



## IMPLEMENTASI ALGORITMA *K-MEANS* UNTUK PENGELOMPOKAN KECAMATAN BERDASARKAN PRODUKTIVITAS TANAMAN PADI DI KABUPATEN CIREBON

Rosa Rosiana<sup>1</sup>, Willy Prihartono<sup>2</sup>, Fathurrohman<sup>3</sup>

<sup>1</sup> Teknik Informatika, STMIK IKMI Cirebon

<sup>2</sup> Komputerisasi Akuntansi, STMIK IKMI Cirebon

<sup>3</sup> Rekayasa Perangkat Lunak, STMIK IKMI Cirebon  
Cirebon, Jawa Barat, Indonesia 45135

rosarosiana101@gmail.com, willy@ikmi.ac.id, fathurrohman.ikmi@gmail.com

### Abstract

*The productivity of rice plants in Cirebon Regency varies in each sub-district, which causes an imbalance in rice production. This study aims to group sub-districts in Cirebon Regency based on rice crop productivity using the K-Means Clustering algorithm to support strategic decision-making in the agricultural sector. The research methods applied include Knowledge Discovery in Databases (KDD), which provides data selection, preprocessing, transformation, analysis using K-Means, and evaluation using the Davies-Bouldin Index (DBI). The data used is rice productivity in 2023 from 40 sub-districts, which includes planting area, harvest area, and production yield. The analysis showed that the DBI value was optimal at  $k=3$ , with three productivity categories: high, medium, and low. Compared to other methods, the K-Means algorithm has proven to be efficient and accurate in grouping data. This research contributes to local governments in formulating policies to increase rice productivity in areas that require further intervention. These findings also provide a basis for further study by comparing other algorithms to improve the accuracy of the results.*

**Keywords:** Algorithm K-Means, Clustering, Davies-Bouldin Index, Knowledge Discovery in Databases, Rice Productivity

### Abstrak

Produktivitas tanaman padi di Kabupaten Cirebon bervariasi di setiap kecamatan, yang menyebabkan ketidakseimbangan dalam hasil produksi padi. Penelitian ini bertujuan untuk mengelompokkan kecamatan di Kabupaten Cirebon berdasarkan produktivitas tanaman padi menggunakan algoritma *K-Means Clustering* untuk mendukung pengambilan keputusan strategis di bidang pertanian. Metode penelitian yang diterapkan mencakup *Knowledge Discovery in Databases* (KDD), yang meliputi seleksi data, *preprocessing*, transformasi, analisis menggunakan *K-Means*, serta evaluasi dengan *Davies-Bouldin Index* (DBI). Data yang digunakan adalah produktivitas padi tahun 2023 dari 40 kecamatan yang mencakup luas tanam, luas panen, dan hasil produksi. Hasil analisis menunjukkan nilai DBI optimal pada  $k=3$ , dengan tiga kategori produktivitas: tinggi, sedang, dan rendah. Algoritma *K-Means* terbukti efisien dan akurat dalam mengelompokkan data dibandingkan metode lain. Penelitian ini memberikan kontribusi bagi pemerintah daerah dalam merumuskan kebijakan peningkatan produktivitas padi di wilayah yang membutuhkan intervensi lebih lanjut. Temuan ini juga memberikan dasar bagi penelitian lebih lanjut dengan membandingkan algoritma lain untuk meningkatkan akurasi hasil.

**Kata kunci:** Algoritma *K-Means*, Clustering, Davies-Bouldin Index, Knowledge Discovery in Databases, Produktivitas Tanaman Padi

### 1. PENDAHULUAN

Produktivitas tanaman padi merupakan salah satu indikator utama keberhasilan sektor pertanian, khususnya di Kabupaten Cirebon yang dikenal sebagai salah satu lumbung padi di Provinsi Jawa Barat. Namun, produktivitas padi di setiap kecamatan tidak merata akibat perbedaan luas tanam, luas panen, dan hasil produksi yang signifikan.

Ketidakeimbangan ini menyulitkan pemerintah dalam merumuskan kebijakan yang tepat untuk meningkatkan produktivitas padi di kecamatan yang kurang berkembang.

Upaya manual dalam menganalisis data produktivitas padi tidak efektif karena besarnya jumlah data yang tersedia. Untuk itu, dibutuhkan metode analisis yang lebih efisien dan

akurat dalam mengelompokkan kecamatan berdasarkan produktivitas padi. Salah satu metode yang diusulkan adalah *K-Means Clustering*, algoritma yang populer dalam pengelompokan data numerik karena kemampuannya dalam mengelompokkan data dengan karakteristik serupa ke dalam kelompok yang berbeda.

