan, Data Mining, Hasil Prediksi, K-Medoids, Pemetaan.

LATAR BELAKANG

Pembangunan ekonomi pada suatu daerah dapat dilakukan dengan peningkatan pada sektor pertanian yang ada pada daerah tersebut. Sektor pertanian memiliki peran yang sangat penting dalam pembangunan ekonomi Indonesia, khususnya bagi pembangunan daerah itu sendiri. Hal ini karena sektor pertanian berperan penting sebagai penyedia bahan pangan, bahan baku industri, penggerak UMKM dan sebagai penyumbang produksi domestik bruto di Indonesia. Salah satu sektor pertanian yang menjadi sangat penting adalah sektor Hortikultura. Sektor Hortikultura merupakan sektor pertanian khusus untuk budidaya tanaman kebun, salah satunya yaitu buah-buahan. Sektor Hortikultura menjadi sangat penting karena sektor ini berperan dalam pemenuhan gizi dan nutrisi masyarakat Indonesia.

Hasil produk dari sektor Hortikultura merupakan produk yang prospektif, baik dalam memenuhi kebutuhan pasar domestik maupun internasional. Permintaan pasar, baik dalam negeri maupun pasar internasional sangat besar terhadap produk ini. Sudah seharusnya sektor Hortikultura mendapatkan perhatian yang lebih dari pemerintah daerah maupun pemerintah pusat. Karena dengan meningkatkan industri pertanian khususnya pada sektor Hortikultura ini, maka akan meningkatkan pendapatan masyarakat dan dapat meningkatkan kesejahteraan petani.

Indonesia merupakan salah satu negara penghasil buah tropis yang memiliki keanekaragaman dan keunggulan cita rasa yang cukup baik bila dibandingkan dengan buah-buahan dari negara-negara penghasil buah tropis lainnya. Dengan berhasilnya sektor Hortikultura di Indonesia berarti pemerintah tidak memerlukan tindakan untuk mengimpor buah dari negara lain. Namun kenyataannya pada sekarang ini, Indonesia masih terus mengimpor buah dari negara lain dalam rangka pemenuhan kebutuhan buah-buahan.

Buah mengandung banyak vitamin, protein, serat, nabati, air, karbohidrat, dan mineral. Banyak sekali manfaat yang didapat jika mengkonsumsi buah-buahan segar seperti mencegah berbagai penyakit kronis seperti kanker, diabetes, dan obesitas. Banyaknya manfaat dari mengkonsumsi buah membuat pemerintah sudah seharusnya memprioritaskan pengembangan serta meningkatkan hasil produksi buah-buahan.

Pada Provinsi Sumatera Utara, berdasarkan Badan Pusat Statistik hasil produksi sektor Hortikultura mengalami fluktuasi di tahun 2017-2019. Peningkatan paling drastis adalah produksi buah salak pada tahun 2019, yaitu meningkat sebesar 41,051 ton. Sedangkan penurunan paling drastis adalah produksi jeruk siam pada tahun 2019 yaitu sebesar 112,749 ton. Secara garis besar, produksi buah-buahan di Provinsi Sumatera Utara selama periode 2018-2020 mengalami penurunan. Penurunan jumlah produksi buah-buahan dapat mengakibatkan harga buah menjadi mahal, dan stok buah-buahan menjadi langka. Untuk itu dalam rangka memenuhi kebutuhan buah-buahan, dilakukan pengelompokan produksi buah-buahan menurut jenis produksi di Provinsi Sumatera Utara agar dapat mengoptimalkan dan meningkatkan produksi buah-buahan. Dengan alasan tidak hanya semata mengutamakan keuntungan pribadi, tetapi mendukung peningkatan nilai tambah produk dan peningkatan pendapatan petani serta bidang usaha (UMM) agribisnis yang membutuhkan bahan baku buah-buahan. Pengelompokan tersebut dapat dilakukan dengan menggunakan metode *K-Medoids clustering*. Pengelompokan hasil produksi buah-buahan akan dibagi menjadi 2 yakni *cluster* tertinggi maupun *cluster* terendah yang dapat mempermudah Dinas Ketahanan Pangan dan Pertanian Provinsi Sumatera dalam menghitung dan meningkatkan hasil pertanian khususnya sektor Hortikultura.

KAJIAN TEORITIS

1. Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis (Riandari & Simangunsong, 2019). Data mining disebut juga dengan *Knowledge Discovery in Database (KDD)*, yang merupakan kegiatan meliputi pengumpulan, pemakaian data dalam menemukan keteraturan dari suatu proses untuk mencari nilai tambahan dari sekumpulan data dengan pola dan hubungan dalam data yang berkapasitas besar.

Menurut Jiawei Han dalam buku *Data Mining Concepts and Techniques KDD* atau *Knowledge Discovery from Data*, merupakan proses terstruktur, yaitu sebagai berikut:

- a) *Data Cleaning* adalah Proses membersihkan data dari data *noise* dan tidak konsisten.
- b) *Data Integration* adalah Proses untuk menggabungkan data dari beberapa sumber yang berbeda.
- c) *Data Selection* adalah Proses untuk memilih data dari database yang sesuai dengan tujuan analisis.
- d) *Data Transformation* adalah Proses mengubah bentuk data menjadi data yang sesuai untuk proses Mining.
- e) Data Mining adalah Proses penting yang menggunakan sebuah metode tertentu untuk memperoleh sebuah pola dari data.
- f) *Pattern Evaluation* adalah Proses mengidentifikasi pola.
- g) *Knowledge Presentation* adalah yang dapat merepresentasikan informasi yang dibutuhkan, proses dimana informasi yang telah didapatkan kemudian digunakan oleh pemilik data (Sulastri & Gufroni, 2017).

2. Clustering

Clustering merupakan satu dari sekian banyak fungsi proses *data mining* untuk menemukan kelompok atau identifikasi kelompok obyek yang hampir sama. Analisis kluster (*clustering*) merupakan usaha untuk mengidentifikasi kelompok obyek yang mirip-mirip dan membantu menemukan pola penyebaran dan pola hubungan dalam sekumpulan data yang besar.

3. Algoritma K-Medoids

Algoritma *K-Medoids* atau *Partitioning Around Medoids (PAM)* ini diusulkan pertama kali tahun 1987. Algoritma *Partitioning Around Medoids (PAM)* dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw. Algoritma *K-Medoids* merupakan algoritma *clustering* yang bisa memecahkan dataset kelompok-kelompok diantara semua objek data dengan menggunakan objek sebagai perwakilan (*medoid*) dalam sebuah *cluster*.

Algoritma *K-Medoids* merupakan teknik partisi klasik dari *clustering* yang melakukan kluster data ke- n ke dalam k *cluster* yang dikenal dengan *Partitioning Around Medoids*. Algoritma *K-medoids* dapat dilakukan dengan langkah-langkah

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*)

- Hitung setiap objek ke *cluster* terdekat menggunakan persamaan ukuran jarak *Eucladian Distance* dengan persamaan (1)

$$d_{ij} = |i - j| = \sqrt{\sum_{i=1}^n (i_i + j_j)} : k = 1, 2, 3 \quad (1)$$

Dimana:

d_{ij} = jarak data ke i ke pusat klaster ke j

i_i = data ke i atribut data ke i

j_j = titik pusat ke j pada atribut ke- i

- Setelah menghitung jarak *eucladian distance*, inialisasikan pusat *cluster* baru secara acak pada masing-masing objek sebagai kandidat non medoids.
- Hitung jarak setiap objek yang berada pada objek *cluster* dengan kandidat non medoids
- Hitung total simpangan dengan menghitung *distance* baru – total *distance* lama jika simpangan kurang dari 0 maka tukar objek dengan data *cluster* non medoids untuk membentuk sekumpulan k objek baru sebagai medoids
Ulangi langkah 3 sampai 5, hingga tidak terjadi perubahan pada medoids sehingga didapatkan *cluster* beserta anggotanya masing-masing.

METODE PENELITIAN

1. Lokasi dan Waktu Penelitian

Penelitian ini dilakukan menggunakan data Badan Pusat Statistik (BPS). BPS diperlukan sebagai tempat sumber pengambilan data sekunder. Untuk memperoleh data sekunder, penulis mengambil data melalui situs <https://bps.go.id> pada tanggal 6 Maret 2022 sampai 13 Maret 2022.

2. Flowchart Penelitian

Flowchart penelitian dengan menggunakan K-Medoids.

3. Prosedur Pengumpulan Data

Pada penelitian ini terdapat teknik yang digunakan dalam pengumpulan data yaitu:

- Penelitian kepustakaan (Library Research) yaitu memanfaatkan perpustakaan, buku atau jurnal sebagai media untuk bahan referensi yang digunakan dalam penelitian.
- Data yang dikumpulkan dalam penelitian berupa data sekunder. Data sekunder yaitu data yang telah tersedia dalam berbagai bentuk seperti statistik atau data yang sudah siap digunakan.
- Sumber data penelitian ini diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara yang merupakan data hasil tanaman buah-buahan dari tahun 2017-2021. Data penelitian ini akan diolah dengan melakukan *clustering*. *Cluster* akan terbagi menjadi 2 *cluster* yaitu *cluster* terendah dan *cluster* tertinggi.

Tabel 1. Data Penelitian

No	Komoditas	Produksi (Ton) Tanaman Buah-Buahan / Tahun				
		2017	2018	2019	2020	2021
1	Stroberi	297	233	102	61	83

No	Komoditas	Produksi (Ton) Tanaman Buah-Buahan / Tahun				
		2017	2018	2019	2020	2021
2	Semangka	63811	68821	41170	63627	19899
3	Melon	493	4700	2604	1259	1754
4	Sukun	600	803	571	927	1775
5	Sirsak	932	1361	1635	1677	2650
6	Markisa	6689	1408	932	837	655
7	Sawo	11707	9573	14040	19886	18633
8	Salak	162622	194455	235506	301932	292881
9	Nenas	160552	145618	138286	158205	165063
10	Nangka	9971	10881	11394	15664	15907
11	Manggis	9382	7693	13110	19521	25821
12	Mangga	24692	26434	31980	32600	46162
13	Jeruk Besar	18076	1256	1300	4612	322
14	Jeruk Siam	435454	409683	296934	336905	448211
15	Jambu Air	11714	15422	16555	13314	18587
16	Pepaya	29570	24867	30421	30065	50232
17	Pisang	150691	118648	114050	100254	121364
18	Rambutan	15266	16841	17849	18721	34655
19	Alpukat	15123	15863	18525	32012	35378
20	Belimbing	3894	4410	4910	5665	6447
21	Duku	13289	14811	16716	16739	24308
22	Durian	64659	82873	90105	74675	15634
23	Jambu Biji	9807	11382	8456	10862	15634

4. Analisis Data

Proses analisa data dilakukan setelah adanya pengumpulan dan bukti yang mendukung penelitian kemudian diolah ke *Microsoft excel*. Setelah hasil yang telah dapat di *Microsoft excel* kemudian akan diaplikasikan ke *RapdMiner* untuk menyesuaikan hasil yang didapatkan. Untuk melakukan penelitian ini penulis menggunakan analisis data statistik dan deskriptif. Data yang digunakan penulis untuk mendukung terlaksananya penelitian ini merupakan data sekunder yang bersumber dari Dinas Pertanian Hortikultura Provinsi Sumatera Utara.

HASIL DAN PEMBAHASAN

1. Pengolahan Data

Terdapat beberapa bagian langkah-langkah dalam proses perhitungan manual menggunakan metode *K-Medoids* pada data produksi buah-buahan sebagai berikut :

- a) Menentukan jumlah *cluster* (K) dan n objek sebanyak 2 *cluster* dari data sampel, dengan memilih *cluster* secara acak untuk menentukan *medoid* seperti tabel 2 berikut:

Tabel 2. Medoid Awal

Nama	Provinsi	2017	2018	2019	2020	2021
C1	Nenas	160552	145618	138286	158205	165063
C2	Nangka	9971	10881	11394	15664	15907

b) Alokasikan setiap data objek ke *cluster* terdekat menggunakan persamaan jarak *Euclidean Distance*. Berikut adalah perhitungan persamaan jarak *Euclidean Distance* pada data produksi buah-buahan:

$$D_{\text{Stoberi}}(C1) = \sqrt{((297 - 160552)^2 + (233 - 145618)^2 + (102 - 138286)^2 + (61 - 158205)^2 + (83 - 165063)^2)}$$

$$= 343716,7529$$

$$D_{\text{Stoberi}}(C2) = \sqrt{((297 - 9971)^2 + (233 - 10881)^2 + (102 - 11394)^2 + (61 - 15664)^2 + (83 - 15907)^2)}$$

$$= 28780,6885$$

$$D_{\text{Semangka}}(C1) = \sqrt{((63811 - 160552)^2 + (68821 - 145618)^2 + (41170 - 138286)^2 + (63627 - 158205)^2 + (19899 - 165063)^2)}$$

$$= 233892,5025$$

$$D_{\text{Semangka}}(C2) = \sqrt{((63811 - 9971)^2 + (68821 - 10881)^2 + (41170 - 11394)^2 + (63627 - 15664)^2 + (19899 - 15907)^2)}$$

$$= 9756,2842$$

$$D_{\text{Melon}}(C1) = \sqrt{((493 - 160552)^2 + (4700 - 145618)^2 + (2604 - 138286)^2 + (1259 - 158205)^2 + (1754 - 165063)^2)}$$

$$= 339393,9418$$

$$D_{\text{Melon}}(C2) = \sqrt{((493 - 9971)^2 + (4700 - 10881)^2 + (2604 - 11394)^2 + (1259 - 15664)^2 + (1754 - 15907)^2)}$$

$$= 24761,1143$$

$$D_{\text{Sukun}}(C1) = \sqrt{((600 - 160552)^2 + (803 - 145618)^2 + (571 - 138286)^2 + (927 - 158205)^2 + (1775 - 165063)^2)}$$

$$= 341936,8187$$

$$D_{\text{Sukun}}(C2) = \sqrt{((600 - 9971)^2 + (803 - 10881)^2 + (571 - 11394)^2 + (927 - 15664)^2 + (1775 - 15907)^2)}$$

$$= 26896,3129$$

Lalu lakukan perhitungan yang sama sampai komoditas jambu biji seperti yang ditunjukkan sebagai berikut :

$$D_{\text{Jambu Biji}}(C1) = \sqrt{((9807 - 160552)^2 + (11382 - 145618)^2 + (8456 - 138286)^2 + (10862 - 158205)^2 + (15634 - 165063)^2)}$$

$$= 318807,4236$$

$$D_{\text{Jambu Biji}}(C2) = \sqrt{((600 - 9971)^2 + (803 - 10881)^2 + (571 - 11394)^2 + (927 - 15664)^2 + (1775 - 15907)^2)}$$

$$= 5660,69554$$

Untuk melihat keseluruhan hasil perhitungan jarak ke setiap *medoid* awal bisa dilihat pada tabel 3 berikut ini.

Tabel 3. Hasil Perhitungan literasi ke-1

Komoditas	Jarak K-Medoids		Kedekatan	Cluster Terdekat
	C1	C2		
Stroberi	343716,7529	28780,68847	28780,68847	2
Semangka	233892,5025	97256,28416	97256,28416	2
Melon	339393,9418	24761,11425	24761,11425	2
Sukun	341936,8187	26896,31289	26896,31289	2
Sirsak	340354,4029	25277,58335	25277,58335	2
Markisa	339309,8084	25737,64033	25737,64033	2
Sawo	310924,4848	6081,302163	6081,302163	2
Salak	220988,5914	515647,0637	220988,5914	1
Nenas	0	315429,4302	0	1
Nangka	315429,4302	0	0	2
Manggis	310106,904	11252,50043	11252,50043	2
Mangga	271804,7	45657,20279	45657,20279	2
Jeruk Besar	332618,954	25005,18776	25005,18776	2
Jeruk Siam	531579,8168	843619,8425	531579,8168	1
Jambu Air	310436,4758	7937,225649	7937,225649	2
Pepaya	270144,3405	48242,61759	48242,61759	2
Pisang	81730,80765	245419,5305	81730,80765	1
Rambutan	297756,0095	21588,38584	21588,38584	2
Alpukat	291670,231	27360,40084	27360,40084	2
Belimbing	332713,1988	17615,83455	17615,83455	2
Duku	305615,3244	11247,67238	11247,67238	2
Durian	211565,8247	133609,2047	133609,2047	2
Jambu Biji	318807,4236	5660,69554	5660,69554	2
Jumlah	6652496,744	2510083,721		
Total Cost	9162580,465			

- c) Setelah diperoleh hasil jarak dari setiap objek (*cost*) pada literasi ke 1 maka lanjut literasi ke 2 dengan cara memilih kandidat *medoid* baru seperti yang terdapat pada tabel 4 sebagai berikut:

Tabel 4. Medoid Baru (Non-Medoid 1)

Nama	Komoditas	2017	2018	2019	2020	2021
C1	Pisang	150691	118648	114050	100254	121364
C2	Rambutan	15266	16841	17849	18721	34655

d) Kemudian hitung kembali jarak dari setiap objek yang berada pada masing-masing *cluster* dengan kandidat baru. Perhitungannya sebagai berikut:

$$D_{\text{Stoberi}}(C1) = \sqrt{\begin{aligned} &((297 - 150691)^2 + (233 - 118648)^2 + (102 - 114050)^2 \\ &+ (61 - 100254)^2 + (83 - 121364)^2 \end{aligned}}$$

$$= 272712,9$$

$$D_{\text{Stoberi}}(C2) = \sqrt{\begin{aligned} &((297 - 15266)^2 + (233 - 16841)^2 + (102 - 17849)^2 \\ &+ (61 - 18721)^2 + (83 - 34655)^2 \end{aligned}}$$

$$= 48562,04$$

$$D_{\text{Semangka}}(C1) = \sqrt{\begin{aligned} &((63811 - 150691)^2 + (68821 - 118648)^2 + (41170 - 114050)^2 \\ &+ (63627 - 100254)^2 + (19899 - 121364)^2 \end{aligned}}$$

$$= 164253$$

$$D_{\text{Semangka}}(C2) = \sqrt{\begin{aligned} &((63811 - 15266)^2 + (68821 - 16841)^2 + (41170 - 17849)^2 \\ &+ (63627 - 18721)^2 + (19899 - 34655)^2 \end{aligned}}$$

$$= 88525,11$$

$$D_{\text{Melon}}(C1) = \sqrt{\begin{aligned} &((493 - 150691)^2 + (4700 - 118648)^2 + (2604 - 114050)^2 \\ &+ (1259 - 100254)^2 + (1754 - 121364)^2 \end{aligned}}$$

$$= 268459,2$$

$$D_{\text{Melon}}(C2) = \sqrt{\begin{aligned} &((493 - 15266)^2 + (4700 - 16841)^2 + (2604 - 17849)^2 \\ &+ (1259 - 18721)^2 + (1754 - 34655)^2 \end{aligned}}$$

$$= 44558,42$$

$$D_{\text{Sukun}}(C1) = \sqrt{\begin{aligned} &((600 - 150691)^2 + (803 - 118648)^2 + (571 - 114050)^2 \\ &+ (927 - 100254)^2 + (1775 - 121364)^2 \end{aligned}}$$

$$= 271034,3$$

$$D_{\text{Sukun}}(C2) = \sqrt{\begin{aligned} &((600 - 15266)^2 + (803 - 16841)^2 + (571 - 17849)^2 \\ &+ (927 - 18721)^2 + (1775 - 34655)^2 \end{aligned}}$$

$$= 46567,79$$

Lalu lakukan perhitungan yang sama sampai komoditas jambu biji seperti yang ditunjukkan sebagai berikut :

$$D_{\text{Jambu Biji}}(C1) = \sqrt{\begin{aligned} &((9807 - 150691)^2 + (11382 - 118648)^2 + (8456 - 114050)^2 \\ &+ (10862 - 100254)^2 + (15634 - 121364)^2 \end{aligned}}$$

$$= 248342,8$$

$$D_{\text{Jambu Biji}}(C2) = \sqrt{\begin{aligned} &((600 - 15266)^2 + (803 - 16841)^2 + (571 - 17849)^2 \\ &+ (927 - 18721)^2 + (1775 - 34655)^2 \end{aligned}}$$

$$= 23903,81$$

Untuk melihat seluruh hasil perhitungan jarak ke setiap medoid baru dapat dilihat pada tabel 5. berikut:

Tabel 5. Hasil Perhitungan Literasi ke-2

Komoditas	Jarak Ke Medoid		Kedekatan	Cluster Terdekat
	C1	C2		
Stroberi	272712.91	48562.037	48562.037	2
Semangka	164252.98	88525.109	88525.109	2
Melon	268459.23	44558.419	44558.419	2
Sukun	271034.35	46567.791	46567.791	2
Sirsak	269502.45	44975.612	44975.612	2
Markisa	267828.08	45538.607	45538.607	2
Sawo	241305.99	18386.421	18386.421	2
Salak	301218.7	497518.07	301218.7	1
Nenas	81730.808	297756.01	81730.808	1
Nangka	245419.53	21588.386	21588.386	2
Manggis	241069.31	14813.677	14813.677	2
Mangga	203339.76	26563.094	26563.094	2
Jeruk Besar	260535.49	43617.388	43617.388	2
Jeruk Siam	601722.99	825251.99	601722.99	1
Jambu Air	239994.09	17427.592	17427.592	2
Pepaya	201110.02	28256.071	28256.071	2
Pisang	0	228309.68	0	1
Rambutan	228309.68	0	0	2
Alpukat	223872.79	13364.405	13364.405	2
Belimbing	262038.58	37648.642	37648.642	2
Duku	235647.96	10968.22	10968.22	2
Durian	145216.52	124553.09	124553.09	2
Jambu Biji	248342.81	23903.81	23903.81	2
Jumlah	5474665	2548654.1		
Total Cost	8023319			

- e) Setelah diperoleh nilai jarak antara literasi 1 dan literasi ke 2, maka hitung total simpangan (*S*) dengan menghitung nilai sebagai berikut:

$$\begin{aligned}
 S &= \text{Total Cost baru} - \text{Total cost lama} \\
 &= 8023319 - 9162580,465 \\
 &= -1139261
 \end{aligned}$$

- f) Jika hasil $S < 0$, maka tukarlah objek dengan data *cluster* untuk membentuk sekumpulan *k* objek baru sebagai *medoid* baru. Karena hasil simpangan masih di bawah 0 maka perhitungan akan dilanjutkan ke literasi ke 3 sebagai *medoid* baru yang terdapat pada tabel 6 sebagai berikut:

Tabel 6. Medoid Baru (Non Medoid) 2

Nama	Komoditas	2017	2018	2019	2020	2021
C1	Jambu Biji	9807	11382	8456	10862	15634
C2	Jeruk Siam	435454	409683	296934	336905	448211

- g) Selanjutnya masuk ke proses perhitungan yang sama seperti sebelumnya yaitu menentukan jarak dari setiap objek pada masing-masing *cluster* dengan menggunakan *medoid* baru seperti berikut ini:

$$D_{\text{Stoberi}}(C1) = \sqrt{\begin{matrix} ((297 - 9807)^2 + (233 - 11382)^2 + (102 - 8456)^2 \\ + (61 - 10862)^2 + (83 - 15634)^2 \end{matrix}}$$

$$= 253557,934$$

$$D_{\text{Stoberi}}(C2) = \sqrt{\begin{matrix} ((297 - 435454)^2 + (233 - 409683)^2 + (102 - 296934)^2 \\ + (61 - 336905)^2 + (83 - 448211)^2 \end{matrix}}$$

$$= 871437,17$$

$$D_{\text{Semangka}}(C1) = \sqrt{\begin{matrix} ((63811 - 9807)^2 + (68821 - 11382)^2 + (41170 - 8456)^2 \\ + (63627 - 10862)^2 + (19899 - 15634)^2 \end{matrix}}$$

$$= 100440,09$$

$$D_{\text{Semangka}}(C2) = \sqrt{\begin{matrix} ((63811 - 435454)^2 + (68821 - 409683)^2 + (41170 - 296934)^2 \\ + (63627 - 336905)^2 + (19899 - 448211)^2 \end{matrix}}$$

$$= 760166,22$$

$$D_{\text{Melon}}(C1) = \sqrt{\begin{matrix} ((493 - 9807)^2 + (4700 - 11382)^2 + (2604 - 8456)^2 \\ + (1259 - 10862)^2 + (1754 - 15634)^2 \end{matrix}}$$

$$= 21225,401$$

$$D_{\text{Melon}}(C2) = \sqrt{\begin{matrix} ((493 - 435454)^2 + (4700 - 409683)^2 + (2604 - 296934)^2 \\ + (1259 - 336905)^2 + (1754 - 44821)^2 \end{matrix}}$$

$$= 867072,4$$

$$D_{\text{Sukun}}(C1) = \sqrt{\begin{matrix} ((600 - 9807)^2 + (803 - 11382)^2 + (571 - 8456)^2 \\ + (927 - 10862)^2 + (1775 - 15634)^2 \end{matrix}}$$

$$= 23444,262$$

$$D_{\text{Sukun}}(C2) = \sqrt{\begin{matrix} ((600 - 435454)^2 + (803 - 409683)^2 + (571 - 296934)^2 \\ + (927 - 336905)^2 + (1775 - 448211)^2 \end{matrix}}$$

$$= 867072,4$$

Lalu lakukan perhitungan yang sama sampai komoditas jambu biji seperti yang ditunjukkan sebagai berikut :

$$D_{\text{Jambu Biji}}(C1) = \sqrt{\begin{matrix} ((9807 - 9807)^2 + (11382 - 11382)^2 + (8456 - 8456)^2 \\ + (10862 - 10862)^2 + (15634 - 15634)^2 \end{matrix}}$$

$$= 0$$

$$D_{\text{Jambu Biji}}(C2) = \sqrt{\begin{matrix} ((600 - 435454)^2 + (803 - 409683)^2 + (571 - 296934)^2 \\ + (927 - 336905)^2 + (1775 - 448211)^2 \end{matrix}}$$

$$= 846442,86$$

Untuk melihat hasil seluruh perhitungan jarak ke setiap medoid dapat dilihat pada tabel 7 berikut ini:

Tabel 7. Hasil Perhitungan Literasi Ke-3

Komoditas	Jarak Ke Medoid		Kedekatan	Cluster Terdekat
	C1	C2		
Stroberi	25357.934	871437.168	25357.93404	1
Semangka	100440.092	760166.219	100440.0915	1
Melon	21225.4007	867072.397	21225.40066	1
Sukun	23444.262	869654.07	23444.26201	1
Sirsak	21878.3132	868124.496	21878.31319	1
Markisa	22151.3937	866829.371	22151.39368	1
Sawo	11335.348	839474.152	11335.34799	1
Salak	519622.453	387148.952	387148.952	2
Nenas	318807.424	531579.817	318807.4236	1
Nangka	5660.69554	843619.842	5660.69554	1
Manggis	14635.6521	838352.723	14635.65209	1
Mangga	49051.034	800031.731	49051.03396	1
Jeruk Besar	22262.9831	859906.777	22262.98311	1
Jeruk Siam	846442.857	0	0	2
Jambu Air	10014.218	838389.9	10014.21804	1
Pepaya	51192.3767	797742.079	51192.3767	1
Pisang	248342.811	601722.986	248342.8107	1
Rambutan	23903.81	825251.993	23903.81001	1
Alpukat	31414.5192	820167.508	31414.51916	1
Belimbing	14406.8125	860606.544	14406.81252	1
Duku	14208.7075	833588.227	14208.70754	1
Durian	137325.756	736857.773	137325.7556	1
Jambu Biji	0	846442.857	0	1
Jumlah	2533124.85	17364167.6		
Total Cost	19897292.4			

h) Setelah didapatkan nilai *n total cost* baru dan *cost* lama lalu dilanjutkan menghitung selisih antara *cost* baru dengan *total cost* lama yaitu :

$$\begin{aligned}
 S &= \text{Total cost baru} - \text{total cost lama} \\
 &= 19897292 - 8023319 \\
 &= 11873973
 \end{aligned}$$

i) Dengan hasil *simpangan* sudah $S > 0$ maka proses perhitungan *cluster* dihentikan sehingga hasil *cluster* yang telah didapat bisa dilihat dari tabel 8 berikut ini :

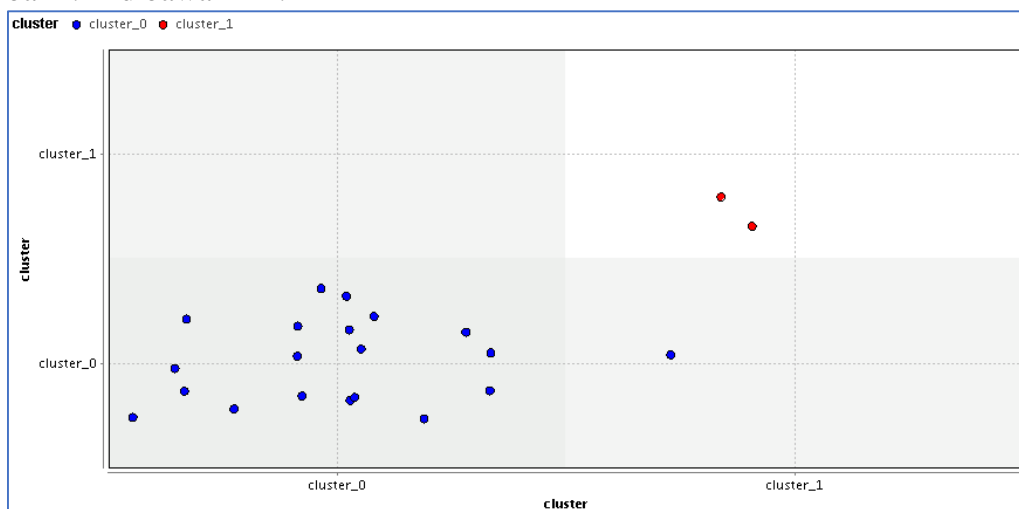
Tabel 8. Hasil Cluster Perhitungan Manual

Komoditas	Cluster Terdekat
Stroberi	C1
Semangka	C1
Melon	C1
Sukun	C1
Sirsak	C1
Markisa	C1
Sawo	C1
Salak	C2

