

SENASTIKA Universitas Malikussaleh

ANALISIS PERBANDINGAN KINERJA ALGORITMA K-MEANS DAN K-MEDOIDS UNTUK KLASTERISASI PRODUKSI PADI DI PULAU SUMATERA

Siti Nadilla^{*1}, Wirda Syahrifa², Fakhrul Razi³, Rizki Maulana⁴, Munirul Ula⁵

^{1,2,3,4} Program Studi Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: ¹siti.210170032@mhs.unimal.ac.id, ²wirda.210170033@mhs.unimal.ac.id,
³fakhrul.210170029@mhs.unimal.ac.id, ⁴rizki.210170037@mhs.unimal.ac.id,
⁵munirulula@unimal.ac.id

Abstrak

Penelitian ini bertujuan untuk melakukan analisis pengelompokan data tanaman padi di Pulau Sumatra selama periode 2010 hingga 2020 dengan menerapkan algoritma K-Means dan K-Medoids. Pengelompokan ini penting untuk mengidentifikasi pola produksi padi, sehingga dapat membedakan daerah dengan hasil produksi tertinggi dan terendah. Data yang dianalisis mencakup variabel kunci seperti luas lahan pertanian, tingkat produktivitas, dan total hasil panen, yang diperoleh dari sumber resmi pertanian. Metodologi yang digunakan mencakup langkah-langkah preprocessing data, seleksi fitur yang relevan, dan penerapan algoritma pengelompokan. Hasil analisis menunjukkan bahwa algoritma K-Means menawarkan kecepatan dan efisiensi yang lebih tinggi dalam membentuk kluster berdasarkan karakteristik produksi, sementara K-Medoids menunjukkan ketahanan yang lebih baik terhadap data yang tidak biasa. Temuan dari pengelompokan ini berhasil mengidentifikasi sejumlah daerah dengan produksi padi tertinggi dan terendah, memberikan wawasan berharga bagi kepentingan pembaca dalam meningkatkan strategi pengelolaan sumber daya pertanian. Penelitian ini diharapkan dapat menjadi dasar bagi pengambilan keputusan yang lebih baik dalam upaya meningkatkan produktivitas dan keberlanjutan sektor pertanian di Pulau Sumatra.

Keywords: *Pengelompokan, K-Means, K-Medoids, Tanaman Padi, Sumatera.*

1. PENDAHULUAN

Pertanian padi di Indonesia, khususnya di Pulau Sumatra, merupakan salah satu sektor penting dalam mendukung ketahanan pangan nasional. Pulau Sumatra memiliki potensi lahan yang luas dan beragam, yang jika dikelola dengan baik dapat meningkatkan hasil produksi padi secara signifikan. Namun, meskipun terdapat berbagai sumber daya alam yang melimpah, disparitas dalam hasil produksi padi masih terjadi di berbagai daerah. Oleh karena itu, analisis pengelompokan data produksi padi menjadi penting untuk memahami faktor-faktor yang memengaruhi kinerja produksi dan untuk merumuskan strategi peningkatan produktivitas [1].

Padi (*Oryza sativa*) merupakan komoditas pangan utama di Indonesia, khususnya di Pulau Sumatra yang memiliki potensi besar dalam produksi beras. Sebagai salah satu sumber pangan utama, pengelolaan data terkait tanaman padi menjadi penting untuk mendukung peningkatan produktivitas dan efisiensi pertanian [1]. Salah satu pendekatan yang dapat digunakan untuk analisis data tanaman padi adalah pengelompokan data atau *clustering*, yang bertujuan untuk mengidentifikasi pola dan mengelompokkan area berdasarkan karakteristik yang relevan seperti jenis lahan, cuaca, dan hasil panen [2].

Algoritma K-Means dan K-Medoids merupakan dua metode pengelompokan yang sering digunakan dalam analisis data. Algoritma K-Means bekerja dengan meminimalkan variasi dalam kluster dengan menghitung centroid, yaitu titik tengah dari setiap kluster yang terbentuk [3]. Sementara itu, K-Medoids bekerja dengan memilih titik data yang aktual sebagai pusat kluster, sehingga lebih robust terhadap outlier dibandingkan K-Means [4]. Penggunaan kedua algoritma ini sangat relevan dalam konteks pengelolaan data tanaman padi, karena seringkali data pertanian memiliki distribusi yang tidak merata dan rentan terhadap outlier [5].

