

## Optimasi Metode Kmeans Dengan Algoritma Artificial Bee Colony Untuk Pengelompokan Penyebaran Covid-19 Di Provinsi Indonesia

### *Optimization Of The Kmeans Method Using The Artificial Bee Colony Algorithm For Classification Of The Spread Of Covid-19 In The Province Of Indonesia*

Intan Dian Puji Lestari<sup>1</sup>, Deni Arifianto<sup>2\*</sup>, Hardian Oktavianto<sup>3</sup>.

<sup>1</sup>Mahasiswa Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember  
E-mail : [intandianpujilestari2@gmail.com](mailto:intandianpujilestari2@gmail.com)

<sup>2</sup>Dosen Fakultas Teknik, Universitas Muhammadiyah Jember\* Koresponden Author  
E-mail: [deniarifianto@unmuhjember.ac.id](mailto:deniarifianto@unmuhjember.ac.id)

<sup>3</sup>Dosen Fakultas Teknik, Universitas Muhammadiyah Jember  
E-mail: [hardian@unmuhjember.ac.id](mailto:hardian@unmuhjember.ac.id)

#### Abstrak

Penyebaran covid-19 di Indonesia diawali oleh kasus impor yang kemudian berkembang menjadi transmisi local. Dampak wabah covid-19 yang terjadi di kehidupan masyarakat salah satunya yaitu aktivitas sosial ditunda sementara waktu, melemahnya perekonomian, pelayanan transportasi dikurangi dan diatur dengan ketat, pariwisata ditutup, pusat pembelanjaan sepi pengunjung dan ditutupnya sektor informal seperti ojek online, supir angkot, pedagang kaki lima, pedagang keliling, UMKM dan kuli pasar mengalami penurunan pendapatan. Pusat-pusat perdagangan seperti, mall, pasar yang biasanya ramai dikunjungi oleh masyarakat mendadak sepi, bekerja dan belajar pun dilakukan dirumah secara online. Pesatnya penyebaran virus dikarenakan protokol covid di daerah/provinsi masih kurang dijalankan, ketika penyebaran covid19 berkembang pesat maka akan meningkatkan angka perkembangan covid di daerah tersebut. Metode ABCKM merupakan gabungan *clustering K-means* dengan optimasi *Artificial Bee Colony*. Dengan memadukan *K-means* dan metode *Artificial Bee Colony* maka akan dapat meningkatkan kemampuan KM dalam menentukan titik sampel data dan kemudian menemukan cluster pada area global yang optimal, dengan uji coba 2 sampai 10 klaster dan menghasilkan nilai *davies bouldin index* sebesar 0. 8637.

**Kata Kunci:** *Covid19/Coronavirus, Clustering, Kmeans, Artificial Bee Colony, Davies Bouldin Index(DBI)*

#### Abstract

The spread of COVID-19 in Indonesia was initiated by imported cases which later developed into local transmission. One of the impacts of the COVID-19 outbreak in people's lives is that social activities are temporarily suspended, the economy is weakening, transportation services are reduced and strictly regulated, tourism is closed, shopping centers are empty of visitors and the closure of the informal sector such as online motorcycle taxis, public transportation drivers, street vendors. , traveling traders, MSMEs and market coolies experienced a decrease in income. Trade centers such as malls, markets which are usually crowded with people are suddenly deserted, work and study are also done online at home. The rapid spread of the virus is due to the lack of implementation of the covid protocol in the area/province, when the spread of covid19 grows rapidly it will increase the number of covid developments in the area. The ABCKM method is a combination of *K-means* clustering with *Artificial Bee Colony* optimization. By combining *K-means* and the *Artificial Bee Colony method*, it will be able to increase KM's ability to determine data sample points and then find clusters in the optimal global area, by testing 2 to 10 clusters and producing a *Davies bouldin index* value of 0.8637.

**Keywords:** *Covid19/Coronavirus, clustering, kmeans, artificial bee colony, DBI*

## 1. PENDAHULUAN

### A. Latar Belakang

Dampak virus covid19 hampir terjadi di kehidupan masyarakat. Aktivitas sosial ditunda, dikarenakan melemahnya perekonomian, pelayanan transportasi dikurangi dan peraturan yang sangat ketat, pariwisata ditutup, pusat perbelanjaan sepi pengunjung dan di tutupnya sektor informal (Syafriada, 2020). Data dari Gugus Tugas Covid-19 pusat per 31 Juli 2020 mencatat Jawa Timur melaporkan 21.772 kasus dan melaupai DKI yang melaporkan sebanyak 20.969 kasus. Kemudian posisi ketiga ditempati Provinsi Sulawesi Selatan dengan 9.346 kasus dan Jawa Tengah sebanyak 9.281 kasus positif.

Hal yang mendesak saat ini adalah pengaturan yang tegas terhadap pelaksanaan imbauan social distancing, yang masih banyak masyarakat belum mematuhi. Sehingga hal tersebut dapat membuat suatu daerah/provinsi akan mengalami pesatnya persebaran virus dikarenakan protokol covid di daerah/provinsi masih kurang dijalankan, ketika penyebaran covid19 berkembang pesat maka akan meningkatkan angka perkembangan covid di daerah tersebut.

