

## Algoritma K-Medoids untuk Pengelompokan Produksi Padi dan Beras sebagai Upaya Optimalisasi Ketahanan Pangan di Provinsi Riau

Luth Fimawahib<sup>1</sup>, Imam Rangga Bakti<sup>2</sup>, Asep Supriyanto<sup>3</sup>

<sup>1</sup>Universitas Pasir Pengaraian, luthfimawahib@upp.ac.id, Rokan Hulu, Riau, Indonesia

<sup>2</sup>Universitas Pasir Pengaraian, imamranggabakti@gmail.com, Rokan Hulu, Riau, Indonesia

<sup>3</sup>Universitas Pasir Pengaraian, asep.tif@gmail.com, Rokan Hulu, Riau, Indonesia

### Informasi Makalah

Submit : Oktober 04, 2022  
Revisi : Oktober 27, 2022  
Diterima : Desember 01, 2022

### Kata Kunci :

*K-Medoids*  
Padi  
Beras  
Provinsi Riau

### Abstrak

Tanaman pangan seperti padi masih menjadi perhatian dunia, termasuk di negara Indonesia khususnya Provinsi Riau. Kendala yang dihadapi Provinsi Riau adalah terbatasnya pengetahuan dalam pengelompokan produksi padi dan beras. Data mining memiliki beberapa algoritma dalam *Clustering* data, salah satunya *K-Medoids*. Produksi padi dan beras di Provinsi Riau dalam penelitian ini dikelompokkan menggunakan algoritma *K-Medoids* dengan menetapkan  $k$  sejumlah 2 (dua). Perhitungan manual metode dan implementasi dalam RapidMiner untuk produksi padi dan beras dari tahun 2019 sampai 2021, maka diperoleh *Cluster* 1 (tinggi) sebanyak 4 kabupaten, sementara *Cluster* 2 (rendah) sebanyak 8 kabupaten. Hasil ini bisa menjadi informasi bagi pemerintah Provinsi Riau bahwa masih banyak kabupaten yang belum optimal dalam produksi padi dan beras, sehingga pemerintah bisa lebih fokus dalam meningkatkan produksi padi dan beras pada tahun-tahun berikutnya. Evaluasi *Cluster* menggunakan *Davies-Bouldin Index* (DBI) dan didapat nilai sebesar 0,626, sehingga bisa dikatakan bahwa evaluasi *Cluster* cukup baik, karena mendekati nol.

### Abstract

Food crops such as rice are still a world concern, including in Indonesia, especially Riau Province. The obstacle faced by Riau Province is the limited knowledge in classifying rice and rice production. Data mining has several algorithms in data clustering, one of which is K-Medoids. Rice and rice production in Riau Province in this study were grouped using the K-Medoids algorithm by setting  $k$  to 2 (two). Manual calculation of methods and implementation in RapidMiner for rice and rice production from 2019 to 2021, then obtained Cluster 1 (high) as many as 4 districts, while Cluster 2 (low) as many as 8 districts. This result can be information for the Riau Provincial government that there are still many districts that are not optimal in rice and rice production, so that the government can focus more on increasing rice and rice production in the following years. Cluster evaluation uses the Davies-Bouldin Index (DBI) and a value of 0.626 is obtained, so it can be said that the Cluster evaluation is quite good, because it is close to zero.

Luth Fimawahib, M.Kom  
Email: luthfimawahib@upp.ac.id

Direktorat Jenderal Pendidikan Tinggi,  
Riset, dan Teknologi - Kemdikbud

## 1. Pendahuluan

Ketahanan pangan selalu menjadi hal yang penting untuk diperhatikan, karena berkaitan dengan hajat hidup orang banyak. Dekade pertama pada abad 21 menunjukkan lambatnya produktivitas beras di seluruh dunia. Kemiskinan di dunia banyak dikaitkan dengan masalah pertanian padi. Sekitar 900 juta dari orang miskin di dunia bergantung pada beras, baik sebagai petani maupun konsumen (Mishra et al., 2022). Ketahanan pangan mempertimbangkan kemampuan manusia untuk menghasilkan pangan yang cukup setiap hari, menyediakan nutrisi dan meminimalkan dampak lingkungan. (Mukhtar et al., 2022) Pemerintah Indonesia, sejak dulu kala sudah memperhatikan hal ini. Bahkan, pada tahun 1984 Indonesia pernah swasembada beras. Namun, saat ini justru Indonesia adalah negara yang selalu mengimpor pangan, termasuk beras. Produksi padi dalam negeri dirasa kurang cukup. Oleh sebab itu, agar stok beras dalam negeri terjaga dan harga stabil, perlu melakukan impor.

Riau merupakan provinsi di Indonesia yang hampir seluruh wilayahnya bisa ditanami padi. Sebagian besar wilayah di provinsi ini adalah dataran rendah dan memiliki curah hujan yang baik, cocok untuk budidaya tanaman padi. Namun, produksi padi di Riau belum optimal, karena sebagian besar wilayah di provinsi ini banyak dimanfaatkan untuk perkebunan kelapa Sawit.

