

## Penerapan Data Mining Pada Daerah Penyebaran Dbd Di Jawa Barat Dengan Metode K-Means Clustering

Maharanisa

STIKOM CKI Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika

\*email: [mahara.nisa16@gmail.com](mailto:mahara.nisa16@gmail.com)

Received: 16-08-2022, Revised: 15-08-2022, Accepted: 18-10-2022

### Abstrak

*Demam Berdarah Dengue (DBD) adalah penyakit yang disebabkan oleh virus Dengue yang ditularkan melalui gigitan nyamuk. Penyakit DBD ini dapat berkenaan kondisi lingkungan dan perilaku masyarakat. Dengan ini membahas tentang bagaimana pengelompokan jumlah daerah yang terjangkit DBD di daerah Provinsi Jawa Barat tahun 2016. Metode yang diterapkan ialah Data Mining Algoritma K-Means Clustering dimana penerapan proses K-Means Clustering menggunakan tools RapidMiner. Dengan menggunakan metode ini maka data set yang sudah telah diperoleh dapat dikelompokkan menjadi beberapa cluster, dimana penerapan metode K-Means Clustering Penelitian ini mencari sumber data search di web site Open Data Jawa Barat. Hasil yang diperoleh bahwa terdapat 3 penyebaran penyakit DBD dengan Cluster tingkat tinggi (C1) yaitu Kota Bandung dengan data kasus DBD berdasarkan Jenis Kelamin Laki-Laki (L) : 2054 Jiwa dan Perempuan (P) : 1827 Jiwa dengan jumlah kasus 3881 Jiwa., Cluster tingkat sedang (C2) yaitu Kota Bogor dengan data kasus DBD berdasarkan Jenis Kelamin Laki-Laki (L) : 658 Jiwa dan Perempuan (P) : 566 Jiwa dengan jumlah kasus 1224 Jiwa., dan Cluster tingkat rendah (C3) yaitu Kabupaten Pangandaran dengan data kasus DBD berdasarkan Jenis Kelamin Laki-Laki (L) : 73 Jiwa dan Perempuan (P) : 70 Jiwa dengan jumlah kasus 143 Jiwa. Hal ini dapat menjadi himbauan kepada masyarakat untuk menjaga kesehatan dengan meningkatkan pencegahan terhadap penularan penyakit demam berdarah, sehingga meningkatkan kepedulian peran aktif masyarakat dengan bergotong royong dalam kebersihan dan dengan melalui aktivitas pemberantasan nyamuk dilingkungannya dan jentik nyamuk secara berkala.*

**Kata Kunci :** Demam Berdarah Dengue (DBD); Data Mining; K-Means; Clustering

### Abstract

*Dengue Hemorrhagic Fever (DHF) is a disease caused by the dengue virus which is transmitted through mosquito bites. This DHF disease can be related to environmental conditions and community behavior. Hereby discusses how to group the number of areas affected by DHF in the West Java Province in 2016. The method applied is the Data Mining Algorithm K-Means Clustering. where the application of the K-Means Clustering process uses RapidMiner tools. By using this method, the data set that has been obtained can be grouped into several clusters, where the application of the K-Means Clustering method. The results obtained that there are 3 distributions of DHF disease with a high level Cluster (C1), namely the City of Bandung with DHF case data based on Gender Male (L) : 2054 souls and Female (F) : 1827 people with the number of cases 3881 people., The medium-level cluster (C2) is Bogor City with data on DHF cases based on Gender Male (M): 658 people and Female (F): 566 people with 1224 cases, and the low-level cluster (C3) is Pangandaran*





*Regency. with data on cases of DHF based on Gender Male (M): 73 people and Female (F): 70 people with a total of 143 people. This can be an appeal to the community to maintain health by increasing prevention of transmission of dengue fever, thereby increasing awareness of the active role of the community by working together in cleanliness and through activities to eradicate mosquitoes in their environment and mosquito larvae on a regular basis.*

**Key Words :** *Dengue Hemorrhagic Fever (DHF); Data Mining; K-Means; Clustering*

## 1. Pendahuluan ( or Introduction )

**Demam Berdarah Dengue (DBD)** merupakan penyakit yang dapat disebabkan oleh virus dengue yang dapat ditularkan ke manusia melalui gigitan nyamuk *Aedes aegypti* dan *Aedes albopictus*. Di negara kita, Indonesia merupakan daerah endemis dengan penyebaran yang beragam di seluruh tanah air. Peran masyarakat terhadap lingkungan yang kurang memperhatikan kebersihan sangat mempengaruhi kesehatan.

Konsep Data Mining ini yang diterapkan dalam penelitian ini untuk mengetahui penyakit demam berdarah dengue (DBD) di provinsi Jawa Barat. Pengelompokan berdasarkan titik lokasi di beberapa daerah provinsi Jawa Barat yang biasa dijadikan tempat beraktivitas seperti pasar, tempat pembuangan sampah, tempat-tempat yang lembab.

**Data mining** suatu proses pengumpulan dan pengolahan data untuk mendapatkan berbagai model dalam data. Proses ini otomatis atau sering semi-otomatis. model yang dapat ditemukan harus bermakna dan memberikan manfaat, biasanya manfaat secara ekonomi. Sejumlah besar data diperlukan (Witten, 2005).

**Clustering** yaitu kluster yang membentuk pada pengelompokan seperti record, observasi, atau mengamati dan membentuk kelas objek yang memiliki kesamaan. Cluster ialah kumpulan record yang memiliki kesamaan satu sama lain, dan tidak sama dengan record di klaster lain. Clustering mencoba membagi seluruh kumpulan data kedalam kelompok yang relatif sama, di mana kesamaan record dalam satu kelompok akan bernilai maksimal, sedangkan kesamaan record pada kelompok lain akan bernilai minimal.

**K-Means** adalah Algoritma yang digunakan untuk mengclustering file atau dokumen. Prinsip utama K-Means yaitu menyusun k prototipe atau pusat massa (centroid) dari sekumpulan data dimensional, sebelum menggunakan proses algoritma K-Means, file atau dokumen akan di proses terlebih dahulu. Kemudian file atau data tersebut direpresentasikan sebagai vektor yang terdapat term dengan nilai tertentu.

Untuk membentuk sistem yang dapat mengelompokan data penyebaran penyakit demam berdarah dengan menerapkan metode K-Means Clustering akan menghasilkan titik-titik pusat penyebaran penyakit demam berdarah, untuk mencegah semakin meningkatnya penyebaran penyakit demam berdarah.

Penelitian ini telah ditentukan agar tidak melenceng dari pokok bahasan yang akan dibahas oleh penulis dan demi terciptanya tujuan penelitian yang peneliti rencanakan hanya meliputi





daerah Provinsi Jawa Barat berdasarkan jenis kelamin pada tahun 2016. dari website <https://opendata.jabarprov.go.id>

## 2. Tinjauan Literatur (or Literature Review)

Kajian pustaka yang digunakan penulis yaitu Dalam melakukan pencarian dari sumber-sumber penelitian sebelumnya dilakukan tahapan dalam mendapatkan jurnal-jurnal penelitian yang telah di publikasikan untuk itu dilakukan Strategi Seleksi Studi atau Studies Selection Strategy :

Identifikasi Masalah, dan Tujuan Penelitian dari hasil tinjauan pustaka terkait judul yang ditulis maka penulis merangkum secara keseluruhan adalah sebagai berikut :

### 2.1 Identifikasi Masalah (Research Problem)

Terjadinya penumpukan data yang mengakibatkan penurunan akurasi dalam mengantisipasi penyebaran Demam Berdarah Dengue (DBD) di Jawa Barat. Dari penelitian ini terdapat banyak data yang dilakukan mengenai pengelompokan data.

