

EVALUASI KINERJA K-MEDOIDS CLUSTERING MODEL UNTUK KLASTERISASI DAERAH PRODUKTIVITAS PANEN PADI DI KABUPATEN BIREUEN

Hayatun Nisa¹⁾, Muhammad Daud²⁾, Sujacka Retno³⁾

^{1,3}Fakultas Teknik, Teknik Informatika, Universitas Malikussaleh

²Fakultas Teknik, Teknik Elektro, Universitas Malikussaleh

Jl. Kampus Unimal Bukit Indah, Blang Pulo, Kec. Muara Satu, Kota Lhokseumawe, Aceh 24355

email:¹nisahayatun03022002@gmail.com, ²mdaud@unimal.ac.id, ³sujacka@unimal.ac.id

Abstract

Rice productivity is a crucial indicator for monitoring and improving rice production in a region. In Bireuen District, Aceh, the uneven allocation of subsidized fertilizers has hindered efforts to optimize rice productivity. This study aims to cluster priority and non-priority rice productivity areas in Bireuen District using the K-Medoids and Purity K-Medoids algorithms. The dataset consists of 204 historical rice farming records from 2012 to 2023, sourced from the Bireuen District Central Bureau of Statistics and the Department of Agriculture and Plantations, including variables such as number of villages, planting area, harvest area, productivity, production volume, and planting area percentage. Clustering performance was evaluated using the Davies Bouldin Index (DBI). The results show that Purity K-Medoids achieved an average DBI value of 0.786911, lower than K-Medoids at 0.907856, indicating better clustering validity. Based on the clustering results, Peusangan Subdistrict was identified most frequently as a priority area, appearing in 2013, 2015, 2016, 2018, 2019, 2020, 2022, and 2023. These findings are expected to inform agricultural resource allocation policies in Bireuen District.

Keywords: Rice Productivity, Clustering, K-Medoids, Purity K-Medoids, Davies Bouldin Index

Abstrak

Produktivitas padi merupakan indikator penting dalam memantau dan meningkatkan produksi padi di suatu wilayah. Di Kabupaten Bireuen, Aceh, alokasi pupuk subsidi yang tidak merata menjadi kendala dalam optimalisasi produktivitas padi. Penelitian ini bertujuan untuk mengklaster daerah prioritas dan non-prioritas produktivitas panen padi di Kabupaten Bireuen menggunakan algoritma K-Medoids dan Purity K-Medoids. Data yang digunakan adalah data historis pertanian padi tahun 2012–2023 sebanyak 204 record, yang diperoleh dari Badan Pusat Statistik Kabupaten Bireuen dan Dinas Pertanian dan Perkebunan Kabupaten Bireuen, dengan variabel seperti jumlah desa, luas tanam, luas panen, produktivitas, jumlah produksi, dan persentase luas tanam. Proses klasterisasi dievaluasi menggunakan Davies Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa Purity K-Medoids menghasilkan nilai rata-rata DBI sebesar 0,786911, lebih rendah dibandingkan K-Medoids yang sebesar 0,907856, menandakan validitas klaster yang lebih baik. Berdasarkan hasil klasterisasi, Kecamatan Peusangan paling sering muncul sebagai daerah prioritas, yaitu pada tahun 2013, 2015, 2016, 2018, 2019, 2020, 2022, dan 2023. Temuan ini diharapkan dapat menjadi pertimbangan dalam kebijakan alokasi sumber daya pertanian di Kabupaten Bireuen.

Kata Kunci: Produktivitas Padi, Klasterisasi, K-Medoids, Purity K-Medoids, Davies Bouldin Index

1. PENDAHULUAN

Pertanian merupakan sektor vital bagi Indonesia, mengingat mayoritas penduduknya menggantungkan hidup pada aktivitas ini. Optimalisasi sektor pertanian menjadi kunci dalam pemanfaatan sumber daya alam dan geografi Indonesia, serta berperan penting dalam penyediaan pangan, bahan baku industri, dan ekspor (Lailiyah et al., 2020). Salah satu komoditas utama adalah padi, di mana produktivitasnya menjadi indikator penting untuk memantau dan meningkatkan produksi di berbagai wilayah. Pemerintah telah menyediakan subsidi pupuk sebagai upaya mendukung petani dalam meningkatkan produktivitas padi dan menekan biaya produksi. Namun, alokasi pupuk subsidi yang tidak

tepat sasaran masih menjadi kendala utama, sehingga menghambat peningkatan produktivitas padi di berbagai daerah, termasuk Kabupaten Bireuen, Aceh.

Kabupaten Bireuen merupakan wilayah otonom di Provinsi Aceh yang terbentuk pada tahun 1999, terdiri dari 17 kecamatan dan 609 desa. Wilayah ini memiliki peran strategis dalam produksi padi di Aceh. Namun, distribusi pupuk subsidi yang tidak merata di beberapa kecamatan menyebabkan ketimpangan hasil panen. Permasalahan ini menuntut adanya pendekatan berbasis data untuk mengidentifikasi daerah prioritas yang membutuhkan intervensi lebih lanjut.

Ada berbagai metode yang bisa digunakan untuk pengelompokan data (*clustering data*), diantaranya *Single Linkage*, *Fuzzy C-Means*, *LVQ*, *K-Means* dan *K-Medoids*. Metode yang digunakan dalam penelitian ini ada dua yaitu *Purity K-Medoids* dan *K-Medoids*. *K-Medoids* merupakan suatu algoritma pengelompokan yang *non-hierarchical* yang merupakan varian dari algoritma *K-Means*. Dengan menggunakan pendekatan partisi, algoritma ini mengelompokkan x objek ke dalam k cluster. Dalam setiap cluster, objek yang paling representatif disebut sebagai *medoids*. Proses pembentukan cluster dilakukan dengan mengukur jarak antara *medoids* dan objek *non-medoids* (Dinata et al., 2021). Sedangkan *Purity* sendiri digunakan untuk mengevaluasi tingkat kemurnian suatu cluster berdasarkan kesesuaian anggotanya dengan kelas tertentu. *Purity* dihitung dengan cara memperhitungkan kemurnian setiap cluster dan nilai *Purity* tersebut akan digunakan sebagai *initial medoids* dalam perhitungan menggunakan *K-Medoids* (Hasdyna & Retno, 2022). Kemudian akan dilakukan perbandingan metode mana yang lebih optimal dengan melakukan evaluasi hasil *clustering* menggunakan *Davies Bouldin Index* (DBI). Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan metode *Purity K-Medoids* dan *K-Medoids* dalam clustering untuk penentuan daerah produktivitas panen padi di Kabupaten Bireuen.
2. Menemukan hasil clustering terbaik antara *Purity K-Medoids* dengan *K-Medoids* menggunakan evaluasi *Davies Bouldin Index* (DBI).

Penelitian ini menggunakan data pertanian padi Kabupaten Bireuen periode 2012–2023 sebanyak 204 record, yang diperoleh dari Badan Pusat Statistik Kabupaten Bireuen dan Dinas Pertanian dan Perkebunan Kabupaten Bireuen. Variabel yang digunakan meliputi jumlah desa, luas tanam, luas panen, jumlah produksi, produktivitas, dan persentase luas tanam. Berbagai penelitian sebelumnya telah mengkaji pengelompokan wilayah produktivitas padi menggunakan berbagai metode *clustering* di berbagai daerah di Indonesia. Sebagian besar penelitian berfokus pada optimalisasi metode clustering atau pada wilayah studi yang berbeda, namun umumnya belum mengaplikasikan *Purity K-Medoids* secara spesifik pada kasus Kabupaten Bireuen.

