

KLASTERISASI DATA PRODUKSI PERTANIAN DI KABUPATEN CIREBON DENGAN ALGORITMA K-MEANS

Muhamad Arif Septianto^{1*}, Ahmad Faqih², Ade Rizki Rinaldi³

1,2,3Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) IKMI Cirebon; Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135; Telp. (0231) 490480

Received: 2 Februari 2025

Accepted: 29 Maret 2025

Published: 14 April 2025

Keywords:

clustering, k-means, agricultural products, knowledge discovery in databases (KDD), davies-bouldin index (DBI)

Corespondent Email:
riefwp99@gmail.com

Abstrak: Kabupaten Cirebon merupakan salah satu daerah pertanian terpenting di Indonesia yang memiliki potensi besar di sektor pertanian. Namun, pengelolaan data produksi pertanian di daerah ini sering kali tidak optimal sehingga mengakibatkan rendahnya efisiensi dan strategi pengelolaan sumber daya yang kurang tepat. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam mengelompokkan data produksi pertanian berdasarkan hasil panen dari 40 kecamatan di Kabupaten Cirebon. Metode Knowledge Discovery in Database (KDD) digunakan dalam penelitian ini, meliputi tahap pemilihan data, preprocessing, transformasi, data mining, dan evaluasi hasil cluster. Data yang dianalisis meliputi hasil panen dari tiga komoditas utama, yaitu padi, jagung, dan melinjo. Hasil analisis menunjukkan bahwa data dapat dibagi menjadi empat klaster optimal dengan nilai Davies-Bouldin Index (DBI) sebesar 0,324, yang mengindikasikan bahwa klaster yang terbentuk memiliki kualitas yang baik. Setiap klaster mencerminkan karakteristik produksi yang berbeda-beda di setiap wilayah. Beberapa daerah menunjukkan keunggulan dalam hasil jagung, sementara daerah lain lebih unggul dalam memproduksi padi atau melinjo. Selain itu, terdapat daerah-daerah dengan tingkat produktivitas yang relatif rendah, yang membutuhkan perhatian lebih dalam mengembangkan sektor pertanian. Temuan-temuan ini memberikan wawasan penting dalam merancang pengelolaan sektor pertanian yang lebih efisien dan berbasis data.

Abstract: Cirebon Regency is one of the most important agricultural regions in Indonesia that has great potential in the agricultural sector. However, the management of agricultural production data in this area is often not optimal, which results in low efficiency and inappropriate resource management strategies. Thus, this study aims to apply the K-Means algorithm in clustering agricultural production data based on crop yields from 40 sub-districts in Cirebon Regency. The Knowledge Discovery in Databases (KDD) method is used in this research, including the phases of data selection, preprocessing, transformation, data mining, and evaluation of cluster results. The data analyzed included crop yields of three main commodities, namely rice, corn, and melinjo. The analysis results indicate that the data can be divided into four optimal clusters with a Davies-Bouldin Index (DBI) value of 0.324, which indicates that the cluster formation is of good quality. Each cluster reflects the varying production characteristics of each region. Some regions show superiority in maize yields, while other regions excel in producing rice or melinjo. In addition, there are regions with relatively low productivity levels, which require more attention in developing the agricultural sector. These findings provide crucial insights in designing a more efficient and data-driven management of the agricultural sector.

1. PENDAHULUAN

Sektor pertanian adalah salah satu dari banyak bidang kehidupan yang sangat dipengaruhi oleh

kemajuan teknologi informasi. Sebagai negara yang masih bergantung pada sektor agraris, pertanian sangat penting untuk sumber perekonomian dan pemenuhan kebutuhan makanan. Kabupaten Cirebon memiliki banyak potensi pertanian, dan pertanian sendiri adalah salah satu industri unggulan di Indonesia. Meskipun kabupaten ini memiliki potensi pertanian yang luar biasa, pengelolaan hasil data pertanian masih jauh dari persepsiyaan, Kata Pengantar Pertanian adalah salah satu sektor prioritas yang dipilih untuk memajukan sektor ekonomi lokal Kabupaten. Penggunaan teknologi informasi akan menghalangi pengambilan keputusan tentang komponen strategi dalam pengembangan. Penelitian ini bertujuan untuk menganalisis data produksi pertanian Kabupaten Cirebon menggunakan metode *K-Means*. Temuan yang dihasilkan diharapkan dapat mengidentifikasi pola-pola produksi serupa di berbagai wilayah, sehingga dapat menjadi dasar pengembangan kebijakan pertanian yang lebih efisien dan terarah. Studi sebelumnya, seperti penelitian oleh [1], menunjukkan efektivitas *K-Means* dan *K-Medoids* dalam analisis data pertanian menggunakan *Indeks Davies-Bouldin (DBI)*. Meskipun sektor pertanian Kabupaten Cirebon memiliki potensi besar, masih ada banyak tantangan untuk mengelola data produksi pertanian. Kurangnya pemanfaatan teknologi informasi dalam pengolahan data menyebabkan pengambilan keputusan yang tidak optimal tentang strategi pengembangan pertanian [2]. Pola produksi pertanian juga berbeda-beda di setiap wilayah karena perbedaan kondisi geografis dan iklim. Karena itu, sulit untuk menentukan metode yang paling efektif untuk meningkatkan produktivitas pertanian. menggunakan *K-Means* untuk analisis penjualan dan menemukan performa klastering yang baik. Temuan serupa juga ditunjukkan oleh [3], yang membahas penerapan klastering dalam sektor retail. Sementara itu, penelitian oleh [4], menyoroti pentingnya pemanfaatan metode klastering dalam pengambilan keputusan berbasis data. Dalam situasi ini, metode analisis data yang dapat mengelompokkan wilayah berdasarkan karakteristik produksi pertaniannya diperlukan. Dengan demikian, dapat diidentifikasi wilayah mana yang membutuhkan perhatian khusus. Salah satu metode analisis data yang dapat

