

KLASTERISASI DATA HASIL PRODUKSI PERTANIAN DAN PETERNAKAN PROVINSI NUSA TENGGARA TIMUR MENGGUNAKAN METODE K-MEANS

Bruno Ortlund Ndasak¹, Patrisius Batarius², Yovinia Carmeneja Hoar Siki³

^{1,2,3}Ilmu Komputer, Fakultas Teknik, Universitas Katolik Widya Mandira, Kupang
¹brunoortlhand@gmail.com, ²patrisbatarius@unwira.ac.id, ³yoviniacarmeneja@gmail.com

Abstrak

Klasterisasi dan pemetaan produksi pertanian dan peternakan berdasarkan kabupaten/kota di provinsi Nusa Tenggara tidak pernah dilakukan. Setiap tahun hasil produksi pertanian dan peternakan mengalami peningkatan. Namun saat ini belum ada sistem untuk mengelompokkan kabupaten/kota di provinsi Nusa Tenggara Timur yang memiliki hasil produksi pertanian dan peternakan. Dengan dibuatnya sistem ini pihak dinas pertanian dan peternakan dapat mengetahui tingkatan hasil produksi berdasarkan hasil pertanian dan peternakan. Tujuan penelitian adalah membangun sebuah sistem pengelompokan dan pemetaan kabupaten/kota berdasarkan hasil produksi pertanian dan peternakan di Provinsi Nusa Tenggara Timur (NTT). Metode pengelompokan yang digunakan dalam penelitian ini adalah metode K-Means *Clustering*. Data yang digunakan data tahun 2019-2021 yang bersumber dari Badan Pusat Statistik (BPS) Provinsi NTT berupa data pertanian dan peternakan. Jumlah data yang digunakan sebanyak 396 data. Data produksi pertanian diantaranya padi, jagung, ubi kayu dan produksi peternakan yaitu ternak sapi, ternak babi, dan ternak ayam kampung. Dalam penelitian ini dikelompokkan dalam 3 cluster yaitu produksi sedang (C1), produksi sedang (C2) dan produksi tinggi (C3). Model pengembangan sistem yang digunakan dalam penelitian ini adalah model *waterfall* dan sistem yang dibuat menggunakan bahasa pemrograman PHP dengan database MySQL. Berdasarkan hasil pengujian sistem menunjukkan bahwa metode *K-Means* yang diterapkan pada sistem klasterisasi hasil produksi pertanian dan peternakan dapat mengelompokkan tingkatan hasil produksi pertanian dan peternakan di Provinsi Nusa Tenggara Timur.

Kata kunci : klasterisasi, produksi, pertanian, peternakan

1. Pendahuluan

Provinsi Nusa Tenggara Timur salah satu provinsi berkeulauan dan iklim kering. Sektor pertanian dan peternakan masih memainkan peran penting dalam struktur perekonomian Setiap wilayah atau kabupaten di Provinsi Nusa Tenggara Timur memiliki mata pencarian dengan ciri khas yang berbeda. Namun, pada dasarnya setiap wilayah selalu mengutamakan komoditas pertanian dan peternakan untuk produksi bahan makanan pokok dan daging.

Menurut informasi Badan Pusat Statistik kontribusi sektor pertanian memberikan kontribusi sebesar 29,17% terhadap pembentukan Produksi Domestik Regional Bruto (PDRB) Provinsi NTT pada Tahun 2021. Dari nilai tersebut tersebut 10,18% disumbang oleh subsektor peternakan dan 7,58% oleh tanaman pangan. Setiap tahun Provinsi NTT mengalami peningkatan produksi peternakan dan hasil pertanian. Untuk memenuhi makanan pokok dan daging, peran dinas pertanian dan peternakan perlu untuk mengelompokkan kabupaten / kota yang menghasilkan produksi pertanian dan peternakan di provinsi NTT agar dapat mengoptimalkan produksi hasil pertanian dan peternakan. Pengelompokan ini bertujuan untuk melihat dan mengetahui kabupaten/

kota yang berpotensi dalam memproduksi hasil pertanian dan peternakan sekaligus mengetahui kabupaten yang kurang maksimal. Untuk mengetahui kabupaten dan kota yang memiliki hasil produksi pertanian di Provinsi Nusa Tenggara Timur dapat menggunakan data mining.

Data mining digunakan dalam mengekstraksi informasi dari kumpulan data kemudian memperkenalkannya untuk dimengerti oleh manusia dalam mengambil keputusan (Ginting & Riandari, 2020). Teknik data mining yang dapat digunakan untuk mengelompokkan data yang memiliki kemiripan yang sama yaitu *clustering*. Clustering merupakan prosedur analisis data dalam mengelompokkan objek-objek yang sifat dan bentuknya sedemikian rupa sehingga muncul kelompok-kelompok sejenis diantara anggota kelompok yang sama (Hedyati & Suartana, 2021). Dibandingkan teknik data mining lainnya clustering memiliki manfaat karena dapat mengkategorikan data tanpa informasi sebelumnya. Berdasarkan kesamaan dari berbagai tipe data, clustering memisahkan dan mengatur data ke dalam cluster. Ada beberapa algoritma yang termasuk dalam teknik *clustering*, salah satu yang populer adalah K-Means. K-Means adalah salah satu teknik dari data mining dengan metode clustering yang mencoba

mengelompokkan data kedalam kelompok (cluster) (Muliono & Sembiring, 2019). Teknik K-Means clustering dapat mengatur data dengan memaksimalkan kesamaan atribut di dalam cluster dan memaksimalkan perbedaan antara cluster (Manalu et al., 2022). Untuk itu, data yang mirip disatukan dalam cluster yang sama, sedangkan data dengan fitur berbeda ditempatkan pada kelompok yang berbeda.

Penelitian ini merujuk pada beberapa penelitian terdahulu yang dilakukan oleh (Wijayanto & Fathoni, M, 2021) digunakan metode Clustering K-Means dalam mengelompokkan produksi padi dimana hasil penelitian ini mengelompokkan 12 daerah masuk dalam C1, 10 daerah masuk dalam C1 dan 5 daerah masuk dalam C2. Penelitian lainnya dilakukan oleh (Novany & Safii, 2021) digunakan metode k-means dapat mengelompokkan hasil produksi daging sapi dimana hasil penelitiannya menunjukkan 5 provinsi *cluster* tingkat tertinggi sedangkan 5 provinsi tingkat terendah. Penelitian lainnya dilakukan oleh (Pratama et al., 2022) digunakan metode k-means untuk mengelompokkan data pertanian. Hasil penelitian ini dapat mengelompokkan data pertanian dari 12 kabupaten. Kajian lainnya yang dilakukan oleh (Prasetya et al., 2021) penelitian ini membangun sistem informasi geografis menggunakan metode k-means clustering dalam mengelompokkan hasil produksi padi, dimana hasil penelitian ini terdapat hasil produksi kategori banyak sebesar 45%, hasil produksi sedang sebesar 40% dan produksi kurang memuaskan sebesar 15%.