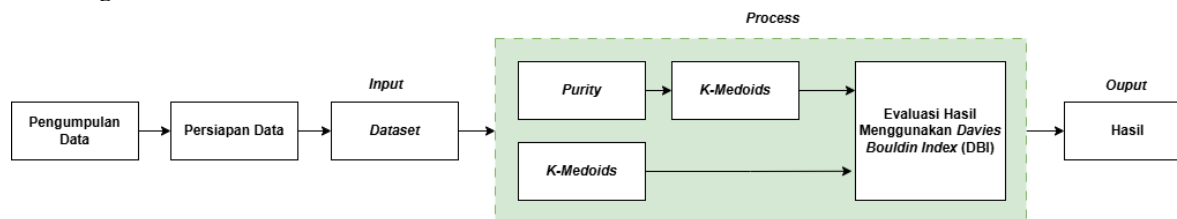
Berdasarkan uraian di atas, penelitian ini berbeda dengan riset sebelumnya dari segi metode yang digunakan, studi kasus yang diangkat, serta lokasi penelitian yang dilakukan di Kabupaten Bireuen. Penelitian ini mengaplikasikan metode *Purity K-Medoids*, yang belum banyak digunakan dalam studi serupa di Indonesia, khususnya untuk kasus produktivitas padi di Kabupaten Bireuen. Selain itu, penelitian ini menggunakan *Davies Bouldin Index* (DBI) untuk mengevaluasi validitas hasil klasterisasi secara objektif, serta melakukan analisis konsistensi kemunculan kecamatan prioritas selama periode 2012–2023. Temuan ini diharapkan dapat menjadi dasar pertimbangan dalam pengambilan kebijakan alokasi pupuk subsidi yang lebih tepat sasaran di Kabupaten Bireuen.

2. METODE PENELITIAN

Penelitian ini menggunakan algoritma *K-Medoids* dan *Purity K-Medoids*. *K-Medoids* merupakan salah satu metode yang digunakan untuk melakukan clustering data. *K-Medoids* dan *Purity K-Medoids* dalam penelitian akan digunakan untuk mengklaster daerah produktivitas Padi di Kabupaten Bireuen Provinsi Aceh untuk melihat kecamatan mana saja yang memiliki tingkat produktivitas padi tinggi dan rendah. Hasil clustering kedua metode tersebut akan dievaluasi menggunakan alat evaluasi *Davies Bouldin Index* (DBI). Hasil dari metode dengan DBI terbaik akan dijadikan sebagai analisis selanjutnya untuk mengetahui daerah mana saja yang banyak muncul tergolong ke dalam daerah produktivitas padi tinggi dalam periode 12 tahun yaitu tahun 2012-2023.

Langkah-langkah penelitian dapat dilihat pada Gambar 1. Langkah pertama yang dilakukan adalah pengumpulan data. Data yang dikumpulkan berupa data historis pertanian padi di kabupaten Bireuen dalam periode 2012-2023. Data tersebut diperoleh dari Badan Pusat Statistik Kabupaten Bireuen dan Dinas Pertanian dan Perkebunan Kabupaten Bireuen. Data dikumpulkan diinputkan ke

dalam aplikasi Microsoft Excel untuk dilakukan normalisasi data. Kemudian data tersebut dikonversi ke dalam format file .csv. *Dataset* tersebut di *input*/di *import* ke dalam program untuk dilakukan proses *clustering*.



Gambar 1. Langkah-Langkah Penelitian

2.1 DATASET

Untuk mendukung penelitian ini, berikut merupakan data asli Pertanian Padi di Kabupaten Bireuen selama periode (2012 – 2023) untuk perhitungan manual dari proses *clustering* dengan jumlah 204 *record* data dimana terdapat 17 *record* data setiap tahunnya dimana:

- (x_1) = Jumlah Desa.
- (x_2) = Luas Tanam.
- (x_3) = Luas Panen.
- (x_4) = Produktivitas.
- (x_5) = Jumlah Produksi.
- (x_6) = Persentase Luas Tanam.
- (p_i) = Data ke-*i* *purity* ke-*i*.

Berikut ini adalah data pertanian padi di Kabupaten Bireuen pada tahun 2016 sebelum dinormalisasikan yang dapat dilihat pada Tabel 1.

Tabel 1. Data Hasil Panen Padi di Kabupaten Bireuen Tahun 2016

Kecamatan	Jumlah Desa	Luas Tanam	Luas Panen	Produktivitas	Jumlah Produksi	Persentase
Samalanga	46	3046	3112	61.59	19168	102.2
Simpang, Mamplam	41	1495	2712	61.23	16605	181.4
Pandrah	19	1695	1848	59.68	11029	109.0
Jeunieb	43	2190	2782	73.90	20558	127.0
Peulimbang	22	1254	1490	61.61	9180	118.8
Peudada	52	2415	3102	59.71	18523	128.4
Juli	36	775	964	58.01	5592	124.4
Jeumpa	42	1814	2561	61.39	15723	141.2
Kota Juang	23	878	1062	63.17	6709	121.0
Kuala	20	1850	2037	65.50	13342	110.1
Jangka	46	2322	3055	62.79	19182	131.6
Peusangan	69	3067	3600	65.92	23731	117.4
Peusangan Selatan	21	971	1128	58.78	6630	116.2
Peusangan, Sb. Krueng	21	1538	1509	56.27	8491	98.1
Makmur	27	926	1569	57.64	9044	169.4
Gandapura	40	1135	1874	58.93	11043	165.1
Kuta Blang	41	2183	2756	63.66	17545	126.2

Normalisasi data adalah salah satu langkah yang dilakukan dalam tahap pra-pemrosesan data. Proses ini mengatur kembali nilai-nilai agar pemrosesan menjadi lebih sederhana. Selain itu, normalisasi menurunkan beban memori dan konsumsi selama pemrosesan (Permana & Salisah, 2022). Metode normalisasi data yang akan digunakan adalah *MinMax Normalization*.

MinMax Normalization adalah teknik normalisasi yang mengubah rentang nilai data menjadi antara 0 dan 1. Metode ini mengubah data yang kompleks tanpa menghilangkan kontennya, sehingga data tersebut lebih mudah untuk diproses (Tohendry & Jollyta, 2023). Rumus untuk menghitung MinMax Normalization dapat dilihat dalam persamaan (1) berikut (Izonin et al., 2022).

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \dots\dots\dots (1)$$

Keterangan:

- x' = Nilai hasil normalisasi.
 x_i = Nilai yang akan dinormalisasi.
 $\min(x)$ = Nilai minimal dari atribut data.
 $\max(x)$ = Nilai maksimal dari atribut data.

