

PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS UNTUK PENGELOMPOKAN HASIL PERTANIAN DI KABUPATEN CIREBON

Mala Nafilah, Nining Rahaningsih, Raditya Danar Dana

Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No. 10B, Karyamulya, Kec. Kesambi, Kota Cirebon 45131, Jawa Barat

malanafilah1@gmail.com

ABSTRAK

Dataset mengenai pengelompokan hasil produksi pertanian Dinas Pertanian Kabupaten Cirebon saat ini masih belum tersedia. Keterbatasan dataset ini membuat kurangnya informasi mengenai perbedaan dan keunggulan hasil pertanian antar kecamatan. Oleh karena itu, sebagai salah satu usaha dalam membantu Dinas Pertanian untuk memudahkan pemberian informasi kepada masyarakat dan pihak tertentu dalam mencari informasi, dengan melakukan pengolahan data yaitu pengelompokan yang dapat menyelesaikan permasalahan tersebut. Diperlukan adanya pengelompokan yaitu untuk pemetaan kecamatan dengan hasil produksi pertanian menggunakan metode *clustering*. Metode *clustering* yang diterapkan adalah algoritma *K-Means* dan *K-Medoids*. Kemudian dilakukan perbandingan berdasarkan nilai *Davies Bouldin Index* (DBI) kedua algoritma agar bisa memilih algoritma *clustering* terbaik. Tujuannya untuk mengetahui setiap Kecamatan di Kabupaten Cirebon dengan hasil pertanian yang terkumpul dalam suatu *cluster* tertentu. Microsoft Excel dan RapidMiner digunakan untuk proses pengolahan data. Metode pendekatan menggunakan teknik data mining *Knowledge Discovery in Database Process* (KDD) dengan langkah dalam prosesnya adalah *data selection*, *preprocessing*, transformasi, data mining dan evaluasi. Hasil perbandingan antara *K-Means* dan *K-Medoids* menunjukkan bahwa nilai DBI pada *K-Means* sebesar 0,368 mendapatkan 3 *cluster* sedangkan nilai DBI *K-Medoids* sebesar 0,706 mendapatkan 8 *cluster*. Dapat disimpulkan nilai optimal *K-Means* menghasilkan 0,368 dengan 3 *cluster*.

Kata kunci : Hasil pertanian, *Clustering*, *K-Means*, *K-Medoids* dan *Knowledge Discovery in Database Process* (KDD)

1. PENDAHULUAN

Manusia membutuhkan sumber energi dalam setiap aktivitasnya, mulai dari bekerja, berolahraga, bahkan kebutuhan yang paling mendasar yang harus dipenuhi dalam setiap saat adalah pangan [1]. Pangan merupakan kebutuhan yang sangat penting untuk diperhatikan dalam kehidupan banyak orang [2]. Sebagian besar penduduk Indonesia bekerja di sektor pertanian karena Indonesia adalah negara agraris [1]. Sektor pertanian menjadi salah satu kekuatan untuk membangun perekonomian suatu wilayah. Salah satu yang memberikan kontribusi terhadap perekonomian daerah di Kabupaten Cirebon adalah produksi pertanian. Subsektor yang memberikan kontribusi dalam pertanian mencakup hortikultura, peternakan, tanaman pangan, perikanan, dan perkebunan. Terdapat 7 jenis tanaman pangan yang dihasilkan di wilayah ini yaitu padi, kacang tanah, jagung, kacang hijau, ubi jalar, kedelai dan ubi kayu.

Permasalahan yang terjadi di Dinas Pertanian Kabupaten Cirebon ini ketersediaan dataset untuk pengelompokan hasil pertanian masih belum memadai. Keterbatasan dataset ini membuat tidak adanya informasi mengenai perbedaan dan keunggulan hasil pertanian antar kecamatan. Oleh karena itu, perlu adanya pengolahan dataset untuk pengelompokan terhadap hasil pertanian setiap kecamatan, sehingga dapat menghasilkan pemetaan dari hasil pertanian kedalam suatu *cluster* tertentu yang dapat membedakan kecamatan-kecamatan dengan keunggulan hasil produksi pertanian

berdasarkan jumlah *cluster* yang optimal. Penggunaan metode pengelompokan dengan penerapan algoritma *K-Means* dan *K-Medoids* dapat digunakan sebagai pemecah masalah, sehingga mempermudah mencari informasi tentang kelompok kecamatan yang menghasilkan keunggulan hasil produksi pertanian.

Penelitian yang pernah dilakukan dengan kedua algoritma yaitu perbandingan metode *clustering* yaitu algoritma *K-Means* dan *K-Medoids* dalam menentukan hasil *cluster*, pernah dilakukan oleh [3] dengan judul “Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Pengelompokan Daerah Produksi Kakao” tujuan pada penelitian ini melakukan proses pengelompokan terhadap wilayah produksi kakao di Provinsi Sulawesi Selatan, perbandingan kinerja dua algoritma yaitu *K-Means* dan *K-Medoids* dilakukan dengan satu dataset. Hasil dari analisis menunjukkan bahwa, berdasarkan perbandingan nilai DBI menggunakan RapidMiner diperoleh algoritma *K-Means* menghasilkan nilai DBI 0,292 dan *K-Medoids* menghasilkan nilai DBI 0,365 dengan masing-masing nilai *k* adalah 4. Kemudian penelitian ini menyimpulkan bahwa nilai algoritma *K-Means* lebih efektif dibandingkan dengan algoritma *K-Medoids*.