### 2.2 Rumusan Masalah (Research Question)

Bagaimana implementasi Data Mining untuk mengelompokkan penyebaran penyakit demam berdarah (DBD) menggunakan Metode K-Means Clustering ?

### 2.3 Tujuan Penelitian (Research Objective)

Untuk dapat mengelompokkan data penyebaran penyakit demam berdarah, maka penelitian dengan menerapkan metode K-Means Clustering ini akan dapat menghasilkan titik-titik pusat penyebaran penyakit demam berdarah, sehingga memudahkan berbagai pihak untuk melakukan pencegahan meningkatnya penyebaran penyakit demam berdarah.

## 3. Metode Penelitian (or Research Method)

Berdasarkan penelitian ini penulis memanfaatkan alat penelitian berupa perangkat keras dan perangkat lunak. Perangkat keras yang berupa Laptop penulis ASUS I Core 3 dan Perangkat Lunak berupa :

1. Sistem Operasi windows 11
  2. Broser Mozila Firefox dan Google chrome
  3. Microsoft Excel & Microsoft Word
  4. Notepat ++
  5. Rapidminer Studio
- Studi Literatur  
Mengumpulkan dan Mempelajari berbagai informasi berupa artikel, dan jurnal, atau dokumen-dokumen yang berkaitan dengan materi metode ini yaitu K-Means Clustering yang di kaji. Dapatkan berbnagai informasi yang diperoleh dari studi kepustakaan ini dibuat rujukan untuk menambah pengetahuan yang ada.
  - Pengumpulan Data  
Dalam melakukan penelitian, penulis mengambil data set pada open data jawa barat





di website <https://opendata.jabarprov.go.id>

Jawa barat secara langsung tentang penyebaran penyakit dbd pada tahun 2016

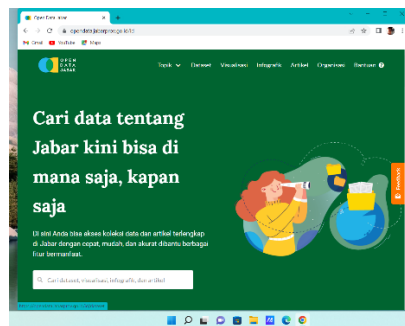
- Standarisasi Data

Jika besaran angka cukup jauh antara variabel satu dengan yang lainnya yang dapat menyulitkan proses pengelompokan yang membuat data menjadi tidak valid Clustering

dengan ini jumlah clustering dibagi menjadi tiga kelompok cluster yaitu cluster 1 atau C1 yaitu bulan tingkat penyebaran relatif tinggi, cluster 2 atau C2 yaitu bulan tingkat penyebaran sedang, cluster 3 atau C3 tingkat bulan penyebaran relatif rendah.

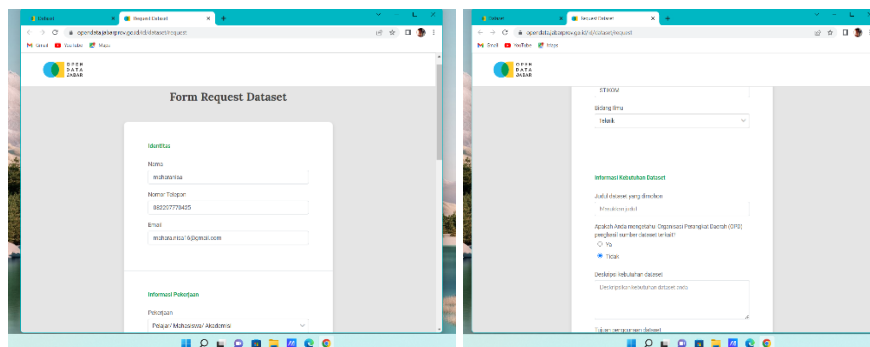
- Kelompokan setiap data ke jarak terdekat pusatnya Pengalokasikan ke setiap kelompok centroid.

Berikut proses penelitian pencarian data set DBD 2016 provinsi Jawa Barat, Kasus DBD Berdasarkan Jenis Kelamin,



Gambar 1 Analisa pencarian data set di open data jawa barat

Sebelum mencari data, mengisi form akun pengguna pencarian data set,



Gambar 2 Mengisi Form Request Dataset

Setelah Menemukan data set sesuai pencarian informasi yang dibutuhkan, di unduh data set file ekstraksi Excel.

#### 4. Hasil dan Pembahasan (or Results and Analysis)

Jawa Barat ialah suatu provinsi yang berada di negara Indonesia. Ibu kotanya Jawa Barat berada di Kota Bandung. Jawa Barat sebagai provinsi dengan jumlah masyarakat



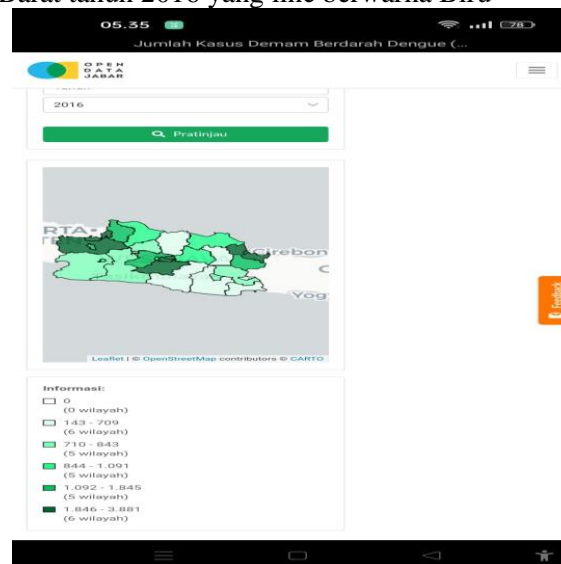


terbanyak di negara Indonesia. Provinsi Jawa Barat terdiri atas 18 kabupaten dan 9 kota, dengan ini penulis menggunakan data set provinsi jawa barat yang akan dijadikan sampel sebagai data untuk penerapan Data Mining.

Berikut ini gambar hasil pencarian data set kasus DBD berdasarkan jenis kelamin pada tahun 2016 di Jawa Barat :



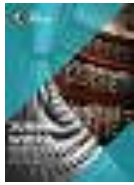
Gambar 3 Grafik Line Chart kasus DBD Berdasarkan Jenis Kelamin di Jawa Barat tahun 2016 yang line berwarna Biru



Gambar 4 Peta StreetMap Hasil Kasus DBD 2016 berdasarkan jenis kelamin

Data ini didapatkan dari web site pemerintahan Jawa Barat yaitu open data jabar. Data set ini dalam bentuk file Microsoft Excel yang didapat oleh penulis. Dalam hal ini dikarenakan pada penelitian dengan pencarian data set mengunduh file data set, saat penulis mengerjakan pengolahan data pada microsoft excel dan menggunakan Aplikasi RapidMiner. Adapun data yang diolah pada penelitian bagian ini berikut tabel data set kasus DBD pada tahun 2016 berdasarkan jenis kelamin dan jumlah kasusnya.