2.2 CLUSTERING DATA

Clustering adalah proses pengelompokan objek data ke dalam beberapa *cluster* yang terpisah dan data dalam setiap *cluster* menjadi sekelompok data yang relatif sama. *Cluster* adalah kumpulan objek data yang memiliki kemiripan karakteristik yang berada pada *cluster* yang sama dan data tersebut memiliki karakteristik yang berbeda dengan objek data pada *cluster* lainnya (Hasdyna & Retno, 2022). Tujuan dari *clustering* adalah mengelompokkan data ke dalam kelompok data yang memiliki sifat serupa, namun berbeda dari kelompok lainnya (Pratiwi & Wahyu Wibowo, 2022).

2.3 K-MEDOIDS

K-Medoids adalah algoritma *non-hierarchical clustering* yang dikembangkan dari algoritma *K-Means*. *K-Medoids* membentuk *cluster* menggunakan *medoids* sebagai pusatnya. Proses *clustering* dimulai dengan memilih k objek dari *dataset* secara acak sebagai *medoids*. *Medoids* merupakan titik yang sebenarnya atau berada dalam *dataset* sehingga lebih tahan terhadap *outlier*. *Cluster* dibentuk dengan menghitung jarak setiap objek *non-medoids* terhadap *k-medoids* yang dipilih, kemudian objek *non-medoids* akan dimasukkan ke dalam *cluster* yang *medoids* nya memiliki jarak terdekat (Hardiyanti et al., 2019).

Langkah-langkah dari *K-Medoids* adalah sebagai berikut (Dinata et al., 2021):

1. Tentukan k atau jumlah *cluster* yang diinginkan.
2. Tentukan secara acak *medoids* awal sebanyak k dari n data.
3. Hitung jarak antara data dan *medoids* dalam *clusters* menggunakan metode *Euclidean Distance*. *Euclidean Distance* merupakan salah satu metode yang digunakan untuk menghitung jarak antara dua titik dalam ruang *Euclidean* (Nishom, 2019). *Euclidean Distance* dapat dihitung menggunakan persamaan (2) berikut.

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

Keterangan:

- d = *Euclidean Distance*.
 i = Banyak objek.
 x_i = Data kriteria.
 μ_j = Nilai titik pusat pada *cluster* ke- j

4. Lakukan iterasi *medoids*.
5. Menghitung total simpangan (S).
 Menghitung total simpangan adalah menghitung total jarak antara setiap objek dalam *cluster* dengan *medoids* nya. Dengan ketentuan jika $S < 0$, maka tukar objek dengan data dari *cluster non-medoids* untuk membentuk kumpulan k objek baru sebagai *medoids*. Total simpangan dapat dihitung menggunakan persamaan (3) berikut:

$$S = b - a \dots\dots\dots (3)$$

Keterangan:

<https://doi.org/10.35145/joisie.v9i2.4965>

- S = Total Simpangan.
 a = jumlah jarak terdekat antara objek dengan *medoids* awal.
 b = jumlah jarak terdekat antara objek dengan *medoids* baru.

6. Menghitung ulang *k-medoids* sampai tidak ada perubahan pada *medoids*.

2.4 PURITY K-MEDOIDS

Purity digunakan untuk mengukur kemurnian satu atau beberapa *cluster* data yang dikategorikan sebagai bagian dari *cluster* yang paling sesuai dalam suatu kelas. Sebuah data dapat dikatakan murni jika semua objek dalam data dengan kelas yang sama berada dalam kelompok data yang juga sama. *Purity* dapat dihitung menggunakan persamaan (4) berikut (Dinata et al., 2021).

$$Purity(y) = \frac{1}{N_y} \max(n_{xy}) \dots\dots\dots (4)$$

Keterangan:

- $Purity(y)$ = Nilai *purity* untuk variabel ke- y .
 N_y = Jumlah data anggota *cluster* ke- y .
 xy = Indeks dari *cluster*.

Langkah-langkah dari metode *Purity K-Medoids* adalah sebagai berikut:

1. Menginout *dataset* yang berisi data yang ingin dikelompokkan.
2. Menentukan jumlah k *cluster*.
3. Menghitung nilai *purity* untuk setiap data menggunakan persamaan 3 di atas.
4. Jika jumlah k bernilai genap, maka pada iterasi pertama data yang akan digunakan sebagai *medoids* adalah data yang memiliki *purity minimum*, sedangkan jika k bernilai ganjil maka *medoids* yang digunakan adalah yang memiliki nilai *purity maksimum*.
5. Menghitung jarak antar data menggunakan persamaan Euclidean Distance dan kemudian menglompokkan nya. Jika jumlah k bernilai genap, maka data yang akan digunakan sebagai *medoids* adalah data yang memiliki *purity maksimum*, sedangkan jika k bernilai ganjil maka *medoids* yang digunakan adalah yang memiliki nilai *purity minimum*.
6. Menghitung total simpangan (S). Jika $S < 0$, maka tukar objek dengan data dari *cluster non-medoids* untuk membentuk kumpulan k objek baru sebagai *medoids*.

2.5 DAVIES BOULDIN INDEX (DBI)

DBI adalah metode untuk mengevaluasi kluster yang diperkenalkan oleh Davies dan Bouldin pada tahun 1979 (Jollyta & Siddik, 2023). *Davies Bouldin Index* (DBI) merupakan teknik pengukuran yang digunakan untuk menentukan hasil *cluster* terbaik setelah proses *clustering* selesai dilakukan. Metode ini bertujuan untuk mengoptimalkan perbedaan antara satu *cluster* dengan *cluster* yang lain dan berusaha meminimalkan jarak antar objek dalam satu *cluster*. DBI dapat menjadi parameter untuk setiap algoritma *clustering* untuk membandingkan proses *clustering* algoritma mana yang lebih optimal (Kaligis & Yulianto, 2022). Semakin kecil nilai DBI mendekati 0, maka semakin optimal *clustering* yang dihasilkan.

Davies Bouldin Index dapat dihitung dengan 4 langkah berikut (Sujacka Retno et al., 2024):

1. Menghitung SSW (Sum of Squares Within Cluster) menggunakan persamaan (5) berikut:

$$SSWi = \frac{1}{k} \sum_{i=1}^k d(x_i, \mu_j) \dots\dots\dots (5)$$
2. Menghitung SSB (Sum of Squares Between Cluster) menggunakan persamaan (6) berikut:

$$SSB_{i,j} = d(x_i, \mu_j) \dots\dots\dots (6)$$

3. Menghitung *Ratio* menggunakan persamaan (7) berikut:

$$R_{ij} = \frac{SSWi + SSW_j}{SSB_{ij}} \dots\dots\dots (7)$$

4. Menghitung SSW (Sum of Squares Within Cluster) menggunakan persamaan (8) berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad \dots\dots\dots (8)$$

2.6 PYTHON DAN GOOGLE COLABORATORY

Python adalah bahasa pemrograman yang populer dan mudah dipelajari. Python banyak digunakan dalam pengembangan perangkat lunak, kecerdasan buatan, pengembangan web, *machine learning*, dan analisis data. Python menyediakan berbagai library seperti *NumPy* untuk komputasi numerik, *Pandas* untuk analisis data, dan *Scikit-learn* yang menyediakan algoritma machine learning siap pakai untuk pengembangan model, pemrosesan data, dan evaluasi kinerja, sehingga memudahkan seseorang dalam menyelesaikan tugas tertentu dengan cepat dan efisien (Angelina M. T. I. Sambu et al., 2023).