Dengan menggunakan metode K-Means nantinya akan membantu pemerintah dan masyarakat dalam mengoptimalkan penanganan covid19 dalam menemukan provinsi – provinsi yang terdapat dalam cluster terendah hingga tertinggi. Akan tetapi metode *k-means* memiliki kekurangan dalam pencarian titik sampel data yang dipilih secara acak dan jika mempunyai banyak titik sampel data, maka perhitungan pencarian titik terdekat akan membutuhkan waktu yang lama. Dengan memadukan *K-means* dan metode *Artificial Bee Colony* maka akan dapat meningkatkan kemampuan KM dalam menentukan titik sampel data dan kemudian menemukan *cluster* pada area global yang optimal.

### B. Rumusan Masalah

Rumusan masalah yang didapatkan yaitu sebagai berikut :

1. Berapa jumlah cluster terendah untuk pengelompokan provinsi di Indonesia berdasarkan persebaran kasus covid-19 di provinsi indonesia menggunakan Algoritma *Artificial Bee Colony KM* (ABCKM) ?
2. Provinsi mana yang termasuk dalam *cluster optimum* ?

### C. Batasan Masalah

Supaya pembahasan ini dapat dilakukan secara terarah, maka perlu adanya batasan masalah :

1. Dataset yang digunakan yaitu dataset persebaran kasus covid19 dengan rentang bulan april sampai bulan Agustus pada tahun 2020 dan dataset diperoleh dari twitter @kawalcovid19 .
2. Dataset yang digunakan terdiri dari 34 atribut dan 153 record .
3. *Tools* yang digunakan pada proses *clustering* menggunakan Matlab
4. Pencarian *cluster optimum* menggunakan metode *Davies Bouldin Index*.
5. Penentuan *cluster optimum* berdasarkan 2 *cluster* sampai 10 *cluster* .

## 2. STUDI PUSTAKA

### A. Data Mining

Untuk mendapatkan sebuah informasi/ pengetahuan dari berbagai database yang besar dan bermanfaat diperlukanlah sebuah data mining, dimana proses pada data mining menggunakan teknik *statistik*, matematika, kecerdasan buatan, dan *machine learning*.

### B. Min-max Normalization

Model *Min-Max Normalization* merupakan salah satu metode yang mengubah data kompleks dengan tidak menghilangkan isi, sehingga data lebih mudah untuk diolah (Wimmer, 2018). Berikut persamaan *Min-max Normalization* :

$$x^l = \frac{x - x_{min}}{x_{max} - x_{min}} \dots \dots \dots (1)$$

Dimana :

- $x^l$  = Dataset asli
- $x_{min}$  = Nilai terkecil dari dataset tersebut
- $x_{max}$  = nilai tertinggi dari dataset tersebut

### C. Artificial Bee Colony Algorithm

Menurut (Amri, Nababan, & Syahputra, 2012) *Artificial Bee Colony Algorithm* yang diusulkan oleh Karaboga dan Basturk ialah pendekatan *population-based metaheuristic* yang mana pada pendekatan ini terinspirasi dari perilaku kawanan lebah yang mencari makanan. Metode *Artificial Bee Colony Algorithm* terdiri 3 tahapan yaitu:

1. *Employed Bee Phase* berkaitan dengan sumber makanan.
2. *onlooker bee* memilih sumber makanan.
3. *scout bee* mencari secara acak sebuah sumber makanan

Ketiga tahapan pada metode *ABC* tersebut saling berkaitan. *Scout bee phase* akan menemukan posisi pada sumber makanan, kemudian *employed bee phase* secara probabilitas akan memodifikasi dalam menargetkan sumber makanan baru dan kemudian mendapatkan nilai *fitness*. *Scout bee phase* mengevaluasi informasi yang diambil dari semua *employed bee* dan memilih nilai probabilitas tertinggi berdasarkan jumlah *fitness*. *Employed bee* yang sumber makanannya telah habis akan menjadi *scout bee* (Amri, Nababan, & Syahputra, 2012).

Menurut (Amri, Nababan, & Syahputra, 2012) *ABC* memiliki 3 tahapan yaitu menghasilkan inialisasi solusi dari sumber makanan secara acak, masing-masing *employed bee phase* akan memilih posisi pada sumber makanan. Kemudian akan dihitung dengan persamaan dibawah inii

$$X_{ij} = \theta_{ij} + \phi(\theta_{ij} - \theta_{kj}) \dots \dots \dots (2)$$

Dimana :

- $X_i$  : Solusi pada calon sumber makanan dari  $\theta_i$
- $\theta_i$  : Posisi pada *employed bee* ke- $i$
- $\theta_k$  : Tetangga *employed bee* dari  $\theta_i$
- $\phi$  : bilangan acak rentang  $[-1,1]$
- $i \in \{1, 2, 3, \dots, n\}$
- $k \in \{1, 2, 3, \dots, n\}$

Dimana nilai dari  $i \neq k$

- $n$  : jumlah *employed bee*
- $D$  : Dimensi penyelesaian.

