



Tersedia Online : <http://e-journals.unmul.ac.id/>

ADOPSI TEKNOLOGI DAN SISTEM INFORMASI (ATASI)

Alamat Jurnal : <http://e-journals2.unmul.ac.id/index.php/atasi/index>



Implementasi Metode K-Means Untuk Pengelompokan Potensi Produksi Komoditas Perkebunan

Herman Santoso Pakpahan ^{1)*}, Joan Angelina Widians ²⁾, Haga Daffa Aska Firmanda ³⁾, Yuniarta Basani ⁴⁾

Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman
E-Mail : pakpahan.herman891@gmail.com^{1)*}

ARTICLE INFO

Article history:

Received : 13 Juni 2022

Revised : 17 Juni 2022

Accepted : 28 Juni 2022

Available online : 30 Juni 2022

Keywords :

Clustering,
Plantation Commodities,
K-Means,
SSE,
Euclidean Distance,
Data Mining

Kata Kunci :

Clustering,
Komoditas Perkebunan,
K-Means,
SSE,
Euclidean Distance,
Data Mining

APA style in citing this article:

Pakpahan, H. S., Widians, J. A., & Firmanda, H. D. A. (2022). Implementasi Metode K-Means Untuk Pengelompokan Potensi Produksi Komoditas Perkebunan. *Adopsi Teknologi Dan Sistem Informasi (ATASI)*, 1(1), 52 - 60. <https://doi.org/10.30872/atasi.v1i1.49>

ABSTRACT

This study aims to classify plantation commodities in East Kalimantan because of the large number of commodities and the large amount of data, a grouping system is needed in order to provide information about the potential in an area that is needed by the community as well as farmers and institutions related to the information. Clustering of plantation commodity data uses the K-Means algorithm which is the right method to be used as a reference in determining plantation commodities in an area that has high, medium, and low levels of potential production. The data used in this study is data obtained from the Plantation Office of East Kalimantan as much as 400 data using 5 attributes. In this study, the results obtained in cluster 1 as many as 16 plantation commodity data, cluster 2 as many as 14 plantation commodity data, and cluster 3 as many as 370 plantation commodity data. Based on the calculation of the level of accuracy that the Euclidean Distance measurement method shows an accuracy rate of 93.75% and the calculation of the Sum of Squared Error (SSE) shows an error rate of 6.25%. This shows that the Euclidean Distance measurement method is quite accurate in grouping plantation commodities, because the Sum of Squared Error (SSE) value does not exceed 50%. A web-based system for grouping plantation commodities in East Kalimantan using the K-Means method has been successfully developed.

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan komoditas perkebunan di Kalimantan Timur karena dilihat banyaknya komoditas dan jumlah data yang banyak maka diperlukan suatu sistem pengelompokan agar dapat memberikan informasi tentang potensi disuatu wilayah yang diperlukan oleh masyarakat maupun petani dan lembaga-lembaga yang terkait dengan informasi tersebut. Clustering data komoditas perkebunan menggunakan algoritma K-Means yang merupakan metode yang tepat untuk digunakan sebagai acuan dalam menentukan komoditas perkebunan disuatu wilayah yang memiliki tingkat potensi produksi banyak, sedang, dan sedikit. Data yang digunakan pada penelitian ini merupakan data yang diperoleh dari Dinas Perkebunan Kalimantan Timur sebanyak 400 data dengan menggunakan 5 atribut. Dalam penelitian ini didapatkan hasil pada cluster 1 sebanyak 16 data komoditas perkebunan, cluster 2 sebanyak 14 data komoditas perkebunan, dan cluster 3 sebanyak 370 data komoditas perkebunan. Berdasarkan perhitungan tingkat akurasi bahwa metode pengukuran jarak Euclidean Distance menunjukkan tingkat akurasi sebesar 93,75% dan perhitungan Sum of Squared Error (SSE) menunjukkan tingkat error sebesar 6.25%. hal ini menunjukkan bahwa metode pengukuran jarak Euclidean Distance cukup akurat dalam pengelompokan komoditas perkebunan, karena nilai Sum of Squared Error (SSE) tidak melebihi 50%. Sistem pengelompokan komoditi perkebunan daerah Kalimanta Timur berbasis web dengan menerapkan metode K-Means telah berhasil dibangun.