Tabel 1. Data Set Kasus DBD di Jawa Barat pada Tahun 2016

No.	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_kelamin		jumlah_kasus
					LAKI-LAKI (L)	PEREMPUAN (P)	
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	1725	1699	3424
2	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	419	374	793
3	32	JAWA BARAT	3203	KABUPATEN CIANJUR	434	365	799
4	32	JAWA BARAT	3204	KABUPATEN BANDUNG	1664	1708	3372
5	32	JAWA BARAT	3205	KABUPATEN GARUT	291	225	516
6	32	JAWA BARAT	3206	KABUPATEN TASIKMALAYA	229	250	479
7	32	JAWA BARAT	3207	KABUPATEN CIAMIS	387	361	748
8	32	JAWA BARAT	3208	KABUPATEN KUNINGAN	933	781	1714
9	32	JAWA BARAT	3209	KABUPATEN CIREBON	940	937	1877
10	32	JAWA BARAT	3210	KABUPATEN MAJALENGA	176	152	328
11	32	JAWA BARAT	3211	KABUPATEN SUMEDANG	475	433	908
12	32	JAWA BARAT	3212	KABUPATEN INDRAMAYU	483	477	960
13	32	JAWA BARAT	3213	KABUPATEN SUBANG	428	271	699
14	32	JAWA BARAT	3214	KABUPATEN PURWAKARTA	424	361	785
15	32	JAWA BARAT	3215	KABUPATEN KARAWANG	546	513	1059
16	32	JAWA BARAT	3216	KABUPATEN BEKASI	910	703	1613
17	32	JAWA BARAT	3217	KABUPATEN BANDUNG BARAT	815	739	1554
18	32	JAWA BARAT	3218	KABUPATEN PANGANDARAN	73	70	143
19	32	JAWA BARAT	3271	KOTA BOGOR	658	566	1224
20	32	JAWA BARAT	3272	KOTA SUKABUMI	431	511	942
21	32	JAWA BARAT	3273	KOTA BANDUNG	2054	1827	3881
22	32	JAWA BARAT	3274	KOTA CIREBON	590	521	1111
23	32	JAWA BARAT	3275	KOTA BEKASI	1972	1841	3813
24	32	JAWA BARAT	3276	KOTA DEPOK	1285	1270	2555
25	32	JAWA BARAT	3277	KOTA CIMAH	483	516	999
26	32	JAWA BARAT	3278	KOTA TASIKMALAYA	381	373	754
27	32	JAWA BARAT	3279	KOTA BANJAR	175	193	368

Jarak setiap data ke pusat cluster berikut rumus menghitung jarak ke antara titik tiap objek menentukan persamaan Euclidean Distance, yaitu jarak garis lurus biasa antara dua titik dalam ruang Euclidean selanjutnya dengan mengelompokan data Cluster, dengan Rumus seperti dibawah ini :

$$De = \sqrt{(xi - Si)^2 + yi - ti)^2}$$

Keterangan :

De = Euclidean Distance

xi = Banyaknya objek / Data kriteria

(x,y) = Nilai Koordinat objek x dan Koordinat y

(s,t) = Nilai Koordinat centroid pada cluster ke-

Dari Tabel 1, hasilnya di titik awal cluster berdasarkan jenis kelamin pada tahun 2016 di Jawa Barat. Peneliti menggunakan variabel yang digunakan jumlah kasus pada tahun 2016 berdasarkan jenis kelamin yaitu Jenis kelamin Laki-Laki (L) dan Jenis kelamin Perempuan (P) dengan kasus DBD pada masing-masing Kelompok Cluster berikut Cluster C1 ialah Cluster pada kasus DBD tingkat Rendah , C2 ialah Cluster pada kasus DBD tingkat Sedang, C3 ialah Cluster pada kasus DBD tingkat Tinggi.

Titik koordinat Centroid pada Tabel 1 berikut keterangan sebelum menentukan Iterasi data clustering,

Iterasi ke - 1





Pusat awal cluster :

Cluster I  $S_i, t_i$  : Pada tabel tersebut. Dengan nilai Maksimal (MAX) yaitu dengan kasus :

Laki - laki (L) : 2054 Jiwa

Perempuan (P) : 1841 Jiwa

Cluster II  $S_i, t_i$  : Nilai Tengah Pada tabel tersebut. Dengan Rumus Nilai Rata-Rata (AVARAGE) yaitu dengan kasus :

Laki - laki (L) : 718 Jiwa

Perempuan (P) : 688 Jiwa

Cluster C III  $S_i, t_i$  : Pada tabel tersebut. Dengan nilai Minimal (MIN) yaitu dengan kasus :

Laki-Laki (L) : 73 Jiwa

Perempuan (P) : 70 Jiwa

A . Berikut perhitungan Euclidean Distance Iterasi ke-1 pada Cluster I :

$$\begin{aligned} \text{Data ke-1 pada cluster I: } & \sqrt{(1725 - 2.054)^2 + (1699 - 1.841)^2} \\ & = 358,336 \end{aligned}$$

$$\begin{aligned} \text{Data ke-2 pada cluster I: } & \sqrt{(419 - 2.054)^2 + (374 - 1.841)^2} \\ & = 2196,66 \end{aligned}$$

$$\begin{aligned} \text{Data ke-3 pada cluster I: } & \sqrt{(434 - 2.054)^2 + (365 - 1.841)^2} \\ & = 2191,57 \end{aligned}$$

.....LANJUTKAN PERHITUNGAN.....

$$\begin{aligned} \text{Data ke-21 pada cluster I: } & \sqrt{(2.054 - 2.054)^2 + (1827 - 1.841)^2} \\ & = 14 \end{aligned}$$

.....LANJUTKAN PERHITUNGAN.....

$$\begin{aligned} \text{Data ke-26 pada cluster I: } & \sqrt{(381 - 2.054)^2 + (373 - 1.841)^2} \\ & = 2225,75 \end{aligned}$$

$$\begin{aligned} \text{Data ke-27 pada cluster I: } & \sqrt{(175 - 2.054)^2 + (193 - 1.841)^2} \\ & = 2499,31 \end{aligned}$$

B . Berikut perhitungan Euclidean Distance Iterasi ke-1 pada Cluster II :

$$\text{Data ke-1 pada cluster II : } \sqrt{(1725 - 718)^2 + (1699 - 668)^2}$$





$$= 1441,29$$

$$\text{Data ke-2 pada cluster II : } \sqrt{(419 - 718)^2 + (374 - 668)^2}$$

$$= 419,223$$

$$\text{Data ke-3 pada cluster II : } \sqrt{(434 - 718)^2 + (365 - 668)^2}$$

$$= 415,189$$

.....LANJUTKAN PERHITUNGAN.....

$$\text{Data ke-19 pada cluster II : } \sqrt{(658 - 718)^2 + (566 - 688)^2}$$

$$= 118,277$$

.....LANJUTKAN PERHITUNGAN.....

$$\text{Data ke-26 pada cluster II : } \sqrt{(381 - 718)^2 + (373 - 668)^2}$$

$$= 447,762$$

$$\text{Data ke-27 pada cluster II : } \sqrt{(175 - 718)^2 + (193 - 668)^2}$$

$$= 721,324$$

C . Berikut perhitungan Euclidean Distance Iterasi ke-1 pada Cluster III :

$$\text{Data ke-1 pada cluster III : } \sqrt{(1725 - 73)^2 + (1699 - 70)^2}$$

$$= 2320,07$$

$$\text{Data ke-2 pada cluster III : } \sqrt{(419 - 73)^2 + (374 - 70)^2}$$

$$= 460,578$$

$$\text{Data ke-3 pada cluster III : } \sqrt{(434 - 73)^2 + (365 - 70)^2}$$

$$= 466,204$$

.....LANJUTKAN PERHITUNGAN.....

$$\text{Data ke-18 pada cluster III : } \sqrt{(73 - 73)^2 + (70 - 70)^2}$$







$$= 0$$

.....LANJUTKAN PERHITUNGAN.....

$$\text{Data ke-26 pada cluster III : } \sqrt{(381 - 73)^2 + (373 - 70)^2}$$

$$= 432,057$$

$$\text{Data ke-27 pada cluster III : } \sqrt{(175 - 73)^2 + (193 - 70)^2}$$

$$= 159,79$$

Berikut tabel hasil iterasi ke-1 jarak ke centroid dari perhitungan persamaan Euclidean Distance. Yang merupakan jarak antar garis lurus biasa antara dua titik ke dalam ruang Euclidean tersebut,

