PENERAPAN ALGORITMA *K-MEANS CLUSTERING* DALAM PENGELOMPOKAN PRODUKSI PADI DI PROVINSI KALIMANTAN TIMUR

Abigail Patandianan, Akhmad Irsyad, Muhammad Rivani Ibrahim
Sistem Informasi, Universitas Mulawarman
Jalan Sambaliung Samarinda, Indonesia
abigailpatdn@gmail.com

ABSTRAK

Padi merupakan komoditas utama dalam sektor pertanian yang berperan penting dalam ketahanan pangan nasional. Namun, produksi padi di Provinsi Kalimantan Timur mengalami penurunan dari tahun ke tahun. Berdasarkan data Badan Pusat Statistik Kalimantan Timur, produksi padi menurun sebesar 5,20% pada tahun 2023 dibandingkan tahun sebelumnya, salah satunya akibat konversi lahan pertanian. Oleh karena itu, diperlukan metode analisis data yang lebih efektif untuk membantu pengambilan keputusan dalam upaya peningkatan produksi padi. Penelitian ini menggunakan data produksi padi dari Dinas Pertanian Provinsi Kalimantan Timur selama periode 2019–2023. Algoritma *K-Means* diterapkan untuk pengelompokan berdasarkan luas lahan, produktivitas, dan total produksi. Jumlah *cluster* optimal ditentukan dengan *Elbow Method*, sedangkan evaluasi *clustering* menggunakan *Silhouette Coefficient*. Hasil penelitian menunjukkan bahwa metode *Euclidean Distance* dan *Minkowski Distance* (p=3) menghasilkan 1 *cluster* tinggi, 3 sedang, dan 6 rendah, sedangkan *Manhattan Distance* menghasilkan 1 *cluster* tinggi, 2 sedang, dan 7 rendah. Evaluasi menunjukkan *Euclidean Distance* dan *Minkowski Distance* lebih baik dengan nilai 0.3692 dibandingkan *Manhattan Distance* (0.3607). Analisis ini membantu mengidentifikasi daerah dengan produksi rendah untuk perencanaan kebijakan peningkatan produksi padi.

Kata kunci : *Data Mining, Clustering, K-Means, Produksi Padi*

1. PENDAHULUAN

Indonesia merupakan negara agraris, di mana mayoritas penduduknya bergantung pada sektor pertanian sebagai mata pencaharian utama [1]. Letak geografis Indonesia yang berada di wilayah tropis memberikan manfaat berupa iklim yang mendukung pertumbuhan komoditas pertanian, salah satunya padi. Hal tersebut karena tanaman padi merupakan sumber bahan makanan pokok bagi sebagian besar masyarakat Indonesia [2].

Tanaman padi (Oryza sativa) sebagai pemegang komoditas pangan utama tentunya sangat diperlukan peningkatan pada produktivitasnya dalam upaya untuk memenuhi kebutuhan pangan yang terus meningkat [3]. Kondisi produksi padi di Kalimantan Timur terus mengalami penurunan dari tahun ke tahun. Berdasarkan data pada Badan Pusat Statistik Provinsi Kalimantan Timur, produksi padi mengalami penurunan sebesar 5,20%, di mana pada tahun 2022 total produksi sebesar 239.425,34 ton menjadi 226.972,07 ton pada tahun 2023 [4].

Salah satu faktor penyebab terjadinya penurunan produksi adalah peningkatan populasi pada suatu daerah namun luas lahan pertanian tidak meningkat sebaliknya luas lahan semakin berkurang karena adanya konversi lahan pertanian menjadi lahan terbangun [5]. Situasi tersebut mendorong peneliti untuk meningkatkan produktivitas tanaman padi dengan harapan hasil produksi selanjutnya lebih baik lagi di Provinsi Kalimantan Timur.

Data yang dimiliki oleh Dinas Pertanian Provinsi Kalimantan Timur pada basis data digunakan untuk membuat laporan produksi, perkiraan produksi di masa depan, dan lainnya dengan bentuk data yang ada hanya sebatas angka laporan. Ratusan hingga ribuan data yang tersimpan mungkin sulit untuk diolah dengan metode konvensional, sehingga diperlukan pendekatan baru untuk mengatasi masalah penggalian informasi penting dari kumpulan data tersebut, yang dikenal sebagai data mining, guna mendukung perencanaan strategis dan pengambilan keputusan yang lebih akurat dalam upaya peningkatan produktivitas padi di wilayah tersebut.

Data mining merupakan proses yang digunakan untuk mengekstraksi dan mengidentifikasi informasi yang berguna serta memperoleh berbagai informasi penting dari suatu kumpulan data [6]. Pada penelitian kali ini menggunakan salah satu teknik yang digunakan dalam data mining yaitu *clustering* dengan tujuan untuk mengetahui produksi padi di Kalimantan Timur.

Pada pengelompokan produksi padi dapat dilakukan dengan menggunakan metode *clustering*, di mana salah satu metode *clustering* yang sering digunakan adalah *K-Means*. *K-Means* merupakan metode pengelompokan data non-hierarki (partisi) yang bertujuan untuk membagi data menjadi dua atau lebih kelompok. Metode ini membagi data ke dalam kelompok-kelompok sedemikian rupa sehingga data dengan karakteristik serupa ditempatkan dalam kelompok yang sama, sementara data dengan karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [7]. Metode ini memiliki keunggulan sebagai algoritma *clustering* yang sederhana dan popular dengan kelebihan yang terletak pada kemampuannya untuk mengelompokkan data

dalam jumlah besar secara efisien dengan waktu komputasi yang relatif cepat [8]. Pemilihan algoritma *K-Means* dalam penelitian ini didasarkan pada sifatnya yang sederhana untuk diimplementasikan, memiliki kinerja yang relatif cepat, mudah beradaptasi, serta umum digunakan. Secara historis, *K-Means* juga dikenal sebagai salah satu algoritma clustering yang paling penting dalam bidang *data mining* [6].