Peraturan terbaru yang berkaitan dengan ketahanan pangan ini sebagaimana tertuang dalam Peraturan Presiden Nomor 104 Tahun 2021 tentang Rincian APBN Tahun Anggaran 2022 pada Pasal 5 ayat (4) huruf b bahwa penggunaan Dana Desa yaitu paling sedikit 20% Penggunaan Dana Desa digunakan untuk Program Ketahanan Pangan dan Hewani (Presiden Republik Indonesia, 2021). Adapun secara konsep, ketahanan pangan sebagaimana Undang-Undang Nomor

18 Tahun 2012 tentang Ketahanan Pangan adalah kondisi terpenuhinya pangan bagi negara sampai dengan perseorangan, yang tercermin dari tersedianya pangan yang cukup, baik jumlah maupun mutunya, aman, beragam, bergizi, merata, dan terjangkau serta tidak bertentangan dengan agama, keyakinan, dan budaya masyarakat untuk dapat hidup sehat, aktif, dan produktif secara berkelanjutan. (Undang-Undang, 2012)

*Medoid* adalah algoritma yang diwakili oleh *cluster*, dimana algoritma ini diistilahkan sebagai algoritma *K-Medoids* atau Algoritma PAM (*Partitioning Around Medoids*). (Kamila et al., 2019) Kelemahan metode *K-Means* yaitu ketika objek dengan nilai yang besar, bisa jadi distribusi data bisa menyimpang secara substansial, diperbaiki dengan hadirnya metode *K-Medoids* ini. Pengelompokan sekumpulan  $n$  objek dijadikan ke dalam beberapa  $k$  *cluster*, dengan menggunakan metode pengelompokan partisi, adalah konsep dari *K-Medoids*.

Penelitian sebelumnya yang menerapkan *K-Medoids* adalah penelitian S. Ani Ritonga, et al yaitu mengelompokkan produktivitas padi menurut provinsi dengan hasil *cluster* 1 atau rendah = 17 provinsi, *cluster* 2 atau sedang = 7 provinsi dan *cluster* 3 atau tinggi = 3 provinsi. (Ani Ritonga et al., 2019) Penelitian dengan metode sama juga dilakukan oleh Cepy Sukmayadi, et al yaitu menentukan wilayah di Kabupaten Karawang yang rawan banjir dengan hasil penelitian 1 kecamatan potensi banjir rendah, 24 kecamatan potensi banjir sedang, dan 5 kecamatan potensi banjir tinggi. (Sukmayadi et al., 2021) Penelitian lain dengan objek padi adalah penelitian A. K. Mishra, et al dengan hasil penelitian varietas padi di Asia dan Afrika yang mampu bertahan dalam kondisi stres, secara signifikan meningkatkan hasil dan pendapatan padi. Selain itu, inovasi kelembagaan, pelatihan, dan praktik pengelolaan sumber daya alam, seperti padi, benih langsung, pengendalian hewan

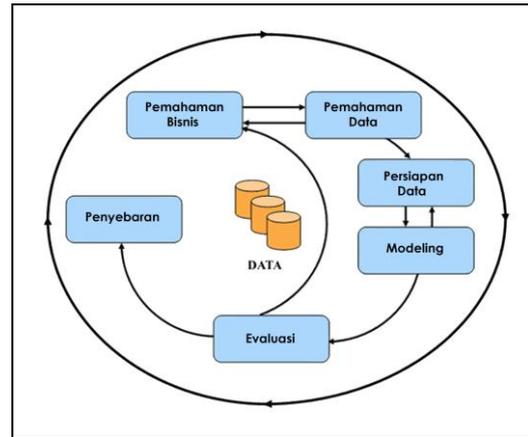
pengerat, dan penghilangan toksisitas besi, telah memiliki dampak yang cukup besar berpengaruh positif terhadap kesejahteraan petani padi sawah (pendapatan dan hasil padi. (Mishra et al., 2022)

Penelitian ini bertujuan memahami algoritma *K-Medoids* dalam mengelompokkan jumlah produksi padi dan beras di Provinsi Riau. Kebaharuan dalam penelitian ini adalah mampu memanfaatkan data mining dalam mengungkap informasi tersembunyi yang selama ini belum diketahui oleh pemerintah provinsi. Ini menjadi hal yang unik dalam kajian penelitian ini, karena dengan data yang selama ini sudah tersaji, ternyata masih bisa diungkap sebuah informasi lain di dalamnya. Data yang diteliti lebih mengerucut di tingkat provinsi sehingga hasil penelitian lebih optimal untuk dijadikan sumber rujukan dalam mengambil kebijakan. Manfaat dari penelitian ini sebagai acuan informasi bagi pemerintah Provinsi Riau dalam mengoptimalkan jumlah produksi padi dan beras sehingga ketahanan pangan bisa dikelola dengan baik. Jika stok produksi padi dan beras cukup, tentu ini berkontribusi untuk menekan jumlah impor.