digunakan adalah metode *Clustering K-Means*, yang dapat digunakan untuk mengelompokkan data produksi pertanian berdasarkan kesamaan atribut [5]. Metode ini diharapkan dapat meningkatkan pemahaman tentang pola produksi di berbagai wilayah. Ini akan membantu dalam perencanaan dan pengelolaan sumber daya pertanian secara lebih strategis dan efisien [6]. Dengan latar belakang ini, tujuan penelitian ini adalah untuk menggunakan metode *klastering K-Means* untuk menganalisis data produksi pertanian di Kabupaten Cirebon. Tujuannya adalah untuk menemukan daerah dengan karakteristik produksi yang sebanding. Diharapkan hasil penelitian ini dapat membantu meningkatkan produktivitas dan efisiensi sektor pertanian dengan pemanfaatan teknologi informasi yang lebih baik [7]. Selain itu, hasil ini akan menjadi dasar bagi para pemangku kebijakan untuk membuat strategi yang lebih tepat sasaran. Selain itu, penelitian ini memiliki potensi untuk mempercepat digitalisasi dan modernisasi sektor pertanian Indonesia, yang akan meningkatkan daya saing pertanian baik di dalam negeri maupun di tingkat global.

2. TINJAUAN PUSTAKA

2.1. *Algoritma K-Means*

Algoritma *K-Means Clustering* merupakan teknik pengelompokan yang mengorganisasi data ke dalam kluster berdasarkan tingkat kemiripan, dengan tujuan agar data dalam satu kluster memiliki karakteristik yang serupa. Langkah-langkahnya meliputi menentukan jumlah kluster (*K*), menginisialisasi titik pusat (*centroid*), mengelompokkan data berdasarkan jarak ke *centroid*, serta memperbarui posisi *centroid* hingga mencapai kestabilan. Keunggulan algoritma ini adalah mudah dipahami dan *efisien*, sementara kekurangannya mencakup perlunya menetapkan jumlah kluster di awal serta kerentanannya terhadap pemilihan titik pusat awal [8].

2.2. *Davies-Bouldin Index (DBI)*

Davies-Bouldin Index (DBI) adalah metrik yang digunakan untuk menilai kualitas hasil pengelompokan data dengan mengukur sejauh mana kluster yang dihasilkan saling terpisah dan seberapa rapat kluster tersebut. *DBI* dihitung dengan membandingkan jarak antara *centroid* kluster dengan tingkat kerapatan dalam

kluster itu sendiri. Nilai *DBI* yang lebih rendah menunjukkan kualitas *kluster* yang lebih baik, karena mencerminkan jarak antar kluster yang lebih besar dibandingkan jarak dalam kluster. Dengan kata lain, semakin kecil nilai *DBI*, semakin baik tingkat pemisahan dan kekompakan kluster yang terbentuk [9].

2.3. Clustering

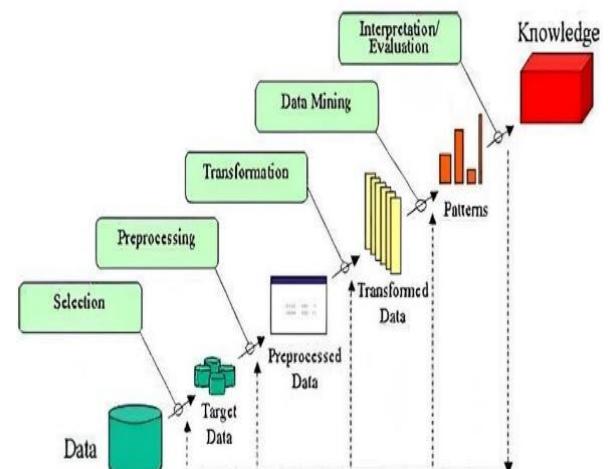
Clustering adalah teknik analisis data yang digunakan untuk mengelompokkan objek atau data berdasarkan karakteristik yang sebanding. Teknik ini tidak membutuhkan label sebelumnya dan bertujuan untuk menemukan pola atau struktur yang tersembunyi dalam data. Data mining, pembelajaran mesin, dan analisis statistik sering menggunakan clustering untuk menemukan hubungan dalam dataset yang besar dan kompleks. *K-Means* adalah salah satu algoritma clustering yang paling populer yang membagi data ke dalam sejumlah kelompok, juga disebut klaster, berdasarkan seberapa dekat atau serupa beberapa fitur [10].