Google Colaboratory atau Colab merupakan layanan cloud berbasis Jupyter Notebooks yang bertujuan untuk mendukung pendidikan dan penelitian machine learning. Jupyter adalah *tools open-source* berbasis browser yang mengintegrasikan bahasa pemrograman yang diinterpretasikan, *library*, dan alat untuk visualisasi. Jupyter notebook dapat berfungsi baik secara lokal maupun di *cloud*. Colab bekerja seperti objek *Google Docs* yaitu dapat dibagikan dan pengguna dapat berkolaborasi pada notebook yang sama. Colab menyediakan runtime Python 2 dan 3 yang sudah dikonfigurasi dengan library penting untuk *machine learning* dan kecerdasan buatan, seperti TensorFlow, Matplotlib, dan Keras (Carneiro et al., 2018).

3. HASIL DAN PEMBAHASAN

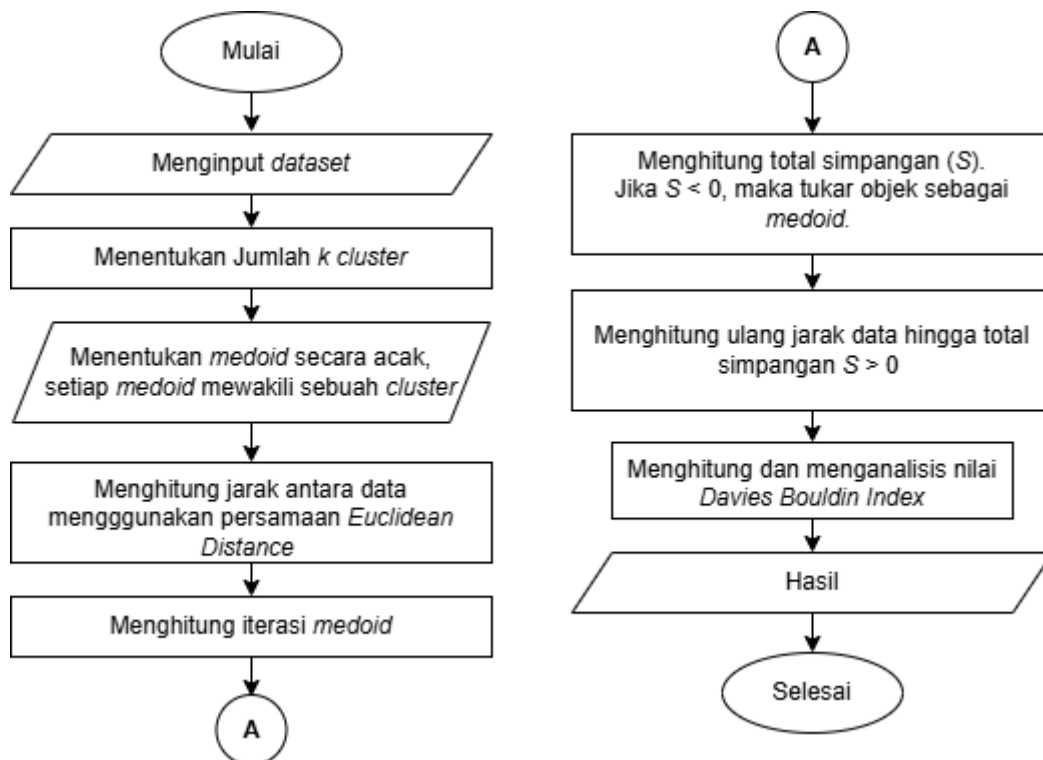
Sebelum data pertanian padi di Kabupaten Bireuen digunakan dalam implementasi metode *K-Medoids* dan *Purity K-Medoids* untuk mengklaster daerah produktivitas padi di Kabupaten Bireuen, data dinormalisasikan terlebih dahulu menggunakan persamaan (1) di atas. *Dataset* pertanian padi di Kabupaten Bireuen 2016 dapat dilihat pada Tabel 2.

Tabel 2. Data Pertanian Padi di Kabupaten Bireuen Tahun 2016

Kecamatan	Jumlah Desa	Luas Tanam	Luas Panen	Produktivitas	Jumlah Produksi	Persentase
Samalanga	0.54	0.99	0.81	0.30	0.75	0.05
Simpang, Mamplam	0.44	0.31	0.66	0.28	0.61	1.00
Pandrah	0.00	0.4	0.34	0.19	0.3	0.13
Jeunieb	0.48	0.62	0.69	1	0.83	0.35
Peulimbang	0.06	0.21	0.2	0.3	0.2	0.25
Peudada	0.66	0.72	0.81	0.2	0.71	0.36
Juli	0.34	0	0	0.1	0	0.32
Jeumpa	0.46	0.45	0.61	0.29	0.56	0.52
Kota Juang	0.08	0.04	0.04	0.39	0.06	0.27
Kuala	0.02	0.47	0.41	0.52	0.43	0.14
Jangka	0.54	0.67	0.79	0.37	0.75	0.4
Peusangan	1	1	1	0.55	1	0.23
Peusangan Selatan	0.04	0.09	0.06	0.14	0.06	0.22
Peusangan, Sb. Krueng	0.04	0.33	0.21	0	0.16	0
Makmur	0.16	0.07	0.23	0.08	0.19	0.86
Gandapura	0.42	0.16	0.35	0.15	0.3	0.8
Kuta Blang	0.44	0.61	0.68	0.42	0.66	0.34

3.1 PERHITUNGAN ALGORITMA K-MEDOIDS

Proses perhitungan *clustering* daerah produktivitas padi di Kabupaten Bireuen menggunakan metode *K-Medoids* dapat dilihat pada Gambar 2.