Metode *clustering* adalah metode analisis atau data mining yang mengimplementasikan pendekatan *unsupervised* untuk pemodelan data. Metode ini termasuk dalam kategori pengelompokan data yang menggunakan sistem partisi. Ada beberapa contoh algoritma *clustering* adalah *K-Means*, *Hierarchical Clustering* dan *Density-Based Spatial Clustering of*

Applications with Noise, K-Medoids. Pengelompokan dengan algoritma *K-Means* dan *K-Medoids* dipilih karena adanya perbedaan cara kerja kedua metode *clustering* tersebut. *K-Means* mengelompokkan data dengan memisahkan data kedalam satu atau lebih kelompok atau *cluster* berdasarkan kesamaan karakteristiknya, sedangkan *K-Medoids* menggolongkan indeks informasi dari setiap objek data kedalam kumpulan-kumpulan yang didefinisikan dengan jelas. Algoritma *K-Medoids* memilih objek sebagai representasi pusat *cluster* untuk setiap *cluster*, sedangkan algoritma *K-Means* menggunakan nilai rata-rata sebagai pusat *cluster* [4].

2. TINJAUAN PUSTAKA

Dalam paper ini membahas mengenai perbandingan dan implementasi metode *clustering* menggunakan algoritma *k-means* dan *k-medoids*. Paper (1) [4] dengan judul “Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Pengelompokan Program BPJS Ketenagakerjaan”. Tujuan penelitian ini adalah untuk membandingkan dua algoritma *clustering* berdasarkan karakteristik bukan penerima upah yang dimiliki. Dalam rangka mempermudah proses pendaftaran bagi peserta baru ke dalam program BPJS Ketenagakerjaan, dalam penelitian ini akan mengelompokkan peserta menggunakan metode *clustering* guna memberikan rekomendasi kepada peserta baru yang akan mendaftar. Hasil penelitian menunjukan bahwa *K-Medoids* menghasilkan kelompok yang lebih kuat dan stabil daripada *K-Means*.

Paper (2) [5] dengan judul “Evaluasi Clustering *K-Means* dan *K-Medoid* Pada Persebaran Covid-19 di Indonesia dengan Metode Davies-Bouldin Index (DBI)”. Pengujian komparasi kedua algoritma dibuktikan melalui hasil nilai Devies Bouldin Index (DBI) sebagai parameter evaluasi. Simpulan dari penelitian ini *K-Means* menghasilkan nilai 0,976 sedangkan *K-Medoids* menghasilkan nilai 0,98. Jadi algoritma yang memiliki nilai DBI paling rendah dan dapat disebut menghasilkan klaster terbaik dalam klaster data sebaran Covid-19 di Indonesia adalah *K-Means*.

Paper (3) [6] dengan judul Implementasi *K-Means* dan *K-Medoids* dalam “Pengelompokan Wilayah Potensial Produksi Daging Ayam” penelitian ini, mengimplementasikan algoritma *K-Means* dan *K-Medoids* menggunakan pengukuran *Davies Bouldin Index* (DBI). Hasilnya menunjukan bahwa algoritma *K-Means* lebih optimal dengan pengelompokan menjadi 5 *cluster* dapat mengelompokkan wilayah potensial produksi daging ayam di Provinsi Jawa Barat. Penilaian kinerja dilakukan menggunakan *Davies Bouldin Index* (DBI) untuk menentukan algoritma paling optimal diantara penggunaan *K-Means* dan *K-Medoids*.

Paper (7) [7] dengan judul “Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk Pengelompokkan Daerah Rawan Tanah Longsor di

Provinsi Jawa Barat”. Penelitian ini membahas tentang Badan Penanggulangan Bencana Daerah (BPBD) yang belum dapat mengoptimalkan layanan kepada korban bencana, ini terkait dengan lamanya waktu yang dibutuhkan untuk mendapatkan bantuan karena kekurangan peralatan dan makanan di lokasi bencana. Analisa dan hasil yang ditemukan pada data kejadian tanah longsor, *K-Means* lebih baik dibandingkan menggunakan *K-Medoids*.

Paper (11) [1] “Analisis *Cluster* Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma *K-Means*”. Dalam penelitian ini, membahas masalah yang ada di sektor pertanian. Hasilnya menunjukkan bahwa tiga provinsi (Jawa Barat, Jawa Tengah, dan Jawa Timur) memiliki tingkat produksi bahan pangan yang tinggi, sedangkan empat provinsi memiliki tingkat produksi sedang, dan tingkat produksi bahan pangan yang rendah sebanyak 27 provinsi.