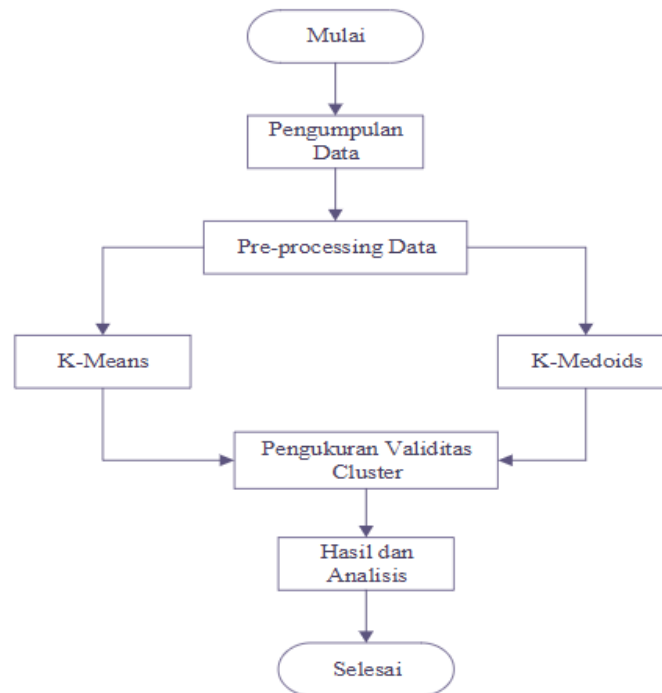
Pulau Sumatra memiliki variasi kondisi agroklimat yang mempengaruhi hasil panen padi di setiap wilayahnya. Dengan demikian, analisis clustering dapat membantu mengelompokkan wilayah-wilayah yang memiliki kesamaan karakteristik dan mengidentifikasi faktor-faktor yang mempengaruhi produktivitas padi. Studi sebelumnya telah menunjukkan bahwa teknik pengelompokan berbasis algoritma seperti K-Means dan K-Medoids efektif dalam mengidentifikasi pola tersembunyi dalam data pertanian, serta memberikan informasi yang bermanfaat untuk pengambilan keputusan di sektor pertanian.

Penelitian ini bertujuan untuk melakukan analisis pengelompokan data tanaman padi di Pulau Sumatra menggunakan algoritma K-Means dan K-Medoids. Dengan mengimplementasikan dua algoritma tersebut dengan mencoba beberapa kluster untuk mengelompokkan wilayah hasil produksi tanaman padi di Provinsi Sumatera, yang lebih jelas mengenai distribusi dan potensi wilayah tanam padi di Sumatera [6]. Pada hasil akhir penelitian, kedua algoritma, K-Means dan K-Medoids, berhasil mengelompokkan data dengan baik. Namun, hasil

menunjukkan bahwa algoritma K-Medoids memberikan performa yang lebih baik dibandingkan K-Means. Algoritma K-Medoids mampu menghasilkan kelompok yang lebih stabil dan akurat. Selain itu, hasil dari studi ini juga diharapkan dapat memberikan wawasan bagi pengambil kebijakan dalam merancang strategi peningkatan produktivitas pertanian yang lebih efektif dan efisien [7].

2. METODE PENELITIAN

Penelitian ini akan melalui lima tahap utama, yaitu pengumpulan data, kemudian dataset yang diperoleh akan melalui proses *pre-processing*. Data yang telah melalui tahap peningkatan kualitas akan diproses menggunakan dua algoritma, yaitu K-Means dan K-Medoids. Setelah data diproses oleh masing-masing algoritma, validitas kluster akan dihitung menggunakan metode DBI. Hasil klustering dan perhitungan validitas DBI kemudian akan dianalisis untuk mendapatkan kesimpulan. Metodologi penelitian ini dapat dilihat pada gambar 1.

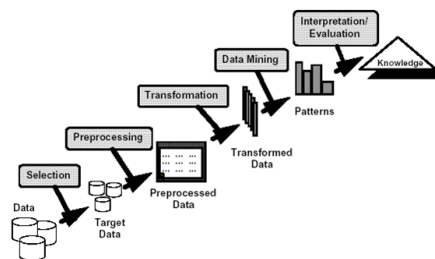


Gambar 1. Metodologi Penelitian

Dalam penelitian ini, dataset yang digunakan adalah data tanaman padi di pulau Sumatera yang diperoleh dari kaggle.com, dengan rentang waktu dari tahun 2010 hingga tahun 2020. Data tersebut akan melalui tahap *pre-processing*, dimulai dengan pembersihan data (cleaning). Setelah kualitas data dianggap memadai, data akan diproses menggunakan dua algoritma, yaitu K-Means dan K-Medoids. Tahap berikutnya adalah analisis hasil, di mana validitas kluster dihitung menggunakan Davies Bouldin Index (DBI) untuk menentukan algoritma terbaik.