Gambar 2. Skema Algoritma K-Medoids

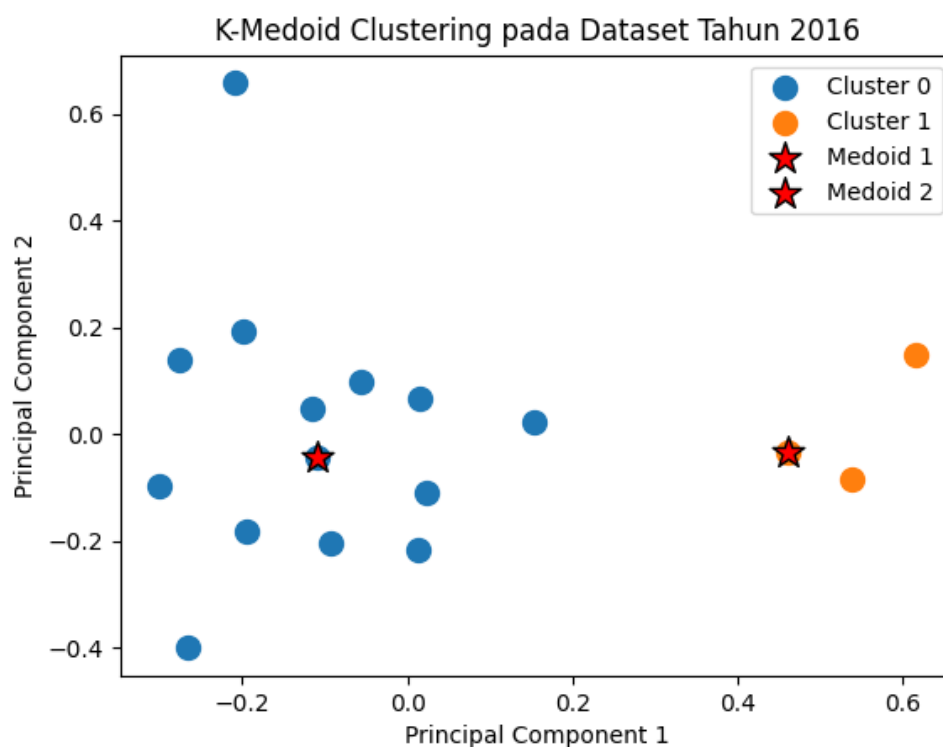
Berdasarkan Gambar 2 di atas, proses diawali dengan menginput dataset. Kemudian dilanjutkan dengan menentukan jumlah *k cluster* yang diinginkan. Setelah nilai *k* ditetapkan, langkah berikutnya adalah menentukan *medoids* awal secara acak dari *n* data. Setiap *medoids* yang terpilih akan mewakili satu *cluster* yang berbeda. Selanjutnya, proses berlanjut dengan menghitung jarak antara setiap data dengan *medoids* yang telah ditentukan. Jarak ini diperlukan untuk mengelompokkan data ke dalam *cluster* yang sesuai berdasarkan kedekatannya dengan *medoids*. Setelah semua jarak dihitung, langkah berikutnya adalah melakukan iterasi pada posisi *medoids*. Di sini, total simpangan (*S*) dari semua *cluster* dihitung dengan ketentuan jika nilai simpangan ini kurang dari nol ($S < 0$), maka tukar objek dengan data lain untuk membentuk *k* baru sebagai *medoids*. Proses ini terus diulang dengan menghitung ulang *medoids* hingga tidak ada perubahan yang terjadi pada *medoids*. Selanjutnya menghitung nilai *Davies Bouldin Index* (DBI). Kemudian mendapatkan hasil berupa pengelompokan daerah produktivitas padi di Kabupaten Bireuen beserta nilai DBI dari *clustering* tersebut.

Hasil *clustering* daerah produktivitas padi di Kabupaten Bireuen pada tahun 2016 dapat dilihat pada Tabel 3. Pada tabel tersebut, terdapat tiga kecamatan yang tergolong ke dalam kategori prioritas dan 14 kecamatan yang tergolong ke dalam kategori non-prioritas. Pada tahun 2016 terdapat 3 kecamatan yang memiliki produktivitas padi tinggi yaitu kecamatan Simpang Mamplam, Makmur, dan Gandapura. Visualisasi hasil *clustering* menggunakan *K-Medoids* pada *dataset* tahun 2019 dapat dilihat pada Gambar 3.

Tabel 3. Hasil *Clustering K-Medoids* Tahun 2016

No	Kecamatan	Cluster
1	Samalanga	Non-Prioritas
2	Simpang, Mamplam	Prioritas
3	Pandrah	Non-Prioritas
4	Jeunieb	Non-Prioritas

5	Peulimbang	Non-Prioritas
6	Peudada	Non-Prioritas
7	Juli	Non-Prioritas
8	Jeumpa	Non-Prioritas
9	Kota Juang	Non-Prioritas
10	Kuala	Non-Prioritas
11	Jangka	Non-Prioritas
12	Peusangan	Non-Prioritas
13	Peusangan Selatan	Non-Prioritas
14	Peusangan, Sb. Krueng	Non-Prioritas
15	Makmur	Prioritas
16	Gandapura	Prioritas
17	Kuta Blang	Non-Prioritas



Gambar 3. Visualisasi hasil *clustering K-Medoids* pada dataset tahun 2016

3.2 PERHITUNGAN ALGORITMA *PURITY K-MEDOIDS*

Untuk menghitung nilai *purity* dapat menggunakan persamaan (4) di atas. Proses perhitungan nilai *purity* data pada dataset tahun 2016 dapat dilihat sebagai berikut:

1. *Purity* data ke-1

$$\begin{aligned}
 P_1 &= \frac{1}{(0.54 + 0.99 + 0.81 + 0.3 + 0.75 + 0.05)} \max(0.54 + 0.99 + 0.81 + 0.3 + 0.75 + 0.05) \\
 &= \frac{1}{(3.44)} \max(0.99) \\
 &= 0.287791
 \end{aligned}$$

2. *Purity* data ke-2

<https://doi.org/10.35145/joisie.v9i2.4965>

$$\begin{aligned}
 P_1 &= \frac{1}{(0.44 + 0.31 + 0.66 + 0.28 + 0.61 + 1.00)} \max (0.44 + 0.31 + 0.66 + 0.28 + 0.61 + 1) \\
 &= \frac{1}{(3,30)} \max(0.66) \\
 &= 0.30303
 \end{aligned}$$

3. *Purity* data ke-3

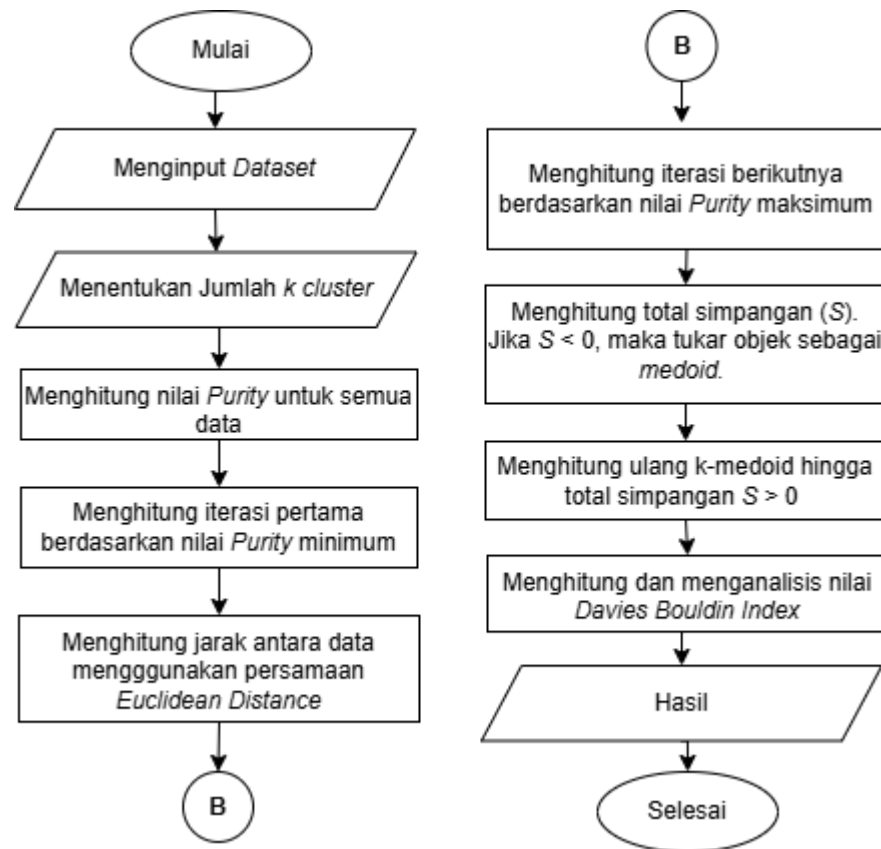
$$\begin{aligned}
 P_1 &= \frac{1}{(0.00 + 0.40 + 0.34 + 0.19 + 0.3 + 0.13)} \max (0.00 + 0.40 + 0.34 + 0.19 + 0.3 + 0.13) \\
 &= \frac{1}{(1.36)} \max(0.40) \\
 &= 0.294118
 \end{aligned}$$

Nilai *purity* untuk seluruh pada tahun 2016 dapat dilihat pada Tabel 4. Berdasarkan tabel tersebut, dua data yang memiliki nilai *purity minimum* berturut-turut adalah data ke-12 dengan nilai 0.211073 dan data ke-8 dengan nilai 0.209205, sedangkan dua data yang memiliki nilai *purity maksimum* berturut-turut adalah data ke-15 dengan nilai 0.540881 dan data ke-14 dengan nilai 0.445946.