2.4. Data Mining

Data Mining adalah serangkaian langkah yang digunakan untuk mengekstrak informasi berharga yang sebelumnya tidak diketahui dari pengolahan data secara manual dalam sebuah basis data. Konsep *Data Mining* mulai dikenal pada tahun 1990-an. Data mining sering diterapkan dalam berbagai bidang, seperti bisnis, kesehatan, keuangan, dan pertanian, termasuk dalam analisis produksi pertanian di Kabupaten Cirebon. Dengan data mining, pola tersembunyi dalam data dapat diidentifikasi untuk meningkatkan efisiensi dan strategi pengelolaan sumber daya [11]. Beberapa teknik utama dalam data mining meliputi clustering untuk mengelompokkan data serupa, classification untuk memprediksi kategori data, serta association rule learning yang mengungkap hubungan antar variabel. Penerapan data mining sangat luas, mulai dari bisnis untuk menganalisis perilaku pelanggan, kesehatan untuk mendeteksi penyakit berdasarkan riwayat medis, hingga keuangan dalam memprediksi pergerakan pasar. Dengan kemampuannya mengolah data secara mendalam, data mining menjadi alat yang sangat berharga dalam era digital dan big data.

3. METODE PENELITIAN

Dalam penelitian ini menggunakan metode pendekatan kuantitatif dengan teknik data mining yaitu *klastering K-Means*. Adapun tujuan dari penelitian ini adalah untuk menganalisis dan mengklasifikasikan setiap kecamatan yang ada di Kabupaten Cirebon berdasarkan hasil produksi pertanian pada setiap Kecamatan tersebut yang terkumpul dalam suatu kategori tertentu. Pada tahap ini, juga digunakan aplikasi Microsoft Excel dan menggunakan RapidMiner untuk menjamin ketepatan penyajian data. Adapun Model pendekatan penelitian ini berasal dari Proses Penemuan Pengetahuan dalam *Database KDD* [12]. *KDD* sendiri terbagi dalam lima bagian, yaitu pemilihan data, praproses, transformasi, data mining serta evaluasi, tahapannya mulai dari penumpukan data hasil dari pertanian pada masing-masing Kecamatan kabupaten cirebon, data ini didapat dari beberapa observasi terhadap teknologi pertanian, dukumentasi pustaka, dan dokumen resmi yang diperoleh dari Dinas Pertanian kabupaten Cirebon. Langkah selanjutnya Proses *K-Means* untuk menentukan klaster sesuai dengan tujuannya.



Gambar 3. 1 Tahapan KDD

3.1 Data Selection

Pertama adalah mengumpulkan data dari Dinas Pertanian Kabupaten Cirebon. Tiga elemen

utama termasuk dalam data yang dikumpulkan. Kecamatan, jenis tanaman, dan jumlah hasil produksi menunjukkan keanekaragaman komoditas yang dikelola petani di setiap wilayah. Sementara itu, jumlah hasil produksi menunjukkan tingkat produktivitas masing-masing kecamatan. Pengumpulan data ini dilakukan secara terstruktur untuk memastikan bahwa data tersebut lengkap dan relevan, dan bahwa mereka dapat digunakan untuk mendukung analisis mendalam tentang sektor pertanian Kabupaten Cirebon.

3.2 Data Preprocessing

Untuk menjamin kualitas data, proses pengolahan data dimulai dengan pembersihan. Dalam langkah ini, kesalahan data diidentifikasi dan diperbaiki, mengatasi data hilang, untuk menghapus outlier. Data yang telah dibersihkan kemudian dinormalisasi agar setiap variabel memiliki skala yang sama dan siap untuk proses klastering.

3.3 Data Transformation

Pada tahap ini, data diolah lebih lanjut untuk menjamin bahwa ia sesuai dengan algoritma pengelompokan yang akan diterapkan. Operator *Nominal to Numerical* berfungsi untuk mengonversi atribut nama kecamatan dari format nominal menjadi format numerik. Algoritma *klastering*, yang dirancang untuk menangani angka, melakukan transformasi ini untuk memungkinkan pengolahan data yang efisien.

3.4 Data Mining

Di fase ini, metode kluster diterapkan dengan memanfaatkan algoritma *K-Means*, yang merupakan algoritma terkenal dalam penambangan data yang dibuat untuk mengelompokkan data berdasar kesamaan tertentu. *K-Means* berfungsi dengan cara mengelompokkan data ke dalam sejumlah kelompok atau kluster, di mana setiap kluster terdiri dari data yang memiliki ciri-ciri yang sama. Sebagai bagian dari pendekatan pembelajaran tanpa pengawasan, algoritma ini tidak memerlukan label atau kategori yang telah ada sebelumnya. Dengan kata lain, *K-Means* secara otomatis menemukan pola-pola tersembunyi dalam data dan membentuk *kluster* berdasarkan kemiripan atribut. Metode ini sangat efisien untuk menyelidiki dan

memahami struktur alami data, sehingga menjadi alat yang handal untuk analisis rumit tanpa keterlibatan langsung dari pengguna

3.5 Evasluasi

Evaluasi klastering adalah memastikan bahwa klaster yang terbentuk memiliki kualitas optimal. Metode *Indeks Davies-Bouldin (DBI)* diterapkan untuk menilai kualitas *klaster* yang terbentuk, dengan mempertimbangkan tingkat homogenitas data dalam setiap klaster serta jarak antarklaster. Nilai *DBI* yang lebih kecil menunjukkan bahwa data dalam tiap *klaster* sangat mirip, sedangkan klaster yang berbeda memiliki. Penilaian ini merupakan tahap krusial dalam penelitian sebab hasilnya akan menentukan apakah pola *klaster* yang terbentuk benar-benar merepresentasikan karakteristik produksi pertanian di setiap daerah. Penelitian ini dapat memberikan wawasan mendalam tentang pola distribusi hasil pertanian serta potensi unggulan di masing-masing daerah Kabupaten Cirebon dengan memanfaatkan *klaster* yang sesuai.