2.1 Data Mining

Data mining adalah proses yang digunakan untuk menemukan pola, hubungan, atau informasi penting dari sejumlah besar data yang tersebar secara tidak terstruktur. Melalui data mining, teknik-teknik analisis digunakan untuk mengekstrak informasi tersembunyi yang dapat membantu pengambilan keputusan dan pengembangan model prediksi [8]. Menurut Han et al. (2019), data mining melibatkan serangkaian proses seperti pengumpulan data, pembersihan data, transformasi, dan penambangan itu sendiri, yang bertujuan untuk menghasilkan pola yang bermakna dari kumpulan data yang besar. Teknik-teknik yang digunakan dalam data mining meliputi algoritma klasifikasi, clustering, regresi, asosiasi, dan analisis deret waktu.



Gambar 2. Proses Data Mining

2.2 Clustering

Clustering adalah salah satu teknik dalam data mining yang memiliki tujuan untuk mengelompokkan objek-objek ke dalam kelompok atau *cluster* berdasarkan kesamaan karakteristiknya. Objek-objek yang berada dalam satu kelompok memiliki karakteristik yang lebih mirip satu sama lain dibandingkan dengan objek-objek yang berada di kelompok lain. Proses clustering digunakan untuk menemukan struktur tersembunyi dalam data yang tidak memiliki label atau kategori sebelumnya, sehingga dapat memberikan pemahaman lebih baik tentang data tersebut.

2.3 K-Means

K-Means adalah salah satu algoritma *clustering* yang paling populer dan sederhana dalam data mining. Algoritma ini bertujuan untuk membagi sekumpulan data ke dalam sejumlah kelompok (*cluster*) berdasarkan kesamaan karakteristik. Setiap *cluster* direpresentasikan oleh sebuah titik pusat, yang disebut centroid, yang merupakan rata-rata dari semua data dalam *cluster* tersebut. Algoritma K-Means bekerja dengan cara meminimalkan jarak antara data dengan centroid *cluster*-nya [9]. Dalam kata lain, K-Means mengelompokkan data sedemikian rupa sehingga objek-objek yang mirip berada dalam kelompok yang sama, dan objek-objek yang berbeda ditempatkan dalam kelompok yang berbeda. Ada beberapa langkah dalam melakukan perhitungan menggunakan algoritma K-Means yaitu:

1. Tentukan nilai K untuk titik awal klaster secara acak
2. Setiap data akan dibagi ke dalam klaster berdasarkan jarak Euclidean ke titik pusat klaster. Rumus untuk menghitung jarak Euclidean dapat dilihat pada persamaan. (1):

$$d(x,y) = \|x-y\|^2 = \dots\dots\dots(1)$$

dengan :

d= jarak data ke pusat cluster

x= data ke pada atribut ke

y= titik pusat ke, pada atribut

3. Pusat klaster akan dihitung kembali berdasarkan nilai rata-rata dari data yang ada dalam klaster tersebut.
4. Proses di tahap 2 dan 3 akan diulang hingga ditemukan kelompok yang stabil, di mana tidak ada perubahan anggota dalam klaster dibandingkan dengan percobaan sebelumnya.

2.4 K-Medoids

K-Medoids adalah algoritma *clustering* yang mirip dengan K-Means, tetapi memiliki perbedaan mendasar dalam cara menentukan pusat klaster. Sementara K-Means menggunakan centroid (yang merupakan rata-rata dari semua titik dalam kluster) sebagai representasi dari setiap kluster, K-Medoids menggunakan titik data aktual yang terletak di dalam kluster tersebut sebagai pusat kluster, yang dikenal sebagai medoid. Dengan demikian, K-Medoids lebih robust terhadap outlier dan noise dalam data, karena medoid adalah titik yang paling representatif dari sekumpulan data dalam kluster [10].

Tahap-tahap penyelesaian K-Medoids adalah sebagai berikut:

1. Tentukan jumlah K sebagai pusat klaster.
2. Setiap data dimasukkan ke dalam klaster terdekat dengan cara mengukur jarak menggunakan rumus Jarak Euclidean sesuai dengan persamaan (1).
3. Pilih secara acak satu data dari setiap klaster untuk dijadikan medoid yang baru.
4. Objek yang termasuk dalam masing-masing klaster akan dihitung menggunakan medoid yang baru.
5. Total simpangan akan dihitung dengan membandingkan total jarak baru dengan total jarak awal. Jika hasilnya $S < 0$, cari data baru untuk membentuk k objek yang baru sebagai medoid.
6. Ulangi langkah 3 hingga medoid tidak lagi berubah, sehingga diperoleh anggota untuk masing-masing klaster.[11]

2.5 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) adalah metrik yang digunakan untuk mengevaluasi kualitas clustering yang dihasilkan oleh algoritma pengelompokan. DBI mengukur sejauh mana setiap kluster terpisah dari kluster lainnya, serta seberapa baik kluster tersebut terdefinisi. Nilai DBI yang lebih rendah menunjukkan bahwa kluster yang terbentuk memiliki jarak yang lebih besar satu sama lain (artinya, mereka lebih terpisah) dan bahwa titik-titik dalam kluster tersebut lebih dekat satu sama lain (artinya, kluster tersebut lebih padat).