## 2. Metode Penelitian

### 2.1. Metode dan Tahap Penelitian

Metode yang diterapkan untuk mengelompokkan produksi padi dan beras di Provinsi Riau adalah Algoritma *K-Medoids*. Adapun RapidMiner adalah *tool* yang digunakan untuk pengujian atau implementasi metode. Jumlah *Cluster* ditentukan sebanyak 2 (dua), yaitu tinggi dan rendah. Langkah-langkah metodologi CRISP-DM dalam penelitian yaitu:



Gambar 1. Tahapan CRISP-DM

1. Proses pemahaman bisnis : Proses pemahaman bisnis berfokus pada tujuan dan persyaratan proyek, yang terdiri dari pemahaman tujuan bisnis, kriteria keberhasilan, rencana proyek, dan pengiriman. Penelitian ini mengimplementasikan metode *K-Medoids* dalam mengelompokkan produksi padi dan beras di Provinsi Riau.
2. Proses pemahaman data : Proses pengumpulan data awal dan pengelolaan dilanjutkan dengan deskripsi data dan eksplorasi data.
3. Proses persiapan data : Proses persiapan data meliputi pembersihan data, pengambilan sampel, normalisasi, dan pemilihan fitur yang sudah ditetapkan.
4. Proses pemodelan : Pemodelan meliputi pemilihan teknik pemodelan, membangun, dan melatih model, selain membuat prediksi. Dalam pengelompokan produksi padi dan beras di Provinsi Riau, *K-Medoids* sebagai model yang digunakan dalam penelitian ini.
5. Proses Evaluasi : Proses evaluasi meliputi validasi model, review hasil, dan evaluasi kriteria keberhasilan.
6. Proses Penyebaran : Proses penyebaran mencakup visualisasi hasil, dan pembuatan laporan.

## 2.2. Tempat dan Waktu Penelitian

Tempat penelitian dilakukan di Badan Pusat Statistik Provinsi Riau. Pelaksanaan penelitian pada bulan Juni 2022 sampai dengan September 2022.

## 2.3. Data

Penelitian ini menggunakan data Badan Pusat Statistik hasil Produksi Padi dan Beras di Provinsi Riau. Data terdiri dari 12 (dua belas) kabupaten dari tahun 2019 sampai 2021 dengan kriteria Produksi Padi (ton GKG) dan Produksi Padi Setara Beras (ton).

## 2.4 Data Mining

Data mining (DM) adalah area penelitian di mana kumpulan data besar dalam basis data dan penyimpanan data berada dijelajahi dan ditambang untuk menemukan pola baru dan berguna. (Bakar et al., 2020). Proses penggalian data diumpamakan seperti penambangan minyak di laut. Minyak adalah pengetahuan baru yang diperoleh, sementara laut adalah *data base*. (Fimawahib & Rouza, 2021). Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa mengetahui yang selama ini tidak diketahui secara manual. (Pratiwi & Adrianto, 2017). Konsep lain dari data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan. (Desyanti, 2019)

## 2.5 CRISP-DM

*CRISP-DM* merupakan kepanjangan dari *Cross-Industry Standard Process for Data Mining* merupakan suatu standarisasi pemrosesan data mining yang telah dikembangkan dimana data yang ada akan melewati setiap fase terstruktur dan terdefenisi dengan jelas dan efisien. (Fimawahib & Rouza, 2021). Proses pemahaman bisnis berfokus pada tujuan dan persyaratan proyek, yang terdiri dari pemahaman tujuan bisnis, kriteria

keberhasilan, rencana proyek, dan pengiriman. Proses pemahaman data dimulai dengan pengumpulan data awal dan pengelolaan dilanjutkan dengan deskripsi data dan eksplorasi data. Proses persiapan data meliputi pembersihan data, pengambilan sampel, normalisasi, dan pemilihan fitur. Proses pemodelan meliputi pemilihan teknik pemodelan, membangun, dan melatih model, selain membuat prediksi. Proses evaluasi meliputi validasi model, review hasil, dan evaluasi kriteria keberhasilan. Terakhir, proses penerapan mencakup visualisasi hasil, dan pembuatan laporan. (Utomo, 2021)

## 2.6 Clustering

*Clustering* diartikan sebagai pengelompokan, yaitu suatu proses menempatkan data yang sama ke dalam sebuah kelompok-kelompok berdasarkan pertimbangan pendekatan penting. (Kamila et al., 2019)

## 2.7 Proses Algoritma K-Medoids

*Medoid* adalah algoritma yang diwakili oleh *cluster*, dimana algoritma ini diistilahkan sebagai algoritma *K-Medoids* atau Algoritma PAM (*Partitioning Around Medoids*). (Kamila et al., 2019) Kelemahan metode *K-Means* yaitu ketika objek dengan nilai yang besar, bisa jadi distribusi data bisa menyimpang secara substansial, diperbaiki dengan hadirnya metode *K-Medoids* ini. Pengelompokan sekumpulan  $n$  objek dijadikan ke dalam beberapa  $k$  *cluster*, dengan menggunakan metode pengelompokan partisi, adalah konsep dari *K-Medoids*. (Sukmayadi et al., 2021) Algoritma *K-Medoids* merupakan bagian dari group clustering. Metode *K-Medoids* cukup efisien dalam mengumpulkan data yang kecil. Langkah pertama adalah menentukan titik (*medoid*) yang paling representatif pada kelompok data dengan menghitung jarak dalam suatu cluster dari semua kombinasi *medoid* sehingga jarak antar titik dalam suatu kelompok kecil, sedangkan jarak titik antar kelompok adalah

besar. (Utomo, 2021) Langkah-langkah algoritma K-Medoids adalah sebagai berikut:

- A. Inisialisasi pusat cluster sebanyak  $k$  (jumlah cluster)
- B. Gunakan persamaan ukuran jarak *Euclidian Distance* untuk mengalokasikan setiap data (objek) ke cluster terdekat.

$$d_{ij} = \frac{\sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2}}{\sqrt{(x_i - x_j)'(x_i - x_j)}} \quad (1)$$

dimana  $i=1, \dots, n$ ;  $j=1, \dots, n$  dan  $p$  adalah banyak variabel, serta  $V$  adalah matrik varian kovarian.

