# Penerapan K-Means Cluster pada Daerah Penggunaan Teknologi di Indonesia

**Silvana Fitria Mandang[1], Betha Nurina Sari[2]**
Fakultas Ilmu Komputer, Teknik Informatika
Universitas Singaperbangsa Karawang, Indonesia
e-mail: [1] silvana.fmandang17027@student.unsika.ac.id, [2] betha.nurina@staff.unsika.ac.id

Diterima: 21 Maret 2021; Direvisi: 19 Mei 2021; Disetujui: 25 Mei 2021

***Abstrak***

*Indonesia saat ini sedang mengalami kondisi yang tidak stabil akibat adanya virus Covid-19. Virus Covid-19 telah menyebar ke seluruh wilayah Indonesia dan menginfeksi ribuan orang. Akibat adanya virus ini hampir semua aspek kehidupan berubah termasuk pendidikan. Pemerintah akhirnya mengeluarkan kebijakan baru dengan mengubah proses belajar dan mengajar tatap muka menjadi daring. Akan tetapi, di Indonesia sendiri perkembangan dan pemanfaatan teknologi komputer dan internet masih belum merata. Umumnya hanya masyarakat perkotaan yang memiliki akses teknologi tinggi dibandingkan dengan pedesaan. Pada penelitian ini akan menerapkan metode clustering k-means pada penggunaan teknologi siswa umur 5-24 tahun selama pembelajaran daring. Dari hasil penelitian menggunakan 34 data provinsi di Indonesia menghasilkan 3 cluster, cluster pertama dengan kategori tinggi sebanyak 7 provinsi, cluster kedua dengan kategori sedang sebanyak 19 provinsi dan cluster ketiga dengan kategori rendah sebanyak 8 provinsi.*

***Kata kunci:*** *Klastering, Data Mining, K-Means, Teknologi*

***Abstract***

*Indonesia is currently experiencing an unstable condition due to the Covid-19 virus. The Covid-19 virus has spread throughout Indonesia and infected thousands of people. As a result of this virus, almost all aspects of life have changed, including education. The government finally issued a new policy by changing the face-to-face learning and teaching process to be bold. However, in Indonesia itself, the development and use of computer and internet technology is still not evenly distributed. Only urban communities have access to high technology compared to rural areas. This study will apply the k-means clustering method to the use of technology for students aged 5-24 years of age during bold learning. From the results of the study using 34 provincial data in Indonesia resulted in 3 clusters, the first cluster with the high category was 7 provinces, the second cluster with the medium category was 19 provinces and the third cluster with the low category was 8 provinces.*

***Keywords****: Clustering, Data Mining, K-Means, Technology*

## 1. PENDAHULUAN

Indonesia saat ini sedang mengalami kondisi yang tidak stabil akibat adanya virus Covid-19. Virus Covid-19 telah menyebar ke seluruh wilayah Indonesia dan menginfeksi ribuan orang. Akibat adanya virus ini hampir semua aspek kehidupan berubah termasuk pendidikan. Pemerintah akhirnya mengeluarkan kebijakan baru dengan mengubah proses belajar dan mengajar tatap muka menjadi daring. Hal ini bertujuan agar meminimalisir penyebaran virus di lingkungan pendidikan. Pembelajaran daring tentunya memerlukan media pendukung seperti teknologi. Teknologi yang

dibutuhkan untuk mendukung pembelajaran daring adalah komputer dan akses internet. Dampak positif yang bisa diambil dari pembelajaran daring adalah siswa menjadi lebih mandiri dalam mempelajari materi yang bisa diakses melalui internet menggunakan komputer atau *gadget*. Akan tetapi, di Indonesia sendiri perkembangan dan pemanfaatan teknologi komputer dan internet masih belum merata. Umumnya hanya masyarakat perkotaan yang memiliki akses teknologi tinggi dibandingkan dengan pedesaan. Sehingga pembelajaran daring menjadi kurang efektif bagi beberapa siswa.

Berdasarkan permasalahan maka dibuatlah suatu penelitian untuk menemukan kelompok daerah dengan pemanfaatan teknologi dari rendah, sedang, dan tinggi. Data yang digunakan dalam penelitian ini adalah data-data yang terdapat di website resmi Badan Pusat Statistik tahun 2020 tentang pendidikan pada bulan Agustus sampai November. Metode yang digunakan dalam pengelompokan adalah *clustering.* Salah satu metode dalam pengelompokkan *clustering* adalah algoritma k-means. K-means dalam beberapa penelitian dapat digunakan untuk mengelompokkan suatu objek ke dalam *cluster* berdasarkan nilai rata-rata, seperti pada penelitian yang telah dilakukan oleh Alkhairi yaitu menerapkan algoritma k-means dalam mengelompokkan daerah yang memiliki potensi produksi karet di Indonesia[1]

Untuk memperoleh informasi provinsi mana saja di Indonesia yang mengakses teknologi rendah, sedang, tinggi maka digunaknan teknik pembagiang *cluster* menggunakan algoritma k-means. Tahapan penelitian ini berdasarkan tahapan yang terdapat pada data mining dan diimplementasikan menggunakan aplikasi rapidminer. Penelitian ini diharapkan dapat dijadikan bahan evaluasi pemerintah daerah terhadap kebijakan pembelajaran daring.

## 2. METODOLOGI PENELITIAN

### 2.1. Data Mining

Data mining merupakan proses mencari pengetahuan pada basis data menggunakan beberapa metode seperti klasifikasi, asosiasi, klastering dan lain sebagainya. Pemilihan tekink atau algoritma sangat bergantung pada tujuan yang ingin dicapai, sehingga hasil dari pengolahan data menggunakan data mining ini dapat digunakan untuk pengambilan keputusan.

### 2.2. Clustering

*Clustering* merupakan suatu proses membagi data kedalam beberapa kelompok yang karakternya memiliki kemiripan antara data satu dengan yang lain. *Clustering* memiliki sifat *unsupervised* yang artinya metode ini dapat diterapkan tanpa latihan karena tidak memerlukan target keluaran. *Clustering* memiliki dua jenis teknik dalam proses pengelompokan, yaitu hirarki dan non-hirarki. Umumnya, *clustering* banyak diimplementasikan kedalam beberapa bidang, seperti bidang ekonomi pada sebuah bisnis untuk menganalisa pasar[2]. Metode *clustering* juga dapat digunakan untuk pemetaan daerah yang padat, menentukan pola-pola distribusi secara keseluruhan dan menemukan keterkaitan antara atribut-atribut data. Dalam data mining, metode *clustering* difokuskan untuk menemukan *cluster* pada basis data berukuran besar secara efektif dan efisien[3].

### 2.3. Algoritma K-Means

Algoritma K-Means merupakan proses yang telah banyak digunakan untuk membagi dan memisahkan data ke dalam beberapa kelompok. Ada dua jenis *clustering* dalam pengelompokkan data, yaitu *hierarchical* dan *non-hierarchical*. *Hierarchical* merupakan metode pengelompokkan yang memiliki urutan berdasarkan tingkatan, berbeda dengan *non-hirarchical*, dan algoritma k-means sendiri merupakan salah satu metode dari *non-hierarchical*. *Clustering K-Means* merupakan metode pengelompokan data kedalam beberapa klaster. Pengelompokkan data tersebut berdasarkan karakteristik dari data yang dimiliki, sehingga data yang memiliki

karakteristi mirip akan dikumpulkan dalam satu klaster dan data yang memiliki karakteristik berbeda akan dikumpulkan ke klaster yang lain. Untuk mengelompokkan data menggunakan algoritma *clustering k-means*, langkah-langkah yang dilakukan sebagai berikut [4]:

1. Langkah pertama adalah penentuan jumlah *cluster*. Jumlah *cluster* yang akan dibentuk sebanyak 3.
2. Menentukan centroid awal. Penentuan centroid awal dapat dilakukan dengan memilih data secara acak.
3. Menghitung jarak data ke centroid menggunakan rumus jarak *Euclidean* yang dirumuskan sebagaimana dalam persamaan 1 berikut :

$$D_{(a,b)} = \sqrt{(x_{1a} - y_1 b)^2 + (x_{2a} - y_2 b)^2 + \cdots . + (x_{ca} - y_{cb})^2} \quad (1)$$

Keterangan:

$D_{(a,b)}$ = Jarak data a ke centroid b
$X_{1a}$ = Data ke 1 pada atribut data ke a
$X_{2a}$ = Data ke 2 pada atribut data ke a
$Y_1b$ = Data centroid ke 1 pada atribut b
$Y_2b$ = Data centroid ke 2 pada atribut b
$X_{ca}$ = Data ke c pada atribut data ke a
$Y_{cb}$ = Data centroid ke c pada atribut b

4. Kelompokkan data yang telah dihitung menggunakan rumus *euclidean* tersebut berdasarkan jarak paling dengat dengan nilai centroid.
5. Hitung kembali keanggotaan *cluster* dengan nilai centroid baru. Nilai centroid baru dapat diperoleh dari rata-rata *cluster*.
6. Lakukan perulangan pada langkah tiga sampai lima, sampai data tidak berpindah ke *cluster* lain.

### 2.4. *Davies-Bouldin Index* (DBI)

*Davies-Bouldin Index* merupakan salah satu metode untuk mengukur evaluasi jumlah *cluster* pada suatu metode pengelompokan data[5]. Suatu *cluster* akan dianggap memiliki skema *clustering* yang baik apabila memiliki nilai DBI paling rendah [6].

## 3. HASIL DAN PEMBAHASAN

Pada penelitian ini data yang digunakan didapat dari situs Badan Pusat Statistik. Data yang dipilih adalah data persentase siswa umur 5-24 tahun yang menggunakan telepon seluler, menggunakan komputer dan menggunakan internet selama bulan Agustus sampai November 2020 menurut Provinsi. Kemudian data di preprocessing sesuai dengan kebutuhan dengan memanfaatkan *software microsoft excel*. Data tersebut akan dikelompokkan berdasarkan *cluster* 1 dengan tertinggi, *cluster* 2 dengan sedang, *cluster* 3 dengan rendah. Data yang akan digunakan menurut provinsi ditunjukkan pada tabel 1:

Tabel 1. Persentase siswa umur 5-24 tahun yang menggunakan teknologi

| Provinsi | Persentase Siswa yang menggunakan teknologi | | |
|---|---|---|---|
| | Telepon Seluler | Menggunakan Komputer | Menggunakan Internet |
| Bengkulu | 75,44 | 22,15 | 54,49 |
| Sulawesi Barat | 69,58 | 16,08 | 41,82 |
| Jawa Tengah | 81,62 | 27,56 | 69,33 |

| | | | |
|---|---|---|---|
| Aceh | 67,53 | 14,08 | 40,48 |
| Sulawesi Tengah | 74,35 | 17,92 | 50,28 |
| Bali | 85,5 | 30,68 | 71,49 |
| Jambi | 72,8 | 19,68 | 53,92 |
| Lampung | 77,94 | 18,14 | 56,15 |
| Gorontalo | 83,15 | 23,95 | 56,77 |
| Kep. Riau | 72,66 | 24,83 | 52,58 |
| Jawa Barat | 75,78 | 24,36 | 64,39 |
| Nusa Tenggara Timur | 69,48 | 14,84 | 31,39 |
| Riau | 78,68 | 21,18 | 53,86 |
| Sumatera Selatan | 76,14 | 18,14 | 52,97 |
| Sumatera Utara | 77,15 | 22,23 | 51,54 |
| DKI Jakarta | 80,81 | 34,37 | 69,93 |
| Sumatera Barat | 77,74 | 26,04 | 51,21 |
| Sulawesi Tenggara | 77,56 | 18,78 | 50,32 |
| Kalimantan Timur | 83,18 | 23,34 | 68,51 |
| Sulawesi Utara | 74,6 | 18,96 | 57,24 |
| Jawa Timur | 82,05 | 31,01 | 67,53 |
| Sulawesi Selatan | 82,43 | 22,99 | 57,99 |
| DI Yogyakarta | 90,55 | 45,7 | 83,21 |
| Papua Barat | 66,25 | 13,28 | 47,33 |
| Kalimantan Utara | 84,7 | 21,41 | 59,47 |
| Maluku Utara | 58,98 | 12,54 | 35,72 |
| Nusa Tenggara Barat | 81,96 | 20,26 | 53,58 |
| Kalimantan Selatan | 83,64 | 37,3 | 65,21 |
| Maluku | 66,45 | 17,88 | 41,67 |
| Kep. Bangka Belitung | 79,81 | 20,92 | 58,92 |
| Kalimantan Barat | 67,8 | 14,58 | 49,63 |
| Kalimantan Tengah | 74,78 | 18,06 | 54,26 |
| Papua | 41,24 | 9,7 | 26,46 |

1. Penerapan Algoritma K-Means

Langkah yang perlu dilakukan untuk mengelompokkan data ke beberapa cluster menggunakan algoritma k-means:

a. Tentukan jumlah *cluster*. Pada penelitian ini data di atas dikelompokkan menjadi 3 *cluster*.
b. Selanjutnya tentukan centroid awal. Untuk menentukan centroid awal, diambil dari data di atas secara acak dan centroid awal yang dipilih adalah pada tabel 2:

Tabel 2. Centroid awal

| | | | |
|---|---|---|---|
| C1 | 75,78 | 24,36 | 64,38 |
| C2 | 74,78 | 18,06 | 54,26 |
| C3 | 74,35 | 17,92 | 50,28 |

c. Hitung jarak data pada masing-masing nilai centroid menggunakan rumus *euclidean distance* sebagai berikut:

$$C1 = \sqrt{(67{,}53 - 75{,}78)^2 + (14{,}08 - 24{,}36)^2 + (40{,}48 - 64{,}39)^2}$$

C1 = 27,30254567

$C2 = \sqrt{(67{,}53 - 74{,}78)^2 + (14{,}08 - 18{,}06)^2 + (40{,}48 - 54{,}26)^2}$

C2 = 16,07144362

$C3 = \sqrt{(67{,}53 - 74{,}35)^2 + (14{,}08 - 17{,}92)^2 + (40{,}48 - 50{,}28)^2}$

C3 = 12,54184994

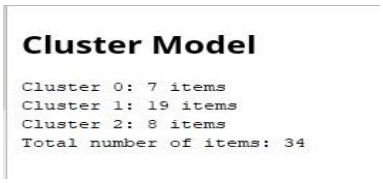
Perhitungan jarak pada nilai centroid dilakukan seterusnya sampai data terakhir yaitu data ke 34. Hasil dari perhitungan data terhadap nilai centroid tersebut akan menjadi anggota dari setiap *cluster* yang memiliki jarak terdekat dari pusat *cluster*nya. Perhitungan terhadap nilai centroid dilakukan sampai proses iterasi *cluster* tidak berubah lagi. Pada penelitian ini proses berhenti pada iterasi keempat. Sehingga didapatkan nilai ditunjukkan pada tabel 3:

Tabel 3. Pengelompokan data berdasarkan *cluster*

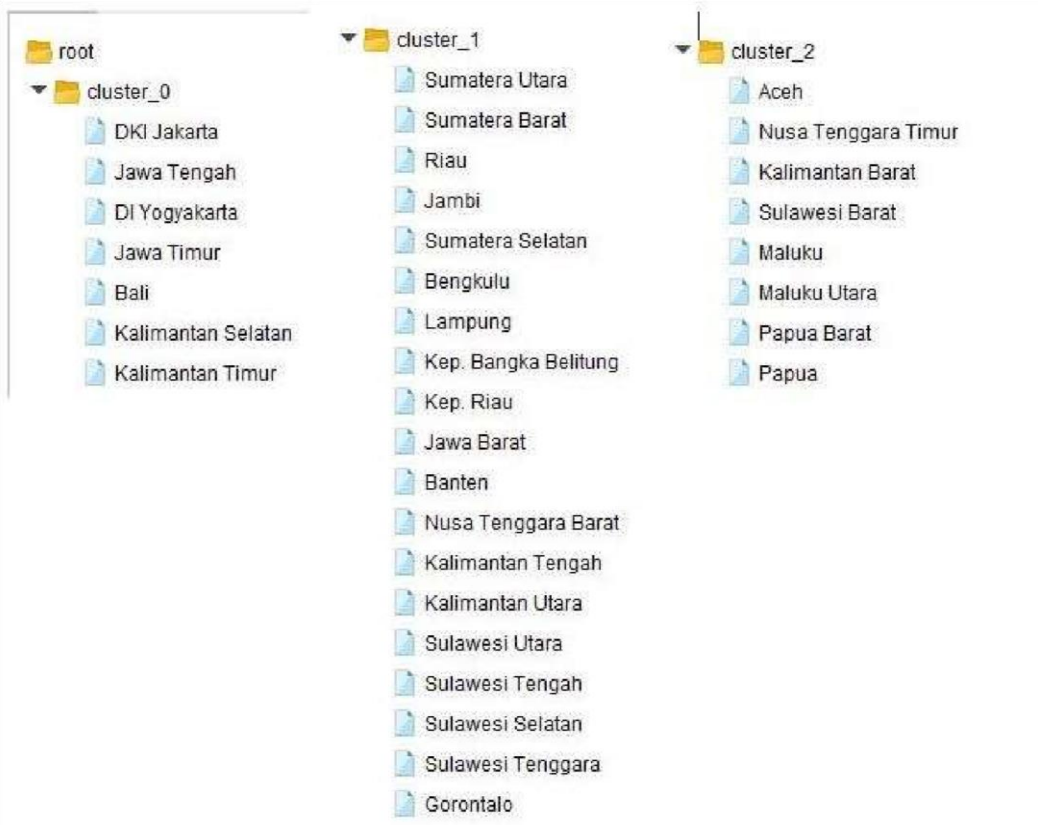
| C1 | C2 | C3 | Cluster |
|---|---|---|---|
| 37,65757562 | 18,35245409 | 1,342203275 | 3 |
| 21,52402499 | 3,269927989 | 16,60009618 | 2 |
| 20,26784341 | 6,160317123 | 18,79132066 | 2 |
| 19,72802135 | 1,626607221 | 18,68812911 | 2 |
| 22,4821335 | 4,616370766 | 15,03472331 | 2 |
| 22,80838172 | 3,311145882 | 15,57485095 | 2 |
| 19,68406276 | 2,198369635 | 17,63986679 | 2 |
| 20,03190896 | 3,259387527 | 19,1186723 | 2 |
| 15,78961372 | 5,1132036 | 22,99642153 | 2 |
| 21,32690734 | 6,346179406 | 16,4056904 | 2 |
| 3,314875799 | 20,81007307 | 37,62445215 | 1 |
| 11,69138044 | 10,5791676 | 26,77774704 | 2 |
| 4,460199162 | 16,84652852 | 34,41421774 | 1 |
| 20,6874878 | 40,22606273 | 57,4568877 | 1 |
| 2,67815264 | 17,19783722 | 34,63387308 | 1 |
| 23,75682274 | 9,220916063 | 15,44863484 | 2 |
| 3,226341049 | 21,28821879 | 39,14260671 | 1 |
| 20,04457599 | 4,865250153 | 20,52692128 | 2 |
| 44,20449894 | 25,11518909 | 10,18032094 | 3 |
| 30,60771678 | 12,33251633 | 8,570297797 | 3 |
| 22,37144109 | 3,727147307 | 15,81310421 | 2 |
| 7,306731592 | 20,63530564 | 37,12488098 | 1 |
| 8,576682142 | 15,41490522 | 33,13420534 | 1 |
| 14,86090093 | 8,989346197 | 26,61309394 | 2 |
| 19,8720363 | 4,237548236 | 18,46401759 | 2 |
| 25,53896536 | 5,900205679 | 12,39983056 | 2 |
| 14,85578512 | 6,614048159 | 24,54869869 | 2 |
| 24,14579107 | 4,691742006 | 14,86105744 | 2 |
| 15,33769707 | 7,054311664 | 24,5582664 | 2 |
| 34,86101514 | 15,56708679 | 3,346625069 | 3 |
| 35,54660605 | 17,02417337 | 3,171617041 | 3 |
| 45,97829136 | 27,48001388 | 9,599676534 | 3 |
| 33,63355024 | 15,14827644 | 6,363446467 | 3 |
| 64,14183367 | 46,97017201 | 29,72350283 | 3 |

2. Implementasi pada Rapidminer

Tahap ini merupakan tahap implementasi dari Algoritma K-Means menggunakan aplikasi RapidMiner. Berdasarkan Gambar 1 pemodelan algoritma *clustering k-means* diperoleh hasil yang serupa dengan perhitungan manual. Dari ketiga cluster yang diterapkan pada model menghasilkan *cluster* 0 sejumlah 7 *items*, *cluster* 1 sejumlah 19 *items*, dan *cluster* 2 sebanyak 8 *items.*




Gambar 1**.** Hasil cluster dalam rapidminer




Gambar 2. Hasil cluster dalam rapidminer

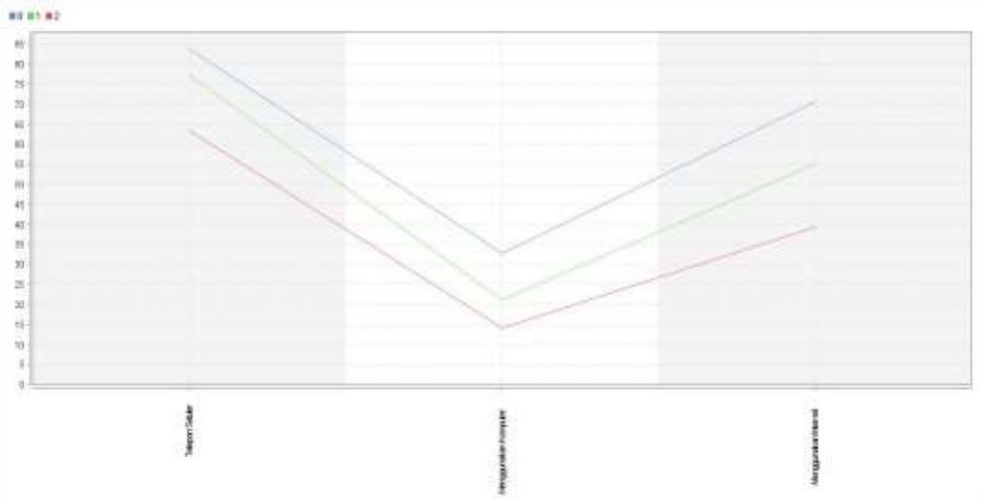
Dapat disimpulkan bahwa *cluster* terhadap persebaran penggunaan teknologi pada siswa usia 5-24 tahun di Indonesia:

a. Cluster 1 (Tinggi) adalah Bali, Jawa Tengah, Jawa Timur, DI Yogyakarta, Kalimantan Selatan, Kalimantan Timur, DKI Jakarta.
b. Cluster 2 (Sedang) adalah Bengkulu, Riau, Sulawesi Utara, Kalimantan Tengah, Sumatera Utara, Jambi, Jawa Barat, Sumatra Barat, Lampung, Kep. Bangka Belitung, Sulawesi Selatan, Kep. Riau, Banten, Gorontalo, Kalinantan Utara, Sulawesi Tengah, Sulawesi Tenggara, Sumatera Selatan, Nusa Tenggara Barat.
c. Cluster 3 (Rendah) adalah Papua Barat, Maluku, Aceh, Sulawesi Barat, Papua, Kalimantan Barat, Maluku Utara, Nusa Tenggara Timur.