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data dan Pre-Processing Data

Pada penelitian ini, Proses pengumpulan data merupakan tahap awal yang sangat penting dalam analisis data. Data yang digunakan dalam penelitian ini diperoleh kaggle terkait produksi padi di Pulau Sumatra selama periode 2010 hingga 2020. Sumber data mencakup laporan tahunan produksi padi, serta luas panen.

Tabel 1. Data awal produksi padi

No	Provinsi	Produksi	Luas Panen
1.	Aceh 2010	1788738	352281
2.	Aceh 2011	1772962	380686
3.	Aceh 2012	1582393	387803
4.	Aceh 2013	2331046	419183
5.	Aceh 2014	1820062	376137
6.	Aceh 2015	1956940	461060
7.	Aceh 2016	2180754	293067
8.	Aceh 2017	2478922	294483
9.	Aceh 2018	1751996,94	329515,78
10.	Aceh 2019	1714437,6	310012,46
11.	Aceh 2020	1861567,1	317869,41
12.	Sumatera Utara 2010	3582302	754674
13.	Sumatera Utara 2011	3607403	757547
14.	Sumatera Utara 2012	3715514	765099
15.	Sumatera Utara 2013	3727249	742968
...
87.	Lampung 2019	2164089,33	464103,42
88.	Lampung 2020	2604913,29	545149,05

Data yang digunakan dalam penelitian ini telah melalui proses normalisasi menggunakan teknik Min-Max Normalization. Teknik ini mengubah nilai-nilai asli setiap variabel ke dalam rentang 0 hingga 1, sehingga memungkinkan setiap fitur memiliki skala yang sebanding dan mencegah bias pada hasil pengelompokan

Tabel 2. Hasil Data Normalisasi

No	Provinsi	Produksi	Luas Panen
1.	Aceh 2010	0,335	0,357
2.	Aceh 2011	0,332	0,392
3.	Aceh 2012	0,291	0,401
4.	Aceh 2013	0,452	0,440
5.	Aceh 2014	0,342	0,387
6.	Aceh 2015	0,371	0,492
7.	Aceh 2016	0,419	0,284
8.	Aceh 2017	0,483	0,286
9.	Aceh 2018	0,327	0,329
10.	Aceh 2019	0,3192	0,305
11.	Aceh 2020	0,351	0,315
12.	Sumatera Utara 2010	0,721	0,854
13.	Sumatera Utara 2011	0,726	0,858
14.	Sumatera Utara 2012	0,749	0,867
15.	Sumatera Utara 2013	0,752	0,840
...
87.	Lampung 2019	0,416	0,495
88.	Lampung 2020	0,511	0,595

3.2 Proses Klasterisasi dan Validitas

Proses klasterisasi atau pengelompokan pada penelitian ini akan dilakukan dengan menggunakan dua algoritma, yaitu algoritma K-Means dan algoritma K-Medoids. Model ini memproses 88 data tanaman padi di Pulau Sumatera dari tahun 2010 sampai 2020 yang berisi hasil panen dan luas lahan.

K-Means

Proses klasterisasi pada penelitian ini menggunakan model K-Means. Percobaan klaster dilakukan dengan menggunakan jumlah klaster 2, 3, 4, 5, 6, 7, 8, 9, dan 10. Dari hasil analisis, klaster ke-3 memberikan hasil terbaik dalam memisahkan data secara optimal. Hasil dari klasterisasi tersebut dapat dilihat pada tabel dibawah ini.

Tabel 3. Hasil Pengelompokan klaster pertama K-Means

No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Sumatera Utara 2010	0.721	0.854	1
2.	Sumatera Utara 2011	0.726	0.858	1
3.	Sumatera Utara 2012	0.749	0.867	1
4.	Sumatera Utara 2013	0.751	0.840	1
5.	Sumatera Utara 2014	0.731	0.808	1
6.	Sumatera Utara 2015	0.820	0.888	1
7.	Sumatera Utara 2016	0.894	0.445	1
...
20.	Lampung 2012	0.617	0.715	1
21.	Lampung 2013	0.640	0.710	1
22.	Lampung 2014	0.664	0.723	1
23.	Lampung 2015	0.734	0.796	1
24.	Lampung 2016	0.774	0.405	1
25.	Lampung 2017	0.830	0.412	1