Penelitian sebelumnya telah banyak membuktikan bahwa algoritma *K-Means* efektif digunakan dalam pengelompokan data. Salah satunya analisis terhadap komoditi sayuran menunjukkan adanya tiga klaster yang potensial dengan evaluasi *clustering* menggunakan *Silhouette Coefficient* yang menghasilkan struktur yang baik (*medium structure*) [9].

Berdasarkan permasalahan di atas, penelitian ini mengangkat judul "Penerapan Algoritma *K-Means Clustering* Dalam Pengelompokan Produksi Padi Di Provinsi Kalimantan Timur". Hal ini bertujuan untuk mengekstraksi informasi data produksi agar pihak instansi dapat mengetahui daerah dengan hasil produksi yang masih rendah. Hasil *clustering* dibedakan menjadi tiga kelompok, yaitu produksi tinggi, sedang, dan rendah**.**

2. TINJAUAN PUSTAKA

2.1. Tanaman Padi

Tanaman padi atau dalam bahasa latin Oriza sativa merupakan salah satu jenis tanaman pangan yang paling penting di dunia [3]. Di Indonesia, masyarakat sering mengidentikkan pangan dengan beras yang merupakan sumber makanan pokok utama mayoritas masyarakat [2].

2.2. *Data Mining*

Data mining adalah proses yang memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrak serta mengidentifikasi informasi yang berguna serta pengetahuan yang terkait dari berbagai kumpulan data besar [10]. Proses dalam data mining mencakup beberapa tahapan, mulai dari pemilihan data, *pre-prosesing* data, pemodelan, dan evaluasi hasil [11].

2.3. Algoritma *K-Means*

K-Means merupakan salah satu metode pengelompokan data non-hierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain [7].

Algoritma *K-Means* menentukan jumlah *cluster* yang diinginkan sesuai dengan karakteristik data yang akan dikelompokkan. Selanjutnya, algoritma secara acak memilih titik pusat awal (*centroid*) untuk setiap *cluster*. Setiap data kemudian dihitung jaraknya terhadap seluruh *centroid*, dan dikelompokkan ke dalam *cluster* dengan jarak terdekat. Setelah proses pengelompokan, *centroid* setiap *cluster* diperbarui berdasarkan rata-rata posisi data yang tergabung dalam *cluster* tersebut. Proses ini diulangi dengan menggunakan *centroid* yang baru, hingga tidak terjadi perubahan pada posisi *centroid*, yang menandakan bahwa proses iterasi telah mencapai konvergensi dan algoritma selesai dijalankan[6]. Rumus untuk menentukan *centroid* baru dapat dilihat pada persamaan 1 [12].

$$C(i) = \frac{x_1+x_2+x\ldots}{x} \quad (1)$$

Keterangan:
$x1 =$ nilai data *record* ke-1
$x2 =$ nilai data *record* ke-2
$x =$ jumlah data *record*

Adapun beberapa metode perhitungan jarak yang digunakan pada perhitungan *K-Means* antara lain.

a. *Euclidean Distance*

Metode ini merupakan perhitungan jarak yang paling umum digunakan. Untuk dua titik data, x dan y, dalam data berdimensi d, rumus perhitungan jaraknya menggunakan persamaan 2 [12].

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

Keterangan:
d = Jarak antara x dan y
x = data pusat *Cluster*
y = data pada atribut
x_1= data pada pusat klaster i
y_1= data pada pusat data ke i

b. *Manhattan Distance*

Metode ini juga disebut sebagai *city blok* distance yang merupakan jarak dari semua atribut dengan menggunakan persamaan 3 [13].

$$d(x,y) = \sum_{i=1}^{n}(x_i - y_i) \quad (3)$$

Keterangan:
d = Jarak antara x dan y
x = data pusat *Cluster*
y = data pada atribut
i = setiap data
n = jumlah data
$x_i =$ data pada pusat klaster i
$y_i =$ data pada pusat data ke i

c. *Minkowski Distance*

Metode ini dianggap sebagai generalisasi dari *Euclidean Distance* dan *Manhattan Distance* yang dalam pengukuran jarak objek nilai p adalah 1 atau 2. Rumus untuk mengshitung jarak dalam metode ini terdapat persamaan 4.

$$d(x,y) = (\sum_{i=1}^{n} |x_i - y_i|^p)^{1/p} \quad (4)$$

Keterangan:
d = Jarak antara x dan y
x = data pusat *cluster*
y = data pada atribut
i = setiap data
n = jumlah data
x_i = data pada pusat klaster i
y_i = data pada pusat data ke i
p = power

2.4. Normalisasi *Min-Max*

Metode normalisasi *Min-Max* adalah teknik normalisasi yang mengubah data secara linear berdasarkan nilai minimum dan maksimum dari data tersebut. Tujuannya adalah untuk menskalakan nilai-nilai agar berada dalam rentang 0 hingga 1, sehingga data menjadi lebih seimbang[14]. Normalisasi *Min-Max* dapat dilakukan dengan menggunakan Persamaan 6

$$x = \frac{x - nilai_{min}}{nilai_{max} - nilai_{min}} \quad (6)$$

Keterangan:
x = Data per kolom
$nilai_{min}$ = Nilai minimum dari data per kolom
$nilai_{max}$ = Nilai maximum dari data per kolom