Tahap berikutnya, jika sudah mendapatkan sumber makanan dari *employed bee*, langkah selanjutnya yaitu mencari nilai probabilitas. Berikut rumus untuk mencari nilai probabilitas :

$$P_{ij} = \frac{F(\theta_i)}{\sum_{k=1}^S F(\theta_k)} \dots \dots \dots (3)$$

Keterangan :

- $P_{ij}$  : kemungkinan memilih *employed bee* ke-  $i$
- $F(\theta)$ : nilai *fitness value* dari *employed bee* ke-  $i$
- $S$  : jumlah *employed bee*
- $\theta_i$  : posisi dari *employed bee*

*ABC Algorithm* memiliki limit atau batasan yang ditetapkan. Jadi setiap makanan akan melewati limit.

### D. K-Means

*K-means* berfungsi membagi data ke dalam suatu kelompok yang berkarakteristik sama dan karakteristik yang beda akan dimasukkan pada kelompok lainnya. Berikut beberapa tahapan proses *K-means*:

1. Inialisasi data
2. Menentukan jumlah kelompok
3. Menentukan nilai titik pusat kelompok
4. Mencari jarak pada masing-masing data dengan titik pusat menggunakan *Euclidean Distance*. Berikut persamaan *Euclidean Distance* ::

$$deuc(x, y) = \left[ \sum_{j=1}^p (x_j - y_j)^2 \right]^{\frac{1}{2}} \dots \dots (4)$$

keterangan,

- $p$  = dimensi pada dataset
- $x_j$  =  $x$  dimensi ke- $j$
- $y_j$  =  $y$  dimensi ke- $j$

5. Menentukan / mengukur titik pusat kelompok baru

Untuk memperbarui titik pusat, *k-means* akan melakukan perhitungan *average*(rata-rata) dari suatu keanggotaan pengelompokan yang sebelumnya (Maimon & Rokach, 2005):

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \dots \dots \dots (5)$$

### E. Davies Bouldin Index

*Davies bouldin index* ialah fungsi rasio yang ada dalam *cluster* sebagai pemisah *cluster* satu dengan *cluster* yang lain, (Sujacka, 2019). *Davies Bouldin Index* pada penelitian ini difungsikan sebagai metode yang memvalidasi data pada setiap *cluster*. Berikut langkah-langkahnya :

1. Menghitung nilai SSW berdasarkan rumus berikut ini :

$$SSWi = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j c_i) \dots\dots\dots (6)$$

Keterangan:

$m_i$  = banyak data *clusters* ke- $i$

$c_i$  = *clusters* ke- $i$

$d(x_j c_j)$  = Jarak per data ke *centroid*

2. Mengitung nilai SSB

$$SSB_{i,j} = d(c_i, c_j) \dots\dots\dots (7)$$

3. Menghitung nilai Rasio, untuk menentukan perbandingan antara *cluster* ke- $i$  dan *cluster* ke- $j$ . dengan persamaan berikut ini:

$$R_{ij} = \frac{SSWi + SSWj}{SSB_{ij}}$$

4. Menghitung nilai DBI. Hasil dari rasio yang telah diketahui selanjutnya akan dipakai dalam mencari sebuah rasio *Davies Bouldin Index*, sesuai rumus DBI berikut ini :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij})$$

Berdasarkan rumus diatas  $k$  = jumlah *cluster* yang ditentukan.

Menurut (Salazar,dkk) sebuah *clustering* yang optimum pada *indeks* pengukuran *DBI* terjadi jika *cluster* tersebut memiliki sebuah nilai *indeks* yang minimum.

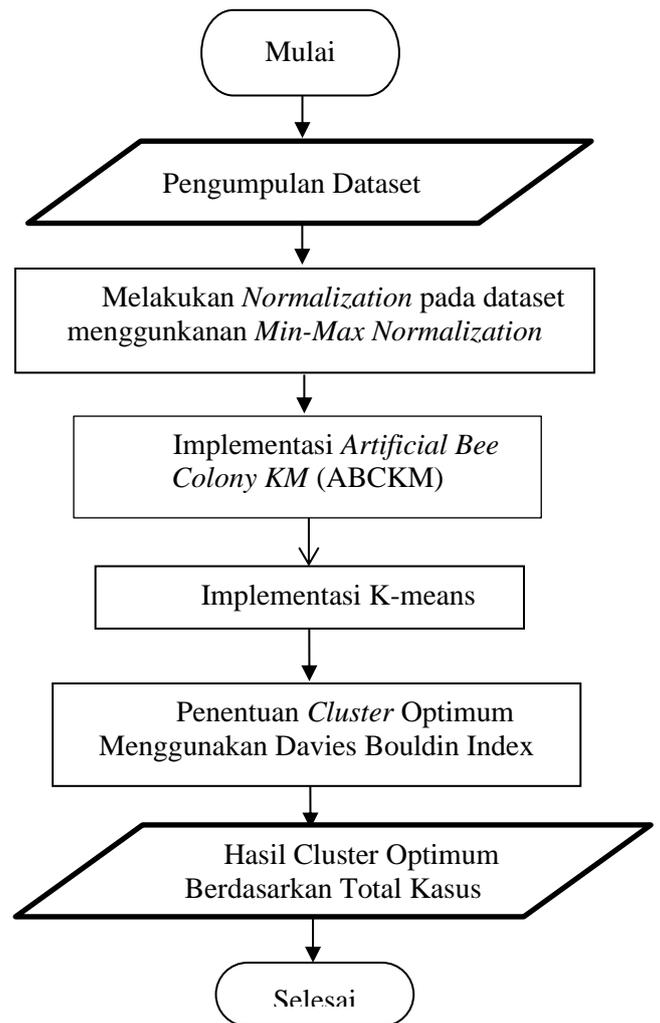
### 3. METODOLOGI PENELITIAN

#### A. Pengumpulan Data

Pada bab ini akan melakukan pengumpulan dataset dan memperoleh sebuah informasi untuk mencapai tujuan penelitian, disini peneliti memperoleh data persebaran kasus covid-19 di provinsi Indonesia dari Twiter (@kawalcovid19) pada bulan april sampai bulan agustus pada tahun 2020.

#### B. Tahapan Penelitian

Pada penelitian yang mengimplementasikan *Artificial Bee Colony k-means* mengelompokkan provinsi di Indonesia berdsarkan indeks kinerja *Davies Bouldin* pada kasus covid19 ini terdapat beberapa tahapan penelitian seperti yang dijabarkan pada diagram metode penelitian dibawah ini :



**Gambar 1.** Flowchart Metodologi Penelitian  
 Sumber : Hasil Perencanaan

#### C. *Min-Max Normalization*

Pada dataset covid-19 terdapat nilai asli sebelum dilakukan proses *normalization*. Banyaknya dataset yang memiliki rentang berbeda, sehingga perlu dilakukan proses normalisasi. Dataset persebaran covid-19 akan dirubah menggunakan metode *min-max normalization* dengan cara mengolah nilai minimum dan nilai maksimum dari setiap atribut. Rentang yang digunakan antara 0 sampai 1. Berikut dataset yang sudah mengalami proses *normalization* :

**Tabel 1.** Dataset Covid19 Pada Bulan April Sampai Agustus

Total Kasus	Aceh	Bali	Banten	..	Gorontalo
01-Apr	5	25	152		0
02-Apr	5	25	164		0
03-Apr	5	27	170		0
04-Apr	5	32	173		0
05-Apr	5	35	177		0
06-Apr	5	43	187		0
07-Apr	5	43	194		0
08-Apr	6	49	212		0
09-Apr	6	63	218		0
10-Apr	5	75	243		1
11-Apr	5	79	279		1
12-Apr	5	81	281		1
13-Apr	5	86	285		1
14-Apr	5	92	280		1
15-Apr	5	98	281		1
16-Apr	5	113	297		4
17-Apr	5	124	311		4
18-Apr	6	131	321		4
19-Apr	7	135	324		4
20-Apr	7	140	341		4
21-Apr	7	150	341		7
22-Apr	7	152	337		7
23-Apr	7	167	337		7
24-Apr	8	177	359		12
25-Apr	9	183	370		14
26-Apr	9	186	370		15
27-Apr	9	194	382		14
28-Apr	9	215	388		15
29-Apr	9	215	388		15
.....					
31 Aug	1,633	5,207	2,903		2,041

Sumber : twitter @kawalcovid19

#### D. Metode Artificial Bee Colony Kmeans

Pada metode *Artificial Bee Colony Algorithm* mempunyai 3 tahapan kelompok lebah, yaitu: *employed bee phase*, *onlooker bee*

*phase*, dan *scouts bee phase*. Pada metode *Artificial Bee Colony K-means*, langkah pertama yaitu melakukan proses inialisasi pada semua parameter yang diperlukan (*colony size*, *maksiterasi*, dan *limit*). Pada inialisasi untuk setiap sumber makanan dan solusi dihasilkan secara acak, kemudian menghitung nilai *fitness* sesuai dengan inialisasi solution yang diinginkan.

Kemudian pada *fase employed bee phase*, pada setiap *employed bee* akan mengupdate *employed bee* dengan menggunakan *neighborhood operator*, *swab operator* dan *swab sequence*, setelah itu menghitung nilai *fitness* pada setiap *employed bee* yang diperoleh, jika pada solusi yang baru nilai *fitness* hasilnya lebih baik dari sebelumnya, maka lebah tersebut akan mengubah solusi lama dengan solusi baru. Setelah mendapatkan nilai *fitness employed bee* kemudian menghitung nilai probabilitas pada tiap *employed bee phase*.

Pada *fase onlooker bee phase* untuk setiap *onlooker bee* memilih solusi dengan menggunakan teknik *roulette wheel selection* kemudian menentukan solusi baru dari *employed bee* yang terpilih menggunakan *neighborhood operator*, *insert operator* dan *insert sequence*. Kemudian menghitung nilai *fitness* pada masing-masing *onlooker bee*, jika solusi yang baru menghasilkan sumber makanan yang lebih baik, maka solusi lama diganti dengan solusi yang baru.

Pada *fase scout bee phase* menghitung jumlah *trial* kemudian akan menyimpan jumlah nilai maksimal untuk masing-masing sumber makanan yang tidak mengalami peningkatan solusi.

#### E. Metode Davies Bouldin Index

Pada proses pengelompokan menggunakan algoritma *Artificial Bee Colony K-means* yang telah dilakukan sebelumnya, di dapat hasil 2 *cluster* sampai *cluster* ke 10. Selanjutnya diterapkan model *Davies Bouldin Index* fungsinya untuk menentukan jumlah *cluster* mana yang merupakan *cluster optimum*.

#### 4. HASIL DAN PEMBAHASAN

Data yang digunakan merupakan data penyebaran Covid-19 diseluruh provinsi Indonesia pada bulan April sampai bulan Agustus 2020. Variable yang digunakan yaitu dari total kasus yang didapat selama pada bulan april sampai desember. Setelah itu data tersebut akan dinormalisasikan terlebih dahulu kemudian diolah menggunakan Algoritma *Artificial Bee Colony Kmeans* (ABCKM) untuk mendapatkan nilai sumber makanan terbaik.

**Tabel 2.** Dataset covid 19 yang sudah di normalisasi.

Total Kasus	01-Apr	02-Apr	..	31 Aug
Aceh	0,0045	0,0045		0,0007
Bali	0,0246	0,0246		0,0043
Banten	0,1528	0,1650		0,0019
Babel	0,0000	0,0000		1,0000
Bengkulu	0,0000	0,0000		1,0000
DIY	0,0272	0,0262		0,0004
Jakarta	0,8320	0,9237		0,0403
Jambi	0,0000	0,0000		1,0000
Jabar	0,2262	0,2293		0,0104
Jateng	0,1042	0,1042		0,0131
Jatim	0,1076	0,1076		0,0340
Kalbar	0,0000	0,0000		1,0000
Kaltim	0,0201	0,0201		0,0032
Kalteng	0,0080	0,0080		0,0016
Kalsel	0,0072	0,0072		0,0075
Kaltara	0,0000	0,0000		1,0000
Kep Riau	0,0000	0,0000		1,0000
NTB	0,0051	0,0051		0,0018
Sumsel	0,0040	0,0100		0,0035
Sumbar	0,0071	0,0071		0,0012
Sulut	0,0020	0,0020		0,0029
Sumut	0,0211	0,0211		0,0058
Sultra	0,0020	0,0020		0,0005
Sulsel	0,0684	0,0684		0,0115
Sulteng	0,0000	0,0000		1,0000
Lampung	0,0000	0,0000		1,0000
Riau	0,0020	0,0061		0,0008
Malut	0,0000	0,0000		0,0009
Maluku	0,0000	0,0000		0,0009
Papbar	0,0000	0,0000		1,0000
Papua	0,0098	0,0098		0,0030

Total Kasus	01-Apr	02-Apr	..	31 Aug
Sulbar	0,0000	0,0000		1,0000
NTT	0,0000	0,0000		1,0000
Gorontalo	0,0000	0,0000		0,0021

Sumber : Hasil Perhitungan

##### A. Algoritma *Artificial Bee Colony K-means*

Dataset sumber makanan terbaik akan di *cluster* dengan melakukan pembentukan *cluster* menggunakan *K-means* dengan sumber makanan terbaik yang akan digunakan sebagai *centroid* awal, dengan menghitung jumlah data dari tiap *cluster* yang terbentuk.

Columns 145 through 153

0.6351	0.4656	0.2031	0.2768	0.4235	0.4241	0.8225	0.8675	0.6804
0.8009	0.0817	0.5729	0.8285	0.0428	0.6673	0.5733	0.1711	0.2883
0.2927	0.3835	0.3497	0.2751	0.6922	0.7469	0.7429	0.7844	0.5931
0.7822	0.3785	0.5258	0.3925	0.0070	0.9598	0.7466	0.7263	0.4987
0.0241	0.9186	0.7157	0.9337	0.6359	0.2512	0.4077	0.0088	0.3621
0.3639	0.9250	0.8020	0.8147	0.4538	0.2537	0.4802	0.5200	0.9521
0.5740	0.2441	0.4534	0.3650	0.4943	0.6502	0.1989	0.3702	0.2633
0.2640	0.2195	0.5652	0.0966	0.1985	0.9312	0.9918	0.4810	0.0873
0.4389	0.7562	0.4333	0.6302	0.4541	0.8779	0.6387	0.3443	0.3667
0.3159	0.4297	0.6601	0.5964	0.7679	0.7592	0.5820	0.6505	0.5298

**Gambar 2.** Hasil Centroid Pada ABCKM

Sumber : Hasil Perhitungan

```

Replicate 1, 1 iterations, total sum of distances = 825.843.
Replicate 2, 1 iterations, total sum of distances = 807.559.
Replicate 3, 1 iterations, total sum of distances = 815.992.
Replicate 4, 2 iterations, total sum of distances = 790.675.
Replicate 5, 1 iterations, total sum of distances = 829.037.
Best total sum of distances = 790.675
    
```

**Gambar 3.** Hasil Nilai Iterasi Pada ABCKM

Sumber : Hasil Perhitungan

##### B. Implementasi *Davies Bouldin Index* Pada ABCKM

Hasil centroid yang di dapat dari metode ABCKM kemudian diterapkan pada metode DBI untuk mendapatkan nilai *cluster* yang optimum. Berikut hasil dari *davies bouldin index* dari 2 sampai 10 *cluster*.

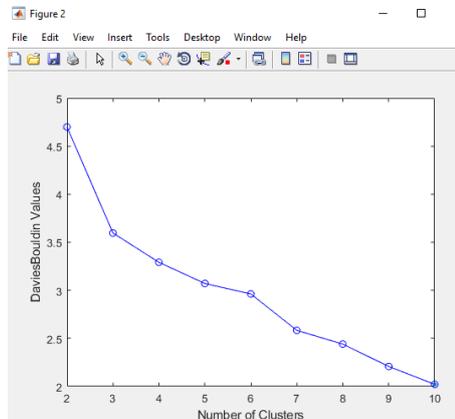
**Tabel 3.** Hasil DBI 2 sampai 10 klaster

Cluster	Nilai DBI
2	4.6991
3	5.5962
4	3.2914
5	3.0710

Cluster	Nilai DBI
6	2.9622
7	2.5835
<b>8</b>	<b>2.4406</b>
9	2.2068
10	<b>2.0218</b>

Sumber : Hasil Perhitungan

Selain itu, perintah pada Matlab juga dapat menampilkan *cluster* kedalam bentuk *grafik*. Adapun contoh *grafik* yang merupakan hasil dari *Artificial Bee Colony Kmeans* pada Matlab yaitu berupa **Gambar 4.15** Plot Hasil Cluster Davies Bouldin Index Pada ABCKM



Sumber : Hasil Perhitungan Plot Matlab

### C. Implementasi K-means Pada Matlab

Pada tahap ini, dataset yang sudah dinormalisasi akan melakukan clusterisasi untuk mendapatkan nilai centroid awal secara random.

```
stream = RandStream('mlfg6331_64'); % Random number stream
options = statset('UseParallel',1,'UseSubstreams',1,...
    'Streams',stream);
tic; % Start stopwatch timer
[idx,C,sumd,D] = kmeans(covid19,10,'Options',options,'MaxIter',10000,...
    'Display','final','Replicates',5);
```

**Gambar 4.** Implementasi Kmeans Pada Matlab

Pada fungsi input *Randstream* digunakan untuk menentukan algoritma *generator fibonanci* sedangkan pada fungsi *options* untuk menentukan jumlah cluster dalam data dan meningkatkan jumlah iterasi untuk menemukan nilai minimum atau nilai yang lebih rendah. Pada fungsi `[idx,C,sumd,D]` untuk menghitung

jarak dari setiap titik ke setiap titik centroid. Berikut hasil centroid yang didapat oleh *k-means* :

Columns 145 through 153

0.8545	0.8625	0.8821	0.8990	0.9174	0.9431	0.9601	0.9742	1.0000
0.0113	0.0114	0.0115	0.0116	0.0118	0.0119	0.0121	0.0122	0.0123
0.0196	0.0198	0.0201	0.0204	0.0207	0.0212	0.0216	0.0219	0.0222
0.0113	0.0115	0.0118	0.0120	0.0124	0.0127	0.0131	0.0135	0.0139
0.0016	0.0016	0.0017	0.0017	0.0018	0.0019	0.0020	0.0020	0.0021
0.9518	0.9581	0.9646	0.9709	0.9723	0.9841	0.9890	0.9966	1.0000
0.0016	0.0016	0.0016	0.0016	0.0017	0.0017	0.0017	0.0017	0.0018
0.0002	0.0002	0.0003	0.0003	0.0004	0.0004	0.0004	0.0005	0.0005
0.0006	0.0007	0.0007	0.0008	0.0009	0.0009	0.0010	0.0011	0.0012
0.0035	0.0036	0.0036	0.0037	0.0038	0.0038	0.0039	0.0039	0.0040

**Gambar 5.** Hasil Centroid Kmeans  
 Sumber : Hasil Perhitungan Matlab

```
Starting parallel pool (parpool) using the 'local' profile ...
connected to 2 workers.
Replicate 1, 2 iterations, total sum of distances = 69.8218.
Replicate 3, 1 iterations, total sum of distances = 74.2223.
Replicate 2, 1 iterations, total sum of distances = 66.2855.
Replicate 4, 1 iterations, total sum of distances = 72.4372.
Replicate 5, 4 iterations, total sum of distances = 67.8294.
Best total sum of distances = 66.2855
```

**Gambar 6.** Hasil Nilai Iterasi Pada Kmeans  
 Sumber : Hasil Perhitungan Matlab

### D. Implementasi Davies Bouldin Index Pada K-means

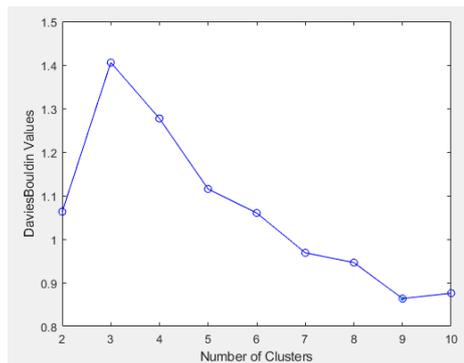
Hasil centroid yang di dapat dari metode K-means kemudian diterapkan pada metode DBI untuk mendapatkan nilai *cluster* yang *optimum*. Berikut hasil dari *davies bouldin index* dari 2 sampai 10 *cluster*.

**Tabel 4.** Hasil DBI 2 sampai 10 cluster

Cluster	Nilai DBI
2	1.0631
3	1.4056
4	1.2771
5	1.1153
6	1.0602
7	0.9687
8	0.9462
9	<b>0.8637</b>
10	0.8761

Sumber : Hasil Perhitungan

Selain itu, perintah pada Matlab juga dapat menampilkan *cluster* kedalam bentuk *grafik*. Adapun contoh *grafik* yang merupakan hasil dari *Kmeans* pada Matlab yaitu berupa Plot Hasil DBI Pada Kmeans



Sumber : Hasil Perhitungan Plot Matlab

### E. Hasil Cluster Optimal pada ABCKM Dan Kmeans

Pada metode ABCKM hasil *cluster* yang optimal ditunjukkan pada *cluster* 10 dengan nilai 2.0218 sedangkan pada *Kmeans* hasil *cluster* optimal ditunjukkan pada *cluster* 9 dengan nilai 0.8637. Berikut hasil pada metode ABCKM dan *Kmeans*.

**Tabel 5.** Cluster Optimal Nilai DBI Metode ABCKM.

Cluster	Nilai DBI
2	4,6991
3	5,5962
4	3,2914
5	3,0710
6	2,9622
7	2,5835
8	2,4406
9	2,2068
10	<b>2,0218</b>

Sumber : Hasil Perhitungan

**Tabel 6.** Cluster Optimal Nilai DBI Metode KM.

Cluster	Nilai DBI
2	1,0631
3	1,4056
4	1,2771
5	1,1153
6	1,0602
7	0,9687
8	0,9462
9	<b>0,8637</b>
10	0,8761

Sumber : Hasil Perhitungan

Pada metode *Davies Bouldin Index* nilai yang bagus terjadi jika nilai *cluster* memiliki nilai yang rendah. Sedangkan Nilai *cluster Davies Bouldin Index* pada metode *Artificial Bee Colony KM* lebih besar daripada nilai *cluster* pada metode *Kmeans*. Jadi Optimasi *Artificial Bee Colony KM* tidak dapat digunakan pada metode *Kmeans*, dikarenakan nilai *cluster* yang optimum terjadi pada metode *Kmeans* dengan nilai DBI 0.8637.

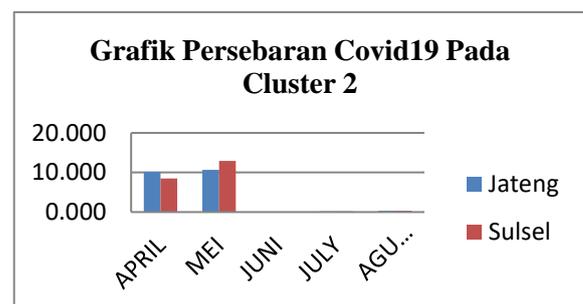
### F. Profiling Cluster K-means

Berdasarkan proses *clustering* menggunakan algoritma *K-Means* dalam proses pencarian *cluster* optimal menggunakan metode *Davies Bouldin Index* (DBI) didapat *cluster* optimal pada *cluster* 9. Supaya dapat mengetahui karakteristik pada *cluster*, maka dilakukan *profiling cluster* dari setiap provinsi yang masuk pada *cluster* 1 sampai *cluster* 10. Salah satu contoh hasil dari 1 sampai 10 *cluster* adalah sebagai berikut:

**Tabel 7.** Anggota Pada Cluster 2

Bulan	Jateng	Sulsel
APRIL	10,191	8,528037
MEI	10,690	12,94608
JUNI	0,039	0,066284
JULY	0,172	0,211066
AGUSTUS	0,332	0,322907

Sumber : Hasil Perhitungan



Sumber : Hasil Perhitungan

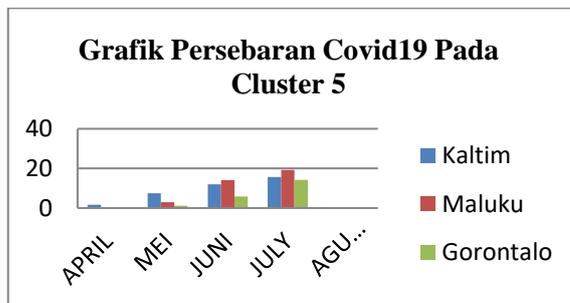
Pada tabel 7 menunjukkan hasil *cluster* 2 yaitu pada provinsi Jawa Tengah dan Sulawesi Selatan berdasarkan grafik total kasus yang diperoleh pada bulan april sampai agustus. Pada bulan april dan mei di provinsi Jawa Tengah

masih mengalami kenaikan persebaran covid19, sedangkan di provinsi Sulawesi Selatan pada bulan april mengalami penurunan persebaran covid19 tetapi pada bulan mei provinsi tersebut mengalami kenaikan. Sedangkan pada bulan juni, july dan agustus kedua provinsi tersebut mengalami penurunan persebaran covid19 yang stabil .