Tabel 4. Hasil Pengelompokan klaster kedua K-Means

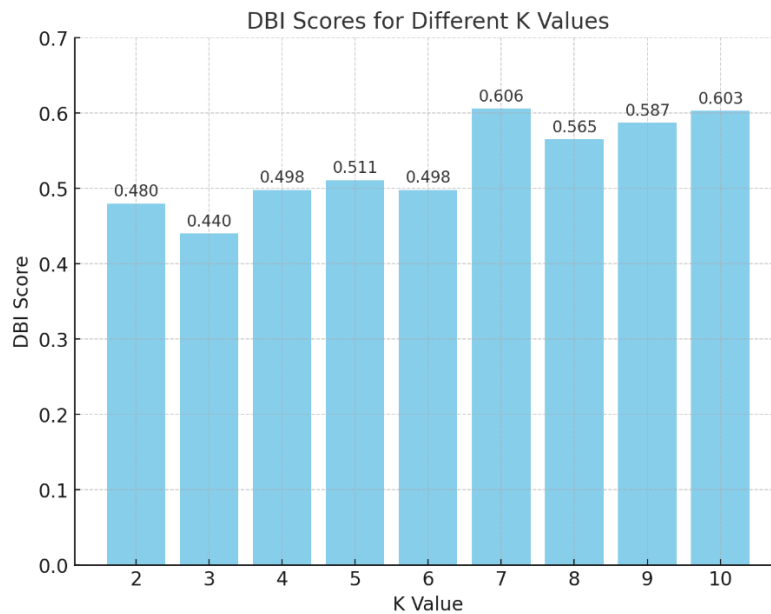
No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Aceh 2010	0.335	0.357	2
2.	Aceh 2011	0.332	0.392	2
3.	Aceh 2012	0.291	0.401	2
4.	Aceh 2013	0.452	0.440	2
5.	Aceh 2014	0.342	0.387	2
6.	Aceh 2015	0.371	0.492	2
7.	Aceh 2016	0.419	0.284	2
...
25.	Sumatera Barat 2020	0.262	0.287	2
26.	Sumatera Selatan 2019	0.510	0.588	2
27.	Sumatera Selatan 2020	0.530	0.603	2
28.	Lampung 2018	0.486	0.554	2
29.	Lampung 2019	0.416	0.495	2
30.	Lampung 2020	0.511	0.595	2

Tabel 5. Hasil Pengelompokan klaster ketiga K-Means

No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Riau 2010	0.074	0.115	3
2.	Riau 2011	0.066	0.101	3
3.	Riau 2012	0.060	0.100	3
4.	Riau 2013	0.044	0.068	3
5.	Riau 2014	0.0333	0.053	3
6.	Riau 2015	0.035	0.055	3
7.	Riau 2016	0.020	0.011	3
...
27.	Bengkulu 2014	0.078	0.104	3
28.	Bengkulu 2015	0.075	0.081	3
29.	Bengkulu 2016	0.086	0.025	3
30.	Bengkulu 2017	0.104	0.024	3
31.	Bengkulu 2018	0.012	0.003	3

32.	Bengkulu 2019	0.014	0.002	3
33.	Bengkulu 2020	0.014	0.001	3

Selanjutnya, dilakukan analisis terhadap 3 cluster tersebut untuk menentukan label wilayah produksi tinggi, sedang, dan rendah. Penentuan label ini didasarkan pada karakteristik data provinsi di setiap cluster. Setelah melakukan proses klusterisasi menggunakan model K-Means, validasi kluster akan dilakukan dengan menggunakan teknik DBI. Nilai DBI yang diperoleh berdasarkan hasil klusterisasi data menggunakan algoritma K-Means ditunjukkan pada Gambar 3.



Gambar 3. Nilai DBI K-Means

Berdasar pada nilai DBI terbaik, algoritma K-Means menunjukkan bahwa kluster yang optimal terjadi pada pengujian kalster 3, dengan nilai DBI sebesar 0,440. Pengujian ini membagi data menjadi 3 cluster, di mana cluster 1 memiliki jumlah sebanyak 25 wilayah untuk hasil produksi tinggi, cluster 2 memiliki total 30 wilayah untuk hasil produksi sedang, dan cluster 3 terdiri dari 33 wilayah untuk hasil produksi rendah. elemen,

K-Medoids

Proses pengelompokan dalam penelitian ini juga menggunakan algoritma K-Medoid, yang menerapkan uji kelompok yang sama seperti pada algoritma sebelumnya. Hasil dari klusterisasi dapat dilihat pada tabel dibawah.