2.5. *Silhouette coefficient*

Silhouette Coefficient merupakan metode evaluasi klaster yang menggabungkan pendekatan *cohessian* (kohesi) dengan separation (separasi). Kohesi diukur dengan menghitung jarak rata-rata antar objek dalam satu klaster, sementara separasi dihitung sebagai jarak rata-rata setiap objek dalam klaster terhadap klaster terdekat [9]. Langkah pertama dalam menghitung koefisien *Silhouette* dimulai dengan menentukan jarak rata-rata antara data ke-i dan seluruh data lain di klaster yang sama (misalnya, data ke-i berada di klaster A). Rumus ini biasanya dinotasikan sebagai $a(i)$ dan menggambarkan tingkat kohesi data ke-i dalam klaster A, yang digunakan untuk mengukur seberapa dekat data ke-i dengan anggota klasternya sendiri [13]. Rumus $a(i)$ terdapat pada Persamaan 7.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq 1} d(i,j) \quad (7)$$

Keterangan:
A = Banyaknya klister A
j = Objek lain dalam satu *cluster* A
a(i) = Perbedaan rata-rata objek (i) ke semua objek lain pada A.
d(i,j) = Jarak antara objek i dengan j

Selanjutnya, hitung nilai minimum dari rata-rata jarak data ke-i ke seluruh data di klaster yang berbeda menggunakan Persamaan 8.

$$d(i,C) = \frac{i}{|C|} \sum_{j \in C} d(i,j) \quad (8)$$

Keterangan:
$d(i,C) =$ Jarak rata-rata antar objek i dengan semua objek pada *cluster* lain.
$C =$ *Cluster* lain selain *cluster* A atau *cluster* C tidak sama dengan *cluster* A.

Setelah menghitung nilai $d(i,C)$, selanjutnya memilih nilai jarak paling minimum sebagai $b(i)$ menggunakan Persamaan 9.

$$b(i) = \min_{C \neq A} d(i,j) \quad (9)$$

Langkah terakhir yaitu menghitung nilai *Silhouette Coefficient* menggunakan Persamaan 10.

$$S(i) = \frac{b(i) - a(i)}{\max(a(i) - b(i))} \quad (10)$$

Nilai $S(i)$ berada antara -1 dan 1, di mana setiap nilai diinterpretasi pada tabel 1.

Tabel 1. Interpretasi nilai *Silhouette coefficient*

Silhouette Coefficient	**Interpretasi**
0.71 – 1.00	Struktur yang dihasilkan kuat
0.51 – 0.70	Struktur yang dihasilkan baik
0.26 – 0.50	Struktur yang dihasilkan lemah
≤0.25	Tidak terstruktur

2.6. *Elbow Method*

Metode *Elbow* adalah teknik yang digunakan untuk menentukan jumlah klaster optimal dengan mengidentifikasi titik siku pada grafik data [15]. Umumnya, metode *Elbow* divisualisasikan dalam bentuk grafik untuk mempermudah pengamatan titik siku yang terbentuk. Tujuan utama dari metode ini adalah memilih nilai k yang relatif kecil namun tetap menghasilkan nilai withinss yang rendah. Grafik *Elbow* menggambarkan hubungan antara jumlah *cluster* dan penurunan *error*, sehingga membantu dalam menentukan jumlah *cluster* yang tepat pada algoritma *K-Means* [16].

2.7. *Google Colab*

Google Colab (Colaboratory) adalah *platform* berbasis *cloud* dari Google yang memungkinkan pengguna menulis dan menjalankan kode Python langsung melalui browser tanpa instalasi tambahan. *Platform* ini menyediakan akses gratis ke GPU/TPU dan mendukung kolaborasi *real-time* melalui fitur berbagi *notebook*. Karena kemudahan penggunaan dan aksesibilitasnya, *Google Colab* banyak dimanfaatkan oleh pengembang, peneliti, dan pelajar untuk mengerjakan proyek *Python* tanpa perlu konfigurasi perangkat keras sendiri [17].

2.8. *Python*

Python merupakan bahasa pemograman berorientasi objek yang bersifat *scripting*. Pemograman ini dapat digunakan untuk pengembangan perangkat lunak dan dapat dijalankan melalui berbagai sistem operasi. *Python* juga

merupakan bahasa pemograman yang popular dalam bidang data *science* dan analisis data dikarenakan dukungan *library* yang menyediakan fungsi analisis data dan fungsi *machine learning*, data *preprocessing tools*, serta visualisasi data [18].

2.9. *Scikit-Learn*

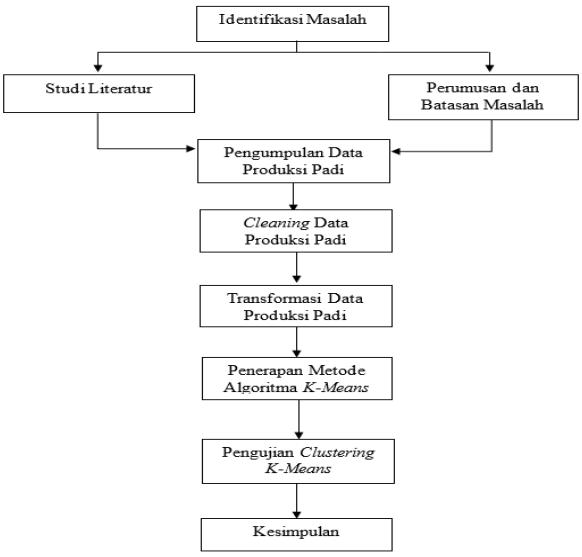
Scikit-Learn merupakan *library Python* yang dirancang untuk memudahkan penerapan teknik *machine learning* dalam pemrograman. *Library scikit-learn* memiliki performa yang optimal karena dibangun di atas modul *Numpy* (*Numerical Python*) dan *SciPy* (*Scientific Python*), sehingga membuat proses perhitungannya menjadi lebih efisien [19].