Tabel 2 Hasil Iterasi ke - 1 Jarak ke Centroid





5	Pusat awal cluster							
6	C1	2.054	1.841					
7	C2	718	668					
8	C3	73	70					
9								
10	KABUPATEN/KOTA	L	P	C1	C2	C3	Jarak terdekat	Cluster
11	KABUPATEN BOGOR	1725	1699	358,336	1441,29	2320,07	358,3364341	1
12	KABUPATEN SUKABUMI	419	374	2196,66	419,223	460,578	419,2231777	2
13	KABUPATEN CIANJUR	434	365	2191,57	415,189	466,204	415,1894687	2
14	KABUPATEN BANDUNG	1664	1708	412,055	1405,98	2283,49	412,054608	1
15	KABUPATEN GARUT	291	225	2391,57	615,185	267,486	267,4864483	3
16	KABUPATEN TASIKMALAYA	229	250	2421,14	643,191	238,193	238,1931989	3
17	KABUPATEN CIAMIS	387	361	2229,19	451,343	428,109	428,1086311	3
18	KABUPATEN KUNINGAN	933	781	1542,8	243,034	1115,85	243,0335263	2
19	KABUPATEN CIREBON	940	937	1434,65	348,866	1226,12	348,8657793	2
20	KABUPATEN MAJALENGKA	176	152	2525,79	748,236	131,655	131,6548518	3
21	KABUPATEN SUMEDANG	475	433	2115,59	337,937	541,639	337,9370401	2
22	KABUPATEN INDRAMAYU	483	477	2080,51	302,71	577,71	302,7096741	2
23	KABUPATEN SUBANG	428	271	2260,26	491,56	407,953	407,9534287	3
24	KABUPATEN PURWAKARTA	424	361	2201,66	424,969	455,941	424,9692783	2
25	KABUPATEN KARAWANG	546	513	2009,39	231,423	648,057	231,4234505	2
26	KABUPATEN BEKASI	910	703	1613,62	195,34	1049,41	195,3395868	2
27	KABUPATEN BANDUNG BARAT	815	739	1658,17	120,336	999,062	120,3357899	2
28	KABUPATEN PANGANDARAN	73	70	2657,22	879,451	0	0	3
29	KOTA BOGOR	658	566	1890,62	118,277	766,969	118,276663	2
30	KOTA SUKABUMI	431	511	2098,34	326,991	568,018	326,9913898	2
31	KOTA BANDUNG	2054	1827	14	1768,78	2647,91	14	1
32	KOTA CIREBON	590	521	1971,22	194,824	686,069	194,8243238	2
33	KOTA BEKASI	1972	1841	82	1717,21	2596,66	82	1
34	KOTA DEPOK	1285	1270	957,811	827,078	1705,56	827,0782569	2
35	KOTA CIMAHI	483	516	2055,16	279,738	605,818	279,7378378	2
36	KOTA TASIKMALAYA	381	373	2225,75	447,762	432,057	432,0567092	3
37	KOTA BANJAR	175	193	2499,31	721,324	159,79	159,7904878	3

Setelah menghitung persamaan Euclidean Distance, yakni jarak antar garis lurus biasa antara dua titik dalam ruang Euclidean dan menentukan hasil iterasi jarak terdekat ke centroid dengan nilai terkecil (MIN) pada setiap nilai centroid C1,C2,C3 pada masing-masing data kabupaten/kota.

Selanjutnya nilai terkecil yang telah di tabulasikan pada Tabel 2 Hasil Iterasi ke-1 Jarak ke Centroid tersebut. Berdasarkan nilai terkecil (minimum) pada setiap data jarak ke centroid yang telah ditentukan

C1 Hasil Iterasi pada nilai tertinggi (maximum) berwarna merah

C2 Hasil Iterasi pada nilai sedang berwarna kuning

C3 Hasil Iterasi pada nilai terkecil (minimum) berwarna hijau

Untuk menentukan iterasi ke-2 dan selanjutnya, menentukan centroid ke-2 dan selanjutnya.

Rumus menentukan centroid selanjutnya :

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_j \epsilon S_j x_j \dots \dots 2 \quad / \dots \dots 3 \quad / \dots \dots \text{-iterasi akhir.}$$





Keterangan :

$\mu_j (t + 1)$  : Centroid baru pada iterasi ke-1 /ke-2..... /ke-3.... .....-iterasi akhir.

$N_{sj}$  : Banyak data pada cluster sj;

Iterasi ke - 2

Pusat awal cluster pada iterasi ke-1

Cluster I  $S_{i,t_i}$  : Pada tabel tersebut. Dengan nilai Maksimal (MAX) yaitu dengan kasus berdasarkan jenis kelamin:

Laki-Laki (L) : 1.854 Jiwa  $\frac{\text{penjumlahan data ter MAX pada setiap Centroid Cluster 1}}{4}$

Perempuan (P) : 1.769 Jiwa

Cluster II  $S_{i,t_i}$  : Nilai Tengah Pada tabel tersebut. Dengan Rumus Nilai Rata-Rata (AVARAGE) yaitu dengan kasus berdasarkan jenis kelamin:

Laki-Laki (L) : 655 Jiwa  $\frac{\text{penjumlahan data rata-rata pada setiap Centroid Cluster 2}}{15}$

Perempuan (P) : 604 Jiwa

Cluster C III  $S_{i,t_i}$  : Pada tabel tersebut. Dengan nilai Minimal (MIN) yaitu dengan kasus berdasarkan jenis kelamin:

Laki-Laki (L) : 268 Jiwa  $\frac{\text{penjumlahan data ter MIN pada setiap Centroid Cluster 3}}{8}$

Perempuan (P) : 237 Jiwa

A . Berikut perhitungan Euclidean Distance Iterasi ke-2 pada Cluster I

$$\text{Data ke-1 pada cluster I : } \sqrt{(1725 - 1.854)^2 + (1699 - 1.769)^2} = 146,43$$

.....LANJUTKAN PERHITUNGAN.....

$$\text{Data ke-27 pada cluster I : } \sqrt{(175 - 1.854)^2 + (193 - 1.769)^2} = 2302,43$$

B . Berikut perhitungan Euclidean Distance Iterasi ke-1 pada Cluster II :

$$\text{Data ke-1 pada cluster II : } \sqrt{(1725 - 665)^2 + (1699 - 604)^2} = 1530,61$$

.....LANJUTKAN PERHITUNGAN.....

$$\text{Data ke-27 pada cluster II : } \sqrt{(175 - 665)^2 + (193 - 604)^2} = 632,2727$$

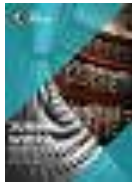
C . Berikut perhitungan Euclidean Distance Iterasi ke-1 pada Cluster III :

$$\text{Data ke-1 pada cluster III : } \sqrt{(1725 - 268)^2 + (1699 - 237)^2} = 2064,489$$

.....LANJUTKAN PERHITUNGAN.....

$$\text{Data ke-27 pada cluster III : } \sqrt{(175 - 268)^2 + (193 - 237)^2} = 102,3781$$





Berikut tabel hasil iterasi ke-2 jarak ke centroid dari perhitungan persamaan Euclidean Distance. Yang merupakan jarak antar garis lurus biasa antara dua titik ke dalam ruang Euclidean tersebut,