3.6 Knowledge

Hasil proses *KDD* dapat disajikan dalam bentuk peta tematik, tabel, atau grafik. Ini adalah representasi dari setiap klaster penerima bantuan. Visualisasi ini memudahkan pemahaman pola distribusi dan ciri penerima. Penelitian akan ditafsirkan secara menyeluruh sehingga pemangku kebijakan dapat memperoleh rekomendasi seperti strategi distribusi yang lebih tepat sasaran dan efisien. Diharapkan pengalaman ini akan membantu meningkatkan kualitas program bantuan sosial di Cirebon.

4. HASIL DAN PEMBAHASAN

Dalam hasil ini, saya akan menjelaskan hasil penelitian saya tentang metode algoritma *K-Means* untuk mengelompokkan data produksi pertanian di Kabupaten Cirebon. Saya menggunakan metode analisis data yang sangat populer di kalangan peneliti, *Knowledge Discovery Databases (KDD)*. Seperti yang digambarkan pada gambar 3., tahap awal penelitian terdiri dari *Data Selection*, lalu lanjut pada tahap *Data Preprocessing*, setelah itu *Data Transformation*, *Data mining*, dan *Evaluation*

4.1. Data Selection

Dinas Pertanian Kabupaten Cirebon menyediakan data untuk penelitian ini. Dataset ini terdiri dari nama empat puluh 40 kecamatan di Kabupaten Cirebon dan mencakup tiga jenis tanaman makanan dengan hasil produksi dari Januari hingga September 2024. Data yang telah diperoleh sebelumnya tidak diproses secara langsung. Data yang diolah melalui metode clustering termasuk nama kecamatan, hasil produksi jagung, dan Melinjo. Gambar 3.2 menunjukkan cara penerapan data selection pada RapidMiner. Pertama, dataset dibaca dari file *Excel*, dan kemudian ditambahkan operator *Read Excel*.



Gambar 3. 2 Mengimpor Data

Tahap awal adalah menambahkan operatoar *Read Excel* untuk mengimpor Data Pertanian Kabupaten Cirebon, setelah itu kita melimilih data apa saja yang akan diamasukan pada penelitian.

Import Data: Select the cells to import.				
Select the cells to import.				
Sheet:	Sheet1	Cell range:	A:D	Select All
		Define header row:	1	
1	Kecamatan	B	Jagung	Melinjo
2	Ward	11.988	139.000	0.800
3	Puncakmas	539.000	0.000	0.000
4	Giring	4.198	1.490	57.000
5	Pluitur	3.007	4.695	114.000
6	Luter	8.698	223.000	323.000
7	PulauBesi	16.001	12.120	53.000
8	Dukuh	12.470	7.120	242.000
9	Cedek	11.247	1.550	159.000
10	Kronengsing	7.181	110.000	11.000
11	Kronengsing	91.389	363.000	98.000
12	Lembang	7.020	173.000	17.000
13	Sukurak	11.010	60.000	6.000
14	Sedep	13.079	25.000	110.000
15	Antengreja	8.002	0.000	19.000
16	Pangraran	8.037	692.000	9.000
17	Munda	7.771	370.000	29.000
18	Huter	8.905	289.000	1230.000
19	Grogot	8.144	293.000	1730.000
20	Latuh	7.679	160.000	46.000

Gambar 3. 3 Tampilan Dataset dari hasil pertanian

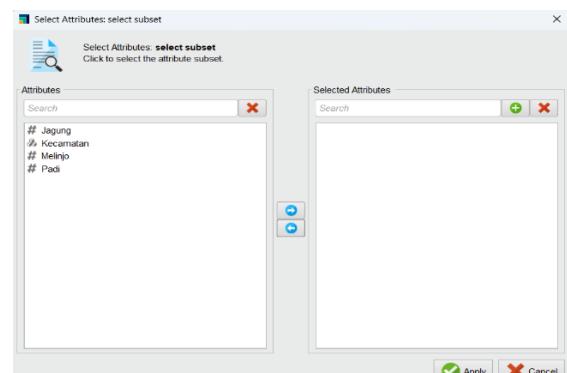
Setelah melalukan pemilahan cells kita beralih ke operator *select attributes* untuk melakukan pemilihan attributes mana saja yang kita gunakan bisa dilihat pada gambar. Operator ini sangat penting dalam proses praproses data, terutama saat kita ingin mengurangi jumlah

atribut yang akan digunakan, sehingga model menjadi lebih sederhana dan efisien.



Gambar 3. 4 Operator Select Atributes

Select attributes adalah operator untuk memelilih attributes apa saja yang akan kita pakai pada Raspidminer.