2.10. Produktivitas Padi

Produktivitas padi merupakan rata-rata hasil produksi padi, baik dari padi sawah maupun padi ladang, yang dihitung per satuan luas lahan. Produktivitas diukur berdasarkan total produksi padi dalam bentuk Gabah Kering Giling (GKG) per satuan luas lahan, dengan satuan kuintal per hektar (Ku/Ha) [20]. Adapun rumus untuk menghitung produktivitas dapat dilihat pada Persamaan 11.

$$Produktivitas\ (Ku/Ha) = \frac{Produksi\ (Ton)}{Luas\ Panen(Ha)} \times 10 \quad (11)$$

3. METODE PENELITIAN

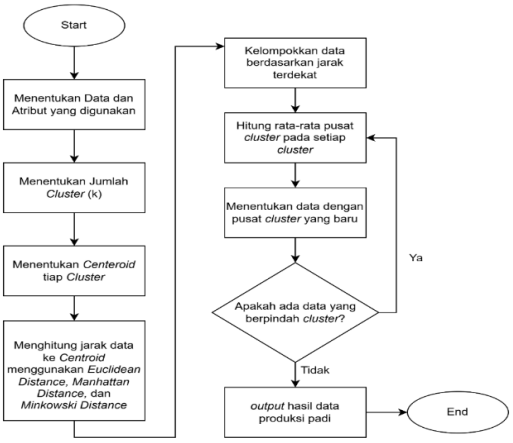


Gambar 1. Metode penelitian

Secara umum, alur dari motede penelitian dapat dilihat pada Gambar 1. di bawah ini. Metode penelitian ini diawali dengan identifikasi masalah berdasarkan data produksi padi dari BPS Kalimantan Timur yang menunjukkan adanya penurunan produksi di beberapa kota/kabupaten, diduga akibat berkurangnya luas lahan. Untuk menindaklanjuti hal tersebut, dilakukan studi literatur dari berbagai sumber seperti jurnal, artikel, buku, dan situs web. Data dikumpulkan melalui observasi, wawancara, dan dokumentasi, dengan sumber utama berasal dari BPS Kalimantan Timur dan Dinas Pertanian Provinsi Kalimantan Timur untuk data produksi tahun 2019 hingga 2023. Setelah data terkumpul, dilakukan pembersihan data untuk menghilangkan informasi yang tidak relevan, lalu dilanjutkan dengan transformasi data menggunakan normalisasi *Min-Max* agar seluruh data berada dalam skala [0,1].

Adapun penerapan algoritma *K-Means* pada penelitian ini dilakukan dengan menggunakan modul *clustering* yang terdapat pada library *Scikit-Learn Python*. Adapun alur dari algoritma *K-Means* dapat dilihat pada Gambar 2 dengan penjelasan sebagai berikut

a. Menentukan atribut dari produksi padi yaitu produksi dari tahun 2019 hingga tahun 2023 serta variabel-variabel yang akan digunakan.
b. Menentukan jumlah *cluster* (k) yang optimal menggunakan *Elbow Method*, yaitu dengan memplot jumlah *cluster* terhadap nilai *Within-Cluster Sum of Squares* (WCSS) untuk melihat titik siku atau elbow pada grafik, yang menandakan jumlah *cluster* optimal. Berdasarkan metode ini, penelitian ini mengelompokkan data ke dalam 3 *cluster*, yaitu produksi padi tinggi, sedang, dan rendah.
c. Menentukan *centroid* atau titik pusat awal untuk setiap *cluster* secara acak dari dataset.
d. Menghitung jarak antara setiap titik data dengan pusat *cluster* menggunakan rumus jarak *Euclidean Distance*, *Manhattan Distance*, dan *Minkowski Distance*.
e. Menentukan *cluster* untuk setiap titik data berdasarkan jarak terdekat dari pusat *cluster*.
f. Menghitung kembali posisi *centroid* untuk masing-masing *cluster* berdasarkan data yang telah dikelompokkan.
g. Menghitung kembali jarak antara setiap titik data dengan pusat *cluster* yang baru setelah *centroid* diperbaharui.
h. Memeriksa apakah ada titik data yang berpindah ke *cluster* lain setelah pembaruan *centroid*. Jika terdapat perubahan, maka proses iterasi dimulai lagi dari langkah 3 hingga 5 hingga tidak ada lagi data yang berpindah *cluster*, menandakan bahwa proses *clustering* telah mencapai konvergensi.



Gambar 2. Alur penerapan algoritma *K-Means*

3.1 Pengujian *Clustering K-Means*

Tahapan ini merupakan pengujian atau evaluasi terhadap hasil perhitungan dan pembuatan model yang telah dibuat. Evaluasi tersebut menggunakan metode *Silhouette coefficient* untuk mengukur seberapa baik hasil dari hasil perhitungan algoritma *K-Means*.

4. HASIL DAN PEMBAHASAN

4.1 Pengolahan Data

Seluruh proses pengolahan data dilakukan menggunakan *Python* dengan memanfaatkan library yang terdapat di dalamnya.

Tabel 2. Data hasil *cleaning*

No	Kota / Kabupaten	Luas Panen (Ha)	Produktivitas (Ku/Ha)	Produksi (Ton)
1	Berau	29246.76	35.590876	104091.78
2	Kota Balikpapan	361.88	35.008290	1266.88
3	Kota Bontang	361.67	39.592446	1431.94
4	Kota Samarinda	8375.43	40.032882	33529.26
5	Kutai Barat	2235.36	30.365131	6787.70
6	Kutai Kartanegara	145474.83	38.268773	556714.32
7	Kutai Timur	20102.76	34.009405	68368.29
8	Mahakam Ulu	1596.35	29.473925	4705.07
9	Paser	57224.54	40.087922	229401.29
10	Penajam Paser Utara	66618.09	33.178935	221031.73

Tabel 2 merupakan hasil dari *cleaning* data awal serta selection data yang relevan dan sesuai dengan kebutuhan penelitian yang nantinya akan digunakan pada perhitungan selanjutnya.

4.2 Transformasi Data

Langkah selanjutnya adalah melakukan normalisasi data untuk menyelaraskan skala setiap variabel.