Dalam paper ini membahas mengenai metode *clustering* menggunakan algoritma *K-Medoids*. Paper (12) [8] “Penerapan *K-Medoids* dalam Klasifikasi Persebaran Lahan Kritis di Jawa Barat Berdasarkan Kabupaten”. Permasalahan yang ditemukan dalam penelitian ini adalah adanya lahan kritis yang membutuhkan rehabilitasi dan peningkatan produktivitas untuk kembali ke kondisi sebelumnya sebagai ekosistem yang sehat atau untuk mencapai hasil yang lebih optimal. Dari hasil *clustering* ditemukan dalam penelitian ini diperoleh 5 *cluster* terbaik yaitu 81 wilayah pada C 0, 48 wilayah pada C 1, 229 wilayah pada C 2, 14 wilayah pada C 3, 6 wilayah pada C 4.

Dalam paper ini membahas mengenai beberapa potensi dalam mendukung peningkatan ekonomi daerah. Paper (14) [9] “Potensi Sektor Pertanian Dalam Prioritas Pembangunan Ekonomi Kabupaten Belitung”. Dengan memanfaatkan data PDRB berdasarkan dasar harga konstan, penelitian ini bertujuan untuk mengidentifikaso potensi peran kontribusi sektor pertanian dalam upaya pembangunan ekonomi di Kabupaten Belitung. Menurut analisis penelitian ini, Sektor Pertanian, Perikanan dan Kehutanan dapat menjadi prioritas dalam upaya pembangunan ekonomi di Kabupaten Belitung.

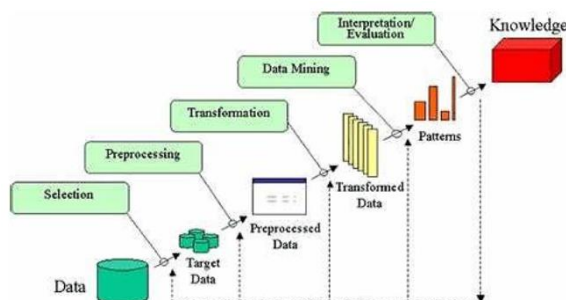
Dalam paper ini membahas mengenai tahapan *Literature Review*. Paper (15) [10] “Legalisasi Dokumen Elektronik Menggunakan Tanda Tangan Digital sebagai Alternatif Pengesahan Dokumen di Masa Pandemi”. Metode dalam *literature review* yang diterapkan dalam penelitian ini adalah dengan mereview kumpulan artikel yang berhubungan dengan Perbandingan Algoritma *K-Means* Dan *K-Medoids* Untuk Pengelompokan Hasil Pertanian di Kabupaten Cirebon. Ada 4 tahapan yang digunakan dalam mereview kumpulan artikel. Tahap pertama formulasi permasalahan Tahap kedua melibatkan pencarian *literature* yang relevan dengan topik penelitian. Ketiga, evaluasi data dilakukan dengan menyaring dan memilih artikel yang paling baru dalam 5 tahun

terakhir. Tahap keempat analisis dan interpretasi adalah tahapan pelaksanaan *literature review* dengan menggunakan teknik yaitu mencari kesamaan (*compare*), mencari ketidaksamaan (*contrast*), memberikan pandangan (*criticize*), membandingkan (*synthesize*), dan meringkas (*summarize*) beberapa penelitian yang relevan. Tujuannya untuk menyusun pemahaman dengan mempertimbangkan temuan-temuan dari berbagai penelitian terkait.

3. METODE PENELITIAN

3.1. Metode Penelitian

Pada penelitian ini metode yang diterapkan adalah metode *Knowledge Discovery in Database Process* (KDD).



Gambar 1. Tahapan metode penelitian [8]

3.2. Sumber Data

Data penelitian ini yang dikumpulkan langsung dari Dinas Pertanian Kabupaten Cirebon pada bulan Oktober 2023. Dataset yang diperoleh mencakup nama kecamatan, jenis tanaman pangan dan hasil produksi selama 5 tahun dari tahun 2018-2022.

3.3. Teknik Pengumpulan Data

Dalam penelitian diperoleh melalui serangkaian tahapan, dimulai dari tahapan observasi, wawancara, studi literatur dan dokumentasi. Dari proses tersebut diperoleh dataset hasil pertanian selama 5 tahun dari tahun 2018-2022. Data produksi pertanian yang digunakan berdasarkan dari 40 Kecamatan di Kabupaten Cirebon.

3.4. Teknik Analisis Data

Penerapan metode *clustering* menggunakan algoritma *K-Means* dan *K-Medoids*. Algoritma yang sudah di implementasikan dalam RapidMiner akan menghasilkan nilai *Davies Bouldin Index* (DBI).

1) Algoritma *K-Means*

Pengelompokan data dengan *K-Means* melibatkan proses sebagai berikut:

- Menentukan jumlah *cluster*
- Pusat awal atau *centroid* (pusat klaster awal) ditentukan secara acak
- Menaksirkan jarak dari tiap *centroid* dengan menggunakan formula *Euclidean distance*

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (1)$$

- Menghitung *centroid* baru dengan kedudukan baru dengan menggunakan metode pencarian objek rata-rata pada *cluster*
- Sampai hasil pengelompokan tidak berubah ulangi langkah sebelumnya [3].