Gambar 3. 5 Tampilan attributes yang akan di pilih

Saya memilih 3 (tiga) *Atribut* utama Padi, Melinjo, dan Jagung untuk digunakan dalam proses pengolahan data ini. Operator Pilih *Atribut* dapat memilih atribut mana saja yang terkait dengan pemodelan atau analisis yang akan dilakukan. Nama kecamatan sebagai identitas wilayah dan jenis tanaman pangan yang menjadi subjek pengolahan data dipilih. Adapun *atribut* ini dipilih untuk memastikan bahwa hanya data yang signifikan dan sesuai dengan tujuan analisis yang digunakan. Ini membuat proses lebih fokus dan efisien.

4.2. Data Preprocessing

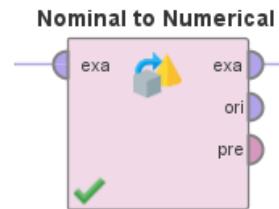
Saat ini, normalisasi data telah selesai, sehingga dataset tidak memiliki nilai yang hilang lagi. Ini memastikan bahwa semua data telah diolah dengan format yang seragam, mengurangi kesalahan yang mungkin terjadi selama pemodelan atau analisis. Dalam proses ini, atribut seperti nomor dan tahun yang dianggap tidak relevan atau tidak diperlukan tidak dipilih untuk dianalisis. Aspek ini dihilangkan agar proses pengolahan data tidak terganggu. Dengan cara ini, model dapat berkonsentrasi pada variabel yang sangat penting. Seperti yang ditunjukkan pada Gambar 3.6, kualitas data yang tidak hilang menunjukkan bahwa dataset tidak ada *missing value*. Jika dataset bersih dan terfokus, langkah selanjutnya dapat dilakukan dengan lebih efisien dan menghasilkan hasil yang optimal.

Name	Type	Missing	Statistics	FB
id	Integer	0	Min: 1 Max: 40 Sum: 20.500	
cluster	Nominal	0	Count: cluster_3 (1) Count: cluster_1 (14) Count: cluster_0 (31), cluster_2 (4), ... [2 more]	
Padi	Real	0	Min: 1.452 Max: 329 Sum: 10.059	
Jagung	Real	0	Min: 0 Max: 992 Sum: 112.000	
Minyak	Integer	0	Min: 0 Max: 1730 Sum: 294.200	

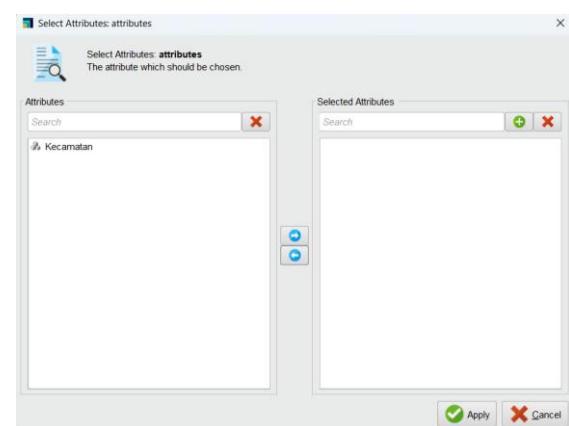
Gambar 3. 6 Hasil dari Preprocessing Data

4.3. Transformasi

Sekarang, atribut nama kecamatan yang awalnya berbentuk nominal dapat diubah menjadi format numerik dengan menggunakan operator *Nominal to Numerical*. Langkah ini sangat penting karena banyak algoritma, seperti yang digunakan untuk clustering, membutuhkan data numerik agar dapat diolah secara optimal. Nilai kategori pada atribut nominal diubah oleh operator ini menjadi representasi numerik, menggunakan angka unik untuk setiap kategori. Tujuan proses ini adalah untuk memastikan bahwa data sesuai dengan algoritma yang akan digunakan tanpa kehilangan informasi penting dari atribut nominal. Gambar 3.8 menunjukkan bagaimana operator ini menampilkan hasil transformasi atribut yang telah dipilih sebelumnya dengan *Select Attributes*. Ini menyediakan data untuk proses clustering berikutnya.

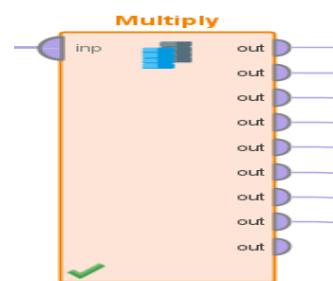


Gambar 3. 7 Operator Nominal To Numerical



Gambar 3. 8 Tampilan dari Select Attribute Nominal To Numerical

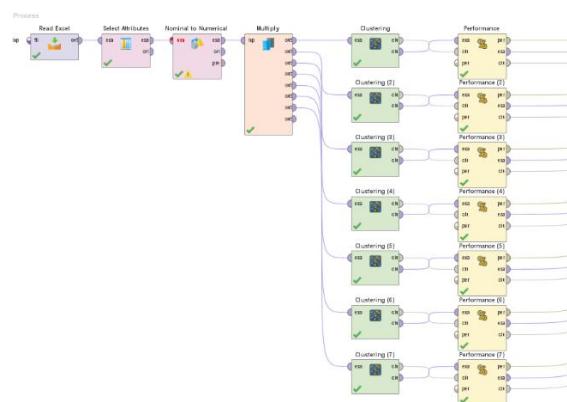
Setelah semua tahap diatas disini saya menambah operator tambahan lagi dikarna saya melakukan beberapa percobaan maka dari itu saya menambahkan *Operator Multiply* untuk menggandakan data atau proses tertentu, sehingga kita bisa membuat beberapa salinan dari satu set data atau hasil proses. Fungsinya mirip dengan *copy-paste* untuk data atau proses, yang sangat berguna dalam berbagai situasi di mana kita ingin menggunakan data atau proses yang sama di beberapa tempat tanpa perlu memuat ulang atau mengulangi proses yang sama.