- C. Kandidat *medoids* baru dipilih secara acak objek pada masing-masing *cluster*.
- D. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru.
- E. Hitung total simpangan ( $S$ ) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika  $S < 0$ , maka tukar objek dengan data *cluster* untuk membentuk sekumpulan  $k$  objek baru sebagai *medoids*.  
Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoids*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

## 2.8 Ketahanan Pangan

Pangan adalah segala sesuatu yang berasal dari sumber hayati dan air, baik yang diolah maupun tidak diolah, yang diperuntukkan sebagai makanan atau minuman bagi konsumsi manusia, termasuk bahan tambahan pangan, bahan baku pangan, dan bahan lainnya yang digunakan dalam proses penyiapan, pengolahan, dan atau pembuatan makanan atau minuman. Adapun, ketahanan pangan adalah kondisi terpenuhinya pangan bagi negara sampai dengan perseorangan, yang tercermin dari tersedianya pangan yang

cukup, baik jumlah maupun mutunya, aman, beragam, bergizi, merata, dan terjangkau serta tidak bertentangan dengan agama, keyakinan, dan budaya masyarakat untuk dapat hidup sehat, aktif, dan produktif secara berkelanjutan. (Fauzin, 2021)

## 2.9 Padi

Salah satu sumber karbohidrat utama bagi tubuh manusia mayoritas penduduk dunia adalah padi. Padi masuk dalam suku padi-padian atau *poaceae*. Padi termasuk tanaman semusim, memiliki akar serabut, batang pendek, struktur serupa batang terbentuk dari rangkaian pelepah daun yang saling menopang daun sempurna dengan pelepah tegak, daun berbentuk lanset, warna hijau muda hingga hijau tua, berurat daun sejajar, tertutupi oleh rambut yang pendek dan jarang, bagian bunga tersusun majemuk, tipe malai bercabang, satuan bunga disebut *floret* yang terletak pada satu spikelet yang duduk pada panikula, tipe buah bulir atau kariopsis yang tidak dapat dibedakan mana buah dan bijinya, bentuk hampir bulat hingga lonjong, ukuran 3 mm hingga 15 mm, tertutup oleh palea dan lemma yang dalam bahasa sehari-hari disebut sekam, struktur dominan padi yang biasa dikonsumsi yaitu jenis *enduspermium* yang berubah-ubah. (Khusna & Mariana, 2021) Ekonomi di Indonesia bisa diukur dengan indikator banyaknya komoditas unggulan berupa padi sebagai bahan baku pangan pokok vital untuk rakyat. (Ani Ritonga et al., 2019)

## 2.10 RapidMiner

Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer merupakan pengembang RapidMiner di *Artificial Intelligence Unit* dari University of Dortmund. Aplikasi ini adalah aplikasi *open source* berlisensi AGPL (*GNU Affero General Public License*) yang dapat digunakan untuk mengolah data mining. (Setiawan, 2016)

### 3. Hasil dan Pembahasan

#### A. Pemahaman Bisnis

Data produksi padi dan beras di Provinsi Riau sudah ada namun informasi tersembunyi dalam data tersebut belum diungkap. Pengumpulan data diperoleh dari Badan Pusat Statistik Provinsi Riau. Hasil yang diharapkan dari luaran atau *output* penelitian ini adalah informasi bermanfaat untuk Provinsi Riau terkait pengembangan daerah yang paling berpotensi untuk budidaya padi sebagai upaya ketahanan pangan.

#### B. Pemahaman Data

*Dataset* awal dalam penelitian adalah laporan Produksi Padi dan Beras dalam ton sepanjang tahun 2019 sampai 2021. Ketika data awal sudah dikumpulkan, pengelolaan dilanjutkan dengan deskripsi data dan eksplorasi data.

Tabel 1. Deskripsi Data Produksi Padi dan Beras

Atribut	Tipe Data	Keterangan
No	Numerik	No Urut Kab/Kota
Kab/Kota	Nominal	Nama Kabupaten/Kota
Produksi Padi (ton GKG)	Numerik	Jumlah Produksi Padi
Produksi Padi setara Beras (ton)	Numerik	Jumlah Produksi Beras
Tahun	Numerik	Tahun Produksi

Tabel I mendeskripsikan data produksi padi dan beras di Provinsi Riau. Tabel I memberikan informasi tentang atribut, tipe data, dan keterangan. Jika deskripsi data sudah dilakukan, maka proses berikutnya adalah *explore the data*, dengan tujuan mudah untuk memahami data agar diperoleh pengetahuan atau pemahaman data. Setelah data diexplore, tahap berikutnya adalah melakukan verifikasi menggunakan *Microsoft Excel* dengan fitur *Conditional Formatting* untuk memeriksa kualitas data.