2) Algoritma *K-Medoids*

Formula *K-Medoids* sebagai berikut:

- Menetapkan jumlah *cluster* (*k*)
- Dengan menggunakan rumus persamaan, persamaan ukuran jarak *Euclidian Distance* digunakan untuk mengalokasikan setiap objek atau data ke *cluster* terdekat

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (2)$$

- Pilih kandidat *medoid* baru dari masing-masing *cluster* secara acak.
- Menghitung menghitung nilai total jarak baru – total jarak lama untuk menghitung total simpangan (*S*). Untuk membuat sekumpulan *k* objek baru sebagai *medoid*, tukar objek dengan *n* data *cluster* jika didapatkan $S < 0$.
- Ulangi langkah 3 sampai dengan 5 hingga tidak ada perubahan *medoid*. Dengan begitu *cluster* dan anggota masing-masing dapat diperoleh [4].

4. HASIL DAN PEMBAHASAN

Pengujian dilakukan delapan kali percobaan untuk mengevaluasi nilai DBI algoritma *K-Means* dan *K-Medoids* menggunakan RapidMiner. Hasil dari analisis ditemukannya nilai DBI dari hasil pengujian data, menunjukkan bahwa penggunaan algoritma *K-Means* dengan 3 klaster dapat secara optimal mengelompokkan hasil pertanian, dengan hasil nilai DBI mencapai 0,368. Sementara, dalam pengelompokan menggunakan algoritma *K-Medoids*, didapat nilai DBI sebesar 0,706 dengan jumlah 8 *cluster*.

4.1. Data Hasil Pertanian

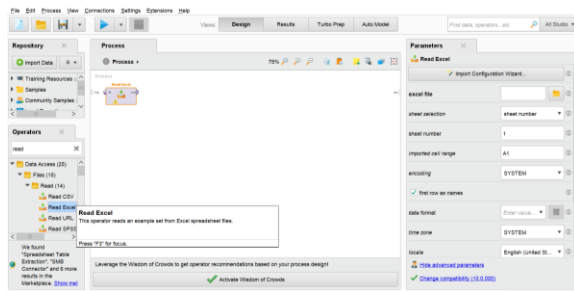
Data yang diperlukan untuk penelitian ini adalah data hasil pertanian, data ini diperoleh langsung dari Dinas Pertanian Kabupaten Cirebon. Dataset ini mencakup nama 40 Kecamatan di Kabupaten Cirebon, terdiri dari 7 jenis tanaman pangan dengan hasil produksi selama 5 tahun dari tahun 2018-2022.

4.2. Data Selection

Data yang sudah diperoleh tidak langsung diolah. Agar dapat diolah menggunakan metode *clustering* diperlukan proses pemilihan data terhadap atribut data. Atribut data yang digunakan terdiri dari nama kecamatan, hasil produksi jagung, kacang hijau, kacang tanah, padi, ubi jalar, ubi kayu dan kedelai.

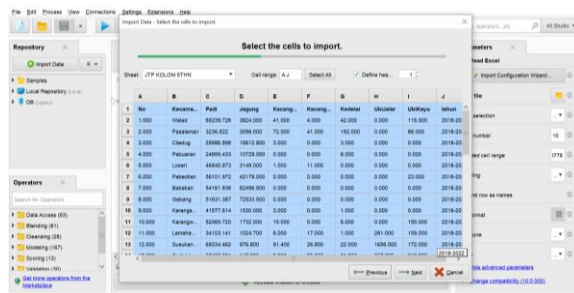
Berikut penerapan data *selection* pada RapidMiner:

- Langkah awal dalam RapidMiner adalah membaca dataset dari file excel, dengan membuka data yang akan diolah adalah tahap penting dalam melakukan analisis. Lakukan penginputan data menggunakan operator *read excel*



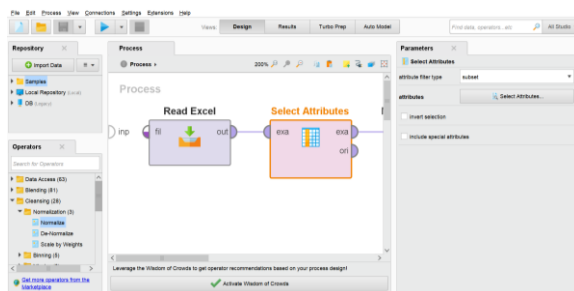
Gambar 2. Cara mengimpor data
[Sumber: Mala, 2023]

Setelah membuka lokasi file, pada tahap ini memilih sel yang akan digunakan dalam proses pengujian

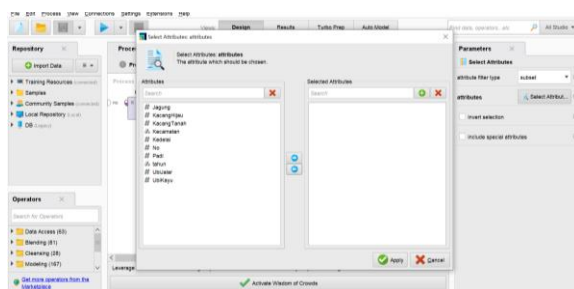


Gambar 3. Tampilan dataset hasil pertanian
[Sumber: Mala, 2023]