Berdasarkan penelitian sebelumnya, belum ada penelitian dalam mengelompokkan hasil produksi pertanian dan peternakan. Penelitian ini bertujuan untuk membuat suatu peta dari data hasil *cluster* dalam mengelompokkan data hasil produksi pertanian dan peternakan di Provinsi Nusa Tenggara Timur. Pengelompokan berdasarkan kabupaten dan variabel – variabel hasil pertanian dan peternakan. Dengan dibuatnya sistem klusterisasi diharapkan hasil akhir dari penelitian ini dapat memberikan informasi pada masyarakat NTT untuk mengetahui Kabupaten/Kota yang memiliki hasil produksi pertanian dan peternakan di Provinsi NTT.

2. Metode

2.1 Data Mining

Data mining merupakan teknik memeriksa data secara mandiri untuk menemukan model dari database yang cukup besar (Dewi et al., 2022). Dalam literatur tentang data mining beberapa metode yang umum digunakan diantaranya: Prediksi, deskripsi, estimasi, klasifikasi, pengelompokan, dan asosiasi (Yuli Mardi, 2019).

2.2 Normalisasi Min-Max

Jika dalam dataset terdapat atribut yang memiliki skala atau range yang berbeda maka dilakukan normalisasi. Salah satu normalisasi data yang digunakan yaitu normalisasi *min-max*. Teknik normalisasi *min-max* digunakan untuk menerapkan penyesuaian linear pada data. Normalisasi bertujuan untuk menerjemahkan semua nilai variabel kedalam rentang yang sama yaitu 0 hingga 1 (Nishom, 2019). Normalisasi min-max dapat menggunakan persamaan berikut :

$$x = \frac{x - \text{nilai}_{\min}}{\text{nilai}_{\max} - \text{nilai}_{\min}} \tag{1}$$

Dimana :

x = data pada kolom
 nilai_{\min} = nilai terkecil pada data
 nilai_{\max} = nilai terbesar pada data

2.3 Clustering

Clustering adalah metode pengelompokan dalam data mining yang menggunakan metode pembelajaran tanpa pengawasan dimana tidak membutuhkan fase pembelajaran dan tidak memberikan pelabelan pada kelompok (Priyatman et al., 2019). Teknik *clustering* bertujuan untuk mengatur sejumlah data atau objek ke dalam satu cluster dengan data yang serupa.

2.4 K-Means

K-means ialah metode pengelompokan yang membagi kumpulan data menjadi sejumlah cluster yang telah diputuskan di awal. Metode k-means cukup mudah digunakan, cepat dieksekusi, serbaguna, dan digunakan secara luas. K-means mengelompokkan data yang bertentangan dengan titik pusat (centroid) menggunakan fungsi jarak. Jarak antara centroid dan data latih adalah dasar dari pendekatan ini (Data et al., 2022). Metode ini membagi data kedalam cluster-cluster sehingga data yang sebanding ditempatkan kedalam cluster yang sama dan data yang beragran dibagi kedalam cluster-cluster yang berbeda (Fernando Ade Pratama et al., 2022).

Menurut (Pakpahan et al., 2022), Algoritma K-Means dapat dihitung menggunakan langkah-langkah berikut :

1. Tentukan jumlah cluster
2. Tentukan centroid dari setiap cluster
3. Gunakan rumus *Euclidean Distance* untuk menentukan jarak antara setiap titik data dan pusat data sebagai berikut:

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

Keterangan :

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

- y_2 = data pada pusat klaster ke i
- Mengelompokan data kedalam cluster berdasarkan jarak terdekat dengan centroid
 - Menentukan centroid baru berdasarkan rata-rata objek di setiap cluster sebagai pusat cluster baru
 - Jarak antara setiap data dengan centroid yang baru dihitung ulang. Proses perhitungan selesai

2.5 Studi Literatur

Pada tahap ini dilakukan pencarian informasi kegunaan, rumus yang berkaitan dengan Metode K-Means *Clustering*, Provinsi NTT, hasil produksi pertanian dan peternakan.

2.6 Metode Pengambilan Data

Data yang digunakan dalam penelitian ini adalah Data Kuantitatif dari Badan Pusat Statistik Nusa Tenggara Timur tahun 2019-2021 melalui halaman website www.ntt.bps.go.id. Jenis data tersebut merupakan data sekunder yang di peroleh dari data yang telah di publikasi tahun 2019-2021 yang diterbitkan oleh BPS Nusa Tenggara Timur. Adapun data produksi yang diambil diantaranya produksi padi, produksi jagung, produksi ubi kayu, produksi sapi, produksi babi dan produksi ayam kampung tahun 2019-2021. Adapun indikator dari data hasil produksi pertanian dan peternakan yang akan digunakan dalam proses klasterisasi adalah sebagai berikut:

Tabel 1 Keterangan Data Hasil Produksi Pertanian

No	Indikator	Keterangan
1	Hasil Produksi	Jumlah hasil produksi pertanian (Ton)
2	Luas Lahan	luas lahan produksi (ha)

Tabel 2 Keterangan Data Hasil Produksi Peternakan

No	Indikator	Keterangan
1	Hasil Produksi	Jumlah hasil produksi daging (Kg)
2	Populasi Ternak	Jumlah populasi ternak (ekor)

2.7 Normalisasi Data

Pada tahap ini akan dilakukan normalisasi data. Dapat dilihat bahwa variabel yang digunakan memiliki skala atau rentang yang berbeda untuk itu penulis ingin melakukan normalisasi data terhadap variabel-variabel yang akan digunakan terlebih dahulu sebelum melakukan pengolahan data.