Penentuan *cluster* didasarkan pada hasil perhitungan dimana hasil yang paling kecil atau mendekati dengan nilai centroid merupakan *cluster* tersebut. Hasil akhir yang didapatkan dari centroid akhir ditunjukkan oleh tabel 4:

Tabel 4. Hasil centroid akhir

| Attribute | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Telepon Seluler | 83,907 | 77,147 | 63,414 |
| Menggunakan Komputer | 32,851 | 21,048 | 14,122 |
| Menggunakan Internet | 70,744 | 55,030 | 39,312 |



Gambar 3. Grafik hasil clustering k-means

3. Evaluasi

Evaluasi yang digunakan adalah DBI atau *Davies-Bouldin Index*. Metode evaluasi DBI umum digunakan dalam mengukur kinerja dari pengelompokkan *cluster*. Pada rapidminer untuk melakukan hasil uji *performance* operator yang digunakan adalah *cluster distance performance*. Operator tersebut juga berfungsi untuk mengukur seberapa baik kinerja dari centroid yang dihasilkan. Hasil uji menggunakan rapidminer nilai DBI yang didapatkan sebesar 0,222. Hal ini dikatakan cukup optimal karena apabila nilai DBI mendekati sama dengan nol maka kinerja dari sebuah *cluster* adalah optimal. Nilai DBI yang dihasilkan pada rapidminer ditunjukkan pada gambar 4:

**PerformanceVector**

```
PerformanceVector:
Avg. within centroid distance: 22.871
Avg. within centroid distance_cluster_0: 27.803
Avg. within centroid distance_cluster_1: 11.283
Avg. within centroid distance_cluster_2: 46.077
Davies Bouldin: 0.222
```

Gambar 4. Hasil uji performance

## 4. KESIMPULAN

Hasil pada penelitian ini didapat 3 *cluster* dengan kelompok yang terbentuk berdasarkan nilai tinggi, sedang, rendah dari persentase siswa 5-24 tahun yang menggunakan teknologi selama pembelajaran daring bulan agustus sampai dengan November 2020. Hasil masing-masing *cluster* didapat *cluster* 1 sejumlah 7 provinsi, *cluster* 2 sejumlah 19 provinsi, *cluster* 3 sejumlah 8 provinsi. Berdasarkan evaluasi kinerja metode menggunakan *Davies-Bouldin Index (DBI*) didapatkan nilai cukup optimal yaitu 0,222.

## DAFTAR PUSTAKA

[1] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.

[2] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.

[3] Z. Aras and Sarjono, "Analisis Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan Bedah Rumah Menggunakan Metode Clustering K-Means( Studi Kasus : Kantor Kecamatan Bahar Utara)," *J. Manaj. Sist. Inf.*, vol. 1, no. 2, pp. 159–170, 2016.

[4] A. T. R. Saragih, A. S. Sembiring, and M. Sayuthi, "Penerapan Metode Clustering K-Means untuk Proses Seleksi Calon Peserta Lomba MTQ," *Pelita Inform.*, vol. 17, no. April, pp. 117–122, 2018, [Online]. Available: https://ejurnal.stmik-budidarma.ac.id/index.php/pelita/article/download/776/704.

[5] A. F. Muhammad, "Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means (Study Kasus : Tim Hockey Kabupaten Kendal)," *Jur. Tek. Inform. FIK UDINUS*, pp. 1–5, 2015.

[6] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.