Tabel 3 Hasil Iterasi ke - 2 Jarak ke Centroid

ITERASI KE - 1								
		L	P	C1	C2	C3	Jarak terdekat	Cluster
39								
40	C1	1.854	1.769					
41	C2	655	604					
42	C3	268	237					
43								
44								
45	KABUPATEN/KOTA	L	P	C1	C2	C3	Jarak terdekat	Cluster
46	KABUPATEN BOGOR	1725	1699	146,43	1530,61	2064,49	146,4295906	1
47	KABUPATEN SUKABUMI	419	374	2000,96	329,913	204,342	204,3416639	3
48	KABUPATEN CIANJUR	434	365	1996,55	325,906	210,091	210,0910889	3
49	KABUPATEN BANDUNG	1664	1708	199,238	1495,24	2028,4	199,2376094	1
50	KABUPATEN GARUT	291	225	2196,67	525,87	26,3299	26,32993781	3
51	KABUPATEN TASIKMALAYA	229	250	2224,05	554,238	40,6757	40,67573755	3
52	KABUPATEN CIAMIS	387	361	2033,01	362,127	172,3	172,2999293	3
53	KABUPATEN KUNINGAN	933	781	1350,34	329,258	859,629	329,2581898	2
54	KABUPATEN CIREBON	940	937	1235,62	437,91	970,789	437,9102902	2
55	KABUPATEN MAJALENGKA	176	152	2329,96	658,962	124,804	124,8039087	3
56	KABUPATEN SUMEDANG	475	433	1919,68	248,646	285,519	248,6459777	2
57	KABUPATEN INDRAMAYU	483	477	1883,5	214,137	322,646	214,1370797	2
58	KABUPATEN SUBANG	428	271	2067,85	403,434	164,088	164,0876766	3
59	KABUPATEN PURWAKARTA	424	361	2006,48	335,66	199,748	199,7480053	3
60	KABUPATEN KARAWANG	546	513	1813,04	142,344	392,183	142,3435594	2
61	KABUPATEN BEKASI	910	703	1423,55	273,313	793,775	273,3126821	2
62	KABUPATEN BANDUNG BARAT	815	739	1462,66	208,993	742,89	208,992557	2
63	KABUPATEN PANGANDARAN	73	70	2461,06	790,225	256,276	256,2762487	3
64	KOTA BOGOR	658	566	1696	38,5783	510,699	38,57834741	2
65	KOTA SUKABUMI	431	511	1898,99	242,78	319,181	242,7795067	2
66	KOTA BANDUNG	2054	1827	208,55	1857,85	2391,67	208,5500539	1
67	KOTA CIREBON	590	521	1775,93	105,832	429,806	105,8317323	2
68	KOTA BEKASI	1972	1841	138,575	1806,47	2340,63	138,5753405	1
69	KOTA DEPOK	1285	1270	756,458	916,379	1450,05	756,4576161	1
70	KOTA CIMAH	483	516	1856,97	193,477	352,634	193,4768433	2
71	KOTA TASIKMALAYA	381	373	2029,07	358,733	177,235	177,2350576	3
72	KOTA BANJAR	175	193	2302,43	632,273	102,378	102,3780525	3
73								

Setelah menghitung persamaan Euclidean Distance, yakni jarak antar garis lurus biasa antara dua titik dalam ruang Euclidean dan menentukan hasil iterasi jarak terdekat ke centroid dengan nilai terkecil (MIN) pada setiap nilai centroid C1,C2,C3 pada masing-masing data kabupaten/kota.

Selanjutnya nilai jarak terkecil dengan pusat cluster yang telah di tabulasikan pada Tabel 3 dipilih secara random pada Hasil Iterasi ke-2 Jarak ke Centroid tersebut. Berdasarkan nilai terkecil (minimum) pada setiap data jarak ke centroid yang telah ditentukan

C1 Hasil Iterasi pada nilai tertinggi (maximum) berwarna merah

C2 Hasil Iterasi pada nilai sedang berwarna kuning





C3 Hasil Iterasi pada nilai terkecil (minimum) berwarna hijau

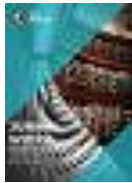
Untuk selanjutnya iterasi ke-3 sampai iterasi ke- sekian ulangi seperti pada langkah iterasi ke-2 dalam perhitungannya, sampai menemukan nilai yang tetap atau hasil yang sama. Maka dari itu pencarian hasil Iterasi dihentikan dikarenakan Hasil telah ditemukan sampai Tabel Hasil Iterasi ke - 7 Jarak ke Centroid.

Tabel 4 Hasil Iterasi ke-7 Jarak ke Centroid

ITERASI KE - 7								
C1	1.740	1.669						
C2	808	708						
C3	365	340						
KABUPATEN/KOTA	L	P	C1	C2	C3	Jarak terdekat	Cluster	
KABUPATEN BOGOR	1725	1699	33,54102	1350,523	1922,624	33,54101966		1
KABUPATEN SUKABUMI	419	374	1849,883	512,3539	63,91162	63,91162111		3
KABUPATEN CIANJUR	434	365	1845,549	507,1109	73,57782	73,57781807		3
KABUPATEN BANDUNG	1664	1708	85,42248	1316,678	1886,473	85,42247948		1
KABUPATEN GARUT	291	225	2045,663	707,158	136,846	136,8459729		3
KABUPATEN TASIKMALAYA	229	250	2072,844	737,8797	162,9952	162,995231		3
KABUPATEN CIAMIS	387	361	1881,88	545,2099	30,4273	30,42729552		3
KABUPATEN KUNINGAN	933	781	1199,914	145,1269	719,1555	145,1268602		2
KABUPATEN CIREBON	940	937	1084,354	264,6308	828,867	264,6308225		2
KABUPATEN MAJALENGKA	176	152	2178,849	841,3997	266,5794	266,5793884		3
KABUPATEN SUMEDANG	475	433	1768,593	431,5095	144,0827	144,0826857		3
KABUPATEN INDRAMAYU	483	477	1732,314	398,3622	180,7798	180,7797564		3
KABUPATEN SUBANG	428	271	1917,224	578,766	93,87596	93,87595705		3
KABUPATEN PURWAKARTA	424	361	1855,457	517,1977	62,81676	62,81675981		3
KABUPATEN KARAWANG	546	513	1661,918	326,2353	250,3901	250,3900863		3
KABUPATEN BEKASI	910	703	1273,6	102,4474	654,9106	102,447412		2
KABUPATEN BANDUNG BARAT	815	739	1311,688	32,01779	601,4425	32,01779019		2
KABUPATEN PANGANDARAN	73	70	2309,911	972,9165	397,6813	397,6813364		3
KOTA BOGOR	658	566	1545,1	206,196	370,0906	206,1960367		2
KOTA SUKABUMI	431	511	1747,697	424,9955	183,1163	183,1162754		3
KOTA BANDUNG	2054	1827	351,511	1675,076	2250,337	351,511024		1
KOTA CIREBON	590	521	1624,932	286,8544	288,8143	286,8544443		2
KOTA BEKASI	1972	1841	288,8044	1624,727	2198,981	288,8044321		1
KOTA DEPOK	1285	1270	605,1661	737,4812	1308,164	605,1660929		1
KOTA CIMAH	483	516	1705,713	377,1054	211,8111	211,8111076		3
KOTA TASIKMALAYA	381	373	1877,897	542,3632	36,53177	36,53177128		3
KOTA BANJAR	175	193	2151,232	815,6718	240,1714	240,1714394		3

Setelah menghitung persamaan Euclidean Distance sampai Hasil Iterasi ke- 7 Jarak ke Centroid. nilai centroid C1,C2,C3 pada masing-masing data kabupaten/kota.