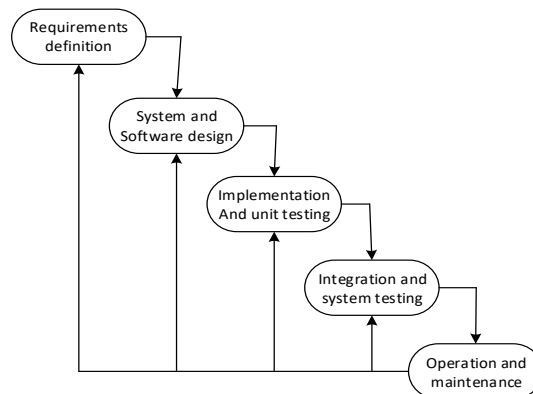
2.8 Metode Pengolahan Data

Dalam metode pengolahan data berisi tentang proses pengolahan data yang akan dilakukan menggunakan algoritma K-Means. Berikut adalah proses pengolahan data dalam penelitian ini :

- Normalisasi data
- Tentukan jumlah cluster (K)
- Tentukan pusat cluster (*centroid*) secara acak dari data yang tersedia
- Menghitung jarak setiap data ke pusat cluster (*centroid*) dengan persamaan *euclidean distance*
- Mengelompokan data berdasarkan jarak terdekat ke *centroid*
- Mengitung rata-rata objek di setiap cluster sebagai pusat cluster baru
- Membangkitkan titik *centroid* berdasarkan cluster baru yang terbentuk
- Melakukan iterasi proses 3-5 sampai kriteria cluster konvergen
- Melakukan pengujian cluster dengan sistem yang di bangun.

2.9 Pengembangan Sistem

Model pengembangan sistem yang digunakan dalam penelitian ini adalah Model *Waterfall*. Model *Waterfall* merupakan jenis model pengembangan aplikasi yang sering digunakan dalam pengembangan perangkat lunak yang bersifat berurutan dan sistematis (Sommerville, 2016). Berikut ini merupakan langkah-langkah dari model *waterfall*:



Gambar 1 Tahapan Model *Waterfall*

1. *Requirement Analysis and Defenition*

Pada tahapan ini peneliti melakukan analisis untuk mengetahui masalah-masalah yang berkaitan dengan penelitian yang mencakup analisis kebutuhan sistem, analisis peran sistem dan analisis peran pengguna sehingga mendapatkan spesifikasi sistem secara rinci.

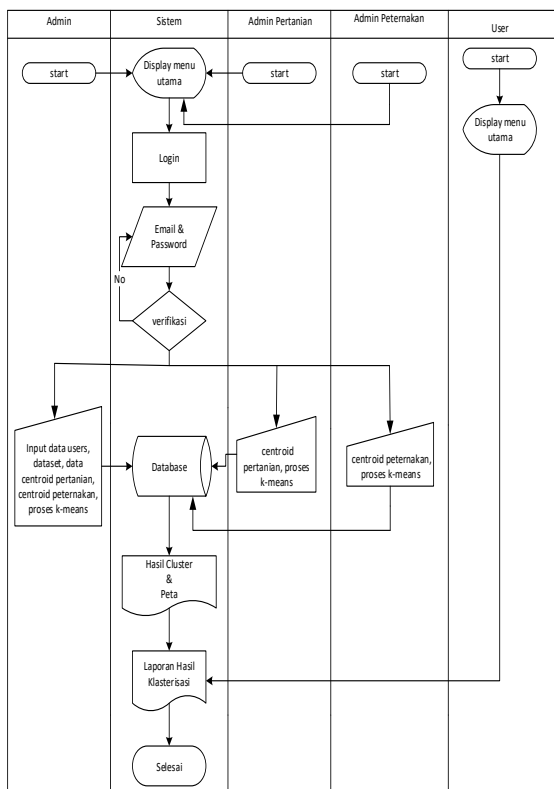
2. *System and Software Design*

Tahap ini bertujuan untuk memberikan gambaran tentang langkah-langkah yang harus dibuat

dan bagaimana tampilan dari sebuah sistem dikembangkan. Dalam tahap ini peneliti merancang kebutuhan-kebutuhan seperti bagan alir (*flowchart*), Diagram konteks, *Data Flow Diagram (DFD)*, *Entity Relations Diagram (ERD)* dan perancangan antarmuka

a. *Flowchart* Sistem

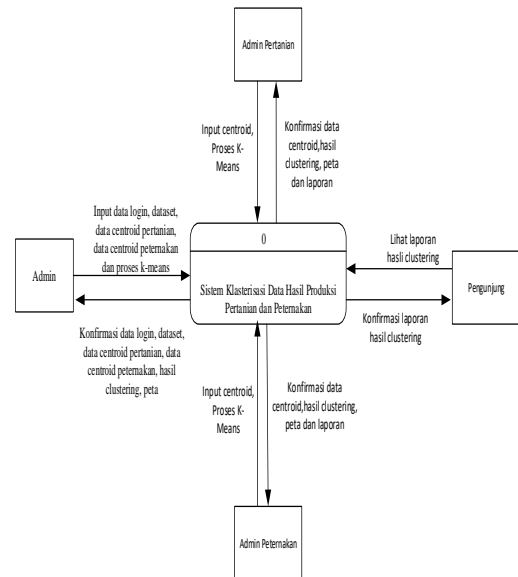
Flowchart adalah diagram yang menggambarkan algoritma atau intruksi berurutan sistem (Hasanah & Purnomo, 2022). Berikut merupakan gambar *flowchart* dari sistem klasterisasi.



Gambar 2 *Flowchart* Sistem

b. Diagram Konteks (*Context Diagram*)

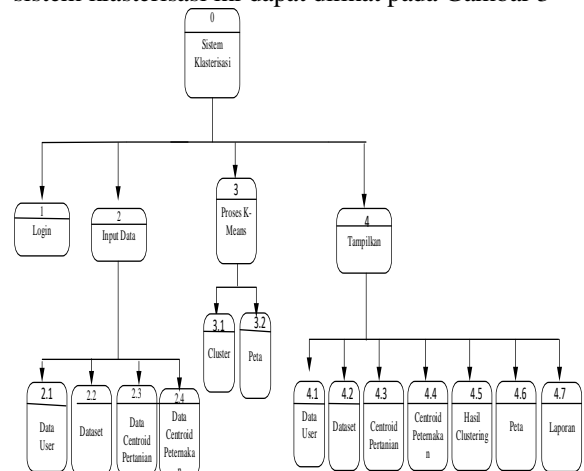
Diagram konteks menunjukkan bagaimana sistem berinteraksi dengan lingkungannya merupakan level tertinggi dari *Data Flow Diagram*. Diagram konteks menunjukkan secara umum hubungan dari proses input dan output (Irrawan et al., 2019).