Dalam proses ini melakukan pemilihan atribut yang akan digunakan dalam proses pengolahan data. Operator *Select Attributes* digunakan untuk memilih atribut. Atribut yang dipilih meliputi nama kecamatan, jenis tanaman pangan. Lakukan pencarian *select attributes* pada operators

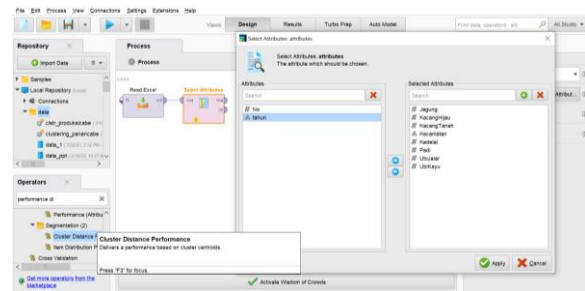


Gambar 4. Tampilan *select attributes*
[Sumber: Mala, 2023]



Gambar 5. Tampilan Atribut yang belum dipindahkan ke *select attributes*
[Sumber: Mala, 2023]

Pada *attributes* klik 2 kali untuk memindahkan atribut yang akan diolah



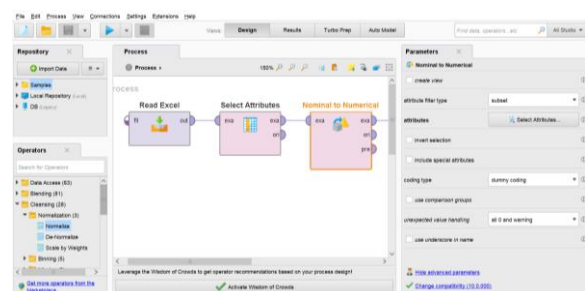
Gambar 6. Tampilan atribut data yang sudah dipilih
[Sumber: Mala, 2023]

4.3. Data Preprocessing

Dalam proses ini, atribut yang tidak dibutuhkan atau tidak diseleksi adalah no dan tahun pada dataset. Hal ini dilakukan agar atribut tersebut tidak mengganggu proses data berlangsung.

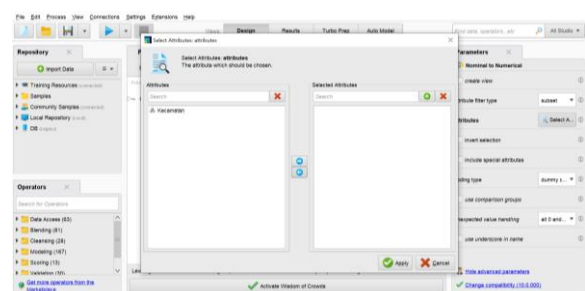
4.4. Transformasi

Tahap ini akan mengubah atribut nama kecamatan yang termasuk dalam nominal ke bentuk numerik dengan menggunakan operator *Nominal To Numerical*. Pada tahap ini, data diubah ke dalam format yang sesuai dengan algoritma *clustering* yang akan digunakan.

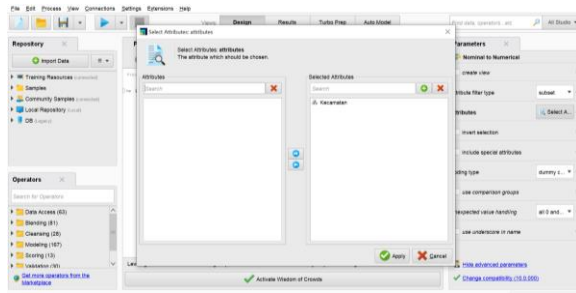


Gambar 7. Tampilan operator *nominal to numerical*
[Sumber: Mala, 2023]

Pada parameters *Nominal To Numerical*, ubah *attribute filter type* menjadi subset agar bisa memilih atribut nominal yang akan dijadikan numerik



Gambar 8. Tampilan sebelum atribut diseleksi
[Sumber: Mala, 2023]



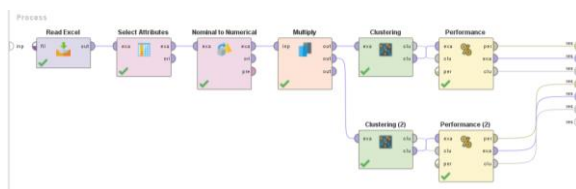
Gambar 9. Tampilan sesudah *selected attribute*
[Sumber: Mala, 2023]

4.5. Data Mining

Pada tahap ini, terdapat dua penggunaan metode *clustering* yaitu menerapkan algoritma *K-Means* dan *K-Medoids*. Proses pada tahapan dengan penggunaan operator tertentu dilakukan langkah sebagai berikut:

1. Pilih *import* data menggunakan operator *Read Excel* untuk menginput data dari excel yang akan diolah ke RapidMiner
2. Operator *Select Attributes* digunakan untuk memilih atribut tertentu dan atribut yang tidak dibutuhkan harus dihilangkan agar tidak mengganggu ketika proses data berlangsung. Atribut yang dipilih meliputi nama kecamatan, jagung, ubi jalar, kacang hijau, padi, kedelai, kacang tanah, dan ubi kayu
3. Dataset hasil pertanian ini memiliki atribut nama kecamatan yang bersifat nominal, dalam hal ini operator *Nominal To Numerical* digunakan untuk mengonversi atribut nominal menjadi bentuk numerik agar dapat digunakan dalam analisis *clustering*.
4. Penelitian ini menggunakan dua algoritma untuk melakukan *clustering*. Oleh karena itu, penggunaan operator *Multiply* dipilih untuk membagi data yang sama, sehingga tidak menambah operator yang sama.
5. Penggunaan operator *K-Means* dan *K-Medoids* dalam pengujian dengan 8 kali percobaan.

Untuk mengevaluasi hasil algoritma *clustering*, digunakan operator *Performance Distance*. Setiap operator *K-Means* dan *K-Medoids* dievaluasi dengan mengukur *performance* jarak antar titik dalam *cluster*.

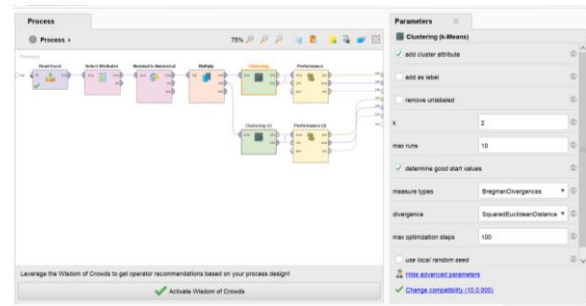


Gambar 10. Tampilan penerapan metode *clustering* menggunakan rapidminer
[Sumber: Mala, 2023]

4.6. Evaluasi

Dalam tahap evaluasi ini, penggunaan operator *Cluster Distance Performance*, dilakukan pengujian mencari hasil terbaik dari 2,3,4,5,6,7,8 *cluster* yang

diuji. Dalam operator *Cluster Distance Performance* yang digunakan untuk menilai hasil DBI dari setiap *cluster* yang diuji. Berikut tampilan hasil pengujian:



Gambar 11. Algoritma *k-means* dan *k-medoids* dengan k 2
[Sumber: Mala, 2023]

PerformanceVector

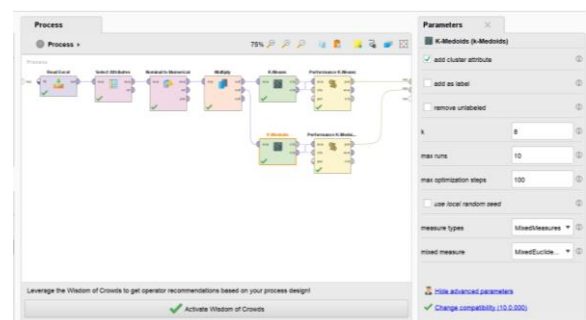
```
PerformanceVector:
Avg. within centroid distance: 1308655751.550
Avg. within centroid distance_cluster_0: 1126573758.145
Avg. within centroid distance_cluster_1: 3554333670.222
Davies Bouldin: 0.397
```

Gambar 12. Hasil DBI algoritma *k-means* dengan k 2
[Sumber: Mala, 2023]

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 2172175829.475
Avg. within centroid distance_cluster_0: 10566245389.400
Avg. within centroid distance_cluster_1: 973023035.200
Davies Bouldin: 1.255
```

Gambar 13. Hasil DBI algoritma *k-medoids* dengan k 2
[Sumber: Mala, 2023]



Gambar 14. Algoritma *k-means* dan *k-medoids* dengan k 8
[Sumber: Mala, 2023]

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 112654664.805
Avg. within centroid distance_cluster_0: 108851775.557
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 0.000
Avg. within centroid distance_cluster_3: 178104951.250
Avg. within centroid distance_cluster_4: 98088073.802
Avg. within centroid distance_cluster_5: 162781515.333
Avg. within centroid distance_cluster_6: 154726549.167
Avg. within centroid distance_cluster_7: 0.000
Davies Bouldin: 0.380
```

Gambar 15. Hasil DBI algoritma *k-means* dengan k 8
[Sumber: Mala, 2023]