Gambar 3. 9 Operator Multiply

4.4. Data Mining

Selanjutnya, yaitu tahap penerapan Algoritma *K-Means* menggunakan metode *Clustering* untuk menganalisis data yang telah diproses. Pada tahap pertama tersebut dengan menentukan jumlah K pada klaster. Tujuan dari tahap tersebut yaitu untuk menentukan apakah jumlah klaster apakah paling efektif pada proses *Clustering*. Selanjutnya adalah pengujian klaster dengan 8 kelompok yang berbeda, salah satunya pada kelompok satu dengan jumlah k=2, kelompok dua dengan k=3, kelompok tiga dengan k=4, kelompok empat dengan k=5, kelompok lima dengan k=6, kelompok enam dengan k=7, kelompok tujuh dengan k=8.



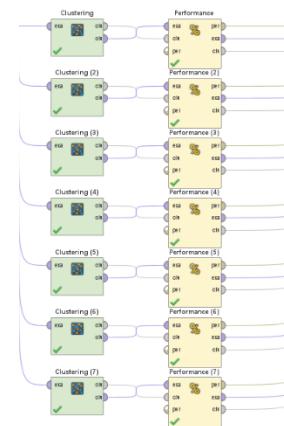
Gambar 3. 10 Model Metode Clustering *K-Means*

Bisa dilihat pada gambar 3.10, saya telah membuat model *clustering K-Means* menggunakan *RapidMiner*. Pada tahap ini, saya melakukan delapan percobaan *clustering* dengan tujuan untuk menghasilkan nilai cluster terbaik yang mungkin. Dengan melakukan beberapa percobaan, saya dapat membandingkan hasil untuk memastikan bahwa model yang dihasilkan memiliki tingkat akurasi dan efisiensi terbaik, sekaligus memberikan pemahaman yang lebih mendalam tentang pola data yang dikelompokkan

4.5. Evaluasi

Tahap *evaluasi* ini melibatkan penggunaan operator kinerja cluster jarak jauh untuk menilai kualitas hasil *clustering*. Konfigurasi dengan jumlah *cluster* 2, 3, 4, 5, 6, 7, dan 8 dicoba bisa dilihat pada Gambar 3.10. Untuk mengukur kualitas clustering, operator ini menggunakan nilai *Davies-Bouldin Index (DBI)* sebagai

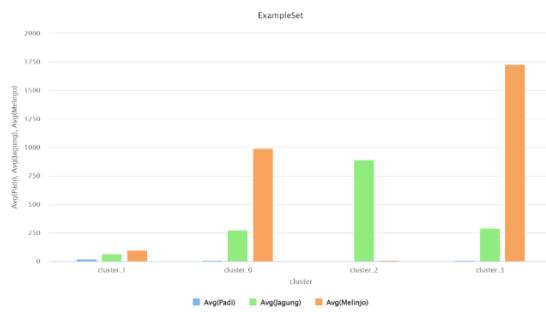
metrik utama. Nilai *DBI* yang lebih rendah menunjukkan jarak antar-cluster yang lebih kecil dan jarak intra-cluster yang lebih kecil, yang menunjukkan hasil clustering yang lebih baik. Hasil pengujian untuk masing-masing jumlah cluster ditunjukkan di sini. Ini menunjukkan bagaimana kinerja model berubah dengan jumlah cluster yang digunakan. Analisis hasil ini digunakan untuk menentukan jumlah cluster yang tepat untuk dataset.



Gambar 3. 11 Operator Cluster distance performance

Berdasarkan hasil pada masing-masing nilai *DBI* (*Davies Bouldin Index*) hasil yang paling rendah diantara masing-masing *Cluster* adalah dengan jumlah k=3 dengan nilai DBI sebesar 0., maka *Cluster* dengan jumlah k=3 lebih baik dibandingkan dengan *Cluster* jumlah k=2, k=4, k=5 dan k=6. Maka dari itu nilai k=3 akan dilakukan analisis untuk melihat *Cluster* dengan jumlah kelurahan, kecamatan, jarak, dan umur tertinggi dan terendah.