Tabel 3. Data hasil normalisasi

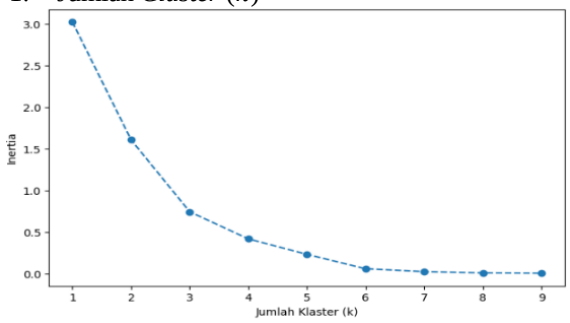
No	Kota/Kabupaten	Luas Panen (Ha)	Produktivitas (Ku/Ha)	Produksi (Ton)
1	Berau	0.199052	0.576310	0.185121
2	Kota Balikpapan	0.000001	0.521421	0.000000
3	Kota Bontang	0.000000	0.953319	0.000297
4	Kota Samarinda	0.055224	0.994814	0.058084
5	Kutai Barat	0.012912	0.083965	0.009939
6	Kutai Kartanegara	1.000000	0.828608	1.000000
7	Kutai Timur	0.136039	0.427311	0.120806
8	Mahakam Ulu	0.008508	0.000000	0.006190
9	Paser	0.391852	1.000000	0.410722
10	Penajam Paser Utara	0.456585	0.349068	0.395654

Tabel 3 merupakan hasil dari normalisasi data, di mana setiap data berada dalam skala 0 hingga 1.

4.3 Perhitungan *K-Means Clustering*

Langkah berikutnya adalah melakukan *clustering* terhadap produksi padi.

1. Jumlah *Cluster* (*k*)



Gambar 3. *Elbow method* untuk *cluster* optimal

Berdasarkan Gambar 3, nilai *inertia* menurun tajam hingga k = 3, kemudian mulai melandai. Titik siku (*Elbow* point) yang terlihat pada k=3 menunjukkan jumlah klaster optimal. Dengan demikian, jumlah klaster yang baik digunakan adalah 3 klaster.

2. *Centroid* Awal

Tabel 4 *Centroid* awal

Data	Luas Panen (Ha)	Produktivitas (Ku/Ha	Produksi (Ton)
9	0.391852	1.000000	0.410722
2	0.000001	0.521421	0.000000
6	1.000000	0.828608	1.000000

Tabel 4 merupakan *centroid* awal yang ditentukan secara acak, di mana data merupakan data ke-9, ke-2, dan ke-6 dalam dataset yang telah dinormalisasi.

3. Perhitungan Jarak

Pada proses ini, digunakan tiga metode perhitungan jarak, yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Minkowski Distance*.

a. *Euclidean Distance*

Tabel 5. Iterasi 4 *Euclidean Distance*

Data Ke	C1	C2	C3	Jarak MIN
1	0.410477	0.266100	1.170128	0.266100
2	0.509359	0.265946	1.447191	0.265946
3	0.217785	0.652455	1.419492	0.217785
4	0.136400	0.676079	1.344408	0.136400
5	0.920713	0.292933	1.584000	0.292933
6	1.208148	1.332168	0.000000	0.000000
7	0.556689	0.100974	1.296322	0.100974
8	1.004002	0.368101	1.630125	0.368101
9	0.352079	0.777341	0.863984	0.352079
10	0.743877	0.424024	0.943660	0.424024

Berdasarkan Tabel 5, hasil iterasi ke-4 (iterasi akhir) menunjukkan bahwa *cluster* C1 terdiri dari 3 data, *cluster* C2 memiliki jumlah terbanyak dengan 6 data, sedangkan *cluster* C3 hanya terdiri dari 1 data.

Tabel 6. *Centroid* akhir *Euclidean Distance*

Cluster	Luas Panen (Ha)	Produktivitas (Ku/Ha	Produksi (Ton)
C1	0.149025	0.982711	0.156368
C2	0.135516	0.326346	0.119618
C3	1.000000	0.828608	1.000000

Tabel 6 merupakan *centroid* akhir yang nantinya akan dilakukan pemetaan *cluster* ke dalam kategori Sedang, Rendah, dan Tinggi. Pemetaan ini didasarkan pada jumlah *cluster* optimal yang telah ditentukan.

Tabel 7. Pemetaan *cluster Euclidean Distance*

Cluster	Kategori	Nilai Rata-rata
C1	Sedang	0.4294
C2	Rendah	0.1938
C3	Tinggi	0.9429

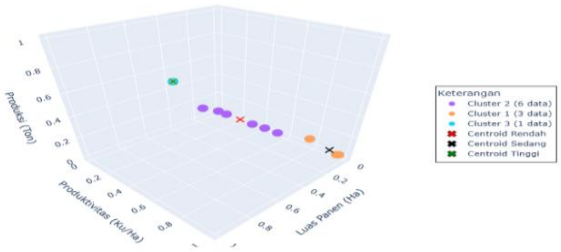
Tabel 7 dilakukan berdasarkan nilai rata-rata hasil normalisasi menggunakan metode *Min-Max*, di mana data berada dalam skala 0 hingga 1. Nilai yang mendekati 1 menunjukkan kategori Tinggi, sedangkan nilai yang mendekati 0 menunjukkan kategori Rendah.

Tabel 8. Hasil akhir *clustering Euclidean Distance*

Data Ke	Kota/Kabupaten	C1	C2	C3	*Cluster*
1	Berau		*		Rendah
2	Kota Balikpapan		*		Rendah
3	Kota Bontang	*			Sedang
4	Kota Samarinda	*			Sedang
5	Kutai Barat		*		Rendah
6	Kutai Kartanegara			*	Tinggi
7	Kutai Timur		*		Rendah
8	Mahakam Ulu		*		Rendah
9	Paser	*			Sedang
10	Penajam Paser Utara		*		Rendah

Berdasarkan Tabel 8, hasil akhir *clustering* menunjukkan bahwa setiap kota/kabupaten telah dikelompokkan ke dalam kategori Rendah, Sedang, atau Tinggi sesuai dengan jarak terdekatnya terhadap *centroid* akhir.

