

# ALGORITMA PARTITIONING AROUND MEDOIDS UNTUK MENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN LUAS PANEN PRODUKTIVITAS DAN PRODUKSI PADI

Ida Safitri<sup>1</sup>, Hardian Oktavianto<sup>2</sup>, Ilham Saifudin<sup>3</sup>  
Program Studi Teknik Informatika, Fakultas Teknik,  
Universitas Muhammadiyah Jember e-mail:  
[idasafitri860@gmail.com](mailto:idasafitri860@gmail.com)<sup>1</sup>

## ABSTRAK

Indonesia merupakan negara penghasil beras terbesar di dunia, namun Indonesia juga merupakan importir beras dengan konsumsi beras perkapita sebanyak 140 kilogram beras per tahun. Dinas Perantaraan berupaya untuk terus mengoptimalkan hasil pertanian beras dalam rangka memenuhi kebutuhan beras. Dalam hal ini diperlukan metode untuk mengelompokkan data hasil panen di Indonesia berdasarkan Luas panen Produktivitas dan Produksi masing-masing Provinsi, salah satu algoritma pengelompokkan yang dapat digunakan adalah *Partitioning Around Medoids* (PAM). Data yang digunakan adalah Data Luas lahan Produktivitas dan Produksi Padi tahun 2019 di 34 Provinsi di Indonesia. Dari rangkaian pengujian mulai dari 2 *cluster* hingga 10 *cluster*, didapatkan *cluster* optimum yang dihasilkan berada dalam 3 *cluster* berdasarkan nilai terendah yang dihitung dengan metode *Davies Bouldin Index*. *cluster* 1 terdapat 29 provinsi, *cluster* 2 terdapat 2 provinsi, dan *cluster* 3 terdapat 3 provinsi. Berdasarkan karakteristik dari masing-masing *cluster*, ditemukan *cluster* 1 memiliki komposisi lebih rendah dibandingkan dengan *cluster* 2 dan *cluster* 3.

**Kata Kunci:** Luas panen Produktivitas dan Produksi Padi, *Clustering*, *Partitioning Around Medoids*, *Davies Bouldin Index*

**ALGORITHM OF PARTITIONING AROUND MEDOIDS (PAM) FOR GROUPING  
PROVINCES IN INDONESIA  
BASED ON HARVEST AREA, RICE PRODUCTIVITY AND PRODUCTION**

Ida Safitri<sup>1</sup>, Hardian Oktavianto<sup>2</sup>, Ilham Saifudin<sup>3</sup>  
Informatics Engineering Study Program , Faculty of Engineering,  
University of Muhammadiyah Jember e-mail:  
[idasafitri860@gmail.com](mailto:idasafitri860@gmail.com)

**ABSTRACT**

*Indonesia is one of the countries producing rice with the highest proportion in the world, But Indonesia was also still is a pengimport rice having rice consumption per capita 142 per year. In order meet rice need, office for agricultural products try to keep in optimize rice. In this case should be a method for harvesting based on data for the group broad harvest the production and productivity of harvest for every province,one clustering algorithms that can be used is Partitioning Around Medoids (PAM).The data used the data broad rice hearvest the production and productivity of years 2019 in 34 provinces in Indonesia.Of a series of testing starts from 2 until 10 clusters of clusters,produced clusters steady that ist at 3 clusters based on the lowest value calculated with the methods Davies Bouldin Index. Clusters of 1 consisting 29 province, clusters of 2 consisting of 2 province,and clusters of 3 consisting 3 province.Based on the has lower than with broad harvest in cluster 2 and cluster of 3 while broad harvest in clusters of 3 had a composition of higher than broad harvest in clusters 1 and clusters of 2.*

**Keywords :** *Productivity and Production Harvested Areas, Clustering, Partitioning Around the Medoid, Davies Bouldin Index.*

## 1. PENDAHULUAN

Departemen pertanian telah membuat rencana pembangunan pertanian yaitu “Pangan kebutuhan nasional yang harus dipenuhi oleh produksi dalam negeri, karena kekurangan pangan dapat menyebabkan kekacauan politik, sosial ekonomi”.