Pada tahap ini saya akan menjelaskan karakteristik dari setiap klaster hasil penerapan algoritma *K-Means* pada data produksi pertanian dapat dianalisis dari rata-rata hasil produksi (Padi, Jagung, dan Melinjo). Berdasarkan grafik rata-rata (*average*) bisa dilihat pada gambar 4.12 hasil produksi padi, jagung, dan melinjo untuk setiap klaster, berikut adalah analisis detail untuk masing-masing wilayah klaster:



Gambar 3. 12 Grafik Averege Cluster

Berdasarkan grafik rata-rata (*average*) bisa dilihat pada gambar 3.13 hasil produksi padi, jagung, dan melinjo untuk setiap klaster, berikut adalah analisis detail untuk masing-masing wilayah klaster:

- a) *Cluster 0* Karakteristik jagung mendominasi produksi di klaster ini dengan nilai rata-rata yang cukup tinggi. Produksi padi terlihat rendah, dan melinjo hanya memiliki kontribusi kecil dibandingkan jagung. Wilayah ini merupakan daerah yang cocok untuk budidaya jagung, menunjukkan spesialisasi tanaman tertentu. Strategi yang dapat diterapkan meliputi pengoptimalan teknologi pertanian untuk jagung, seperti varietas unggul dan manajemen hama yang efisien.
- b) *Cluster 1* Karakteristik Produksi padi, jagung, dan melinjo hampir tidak signifikan, terlihat dari rata-rata hasil yang sangat rendah pada semua jenis tanaman. Wilayah ini memiliki tingkat produksi yang sangat rendah, menunjukkan kemungkinan keterbatasan dalam infrastruktur pertanian, kualitas tanah, atau akses ke sumber daya seperti air dan pupuk. Fokus kebijakan dapat diarahkan pada pengembangan fasilitas atau penelitian untuk meningkatkan produktivitas.
- c) *Cluster 2* Karakteristik Produksi jagung sangat signifikan, hampir mendekati angka produksi maksimal, menunjukkan dominasi hasil jagung di wilayah ini. Melinjo berada di tingkat rata-rata yang moderat, sementara padi hampir tidak ada produksi. Wilayah ini menunjukkan potensi besar untuk pengembangan jagung sebagai komoditas utama, dengan melinjo
- d) *Cluster 3* Karakteristik Produksi melinjo adalah komoditas utama di klaster ini dengan rata-rata produksi yang sangat tinggi. Produksi jagung dan padi rendah atau tidak signifikan. Wilayah ini menunjukkan fokus pada budidaya melinjo, menjadikannya sebagai pusat produksi melinjo. Pengelolaan rantai pasok melinjo dan akses pasar yang lebih baik dapat menjadi prioritas untuk memaksimalkan keuntungan.

sebagai komoditas sekunder. Fokus pengembangan dapat diarahkan pada diversifikasi tanaman untuk meningkatkan hasil melinjo dan mengatasi risiko dari ketergantungan pada satu komoditas.

- d) *Cluster 3* Karakteristik Produksi melinjo adalah komoditas utama di klaster ini dengan rata-rata produksi yang sangat tinggi. Produksi jagung dan padi rendah atau tidak signifikan. Wilayah ini menunjukkan fokus pada budidaya melinjo, menjadikannya sebagai pusat produksi melinjo. Pengelolaan rantai pasok melinjo dan akses pasar yang lebih baik dapat menjadi prioritas untuk memaksimalkan keuntungan.

4.6. Knowlede

Berdasarkan hasil simulasi mulai dari $K=2$ hingga $K=8$ diperoleh hasil DBI sesuai yaitu kluster 4, terlihat pada Tabel 3.1, bahwa nilai DBI terkecil ada pada $K=4$, yaitu 0.324, yang menunjukkan bahwa $K=4$ adalah jumlah *klaster* yang optimal untuk penelitian ini, karena memberikan pemisahan antar *klaster* yang terbaik dengan kepadatan yang cukup tinggi, sehingga memastikan pembentukan kelompok wilayah produksi pertanian yang lebih efektif.

Tabel 3. 1 Hasil sari DBI Algoritma K-Means

JUMLAH CLUSTER	DBI Algoritma K-Means
2	0.463
3	0.378
4	0.324
5	0.434
6	0.437
7	0.413
8	0.374

5. KESIMPULAN

- a) Penelitian ini berhasil mengelompokkan data produksi pertanian dari 40 kecamatan di Kabupaten Cirebon menggunakan algoritma *K-Means*. Dengan pendekatan *Knowledge Discovery in Databases* (*KDD*), data pertanian diproses dalam beberapa tahapan, termasuk pemilihan data, preprocessing, transformasi, data mining, dan evaluasi. Hasil klasterisasi menunjukkan bahwa empat klaster optimal terbentuk dengan nilai *Davies-Bouldin Index* (*DBI*) sebesar 0.324, yang mengindikasikan bahwa pembentukan klaster memiliki kualitas yang baik.
- b) Hasil klasterisasi menunjukkan bahwa setiap wilayah memiliki karakteristik produksi yang berbeda. Beberapa wilayah lebih unggul dalam produksi jagung, sementara wilayah lain lebih produktif dalam menanam padi atau melinjo. Selain itu, terdapat daerah dengan tingkat produktivitas yang rendah, yang memerlukan perhatian khusus dalam perencanaan strategi peningkatan produksi pertanian. Dengan memahami pola ini, strategi pengelolaan pertanian dapat disesuaikan untuk meningkatkan efisiensi dan produktivitas di masing-masing daerah.
- c) Hasil penelitian ini memberikan wawasan penting dalam pengelolaan sektor pertanian berbasis data. Dengan memanfaatkan metode klasterisasi, pemerintah daerah dan pemangku kepentingan dapat membuat kebijakan yang lebih terarah, seperti optimalisasi distribusi sumber daya, pengembangan teknologi pertanian spesifik untuk setiap klaster, dan peningkatan kesejahteraan petani. Pendekatan ini juga dapat menjadi dasar untuk penelitian lanjutan yang berfokus pada pengembangan strategi pertanian berbasis teknologi dan data mining.