Gambar 4 menampilkan visualisasi hasil *clustering* yang memetakan kota/kabupaten berdasarkan variabel Luas Panen (Ha), Produktivitas (Ku/Ha), dan Produksi (Ton).;



Gambar 4. Visualisasi *clustering* (*Euclidean Distance*)

b. *Manhattan Distance*

Tabel 9. Iterasi 2 *Manhattan Distance*

Data Ke	C1	C2	C3	Jarak MIN
1	0.541457	0.339152	1.868126	0.339152
2	0.980517	0.126425	2.307186	0.126425
3	1.105893	0.558026	2.124413	0.558026
4	1.034378	0.493982	2.052898	0.493982
5	1.395124	0.514514	2.721792	0.514514
6	1.326668	2.207278	0.000000	0.000000
7	0.817784	0.224529	2.144452	0.224529
8	1.487242	0.606632	2.813910	0.606632
9	0.365366	1.181243	1.368818	0.365366
10	0.365366	0.898165	1.627302	0.365366

Berdasarkan Tabel 9, hasil iterasi ke-2 menunjukkan bahwa *Cluster* C1 terdiri dari 2 data, *Cluster* C2 memiliki jumlah terbanyak dengan 7 data, sedangkan *Cluster* C3 hanya terdiri dari 1 data

Tabel 10. *Centroid* akhir *Manhattan Distance*

Cluster	Luas Panen (Ha)	Produktivitas (Ku/Ha	Produksi (Ton)
C1	0.149025	0.982711	0.156368
C2	0.135516	0.326346	0.119618
C3	1.000000	0.828608	1.000000

Tabel 10 merupakan *centroid* akhir yang nantinya akan dilakukan pemetaan *cluster* ke dalam kategori Sedang, Rendah, dan Tinggi. Pemetaan ini didasarkan pada jumlah *cluster* optimal yang telah ditentukan.

Tabel 11. Pemetaan *cluster Manhattan Distance*

Cluster	Kategori	Nilai Rata-rata
C1	Sedang	0.5006
C2	Rendah	0.2071
C3	Tinggi	0.9429

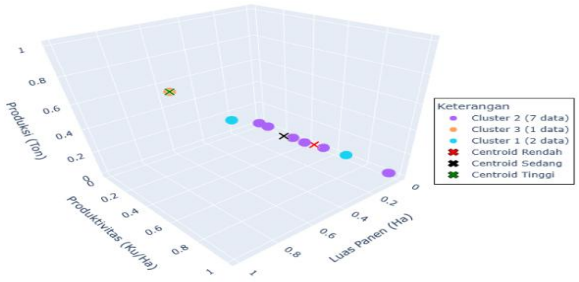
Tabel 11 menunjukkan bahwa *Cluster* C3 dikategorikan sebagai tinggi karena mendekati 1. *Cluster* C2 dikategorikan sebagai rendah karena mendekati 0. Sementara itu, *Cluster* C1 sedang karena berada di antara keduanya.

Tabel 12. Hasil akhir *clustering Manhattan Distance*

Data Ke	Kota/Kabupaten	C1	C2	C3	*Cluster*
1	Berau		*		Rendah
2	Kota Balikpapan		*		Rendah
3	Kota Bontang	*			Sedang
4	Kota Samarinda	*			Sedang
5	Kutai Barat		*		Rendah
6	Kutai Kartanegara			*	Tinggi
7	Kutai Timur		*		Rendah
8	Mahakam Ulu		*		Rendah
9	Paser	*			Sedang
10	Penajam Paser Utara		*		Rendah

Tabel 12 merupakan hasil akhir *clustering* menunjukkan bahwa setiap kota/kabupaten telah dikelompokkan ke dalam kategori Rendah, Sedang, atau Tinggi sesuai dengan jarak terdekatnya terhadap *centroid* akhir.

Gambar 5 menampilkan hasil visualisasi *clustering* menggunakan *Manhattan Distance*, yang memetakan kota/kabupaten berdasarkan setiap variabel.



Gambar 5. Visualisasi *clustering* (*Manhattan Distance*)

c. *Minkowski Distance*

Tabel 13. iterasi 3 *Minkowski Distance*

Data Ke	C1	C2	C3	Jarak MIN
1	0.406702	0.252799	1.023123	0.252799
2	0.472203	0.226524	1.265978	0.226524
3	0.192529	0.630504	1.260141	0.192529
4	0.121112	0.669028	1.189627	0.121112
5	0.901076	0.259138	1.328586	0.259138
6	1.068625	1.133182	0.000000	0.000000
7	0.555451	0.100965	1.115781	0.100965
8	0.984833	0.336976	1.361745	0.336976
9	0.313390	0.702861	0.757455	0.313390
10	0.667346	0.378307	0.789163	0.378307

Berdasarkan Tabel 13, hasil iterasi ke-3 menunjukkan bahwa *Cluster* C1 terdiri dari 3 data, *Cluster* C2 memiliki jumlah terbanyak dengan 6 data, sedangkan *Cluster* C3 hanya terdiri dari 1 data.

Tabel 14. *Centroid* akhir *Minkowski Distance*

Cluster	Luas Panen (Ha)	Produktivitas (Ku/Ha	Produksi (Ton)
C1	0.149025	0.982711	0.156368
C2	0.135516	0.326346	0.119618
C3	1.000000	0.828608	1.000000

Tabel 14 merupakan *centroid* akhir yang nantinya akan dilakukan pemetaan *cluster* ke dalam kategori Sedang, Rendah, dan Tinggi. Pemetaan ini didasarkan pada jumlah *cluster* optimal yang telah ditentukan.