Tabel 2. Hasil Explore Data Produksi Padi dan Beras

Atribut	Tipe Data	Keterangan
No	Numerik	No Urut Kab/Kota Data 2019-2021: 1-12
Kab/Kota	Nominal	Nama Kab/Kota Data 2019-2021: 12 Kab/Kota
Produksi Padi (ton GKG)	Numerik	Jumlah Produksi Padi Data 2019-2021: dalam ton
Produksi Padi setara Beras (ton)	Numerik	Jumlah Produksi Beras Data 2019-2021: dalam ton
Tahun	Numerik	Tahun Produksi : 2019, 2020, 2021

Tidak ditemukan *missing value* atau hilangnya nilai/data dan data ganda sebagaimana yang terlihat pada Tabel II Explore Data Produksi Padi dan Beras.

#### C. Persiapan Data

Tabel III menampilkan data produksi padi dan beras sesudah proses pembersihan atau *cleaning* dan masih memiliki 8 atribut.

Tabel 3. Explore Data Produksi Padi dan Beras Cleaning

No	Kab/Kota	Produksi Padi (ton GKG)			Produksi Padi Setara Beras (ton)		
		2019	2020	2021	2019	2020	2021
1	Kuantan Singingi	19.322	27.198	20.950	11.032	15.609	12.023
2	Indragiri Hulu	7.842	6.578	4.443	4.478	3.775	2.550
3	Indragiri Hilir	70.139	78.459	65.755	40.046	45.028	37.737
4	Pelalawan	16.897	15.155	19.001	9.647	8.698	10.905
5	Siak	28.292	26.549	29.382	16.153	15.236	16.863
6	Kampar	9.568	18.947	9.742	5.463	10.874	5.591

7	Rokan Hulu	11.757	11.638	5.926	6.713	6.679	3.401
8	Bengkalis	21.574	13.456	14.997	12.318	7.723	8.607
9	Rokan Hilir	39.557	37.209	37.724	22.585	21.354	21.650
10	Kepulauan MeraNTI	4.834	4.891	6.882	2.760	2.807	3.950
11	Pekanbaru	1	0	0	1	0	0
12	Dumai	1.091	3.606	2.663	623	2.069	1.525

### Implementasi K-Medoids

Implementasi *K-Medoids* sesuai dengan studi kasus dalam penelitian ini yaitu:

- Pusat *cluster* yang diinisialisasikan sebanyak 2 (dua), dipilih secara random untuk setiap *medoid*. Tabel IV menampilkan Pusat *Cluster Medoid* I:

Tabel 4. Pusat Cluster Medoid 1

No	Kab/Kota	Produksi Padi (ton GKG)			Produksi Padi Setara Beras (ton)		
		2019	2020	2021	2019	2020	2021
1	Rokan Hilir	39.57	37.09	37.24	22.85	21.54	21.50
		.5	.2	.7	.5	.3	.6
2	Dumai	1.09	3.60	2.65	623	2.06	1.52
		1	6	7	3	9	5

- Gunakan persamaan ukuran jarak *Euclidian Distance* untuk mengalokasikan setiap data (objek) ke cluster terdekat:

$$\begin{aligned}
 & D(\text{Kuantan Singingi}) \\
 &= \sqrt{(19.322 - 39.557)^2 + (27.198 - 37.209)^2 +} \\
 & \quad (20.950 - 37.724)^2 + (11.032 - 22.585)^2 +} \\
 & \quad (15.609 - 21.354)^2 + (12.023 - 21.650)^2} \\
 & \quad = 32.407 \quad (1)
 \end{aligned}$$

$$\begin{aligned}
 & D(\text{Indragiri Hulu}) \\
 &= \sqrt{(7.942 - 39.557)^2 + (6.578 - 37.209)^2 +} \\
 & \quad (4.443 - 37.724)^2 + (4.478 - 22.585)^2 +} \\
 & \quad (3.775 - 21.354)^2 + (2.550 - 21.650)^2} \\
 & \quad = 63.667 \quad (2)
 \end{aligned}$$

$$\begin{aligned}
 & D(\text{Dumai}) \\
 &= \sqrt{(1.091 - 39.557)^2 + (3.606 - 37.209)^2 +} \\
 & \quad (2.657 - 37.724)^2 + (623 - 22.585)^2 + (2.069} \\
 & \quad - 21.354)^2 + (1.525 - 21.650)^2} \\
 & \quad = 71.398 \quad (3)
 \end{aligned}$$

Perhitungan ini terus dilakukan sampai data kabupaten/kota ke 12. Tabel V menampilkan hasil data setelah dilakukan perhitungan.