Berdasarkan nilai jarak terdekat centroid pada setiap data jarak ke centroid yang telah ditentukan keterangan Hasil Akhir pada Iterasi ke – 7 Cluster :

C1 Hasil Iterasi pada nilai tertinggi (maximum) berwarna merah yaitu Cluster 1

C2 Hasil Iterasi pada nilai sedang berwarna kuning yaitu Cluster 2

C3 Hasil Iterasi pada nilai terkecil (minimum) berwarna hijau yaitu Cluster 3

## 4.1 Hasil Akhir Pengujian

Berikutnya setelah menentukan Hasil Iterasi ke- 7 Jarak ke Centroid berdasarkan nilai terkecil (minimum) yang telah ditentukan menggunakan rumus dan hasilnya pada tabel 4 tersebut, maka dengan tersebut bisa memperoleh hasil pengelompokan data cluster anggota C1 , C2 , C3 dan jumlah anggota, berikut Hasil Pengelompokan data cluster :

CLUSTER 1	5	KABUPATEN BOGOR KABUPATEN BANDUNG KOTA BANDUNG KOTA BEKASI KOTA DEPOK
CLUSTER 2	6	KABUPATEN KUNINGAN KABUPATEN CIREBON KABUPATEN BEKASI KABUPATEN BANDUNG BARAT KOTA BOGOR KOTA CIREBON
CLUSTER 3	16	KABUPATEN SUKABUMI KABUPATEN CIANJUR KABUPATEN GARUT KABUPATEN TASIKMALAYA KABUPATEN CIAMIS KABUPATEN MAJALENGKA KABUPATEN SUMEDANG KABUPATEN INDRAMAYU KABUPATEN SUBANG KABUPATEN PURWAKARTA KABUPATEN KARAWANG





KABUPATEN PANGANDARAN  
KOTA SUKABUMI  
KOTA CIMAHI  
KOTA TASIKMALAYA  
KOTA BANJAR

Berikutnya setelah menentukan Hasil Iterasi Jarak ke Centroid dengan nilai terkecil ( MIN) yang telah menentukan rumus dan hasilnya pada tabel 4 tersebut, dengan tersebut bisa memperoleh hasil pengelompokan data cluster anggota C1 , C2 , C3 dan jumlah anggota, berikut Hasil Pengelompokan data cluster :

Tabel 4 Hasil Pengelompokan Cluster

Kelompok ( Cluster )	Anggota kelompok ( Cluster )	Jumlah Anggota Cluster
C1	{ 1, 4, 21, 23, 24 }	5
C2	{ 8, 9, 16, 17, 19, 22 }	6
C3	{ 2, 3, 5, 6, 7, 10, 11, 12, 13, 14, 15, 18, 22, 25, 26, 27 }	16

Hasil pengelompokan data cluster pada tabel 4 diatas telah menemukan hasil anggota cluster pada setiap kelompok cluster mempunyai anggota cluster , dimana hasil dari pengelompokan tersebut pada kelompok cluster C1 adalah kelompok daerah penyebaran penyakit DBD tingkat Tinggi dengan berjumlah 5 Daerah Kabupaten/Kota yaitu :

1. Kota Bandung dengan data kasus DBD berdasarkan Jenis Kelamin Laki - laki (L) : 2054 Jiwa dan Perempuan (P) : 1827 Jiwa dengan jumlah kasus 3881 Jiwa.
2. Kota Bekasi dengan data kasus DBD berdasarkan Jenis Kelamin Laki - laki (L) : 1972 Jiwa dan Perempuan (P) : 1841 Jiwa dengan jumlah kasus 3813 Jiwa.
3. Kabupaten Bogor dengan data kasus DBD berdasarkan Jenis Kelamin Laki - laki (L) : 1725 Jiwa dan Perempuan (P) : 1699 Jiwa dengan jumlah kasus 3424 Jiwa.
4. Kabupaten Bandung dengan data kasus DBD berdasarkan Jenis Kelamin Laki - laki (L) : 1664 Jiwa dan Perempuan (P) : 1708 Jiwa dengan jumlah kasus 3372 Jiwa.
5. Kota Depok dengan data kasus DBD berdasarkan Jenis Kelamin Laki - laki (L) : 1285 Jiwa dan Perempuan (P) : 1270 Jiwa dengan jumlah kasus 2555 Jiwa.

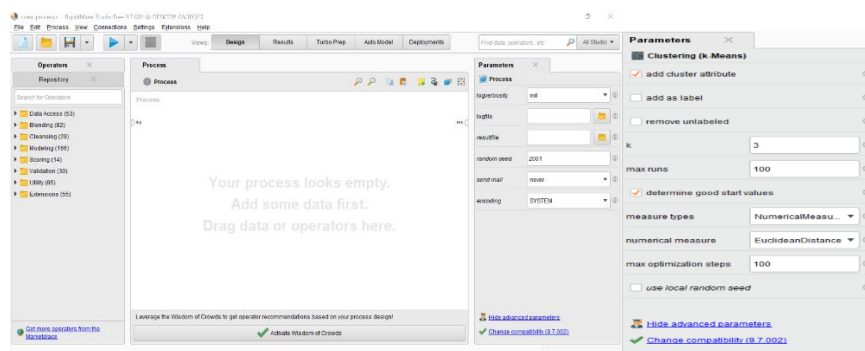
## 4.2 Clustering K-Means dengan Aplikasi RapidMiner Studio



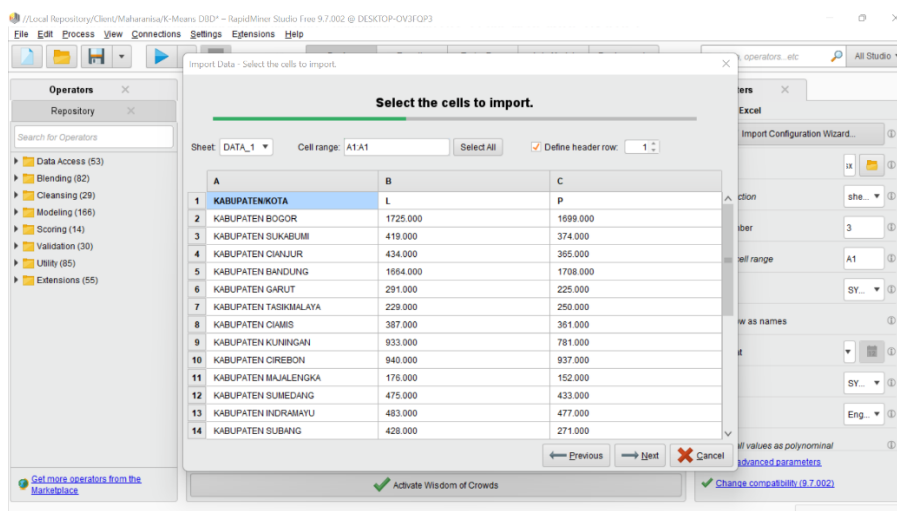


Sebelum menentukan hasil clustering K-Means dengan aplikasi RapidMiner Studio kita siapkan terlebih dahulu data yang mau di input / impor data dengan menggunakan data ms excel seperti Tabel 2 Hasil Iterasi ke - 1 Jarak ke Centroid sampai Hasil Iterasi ke - 7 Jarak ke Centroid yang dimana sudah menemukan Hasil Iterasi dalam tabel Hasil Pengelompokkan data cluster, data tersebut disave format .xls atau ms excel untuk membantu data yang akan dibuat dahulu. Setelah itu jalankan aplikasi RapidMiner Studio yang sudah diinstal.