Tabel 15. Pemetaan *cluster minkowski distance*

Cluster	Kategori	Nilai Rata-rata
C1	Sedang	0.5006
C2	Rendah	0.2071
C3	Tinggi	0.9429

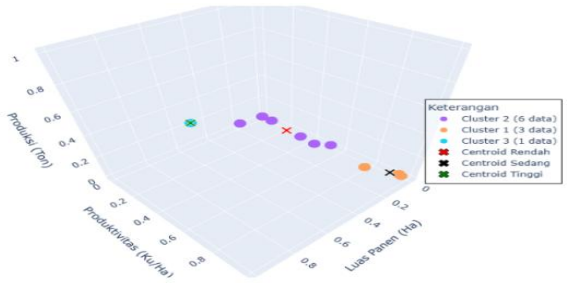
Tabel 15 menunjukkan bahwa *Cluster* C3 dikategorikan sebagai Tinggi karena mendekati 1. *Cluster* C2 dikategorikan sebagai Rendah karena mendekati 0. Sementara itu, *Cluster* C1 masuk dalam kategori Sedang, karena berada di antara keduanya.

Tabel 16. Hasil akhir clustering minkowski distance

Data Ke	Kota/Kabupaten	C1	C2	C3	*Cluster*
1	Berau		*		Rendah
2	Kota Balikpapan		*		Rendah
3	Kota Bontang	*			Sedang
4	Kota Samarinda	*			Sedang
5	Kutai Barat		*		Rendah
6	Kutai Kartanegara			*	Tinggi
7	Kutai Timur		*		Rendah
8	Mahakam Ulu		*		Rendah
9	Paser	*			Sedang
10	Penajam Paser Utara		*		Rendah

Tabel 16 merupakan hasil akhir *clustering* menunjukkan bahwa setiap kota/kabupaten telah dikelompokkan ke dalam kategori Rendah, Sedang, atau Tinggi sesuai dengan jarak terdekatnya terhadap *centroid* akhir.

Gambar 6 menampilkan hasil visualisasi *clustering* menggunakan metode *Minkowski* dengan parameter $p = 3$, yang memetakan kota/kabupaten berdasarkan variabel



Gambar 6. Visualisasi *clustering* (*Minkowski Distance*)

4.4 Evaluasi hasil *cluster*

Tabel 17. Hasil evaluasi *Silhouette Coefficient*

Metode	Nilai *Silhouette coefficient*
Euclidean Distance	0.3692
Manhattan Distance	0.3607
Minkowski Distance	0.3692

Tabel 17 menunjukkan bahwa metode *Euclidean Distance* dan *Minkowski Distance* (p=3) memiliki nilai *Silhouette Coefficient* sebesar 0.3692, sedangkan metode *Manhattan Distance* menghasilkan nilai 0.3607. Berdasarkan interpretasi *Silhouette Coefficient* pada Tabel 1, nilai tersebut berada dalam rentang 0.26 – 0.50, yang mengindikasikan bahwa struktur *cluster* yang terbentuk masih tergolong lemah.

Nilai *Silhouette Coefficient* yang relatif rendah ini dapat disebabkan oleh beberapa faktor. Salah satunya adalah distribusi data yang kurang jelas, di mana pola data yang menyebar atau saling tumpang tindih membuat pemisahan antar cluster menjadi kurang optimal. Meskipun telah dilakukan proses normalisasi data, hasil *clustering* tetap menunjukkan nilai yang rendah, yang yang menandakan bahwa algoritma masih kesulitan mengenali pola yang jelas dalam data. Selain itu, pemilihan variabel (Luas Panen, Produktivitas, dan Produksi) yang memiliki keterkaitan tinggi satu sama lain dapat memengaruhi kualitas *clustering*, karena perbedaan antar *cluster* menjadi kurang signifikan.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, hasil *clustering* menggunakan algoritma *K-Means* menunjukkan bahwa metode *Euclidean Distance* dan *Minkowski Distance* (p=3) menghasilkan nilai *Silhouette Coefficient* tertinggi sebesar 0,3692 dengan pembagian *cluster* terdiri dari 1 *cluster* tinggi, 3 *cluster* sedang, dan 6 *cluster* rendah, di mana Kabupaten/Kota yang termasuk dalam *cluster* produksi padi rendah antara lain Berau, Balikpapan, Kutai Barat, Kutai Timur, Mahakam Ulu, dan Penajam Paser Utara. Sementara itu, metode Manhattan Distance menghasilkan nilai Silhouette Coefficient sebesar 0,3607 dengan pembagian 1 *cluster* tinggi, 2 *cluster* sedang, dan 7 *cluster* rendah, dengan daerah rendah meliputi Berau, Balikpapan, Bontang, Samarinda, Kutai Barat, Kutai Timur, dan Mahakam Ulu. Nilai *Silhouette Coefficient* pada seluruh metode menunjukkan struktur *cluster* yang lemah namun masih dapat diterima, dan metode Euclidean Distance dinilai paling optimal untuk digunakan dalam analisis akhir karena memberikan hasil yang konsisten dan efisien. Hasil *clustering* ini dapat dimanfaatkan sebagai dasar dalam perencanaan strategi peningkatan produksi padi pada wilayah dengan kategori rendah. Diharapkan untuk penelitian selanjutnya untuk malakukan perbandingan algoritma *K-Means* dengan metode *clustering* lain, seperti DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) dan *Hierarchical Clustering*, untuk mengevaluasi apakah terdapat pendekatan yang lebih optimal dalam mengelompokkan data produksi padi.