2022 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

*) Correspondenting Author

<https://doi.org/10.30872/atasi.v1i1.49>

1. PENDAHULUAN

Sektor perkebunan sangat berperan penting dalam mensukseskan pelaksanaan strategi transformasi ekonomi di Indonesia, istilah komoditas perkebunan umumnya merujuk kepada sekelompok tanaman atau komoditas tertentu. Namun permasalahan yang timbul adalah aktor yang berperan di sini masih sangat minim, terutama dari masyarakat lokalnya yang memiliki pengetahuan terbatas mengenai potensi lokal yang ada di daerahnya (Priatna, Y, 2014). Tugas penulis adalah mengelompokkan potensi produksi komoditas perkebunan di daerah Kalimantan Timur. karena dilihat dari banyaknya komoditas dan jumlah data yang banyak maka diperlukan suatu sistem pengelompokkan untuk mempermudah instansi maupun pemerintah dalam mengelompokkan data dengan jumlah yang banyak dan agar dapat memberikan informasi tentang potensi disuatu wilayah yang diperlukan oleh masyarakat maupun petani dan lembaga-lembaga yang terkait dengan informasi tersebut. Era teknologi dan globalisasi yang terjadi saat ini mendorong timbulnya kebutuhan manusia akan informasi. Kemajuan teknologi yang begitu cepat mendorong manusia dalam memanfaatkan teknologi tersebut untuk melakukan pekerjaan yang dahulu dikerjakan secara manual (Widians, J.A, Rachman,W & Masnawati, 2017).

Oleh karena masalah di atas, penelitian ini dibuat dengan mengembangkan informasi dengan memanfaatkan teknologi komputer terutama dalam bidang data mining. Salah satu jenis pengelompokkan dalam data mining adalah Clustering. Clustering ini biasa dilakukan agar sebuah data yang tadinya masih dalam keadaan tidak terusun dapat dikelompokkan menjadi satu kesatuan agar data tersebut menjadi lebih tertata. Penelitian ini dibuat untuk mempermudah pengelompokkan potensi komoditas perkebunan khususnya di Kalimantan Timur.

2. TINJAUAN PUSAKA

A. Data Mining

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Hermawan, T.N, Ugiarto, M, & Puspitasari, N, 2017). Data Mining adalah kenyataan bahwa Data Mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, Data Mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani jumlah data yang sangat besar, dimensi data yang tinggi, data yang heterogen dan berbeda sifat pengelompokan (Maulana, A, Fajrin, A, 2018)

B. Clustering

Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain (Anggara, M., Sujiani, H, Helfi, 2016). Klasifikasi adalah algoritma yang mampu mengklasifikasikan atau meng-cluster objek berdasarkan pada karakteristik ciri-ciri yang diberikan (Wati, M, Indrawan W, & Widians, 2017).

C. Website

Website adalah kumpulan dari halaman-halaman situs yang terangkum dalam sebuah domain atau subdomain, yang tempatnya berada di dalam World Wide Web (WWW) di dalam internet) (Widians, J, Wati, M, & Awaludin, 2107). Website sebagai salah satu contoh penerapan teknologi informasi yang merupakan suatu aplikasi yang dapat memberikan kemudahan bagi setiap penggunaanya untuk memperoleh informasi secara cepat. Saat ini website bukan hal yang asing lagi, karena sebagai media atau sarana untuk memperoleh informasi maupun memberikan informasi yang berguna bagi semua pihak yang berkepentingan (Dengen, N, Maulana, S, & Hairah, 2018).

D. Informasi

Informasi adalah data yang sudah diolah menjadi bentuk yang lebih berguna dan lebih berarti bagi yang menerimanya. Informasi mengandung arti yang dihubungkan dengan kenyataan, simbol-simbol, gambar-gambar, kata-kata, angka-angka, huruf-huruf atau simbol yang menunjukkan ide, objek, kondisi dan situasi (Widians, J.A, Rahman & Hairah, 2017).