Tabel 5. Hasil Perhitungan

No	Kab/Kota	Hasil Perhitungan C1	Hasil Perhitungan C2
1	Kuantan Singingi	32.407	40.317
2	Indragiri Hulu	63.667	8.742
3	Indragiri Hilir	67.429	138.068
4	Pelalawan	42.349	29.388
5	Siak	20.294	51.286
6	Kampar	51.737	21.792
7	Rokan Hulu	56.905	15.837
8	Bengkalis	43.190	29.795
9	Rokan Hilir	-	71.398
10	Kepulauan Meranti	65.205	6.670
11	Pekanbaru	76.204	5.315
12	Dumai	71.398	-

Adapun hasil keseluruhan dapat dilihat pada tabel VI berikut ini:

Tabel 6. Hasil Perhitungan Algoritma K-Medoids Iterasi ke-1

No	Kab/Kota	Jarak ke medoids		Terdekat	Klaster yang diikuti
		C1	C2		
1	Kuantan Singingi	32.407	40.317	32.407	1

i					
2	Indragiri Hulu	63.6 67	8.742	8.742	2
3	Indragiri Hilir	67.4 29	138.0 68	67.429	1
4	Pelalawan	42.3 49	29.38 8	29.388	2
5	Siak	20.2 94	51.28 6	20.294	1
6	Kampar	51.7 37	21.79 2	21.792	2
7	Rokan Hulu	56.9 05	15.83 7	15.837	2
8	Bengkalis	43.1 90	29.79 5	29.795	2
9	Rokan Hilir	-	71.39 8	-	1
10	Kepulauan Meranti	65.2 05	6.670	6.670	2
11	Pekanbaru	76.2 04	5.315	5.315	2
12	Dumai	71.3 98	-	-	2
<b>Tota</b>		<b>237.66</b>		<b>7</b>	

- c. Kandidat *medoids* baru dipilih secara acak objek pada masing-masing *cluster*.

Tabel 7. Pusat Cluster Medoids II

No	Kab/Kota	Produksi Padi (ton GKG)			Produksi Padi Setara Beras (ton)		
		2019	2020	2021	2019	2020	2021
1	Indragiri Hulu	7.	6.	4.	4.	3.	2.
		84	57	44	47	77	55
2	Kampar	9.	18	9.	5.	10	5.
		56	.9	74	46	.8	59
		8	47	2	3	74	1

- d. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru. Penentuan jarak dari setiap objek pada iterasi ke-2 ini masih menggunakan langkah-langkah yang sama seperti sebelumnya. Berikut ini hasil perhitungan iterasi 2

Tabel 8. Hasil Perhitungan 2

No	Kab/Kota	Hasil Perhitungan C1	Hasil Perhitungan C2
1	Kuantan Singingi	33.199	19.587
2	Indragiri Hulu	-	15.642
3	Indragiri Hilir	130.430	117.234
4	Pelalawan	22.096	14.293
5	Siak	43.722	32.473
6	Kampar	15.642	-
7	Rokan Hulu	7.568	9.835
8	Bengkalis	21.472	16.369
9	Rokan Hilir	63.667	51.737
10	Kepulauan Meranti	4.868	17.414
11	Pekanbaru	12.856	26.921
12	Dumai	8.742	21.792

Hingga keseluruhan data sebagaimana tabel 9 di bawah ini

Tabel 9. Hasil Perhitungan Algoritma K-Medoids Iterasi ke-2

No	Kab/Kota	Jarak ke medoids		Terdekat	Klaster yang diikuti
		C1	C2		
1	Kuantan Singingi	33.19	19.58	19.587	2
		9	7		
2	Indragiri Hulu	-	15.64	-	1
3	Indragiri Hilir	130.4	117.2	117.23	2
4	Pelalawan	22.09	14.29	14.293	2
5	Siak	43.72	32.47	32.473	2
6	Kampar	15.64	-	-	2
7	Rokan Hulu	7.568	9.835	7.568	1
8	Bengkalis	21.47	16.36	16.369	2
9	Rokan Hilir	63.66	51.73	51.737	2
10	Kepulauan	4.868	17.41	4.868	1
			4		

Meranti				
1	Pekanb	12.85	26.92	12.856
1	aru	6	1	1
1	Dumai	8.742	21.79	8.742
2			2	1
<b>Total</b>			<b>285.72</b>	<b>8</b>

e. Hitung total simpangan ( $S$ ) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika  $S < 0$ , maka tukar objek dengan data *cluster* untuk membentuk sekumpulan  $k$  objek baru sebagai *medoids*.

$$\begin{aligned}
 S &= \text{total jarak terdekat baru} - \text{total jarak terdekat lama} \\
 &= 285.728 - 237.667 \\
 &= 48.061
 \end{aligned}$$

Total simpangan memiliki nilai lebih dari 0, maka iterasi yang digunakan adalah iterasi pertama. Hasil cluster tersebut yaitu *cluster* 1 sebanyak 4 kab/kota, *cluster* 2 sebanyak 8 kab/kota.

#### D. Pemodelan

Jika data sudah selesai dipersiapkan, kemudian pengujian dengan menghitung secara manual telah dilakukan, maka langkah berikutnya adalah pemodelan (*modeling*). RapidMiner merupakan aplikasi yang dimanfaatkan untuk modeling dengan algoritma *K-Medoids*. Dari 12 data kab/kota di Provinsi Riau, dibentuklah 2 (dua) *cluster* untuk produksi padi dan beras yaitu tinggi dan rendah.

## Model Klaster

**Klaster 0 : 4 buah**

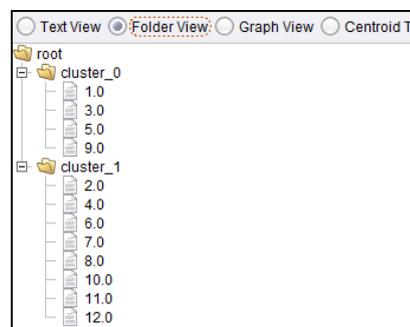
**Klaster 1 : 8 buah**

**Jumlah total : 12 buah**

Gambar 1. Rapid miner menampilkan klaster model

Gambar 1 menampilkan hasil pengelompokan (*cluster*) dengan pengujian

RapidMiner. Kelompok (*cluster*) pertama berjumlah 4 *items*, dan kelompok (*cluster*) kedua berjumlah 8 *items*.