Gambar 3 *Context Diagram*

c. Diagram Berjenjang (HIPO)

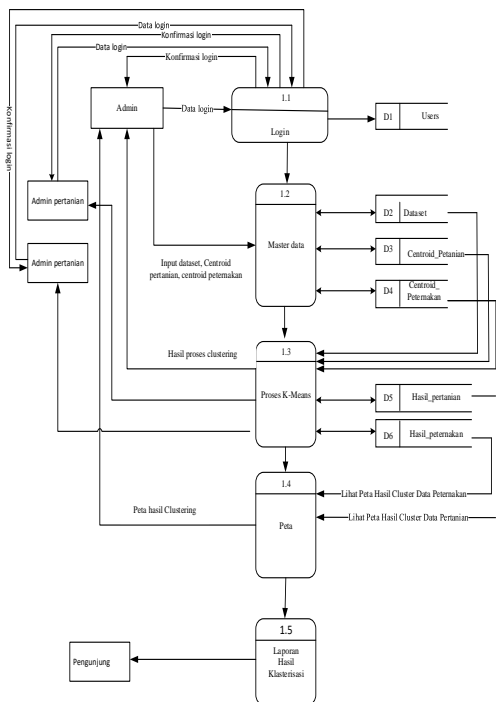
Diagram berjenjang merupakan diagram yang memberikan deskripsi terstruktur tentang program yang bekerja untuk memelihara program dan menjabarkan batasannya (Indarwati et al., 2019). HIPO bertujuan untuk membuat kerangka kerja dan strategi dokumentasi program. Diagram berjenjang sistem klasterisasi ini dapat dilihat pada Gambar 3



Gambar 4 Diagram Berjenjang (HIPO)

d. *Data Flow Diagram*

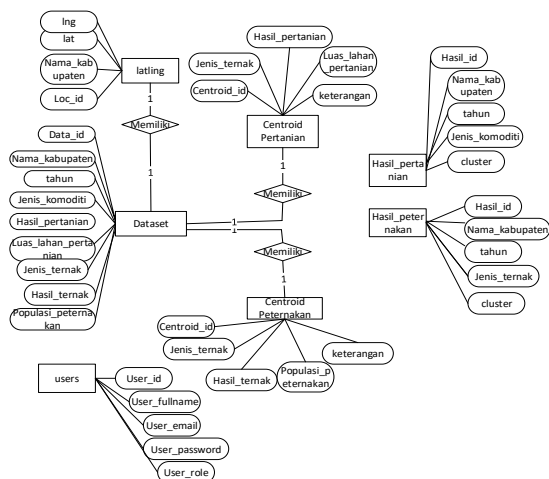
Data flow Diagram adalah alat pemodelan yang menggambarkan sistem sebagai jaringan proses fungsional yang dihubungkan oleh aliran data, baik secara manual ataupun elektronik (Suryati et al., 2020). *Data flow diagram* ini berisi tentang data secara umum dari sistem klasterisasi ini. Berikut ini tampilan *data flow diagram* pada sistem klasterisasi hasil produksi pertanian dan peternakan ini.



Gambar 3 Data Flow Diagram

e. Entity Relationship Diagram

Entity Relationship Diagram (ERD) adalah konsep umum basis data yang menggambarkan data yang akan disimpan dalam sebuah sistem maupun batasannya (Andry & Stefanus, 2020).



Gambar 4 Entity Relationship Diagram

3. Implementation and Unit Testing

Pada tahap ini peneliti mulai merancang perangkat lunak dengan mengimplementasikan kode program dengan menggunakan bahasa pemrograman PHP dan database MySQL sesuai dengan kebutuhan.

4. Integration and System Testing

Seluruh unit yang dikembangkan dalam tahap implementasi diintegrasikan kedalam sistem setelah pengujian yang dilakukan masing-masing unit.

Setelah integrasi seluruh sistem diuji untuk mengecek setiap kegagalan maupun kesalahan

5. Operation & Maintenance

Tahap akhir dalam model waterfall. Perangkat lunak yang sudah jadi, dijalankan serta dilakukan pemeliharaan. Pemeliharaan termasuk dalam memperbaiki kesalahan yang tidak ditemukan pada langkah sebelumnya. Perbaikan implementasi unit sistem dan peningkatan jasa sistem sebagai kebutuhan baru

3. Hasil dan Pembahasan

3.1 Perhitungan K-Means

Metode K-Means Clustering digunakan untuk mengelompokkan data hasil produksi pertanian dan peternakan di Provinsi Nusa Tenggara Timur dibagi menjadi 3 cluster. Ada 396 data yang digunakan dalam penelitian ini. Namun data yang akan dihitung menggunakan metode k-means yaitu data produksi padi tahun 2019 dapat dilihat pada Gambar 3.

Tabel 3 Data Produksi Padi 2019

Kabupaten/Kota	Komoditi	Hasil Produksi	Luas Lahan
Sumba Barat	Padi	32013	9812
Sumba Timur	Padi	68917	16678
Kupang	Padi	61593	18804
Timor Tengah Selatan	Padi	20032	4730
Timor Tengah Utara	Padi	40135	8939
Belu	Padi	17107	4773
Alor	Padi	2382	760
Lembata	Padi	2730	1041
Flores Timur	Padi	10997	3394
Sikka	Padi	13051	3806
Ende	Padi	19561	4604
Ngada	Padi	55563	12163
Manggarai	Padi	92029	19614
Rote Ndao	Padi	37776	9589
Manggarai Barat	Padi	121440	22305
Sumba Tengah	Padi	25883	6697
Sumba Barat Daya	Padi	35858	13609
Nagekeo	Padi	40087	10506
Manggarai Timur	Padi	88988	20118
Sabu Raijua	Padi	4063	1205
Malaka	Padi	19728	5300
Kota Kupang	Padi	1792	419

Dari Tabel 3 data produksi pertanian maka akan dilakukan normalisasi data. Metode normalisasi data

yang digunakan yaitu normalisasi min-max menggunakan rumus persamaan 1.