UCAPAN TERIMA KASIH

Saya dengan tulus menyampaikan rasa terima kasih kepada Dinas Pertanian Kabupaten Cirebon atas dukungan luar biasa yang diberikan dalam bentuk data primer, laporan resmi, dan informasi lapangan yang menjadi landasan utama penelitian ini. Dukungan ini sangat membantu kami dalam mengidentifikasi pola produksi pertanian di 40 kecamatan di Kabupaten Cirebon. Kami juga berterima kasih kepada STMIK IKMI Cirebon atas penyediaan fasilitas penelitian, termasuk perangkat lunak dan ruang diskusi, yang mendukung proses analisis data menggunakan algoritma K-Means. Apresiasi setinggi-tingginya kami sampaikan kepada rekan-rekan sejawat yang telah memberikan masukan konstruktif selama tahapan praproses data, transformasi, dan evaluasi hasil klasterisasi. Tidak lupa, kami menghargai peran berbagai pihak yang mendukung di balik layar, termasuk kolega yang memberikan saran praktis terkait penerapan metode *Knowledge Discovery in Databases* (*KDD*) dan keluarga yang senantiasa memberikan motivasi dan semangat selama penelitian berlangsung. Semua bentuk kontribusi ini menjadi fondasi keberhasilan penelitian ini, yang kami harapkan dapat memberikan manfaat nyata bagi pengelolaan sektor pertanian berbasis data di Kabupaten Cirebon dan menjadi referensi dalam pengembangan kebijakan pertanian yang lebih efisien dan terarah di masa depan.

DAFTAR PUSTAKA

- [1] M. Nafilah, N. Rahaningsih, and R. D. Dana, “Perbandingan Algoritma K-Means dan K-Medois Untuk Pengelompokan Hasil Pertanian di Kabupaten Cirebon,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 223–229, 2024.
- [2] N. B. Nugraha, E. Alimudin, and B. V. Indriyono, “Implementasi K-Means Clustering Pada Sistem Pakar Penentuan Jenis Sayuran,” *J. Innov. Inf. Technol. Appl.*, vol. 4, no. 2, pp. 133–144, 2022.
- [3] A. Nugraha, O. Nurdianwan, and G. Dwilestari, “Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Yana Sport,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 849–855, 2022.
- [4] C. J. Silalahi, A. Situmorang, and J. F. Naibaho, “Implementasi Metode K-Means Clustering Untuk Memetakan Daerah

- Potensial Penghasil Padi di Provinsi Sumatera Utara,” *Methotika J. Ilm. Tek. Inform.*, vol. 2, no. 2, pp. 49–57, 2022.
- [5] A. Rohmatullah, D. Rahmalia, and M. S. Pradana, “Klasterisasi Data Pertanian di Kabupaten Lamongan Menggunakan Algoritma K-Means Dan Fuzzy C-Means,” *J. Ilm. Teknosains*, vol. V, no. 2, pp. 86–93, 2019.
- [6] J. S. Hariyono, H. Z. Zahro, and R. Primaswara, “Perancangan Sistem Informasi Geografis Hasil Produksi Pertanian Bawah Merah di Kabupaten Nganjuk Menggunakan Metode K-Means,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 487–494, 2021.
- [7] E. Suprapto, “Pengelompokan Potensi Padi di Indonesia Menggunakan K-Means Cluster,” *J. Ilm. Pop.*, vol. 5, no. 2, pp. 28–34, 2022.
- [8] I. Febriani, M. Safii, and O. Alfina, “IMPLEMENTASI DATA MINING PENINGKATAN PRODUKSI BERAS MENGGUNAKAN METODE K-MEANS CLUSTERING,” *Maj. Ilm. METHODA*, vol. 12, no. 3, pp. 258–268, Dec. 2022.
- [9] A. S. Ilmananda and H. D. Ranglalin, “Klasterisasi Negara Pengekspor Beras ke Indonesia Menggunakan Algoritma K-Means Clustering,” *J. SISFOTENIKA*, vol. 13, no. 2, pp. 139–150, 2023.
- [10] E. R. Arini, “Penerapan K-Means Cluster di Provinsi Jawa Timur Berdasarkan Ketahanan Pangan,” *JSNu J. Sci. Nusant.*, vol. 3, no. 1, pp. 32–36, 2023.
- [11] A. Aziz, A. M. Siregar, and C. Zonyfar, “Penerapan Algoritma K-Means dan Fuzzy C-Means untuk Pengelompokan Kabupaten Kota Berdasarkan Produksi Padi di Provinsi Jawa Barat,” *Sci. Student J. Information, Technol. Sci.*, vol. III, no. 1, pp. 1–8, 2022.
- [12] P. S. Rosiana, A. A. Mohsa, M. A. Fadila, and J. H. Jaman, “Visualisasi Data Tindak Kejahatan Berdasarkan Jenis Kriminalitas Di Kabupaten Karawang Dengan Menggunakan Algoritma Clustering K-Means,” *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, pp. 822–828, 2023.