Tabel 6. Hasil Pengelompokkan kluster pertama K-Medoids

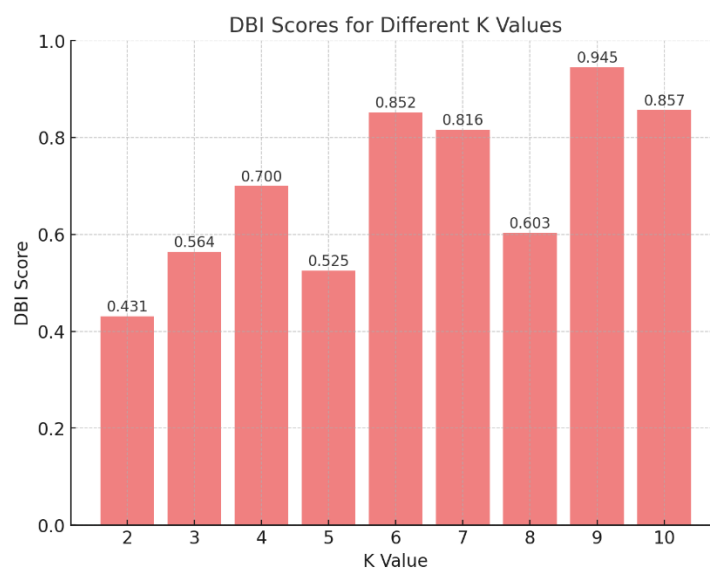
No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Sumatera Selatan 2014	0.740	0.924	1
2.	Sumatera Selatan 2015	0.864	1	1
3.	Sumatera Selatan 2016	1	0.682	1
4.	Sumatera Selatan 2017	0.984	0.690	1
5.	Sumatera Selatan 2018	0.594	0.640	1
6.	Sumatera Selatan 2019	0.510	0.588	1
7.	Sumatera Selatan 2020	0.530	0.603	1
...
50.	Sumatera Utara 2013	0.752	0.840	1
51.	Sumatera Utara 2014	0.731	0.808	1
52.	Sumatera Utara 2015	0.820	0.888	1
53.	Sumatera Utara 2016	0.894	0.445	1
54.	Sumatera Utara 2017	0.955	0.435	1

Tabel 7. Hasil Pengelompokkan kluster kedua K-Medoids

No.	Provinsi dan Tahun	Produksi	Luas Panen	Cluster
1.	Sumatera Barat 2020	0.262	0.287	2
2.	Riau 2010	0.074	0.115	2

3.	Riau 2011	0.066	0.101	2
4.	Riau 2012	0.060	0.100	2
5.	Riau 2013	0.044	0.068	2
6.	Riau 2014	0.033	0.053	2
7.	Riau 2015	0.035	0.055	2
...
29.	Bengkulu 2015	0.075	0.081	2
30.	Bengkulu 2016	0.086	0.025	2
31.	Bengkulu 2017	0.104	0.024	2
32.	Bengkulu 2018	0.012	0.003	2
33.	Bengkulu 2019	0.014	0.002	2
34.	Bengkulu 2020	0.014	0.001	2

Selain itu, validasi kluster juga dilakukan dengan menggunakan Davies Bouldin Index (DBI). Nilai DBI yang didapat dari hasil klusterisasi data dengan algoritma K-Medoid dilihat pada Gambar 4.

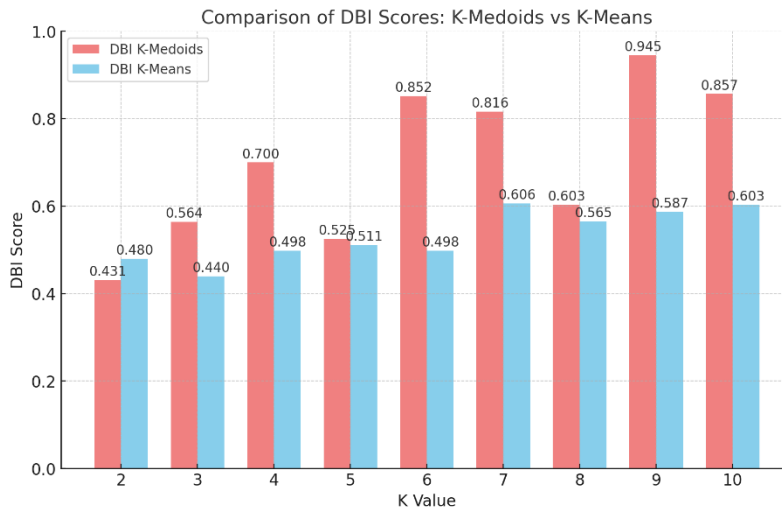


Gambar 4. Nilai DBI K-Medoids

Berdasarkan nilai DBI terbaik, algoritma K-Medoids menunjukkan bahwa cluster yang optimal terjadi pada pengujian kluster 2, dengan nilai DBI sebesar 0,431. Dengan kluster ke-2, data dibagi menjadi dua kelompok, di mana kelompok pertama berjumlah 54 wilayah dengan hasil produksi tinggi dan kelompok kedua terdiri dari 34 wilayah dengan hasil produksi rendah.