Tabel 4 Hasil Normalisasi Data

Kabupaten/Kota	Komoditi	Hasil Produksi	Luas Lahan
Sumba Barat	Padi	0,253	0,429
Sumba Timur	Padi	0,561	0,743
Kupang	Padi	0,500	0,840
Timor Tengah Selatan	Padi	0,152	0,197
Timor Tengah Utara	Padi	0,320	0,389
Belu	Padi	0,128	0,199
Alor	Padi	0,005	0,016
Lembata	Padi	0,008	0,028
Flores Timur	Padi	0,077	0,136
Sikka	Padi	0,094	0,155
Ende	Padi	0,149	0,191
Ngada	Padi	0,449	0,537
Manggarai	Padi	0,754	0,877
Rote Ndao	Padi	0,301	0,419
Manggarai Barat	Padi	1,000	1,000
Sumba Tengah	Padi	0,201	0,287
Sumba Barat Daya	Padi	0,285	0,603
Nagekeo	Padi	0,320	0,461
Manggarai Timur	Padi	0,729	0,900
Sabu Raijua	Padi	0,019	0,036
Malaka	Padi	0,150	0,223
Kota Kupang	Padi	0,000	0,000

Setelah normalisasi data dilakukan, maka tahap selanjutnya yaitu menentukan jumlah *cluster*. Jumlah *cluster* yang akan dibentuk dalam proses pengelompokan ini adalah 3 *cluster*. Setelah *cluster* ditentukan, proses selanjutnya yaitu penentuan *centroid*. Penentuan *centroid* dilakukan secara acak setiap data produksi pertanian dan peternakan tahun 2019-2021 berdasarkan variabel penelitian. Berdasarkan penentuan *centroid* dari setiap hasil produksi dapat dilihat pada Tabel 5.

Tabel 5 Nilai Awal Centroid

Variabel	Cluster					
	1		2		3	
Padi	765	220	10997	3394	51771	10061
Jagung	565	385	10078	3322	50719	152007
Ubi Kayu	39	3	10688	759	50062	2705
Sapi	21045	1692	101405,28	38381	1584067,5	132675
Babi	6549,64	125575	1008415,1	70343	2021662,5	119459
Ayam Kampung	94593,5	90894	205753,67	285972	508257,75	471482

Berdasarkan Tabel 5 diperoleh pusat titik *cluster (centroid)*, selanjutnya *centroid* yang sudah terbentuk akan dilakukan proses normalisasi data.

Tabel 6 Nilai Awal Centroid Setelah di Normalisasi

Variabel	Cluster					
	1		2		3	
Padi	0	0	0,201	0,323	1	1
Jagung	0	0	0,190	0,198	1	1
Ubi Kayu	0	0	0,213	0,280	1	1
Sapi	0	0	0,051	0,280	1	1
Babi	0	0	0,497	0,137	1	1
Ayam Kampung	0	0	0,269	0,513	1	1

Berdasarkan penentuan banyaknya jumlah *cluster* yang akan dibentuk dan penentuan *centroid* pada Tabel 5, maka selanjutnya adalah hitung jarak data ke masing-masing *centroid* dari masing-masing *cluster*. Untuk proses perhitungan dapat menggunakan persamaan 2. Selanjutnya akan dilakukan proses perhitungan k-means data produksi padi tahun 2019. Dari 22 data produksi padi tahun 2019 telah dipilih pusat *centroid* untuk produksi padi pada Tabel 5. Selanjutnya melakukan perhitungan jarak dari data dengan pusat *centroid*.

Jarak setiap objek ke pusat cluster pertama (C1)

$$d = \sqrt{(0,253 - 0)^2 + (0,429)^2} = 0,498$$

$$d = \sqrt{(0,561 - 0)^2 + (0,743 - 0)^2} = 0,931$$

Hasil perhitungan jarak setiap objek ke pusat *cluster* pertama (C1) secara lengkap dapat dilihat pada tabel 6.

Jarak setiap objek ke titik pusat kedua (C2)

$$d = \sqrt{(0,253 - 0,201)^2 + (0,429 - 0,323)^2}$$

$$= 0,119$$

$$d = \sqrt{(0,561 - 0,323)^2 + (0,743 - 0,323)^2}$$

$$= 0,554$$

Hasil perhitungan jarak setiap objek ke pusat klaster pertama (C2) secara lengkap dapat dilihat pada Tabel 7

Jarak setiap objek ke titik pusat ketiga (C3)

$$d = \sqrt{(0,253 - 1)^2 + (0,429 - 1)^2} = 0,940$$

$$d = \sqrt{(0,561 - 1)^2 + (0,743 - 1)^2} = 0,509$$

Hasil perhitungan jarak setiap objek ke pusat klaster ketiga (C3) secara lengkap dapat dilihat pada Tabel 7

Tabel 7 Jarak Masing-masing Objek Ke Nilai Akhir Centroid

Kabupaten/Kota	C1	C2	C3	Jarak Terdekat	Cluster
Sumba Barat	0,498	0,119	0,940	0,119	C2
Sumba Timur	0,931	0,554	0,509	0,509	C3
Kupang	0,977	0,598	0,525	0,525	C3
Timor Tengah Selatan	0,249	0,134	1,168	0,134	C2
Timor Tengah Utara	0,504	0,137	0,914	0,137	C2
Belu	0,237	0,143	1,184	0,143	C2
Alor	0,016	0,364	1,400	0,016	C1
Lembata	0,029	0,352	1,389	0,029	C1
Flores Timur	0,156	0,224	1,264	0,156	C1
Sikka	0,181	0,199	1,239	0,181	C1
Ende	0,242	0,141	1,174	0,141	C2
Ngada	0,700	0,328	0,720	0,328	C2
Manggarai	1,157	0,784	0,275	0,275	C3
Rote Ndao	0,516	0,139	0,909	0,139	C2
Manggarai Barat	1,414	1,048	0,000	0,000	C3
Sumba Tengah	0,350	0,036	1,071	0,036	C2
Sumba Barat Daya	0,667	0,292	0,818	0,292	C2
Nagekeo	0,561	0,183	0,868	0,183	C2
Manggarai Timur	1,158	0,783	0,289	0,289	C3
Sabu Raijua	0,041	0,339	1,375	0,041	C1
Malaka	0,269	0,112	1,152	0,112	C2
Kota Kupang	0,000	0,380	1,414	0,000	C1

Berdasarkan hasil perhitungan iterasi 1, proses selanjutnya menghitung kembali *centroid* baru yang merupakan rata-rata pada tiap *cluster*. Sehingga didapatkan titik pusat atau *centroid* yang baru yaitu :

Tabel 8 Centroid Baru

Cluster	Padi	
C1	0,034	0,062
C2	0,246	0,358
C3	0,709	0,872

Berdasarkan hasil penentuan centroid baru langkah selanjutnya yaitu menghitung jarak pusat *cluster* baru untuk mengetahui beberapa kali proses *clustering* dari 22 objek data produksi padi tahun 2019 yang uji, maka dilakukan perhitung *Euclidean distance* semua data ke centroid baru (C1, C2, C3) seperti perhitungan pada iterasi 1, lalu bandingkan hasilnya. Jika hasil cluster yang diperoleh pada iterasi 2 identik dengan hasil cluster pada iterasi 1, maka proses perhitungan dihentikan, jika tidak proses perhitungan akan berlanjut ke iterasi ketiga.