Tabel 4. Hasil perhitungan *Purity* Tahun 2016

Data ke-	Nilai <i>Purity</i>
1	0.287791
2	0.30303
3	0.294118
4	0.251889
5	0.245902
6	0.234104
7	0.447368
8	0.211073
9	0.443182
10	0.261307
11	0.224432
12	0.209205
13	0.360656
14	0.445946
15	0.540881
16	0.366972
17	0.215873

Proses perhitungan clustering daerah produktivitas padi di Kabupaten Bireuen menggunakan algoritma *Purity K-Medoids* dapat dilihat pada Gambar 4.



Gambar 4. Skema Algoritma Purity K-Medoids

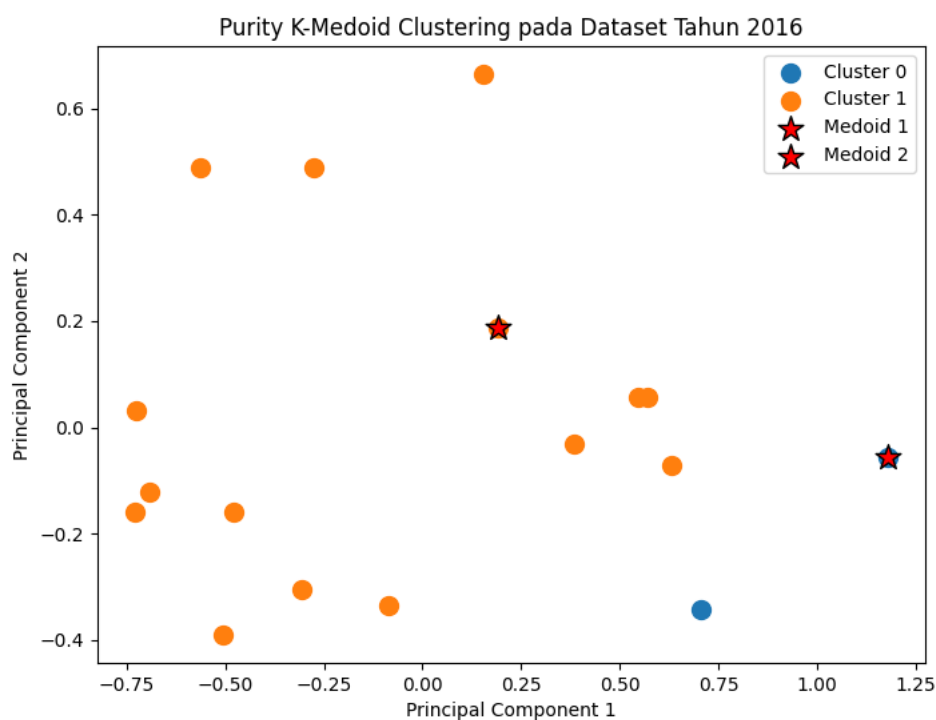
Proses dimulai dengan menginput *dataset* yang berisi data yang ingin dikelompokkan. Setelah *dataset* diinput, langkah pertama yang dilakukan adalah menentukan jumlah *k cluster* yang diinginkan. Setelah *k* ditentukan melakukan proses perhitungan nilai *purity* untuk setiap data. Setelah setiap data memiliki nilai *purity* dilakukan perangkingan data berdasarkan nilai *purity* dimulai dari data dengan nilai *purity* sebagai ranking 1 dan seterusnya untuk memudahkan dalam penentuan data yang memiliki nilai *purity minimum* dan *maksimum*. Setelah nilai *purity* serta perangkingan ditentukan, langkah berikutnya adalah menghitung nilai *purity* untuk setiap data, jika *k* bernilai genap, iterasi pertama dilakukan berdasarkan nilai *purity minimum*, sedangkan jika *k* bernilai ganjil, iterasi pertama dilakukan berdasarkan nilai *purity maksimum*. Selanjutnya menghitung jarak data terhadap *medoids* serta mengelompokkannya dan menghitung jumlahnya. Dilanjutkan dengan melakukan iterasi selanjutnya berdasarkan nilai *purity maksimum*, jika *k* bernilai genap. Sedangkan jika *k* bernilai ganjil, iterasi selanjutnya dilakukan berdasarkan nilai *purity minimum*. Proses selanjutnya adalah menghitung total simpangan (*S*) dari semua *cluster*. Kemudian menghitung ulang *medoids* hingga tidak ada perubahan yang terjadi pada *medoids*. Selanjutnya menghitung nilai *Davies Bouldin Index* (DBI). Kemudian mendapatkan hasil berupa pengelompokan daerah produktivitas padi di Kabupaten Bireuen beserta nilai DBI dari *clustering* tersebut.

Hasil *clustering* daerah produktivitas padi di Kabupaten Bireuen pada tahun 2016 dapat dilihat pada Tabel 5. Pada tabel tersebut, terdapat dua kecamatan yang tergolong ke dalam kategori prioritas dan 15 kecamatan yang tergolong ke dalam non-prioritas. Pada tahun 2016 terdapat 2 kecamatan yang memiliki produktivitas padi tinggi yaitu kecamatan Samalanga dan kecamatan Peusangan. Visualisasi hasil *clustering* menggunakan *K-Medoids* pada *dataset* tahun 2019 dapat dilihat pada Gambar 5.

Tabel 5. Hasil Clustering Purity K-Medoids Tahun 2016

No	Kecamatan	Cluster
1	Samalanga	Prioritas
2	Simpang, Mamplam	Non-Prioritas

3	Pandrah	Non-Prioritas
4	Jeunieb	Non-Prioritas
5	Peulimbang	Non-Prioritas
6	Peudada	Non-Prioritas
7	Juli	Non-Prioritas
8	Jeumpa	Non-Prioritas
9	Kota Juang	Non-Prioritas
10	Kuala	Non-Prioritas
11	Jangka	Non-Prioritas
12	Peusangan	Prioritas
13	Peusangan Selatan	Non-Prioritas
14	Peusangan, Sb. Krueng	Non-Prioritas
15	Makmur	Non-Prioritas
16	Gandapura	Non-Prioritas
17	Kuta Blang	Non-Prioritas



Gambar 5. Viusalisasi hasil clustering K-Medoids pada dataset tahun 2016

3.3 PERBANDINGAN HASIL *CLUSTERING*

Hasil Perbandingan hasil *clustering* daerah produktivitas padi di Kabupaten Bireuen pada tahun 2016 menggunakan *K-Medoids* dan *Purity K-Medoids* dapat dilihat pada Tabel 6.