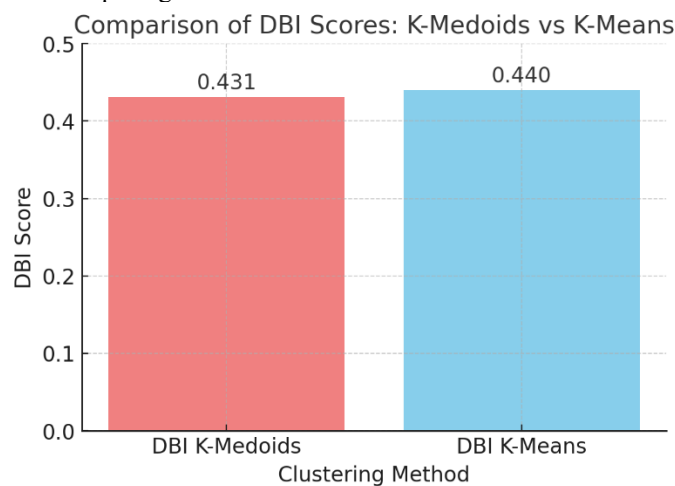
3.3 Perbandingan Kinerja Algoritma K-Means dan K-medoids

Perbandingan antara algoritma K-Means dan K-Medoids pada klusterisasi data tanaman padi di Pulau Sumatera dapat dilihat pada gambar 5.



Gambar 5. Perbandingan Nilai DBI K-Means dan K-Medoids

Berdasarkan hasil analisis dan pengolahan data, perbandingan antara algoritma K-Means dan K-Medoids menunjukkan perbedaan yang signifikan dalam kinerja keduanya. Algoritma K-Medoids berhasil menghasilkan cluster terbaik dengan nilai kluster ke-2, yang membagi data menjadi dua kelompok, yaitu kelompok 1 mencakup 54 wilayah dengan hasil produksi tinggi, sedangkan kelompok 2 terdiri dari 34 wilayah dengan hasil produksi rendah. Di sisi lain, algoritma K-Means menunjukkan hasil yang optimal pada K=3. Perbandingan DBI terbaik K-Means dan K-Medoids dapat dilihat pada gambar 6 dibawah ini:



Gambar 6. Perbandingan Nilai DBI Terbaik K-Means dan K-Medoids

Validasi klusterisasi dilakukan menggunakan Davies Bouldin Index (DBI), di mana K-Medoids menunjukkan nilai DBI terbaik sebesar 0,43, lebih rendah dibandingkan K-Means yang memiliki nilai DBI 0,440. Hal ini menegaskan bahwa K-Medoids lebih baik dalam mengelompokkan data dalam konteks penelitian ini, karena dapat menciptakan kluster yang lebih terpisah dan homogen.

4. DISKUSI

Berdasarkan hasil analisis klusterisasi pada data produksi tanaman padi di Pulau Sumatera dari tahun 2010 hingga 2020, penelitian ini membandingkan dua algoritma klusterisasi yang umum digunakan, yaitu K-Means dan K-Medoids. Proses klusterisasi dilakukan untuk mengelompokkan wilayah berdasarkan produksi padi mereka.

4.1 Perbandingan Dengan Penelitian Terdahulu

Hasil penelitian ini yang membahas klusterisasi data produksi padi di Pulau Sumatera menunjukkan bahwa algoritma K-Medoids dengan K=2 menghasilkan kluster terbaik, dengan nilai Davies Bouldin Index (DBI) sebesar 0,43. Di sisi lain, algoritma K-Means menghasilkan kluster terbaik dengan K=3, namun dengan nilai DBI yang sedikit lebih tinggi, yaitu 0,440. Temuan ini dapat dibandingkan dengan hasil penelitian yang serupa pada data persediaan di PT. Asia Petrocom Services, yang juga menggunakan algoritma K-Means dan K-Medoids untuk analisis klusterisasi.

Pada penelitian PT. Asia Petrocom Services, K-Medoids juga menghasilkan kluster terbaik dengan $K=2$, namun dengan nilai DBI yang jauh lebih rendah, yaitu 0,073. Sementara itu, K-Means menghasilkan kluster terbaik dengan $K=4$, menunjukkan bahwa pada data persediaan yang besar dan kompleks, K-Means cenderung memerlukan lebih banyak kluster untuk mencapai hasil yang optimal. Namun, walaupun jumlah kluster lebih banyak, K-Means tetap tidak mengungguli K-Medoids dalam hal kualitas klusterisasi, sebagaimana diukur dengan nilai DBI.