3. METODE PENELITIAN

Tahapan-tahapan yang digunakan untuk mencapai tujuan penelitian meliputi beberapa langkah antara lain, (1) Objek penelitian, (2) Studi literatur, (3) Wawancara, (4) Pengumpulan data, (5) Analisis, (6) Rumusan masalah, (7) Tujuan, (8) Menghitung dan perhitungan data menggunakan algoritma K-Means, (9) Kesimpulan.

A. Metode K-Means

Metode K-Means adalah salah satu metode pengelompokkan bersifat partitional secara pembelajaran berciri unsupervised. Secara prinsip, metode K-Means bekerja dengan memasukkan K sebagai konstanta jumlah cluster yang diinginkan. Sedangkan, Means dalam hal ini berarti nilai satu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai cluster (Puspitasari, N & Haviluddin, 2016). Algoritma ini didasarkan pada penentuan jarak antara centroid dan data latih. Kemudian, jumlah cluster centroid berdasarkan jumlah yang diinginkan. Sementara itu, centroid inisialisasi yang dihasilkan secara acak dengan mempertimbangkan pelatihan data. Dengan kata lain, centroid harus berada dalam ruang data pelatihan (Haviluddin, Fanany, A, & Gafar, 2018). K-Means Clustering sangat cocok untuk data dengan ukuran yang besar karena memiliki kecepatan yang lebih tinggi. K-Means

Clustering merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok (Adhe, D, Rachman, C, & Goejantoro, 2020).

Pengukuran jarak dapat menggunakan rumus Euclidean Distance. Berikut pengukuran jarak menggunakan rumus di bawah ini.

Rumus Euclidean Distance :

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \dots\dots\dots (1)$$

Keterangan :

- d = jarak antara *x* dan *y*
- x_1 = data pada pusat klaster ke *i*
- y_1 = data pada setiap data ke *i*
- x_2 = data pada pusat klaster ke *i*
- y_2 = data pada setiap data ke *i*

Sedangkan untuk menentukan centroid baru dapat dilakukan dengan mengambil rata-rata dokumen yang masuk pada cluster awal. Iterasi dilakukan terus hingga posisi grup tidak berubah. Berikut penentuan centroid menggunakan rumus di bawah ini menurut (Praja, B, Kusuma, D, & Setianingsih, 2019).

$$\text{Pusat cluster baru} = \frac{x_1 + x_2 + x_3 + \dots}{\text{jumlah } x} \dots\dots\dots (2)$$

Keterangan:

- X_1 = nilai data *record* ke-1
- X_2 = nilai data *record* ke-2

B. Akurasi Perhitungan K-Means

Algoritma K-Means merupakan salah satu algoritma pengelompokan (clustering) berbasiskan metode non-hierarchy yang mempartisi data dan membentuk satu atau lebih kelompok yang memiliki kesamaan. Pengujian model dilakukan untuk mengetahui seberapa dekat relasi antara objek dalam sebuah cluster dan seberapa jauh sebuah cluster terpisah dengan cluster lain (Haviluddin, H, Pakpahan, 2021). Perhitungan SSE dapat dilihat pada persamaan 3, menurut (Gustientiedina, G Adiya, 2019).

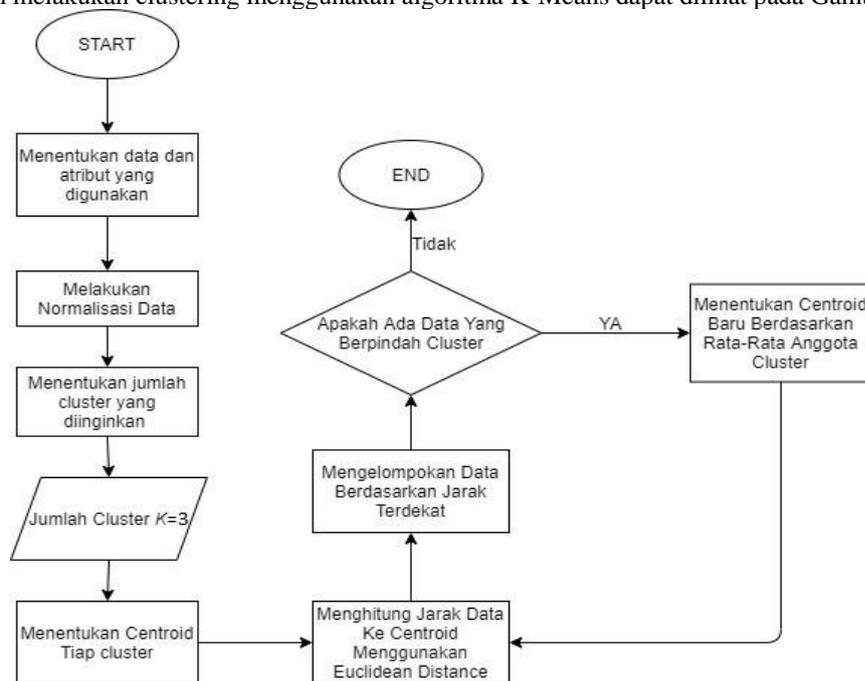
$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} (x_i - C_k)^2 \dots\dots\dots (3)$$