Algoritma *K-Means* merupakan salah satu metode *clustering* yang efektif dan banyak diterapkan di berbagai bidang. Dalam penelitian pengelompokan produktivitas padi di Jawa Tengah yang mengategorikan wilayah menjadi tiga kelompok produktivitas: tinggi, sedang, dan rendah [1]. Di Kabupaten Indramayu, algoritma ini digunakan untuk memetakan produktivitas padi dengan hasil yang sangat memuaskan [2]. Pada penelitian di Kabupaten Malang, penerapan *K-Means* menghasilkan akurasi 100% dalam pemetaan hasil produksi padi [3]. *K-Means* juga digunakan untuk analisis potensi lahan pangan di Kalimantan Selatan [4]. dan pengelompokan desa di Kecamatan Krangkeng berdasarkan tingkat pendidikan [5].

Di luar sektor pertanian, algoritma *K-Means* juga digunakan untuk berbagai tujuan lainnya. Penelitian di Jawa Barat menunjukkan bahwa wilayah dengan potensi objek daya tarik wisata (ODTW) berhasil dikelompokkan menjadi tiga *cluster*: tinggi, sedang, dan rendah [6]. Penelitian lain di Kecamatan Sukagumiwang berhasil mengelompokkan tingkat efektivitas pelayanan publik menggunakan nilai optimal  $k = 15$  [7]. Penelitian telah membuktikan efektivitasnya dalam mengelompokkan data untuk berbagai keperluan. Misalnya, dalam pengelompokan data pengiriman paket di Kantor Pos Cirebon [8]. Pada kasus lain, algoritma *K-Means* juga diterapkan untuk mengelompokkan desa rawan bencana di Kabupaten Purbalingga menjadi lima kategori risiko bencana, dari potensi sangat rendah hingga sangat tinggi, yang memberikan panduan penting bagi pemerintah daerah dalam mitigasi bencana [9]. Selain itu, pengelompokan dengan *K-Means* pada data penyebaran *COVID-19* di Jawa Barat menghasilkan tiga *cluster* yang mencerminkan tingkat penyebaran yang berbeda, yaitu tinggi, sedang, dan rendah [10].

Dalam bidang sosial, metode *K-Means* dimanfaatkan untuk pengambilan keputusan strategis, seperti dalam penentuan penerima Bantuan Langsung Tunai di desa, yang mengelompokkan data penduduk berdasarkan kelayakan penerimaan bantuan [11]. *K-Means* juga diterapkan dalam analisis tingkat kemiskinan di provinsi-provinsi di Indonesia [12]. Di bidang geografis dan infrastruktur, algoritma *K-Means* menunjukkan keunggulannya dalam pemetaan kerusakan jalan di Kabupaten Malang dengan tingkat kecocokan metode mencapai 100% [13].

Secara keseluruhan, berbagai penelitian menunjukkan bahwa belum ada penelitian yang secara khusus mengkaji penerapan algoritma *K-Means* pada wilayah Kabupaten Cirebon yang memiliki karakteristik produktivitas padi

yang kompleks dan beragam. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan kecamatan di Kabupaten Cirebon berdasarkan tingkat produktivitas tanaman padi dengan menggunakan algoritma *K-Means Clustering*. Dengan pendekatan ini, diharapkan dapat diperoleh gambaran yang lebih akurat terkait pola produktivitas padi di setiap kecamatan, sehingga dapat memberikan informasi yang bermanfaat bagi pemerintah daerah dalam menyusun kebijakan peningkatan produktivitas padi yang lebih tepat sasaran.

Algoritma *K-Means* dipilih karena kemampuannya yang unggul dalam mengelompokkan data berukuran besar dan kemampuannya mengidentifikasi kelompok dengan karakteristik yang serupa. Selain itu, metode ini memiliki kelebihan dalam mengelompokkan data numerik dengan distribusi yang jelas dibandingkan dengan metode lain seperti DBSCAN, *Fuzzy K-Means*, atau *Fuzzy C-Means* [14][15][16].

Melalui penelitian ini, diharapkan dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dalam bidang pengelompokan data serta memberikan manfaat praktis bagi pemerintah daerah dalam merumuskan kebijakan yang lebih efektif dalam mengelola produktivitas padi di Kabupaten Cirebon.

## 2. METODE PENELITIAN

Metode penelitian yang digunakan adalah pendekatan kuantitatif, yang memungkinkan analisis statistik terhadap data numerik pada komoditas tanaman padi. Algoritma *K-Means* digunakan untuk pengolahan data berupa mengelompokkan kecamatan berdasarkan data produksi padi dari *website* Dinas Pertanian Cirebon.

### 2.1 Metode Pengumpulan Data, Instrumen Penelitian, dan Metode Pengujian

Metode pengumpulan data dalam penelitian ini menggunakan data sekunder yang diperoleh dari *website* resmi Dinas Pertanian Kabupaten Cirebon. Data yang digunakan mencakup luas tanam, luas panen, produksi, dan produktivitas padi pada tahun 2023. Proses pengumpulan data dilakukan dengan mengunduh file *dataset* dalam format *Excel*. Selanjutnya, data ini dianalisis untuk memastikan relevansi dan kelengkapannya sebelum diolah lebih lanjut.