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 523258683.100
Avg. within centroid distance_cluster_0: 0.000
Avg. within centroid distance_cluster_1: 1187000249.333
Avg. within centroid distance_cluster_2: 0.000
Avg. within centroid distance_cluster_3: 4483798668.500
Avg. within centroid distance_cluster_4: 252187766.667
Avg. within centroid distance_cluster_5: 53571135.500
Avg. within centroid distance_cluster_6: 299608645.100
Avg. within centroid distance_cluster_7: 16372083.000
Davies Bouldin: 0.706
```

Gambar 16. Hasil DBI algoritma *k-medoids* dengan k 8
[Sumber: Mala, 2023]

Pada tabel 1 menampilkan hasil perbandingan *K-Means* dan *K-Medoids*, disertakan keterangan jumlah *cluster* dan nilai DBI yang dihasilkan oleh algoritma *K-Means* dan *K-Medoids*. Untuk mengukur seberapa baik pengelompokan yang diperoleh oleh masing-masing algoritma, maka nilai DBI digunakan sebagai evaluasi. Berikut rincian dari hasil perbandingan:

Tabel 1. Hasil Perbandingan DBI Metode *Clustering*

Jumlah Cluster	DBI Algoritma	
	<i>K-Means</i>	<i>K-Medoids</i>
2	0.397	1.255
3	0.368	1.031
4	0.470	2.035
5	0.409	1.218
6	0.442	0.839
7	0.469	0.942
8	0.380	0.706

[Sumber: Mala, 2023]

Berdasarkan hasil dari Tabel 2, ditemukan nilai DBI sebesar 0,368 menggunakan algoritma *K-Means*, dengan jumlah 3 *cluster* menunjukkan bahwa nilai tersebut adalah yang terbaik untuk pengelompokan, karena nilai DBI nya lebih kecil diantara hasil DBI *K-Means* lain. Gambar 4.16 menunjukkan adanya 3 *clusters* dengan 35 anggota pada kelompok C0, 2 anggota pada kelompok C1, dan 3 anggota pada kelompok C2.

Cluster Model

```
Cluster 0: 35 items
Cluster 1: 2 items
Cluster 2: 3 items
Total number of items: 40
```

Gambar 17. Jumlah *cluster* model *k-means*
[Sumber: Mala, 2023]

Cluster Model