Tabel 9 Hasil Iterasi 2

Kabupaten/Kota	C1	C2	C3	Jarak Terdekat	Cluster
Sumba Barat	0,428	0,072	0,636	0,072	C2
Sumba Timur	0,861	0,497	0,196	0,196	C3
Kupang	0,907	0,545	0,211	0,211	C3
Timor Tengah Selatan	0,180	0,186	0,875	0,180	C1
Timor Tengah Utara	0,435	0,081	0,620	0,081	C2
Belu	0,166	0,198	0,889	0,166	C1
Alor	0,054	0,419	1,109	0,054	C1
Lembata	0,042	0,406	1,097	0,042	C1
Flores Timur	0,086	0,279	0,970	0,086	C1
Sikka	0,111	0,254	0,945	0,111	C1
Ende	0,173	0,193	0,882	0,173	C1
Ngada	0,631	0,271	0,424	0,271	C2
Manggarai	1,088	0,726	0,046	0,046	C3
Rote Ndao	0,446	0,082	0,610	0,082	C2
Manggarai Barat	1,347	0,990	0,318	0,318	C3
Sumba Tengah	0,281	0,084	0,775	0,084	C2
Sumba Barat Daya	0,596	0,248	0,502	0,248	C2
Nagekeo	0,491	0,127	0,566	0,127	C2
Manggarai Timur	1,089	0,726	0,034	0,034	C3
Sabu Raijua	0,030	0,394	1,084	0,030	C1
Malaka	0,199	0,16	0,85	0,166	C2
Kota Kupang	0,070	0,43	1,12	0,070	C1

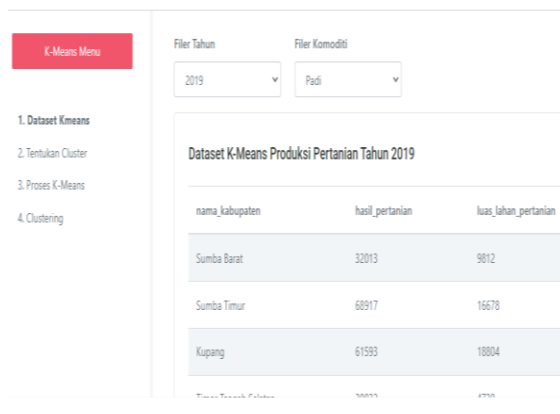
Berdasarkan posisi cluster pada perhitungan iterasi ke 2 pada Tabel 9 tidak sama dengan hasil proses sebelumnya, maka proses dilanjutkan ke proses iterasi ke 3. Langkah selanjutnya adalah menentukan centroid baru. Berdasarkan perhitungan *euclidean distance* dengan bantuan *Microsoft Excel* proses perhitungan berhenti pada iterasi 4. Tabel berikut merupakan hasil proses *clustering*.

Tabel 10 Hasil Cluster Produksi Padi 2019

Wilayah	Cluster
Sumba Barat	C2
Sumba Timur	C3
Kupang	C3
Timor Tengah Selatan	C1
Timor Tengah Utara	C2
Belu	C1
Alor	C1
Lembata	C1
Flores Timur	C1
Sikka	C1
Ende	C1
Ngada	C2
Manggarai	C3
Rote Ndao	C2
Manggarai Barat	C3
Sumba Tengah	C2
Sumba Barat Daya	C2
Nagekeo	C2
Manggarai Timur	C3
Sabu Raijua	C1
Malaka	C1
Kota Kupang	C1

3.2 Implementasi Sistem

Implementasi sistem pada penelitian klasterisasi data hasil produksi pertanian dan peternakan di Provinsi Nusa Tenggara Timur ini berbasis web. Sistem ini dibangun menggunakan bahasa pemrograman PHP dan MySQL. Sistem ini memiliki antar muka berupa tampilan yang terdiri dari halaman login, dataset, user, centroid pertanian, centroid peternakan, proses k-means pertanian, proses k-means peternakan, dashboard pertanian dan dashboard peternakan dan laporan hasil clustering. Berikut ini tampilan output menu proses k-means. Untuk proses clustering admin dapat melakukan proses k-means melalui menu proses k-means. Admin dapat melakukan proses clustering seperti yang terlihat pada Gambar 7.



Gambar 5 Menu Proses K-Means

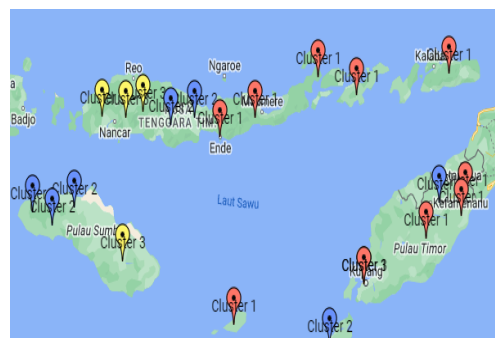
Berikut ini adalah hasil analisis K-Means menggunakan data yang di pilih oleh admin seperti yang ditunjukkan pada Gambar 8.

Hasil Cluster KMeans

Object	Cluster
Sumba Barat	2
Sumba Timur	3
Kupang	3
Timor Tengah Selatan	1
Timor Tengah Utara	2

Gambar 6 Tampilan Hasil Clustering

Berdasarkan Gambar 8 diatas maka proses selanjutnya yaitu menampilkan peta berdasarkan hasil perhitungan cluster. Halaman peta ini digunakan untuk menampilkan lokasi kabupaten dan kota yang memiliki hasil produksi pertanian dan peternakan serta tingkatan hasil produksi pertanian dan peternakan berdasarkan kabupaten dan kota di Provinsi Nusa Tenggara Timur.