Menurut FAO (2014) Sumber penghasilan pangan di Indonesia masih identik dengan nasi. Meski begitu, nasi bukan satu-satunya sumber pangan di Indonesia. Beras merupakan makanan pokok terpenting di dunia terutama di Benua Asia, Untuk mayoritas masyarakat dikalangan menengah kebawah beras menjadi makanan pokok. Di Indonesia tahun 2014, menempati peringkat ke-3 dengan total produksi sebanyak 70,6 juta ton dan merupakan negara penghasil beras terbesar di dunia. Akan tetapi Indonesia masih tetap menjadi negara pengimpor beras. Dikarenakan, konsumsi beras perkapita Indonesia merupakan yang terbesar di dunia, yakni sekitar 140 kilogram beras per tahun.

Menurut Ridlo (2017) melakukan penelitian pengelompokan yaitu Penerapan Algoritma *K-Means* untuk pemetaan Produktivitas Panen Padi di Kabupaten Karawang. Penelitian ini menggunakan K-Means Clustering untuk mengelompokkan wilayah berdasarkan Produktivitas Panen Padi hanya di Kabupaten Karawang yang dikelompokkan menjadi 3 cluster, yaitu Produktivitas Panen tinggi, sedang dan kurang.

Untuk memenuhi kebutuhan beras, Dinas Pertanian terus memaksimalkan hasil panen padi. maka diperlukan suatu metode untuk memperjelas provinsi-provinsi di Indonesia berdasarkan luas panen produktivitas dan produksi padi. Tujuannya adalah untuk mengetahui hasil produksi padi suboptimal di setiap Provinsi di Indonesia. Sehingga Dinas pertanian dapat memberikan perhatian dan penanganan yang efektif terkait pengambilan kebijakan dan penyaluran bantuan.

## 2. PENELITIAN TERKAIT

### **Produksi Padi**

Menurut Lains (1998) (dalam Joko Triyanto, 2006 ) Penggunaan luas lahan untuk pertanian secara umum dapat dibedakan menjadi penggunaan lahan musiman, tahunan dan permanen. Lahan musiman digunakan untuk panen dengan pola tanam berbagai jenis tanaman secara bergilir di satu petak lahan dilakukan dalam jangka waktu kurang dari setahun. Lahan tahunan merupakan tanaman jangka panjang. Lahan Permanen diarahkan pada lahan yang tidak diusahakan untuk pertanian seperti hutan dll. Jadi hubungannya luas lahan dan produksi padi adalah positif

### **A. Data Mining**

Pramudiono (2006) Merupakan rangkaian proses untuk menggali nilai tambah berupa pengetahuan yang tidak diketahui secara manual dari suatu kumpulan data.

### **B. Clustering**

Rui Xu & Donald (2009) (dalam Hasugian, 2018). Pengelompokan data merupakan pendekatan untuk mempertimbangkan atau mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok.

### C. Partitioning Around Medoids

Menurut Leonard Kaufman dan Peter J. Rousseeuw (1987) (dalam Santoso, Februariyanti & Henry, 2016) Algoritma *Partitioning Around Medoids* juga dikenal sebagai *K-Medoids*, adalah algoritma pengelompokan yang terkait dengan algoritma *K-Means*.

Algoritma *Partitioning Around Medoids* menggunakan metode pengelompokan partisi untuk mengelompokkan sekelompok objek  $n$  menjadi sejumlah  $k$  cluster.

Langkah-langkah dari algoritma PAM (Han dan Kamber, 2006) adalah :

1. Pilih  $k$  objek secara acak dari satu set  $n$  sebagai *medoids* .
2. Menempatkan objek *non-medoids* ke kelompok yang paling dekat dengan *medoid* menggunakan *Euclidian Distance*.
3. Secara acak memilih  $o_{random}$  : objek *non-medoid*..
4. Menghitung total *cost*,  $S$  (selisih), dari pertukaran *medoid*  $o_j$  dengan  $o_{random}$ .
5. Jika  $S < 0$  maka tukar  $o_j$  dengan  $o_{random}$  untuk membentuk himpunan  $k$  objek baru sebagai *medoid*.
6. Mengulangi langkah 3 hingga 5 sampai tidak ada perubahan *medoid*, sehingga didapatkan masing-masing *cluster* dan anggota *cluster* .

Nilai *cost* dinyatakan dengan persamaan:

$$Total\ Cost = \sum \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \dots \dots \dots (2.1)$$

Dengan:

$n$  = jumlah data

$k$  = indeks data

$x_k$  = nilai atribut ke- $k$  dari  $x$

$y_k$  = nilai atribut ke- $k$  dari  $y$

Nilai  $S$  dinyatakan dalam persamaan:

$$S = Total\ Cost\ baru - Total\ Cost\ Lama \dots \dots \dots (2.2)$$

Dengan:

$S$  = Selisih

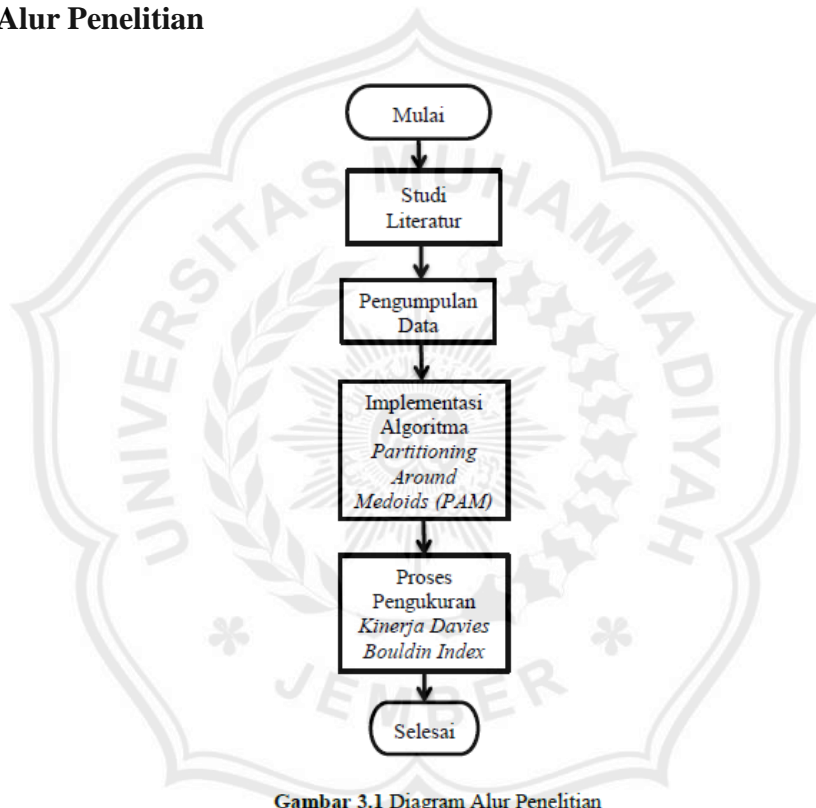
*Total Cost Baru* = Jumlah *cost non-medoids*

*Total Cost Lama* = Jumlah *cost medoids*.

#### D. Metode Analisis Data

Metode analisa data yang digunakan adalah Algoritma *Partitioning Around Medoids*. *Partitioning Around Medoids* yaitu teknik *cluster* atau pengelompokan suatu kelompok objek sebagai perwakilan (*medoid*) dalam pusat *cluster* untuk setiap cluster. Keunggulan PAM adalah dapat mengatasi kelemahan algoritma *K-Means* yang sensitif terhadap *noise* atau *outlier*, dan objek dengan nilai besar yang memungkinkan terjadinya deviasi dari sebaran data.

#### E. Diagram Alur Penelitian



Gambar 3.1 Diagram Alur Penelitian

#### F. Pengumpulan Data

Data yang digunakan adalah data Luas Lahan Produktivitas dan Produksi Padi Menurut Provinsi di Indonesia tahun 2019 yang terdiri dari 34 Provinsi. Data ini diambil dari situs resmi Badan Pusat Statistik yaitu <https://www.bps.go.id/dynamictable/2019/04/15/1608/luas-panen-produksi-dan-produktivitas-padi-menurut-provinsi-2018-2019.html>.

### 3. HASIL DAN PEMBAHASAN

Pada Penelitian ini membahas mengenai hasil yang diperoleh dari pengujian data, data tersebut akan diolah dengan menggunakan algoritma *Partitioning Around Medoids*. Berikut hasil dan pembahasan Algoritma *Partitioning Around Medoids* pada pengelompokan Provinsi di Indonesia berdasarkan Luas Panen Produktivitas dan Produksi Padi. Data yang digunakan adalah data Luas lahan Produktivitas untuk Produksi Padi di Indonesia tahun 2019.

**Tabel 4.1** Data Luas Panen Produktivitas dan Produksi Padi

NO	PROVINSI	LUAS PANEN (ha)	PRODUKTIVITAS (kw/ha)	PRODUKSI (ton)
1	Aceh	0,18	0,83	0,18
2	Sumatera Utara	0,24	0,68	0,22
3	Sumatera Barat	0,18	0,59	0,15
4	Riau	0,04	0,25	0,02
5	Jambi	0,04	0,5	0,03
6	Sumatera Selatan	0,32	0,61	0,27
7	Bengkulu	0,04	0,54	0,03
8	Lampung	0,27	0,56	0,22
9	Kep, Bangka Belitung	0,01	0	0
10	Kepulauan Riau	0	0,12	0
11	DKI Jakarta	0	0,79	0
12	Jawa Barat	0,93	0,9	0,94
...	.....	.....	.....	....
...	.....	.....	.....	....
34	Papua	0,03	0,46	0,02

a. *Partitioning Around Medoids* (PAM) pada RapidMiner

Pengolahan data menggunakan tools RapidMiner yang di cluster menggunakan algoritma PAM dari 2-10 *cluster*. Pengolahan tersebut menghasilkan output berupa pusat *cluster*, keanggotaan *cluster*, centroid tabel, nilai DBI dan total keseluruhan *items*.

b. Penentuan Jumlah Cluster Optimum

Setelah melalui proses cluster dengan menggunakan algoritma *Partitioning Around Medoids*, kemudian dilakukan proses dengan menggunakan metode *Davies Bouldin-Index* untuk menentukan cluster yang optimum. Berikut ini adalah hasil metode *Davies Bouldin-Index* :

**Tabel 4.2** Hasil Nilai Metode *Davies Bouldin-Index*

<i>Cluster</i>	Nilai <i>Davies Bouldin-Index</i>
2	0, 896
3	0, 486
4	0, 945
5	0, 919
6	0, 518
7	0, 631
8	0,557
9	0,751
10	0, 571

c. Hasil Profiling *Cluster*

Dari Tabel 4.2 menunjukkan nilai *Davies Bouldin Index* dengan pengujian skenario *cluster 2* sampai dengan 10 *cluster*, maka *cluster* yang paling optimum berdasarkan nilai *Davies Bouldin Index* terkecil yaitu 3 *cluster* dengan nilai 0,486. Berdasarkan perhitungan yang telah dilakukan sebelumnya. Hasil 3 *cluster* mencakup *cluster 1* terdapat 29 provinsi, *cluster 2* terdapat 2 provinsi dan *cluster 3* terdapat 3 provinsi.

**4. KESIMPULAN**

Berdasarkan Penelitian di atas maka dapat disimpulkan bahwa:

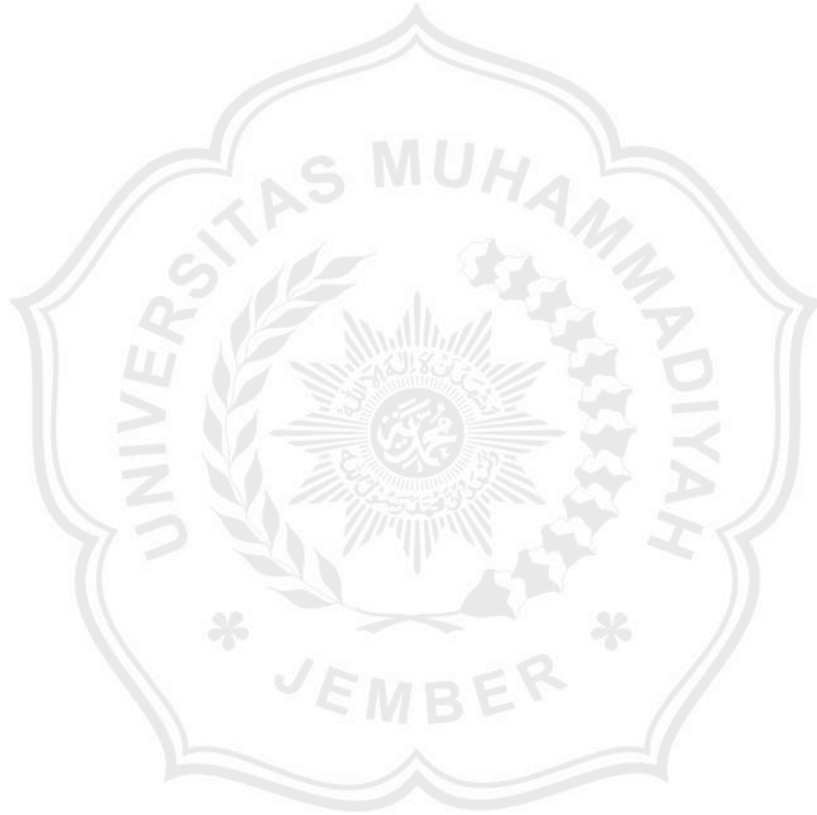
1. Hasil dari penerapan metode *Partitioning Around Medoids* dalam mengelompokkan provinsi di Indonesia berdasarkan luas panen, produktivitas dan produksi padi diperoleh *cluster optimum* dengan hasil 3 *cluster* berdasarkan indeks nilai, *Davies Bouldin- Indeks* terkecil yaitu sebesar 0,486 dengan skenario *cluster 2* sampai dengan 10 *cluster*.
2. Hasil dari pengelompokkan dari 3 *cluster* pada *cluster 1* terdapat 29 provinsi yang terdiri dari Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua barat, Papua. Pada *cluster 2* terdapat 2 provinsi yaitu dari Sumatera Selatan dan Sulawesi Selatan. Sedangkan pada *Cluster 3* terdapat 3 provinsi terdiri dari Jawa Barat, Jawa Tengah, dan Jawa Timur.

**5. SARAN**

Saran dari penelitian ini adalah:

1. Dapat dikembangkan menggunakan metode *clustering* yang berbeda, untuk uji validitas *cluster* dapat menggunakan alternatif lain seperti *Elbow*, *silhouette coefisien*, dll.

2. Perhitungan manual *cluster* pada algoritma *Partitioning Around Medoids* (PAM) pada *Microsoft Excel* dapat digunakan sebagai alternatif pembandingan hasil *cluster* pada tools *RapidMiner*.





## DAFTAR PUSTAKA

- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceeding – 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>
- Jain, A. K., & Duta besar, R. C. (1988). *Clustering\_Jain\_Dubes.pdf* (p. 334). p. 334.
- Ridlo, M. R., Defiyanti, S., & Primajaya, A. (2017). Implementasi Algoritme K-Means Untuk Pemetaan Produktivitas Panen Padi Di Kabupaten Karawang. *Citee 2017*, 426–433.
- Statistik Indonesia. (2014). *Organisasi dan Kegiatan Sensus Pertanian 2013 di Indonesia*.
- Triyanto, J. (2006). Analisis Produksi Beras di Jawa Tengah. *Universitas Diponegoro*, 1–85.