Gambar 2. Rapid miner menampilkan cluster di folder

Kesimpulan yang bisa didapatkan dari produksi padi dan beras di Provinsi Riau adalah:

1. Cluster 1 (*Tinggi*) adalah Kabupaten Kuantan Singingi, Indragiri Hilir, Siak, dan Rokan Hilir
2. Cluster 2 (*Rendah*) adalah Kabupaten Indragiri Hulu, Pelalawan, Kampar, Rokan Hulu, Bengkalis, Kepulauan Meranti, Pekanbaru, Dumai.

Adapun dasar penentuan sebuah *cluster* adalah hasil perhitungan dengan angka paling kecil ataupun mendekati nilai titik pusat (*centroid*).

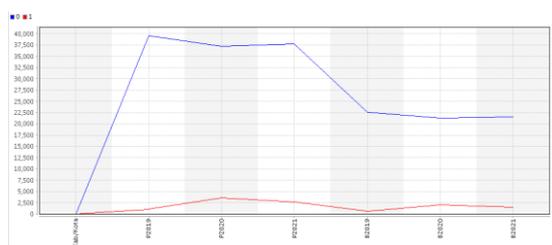
Tabel 10. Hasil Centroid Akhir

No	Atribut	Cluster 1	Cluster 2
1	Kab/Kota	Rokan Hilir	Dumai
2	P2019	39.557	1.091
3	P2020	37.209	3.606
4	P2021	37.724	2.657
5	B2019	22.585	623
6	B2020	21.354	2.069
7	B2021	21.650	1.525

Tabel X merupakan *Centroid* final pada studi kasus penelitian ini. Dari Tabel, terlihat *Cluster* 1 untuk produksi padi dan beras merupakan Kabupaten Rokan Hilir, dengan data produksi padi dari tahun 2019, 2020 dan 2021 adalah 39.557 ton, 37.209 ton, 37.724 ton, sementara produksi beras dari tahun

2019, 2020 dan 2021 adalah 22.585 ton, 21.354 ton, 21.650 ton. *Cluster 2* produksi padi dan beras adalah Kota Dumai, dengan data produksi padi dari tahun 2019, 2020 dan 2021 adalah 1.091 ton, 3.606 ton, 2.657 ton, sementara produksi beras dari tahun 2019, 2020 dan 2021 adalah 623 ton, 2.069 ton, 1.525 ton.

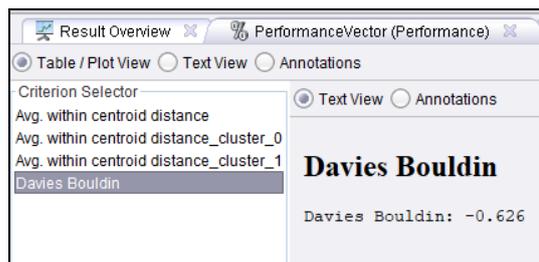
Gambar 3 di bawah menampilkan grafik hasil proses *K-Medoids Centroid final*. Warna biru adalah garis yang merepresentasikan *Cluster 1*, sedangkan warna merah adalah garis yang merepresentasikan *Cluster 2*. *Cluster 1* garis dengan warna biru, puncak tertinggi grafik ada pada atribut P2019 yaitu produksi padi pada tahun 2019, *Cluster 2* garis berwarna merah, atribut P2020 adalah puncak tertinggi grafik yaitu produksi padi pada tahun 2020.



Gambar 3. *K-Means clustering* dalam tampilan grafik

#### E. Evaluasi

DBI atau *Davies-Bouldin Index* adalah evaluasi yang digunakan. Adapun *Cluster distance performance* merupakan operator pada aplikasi RapidMiner yang digunakan untuk hasil uji *performance*. Di satu sisi, operator ini juga bisa mengukur seberapa baik kinerja dari *centroid* yang dihasilkan. Dari hasil pengujian RapidMiner terhadap data produksi padi dan beras, nilai DBI yang diperoleh adalah 0,626 seperti terlihat pada Gambar 4 tampilan uji *performance* di atas. Dari nilai ini, evaluasi *Cluster* adalah baik, karena mendekati nol. Semakin kecil DBI, maka evaluasi terhadap *Cluster* dikatakan baik.



Gambar 4. Tampilan uji *performance*

#### F. Penyebaran

Penyebaran pada penelitian ini maksudnya adalah menyampaikan hasil penelitian kepada pihak terkait. Penyebaran dilakukan dengan cara mempresentasikan dan mendemokan aplikasi yang dipakai, menggunakan aplikasi RapidMiner. Aplikasi dekstop merupakan *tools* yang digunakan untuk pengujian. User merupakan komponen yang turut serta terlibat dalam proses pengujian metode *K-Medoids* untuk pengelompokan produksi padi dan beras di Provinsi Riau.