Keterangan :

- K = Jumlah klaster
- X_i = data ke-*i*
- C_k = nilai rata-rata *cluster* k

C. Flowchart K-Means

Alur dalam melakukan clustering menggunakan algoritma K-Means dapat dilihat pada Gambar 1 :



Gambar 1. Alur Clustering Algoritma K-Means

Alur pada Gambar 1 dapat dijelaskan sebagai berikut:

1. Menentukan data produksi komoditas perkebunan dan atribut-atribut yang akan digunakan
2. Melakukan normalisasi data ke rentang 0 sampai 1.
3. Penetapan jumlah *cluster*, pada penelitian ini berjumlah 3 *cluster*
4. Menentukan centroid masing-masing *cluster*
5. Menghitung jarak antara setiap titik data dan centroid menggunakan perhitungan jarak *Euclidean Distance*. Perhitungan jarak dapat menggunakan persamaan 2.1
6. Mengelompokkan data-data ke kluster berdasarkan jarak terdekat
7. Menentukan centroid baru yang dihitung dengan menggunakan persamaan 2.2
8. Jarak antara setiap titik data dengan centroid yang baru dihitung ulang.

Jika tidak ada data yang berpindah *cluster* maka proses selesai, jika ada maka ulangi langkah 4.

D. Normalisasi

Normalisasi pada Data Mining merupakan proses penskalaan nilai atribut dari data sehingga bisa jatuh pada range tertentu. Normalisasi ini sangat dibutuhkan agar tidak ada parameter yang mendominasi dalam perhitungan jarak antar data. Rumus persamaan yang digunakan untuk menghitung normalisasi z-score yaitu, menurut (Abdurrahman, D, Agus, F, & Putra, 2021):

$$Z_i = \left(\frac{X_i - \bar{X}}{S} \right) \dots \dots \dots (4)$$

Keterangan :

- \bar{X} = Nilai Rata-Rata
- X_i = Data Mentah
- Z_i = Nilai Standar
- S = Simpangan Baku

E. Unified Modelling Language (UML)

Unified Modeling Language (UML) adalah salah satu standar bahasa yang banyak digunakan di dunia industri untuk mendefinisikan *requirement*, membuat analisis dan desain serta menggambarkan arsitektur dalam pemrograman berorientasi objek (Rosa, A & Shalahuddin, 2015).

F. Use Case Diagram

Use case diagram merupakan pemodelan untuk kelakuan (behavior) sistem informasi yang akan dibuat. Use case mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem informasi yang akan dibuat. Secara kasar, use case digunakan untuk mengetahui fungsi apa saja yang ada di dalam sebuah sistem informasi dan siapa saja yang berhak menggunakan fungsi-fungsi itu (Rosa, A & Shalahuddin, 2015). Use case dipergunakan untuk menjelaskan interaksi pengguna dengan sistem tracer study yang disesuaikan dengan hak akses dan fungsinya dalam sistem (Ramadiani, Widagdo, P, & Astuti, 2017).

G. Class Diagram

Class diagram menggambarkan struktur statis dari kelas dalam sistem anda dan menggambarkan atribut, operasi dan hubungan antara kelas. Class diagram membantu dalam memvisualisasikan struktur kelas-kelas dari suatu sistem dan merupakan tipe diagram yang paling banyak dipakai (Haviluddin, 2011). Berikut ini penggambaran struktur *class* dalam suatu *class diagram* yang akan digunakan untuk aplikasi potensi produksi komoditas perkebunan di Kalimantan Timur.