Penelitian ini melibatkan 200 data dari 40 kecamatan di Kabupaten Cirebon sebagai subjek penelitian, dengan karakteristik utama berupa data produktivitas padi yang mencakup luas tanam, luas panen, produksi, dan produktivitas per hektar pada tahun 2023. Alat ukur yang digunakan dalam penelitian ini adalah algoritma *K-Means* untuk pengelompokan data, dengan perangkat lunak *RapidMiner* sebagai media analisis. Data terdiri dari empat item variabel utama yang telah melalui proses normalisasi menggunakan metode *min-max scaling*. Reliabilitas pengelompokan diuji menggunakan *Davies-Bouldin Index*

(DBI) dan metode *elbow* untuk menentukan jumlah *cluster* optimal. Analisis data dilakukan melalui pendekatan kuantitatif dengan teknik *clustering* untuk mengidentifikasi kelompok kecamatan berdasarkan tingkat produktivitas padi.

## 2.2 Tahapan Penelitian

Penelitian ini dilakukan melalui tahapan *Knowledge Discovery in Databases* (KDD) yang meliputi seleksi data, *preprocessing*, transformasi data, *data mining*, dan evaluasi hasil. Tahap seleksi data dilakukan dengan memilih data produktivitas padi dari tahun 2023 yang relevan untuk analisis. Pada tahap *preprocessing*, dilakukan pembersihan data dan normalisasi menggunakan teknik *min-max scaling* untuk menyamakan skala data. Transformasi data dilakukan dengan mengubah data mentah menjadi format yang sesuai untuk proses analisis. Proses *data mining* menggunakan algoritma *K-Means* menghasilkan tiga *cluster* kecamatan dengan produktivitas tinggi, sedang, dan rendah. Evaluasi hasil dilakukan dengan metode *elbow* untuk menentukan jumlah *cluster* optimal dan *Davies-Bouldin Index* (DBI) untuk mengukur validitas pengelompokan. Hasil akhir menunjukkan produktivitas tanaman padi yang bervariasi di Kabupaten Cirebon, memberikan gambaran yang mendalam untuk pengambilan keputusan strategis dalam sektor pertanian.

## 3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, penulis menerapkan tahapan *KDD* (*Knowledge Discovery in Databases*) untuk mengelompokkan kecamatan di Kabupaten Cirebon berdasarkan produktivitas tanaman padi. Adapun tahapan *KDD* tersebut adalah sebagai berikut:

### 3.1 Data Selection

Tahap *data selection* dalam penelitian ini melibatkan pemilihan data produktivitas padi tahun 2023 yang diperoleh dari *website* Dinas Pertanian Kabupaten Cirebon. Data yang dipilih mencakup atribut luas tanam, luas panen, produksi, dan produktivitas padi per hektar di setiap kecamatan. Pemilihan data dilakukan dengan mempertimbangkan relevansi dan kelengkapan informasi untuk mendukung penerapan algoritma *K-Means*. Fokus pada data tahun 2023 dipilih karena data dari tahun-tahun sebelumnya tidak konsisten di semua kecamatan, sehingga memastikan analisis yang lebih akurat dan representatif



Gambar 1. Read Excel

Operator '*Read Excel*' seperti yang terlihat pada Gambar 1, dilakukan untuk mengimpor data produktivitas padi yang telah diunduh dalam format file *Excel* (.xlsx). Proses ini memungkinkan data dimasukkan ke dalam perangkat lunak analisis seperti *RapidMiner* untuk pengolahan lebih lanjut. Setelah data berhasil diimpor, dilakukan pengecekan awal untuk memastikan seluruh atribut seperti luas tanam, luas panen, produksi, dan produktivitas padi telah dimuat dengan benar. Langkah ini penting untuk memastikan data siap digunakan dalam proses *preprocessing* dan analisis menggunakan algoritma *K-Means*.



Gambar 2. Select Attribut

Operator '*Select Attribute*' seperti yang terlihat pada Gambar 2, dilakukan untuk memilih atribut yang relevan dari *dataset* yang diimpor. Dalam penelitian ini, atribut yang dipilih meliputi luas tanam, luas panen, produksi, dan produktivitas padi per hektar, yang digunakan sebagai variabel utama dalam analisis. Atribut-atribut ini dipilih karena memiliki pengaruh signifikan terhadap produktivitas tanaman padi di setiap kecamatan. Sementara itu, atribut yang tidak relevan atau tidak diperlukan untuk analisis diseleksi dari proses pengolahan. Langkah ini memastikan bahwa hanya data yang relevan digunakan dalam proses pengelompokan menggunakan algoritma *K-Means*.

### 3.2 Processing Data

Tahap *Processing Data* dilakukan untuk mempersiapkan data sebelum dianalisis menggunakan algoritma *K-Means*. Langkah ini mencakup *Set Role* untuk menentukan peran masing-masing atribut dalam *dataset* sebelum proses analisis.