### 4. Simpulan

Penelitian ini adalah mengelompokkan produksi padi dan beras di Provinsi Riau menggunakan metode *K-Medoids* dengan jumlah *k* sebanyak 2 (dua). Dari perhitungan manual dan uji data menggunakan RapidMiner pada 12 kabupaten yang ada di Provinsi Riau, produksi padi dan beras dari tahun 2019 sampai 2021 dengan *Cluster 1* (tinggi) sebanyak 4 kabupaten, sementara *Cluster 2* (rendah) sebanyak 8 kabupaten. Evaluasi *Cluster* yang digunakan adalah DBI atau *Davies-Bouldin Index*. Hasil evaluasi adalah 0,626, sehingga bisa dikatakan bahwa evaluasi *Cluster* cukup baik, karena semakin kecil DBI dan tidak bernilai negatif, semakin baik evaluasi sebuah *Cluster*.

### 5. Referensi

Ani Ritonga, S., Safii, M., Parlina, I., Satria Tambunan, H., Studi Sistem Informasi, P., & Tunas Bangsa Pematangsiantar Jln Jendral Sudirman Blok No, S. A.

- (2019). *Prosiding Seminar Nasional Riset Information Science (SENARIS) Teknik Data Mining dalam Mengelompokkan Produktivitas Padi Menurut Provinsi Menggunakan K-Medoids*.
- Bakar, W. A. W. A., Man, M., Man, M., & Abdullah, Z. (2020). i-Eclat: Performance enhancement of Eclat via incremental approach in frequent itemset mining. *Telkomnika (Telecommunication Computing Electronics and Control)*, 18(1), 562–570.  
<https://doi.org/10.12928/TELKOMNIK.A.V18I1.13497>
- Desyanti, D.-. (2019). Penerapan Data Mining Algoritma C4.5 untuk Mengetahui Tingkat Kepuasan Konsumen di Hotel Grand Zuri Dumai. *SATIN - Sains Dan Teknologi Informasi*, 4(2), 36.  
<https://doi.org/10.33372/stn.v4i2.403>
- Fauzin, F. (2021). Pengaturan Impor Pangan Negara Indonesia Yang Berbasis Pada Kedaulatan Pangan. *Pamator Journal*, 14(1), 1–9.  
<https://doi.org/10.21107/pamator.v14i1.10497>
- Fimawahib, L., & Rouza, E. (2021). Penerapan K-Means Clustering pada Penentuan Jenis Pembelajaran di Universitas Pasir Pengaraian. *INOVTEK Polbeng - Seri Informatika*, 6(2), 234.  
<https://doi.org/10.35314/isi.v6i2.2096>
- Indonesia. (2012). Undang-Undang Nomor 18 Tahun 2012 tentang Pangan. Lembaran Negara Republik Indonesia Tahun 2012 Nomor 227.
- Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119.  
<https://doi.org/10.24014/rmsi.v5i1.7381>
- Khusna, I. M., & Mariana, N. (2021). Sistem Pendukung Keputusan Pemilihan Bibit Padi Berkualitas Dengan Metode AHP Dan Topsis. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 10(2), 162–169.  
<https://doi.org/10.32736/sisfokom.v10i2.1145>
- Mishra, A. K., Pedde, V. O., Arouna, A., Labarta, R., Andrade, R., Veetil, P. C., Bhandari, H., Laborte, A. G., Balie, J., & Bouman, B. (2022). Helping feed the world with rice innovations: CGIAR research adoption and socioeconomic impact on farmers. *Global Food Security*, 33(March), 100628.  
<https://doi.org/10.1016/j.gfs.2022.100628>
- Mukhtar, Majahar Ali, M. K., Tahir Ismail, M., Hamundu, F. M., Alimuddin, Akhtar, N., & Fudholi, A. (2022). Hybrid model in machine learning–robust regression applied for sustainability agriculture and food security. *International Journal of Electrical and Computer Engineering*, 12(4), 4457–4468.  
<https://doi.org/10.11591/ijece.v12i4.pp4457-4468>
- Pratiwi, F., & Adrianto, S. (2017). Peningkatan Jumlah Mahasiswa melalui Promosi dengan Penerapan Analisa Data Mining. *SATIN - Sains Dan Teknologi Informasi*, 3(2), 29–37.  
<http://jurnal.stmik-amik-riau.ac.id/index.php/satin/article/view/257/pdf>
- Presiden Republik Indonesia. (2021). *Peraturan Presiden Republik Indonesia Nomor 104 Tahun 2021 tentang Rincian Anggaran Pendapatan dan Belanja Negara Tahun Anggaran 2022*. 4–4b.
- Setiawan, R. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi

Promosi Mahasiswa Baru ( Studi Kasus : Politeknik Lp3i Jakarta ).  
*Jurnal Lentera Ict*, 3(1), 76–92.

Sukmayadi, C., Primajaya, A., & Maulana, I. (2021). Penerapan Algoritma K-Medoids dalam Menentukan Daerah Rawan Banjir di Kabupaten Karawang. *INFORMAL: Informatics Journal*, 6(3), 187.

<https://doi.org/10.19184/isj.v6i3.25423>

Utomo, W. (2021). The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia. *ILKOM Jurnal Ilmiah*, 13(1), 31–35.  
<https://doi.org/10.33096/ilkom.v13i1.763.31-35>