Tabel 6. Perbandingan Hasil *Clustering* Tahun 2016

No	Kecamatan	<i>K-Medoids</i>	<i>Purity K-Medoids</i>
1	Samalanga	Non-Prioritas	Prioritas
2	Simpang, Mamplam	Prioritas	Non-Prioritas
3	Pandrah	Non-Prioritas	Non-Prioritas

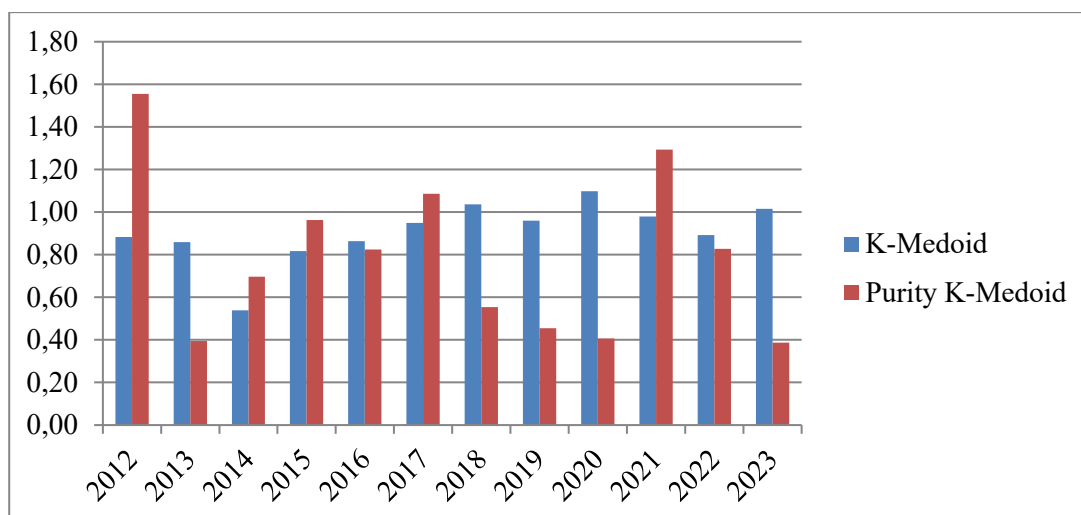
4	Jeunieb	Non-Prioritas	Non-Prioritas
5	Peulimbang	Non-Prioritas	Non-Prioritas
6	Peudada	Non-Prioritas	Non-Prioritas
7	Juli	Non-Prioritas	Non-Prioritas
8	Jeumpa	Non-Prioritas	Non-Prioritas
9	Kota Juang	Non-Prioritas	Non-Prioritas
10	Kuala	Non-Prioritas	Non-Prioritas
11	Jangka	Non-Prioritas	Non-Prioritas
12	Peusangan	Non-Prioritas	Prioritas
13	Peusangan Selatan	Non-Prioritas	Non-Prioritas
14	Peusangan, Sb. Krueng	Non-Prioritas	Non-Prioritas
15	Makmur	Prioritas	Non-Prioritas
16	Gandapura	Prioritas	Non-Prioritas
17	Kuta Blang	Non-Prioritas	Non-Prioritas

3.4 HASIL VALIDASI NILAI *DAVIES BOULDIN INDEX* (DBI)

Hasil validasi nilai *Davies Bouldin Index* (DBI) dari *clustering* menggunakan *K-Medoids* dan *Purity K-Medoids* setelah di rata-rata kan dari 10 kali percobaan dapat dilihat pada Tabel 7. Berdasarkan tabel tersebut, diperoleh rata-rata nilai *Davies Bouldin Index* (DBI) dari *clustering* data setiap tahun dengan masing-masing 10 kali percobaan. Setelah dirata-ratakan keseluruhan, *clustering* menggunakan metode *K-Medoids* memiliki nilai rata-rata DBI sebesar 0.907856. Sedangkan *clustering* menggunakan metode *Purity K-Medoids* memiliki nilai rata-rata DBI sebesar 0.786911. Visualisasi dari hasil *Davies Buldin Index* dapat dilihat pada Gambar 6.

Tabel 7. Hasil Validasi Nilai *Davies Bouldin Index* (DBI)

No	Tahun	<i>Nilai Davies Bouldin Index (DBI)</i>	
		<i>K-Medoids</i>	<i>Purity K-Medoids</i>
1	2012	0.883811	1.554882
2	2013	0.859425	0.394329
3	2014	0.539447	0.697099
4	2015	0.817701	0.963164
5	2016	0.863290	0.823944
6	2017	0.948549	1.085871
7	2018	1.036833	0.554250
8	2019	0.960252	0.455156
9	2020	1.097963	0.407043
10	2021	0.979807	1.293412
11	2022	0.892458	0.827134
12	2023	1.014740	0.386645
Rata-Rata		0.907856	0.786911



Gambar 6. Visualisasi nilai DBI

Nilai Davies Bouldin Index (DBI) yang lebih rendah pada metode *Purity K-Medoids* menunjukkan bahwa hasil klasterisasi lebih optimal dibandingkan *K-Medoids*. *Purity K-Medoids* menginisialisasi *medoids* berdasarkan tingkat kemurnian (*purity*) setiap *cluster*, sehingga anggota *cluster* cenderung lebih homogen sejak awal proses. Inisialisasi ini mengurangi kemungkinan terbentuknya *cluster* yang tumpang tindih atau tidak representatif, sehingga jarak antar *cluster* menjadi lebih besar dan jarak dalam *cluster* menjadi lebih kecil. Hal ini secara langsung berdampak pada penurunan nilai DBI, yang memang mengukur rasio antara jarak *intra-cluster* dan *inter-cluster*. Dengan demikian, *Purity K-Medoids* secara teoretis mampu menghasilkan pembagian *cluster* yang lebih baik dan terpisah jelas dibandingkan *K-Medoids* yang inisialisasinya secara acak.

3.5 ANALISIS KONSISTENSI KEMUNCULAN

Setelah dilakukan proses *clustering*, maka langkah selanjutnya dalam penelitian ini adalah menentukan kecamatan yang sering muncul sebagai daerah prioritas. Dikarenakan *clustering* menggunakan metode *Purity K-Medoids* memiliki nilai rata-rata DBI 0.786911 dimana lebih rendah daripada *clustering* menggunakan metode *K-medoids* yaitu 0.907856, untuk menentukan kecamatan mana saja yang sering muncul sebagai kecamatan prioritas selama 12 tahun dalam periode 2012-2023, maka yang akan dijadikan acuan adalah hasil *clustering* menggunakan metode *Purity K-Medoids*. Kecamatan-kecamatan yang tergolong ke dalam kecamatan prioritas selama 12 tahun dalam periode 2012-2023 dapat dilihat pada Tabel 8.