Gambar 3. Set Role

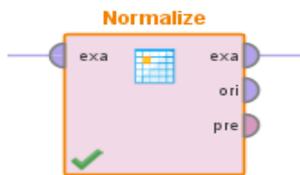
Operator *Set Role* seperti yang terlihat pada Gambar 3, dilakukan untuk menentukan peran masing-masing atribut dalam *dataset* sebelum proses analisis. Dalam penelitian ini, atribut seperti nama kecamatan diberi peran sebagai "*Id*" untuk identifikasi data, tetapi tidak digunakan dalam proses *clustering*. Penetapan peran ini seperti yang dilihat pada Tabel 1, dilakukan untuk memastikan bahwa hanya data numerik yang relevan diproses untuk analisis, sehingga meningkatkan akurasi dan efisiensi pengelompokan.

Tabel 1. Parameter *Set Role*

Parameter	Isi	
Set Role	Edit list:	
	Attribute	Target role
	Nama Kecamatan	Id

### 3.3 Data Transformation

Tahap *Data Transformation* seperti yang terlihat pada Gambar 4, dilakukan untuk mengubah data mentah menjadi format yang sesuai untuk analisis. Dalam penelitian ini, teknik *min-max scaling* digunakan untuk menormalkan nilai atribut seperti luas tanam, luas panen, produksi, dan produktivitas padi ke dalam rentang 0 hingga 1. Transformasi ini penting untuk menyamakan skala data agar tidak ada atribut yang mendominasi perhitungan jarak dalam algoritma *K-Means*. Dengan transformasi ini, proses pengelompokan menjadi lebih akurat, memungkinkan interpretasi yang lebih jelas terkait kecamatan dengan produktivitas tanaman padi tinggi, sedang, dan rendah. Operator yang digunakan untuk tahap normalisasi adalah ‘*Normalize*’



Gambar 4. *Normalize*

### 3.4 Data Mining

*Data mining*, teknik *clustering* yang diimplementasikan adalah algoritma *K-Means Clustering* menggunakan operator ‘*Clustering*’ seperti yang terlihat pada Gambar 5. Operator ini merupakan operator utama pemodelan untuk menghasilkan pengelompokan *dataset*.



Gambar 5. *Clustering K-Means*

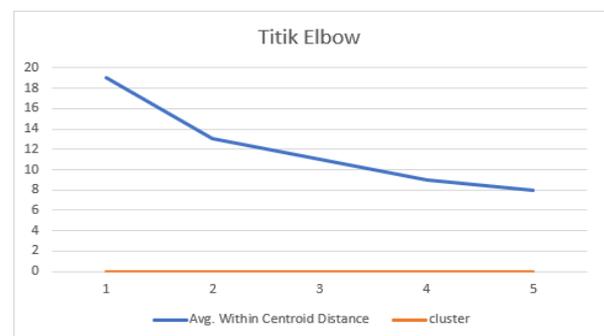
Pada penerapan algoritma *K-Means*, beberapa parameter utama telah diatur untuk memastikan proses *clustering* berjalan optimal. Parameter *k* ditentukan dalam rentang 2 hingga 5 untuk mengeksplorasi jumlah klaster terbaik. Proses *clustering* dilakukan dengan maksimal 10 (*MaxRuns*) pada setiap nilai *K*, menggunakan tipe pengukuran *Bregman Divergences* dengan metode *Square Euclidean Distance* sebagai metrik jaraknya. Selain itu, langkah optimasi maksimum ditetapkan sebanyak 100 langkah (*Max Optimization Steps*) untuk memastikan konvergensi algoritma secara efisien.

Langkah selanjutnya, seperti yang dilihat pada Tabel 2 dilakukan untuk menentukan nilai *k* yang optimal dalam algoritma *K-Means*, langkah penting yang harus dilakukan adalah melakukan evaluasi kualitas hasil *clustering*. Salah satu metode yang sering digunakan adalah metode *elbow*, metode *elbow* merupakan proses untuk mencari titik pada grafik terhadap jumlah *cluster* (*k*) di mana penurunan yang mulai melambat atau membentuk siku (*elbow*).

Tabel 2. Parameter *Elbow*

Cluster	Measure Type	Avg. Centroid Distance
K2	<i>Bregman Divergences</i>	0.019
K3		0.013
K4		0.011
K5		0.009
K6		0.008

Proses penentuan nilai *k* optimal dalam algoritma *K-Means* dilakukan dengan memvisualisasikan data *centroid distance* yang telah dinormalisasi melalui grafik garis. Berdasarkan proses normalisasi menghasilkan beberapa nilai *centroid distance* di bawah nol, untuk keperluan visualisasi titik *elbow* di *Ms. Excel*, peneliti melakukan pembulatan nilai-nilai tersebut. Hal ini dilakukan agar titik *elbow* pada grafik dapat teridentifikasi dengan jelas. Titik *elbow* pada grafik garis yang menunjukkan perubahan kemiringan yang signifikan, kemudian diinterpretasikan sebagai nilai *k* optimal. Dengan demikian, penentuan nilai *k* optimal dalam penelitian ini didasarkan pada interpretasi visual grafik garis yang telah dimodifikasi melalui pembulatan nilai *centroid distance*, sehingga memungkinkan identifikasi titik *elbow* yang representatif. Berikut tampilan titik *elbow* yang telah diperoleh:

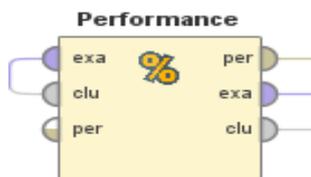


Gambar 6. Titik *Elbow*

Berdasarkan analisis grafik *elbow* pada Gambar 6 menunjukkan nilai *k* optimal berada pada *cluster* 2, penelitian ini bertujuan untuk mengelompokkan kecamatan di Kabupaten Cirebon ke dalam tiga kategori produktivitas padi tinggi, sedang, dan rendah. Oleh karena itu, penelitian ini menetapkan *k = 3* sebagai jumlah *cluster* yang digunakan. Penetapan *k = 3* ini didasarkan pada kebutuhan untuk memperoleh pengelompokan kecamatan yang sesuai dengan tujuan penelitian, yaitu mengidentifikasi dan mengarakterisasi kelompok kecamatan berdasarkan kategori produktivitas tanaman padi.

### 3.5 Evaluasi

Operator ‘Performance Distance’ seperti yang terlihat pada Gambar 7, pada penerapan algoritma *K-Means* mengacu pada evaluasi kinerja *clustering* berdasarkan jarak antar kluster. Evaluasi ini sering digunakan untuk mengukur efisiensi pengelompokan data dengan mempertimbangkan seberapa jauh pusat kluster satu dengan yang lain. Semakin besar jarak antar kluster, semakin baik pengelompokan yang dihasilkan, karena menunjukkan adanya perbedaan yang jelas antara kelompok data yang berbeda.



Gambar 7. Cluster Performance Distance

Dari hasil proses operator *K-Means clustering* dengan menggunakan Parameter *Measure Type Numerical Measures*, *Bregman Divergences*, *Mixed Measures*, serta pembacaan operator *Cluster Distance Performance*, kemudian pada bagian *Main Criterion* pilih menggunakan *Davies Bouldin Index (DBI)*, maka informasi hasil DBI yang diperoleh dapat dilihat pada Tabel 3.

Tabel 3. Perbandingan Nilai DBI

Cluster	Measure Type	Davies Bouldin Index
2	MixedMeasures	0.127
	NumericalMeasures	0.127
	Bregman Measure	0.127
3	MixedMeasures	0.228
	NumericalMeasures	0.228
	Bregman Measure	0.230
4	MixedMeasures	0.291
	NumericalMeasures	0.262
	Bregman Measure	0.222
5	MixedMeasures	0.236
	NumericalMeasures	0.276
	Bregman Measure	0.306

Hasil analisis menunjukkan *cluster* dengan nilai DBI yang paling mendekati 0 diperoleh pada  $k = 2$  yaitu sebesar 0.127 seperti yang terlihat pada Gambar 8. Nilai ini menandakan performa klusterisasi yang cukup baik dalam menghasilkan *cluster* yang optimal.

#### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.019
Avg. within centroid distance_cluster_0: 0.018
Avg. within centroid distance_cluster_1: 0.041
Davies Bouldin: 0.127
```

Gambar 8. Nilai DBI Terkecil

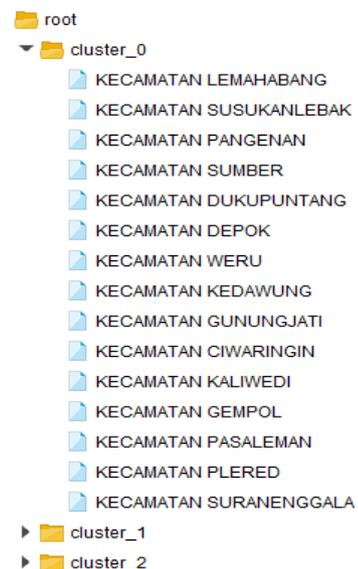
Evaluasi tersebut menunjukkan bahwa nilai  $k$  optimal berada pada *cluster 2*, penelitian ini bertujuan untuk mengelompokkan kecamatan di Kabupaten Cirebon ke dalam tiga kategori produktivitas padi tinggi, sedang, dan rendah. Oleh karena itu, penelitian ini menetapkan  $k = 3$  sebagai jumlah *cluster* yang digunakan. Penetapan  $k = 3$  ini didasarkan pada kebutuhan untuk memperoleh pengelompokan kecamatan yang sesuai dengan tujuan penelitian, yaitu mengidentifikasi dan mengarakterisasi kelompok kecamatan berdasarkan kategori produktivitas tanaman padi. Hasil *cluster* dari  $k = 3$  ini dapat dilihat pada Tabel 4.

Tabel 4. Hasil Cluster k=3

Cluster	Jumlah Anggota
Cluster0	15 items
Cluster1	23 items
Cluster2	2 items
Total numbers of items 40	

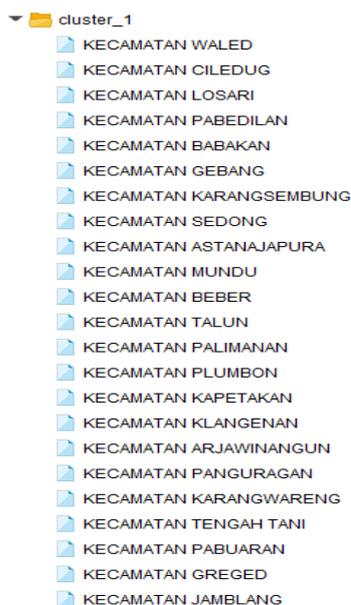
### 3.6 Interpretasi Hasil

Interpretasi merupakan menampilkan hasil proses *clustering K-Means* pada data produktivitas tanaman padi. Kecamatan yang telah dikelompokkan memiliki 3 kategori yaitu, tinggi, sedang dan rendah berdasarkan produktivitas tanaman padi. Pada Gambar 9 menampilkan kecamatan yang terdapat pada *cluster 0*.



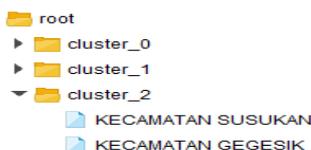
Gambar 9. Cluster 0

*Cluster 0* merupakan kelompok kecamatan yang memiliki produktivitas tanaman padi yang rendah, berdasarkan luas panen dan hasil produksi padi yang rendah dibandingkan kecamatan lain.



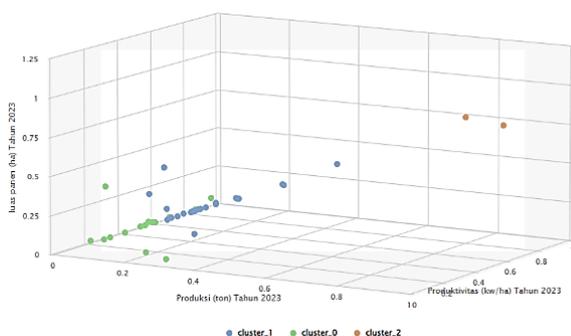
Gambar 10. Cluster 1

Gambar 10 merupakan kelompok kecamatan yang memiliki produktivitas tanaman padi yang sedang. Berdasarkan hasil *clustering* meskipun ada beberapa kecamatan yang memiliki nilai produktivitas tinggi namun luas tanam, panen, dan produksi yang timpang menyebabkan kecamatan tersebut berada di *cluster* 1.



Gambar 11. Cluster 2

Gambar 11 merupakan kelompok kecamatan yang memiliki produktivitas tanaman padi yang tinggi. Berdasarkan hasil *clustering* terdapat 2 kecamatan yaitu kecamatan Susukan dan kecamatan Gegecik. Produktivitas kecamatan Susukan merupakan yang tertinggi dibandingkan kecamatan lain karena wilayah panen yang luas serta produksi padi sangat mempengaruhi produktivitas tanaman padi. Sedangkan kecamatan Gegecik memiliki wilayah tanam dan panen yang luas serta produksi padi yang tertinggi sehingga dapat berada di *cluster* 2.



Gambar 12. Visualisasi Hasil Clustering

Gambar 12 merupakan visualisasi tiga dimensi (*3D scatter plot*) yang menunjukkan hasil *clustering* data produksi padi tahun 2023. Data dikelompokkan menjadi tiga *cluster* ( $k=3$ ) menggunakan algoritma *k-means*.

Setelah klasterisasi dilakukan, karakteristik dari masing-masing klaster dianalisis untuk memberikan wawasan yang lebih mendalam terkait pengelompokan kecamatan berdasarkan produktivitas tanaman padi di kabupaten Cirebon. Klaster pertama (*Cluster 0*) terdiri dari 15 item, yang sebagian besar merupakan kelompok kecamatan yang produksi yang relatif lebih rendah, berkisar antara 0,1 ton hingga 0,4 ton, dengan produktivitas antara 0,1 kw/ha hingga 0,4 kw/ha berdasarkan *dataset* yang telah dinormalisasi. Klaster kedua (*Cluster 1*) memiliki 23 item, yang sebagian besar merupakan kelompok kecamatan yang memiliki kisaran produksi yang lebih luas, dari sekitar 0,2 ton hingga 0,8 ton, dengan produktivitas berkisar antara 0,2 kw/ha hingga 0,8 kw/ha berdasarkan *dataset* yang telah dinormalisasi. Klaster ketiga (*Cluster 2*) berisi 2 item, yang merupakan karakteristik kelompok kecamatan memiliki produksi yang paling tinggi, mencapai sekitar 1 ton, namun dengan produktivitas yang bervariasi, dari 0,6 kw/ha hingga 1 kw/ha berdasarkan *dataset* yang telah dinormalisasi. Dengan membedakan karakteristik klaster ini, dapat memberikan informasi penting untuk perencanaan dan strategi pengembangan pertanian padi di Kabupaten Cirebon, memungkinkan intervensi yang lebih tertarget dan efektif untuk meningkatkan produktivitas padi di masing-masing wilayah. Analisis ini tidak hanya memberikan wawasan tentang pengelompokan kecamatan tetapi juga menjadi dasar untuk pengambilan keputusan strategis yang lebih terfokus dan efisien.