4.2 Keterbatasan Penelitian

Penelitian ini, meskipun berhasil memberikan hasil klusterisasi yang baik menggunakan algoritma K-Means dan K-Medoids, memiliki beberapa keterbatasan yang perlu diperhatikan seperti Data yang digunakan dalam penelitian ini terbatas pada periode 2010-2020 dan hanya mencakup wilayah di Pulau Sumatera. Dengan demikian, hasil penelitian ini tidak dapat digeneralisasi untuk seluruh wilayah di Indonesia atau untuk periode waktu yang lebih Panjang dan penelitian ini hanya berfokus pada data produksi padi tanpa mempertimbangkan faktor-faktor eksternal seperti cuaca, kebijakan pertanian, atau infrastruktur irigasi yang mungkin memiliki pengaruh signifikan terhadap produksi padi di setiap wilayah.

Meskipun penelitian ini memberikan hasil yang valid dan relevan, ada ruang untuk pengembangan lebih lanjut, terutama dalam hal variasi parameter, validasi yang lebih luas, serta penggunaan data dan algoritma yang lebih beragam. Hal ini akan memperkuat keakuratan hasil klusterisasi serta memperluas relevansi penelitian ini dalam konteks yang lebih luas.

5. KESIMPULAN

Berdasarkan hasil analisis dan pengolahan data tanaman padi di Pulau Sumatera dari tahun 2010 sampai tahun 2020 menggunakan algoritma K-Means dan K-Medoids, klasster terbaik diperoleh dengan algoritma K-Medoids pada nilai kluster ke-2. Sementara itu, algoritma K-Means menghasilkan kluster terbaik dengan kluster ke-3. Selain menghasilkan kluster yang lebih kompak, K-Medoids juga berhasil mengidentifikasi dua kelompok wilayah berdasarkan produksi padi. Hasil klusterisasi tersebut divalidasi menggunakan Davies Bouldin Index (DBI). Dalam penelitian ini, K-Medoids menunjukkan kinerja yang lebih baik dibandingkan K-Means, dengan nilai DBI terbaik sebesar 0,43 untuk K-Medoids dan 0,440 untuk K-Means. Pengujian K-Medoids dengan $K=2$ mengelompokkan data menjadi 2 kelompok, di mana kelompok 1 terdiri dari 54 wilayah dengan hasil produksi tinggi dan kelompok 2 terdiri dari 34 wilayah dengan hasil produksi rendah.

DAFTAR PUSTAKA

- [1] Iyan, R., (2020). ANALISIS KOMODITAS UNGGULAN SEKTOR PERTANIAN DI WILAYAH SUMATERA. *JURNAL SOSIAL EKONOMI PEMBANGUNAN*, 11(4), 215-235.
- [2] Yulianto, T., & Susilo, A. (2020). Penggunaan Algoritma K-Means untuk Segmentasi Data Pertanian di Sumatera. *Jurnal Teknik Pertanian Indonesia*, 11(2), 85-94.
- [3] Das, D., & Pradhan, S. (2020). Application of K-means and K-medoids Clustering Algorithms for Data Mining in Agriculture. *Journal of Data Mining and Knowledge Discovery*, 14(3), 214-228.
- [4] Wang, X., Chen, L., & Zhou, F. (2020). Comparative Study of K-Means and K-Medoids Clustering in Agricultural Data Analysis. *Computational Agriculture*, 19(2), 110-118.
- [5] Nurdin, I., & Fatimah, N. (2019). Implementing K-Means and K-Medoids for Agricultural Data Grouping. *Journal of Agricultural Data Science*, 6(4), 198-207.
- [6] Puspita, M., & Ningsih, S. (2021). Penggunaan Metode Clustering dalam Analisis Data Pertanian: Studi Kasus di Sumatera. *Jurnal Informatika dan Sains Data*, 15(4), 210-222.
- [7] Putra, I. R., & Wahyudi, A. (2020). Analisis Pengelompokan Data Produksi Padi di Sumatera Menggunakan Algoritma K-Medoids. *Jurnal Teknik Informatika dan Sistem Informasi*, 10(1), 45-55.
- [8] Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons.
- [9] Han, J., Pei, J., & Kamber, M. (2019). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- [10] Park, H. S., & Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, 36(2), 3336-3341.
- [11] Mubarak, H.A. Penerapan Algoritma K-Means dan K-Medoids Dalam Pengelompokan Data Inventaris Rig. *IJRSE: Indonesian Journal of Informatic Research and Software Engineering*. 3(2), 127-135.