Untuk menentukan clustering k-means hubungkan terlebih dahulu operator ke Res klik pada import configuration wizard di panel parameter pilih file data excel yang telah dibuat dan cari operator select attribute setelah finish seret ke panel proses, setelah berhasil menemukan masing-masing operators sambungkan dengan evaluation exa dan semua sedang disambungkan klik tombol bottoms PLAY berwarna biru di atas menu. Berikut gambar tahapannya :



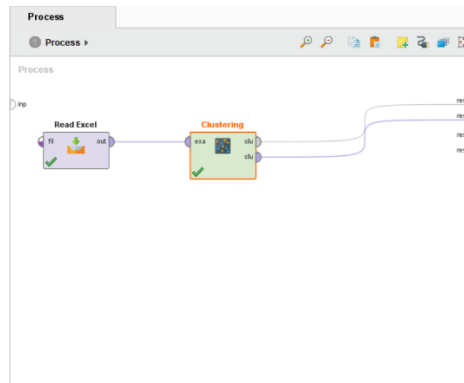
Gambar 5 Halaman pertama membuat clustering K-Means dan input Parameters



Gambar 6 Input data dengan import data .xls / ms excel yang ingin di clustering

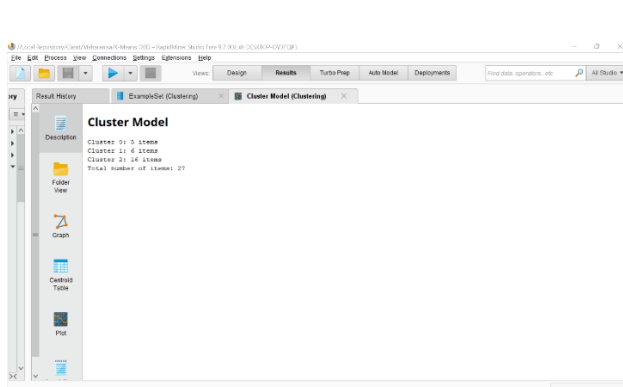






Gambar 7 Proses Modeling

Setelah berhasil menemukan masing-masing operators sambungkan dengan evaluation exa dan semua sedang disambungkan klik tombol buttons PLAY berwarna biru di atas menu dan muncul tampilan berhasil Proses Modeling, berikut hasil clustering k-means pada data yang setelah berhasil di input dan proses modeling :

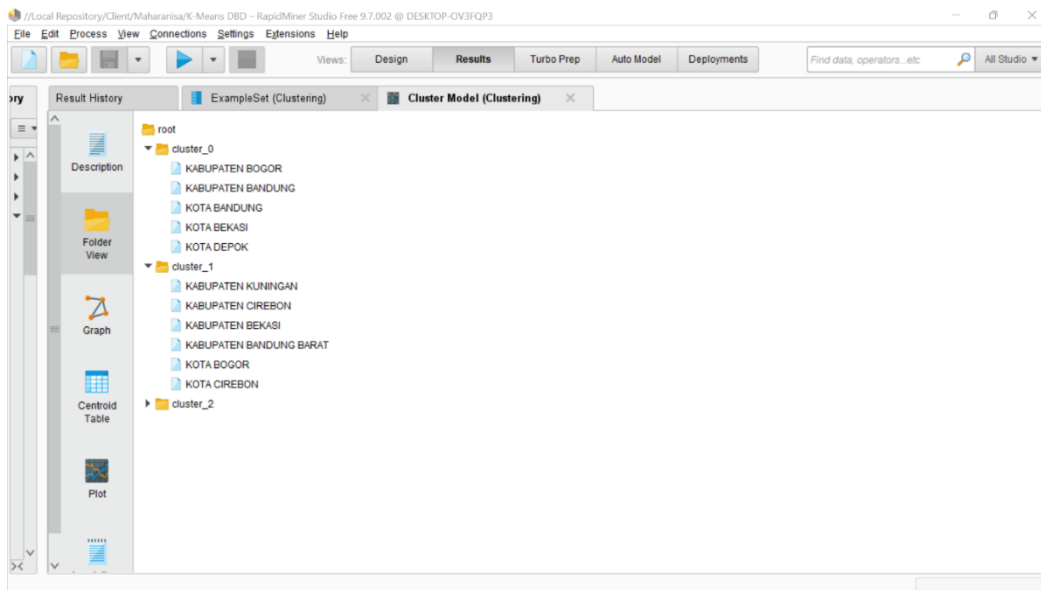


## Cluster Model

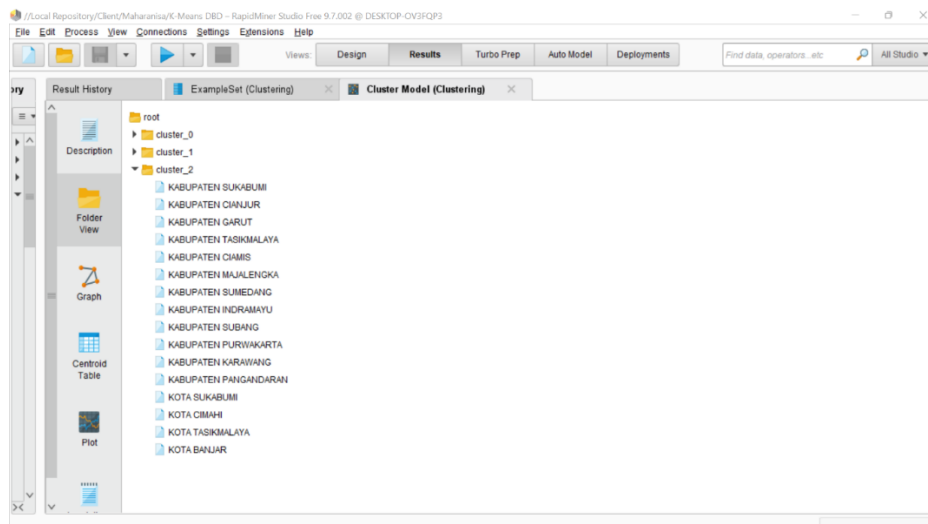
```
Cluster 0: 18 items  
Cluster 1: 5 items  
Cluster 2: 4 items  
Total number of items: 27
```

Gambar 8 Model Cluster (Clustering)



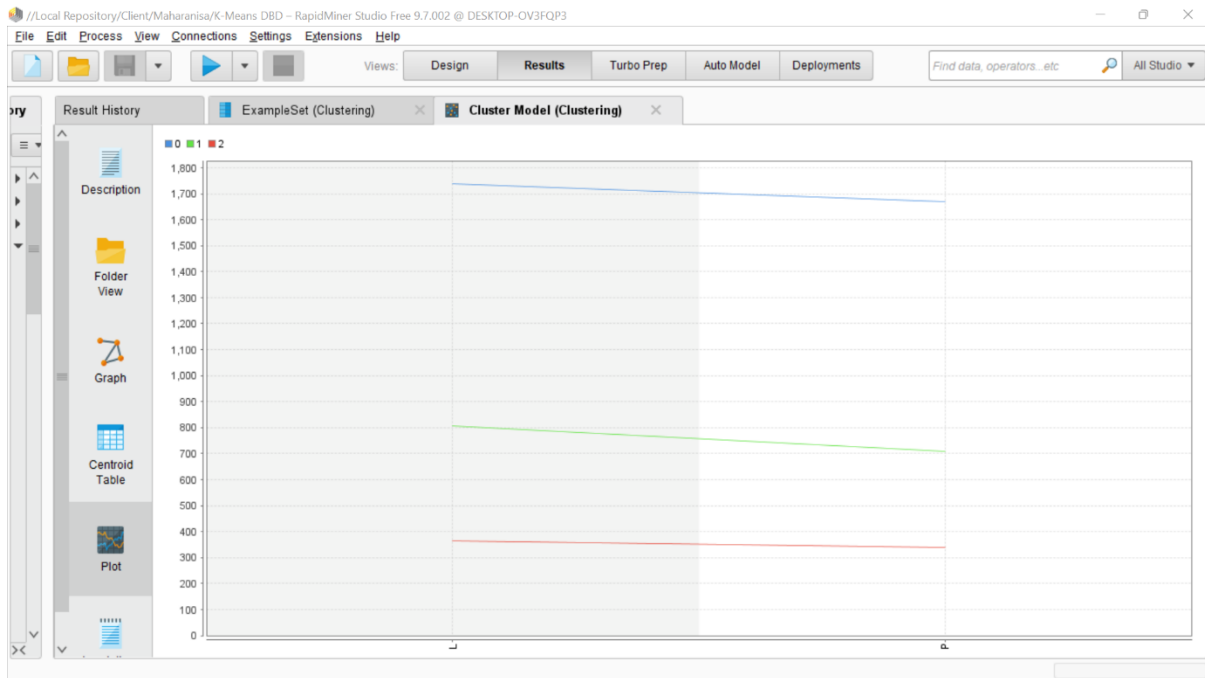


Gambar 9 Hasil pengelompokkan Cluster Cluster 1: Cluster Tinggi, dan Cluster 2: Clustering