### 3.7 Diskusi

Penelitian ini melibatkan proses pengolahan data, penerapan algoritma *K-Means Clustering*, validasi hasil, serta interpretasi dari hasil pengelompokan kecamatan berdasarkan produktivitas padi di Kabupaten Cirebon. Pengolahan data diawali dengan pengumpulan data produktivitas padi yang mencakup luas tanam, luas panen, dan hasil produksi dari seluruh kecamatan. Proses pra-proses meliputi pembersihan, normalisasi, dan transformasi data agar dapat digunakan dengan baik dalam algoritma *K-Means Clustering*. Algoritma ini kemudian diterapkan untuk mengelompokkan kecamatan berdasarkan tingkat produktivitas padi dengan menggunakan metode *Elbow Method* dan *Davies-Bouldin Index (DBI)* untuk menentukan jumlah klaster yang optimal.

Hasil pengelompokan menunjukkan bahwa kecamatan di Kabupaten Cirebon dapat dibagi menjadi tiga kelompok produktivitas: tinggi, sedang, dan rendah. Kecamatan dengan produktivitas tinggi umumnya memiliki luas panen yang lebih besar dan hasil produksi yang lebih tinggi dibandingkan dengan kecamatan lainnya. Hasil ini dapat digunakan sebagai dasar bagi pemerintah daerah dalam

menyusun kebijakan peningkatan produktivitas padi. Validasi hasil pengelompokan dilakukan dengan mengukur nilai DBI, di mana semakin rendah nilai DBI, semakin baik hasil pengelompokan yang diperoleh. Hasil ini memberikan gambaran yang jelas tentang distribusi produktivitas padi di setiap kecamatan, yang dapat digunakan sebagai dasar dalam perumusan kebijakan peningkatan produktivitas padi.

Penelitian ini menunjukkan bahwa algoritma *K-Means Clustering* efektif untuk mengelompokkan wilayah berdasarkan produktivitasnya. Namun, penelitian ini memiliki beberapa keterbatasan, yaitu tidak adanya perbandingan dengan algoritma lain seperti *Fuzzy C-Means* atau DBSCAN dan tidak mengkaji faktor-faktor lain yang dapat mempengaruhi produktivitas padi. Oleh karena itu, penelitian selanjutnya dapat mengintegrasikan algoritma lain serta mengkaji lebih dalam faktor-faktor yang mempengaruhi produktivitas padi. Penelitian ini diharapkan dapat memberikan manfaat bagi pemerintah daerah dalam merumuskan kebijakan yang lebih tepat dalam mengelola produktivitas padi di Kabupaten Cirebon.

#### 4. KESIMPULAN

Penelitian ini berhasil mengelompokkan kecamatan di Kabupaten Cirebon berdasarkan tingkat produktivitas tanaman padi dengan menggunakan algoritma *K-Means Clustering*. Penerapan algoritma ini mampu mengidentifikasi pola produktivitas padi di setiap kecamatan dan menghasilkan pengelompokan yang lebih akurat dan efisien dibandingkan metode konvensional. Hal ini memberikan gambaran yang jelas mengenai kecamatan-kecamatan yang memiliki tingkat produktivitas tinggi, sedang, dan rendah, sehingga dapat digunakan sebagai dasar untuk perumusan kebijakan yang lebih tepat.

Implikasi dari penelitian ini menunjukkan bahwa algoritma *K-Means Clustering* dapat diterapkan secara efektif dalam bidang pertanian untuk mengelompokkan wilayah berdasarkan produktivitasnya. Penggunaan metode ini juga dapat diterapkan dalam pengelompokan wilayah pertanian lainnya yang memiliki karakteristik produktivitas yang berbeda. Dengan demikian, penelitian ini dapat memberikan kontribusi dalam meningkatkan kualitas perencanaan kebijakan pemerintah di bidang pertanian.

Namun, penelitian ini memiliki beberapa batasan. Pertama, penelitian ini hanya menggunakan algoritma *K-Means Clustering* tanpa melakukan perbandingan dengan algoritma lain seperti *Fuzzy C-Means* atau DBSCAN. Kedua, penelitian ini hanya difokuskan pada pengelompokan kecamatan di Kabupaten Cirebon dan belum mengkaji faktor-faktor lain yang mempengaruhi produktivitas padi.

Penelitian selanjutnya dapat mengintegrasikan algoritma lain untuk dibandingkan dengan hasil yang diperoleh dari *K-Means*, serta mengkaji lebih dalam mengenai faktor-faktor yang memengaruhi produktivitas padi di Kabupaten Cirebon. Selain itu, penelitian ini dapat memberikan