**Tabel 8.** Anggota Pada Cluster 5

Bulan	Kaltim	Maluku	Gorontalo
APRIL	1,671739	0,317954	0,153605
MEI	7,440281	2,913741	1,114943
JUNI	11,9591	14,1344	5,872518
JULY	15,61775	19,23616	14,25917
AGUSTUS	0,047286	0,015069	0,055833

Sumber : Hasil Perhitungan



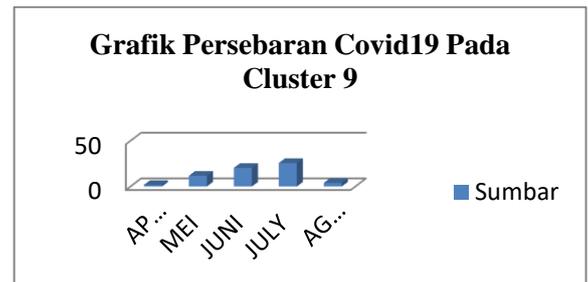
Sumber : Hasil Perhitungan

Pada tabel 8 menunjukkan hasil cluster 5 yaitu pada provinsi Kalimantan Timur, Maluku dan Gorontalo. Pada bulan april, mei, juni dan july persebaran kasus covid19 mengalami kenaikan. Sedangkan pada bulan agustus tidak dapat memunculkan grafik persebaran covid19 dikarenakan nilai data pada bulan agustus tidak cukup untuk ditampilkan.

**Tabel 9.** Anggota Pada Cluster 9

Bulan	Sumbar
APRIL	1,790903
MEI	11,75749
JUNI	20,37519
JULY	25,72613
AGUSTUS	3,929656

Sumber : Hasil Perhitungan



Sumber : Hasil Perhitungan

Pada tabel 9 menunjukkan hasil cluster 9 yaitu provinsi Sumbar, berdasarkan grafik total kasus yang diperoleh pada bulan april sampai bulan july mengalami kenaikan persebaran kasus covid19, sedangkan pada bulan agustus mengalami penurunan persebaran covid19.

## 5. KESIMPULAN DAN SARAN

### A. Kesimpulan

Berdasarkan uraian permasalahan dan pembahasan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Jumlah cluster optimum yang didapat untuk pengelompokkan provinsi di Indonesia berdasarkan kasus persebaran covid19 menggunakan algoritma ABCKM yaitu sebanyak 1 kasus dan terdapat pada cluster 10 dengan nilai Davies Bouldin Index 2.0218.
2. Provinsi yang termasuk dalam cluster terendah yaitu provinsi Sumatra barat yang terdapat pada cluster 9.

### B. Saran

Saran pada penelitian tugas akhir ini yaitu sebagai berikut :

1. Pada validasi *clusters* dalam menemukan *clusters* yang optimal bisa menggunakan alternative lain selain metode *Davies Bouldin Index*, contohnya seperti metode *Silhouette*, *Elbow*, *Gap Statistic*, dll.
2. Pada penelitian selanjutnya dapat menggunakan algoritma *Artificial Bee ColonyKM* tetapi dengan tools yang berbeda contohnya seperti *Phyton*, *Rstudio*, dll.

#### Daftar Pustaka

- Bansal, J. C., Sharma, H., & Jadon, S. S. (2013). Artificial bee colony algorithm: a survey. *Int. J. Advanced Intelligence Paradigms*, 5, 123-157.
- Cahyono, B. (2013). PENGGUNAAN SOFTWARE MATRIX LABORATORY (MATLAB) DALAM PEMBELAJARAN ALJABAR LINIER. *PHENOMENON*, 1, 45-62.
- Darnisa Azzahra Nasution, H. H. (2019). PERBANDINGAN NORMALISASI DATA UNRUK KALSIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN. *CESS (JURNAL OF COMPUTER ENGINEERING SYSTEM AND SCIENCE)*, 4, 78-82.
- Dwitri, N., Tampubolon, J. A., Prayoga, S., Ilmi, F., & Hartama, D. (2020). PENERAPAN ALGORITMA K-MEANS DALAM MENENTUKAN TINGKAT PENYEBARAN PANDEMI COVID-19 DI INDONESIA. *Teknologi Informasi*, 4, 128-132.
- Karaboga, D., 2005. An Idea Based on honey bee swarm for numerical optimization, Turkey: Erciyes University, Engineering Faculty, Computer Engineering Departement.
- Karaboga, D., & Ozturk, C. (2011). A novel clustering approach: Artificial Bee Colony (ABC) algorithm. *Applied Soft Computing*, 652-657.
- Karaboga, D., & Ozturk, C. (2010). Fuzzy clustering with artificial bee colony algorithm. *Scientific Research and Essays*, 5(14), 1899-1902.
- Nugroho, R. (2020, May 01). Kasus Covid-19 di Indonesia Selama April dan Prediksi Bulan Mei Halaman all. Retrieved August 10, 2020, from <https://www.kompas.com/tren/read/2020/05/01/190604465/kasus-covid-19-di-indonesia-selama-april-dan-prediksi-bulan-mei?page=all>.
- Sanur, D. (2020). WACANA KEBIJAKAN LOCKDOWN DALAM MENGHADAPI COVID-19 DI INDONESIA. 25-30.
- Syafrida, R. H. (2020). Bersama Melawan Virus Covid 19 di Indonesia. *FSH UIN Syarif Hidayatullah Jakarta*, 495-508.
- Wimmer, H. (2018). Effects of Normalization Techniques on Logistic Regression in Data Science. *Proceedings of the Conference on Information Systems Applied Research*, 1-9.
- Zahrotunnimah. (2020). Langkah Taktis Pemerintah Daerah Dalam Pencegahan Penyebaran Virus Corona Covid-19 Di Indonesia. *Sosial & Budaya Syar-i*, 7, 247-260.