Tabel 8. Kecamatan Prioritas Tahun 2012-2023

No	Tahun	Kecamatan
1	2012	Samalanga
		Simpang mamplam
		Pandrah
		Jeunieb
		Peudada
		Makmur
2	2013	Peusangan
3	2014	Peusangan Selatan
		Peusangan Sb. Krueng
		Makmur
		Gandapura
4	2015	Samalanga

		Simpang, Mamplam
		Peudada
		Jangka
		Peusangan
5	2016	Samalanga
		Peusangan
6	2017	Samalanga
		Simpang, Mamplam
		Jeunieb
		Peudada
7	2018	Peusangan
8	2019	Peusangan
9	2020	Peusangan
10	2021	Peusangan Selatan
		Makmur
11	2022	Jeunieb
		Peusangan
12	2023	Peusangan

Tabel 9. Jumlah Kemunculan

No	Kecamatan	Jumlah Kemunculan
1	Samalanga	4
2	Simpang, Mamplam	3
3	Pandrah	1
4	Jeunieb	3
5	Peulimbang	0
6	Peudada	3
7	Juli	0
8	Jeumpa	0
9	Kota Juang	0
10	Kuala	0
11	Jangka	1
12	Peusangan	8
13	Peusangan Selatan	2
14	Peusangan, Sb. Krueng	1
15	Makmur	3
16	Gandapura	1
17	Kuta Blang	0

Jumlah kemunculan tiap kecamatan sebagai prioritas dapat dilihat pada Tabel 9. Berdasarkan tabel tersebut, dalam periode 2012-2023, kecamatan Peusangan memiliki total jumlah 8 kali kemunculan yang digolongkan sebagai kecamatan prioritas. Selanjutnya adalah kecamatan Samalanga dengan total 4 kali kemunculan, kecamatan Simpang Mamplam, Jeunieb, peudaa dan makmur dengan total masing-masing 3 kali kemunculan, kecamatan pandrah, Jangka, Peusangan, Sb. Krueng dan gandapura dengan total masing-masing 1 kali kemunculan. Sedangkan kecamatan lainnya tidak pernah

<https://doi.org/10.35145/joisie.v9i2.4965>

JOISIE licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

muncul tergolong sebagai daerah prioritas. Jadi dapat disimpulkan bahwa kecamatan Peusangan adalah kecamatan yang paling sering muncul dalam periode 2012-2023 dengan total jumlah 8 kali kemunculan yang digolongkan sebagai kecamatan prioritas.

4. SIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode *Purity K-Medoids* terbukti memberikan hasil klasterisasi yang lebih baik dibandingkan *K-Medoids* standar dalam mengelompokkan daerah produktivitas panen padi di Kabupaten Bireuen. Hal ini ditunjukkan oleh nilai rata-rata *Davies Bouldin Index* (DBI) pada *Purity K-Medoids* sebesar 0,786911, yang lebih rendah daripada DBI *K-Medoids* sebesar 0,907856. Hasil ini menunjukkan juga bahwa *Purity K-Medoids* mampu menghasilkan *cluster* yang lebih terpisah dan homogen secara internal. Penelitian ini pula secara konsisten mengidentifikasi kecamatan Peusangan sebagai daerah prioritas, dengan kemunculan sebanyak 8 kali pada tahun 2013, 2015, 2016, 2018, 2019, 2020, 2022, dan 2023 selama periode 2012–2023. Temuan ini menegaskan bahwa Kecamatan Peusangan merupakan wilayah yang perlu mendapatkan perhatian khusus dalam alokasi sumber daya pertanian. Diharapkan temuan penelitian ini dapat membantu pemerintah daerah dalam merumuskan kebijakan alokasi pupuk subsidi berbasis data, sehingga distribusi bantuan dapat lebih tepat sasaran dan efektif dalam meningkatkan produktivitas padi di Kabupaten Bireuen.

5. DAFTAR PUSTAKA

- Angelina M. T. I. Sambu Ua, Diandra Lestriani H, Elizabeth Sonia Kristanty Marpaung, Jesslyn Ong, Michelle Savinka, Putri Nurhaliza, & Rahmi Yulia Ningsih. (2023). Penggunaan Bahasa Pemrograman Python Dalam Analisis Faktor Penyebab Kanker Paru-Paru. *Jurnal Publikasi Teknik Informatika*, 2(2), 88–99. <https://doi.org/10.55606/jupti.v2i2.1742>
- Carneiro, T., Da Nobrega, R. V. M., Nepomuceno, T., Bian, G. Bin, De Albuquerque, V. H. C., & Filho, P. P. R. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access*, 6, 61677–61685. <https://doi.org/10.1109/ACCESS.2018.2874767>
- Dinata, R. K., Retno, S., & Hasdyna, N. (2021). Minimization of the Number of Iterations in K-Medoids Clustering with Purity Algorithm. *Revue d'Intelligence Artificielle*, 35(3), 193–199. <https://doi.org/10.18280/ria.350302>
- Hardiyanti, F., Tambunan, H. S., & Saragih, I. S. (2019). Penerapan Metode K-Medoids Clustering Pada Penanganan Kasus Diare Di Indonesia. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 598–603. <https://doi.org/10.30865/komik.v3i1.1666>
- Hasdyna, N., & Retno, S. (2022). Purity Algorithm in Determining System of The Productivity of Rice Harvesting Areas in Kabupaten Aceh Utara. *Journal of Informatics and Telecommunication Engineering*, 5(2), 259–267. <https://doi.org/10.31289/jite.v5i2.6030>
- Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain. *Mathematics*, 10(11), 1–18. <https://doi.org/10.3390/math10111942>
- Jollyta, D., & Siddik, M. (2023). Pengoptimalan Pengukuran Bregman Divergences Menggunakan Davies Bouldin Index. *JOISIE Journal Of Information System And Informatics Engineering*, 7(1), 174–180.
- Kaligis, G. B., & Yulianto, S. (2022). Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokan Kinerja Pegawai. *IT-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, 1(3), 179–193. <https://doi.org/10.24246/itexplore.v1i3.2022.pp179-193>
- Lailiyah, N., Timisela, N. R., & Kaplale, R. (2020). Analisis Produksi Padi Sawah (*Oryza sativa* L) Tadah Hujan di Desa Lea Wai Kecamatan Seram Utara Timur Kobi. *Jurnal Agribisnis Kepulauan*, 5(2), 151–165.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>

- Permana, I., & Salisah, F. N. S. (2022). Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 2(1), 67–72. <https://doi.org/10.57152/ijirse.v2i1.311>
- Pratiwi, H., & Wahyu Wibowo, A. P. (2022). Implementasi Algoritma K-Means Untuk Mengklaster Kelompok Sektor Perkebunan Di Indonesia. *JOISIE Journal Of Information System And Informatics Engineering*, 6(1), 1–10.
- Sujacka Retno, Bustami, & Rozzi Kesuma Dinata. (2024). Enhancing K-Means Clustering Model to Improve Rice Harvest Productivity Areas in Aceh Utara Using Purity. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 13(2), 405–418. <https://doi.org/10.23887/janapati.v13i2.78254>
- Tohendry, D., & Jollyta, D. (2023). Penerapan Algoritma K-Means Cluster Ing Untuk Pengelompokkan Saham Berdasarkan Price Earning Ratio Dan Price To Book Value. *Jurnal Mahasiswa Aplikasi Teknologi Komputer Dan Informasi*, 5(1), 3–9.