Gambar 7 Tampilan Halaman Peta

Gambar 9 merupakan tampilan halaman peta hasil cluster. Pada halaman peta dapat menampilkan

peta wilayah Provinsi Nusa Tenggara Timur dan simbol titik lokasi berdasarkan kabupaten. Simbol titik warna merah menggambarkan kabupaten dengan tingkat hasil produksi rendah (C1), sedangkan simbol titik warna biru menggambarkan kabupaten dengan hasil produksi sedang (C2) dan simbol titik kuning menggambarkan kabupaten dengan hasil produksi tinggi (C3)

3.3 Pengujian Sistem

1. Pengujian Validitas Sistem

Hasil perhitungan manual dan perhitungan *Microsoft Excel* dari data hasil produksi pertanian dan peternakan yang ada kemudian data diolah kedalam sistem menggunakan algoritma K-Means dan kemudian akan dibandingkan hasil perhitungan manual dengan perhitungan sistem yang telah dibangun. Dari hasil perhitungan manual, excel dan sistem dapat disimpulkan bahwa hasil perhitungan menghasilkan hasil yang sama.

2. Analisa Pengujian

Pada tahap ini akan dilakukan analisa pengujian sistem dengan seluruh data yang digunakan yaitu data produksi pertanian dan peternakan 2019-2021 dengan jumlah data yang digunakan berjumlah 396 data. Sistem ini mengelompokkan hasil produksi pertanian dan peternakan menjadi 3 cluster. Berdasarkan penentuan centroid pada Tabel 5 sistem dapat mengelompokkan hasil produksi pertanian dan peternakan. Untuk melihat hasil pengelompokan hasil produksi pertanian dan peternakan dapat dilihat pada Tabel 11 dan Tabel 12

Tabel 11 Hasil Cluster Data Produksi Pertanian

Kabupaten/Kota	Padi			Jagung			Ubi Kayu		
	2019	2020	2021	2019	2020	2021	2019	2020	2021
Sumba Barat	C2	C1	C2	C1	C1	C1	C1	C1	C1
Sumba Timur	C3	C2	C3	C2	C2	C1	C2	C2	C1
Kupang	C3	C2	C3	C2	C2	C2	C2	C2	C2
Timor Tengah Selatan	C1	C1	C1	C3	C3	C3	C3	C3	C3
Timor Tengah Utara	C2	C1	C2	C2	C2	C2	C2	C2	C2
Belu	C1	C1	C2	C2	C2	C1	C3	C2	C2
Alor	C1	C1	C1	C1	C2	C1	C2	C2	C2
Lembata	C1	C1	C1	C1	C1	C1	C2	C3	C1
Flores Timur	C1	C1	C1	C1	C2	C1	C2	C3	C3
Sikka	C1	C1	C1	C1	C2	C1	C3	C3	C3
Ende	C1	C1	C2	C1	C1	C1	C2	C2	C2
Ngada	C2	C2	C2	C2	C1	C1	C1	C1	C1
Manggarai	C3	C3	C3	C1	C1	C1	C1	C1	C1
Rote Ndao	C2	C1	C2	C1	C1	C1	C1	C1	C1
Manggarai Barat	C3	C3	C3	C1	C2	C1	C1	C1	C2
Sumba Tengah	C2	C1	C2	C1	C1	C1	C1	C1	C1
Sumba Barat Daya	C2	C2	C2	C2	C3	C2	C2	C2	C2
Nagekeo	C2	C2	C2	C1	C1	C1	C1	C1	C1
Manggarai Timur	C3	C3	C3	C1	C2	C1	C1	C1	C1
Sabu Raijua	C1	C1	C1	C1	C1	C1	C1	C1	C1
Malaka	C1	C1	C2	C2	C2	C2	C2	C2	C2
Kota Kupang	C1	C1	C1	C1	C1	C1	C1	C1	C1

Tabel 12 Hasil Cluster Data Produksi Peternakan

Kabupaten/Kota	Sapi			Babi			Ayam Kampung		
	2019	2020	2021	2019	2020	2021	2019	2020	2021
Sumba Barat	C1	C1	C1	C1	C1	C1	C1	C1	C1
Sumba Timur	C2	C1	C1	C1	C1	C1	C2	C1	C2
Kupang	C3	C2	C2	C3	C3	C3	C3	C2	C2
Timor Tengah Selatan	C3	C2	C2	C2	C1	C2	C2	C1	C2
Timor Tengah Utara	C3	C2	C2	C2	C1	C1	C1	C1	C1
Belu	C2	C1	C1	C2	C1	C1	C1	C1	C1
Alor	C1	C1	C1	C2	C2	C2	C1	C1	C1
Lembata	C1	C1	C1	C1	C1	C1	C1	C1	C1
Flores Timur	C1	C1	C1	C2	C2	C2	C1	C1	C1
Sikka	C1	C1	C1	C2	C1	C1	C1	C1	C1
Ende	C2	C1	C1	C2	C1	C1	C3	C3	C3
Ngada	C2	C1	C1	C2	C1	C2	C1	C1	C2
Ngada	C2	C1	C1	C2	C1	C2	C1	C1	C2
Manggarai	C2	C1	C1	C1	C1	C1	C1	C1	C1
Rote Ndao	C2	C1	C1	C2	C1	C1	C1	C1	C1
Manggarai Barat	C1	C1	C1	C1	C1	C2	C1	C1	C1
Sumba Tengah	C1	C1	C1	C1	C1	C1	C1	C1	C1
Sumba Barat Daya	C1	C1	C1	C1	C2	C2	C1	C1	C1
Nagekeo	C2	C1	C1	C2	C1	C1	C2	C1	C2
Manggarai Timur	C1	C1	C1	C1	C1	C1	C1	C1	C2
Sabu Raijua	C1	C1	C1	C1	C1	C1	C1	C1	C2
Malaka	C2	C1	C1	C1	C1	C1	C2	C1	C1
Kota Kupang	C3	C3	C3	C2	C1	C1	C1	C1	C1

Keterangan :

- C1 : Produksi rendah
- C2 : Produksi sedang
- C3 : produksi tinggi

4. Kesimpulan

Berdasarkan hasil pengelompokan dan pengujian sistem klasterisasi hasil produksi pertanian dan peternakan Provinsi Nusa Tenggara Timur menggunakan metode k-means dapat disimpulkan sebagai berikut :

