

## DATA MINING: PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDARAH DENGUE (DBD) BERDASARKAN PROVINSI

Page | 173

Kiki Fatmawati<sup>1</sup>, Agus Perdana Windarto<sup>2</sup>

<sup>1,2</sup> Sekolah Tinggi Ilmu Komputer Tunas Bangsa (STIKOM) Tunas Bangsa  
Jl. Jendral Sudirman Blok A No/1/2/3 Pematangsiantar, Medan, Indonesia, 21127  
<sup>1</sup>kikifatmawati@gmail.com, <sup>2</sup>agus.perdana@amiktunasbangsa.ac.id

**Abstrak**—Penyakit Demam Berdarah Dengue (DBD) adalah penyakit yang disebabkan oleh virus Dengue yang tergolong *Arthropod-Borne Virus*, genus *Flavivirus*, dan famili *Flaviviridae*. DBD ditularkan melalui gigitan nyamuk dari genus *Aedes*, terutama *Aedes aegypti* atau *Aedes albopictus*. Penyakit DBD dapat muncul sepanjang tahun dan dapat menyerang seluruh kelompok umur. Penyakit ini berkaitan dengan kondisi lingkungan dan perilaku masyarakat. Penelitian ini membahas tentang pengelompokan jumlah daerah yang terjangkit demam berdarah dengue (DBD) berdasarkan provinsi. Metode yang digunakan adalah Data mining K-Means Clustering. Dengan menggunakan metode ini data-data yang telah diperoleh dapat dikelompokkan ke dalam beberapa cluster, dimana penerapan proses K-Means Clustering menggunakan tools RapidMiner. