

Implementasi Algoritma Partitioning Around Medoids (PAM) untuk Mengelompokkan Hasil Produksi Komoditi Perkebunan (Studi Kasus: Dinas Perkebunan Provinsi Kalimantan Timur)

Dary Daris Abdurrahman¹⁾, Fahrul Agus^{2,*}, Gubtha Mahendra Putra³⁾

^{1,2,3)}Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman

Jl.Sambaliung No.9 Kampus Gunung Kelua, Samarinda, Kalimantan Timur 75119

e-mail : darykun23@gmail.com¹⁾; fahrulagus@unmul.ac.id²⁾; gubthamp@fkti.unmul.ac.id³⁾;

ABSTRAK

Sebagai salah satu Provinsi terluas di Indonesia, tepatnya terluas ke-4, Kalimantan Timur memiliki luas 129.000 Km². Berdasarkan data statistik dari Dinas Perkebunan tahun 2019, luas lahan perkebunan di Provinsi Kalimantan Timur seluas 1,39 juta hektar atau 10,7% dari luas keseluruhan. Dari luas keseluruhan tersebut, Provinsi Kalimantan Timur mampu memproduksi 18,4 juta ton komoditi perkebunan. Akan tetapi, produksi komoditi-komoditi tersebut dari tahun ke tahun mengalami perubahan jumlah produksi yang menunjukkan pola yang tidak tetap. Untuk itu, dalam rangka mengoptimalkan jumlah produksi, Dinas Perkebunan perlu untuk mengelompokkan daerah-daerah berdasarkan jumlah produksinya. *Clustering* adalah metode data mining yang membagi data menjadi kelompok-kelompok yang mempunyai objek yang karakteristiknya sama. Penelitian ini menggunakan metode *clustering Partitioning Around Medoids (PAM)* dengan 3 distance measure yakni *Euclidean Distance*, *Manhattan Distance*, dan *Chebyshev Distance*. Untuk mengukur kualitas hasil *cluster* digunakan metode *Silhouette Coefficient (SC)*. Semakin besar nilai SC, semakin baik kualitas *cluster*. Dari 3 kali uji coba dengan menggunakan 3 *cluster*, 5 *cluster*, dan 7 *cluster* diperoleh nilai rata-rata SC terbesar pada uji coba 5 *cluster* dengan nilai SC 0.954701931 pada distance measure *Manhattan Distance*.

Kata Kunci: *Clustering*, Partitioning Around Medoids, Euclidean Distance, Manhattan Distance, Chebyshev Distance, Silhouette Coefficient

1. PENDAHULUAN

Indonesia merupakan Negara agraris yang artinya Negara yang mengandalkan sektor pertanian serta perkebunan sebagai sumber mata pencaharian maupun sebagai penopang pembangunan nasional. Salah satu provinsi yang cocok untuk daerah perkebunan adalah Provinsi Kalimantan Timur. Berdasarkan data statistik dari Dinas Perkebunan Provinsi Kalimantan Timur tahun 2019, total luas lahan perkebunan tanaman komoditi di Kalimantan Timur mencapai 1.389.845 hektar (ha) yang tersebar di 10 kabupaten/kota di Kalimantan Timur. Luas lahan perkebunan tersebut terdiri dari Perkebunan Pola Rakyat/Swadaya seluas 391.167 hektar (ha), Pola Perkebunan Besar Negara (PTPN XIII) seluas 14.801 hektar (ha), dan Pola Perkebunan Swasta seluas 983.877 hektar (ha). Dari luas lahan tersebut, perkebunan provinsi Kalimantan Timur dapat memproduksi 18.416.921 ton komoditi perkebunan yang terdiri dari Kelapa Sawit, Karet, Lada, Kakao, Kopi, Aren, Jambu Mete, Kelapa Dalam, Kapok (disbun.kaltimprov.go.id, 2019).

Produksi komoditi perkebunan di Kalimantan Timur dari tahun ke tahun mengalami perubahan yang menunjukkan pola yang tidak tetap, ada komoditi yang mengalami kenaikan jumlah produksi di setiap tahunnya, stabil di setiap tahunnya, dan juga mengalami penurunan jumlah produksi di setiap tahunnya. Untuk itu, dalam rangka mengoptimalkan jumlah produksi, peran Dinas Perkebunan untuk mengelompokkan daerah mana saja yang menghasilkan komoditi perkebunan di Kalimantan Timur agar dapat mengoptimalkan produksinya.

Pengelompokan (*clustering*) adalah cara untuk menemukan sekumpulan objek yang memiliki kemiripan satu sama lain dan dapat menemukan pola penyebaran dan pola hubungan dalam kumpulan data yang besar (Narwati, 2010). Salah satu algoritma *clustering* yang banyak digunakan adalah *Partitioning Around Medoids (PAM)*. Penelitian terdahulu yang menggunakan analisis *clustering* oleh Septian Wulandari dan Nurfidah Dwitiyanti dalam implementasi algoritma PAM untuk *clustering* virus *MERS-CoV*. Penelitian tersebut menggunakan 100 data set dan menggunakan *Silhouette Coefficient (SC)* untuk mengetahui kekuatan cluster yang didapatkan. *Cluster* terbaik yang didapat yakni dengan menggunakan 2 *cluster* yang menghasilkan nilai *SC* tertinggi yakni 0,61534 (Wulandari & Dwitiyanti, 2020).

Para peneliti telah memperlihatkan berbagai algoritma dalam data mining yang digunakan untuk melakukan *clustering* dalam berbagai bidang seperti kesehatan (Riyanto, 2019) (Hardiyanti et al., 2019) (Marlina et al., 2018), pendidikan (Haviluddin et al., 2021) (Narwati, 2010), ekonomi (Anggreini, 2019) dan lain-lain. Hal ini menunjukkan bahwa algoritma *clustering* menjadi pilihan bagi para peneliti untuk melakukan analisa. Tujuan penelitian ini adalah mengetahui daerah potensial penghasil komoditi perkebunan dengan menggunakan algoritma *Partitioning Around Medoids (PAM)* dan diharapkan dapat membantu Dinas Perkebunan Provinsi Kalimantan Timur berdasarkan informasi hasil *clustering* wilayah Kabupaten/Kota di Kalimantan Timur sehingga pihak Dinas Perkebunan diharapkan

*) Corresponding Author

dapat memberi perhatian lebih pada daerah yang menghasilkan peroduksi terendah.

2. TINJAUAN PUSAKA

A. Data Mining

Data mining adalah sebuah proses penambangan/pencarian informasi dari jumlah kumpulan data yang besar dengan menggunakan algoritma dan teknik gambar dari statistik, mesin pembelajaran, kecerdasan buatan, dan sistem manajemen database (Narwati, 2010).

B. Clustering

Proses pengelompokan sekumpulan objek ke dalam kelas-kelas objek yang sama disebut clustering/pengelompokan. Pengklasteran merupakan satu dari sekian banyak fungsi pada proses data mining untuk menemukan kelompok atau mengidentifikasi kelompok objek yang memiliki kesamaan karakteristik. Analisis kluster (*cluster*) adalah usaha untuk mengidentifikasi kelompok objek yang memiliki kemiripan satu sama lain serta membantu menemukan pola penyebaran dan pola hubungan dalam sekumpulan data yang besar. Hal penting dalam proses pengklasteran adalah menempatkan sekumpulan pola pada data ke kelompok yang sesuai untuk menemukan kesamaan dan perbedaan sehingga mendapatkan kesimpulan yang berharga (Narwati, 2010).

Menurut (Abdillah et al., 2016), algoritma *clustering* diklasifikasikan ke dalam 5 kategori, yaitu:

1. Partitioning Methods

Pengelompokan objek dimana masing-masing objek dimiliki oleh 1 *cluster*. Algoritma partisi dapat meminimalkan mean kuadrat jarak dari setiap titik data ke pusat terdekat nya. Yang termasuk ke dalam metode Partitioning Methods ini adalah algoritma *K-Means*, *K-Medoids* atau *PAM*, *CLARA*, dan *CLARANS.PAM*.

2. Hierarchical Methods

Pengelompokan objek dapat dilakukan dengan 2 cara, *agglomerative* yang dimulai dengan menggabungkan beberapa *cluster* hingga menjadi satu, atau *divisive* yang dimulai dengan *cluster* yang lebih kecil. Yang termasuk ke dalam metode Hierarchical Methods antara lain algoritma *BIRCH*, *Chameleon*, dan *CURE*.

3. Density-Based Methods

Pengelompokan objek berdasarkan tingkat kerapatan objek atau densitas. Yang termasuk dalam metode ini adalah algoritma *DENCLUE*, *OPTICS*, dan *DBSCAN*.

4. Grid-Based Methods

Pengelompokan objek dengan menggunakan struktur data grid multiresolusi yang mampu menangani data berdimensi tinggi. Yang termasuk dalam metode ini adalah algoritma *STING*, *CLIQUE*, dan *WaveCluster*.

5. Model-Based Methods

Pengelompokan objek dengan memodelkan tiap-tiap *cluster*, dan mencoba mengoptimasikan kesesuaian data dengan model matematika. Yang

termasuk dalam metode ini salah satunya adalah algoritma *COBWEB*.

C. Partitioning Around Medoids (PAM)

Algoritma PAM dikembangkan oleh Leonard Kaufman dan J. Rousseeuw, dan algoritma ini mirip dengan *K-Means*, terutama keduanya algoritma partitional dengan kata lain, keduanya memecah dataset menjadi kelompok-kelompok, dan keduanya bekerja secara acak, lalu proses perhitungan *cost* dilakukan kembali. Apabila total *cost* yang dihasilkan lebih kecil dari *cost* setiap objek dengan *medoid* lama, maka objek baru tersebut dapat menjadi *medoids* baru. *Iterasi* terakhir sampai tidak terdapat perubahan *cost* terhadap *cost* yang dihasilkan oleh *medoid* baru.

Menurut Han dan Kamber, 2006 (Agarwal, 2014) (Data Mining Concepts and Techniques 2nd Edition, Hal 406), tahapan algoritma PAM sebagai berikut:

1. Secara acak, pilih k objek pada sekumpulan n objek sebagai *medoid*.
 2. Ulangi:
 3. Tempatkan objek *non-medoid* ke dalam *cluster* yang paling dekat dengan *medoid* menggunakan metode perhitungan jarak *Euclidean Distance*.
- $$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$
4. Secara acak, pilih O_{random} : sebuah objek *non-medoid*.
 5. Hitung total biaya, S , dari pertukaran *medoid* O_j dengan O_{random} .
 6. Jika $S < 0$ maka tukar O_j dengan O_{random} untuk membentuk sekumpulan k objek baru sebagai *medoid*.
 7. Ulangi proses 3-6 hingga tidak ada perubahan.

D. Distance Measure

Dalam melakukan *pattern matching* ataupun untuk melakukan berbagai jenis pengelompokan (*clustering*), *similarity measure* merupakan bagian penting yang harus diperhatikan karena mempengaruhi perhitungan dalam menentukan anggota suatu *cluster*. Dalam penelitian ini akan digunakan tiga *distance measure* yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Chebyshev Distance*.

E. Normalisasi Z-Score

Normalisasi adalah proses penskalaan nilai atribut dari data sehingga bisa terletak pada rentang tertentu. *Normalisasi Z-score* atau *Z-score Normalization* merupakan metode *normalisasi* berdasarkan *mean* (nilai rata-rata) dan *standard deviation* (deviasi standar) dari data. Metode ini sangat berguna jika tidak diketahui nilai aktual *minimum* dan *maksimum* dari data (Nasution et al., 2019). Rumus persamaan yang digunakan untuk menghitung normalisasi z-score yaitu:

$$Z_i = \frac{x_i - \bar{x}}{s} \quad (2)$$

F. Silhouette Coefficient

Silhouette Coefficient merupakan suatu metode yang digunakan untuk mengukur mutu kualitas dan kekuatan dari *cluster*. Metode *Silhouette Coefficient* merupakan gabungan dari dua metode perhitungan yaitu, metode *cohesion* yang berfungsi untuk mengukur seberapa dekat relasi antara objek dalam sebuah *cluster*, dan metode *separation* yang berfungsi untuk mengukur seberapa jauh jarak sebuah *cluster* terpisah dengan *cluster* lain. Tahapan perhitungan *Silhouette Coefficient* (Handoyo et al., 2014).

1. Hitung rata-rata jarak objek dengan semua objek lain yang berada di dalam satu *cluster* dengan persamaan:

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (3)$$

Dimana:

J = dokumen lain dalam satu *cluster* A

$d(i, j)$ = jarak antara dokumen i dengan j

2. Hitung rata-rata jarak dari dokumen i tersebut dengan semua dokumen di *cluster* lain, dan ambil nilai terkecilnya menggunakan persamaan:

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (4)$$

Dimana:

$d(i, C)$ = jarak rata-rata dokumen i dengan semua objek pada *cluster* lain C dimana $A \neq C$.

$$b(i) = \min_{C \neq A} d(i, C) \quad (5)$$

3. Nilai *Silhouette Coefficient* nya adalah:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Kriteria subjektif pengukuran pengelompokkan berdasarkan *Silhouette Coefficient* (SC) menurut Kauffman dan Roesseeuw, dapat dilihat pada tabel 1.

Tabel 1. Kriteria Pengukuran Silhouette Coefficient

Nilai SC	Kriteria
0,71 – 1,00	Struktur Kuat
0,51 – 0,70	Struktur Baik
0,26 – 0,50	Struktur Lemah
≤ 0,25	Struktur Buruk

3. METODE PENELITIAN

A. Tahapan Pelaksanaan Penelitian



Gambar 1. Tahapan Penelitian

Tahapan pada penelitian ini dilaksanakan pada bulan November 2020 sampai dengan Juli 2021 bertempat di Dinas Perkebunana Provinsi Kalimantan Timur Jalan MT. Haryono, Air Putih, Kecamatan Samarinda Ulu, Kota Samarinda. Alur tahapan penelitian dapat dilihat pada gambar 1.

1. Kajian Pustaka

Penulis mempelajari serta memahami hal-hal yang bersangkutan dan mencari referensi dari buku, jurnal, literature, artikel, website dan lain-lain.

2. Perumusan dan Pembatasan Masalah

Penulis merumuskan rumusan masalah dan memberikan batasan pada masalah-masalah tersebut.

3. Pengumpulan Data

Penulis mengumpulkan data yang bersumber dari Dinas Perkebunan Provinsi Kalimantan Timur yang akan digunakan sebagai *inputan* untuk penelitian. Data yang digunakan adalah data hasil produksi komoditi perkebunan tahun 2000 – 2019.

4. Pre-processing / Cleaning Data

Penulis memilah data yang akan digunakan dengan cara membuang data yang tidak konsisten.

5. Transformasi Data

Data yang sudah melewati proses *Pre-processing / Cleaning Data*, ditransformasikan atau diubah terlebih dahulu dalam bentuk angka/numerikal

untuk memperoleh hasil perhitungan yang optimal.

6. *Clustering Partitioning Around Medoids* (PAM)

Pada tahap ini, penulis melakukan perhitungan dengan algoritma *Partitioning Around Medoids* (PAM) untuk mengelompokkan/*clustering* data.

7. Hasil dan Kesimpulan

Pada tahap ini, diperoleh hasil dan kesimpulan sebagai hasil analisis akhir.

B. Perancangan Data

Jenis data yang digunakan pada penelitian ini adalah data produksi dari komoditi perkebunan selama 6 tahun terakhir di Provinsi Kalimantan Timur tahun 2014 sampai dengan 2019 yang berupa file format *.pdf* yang kemudian diubah ke dalam bentuk dokumen yang berekstensi *.xlsx*.

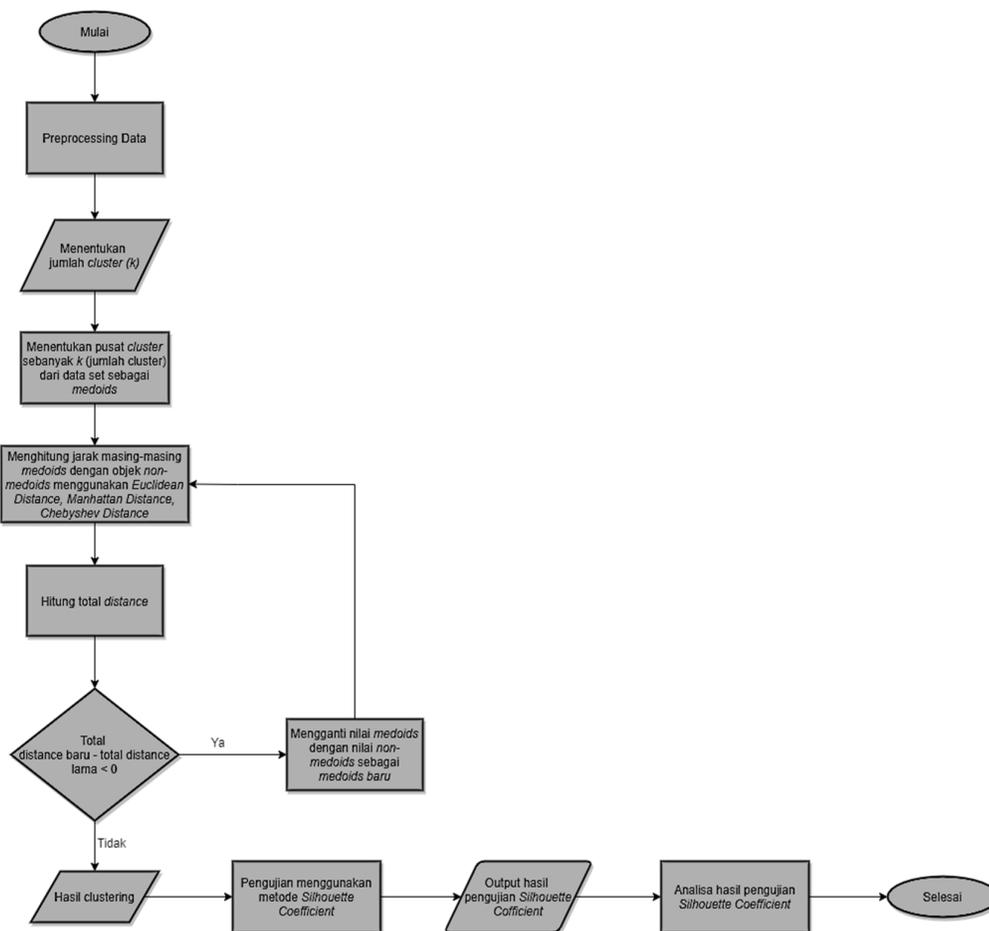
Tabel 2. Contoh Sampel Data Tahun 2015

Kabupaten/ Kota	Luas	Produksi	TKP	Tahun
Berau	115195	1253891	4091	2015
Kutai Barat	114440	523345	52892	2015
Kutai				
Kertanegara	191366	1506913	25458	2015
Kutai Timur	424311	5694529	72106	2015
Mahakam Ulu	12341	0	6171	2015
Paser	182145	1364430	74745	2015
Penajam				
Paser Utara	48865	462616	11972	2015
Balikpapan	21	75	8	2015
Bontang	52	0	33	2015
Samarinda	1370	7094	613	2015

Dari tabel 2 dapat dilihat contoh data produksi dan luas lahan untuk komoditi kelapa sawit tahun 2015. Parameter yang digunakan pada penelitian ini berjumlah 3 parameter, yaitu *luas lahan*, *produksi*, dan *tenaga kerja perkebunan*.

C. Perancangan Algoritma Partitioning Around Medoids (PAM)

Secara garis besar, perancangan algoritma *Partitioning Around Medoids* (PAM) dapat dilihat pada gambar 2



Gambar 2. Flowchart Partitioning Around Medoids

Flowchart pada gambar 2 dapat dijelaskan sebagai berikut:

1. Menentukan jumlah data yang akan di *cluster*, dimana sampel data adalah data produksi komoditas perkebunan Provinsi Kalimantan Timur.
2. Menetapkan nilai *k* jumlah *cluster* data sebanyak *k cluster*.
3. Menentukan pusat *cluster* awal yang ditentukan secara acak berdasarkan variabel data yang di *cluster* sebanyak *k* yang telah ditentukan sebelumnya.
4. Pilih secara acak objek pada tiap-tiap *cluster* sebagai kandidat *medoid* baru.
5. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoid* baru menggunakan *Euclidean Distance*, *Manhattan Distance*, *Chebyshev Distance*.
6. Hitung total selisih jarak (*S*) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan objek data lain di *cluster* untuk membentuk sekumpulan *k* objek sebagai *medoid* baru.
7. Ulangi langkah 3 sampai dengan langkah 5 hingga tidak terjadi perubahan pada *medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.
8. Setelah proses *clustering* selesai, maka selanjutnya pengujian menggunakan metode *Silhouette Coefficient* yang memberikan informasi untuk melihat kualitas dan keakuratan dari *cluster*.
9. Hasil pengujian *Silhouette Coefficient*.
10. Selesai

D. Perancangan Pengujian

Pengujian dalam penelitian ini dilakukan menggunakan metode *Silhouette Coefficient* untuk melihat kualitas dan kekuatan dari *cluster* yang terbentuk. Dalam *Silhouette Coefficient*, suatu *cluster* dikatakan baik ketika dalam pengujian menghasilkan nilai tertinggi atau mendekati 1. Kriteria subjektif pengukuran baik atau tidaknya *cluster* yang terbentuk berdasarkan *Silhouette Coefficient* dapat dilihat pada tabel 1.

Metode perhitungan jarak yang akan digunakan pada penelitian ini adalah metode perhitungan jarak *Euclidean Distance*, *Manhattan Distance*, dan *Chebyshev Distance*. Dari ketiga metode perhitungan jarak ini, akan dilihat metode manakah yang lebih baik dalam analisis *cluster* dengan menggunakan metode *Silhouette Coefficient*.

4. HASIL DAN PEMBAHASAN

A. Pre-processing Data

Preprocessing merupakan salah satu tahapan penting dalam proses *data mining*. Data yang digunakan dalam proses *data mining* tidak selamanya dalam kondisi ideal untuk diproses. Terkadang pada data tersebut terdapat berbagai permasalahan yang dapat mengganggu hasil dari proses mining itu sendiri seperti diantaranya adalah *missing value*, *data*

redundant, *outliers*, ataupun format data yang tidak sesuai. Oleh karena itu, untuk mengatasi permasalahan tersebut dibutuhkan tahap *pre-processing*. Tahapan tersebut diantaranya:

1. Cleaning Data

Cleaning data adalah proses menyiapkan data untuk dilakukan analisis dengan cara menghapus atau memodifikasi data yang tidak relevan dan tidak sesuai. Data yang akan di-*cleaning* pada penelitian ini adalah data hasil produksi komoditi selama 20 tahun terakhir yakni tahun 2000 – 2019. Dikarenakan data hasil produksi komoditi perkebunan tahun 2000 – 2013 tidak relevan, dimana Kabupaten/Kota yang ada di Provinsi Kalimantan Utara masih menjadi bagian dari Provinsi Kalimantan Timur, maka data yang digunakan adalah data hasil produksi komoditi tahun 2014 – 2019 khusus untuk Kabupaten/Kota Provinsi Kalimantan Timur sebagai data yang akan diolah.

2. Normalisasi Data

Berdasarkan data hasil produksi komoditi perkebunan tahun 2014, 2015, 2016, 2017, 2018, dan 2019 Dinas Perkebunan Provinsi Kalimantan Timur, dilakukan *normalisasi* untuk menyamakan rentang nilai tiap data. *Normalisasi* yang digunakan pada penelitian ini adalah *Z-Score normalisasi* dengan rumus persamaan 2.

B. Perhitungan Clustering Partitioning Around Medoids (PAM)

Dilakukan perhitungan *clustering Partitioning Around Medoids* (PAM) secara manual menggunakan *Microsoft Excel*. Menentukan jumlah *cluster* yang akan dibentuk adalah tahapan pertama dalam melakukan perhitungan *clustering PAM*. Pada penelitian ini, jumlah *cluster* yang akan dibentuk sebanyak 3 *cluster*, 5 *cluster*, dan 7 *cluster*.

Tabel 3. Produksi Kelapa Sawit 2014-2019

No	Kabupaten/Kota	TKP	Luas Lahan(Ha)	Produksi (Ton)
1	Berau	37675	107920	919522
2	Kutai Barat	47084	96084	464199
...
...
...
58	Balikpapan	20	36	469
59	Bontang	34	72	336
60	Samarinda	610	1308	10644

1. Langkah awal yang dilakukan dalam perhitungan manual *clustering Partitioning Around Medoids* (PAM) adalah melakukan *normalisasi* data. Dengan menggunakan data asli pada tabel 3, langkah selanjutnya yaitu mencari nilai rata-rata untuk setiap variabel.

Rata-rata *variable* tenaga kerja perkebunan

$$\bar{x} = \frac{37675 + 47084 + 23109 + \dots + 610}{60} = \frac{1429654}{60}$$

$$\bar{x} = 24442,067$$

Rata-rata *variable* luas lahan

$$\bar{X} = \frac{107920 + 96084 + 180210 + \dots + 1308}{60} = \frac{6880484}{60}$$

$$\bar{X} = 114674,7$$

Rata-rata *variable* produksi

$$\bar{X} = \frac{919522 + 464199 + 1231387 + \dots + 10644}{60} = \frac{76765600}{60}$$

$$\bar{X} = 1279426,7$$

Kemudian, menghitung nilai simpangan baku untuk setiap *variable* dengan menggunakan rumus persamaan 4.3

Simpangan baku *variable* tenaga kerja perkebunan

$$s = \sqrt{\frac{(37675 - 23827,6)^2 + (47084 - 23827,6)^2 + \dots + (610 - 23827,6)^2}{60 - 1}} = 25668,809$$

Simpangan baku *variable* luas lahan

$$s = \sqrt{\frac{(107920 - 114674,7)^2 + (96084 - 114674,7)^2 + \dots + (1308 - 114674,7)^2}{60 - 1}}$$

$$s = 133080,355$$

Simpangan baku *variable* produksi

$$s = \sqrt{\frac{(919522 - 1279426,7)^2 + (464199 - 1279426,7)^2 + \dots + (10644 - 1279426,7)^2}{60 - 1}} = 1798003,794$$

Hasil normalisasi data untuk produksi kelapa sawit dapat dilihat pada tabel 4.

Tabel 4. Hasil Normalisasi Data Kelapa Sawit

No	Kabupaten/Kota	TKP	Luas Lahan(Ha)	Produksi (Ton)
1	Berau	0.516	-0.051	-0.200
2	Kutai Barat	0.883	-0.140	-0.453
...
...
58	Balikpapan	-0.95	-0.861	-0.711
59	Bontang	-0.95	-0.861	-0.711
60	Samarinda	-0.93	-0.852	-0.706

- Selanjutnya, tentukan *k* (jumlah *cluster*) yang diinginkan. Pada kasus penelitian ini, jumlah *k* ditetapkan sebanyak 3 (*k* = 3)

Tabel 5. Kandidat Medoid Sementara

No	Kabupaten/Kota	TKP	Luas Lahan (Ha)	Produksi (Ton)
C1	Berau	0.516	-0.051	-0.200
C2	Kutai Barat	0.883	-0.140	-0.453
C3	Kutai Kertanegara	-0.052	0.492	-0.027

- Setelah menetapkan kandidat *medoid* sementara, selanjutnya hitung jarak masing-masing objek ke kandidat *medoid* sementara menggunakan rumus *distance measure*. *Distance measure* yang akan digunakan pada tahap ini yaitu *euclidean distance*. Jarak *Euclidean Distance* data ke-1 dengan *centroid* 1:

$$d_{1,1} = \sqrt{(0,516 - 0,516)^2 + (-0,051 - (-0,051))^2 + (-0,200 - (-0,200))^2} = 0,000$$

Jarak *Euclidean Distance* data ke-1 dengan *centroid* 2:

$$d_{1,2} = \sqrt{(0,906 - 0,516)^2 + (-0,140 - (-0,051))^2 + (-0,453 - (-0,200))^2} = 0,455$$

Jarak *Euclidean Distance* data ke-1 dengan *centroid* 3:

$$d_{1,3} = \sqrt{(-0,028 - 0,516)^2 + (0,492 - (-0,051))^2 + (-0,027 - (-0,200))^2} = 0,805$$

- Setelah didapatkan hasil perhitungan jarak antara objek data dengan *medoid*, hitung total jarak terdekat untuk setiap data ke *medoid*. Selanjutnya, ambil jarak terdekat antar objek dengan *medoid* sementara. Jarak terdekat antara objek data dengan *medoid* menandakan *cluster* dari setiap objek.

Tabel 6. Iterasi ke-1

No	Cost 1	Cost 2	Cost 3	Terdekat	Cluster
1	0.000	0.455	0.805	0.000	1
2	0.455	0.000	1.207	0.000	2
3	0.805	1.207	0.000	0.000	3
...
...
...
58	1.754	1.990	1.764	1.754	1
59	1.754	1.989	1.764	1.754	1
60	1.729	1.964	1.743	1.729	1
Total Kedekatan				34,531	

- Setelah didapatkan hasil perhitungan jarak untuk setiap objek dengan *medoid* sementara, total jarak terdekat, dan juga penentuan *cluster*. Langkah selanjutnya adalah melakukan *iterasi medoid*. Ulangi langkah-langkah 2-4 dengan mengganti *medoid* sementara dengan *medoid* baru.

Tabel 7. Medoid Baru

No	Kabupaten/Kota	TKP	Luas Lahan(Ha)	Produksi (Ton)
C1	Berau	0.171	0.153	0.608
C2	Kutai Barat	-0.19	0.229	-0.088
C3	Kutai Kertanegara	0.168	0.881	1.201

Setelah didapatkan hasil perhitungan jarak antara objek data dengan *medoid* baru, hitung total jarak terdekat untuk setiap data ke *medoid*. Selanjutnya, ambil jarak terdekat antar objek dengan *medoid*

baru. Jarak terdekat antara objek data dengan *medoid* menandakan *cluster* dari setiap objek.

Tabel 8. Iterasi ke-2

No	Cost 1	Cost 2	Cost 3	Terdekat	Cluster
1	0.902	0.776	1.719	0.776	2
2	1.311	1.200	2.072	1.200	2
3	0.753	0.308	1.307	0.308	2
...
...
...
...
...
No	Cost 1	Cost 2	Cost 3	Terdekat	Cluster
58	2.008	1.465	2.820	1.465	2
59	2.008	1.464	2.820	1.464	2
60	1.987	1.444	2.801	1.444	2
Total Kedekatan				72,144	

6. Selanjutnya, jika sudah didapat jumlah *cost* untuk setiap data pada tiap *iterasi*, dilanjutkan dengan menghitung nilai selisih antara jumlah *cost* pada *iterasi* baru dengan *iterasi* lama. Adapun jika selisih yang didapatkan kurang dari 0, maka proses *iterasi* tetap dilakukan hingga didapatkan hasil nilai selisih lebih dari 0 yang dimana proses *iterasi* berhenti.

Tabel 9. Selisih iterasi ke-2 dan ke-1

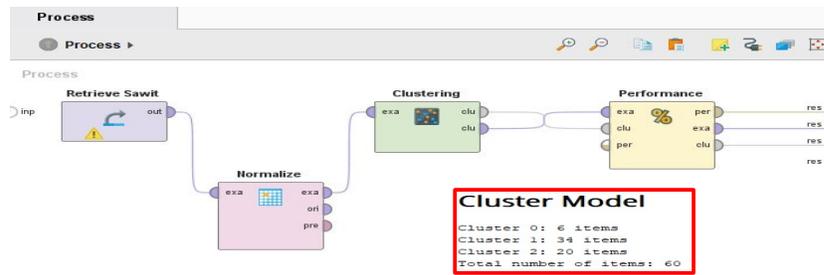
Iterasi ke-	Total Cost	Selisih
1	34,531	
2	72,144	
		37,613

Karena selisih antara *cost* baru – *cost* lama yakni iterasi ke 2 – iterasi ke 1 sudah > 0, maka proses perhitungan dihentikan dan hasil *cluster* ada pada terasi ke 1.

7. Selanjutnya, jika sudah mendapatkan *iterasi* terakhir, peneliti mengelompokkan anggota *cluster* sesuai dengan jarak terdekat dan mengalokasikan ke tiap *cluster*. Dari hasil *iterasi* ke-2, diperoleh data pada *cluster* satu (C1) berjumlah 6 data, *cluster* dua (C2) berjumlah 34 data, dan *cluster* 3 (C3) berjumlah 20 data.

C. Pemodelan Menggunakan Rapid Miner Studio

Setelah melakukan perhitungan algoritma *Partitioning Around Medoids* (PAM) secara manual, peneliti juga melakukan pemodelan algoritma PAM dengan bantuan software *RapidMiner Studio*. Pemodelan menggunakan dataset yang sama dengan data yang digunakan untuk perhitungan manual. Dataset yang digunakan sebanyak 60 data dengan 3 atribut dimana seluruh atribut bertipe data *numeric*. Data dibuat dalam bentuk dataset baru dengan format *.xlsx (excel)* untuk bisa dianalisis.



Gambar 3. Proses Pengujian Algoritma PAM dan Hasil Pengujian RapidMiner

Berdasarkan hasil dari pemodelan menggunakan bantuan software *RapidMiner*, peneliti mencoba membandingkan dengan hasil perhitungan secara manual. Kesimpulan yang didapat dari perbandingan antara sistem dan perhitungan manual yaitu sistem telah mengimplementasikan algoritma dengan baik dengan menghasilkan jumlah dan data anggota *cluster* yang sama dengan perhitungan secara manual

D. Evaluasi Silhouette Coefficient

Dalam evaluasi *cluster*, peneliti menggunakan uji pemodelan menggunakan *tools* pengolah angka yaitu *Microsoft Excel* dalam melakukan perhitungan manual untuk mengetahui kualitas dan kekuatan *cluster* dengan metode *Silhouette Coefficient* (SC) pada *cluster* yang telah terbentuk.

Tabel 10. Nilai Silhouette Coefficient

Data ke-	Kabupaten/Kota	Tahun	SC
1	Berau	2014	0.058184235
2	Kutai Barat	2014	0.061783902
	Kutai	2014	
3	Kertanegara	2014	-0.02204837
4	Kutai Timur	2014	0.457619303
...
...
...
60	Samarinda	2019	0.813517346
Rata-rata Silhouette Coefficient			0.547311963

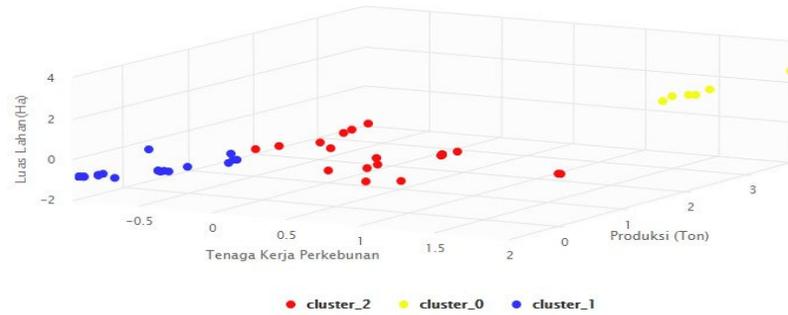
Berdasarkan perhitungan diatas, nilai rata-rata *Silhouette Coefficient* yang didapat senilai 0,547311963. Dapat disimpulkan bahwa pengelompokan data algoritma *Partitioning Around*

Medoids (PAM) berdasarkan tabel 1 menunjukkan struktur baik.

E. Pembahasan

Penelitian ini bertujuan untuk mengetahui daerah potensial penghasil komoditi perkebunan menggunakan algoritma *Partitioning Around Medoids* (PAM) agar dapat membantu Dinas Perkebunan Provinsi Kalimantan Timur berdasarkan informasi hasil *clustering* wilayah Kabupaten/Kota

sehingga pihak Dinas Perkebunan dapat memberi perhatian lebih pada daerah yang menghasilkan produksi terendah. Data yang digunakan pada penelitian ini berjumlah 540 data yang terdiri dari 9 jenis komoditi perkebunan dengan 3 atribut. Metode yang digunakan untuk menentukan *cluster* yaitu PAM serta *Silhouette Coefficient* sebagai metode pengukur validitas dari hasil perhitungan dengan menggunakan metode PAM.



Gambar 4. Scatter Plot Hasil Clustering Produksi Kelapa Sawit

F. Perbandingan Jumlah Cluster

Penelitian ini dilakukan dengan mengambil data selama 6 tahun terakhir untuk hasil produksi komoditi perkebunan di Provinsi Kalimantan Timur yaitu sebanyak 540 data. Data yang akan diuji adalah data dengan atribut tenaga kerja perkebunan, luas lahan, dan hasil produksi untuk setiap komoditi yang ada. Jumlah *cluster* yang akan diuji adalah $k=3$, $k=5$, $k=7$.

Perbandingan jumlah *cluster* dilakukan untuk menentukan jumlah k terbaik dalam proses *clustering* dengan melihat rata-rata *silhouette coefficient* yang didapat untuk setiap *cluster*.

Berikut tabel menampilkan hasil *Silhouette Coefficient* yang didapat untuk 3 *cluster*, 5 *cluster* dan 7 *cluster*.

Tabel 11. Perhitungan Silhouette Coefficient 3 Cluster

Komoditi	Euclidean Distance	Manhattan Distance	Chebyshev Distance
Kelapa Sawit	0.5473119	0.6880784	0.2702499
Karet	0.3121135	0.6641440	0.3281229
Kakao	0.7513280	0.8233865	0.5945630
Kelapa Dalam	0.5477539	0.6657067	0.2637072
Lada	0.7468282	0.8459168	0.5096249
Kopi	0.6028689	0.6765047	0.3186268
Aren	0.6105490	0.7611584	0.4402875
Kemiri	0.5856964	0.6589594	0.3093786
Kapok	0.7520581	0.8992299	0.3186268
Rata-rata	0.60627871	0.742565027	0.3725764

Tabel 12 Perhitungan Silhouette Coefficient 5 Cluster

Komoditi	Euclidean Distance	Manhattan Distance	Chebyshev Distance
Kelapa Sawit	0.5397000	0.5930242	0.2927752
Karet	0.4208304	0.6041030	0.5522554
Kakao	0.6568856	0.705741	0.705741
Kelapa Dalam	0.6274485	0.7226356	0.735797
Lada	0.7732242	0.788060	0.7854559
Kopi	0.5687734	2.892831	0.6073808
Aren	0.6602723	0.7295853	0.5370413
Kemiri	0.5668748	0.6398106	0.4893592
Kapok	0.8114039	0.9165241	0.9165241
Rata-rata	0.625045	0.9547019	0.6247034

Tabel 13 Perhitungan Silhouette Coefficient 7 Cluster

Komoditi	Euclidean Distance	Manhattan Distance	Chebyshev Distance
Kelapa Sawit	0.5788644	0.6248100	-0.068095
Karet	0.5717241	0.6511926	0.6214519
Kakao	0.4812153	0.2680232	0.6398579
Kelapa Dalam	0.6205027	0.6989465	0.6745564
Lada	0.5984533	0.6550636	0.6550636
Kopi	0.5360847	0.6620500	0.6194765
Aren	0.6227916	0.6470452	0.7318458
Kemiri	0.6908314	0.7636240	0.6082910
Kapok	-1	0.2588564	0.6131928
Rata-rata	0.4111631	0.581068	0.566182

5. KESIMPULAN

Berdasarkan hasil yang telah dilakukan maka diperoleh kesimpulan sebagai berikut:

1. Implementasi algoritma *Partitioning Around Medoids* (PAM) terhadap *clustering* komoditi perkebunan pada metode perhitungan jarak *Euclidean Distance*, *Manhattan Distance*, dan *Chebyshev Distance* telah berhasil dilakukan. Hasil dengan 5 *cluster* lebih ideal dibandingkan dengan hasil 3 *cluster* dan 7 *cluster* karena memiliki nilai *Silhouette Coefficient* yang optimal pada jumlah *cluster* 5 yaitu 0.9547019.
2. Analisa perhitungan hasil *clustering Partitioning Around Medoids* (PAM) menggunakan metode *Euclidean Distance*, *Manhattan Distance* dan *Chebyshev Distance*, menghasilkan nilai yang optimal adalah *Manhattan Distance*.

6. DAFTAR PUSTAKA

- Abdillah, G., Putra, F. A., Renaldi, F., Informatika, P. S., Jenderal, U., Yani, A., Barat, J., & Cimahi, K. (2016). Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means. *Sentika 2016, 2016*(Sentika), 18–19.
- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>
- Anggreini, N. L. (2019). Teknik Clustering Dengan Algoritma K-Medoids Untuk Menangani Strategi Promosi Di Politeknik Tedc Bandung. *Jurnal Teknologi Informasi Dan Pendidikan, 12*(2), 1–7. <https://doi.org/10.24036/tip.v12i2.215>
- disbun.kaltimprov.go.id. (2019). *Potensi Daerah Provinsi Kalimantan Timur*. Disbun.Kaltimprov.Go.Id.
- Handoyo, R., Rumani, R., & Nasution, S. M. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K-Means Pada Pengelompokan Dokumen. *JSM*
- STMIK Mikroskil, 15*(2), 73–82.
- Hardiyanti, F., Tambunan, H. S., & Saragih, I. S. (2019). Penerapan Metode K-Medoids Clustering Pada Penanganan Kasus Diare Di Indonesia. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer), 3*(1), 598–603. <https://doi.org/10.30865/komik.v3i1.1666>
- Haviluddin, H., Patandianan, S. J., Putra, G. M., Puspitasari, N., & Pakpahan, H. S. (2021). Implementasi Metode K-Means Untuk Pengelompokan Rekomendasi Tugas Akhir. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer, 16*(1), 13. <https://doi.org/10.30872/jim.v16i1.5182>
- Marlina, D., Lina, N., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi, 4*(2), 64. <https://doi.org/10.24014/coreit.v4i2.4498>
- Narwati, N. (2010). Pengelompokan Mahasiswa Menggunakan Algoritma K-Means. *Jurnal Dinamika Informatika, 2*(2).
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal, 4*(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>
- Riyanto, B. (2019). Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer), 3*(1), 562–568. <https://doi.org/10.30865/komik.v3i1.1659>
- Wulandari, S., & Dwitiyanti, N. (2020). Implementasi Algoritma Clustering Partitioning Around Medoid (PAM) Dalam Clustering Virus MERS-CoV. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi), 5*(1), 70–77.