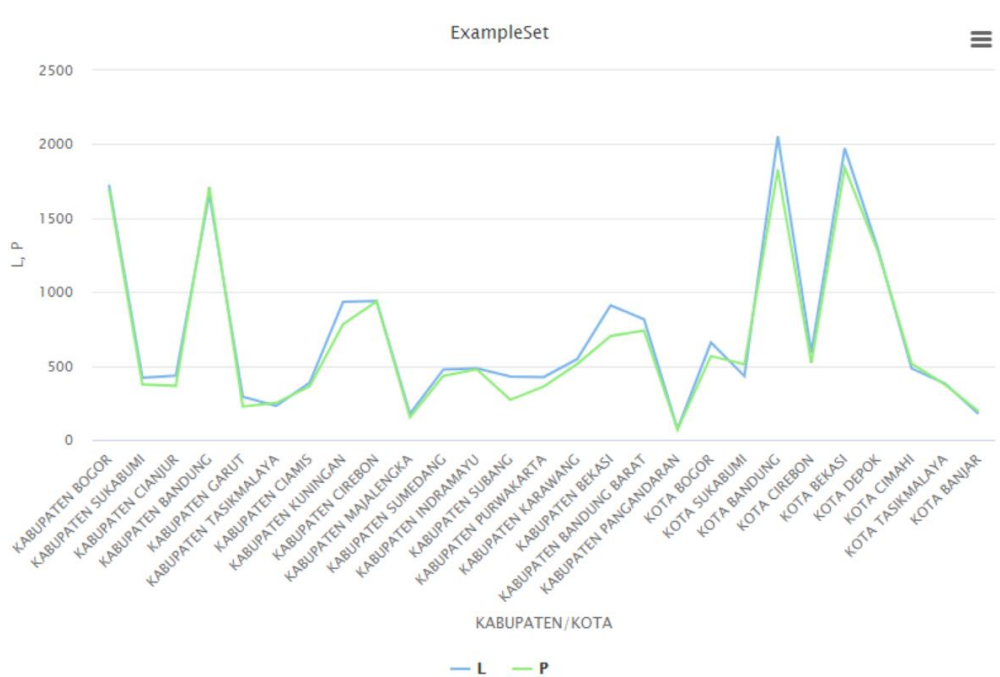


Gambar 10 Hasil pengelompokkan Cluster Cluster 3: Cluster Rendah





Gambar 11 Gambar Plot Cluster



Gambar 12 Visualisasi Data





## 5. Kesimpulan (or Conclusion)

Kesimpulan yang didapat penulis terhadap penelitian ini yaitu terkait bagaimana metode ini dilakukan ke dalam penelitian ini dengan pencarian data set dan juga isi dari penelitian ini yaitu pembagian kelompok daerah penyebaran penyakit DBD tingkat Tinggi atau cluster 1 dengan berjumlah 5 Daerah Kabupaten/Kota yaitu : Kota Bandung dengan jumlah kasus 3881 jiwa, Kota Bekasi dengan kasus 3813 jiwa, Kabupaten Bogor dengan kasus 3424 jiwa, Kabupaten Bandung 3372 jiwa, dan Kota Depok 2555 jiwa. Penulis belum sepenuhnya efektif dan efisien melakukan penulisan ini.

Dengan melakukan pembagian kategori berdasarkan jenis kelamin Laki - laki (L), Perempuan (P), kemudian dikategorikan menjadi Daerah Penyebaran Penyakit DBD Tingkat Tinggi, Sedang, dan Rendah dengan nilai terkecil (minimum) yang di dapat pada iterasi setiap data cluster. Solusi untuk daerah penyebaran penyakit DBD tingkat Tinggi atau cluster 1 perlu dipantau dan penanganan oleh pihak yang berwenang sehingga pencegahan bisa diatasi, sedangkan untuk daerah yang berada di tingkat sedang cluster 2 dan rendah cluster 3 untuk melakukan penanganan dan wawasan dalam pencegahan DBD dengan langkah-langkah preventif.

## 6. Referensi (Reference)

- [1] Harry Gunawan, Vega Purwayoga “ Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Potensi Penyebaran Virus Corona di Kota Cirebon “ Jurnal SISFOKOM (Sistem Informasi dan Komputer), Volume 11, Nomor 1, PP 1-8.
- [2] Haditsah Annur “ Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering “ JURNAL INFORMATIKA UPGRIS Vol. 5, No. 1, (2019) P/E-ISSN: 2460-4801/2447-6645.
- [3] Linda Maulida “ PENERAPAN DATAMINING DALAM MENGELOMPOKKAN KUNJUNGAN WISATAWAN KE OBJEK WISATA UNGGULAN DI PROV. DKI JAKARTA DENGAN K-MEANS “ ISSN 2527-5836 JISKa (Jurnal Informatika Sunan Kalijaga).
- [4] Sri Marliska Hutabarat, Anita Sindar “ Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means “ Jurnal Nasional Komputasi dan Teknologi Informasi Vol. 2 No. 2, Oktober 2019 P-ISSN 2620-8342/E-ISSN 2621-3052.
- [5] Nugroho Arif Sudiby, Ardymulya Iswardani, Kartika Sari, Siti Suprihatiningsih “ PENERAPAN DATA MINING PADA JUMLAH PENDUDUK MISKIN DI Indonesia “ Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika Nugroho Arif Sudiby, Ardymulya Iswardani, Kartika Sari, Siti Suprihatiningsih Volume 1, No. 3, Desember 2020 hal.199-207 DOI Artikel : 10.46306/lb.v1i3.42.
- [6] Sabrina Aulia Rahmah “ KLASTERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI TOKO





- JUANDA TANI KECAMATAN HUTABAYU RAJA) “ Djtechno : Journal of Information Technology Research Vol.1, No.1 Tahun 2020.
- [7] Mhd Gading Sadewo, Agus Perdana Windarto, Dedy Hartama “ PENERAPAN DATAMINING PADA POPULASI DAGING AYAM RAS PEDAGING DI INDONESIA BERDASARKAN PROVINSI MENGGUNAKAN K-MEANS CLUSTERING “ InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan) e-ISSN : 2540-7600 Vol 2, No 1, September 2017 p-ISSN : 2540-7597.
- [8] Normah, Siti Nurajizah, Arinda Salbinda “ Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten “ Jurnal Teknik Komputer AMIK BSI, Volume 7, No.2, Juli 2021 P-ISSN 2442-2436, E-ISSN: 2550-0120.
- [9] Dina Sunia, Kurniabudi, Pareza Alam Jusia “PENERAPAN DATA MINING UNTUK CLUSTERING DATA PENDUDUK MISKIN MENGGUNAKAN ALGORITMA K-MEANS”.
- [10] Suhartini, Ria Yuliani “ Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma KMeans di Dusun Bagik Endep Sukamulia Timur “ Jurnal Informatika dan Teknologi – Vol.4 No. 1 Januari 2021.
- [11] Written, I. H and Frank,. E.2005 Data Mining : Partical Machine Learning Tools and Techniques Second Edition. Morgan Kauffman : San Francisco.
- [12] WHO. World Healt Organization 2010. Dangué and Dangué Hemmoragic Fever.
- [13] <https://opendata.jabarprov.go.id>