manfaat bagi masyarakat umum dengan memberikan rekomendasi kebijakan yang lebih tepat dan efisien dalam meningkatkan produktivitas tanaman padi, khususnya di wilayah-wilayah yang masih tertinggal.

#### DAFTAR PUSTAKA

- [1] S. Wijayanto and M. Yoka Fathoni, "Pengelompokan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means," *Jupiter*, vol. 13, no. 2, pp. 212–219, 2021.
- [2] R. Farismana, "Penerapan K-Means Clustering Untuk Pemetaan Produktivitas Padi Dan Prediksi Panen Di Kabupaten Indramayu," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 8, no. 3, p. 589, 2024, doi: 10.52362/jisamar.v8i3.1572.
- [3] Y. Y. Prasetya, A. Faisol, and N. Vendyansah, "Sistem Informasi Geografis Hasil Produksi Padi Di Kabupaten Malang Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 5, no. 2, pp. 806–814, 2021, doi: 10.36040/jati.v5i2.3788.
- [4] R. P. Harjono and M. A. I. Pakereng, "Penerapan Metode K-Means Clustering Untuk Analisis Potensi Lahan Pangan Pada Provinsi Kalimantan Selatan," *J-SAKTI (Jurnal Sains Komput. Inform.)*, vol. 7, no. 1, pp. 2549–7200, 2023, [Online]. Available: <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/596%0Ahttps://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/viewFile/596/574>
- [5] T. Suprapti *et al.*, "Analisis Desa Di Kecamatan Krangkeng Berdasarkan Tingkat Pendidikan Menggunakan Algoritma K-Means," vol. 8, no. 2, pp. 1861–1868, 2024.
- [6] H. Habiballoh, A. Faqih, and T. Suprapti, "Implementasi Algoritma K-Means Dalam Mengelompokkan Kabupaten/Kota Di Jawa Barat Berdasarkan Jenis Dan Jumlah Potensi Objek Daya Tarik Wisata," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4270.
- [7] S. Oop Sofiyah, N. R., and R. Danar Dana, "Analisis Efektivitas Pelayanan Publik Menggunakan K-Means Clustering Di Kecamatan Sukagumiwang," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1291–1296, 2023, doi: 10.36040/jati.v7i2.6536.
- [8] A. Febrian, Nana Suarna, and Gifthera Dwilestari, "Penerapan Algoritma K-Means Untuk Mengelompokkan Data Pengiriman Paket Di Kantor Pos Cirebon," *J. Teknol. Technoscintia*, vol. 15, no. 1, pp. 23–27, 2022, doi: 10.34151/technoscintia.v15i1.3858.
- [9] D. I. Ramadhani, O. Damayanti, O. Thaushiyah, and A. R. Kadafi, "Penerapan Metode K-Means Untuk Clustering Desa Rawan Bencana Berdasarkan Data Kejadian Terjadinya Bencana Alam," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 3, p. 749, 2022, doi:

- 10.30865/jurikom.v9i3.4326.
- [10] M. W. Goni, D. Gustian, and F. Sembiring, "Implementasi K-Means Dalam Pengelompokan Penyebaran COVID-19 di Jawa Barat," *Progresif J. Ilm. Komput.*, vol. 17, no. 2, p. 107, 2021, doi: 10.35889/progresif.v17i2.648.
- [11] S. Sari and J. N. Utamajaya, "Sistem Pendukung Keputusan Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Metode Algoritma K-Means Clustering," *J. JUPITER*, vol. 14, no. 1, pp. 150–160, 2022.
- [12] A. Bahauddin, A. Fatmawati, and F. Permata Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *J. Manaj. Inform. dan Sist. Inf.*, vol. 4, no. 1, pp. 1–8, 2021, doi: 10.36595/misi.v4i1.216.
- [13] T. Suryani, A. Faisol, and N. Vendyansyah, "Sistem Informasi Geografis Pemetaan Kerusakan Jalan Di Kabupaten Malang Menggunakan Metode K-Means," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 5, no. 1, pp. 380–388, 2021, doi: 10.36040/jati.v5i1.3259.
- [14] M. P. M, C. Dewi, P. S. Emban, G. A. Wijayanti, N. Aulia, and R. Nooraeni, "Comparison of DBSCAN and K-Means Clustering for Grouping the Village Status in Central Java 2020," *J. Mat. Stat. Komputasi*, vol. 17, no. 3, pp. 394–404, 2021, doi: 10.20956/j.v17i3.11704.
- [15] M. B. Johra, "Soft Clustering Dengan Algoritma Fuzzy K-Means (Studi Kasus: Pengelompokan Desa Di Kota Tidore Kepulauan)," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 2, pp. 385–392, 2021, doi: 10.30598/barekengvol15iss2pp385-392.
- [16] L. Rohmaniah, A. Faqih, and T. Suprpti, "Pengelompokan Penyandang Masalah Kesejahteraan Sosial Di Jawa Barat Menggunakan K-Means Dan Fuzzy C-Means," *J. Teknol. Technoscientia*, vol. 15, no. 1, pp. 1–7, 2022, doi: 10.34151/technoscientia.v15i1.3847.