```
Cluster 0: 1 items
Cluster 1: 3 items
Cluster 2: 1 items
Cluster 3: 2 items
Cluster 4: 9 items
Cluster 5: 2 items
Cluster 6: 20 items
Cluster 7: 2 items
Total number of items: 40
```

Gambar 18. Jumlah *cluster* model *k-medoids*
[Sumber: Mala, 2023]

Pada Gambar 19 menunjukkan hasil pengelompokan dari masing-masing *cluster*:

Kecamatan	Padi	Jagung	KacangHijau	KacangTanah	Kedelai	UbiJalar	UbiKayu	Cluster
Waled	65239	3824	41	4	42	0	115	cluster_0
Pasaleman	3236	3098	72	41	192	0	86	cluster_0
Ciledug	28986	18612	3	0	3	0	0	cluster_0
Pabuaran	24855	10729	0	0	8	0	0	cluster_0
Losari	46840	3149	1	11	0	0	0	cluster_0
Pabedilan	56101	42179	0	0	0	0	23	cluster_0
Babakan	54161	62496	0	0	0	0	0	cluster_0
Gebang	51631	72533	0	0	0	0	0	cluster_0
Karangsembung	41577	1530	3	0	1	0	0	cluster_0
Karangwates	52065	1732	15	0	6	0	150	cluster_0
Lemahabang	34103	1024	6	17	1	281	159	cluster_0
Susukan Lebak	69334	976	91	26	22	1656	172	cluster_0
Sedong	87487	147	0	26	31	957	342	cluster_0
Astanajapura	42849	32	0	171	0	0	0	cluster_0
Pangenan	38349	5940	0	0	0	0	0	cluster_0
Mundu	49024	0	2	0	5	18	0	cluster_0
Beber	53974	220	0	22	14	29414	604	cluster_0
Greged	49992	65	0	51	0	480	0	cluster_0
Tahun	42709	96	5	31	8	1822	705	cluster_0
Sumber	46807	17	0	24	0	498	563	cluster_0
Dukupuntang	97711	1204	0	0	0	18	43	cluster_0
Palimanan	76523	0	1159	0	0	0	0	cluster_0
Phumbon	47119	81	51	1	0	24	181	cluster_0
Depok	46432	18	0	1	0	0	20	cluster_0
Weru	13487	156	26	0	0	0	0	cluster_0
Plered	38636	81	1	0	14	0	0	cluster_0
Tengah tani	28367	182	25	0	13	80	0	cluster_0
Kedavung	6888	0	0	0	0	0	133	cluster_0
Gusung jati	44986	0	0	0	0	0	0	cluster_0
Kapetakan	195561	0	0	0	26	0	0	cluster_2
Suranenggala	85355	0	0	0	0	0	0	cluster_0
Klangenan	97657	89	187	0	0	0	0	cluster_0
Jambiang	80437	20	5	0	0	23	0	cluster_0
Arjasmangun	91484	0	329	0	61	0	0	cluster_0
Panguragan	115498	0	0	0	21	0	0	cluster_0
Ciwaringin	72094	0	3113	0	148	0	0	cluster_0
Gempol	60801	659	1779	17	123	0	54	cluster_0
Susukan	244495	44	2451	0	64	0	0	cluster_2
Gegesik	339160	52	12	2	0	223	364	cluster_1
Kaliwedi	142189	0	0	0	0	59	80	cluster_2

Gambar 19. Hasil pengelompokan
[Sumber: Mala, 2023]

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis terhadap data hasil pertanian di Kabupaten Cirebon menggunakan algoritma *clustering* dengan membandingkan kedua algoritma, maka dapat ditarik kesimpulan sebagai berikut: Diterapkannya teknik data mining dengan metode algoritma *K-Means* dan *K-Medoids* untuk pengelompokan hasil pertanian, dengan tahapan pertama mengumpulkan data hasil pertanian secara langsung pada Dinas Pertanian Kabupaten Cirebon. Teridentifikasinya hasil pengelompokan dengan algoritma *K-Means* dan *K-Medoids*, dengan pengujian 8 kali percobaan pada masing-masing algoritma untuk mendapatkan nilai DBI terkecil. Hasilnya menunjukkan bahwa nilai DBI pada algoritma *K-Means* menghasilkan nilai DBI 0,368, dengan 3 *cluster* hasil dengan C0 memiliki 36 anggota, C1 memiliki 1 anggota dan C2 memiliki 4 anggota. Sedangkan nilai DBI *K-Medoids* sebesar 0,706 dengan jumlah 8 *cluster*. C0 memiliki 1 anggota, C1 memiliki 3 anggota, C2 memiliki 1 anggota, C3 memiliki 2 anggota, C4 memiliki 9 anggota, C5 memiliki 2 anggota, C6 memiliki 20 anggota, dan C7 memiliki 2 anggota. Diketahuinya jumlah *cluster* dari hasil evaluasi *performance* perbandingan antara algoritma *K-Means* dan *K-Medoids*. Dapat disimpulkan bahwa nilai DBI pada algoritma *K-Means* menghasilkan nilai DBI 0,368, dengan 3 *cluster* memberikan hasil pengelompokan yang optimal. Informasi ini dapat menjadi dasar bagi para pemangku kebijakan untuk pengambilan keputusan terkait sektor pertanian di Kabupaten Cirebon.

Penelitian tentang data mining menggunakan dataset hasil produksi pertanian dengan algoritma *clustering* dapat dilanjutkan dengan mempertimbangkan penggunaan operator yang digunakan pada RapidMiner. Agar mengoptimalkan proses rancangan RapidMiner penelitian selanjutnya disarankan untuk melakukan tahapan data *selection*, *preprocessing* dan transformasi pada parameter *read excel* yaitu *Import Configuration Wizard*, sehingga operator *Select Atribut* dan *Nominal To Numerical* tidak perlu digunakan. Kemudian agar menghemat waktu *run* pada RapidMiner maka disarankan untuk menggunakan operator *Optimize Parameters*.

DAFTAR PUSTAKA

- [1] T. Tendean and W. Purba, "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan

Pangan Menggunakan Algoritma K-Means," *J. Sains dan Teknol.*, vol. 1, no. 2, pp. 5–11, 2020.

- [2] L. Fimawahib, I. R. Bakti, and A. Supriyanto, "Algoritma K-Medoids untuk Pengelompokan Produksi Padi dan Beras sebagai Upaya Optimalisasi Ketahanan Pangan di Provinsi Riau," *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 13–24, 2022, doi: 10.33372/stn.v8i2.877.
- [3] N. A. S. Z. Abidin, R. D. Avila, A. Hermatyar, and Rismayani, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Produksi Kakao," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, pp. 383–391, 2022.
- [4] A. Meiriza, E. Ali, Rahmiati, and Agustin, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan," *Indones. J. Comput. Sci.*, 2023.
- [5] Fathurrahman, S. Harini, and R. Kusumawati, "Evaluasi Clustering K-Means dan K-Medoid Pada Persebaran Covid-19 di Indonesia dengan Metode Davies-Bouldin Index (DBI)," *J. Mnemon.*, vol. 6, no. 2, pp. 117–128, 2023.
- [6] M. K. Hidayat and R. Fitriana, "Implementasi K-Means dan K-Medoids dalam Pengelompokan Wilayah Potensial Produksi Daging Ayam," *J. Teknol. Ind. Pertan.*, vol. 32, no. 158, pp. 239–247, 2022.
- [7] M. Herviany, S. P. Delima, T. Nurhidayah, and Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. April, pp. 34–40, 2021.
- [8] L. Karlina and O. Nurdian, "Penerapan K-Medoids dalam Klasifikasi Persebaran Lahan Kritis di Jawa Barat Berdasarkan Kabupaten/Kota," 2023. [Online]. Available: <https://opendata.jabarprov.go.id/id>.
- [9] G. S. Tanjung, M. A. Pujiyanto, L. Farichah, and E. F. Anggraini, "Potensi Sektor Pertanian dalam Prioritas Pembangunan Ekonomi Kabupaten Belitung," vol. 19, no. 2016, pp. 110–123, 2022.
- [10] T. Yuniati and M. F. Sidiq, "Literature Review: Legalisasi Dokumen Elektronik Menggunakan Tanda Tangan Digital sebagai Alternatif Pengesahan Dokumen di Masa Pandemi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 6, 2020, doi: 10.29207/resti.v4i6.2502.