DAFTAR PUSTAKA

[1] O. A. Pradipta *et al.*, “Meningkatkan Ketahanan Pangan Provinsi Kalimantan Timur Melalui Haki Atas Varietas Tanaman Padi,” *Risal. Huk.*, vol. 20, no. 2, pp. 112–134, 2023, [Online]. Available: https://www.bps.go.id/pressrelease/2018/02/05/1519/ekonomi-indonesia-

[2] A. Oktavia, Z. Zulfanetti, and Y. Yulmardi, “Analisis produktivitas tenaga kerja sektor pertanian di Sumatera,” *J. Paradig. Ekon.*, vol. 12, no. 2, pp. 49–56, 2020, doi: 10.22437/paradigma.v12i2.3940.

[3] M. A. R. Siregar, “Peningkatan Produktivitas Tanaman Padi Melalui Penerapan Teknologi Pertanian Terkini,” *J. Agribisnis*, vol. 1, no. 1, pp. 1–11, 2023, doi: 10.31219/osf.io/g98xr.

[4] BPS Kalimantan Timur, “Pada 2023, luas panen padi mencapai sekitar 57,08 ribu hektare dengan produksi padi sebesar 226,97 ribu ton gabah kering giling (GKG).,” 2024. https://kaltim.bps.go.id/id/pressrelease/2024/03/01/1118/pada-2023--luas-panen-padi-mencapai-sekitar-57-08-ribu-hektare-dengan-produksi-padi-sebesar-226-97-ribu-ton-gabah-kering-giling--gkg--.html (accessed Nov. 02, 2024).

[5] L. M. A. H. Ramadan, N. Alim, and M. Tahrir, “Analisis Ketahanan Pangan Beras Provinsi Kalimantan Timur Tahun 2023-2032,” *Nusant. Innov. J.*, vol. 1, no. 2, pp. 34–46, 2023, doi: 10.70260/nij.v1i2.20.

[6] S. Laia, D. H. Pane, and E. Affandi, “Implementasi Metode K-Means Untuk Mengelompokkan Kawasan Potensi Pertanian Karet Produktif,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 4, p. 282, 2022, doi: 10.53513/jursi.v1i4.5224.

[7] S. Dewi, S. Defit, and Y. Yuhandri, “Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means,” *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 28–33, 2021, doi: 10.37034/jsisfotek.v3i1.40.

[8] E. Edison, Y. Yupianti, and L. Elfianty, “Application of the K-Means Method in Determining the Clustering of the Economic Status of the Villagers of Gunung Megang,” *J. Komputer, Inf. dan Teknol.*, vol. 1, no. 2, pp. 50–58, 2021, doi: 10.53697/jkomitek.v1i2.230.

[9] L. M. Harahap, W. Fuadi, L. Rosnita, E. Darnila, and R. Meiyanti, “Klastering Sayuran Unggulan Menggunakan Algoritma K-Means,” *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 567–579, 2022, doi: 10.28932/jutisi.v8i3.5277.

[10] Haris Kurniawan, Sarjon Defit, and Sumijan, “Data Mining Menggunakan Metode K-Means

Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 80–89, 2020, doi: 10.52158/jacost.v1i2.102.

[11] M. M. Muttaqin, Wahyu Wijaya Widiyanto, A. W. Green Ferry Mandias, Stenly Richard Pungus, S. A. H. Wiranti Kusuma Hapsari, E. F. B. Aslam Fatkhudin, Pasnur, and N. S. Mochammad Anshori, Suryani, *Pengenalan Data Mining*. 2023. [Online]. Available: https://kitamenulis.id/2023/03/29/pengenalan-data-mining/

[12] H. S. Pakpahan, J. A. Widians, and H. D. A. Firmanda, "Implementasi Metode K-Means Untuk Pengelompokan Potensi Produksi Komoditas Perkebunan," *Adopsi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 52–60, 2022, doi: 10.30872/atasi.v1i1.49.

[13] R. Hidayati, A. Zubair, A. Hidayat Pratama, and L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering," *Techno.COM*, vol. 20, no. 2, pp. 186–197, 2021, doi: doi.org/10.33633/tc.v20i2.4556.

[14] T. Nugroho, U. Enri, and I. Maulana, "Clustering Menu Makanan Berdasarkan Kebutuhan Kalori Harian Menggunakan Algoritme K-Means," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 4411–4417, 2024, doi: 10.36040/jati.v8i4.9958.

[15] I. Pramudya, V. Sufi, H. Sukmawati, Y. Wellyhans, and I. Darwati, "Implementasi Algoritma K-Means pada Klasterisasi Tingkat Kasus Stunting di Kabupaten Batang," *J. Infortech*, vol. 6, no. 2, pp. 101–106, 2024, doi: https://doi.org/10.31294/infortech.v6i2.23435.g6666.

[16] M. Qusyairi, Z. Hidayatullah, and A. Sandi, "Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 2, pp. 500–510, 2024, doi: DOI : 10.29408/jit.v7i2.26375.

[17] R. Nazar, "Implementasi Pemrograman Python Menggunakan Google Colab," *J. Inform. dan Komput.* , vol. 15, no. 1, pp. 50–56, 2024.

[18] R. M. Koretsky, "Python3," in *Raspberry Pi OS System Administration with systemd and Python*, vol. 8, no. 1, Boca Raton: Chapman and Hall/CRC, 2023, pp. 175–305. doi: 10.1201/b23421-3.

[19] M. N. Fahmi, "Implementasi Mechine Learning menggunakan Python Library : Scikit-Learn (Supervised dan Unsupervised Learning)," *Sains Data J. Stud. Mat. dan Teknol.*, vol. 1, no. 2, pp. 87–96, 2023, doi: 10.52620/sainsdata.v1i2.31.

[20] BPS Indonesia, "Analisis Produktivitas Padi di Indonesia 2021," 2022. https://www.bps.go.id/id/publication/2022/12/16/a4fb42fcf25867b707461625/analisis-produktivitas-padi-di-indonesia-2021.html (accessed Mar. 03, 2025).