Data yang digunakan dalam penelitian ini sebanyak 396 data diantaranya data produksi padi, jagung, ubi kayu, sapi, babi, dan ayam kampung. Hasil clustering dibagi menjadi 3 cluster, yaitu : produksi rendah (C1), produksi sedang (C2), dan produksi tinggi (C3). Dari hasil pengujian sistem yang dilakukan maka didapatkan hasil pengelompokan data hasil produksi pertanian dan peternakan tahun 2019-2021 diantaranya untuk produksi padi jumlah data yang masuk dalam kategori produksi rendah (C1) ada 30 data, 22 data masuk dalam kategori produksi sedang (C2), dan 13 data masuk dalam kategori produksi tinggi (C3). Untuk produksi jagung jumlah data yang masuk dalam kategori produksi rendah (C1) ada 43 data, 21 data masuk dalam kategori produksi sedang (C2) dan 4 data masuk dalam kategori produksi tinggi (C3). Untuk produksi ubi kayu jumlah data yang masuk dalam kategori produksi rendah (C1) ada 33 data, 28 data masuk dalam kategori produksi sedang (C2) dan 34 data masuk dalam kategori produksi Tinggi (C3). Untuk produksi sapi jumlah data yang masuk dalam kategori produksi rendah (C1) ada 46 data, 14 data masuk dalam kategori produksi sedang (C2) dan 6 data masuk dalam kategori produksi Tinggi (C3). Untuk produksi babi jumlah data yang masuk dalam

kategori produksi rendah (C1) ada 42 data, 26 data masuk dalam kategori produksi sedang (C2) dan 4 data masuk dalam kategori produksi Tinggi (C3). Untuk produksi ayam kampung jumlah data yang masuk dalam kategori produksi rendah (C1) ada 50 data, 12 data masuk dalam kategori produksi sedang (C2) dan 4 data masuk dalam kategori produksi Tinggi (C3).

Sistem yang dibuat menunjukkan bahwa metode k-means yang diterapkan pada sistem klusterisasi data hasil produksi pertanian dan peternakan Provinsi Nusa Tenggara Timur ini dapat mengelompokkan kabupaten/kota berdasarkan tingkat hasil produksi pertanian dan peternakan.

Berdasarkan hasil penelitian yang telah dilakukan, maka penulis dapat memberikan saran-saran untuk pengembangan sistem selanjutnya yaitu dalam menampilkan peta sebaiknya menampilkan data hasil produksi pertanian dan peternakan, sehingga masyarakat bisa mengetahui hasil produksi pertanian dan peternakan pada tiap kabupaten di Provinsi Nusa Tenggara Timur.

Penelitian selanjutnya dapat menambahkan data hasil produksi pertanian dan peternakan serta melakukan perbandingan perhitungan metode lain seperti K-NN, K-Modes, dan metode clustering lainnya.

Daftar Pustaka:

Andry, J., & Stefanus, M. (2020). Pengembangan Aplikasi E-learning Berbasis Web Menggunakan Model Waterfall Pada SMK Strada 2 Jakarta. *Jurnal Fasilkom*, 10(1), 1–10. <https://doi.org/10.37859/jf.v10i1.1878>

Data, K., Mahasiswa, P., Pada, B., Bina, U., & Clustering, M. A. K. (2022). *Jurnal Pendidikan dan Konseling*, 4, 2692–2701.

Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>

Fernando Ade Pratama, E., Khairil, K., & Jumadi, J. (2022). Implementasi Metode K-Means Clustering Pada Segmentasi Citra Digital. *Jurnal Media Infotama*, 18(2), 291–301.

Ginting, B., & Riandari, F. (2020). Implementasi Metode K-Means Clustering Dalam Pengelompokan Bibit Tanaman Kopi Arabika. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 3(2), 151–157. <https://doi.org/10.32672/jnkti.v3i2.2381>

Hasanah, N. N., & Purnomo, A. S. (2022). Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik LPP Yogyakarta).

Jurnal Teknologi Dan Sistem Informasi Bisnis, 4(2), 300–311. <https://doi.org/10.47233/jteksis.v4i2.499>

Hediyati, D., & Suartana, I. M. (2021). Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro. *Journal of Information Engineering and Educational Technology*, 5(2), 49–54. <https://doi.org/10.26740/jieet.v5n2.p49-54>

Indarwati, T., Irawati, T., & Rimawati, E. (2019). Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomsin)*, 6(2), 2–7. <https://doi.org/10.30646/tikomsin.v6i2.369>

Irrawan, S. N., Simanjuntak, R. A., & Yusuf, M. (2019). PERANCANGAN SISTEM INFORMASI RENTAL & INVENTARIS ALAT MULTIMEDIA BERBASIS WEB MENGGUNAKAN METODE CUSTOMER RELATIONSHIP MANAGEMENT. *Jurnal REKAVASI*, 7(1).

Manalu, D. A., Gunadi, G., & Informatika, T. (2022). *Infotech: journal of technology information*, 8(1), 45–54.

Muliono, R., & Sembiring, Z. (2019). Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klusterisasi Tingkat Tridarma Pengajaran Dosen. *CESS (Journal of Computer Engineering, System and Science)*, 4(2), 2502–2714.

Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>

Novany, A. A., & Safii, M. (2021). Penerapan Data Mining Dalam Mengelompokkan Produksi Daging Sapi di Pulau Sumatera Menggunakan Algoritma K-Means Clustering. *Seminar Nasional Informatika (SENATIKA)*, 237–243.

Pakpahan, H. S., Widiars, J. A., Daffa, H., Firminda, A., & Basani, Y. (2022). Implementasi Metode K-Means Untuk Pengelompokan Potensi Produksi Komoditas Perkebunan. 1(1), 52–60.

Prasetya, Y. Y., Faisol, A., & Vendyansah, N. (2021). SISTEM INFORMASI GEOGRAFIS HASIL PRODUKSI PADI DI KABUPATEN MALANG MENGGUNAKAN METODE K-MEANS CLUSTERING. 5(2), 806–814.

Pratama, Y., Pramudya, Y., Albert, E., S, M. I., Ferdinand, R., & Juniansyah, V. (2022). BULLETIN OF COMPUTER SCIENCE RESEARCH Klusterisasi Data Pertanian di Tingkat Provinsi Jambi Tahun 2021 Menggunakan Algoritma K-Means. 3(1), 57–63. <https://doi.org/10.47065/bulletincsr.v3i1.205>

- Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(1), 62. <https://doi.org/10.26418/jp.v5i1.29611>
- Sommerville, I. (2016). Software engineering (10th edition). In *Pearson Education Limited*.
- Suryati, N. K. A. R., Dharma, E. M., & Estiyanti, N. M. (2020). Model Sistem Informasi Persediaan Barang Berbasis Web Pada MP.Net. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 9(2), 51. <https://doi.org/10.35889/jutisi.v9i2.494>
- Wijayanto, S., & Fathoni, M. Y. (2021). Pengelompokan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means. *Jurnal JUPITER*, 13(2), 212–219.
- Yuli Mardi. (2019). Data Mining: Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2(2), 213–219.