Komoditas	Cluster Terdekat
Nenas	C1
Nangka	C1
Manggis	C1
Mangga	C1
Jeruk Besar	C1
Jeruk Siam	C2
Jambu Air	C1
Pepaya	C1
Pisang	C1
Rambutan	C1
Alpukat	C1
Belimbing	C1
Duku	C1
Durian	C1
Jambu Biji	C1

2. Pengujian

Setelah hasil dari perhitungan manual didapatkan, maka selanjutnya melakukan pengujian data dengan menggunakan software *Rapidminer 5.3*. Adapun hasil akhir pengelompokkan masing-masing anggota *cluster* dimana terdapat 2 *cluster* yaitu *cluster_1* berwarna merah dan *cluster_0* yang berwarna biru dan juga dapat dilihat pada gambar 4.12 dibawah ini :



Gambar 2. Hasil Pengelompokkan

Pada data yang telah dianalisa dapat disimpulkan bahwa data yang digunakan dalam penelitian ini adalah valid. Hal ini dapat dibuktikan dengan hasil akhir perhitungan manual dengan hasil perhitungan menggunakan *software rapidminer 5.3* menampilkan hasil yang sama yaitu *cluster* tinggi yaitu 21 dan *cluster* rendah 2. Berikut ini tampilan dari *cluster* model berupa teks yang terdapat pada gambar 4.13 dibawah ini :

Cluster Model

Cluster 0: 21 items
 Cluster 1: 2 items
 Total number of items: 23

Gambar 3. Cluster Model

Dari gambar 4.13 dapat diketahui jumlah setiap *cluster*. Pada *cluster_0* terdapat 21 *items*, dan *cluster_1* terdapat 2 *items*.

3. Hasil Percobaan dan Pengujian

Setelah melakukan perhitungan baik secara manual maupun menggunakan *software Rapidminer 5.3*. Maka hasil yang didapatkan dari perhitungan algoritma dan pengujian pada *Software Rapidminer 5.3* dapat dilihat pada tabel 9 berikut ini:

Tabel 9. Hasil Perhitungan Dan Pengujian

Komoditas	Perhitungan K-Medoids	Pengujian dengan Rapid Miner
Stroberi	<i>Cluster_1</i>	<i>Cluster_0</i>
Semangka	<i>Cluster_1</i>	<i>Cluster_0</i>
Melon	<i>Cluster_1</i>	<i>Cluster_0</i>
Sukun	<i>Cluster_1</i>	<i>Cluster_0</i>
Sirsak	<i>Cluster_1</i>	<i>Cluster_0</i>
Markisa	<i>Cluster_1</i>	<i>Cluster_0</i>
Sawo	<i>Cluster_1</i>	<i>Cluster_0</i>
Salak	<i>Cluster_2</i>	<i>Cluster_1</i>
Nenas	<i>Cluster_1</i>	<i>Cluster_0</i>
Nangka	<i>Cluster_1</i>	<i>Cluster_0</i>
Manggis	<i>Cluster_1</i>	<i>Cluster_0</i>
Mangga	<i>Cluster_1</i>	<i>Cluster_0</i>
Jeruk Besar	<i>Cluster_1</i>	<i>Cluster_0</i>
Jeruk Siam	<i>Cluster_2</i>	<i>Cluster_1</i>
Jambu Air	<i>Cluster_1</i>	<i>Cluster_0</i>
Pepaya	<i>Cluster_1</i>	<i>Cluster_0</i>
Pisang	<i>Cluster_1</i>	<i>Cluster_0</i>
Rambutan	<i>Cluster_1</i>	<i>Cluster_0</i>
Alpukat	<i>Cluster_1</i>	<i>Cluster_0</i>
Belimbing	<i>Cluster_1</i>	<i>Cluster_0</i>
Duku	<i>Cluster_1</i>	<i>Cluster_0</i>
Durian	<i>Cluster_1</i>	<i>Cluster_0</i>
Jambu Biji	<i>Cluster_1</i>	<i>Cluster_0</i>
Hasil	<i>Cluster_1=21</i>	<i>Cluster_0=21</i>
	<i>Cluster_2= 2</i>	<i>Cluster_1=2</i>

Berdasarkan tabel diatas dapat ketahui bahwa hasil pengelompokan dari *tools Rapidminer 5.3* dengan perhitungan K-Medoids diperoleh *cluster* rendah sebanyak 21

komoditas, yaitu : Stroberi, Semangka, Melon, Sukun, Sirsak, Markisa, Sawo, Nenas, Nangka, Mangga, Manggis, Mangga, Jeruk Besar, Jambu Air, Pepaya, Pisang, Rambutan, Alpukat, Belimbing, Duku, Durian, Jambu Biji. Sedangkan *Cluster* tinggi sebanyak 2, yaitu: Salak dan Jeruk Siam.