4. HASIL DAN PEMBAHASAN

A. Normalisasi

Pada Setelah dilakukan semua proses *cleaning* data, kemudian masuk ketahap normalisasi data agar rentan antar setiap data tidak terlalu jauh, dan untuk metode yang digunakan untuk normalisasi data yaitu metode *Z-Score* menggunakan persamaan (4).

$$\begin{aligned} Z_{1,1} &= \frac{134.507 - 63010.00}{50377540.69} \\ &= 0.001419224 \end{aligned}$$

$$\begin{aligned} Z_{1,2} &= \frac{1.406.913 - 63010.00}{50377540.69} \\ &= 0.02667663 \end{aligned}$$

Setelah dilakukan perhitungan keseluruhan data, maka hasilnya dapat dilihat pada tabel yang tersaji dalam Tabel 1.

Tabel 1. Data hasil *normalisasi*

NO	Kabupaten	Komoditi	Tahun	Tanaman Menghasilkan	Produksi
1	Kutai Kartanegara	Kelapa Sawit	2015	0.001419224	0.02667663
2	Kutai Timur	Kelapa Sawit	2015	0.001970303	0.03238584
3	Kutai Barat	Kelapa Sawit	2015	-0.000694536	0.009137703
4	Penajam Paser Utara	Kelapa Sawit	2015	-0.000473842	0.007932225
5	Paser	Kelapa Sawit	2015	0.000543913	0.023848326
⋮	⋮	⋮	⋮	⋮	⋮
7	Balikpapan	Kemiri	2019	-0.001250101	-0.001250299

B. Penerapan Proses

a) Menentukan Pusat *Cluster*

Cluster yang ditentukan sebanyak 3 *cluster*, maka dari itu penulis memilih secara *random* dengan menghitung rata-rata, nilai minimal, dan maksimal dari variabel tanaman menghasilkan dan produksi, data yang akan dijadikan pusat *cluster* yaitu data ke-21, data ke-109, dan data ke-221 yang disajikan pada Tabel 2.

Tabel 2. Pusat *Cluster*

Data ke -i	Tanaman Menghasilkan	Produksi
21	-0.001049714	-0.00098528
109	-0.001203493	-0.001194798
221	-0.00124056	-0.001247878

b) Menghitung *Centroid*

Pada tahap ini dilakukan perhitungan *centroid*/rata-rata dari data yang ada di masing-masing *cluster* dengan menggunakan *centroid* awal yang telah ditentukan di Tabel 2, pada proses ini perhitungan jarak dilakukan menggunakan *Euclidean Distance* dimana rumus merujuk pada persamaan (1).

Centroid 1 untuk data ke-1

$$\begin{aligned}
 d &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \\
 &= \sqrt{(0.001419224 - (-0.001049714))^2 + (0.02667663 - (-0.00098528))^2} \\
 &= 0.02777187
 \end{aligned}$$

Centroid 2 untuk data ke-1

$$\begin{aligned}
 d &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \\
 &= \sqrt{(0.001419224 - (-0.001203493))^2 + (0.02667663 - (-0.001194798))^2} \\
 &= 0.027994556
 \end{aligned}$$

Centroid 3 untuk data ke-1

$$\begin{aligned}
 d &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \\
 &= \sqrt{(0.001419224 - (-0.00124056))^2 + (0.02667663 - (-0.001247878))^2} \\
 &= 0.028051604
 \end{aligned}$$

Dalam menentukan titik *centroid* terdekat dapat dilakukan dengan mengambil nilai paling rendah berarti jarak tersebut yang paling mendekati titik pusat. Pada iterasi pertama bentuk data dan hasil dapat dilihat pada Tabel 3.

Tabel 3. Tabel 3 Data hasil iterasi 1

NO	C1	C2	C3	Nilai MIN	Cluster
1	0.027771873	0.027994556	0.028051604	0.027771873	C1
2	0.033507495	0.033730287	0.033787347	0.033507495	C1
3	0.010129212	0.010345029	0.01040032	0.010129212	C1
4	0.00893608	0.009156143	0.009212692	0.00893608	C1
5	0.024884687	0.025104013	0.025160099	0.024884687	C1
6	0.022667203	0.022884527	0.022940047	0.022667203	C1
7	0.000333009	7.32464E-05	3.94618E-06	3.94618E-06	C3
8	0.000227099	9.21984E-05	0.000138201	9.21984E-05	C2
9	0.000331703	7.19856E-05	2.86117E-06	2.86117E-06	C3
10	0.000333009	7.32464E-05	3.94618E-06	3.94618E-06	C3
⋮	⋮	⋮	⋮	⋮	⋮
400	0.000333009	7.32464E-05	3.94618E-06	3.94618E-06	C3

c) Menentukan Centroid Baru

Lakukan perhitungan nilai *centroid* yang baru pada setiap *cluster*, setelah mendapatkan *cluster* untuk masing-masing data, selanjutnya kita menghitung kembali nilai *cluster*. Perhitungan dilakukan menggunakan persamaan (2) seperti di bawah ini :

Nilai *cluster* 1

$$C_{1,1} = \frac{0.001419224 + 0.001970303 + (-0.000694536) + \dots}{45}$$

$$= 0.000250232$$

$$C_{1,2} = \frac{0.02667663 + 0.03238584 + 0.009137703 + \dots + \dots}{45}$$

$$= 0.000250232$$

Nilai *cluster* 2

$$C_{2,1} = \frac{-0.001239541 + (-0.001202977) + (-0.009137703) + \dots}{48}$$

$$= -0.001182972$$

$$C_{2,2} = \frac{-0.001109939 + (-0.001226261) + (-0.001204426) + \dots}{48}$$

$$= -0.001180999$$

Nilai *cluster* 3

$$C_{3,1} = \frac{-0.001250756 + (-0.001250557) + (-0.001250756) + \dots}{307}$$

$$= -0.001246873$$

$$C_{3,2} = \frac{-0.001250756 + (-0.001249267) + (-0.001250756) + \dots}{307}$$

$$= -0.001248327$$

Setelah perhitungan di atas dilakukan, maka hasil dari nilai *cluster* 1 sampai dengan *cluster* 3 akan menjadi *centroid* baru. Berdasarkan hasil perhitungan di atas, maka didapat nilai *centroid* baru seperti pada Tabel 4.

Tabel 4. Titik Centroid Baru

Cluster	Tanaman Menghasilkan	Produksi
C1	0.000250232	0.019294147
C2	-0.001182972	-0.001180999
C3	-0.001246873	-0.001248327

Proses perhitungan nilai *centroid* dilakukan berulang-ulang hingga didapat anggota pada kelompok *cluster* tersebut tidak berubah. Pada perhitungan yang dilakukan secara manual, perhitungan dilakukan sampai iterasi ke-

9. Adapun hasil perhitungan titik *centroid* baru pada iterasi ke-9 yang tidak mengalami perubahan lagi ditampilkan pada Tabel 5.

Tabel 5. Hasil Titik Centroid Baru Pada Iterasi Terakhir

<i>Cluster</i>	Tanaman Menghasilkan	Produksi
C1	0.001764693	0.041472841
C2	-0.000195603	0.014908089
C3	-0.00122651	-0.00119991

Adapun hasil perhitungan jarak tiap data ke titik *centroid* baru masing-masing *cluster* pada iterasi ke-9 dengan menggunakan perhitungan jarak Euclidean ditampilkan pada Tabel 6.

Tabel 6. Data Hasil Iterasi Terakhir

NO	C1	C2	C3	Nilai MIN	Cluster
1	0.014800244	0.011878814	0.02800181	0.011878814	C2
2	0.009089327	0.017611444	0.033737549	0.009089327	C1
3	0.032428521	0.005791916	0.010351291	0.005791916	C2
4	0.033615234	0.00698141	0.0091631	0.00698141	C2
5	0.017666744	0.00897077	0.025110725	0.00897077	C2
6	0.019890818	0.00675136	0.022890925	0.00675136	C2
7	0.042829881	0.016193258	5.63311E-05	5.63311E-05	C3
⋮	⋮	⋮	⋮	⋮	⋮
400	0.042829881	0.016193258	5.63311E-05	5.63311E-05	C3

Setelah semua data tidak ada yang berpindah maka perhitungan akan berhenti di iterasi ke-9 dan jumlah data disetiap *cluster* ditampilkan pada Tabel 7.

Tabel 7. Jumlah Data per Cluster

Iterasi ke-i	C1	C2	C3
1	45	48	307
2	26	58	316
3	22	15	363
4	21	10	369
5	18	13	369
6	17	14	369
7	16	14	370
8	16	14	370
9	16	14	370

d) Hasil Pengujian

Pengujian metode dilakukan dengan mengukur keakuratan *cluster*. Metode yang digunakan adalah *Sum of Squared Error* (SSE). Perhitungan dilakukan menggunakan persamaan (3), yang pertama dilakukan adalah hasil dari menghitung rata-rata dari jarak data dengan *centroid* dapat dilihat pada tabel 8.

Tabel 8. Hasil Rata-Rata setiap cluster

<i>Cluster</i>	Hasil
C1	0.040765398
C2	0.016217086
C3	0.002362591

Selanjutnya nilai data ke-i dikurangkan dengan hasil rata-rata setiap cluster :

Cluster 1

$$= (0.014800244 - 0.040765398)^2$$

$$= 0.000674189$$

Cluster 2

$$= (0.011878814 - 0.016217086)^2$$

$$= 0.0000188206069079468$$

Cluster 3

$$= (0.02800181 - 0.002362591)^2$$

$$= 0.00065737$$

Setelah semuanya terhitung maka akan didapatkan hasil seperti tabel 9.

Tabel 9. Hasil Perhitungan Keseluruhan SSE

C1	C2	C3
0.000674189	0.0000188206069079468	0.00065737
0.001003373	0.0000019442328830224	0.001138222
0.0000695035095251478	0.000108684	0.000107149
0.000051124837534282	0.0000852977067465587	0.0000839623940835827
0.000533547785293986	0.0000525090951432449	0.00063054848623891
⋮	⋮	⋮
0.00000426209043451487	0.000000000567773910614598	0.000000003173192636267860

Selanjutnya masing-masing dari tiap cluster dijumlahkan keseluruhannya dan akan didapatkan hasil dari perhitungan *Sum of Squared Error* (SSE) seperti pada tabel 10.

Tabel 10. Hasil Perhitungan SSE

Cluster	Hasil
C1	0.023889017
C2	0.004166765
C3	0.034414884

Dilihat dari tabel 9 di atas, hasil SSE dari tiap cluster kemudian dijumlahkan hingga didapati hasil yaitu 0.062470666 yang dimana jika kita jadikan kebentuk persentasi maka didapati tingkat error dari metode yang digunakan adalah sebesar 6.25%

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dijelaskan di atas maka dapat disimpulkan Sistem pengelompokan komoditi perkebunan daerah Kalimantan Timur berbasis web dengan menerapkan metode K-Means telah berhasil dibangun. Penelitian ini menggunakan 3 cluster yaitu tingkat potensi produksi banyak (Cluster 1), tingkat potensi produksi sedang (Cluster 2), dan tingkat potensi produksi sedikit (Cluster3). Dan mendapatkan hasil perhitungan 16 data yang berada pada kelompok 1 (Cluster 1), 14 data yang berada pada kelompok 2 (Cluster 2), dan 370 data berada pada kelompok 3 (Cluster 3). Hasil perhitungan tingkat akurasi, bahwa metode pengukuran jarak *Euclidean Distance* menunjukkan tingkat akurasi sebesar 93,75% dan perhitungan *Sum of Squared Error* (SSE) menunjukkan tingkat error sebesar 6.25%. hal ini menunjukkan bahwa metode pengukuran jarak *Euclidean Distance* cukup akurat dalam pengelompokan komoditas perkebunan, karena nilai *Sum of Squared Error* (SSE) tidak melebihi 50%.

6. DAFTAR PUSTAKA

Abdurrahman, D., F. Agus, and G. M. Putra. (2021). Implementasi Algoritma Partitioning Around Medoids (PAM) untuk Mengelompokkan Hasil Produksi Komoditi Perkebunan (Studi Kasus : Dinas Perkebunan Provinsi Kalimantan Timur),” vol. 16, no. 2.

Adhe, D., C. Rachman, R. Goejantoro, and D. Tisna. (2020). Implementation Of Text Mining For Grouping Thesis Documents Using K-Means Clustering. *J. EKSPONENSIAL*, vol. 11, no. 2, pp. 167–174.

Anggara, M., Sujiani, H., Helfi, N. (2016). Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness. *J. Sist. dan Teknol. Inf.*, vol. 1, no. 1, pp. 1–6.

Dengen, N., Maulana, S.A., Hairah, U. (2018). Sistem Informasi Penyewaan Dan Penjualan Pada Darmo Music Studio. *J. Rekayasa Teknol. Inf.*, vol. 2, no. 1, p. 72, doi: 10.30872/jurti.v2i1.1417.

Dengen, N., R. Subagyo, and indah fitri Astuti. (2018). Perancangan Sistem Informasi Geografis Wilayah Hutan Lindung di Provinsi Kalimantan Timur,” vol. 3, no. 2, pp. 76–79.

Gustientiedina, G., M. H. Adiya, and Y. Desnelita. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, doi: 10.25077/teknosi.v5i1.2019.17-24.

Haviluddin, “Memahami Penggunaan UML (Unified Modelling Language). (2011). Memahami Pengguna. UML (Unified Model. Lang., vol. 6, no. 1, pp. 1–15, 2011, [Online]. Available: <https://informatikamulawarman.files.wordpress.com/2011/10/01-jurnal-informatika-mulawarman-feb-2011.pdf>.

- Haviluddin, A. Fanany, and O. Gafar. (2018). Comparison Between K-Means and Fuzzy C-Means Clustering in Network Traffic Activities,” *Comp. Between K-Means Fuzzy C-Means Clust. Netw. Traffic Act.*, vol. 2, doi: 10.1007/978-3-319-59280-0.
- Hermawan, T., M. Ugiarto, and N. Puspitasari. (2017). Sistem Evaluasi Kinerja Asisten Laboratorium Menggunakan Metode K-Means. *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, pp. 3–6.
- Maulana and A. A. Fajrin. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor. *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, p. 27, doi: 10.20527/klik.v5i1.100.
- Priatna, Y (2014). Kemelekan Informasi Masyarakat Terhadap Potensi Daerah. *Kemelekan Inf. Masy. Terhadap Potensi Drh.*, pp. 1–16.
- Praja, P. D. Kusuma, and C. Setianingsih. (2019). Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Penumpang Dan Kapal Angkutan Laut Di Indonesia,” *e-Proceeding Eng.*, vol. 06, no. 1, p. 1442.
- Puspitasari, N and Haviluddin. (2016). Penerapan Metode K-Means Dalam Pengelompokkan Curah Hujan. *Semin. Nas. Ris. Ilmu Komput. (SNRIK)*, vol. 1, no. March 2017, pp. 2–7.
- Ramadiani, P. P. Widagdo, and E. Junirianto. (2017). Tracer Study Menggunakan Framework Bootstrap,” vol. 1, no. 1, 2017, doi: 10.31227/osf.io/pxgb7.
- Rosa and M. Shalahuddin. (2015). Rekayasa Perangkat Lunak Menggunakan UML dan JAVA. *Bandung Inform. Bandung*.
- Wati, M., Indrawan, W., Widians, J.A., Puspitasari, N. (2017). Data mining for predicting students’ learning result,” in *2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT)*, pp. 1–4.
- Widians, J.A., W. H. Rachman, and Masnawati. (2017). Sistem Pendukung Keputusan Pemilihan Bibit Cabai Rawit Menggunakan Metode Simple Additive Weighting (SAW) Berbasis Web. *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 2, no. 1, pp. 175–181.
- Widians, J.A., Masnawati, and Awalludin, M. (2017). Apl. Daft. Hadir Perkuliahan Mhs. FKTI Univ. Mulawarman Berbas. Web, vol. 2, no. 1.
- Widians, J.A., Rahman, A.A., Hairah, U. (2017). Negeri Samarinda Berbasis Web,” vol. 2, no. 2, pp. 180–185.