KESIMPULAN DAN SARAN

Berdasarkan seluruh hasil tahapan penelitian yang telah dilakukan pada pemetaan hasil produksi buah-buahan dengan teknik data mining dapat disimpulkan sebagai berikut : Hasil pengelompokan dari tools *Rapidminer* dengan perhitungan K-Medoids diperoleh hasil *cluster* tinggi (C1) sebanyak 2 komoditas, dan *cluster* rendah (C2) sebanyak 21 komoditas, jadi jenis buah yang masih perlu ditingkatkan adalah *cluster* terendah sebanyak 21 jenis buah yaitu : Stroberi, Semangka, Melon, Sukun, Sirsak, Markisa, Sawo, Nenas, Nangka, Mangga, Manggis, Mangga, Jeruk Besar, Jambu Air, Pepaya, Pisang, Rambutan, Alpukat, Belimbing, Duku, Durian, Jambu Biji. Penelitian yang dilakukan menggunakan metode K-Medoids mampu untuk mengelompokkan produksi buah-buahan dengan menggunakan 2 *cluster* yaitu *cluster* tinggi dan *cluster* rendah. Adapun Hasil perhitungan manual dan hasil *Clustering* yang di evaluasi menggunakan software *Rapidminer* 5.3 mendapatkan hasil yang sama yaitu terdapat 2 komoditas buah-buahan dengan *cluster* tinggi dan 21 provinsi dengan *cluster* rendah.

DAFTAR REFERENSI

- Adlin, F. N., & Purwaningsih, T. (2021). *perbandingan cluster k-medoids dan k-median tingkat produksi mangga dan pisang kabupaten indramayu*. 1–11.
- Arminarahmah, N., GS, A. D., Bhawika, G. W., Dewi, M. P., & Wanto, A. (2021). Mapping the Spread of Covid-19 in Asia Using Data Mining X-Means Algorithms. *IOP Conf. Series: Materials Science and Engineering*, 1071(1), 012018. <https://doi.org/10.1088/1757-899X/1071/1/012018>
- Febriyati, N. A., GS, A. D., & Wanto, A. (2020). GRDP Growth Rate *Clustering* in Surabaya City uses the K- Means Algorithm. *International Journal of Information System & Technology*, 3(2), 276–283.
- Gultom, W. T. C., Wanto, A., Gunawan, I., Lubis, M. R., & Kirana, I. O. (2021). Application of The Levenberg Marquardt Method In Predict The Amount of Criminality in Pematangsiantar City. *Journal of Computer Networks, Architecture, and High-Performance Computing*, 3(1), 21–29. <https://doi.org/10.47709/cnahpc.v3i1.926>
- Hanafiah, M. A., & Wanto, A. (2020). Implementation of Data Mining Algorithms for Grouping Poverty Lines by District/City in North Sumatra. *International Journal of Information System & Technology*, 3(2), 315–322.
- Hutagalung, J., Ginantra, N. L. W. S. R., Bhawika, G. W., Parwita, W. G. S., Wanto, A., & Panjaitan, P. D. (2021). COVID-19 Cases and Deaths in Southeast Asia *Clustering* using K-Means Algorithm. *Journal of Physics: Conference Series*, 1783(1), 012027. <https://doi.org/10.1088/1742-6596/1783/1/012027>

- Mahulae, H. (2020). *Pengelompokan Potensi Produksi Buah-Buahan di Provinsi Sumatera Utara dengan Menerapkan K-Clustering (Studi Kasus : Dinas Tanaman Pangan dan Holtikultura)*. 7(2), 312–325. <https://doi.org/10.30865/jurikom.v7i2.2122>
- Mutiara, D., & Hermanto, C. (2017). *ANALISIS ALGORITMA CLUSTERING DALAM KASUS PENENTUAN JENIS BUNGA IRIS*. 9(2), 72–84.
- Pulungan, N., & Suhendro, D. (2019). *PENERAPAN ALGORITMA K-MEDOIDS UNTUK MENGELOMPOKKAN PENDUDUK*. 3, 329–334. <https://doi.org/10.30865/komik.v3i1.1609>
- Riandari, F., & Simangunsong, A. (2019). Penerapan algoritma C4.5 untuk mengukur tingkat kepuasan mahasiswa. *Jurnal Mantik Penusa*, 3(2), 1–7.
- Ritonga, S. A., Safii, M., Parlina, I., & Tambunan, H. S. (2019). *Teknik Data Mining dalam Mengelompokkan Produktivitas Padi Menurut Provinsi Menggunakan*. (September), 651–660.
- Sinaga, T. H., Wanto, A., Gunawan, I., Sumarno, S., & Nasution, Z. M. (2021). Implementation of Data Mining Using C4.5 Algorithm on Customer Satisfaction in Tirta Lihou PDAM. *Journal of Computer Networks, Architecture, and High-Performance Computing*, 3(1), 9–20. <https://doi.org/10.47709/cnahpc.v3i1.923>
- Sulastri, H., & Gufroni, A. I. (2017). Penerapan data mining dalam pengelompokan penderita Thalassaemia. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 3(2), 299–305. <https://doi.org/10.25077/teknosi.v3i2.2017.299-305>
- Sundari, S., Damanik, I. S., Windarto, A. P., Tambunan, H. S., Jalaluddin, J., & Wanto, A. (2019). Analisis K-Medoids Clustering Dalam Pengelompokkan Data Imunisasi Campak Balita di Indonesia. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1, 687–696. <https://doi.org/10.30645/senaris.v1i10.75>
- Wanto, A., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitupulu, D., ... Prianto, C. (2020). *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis.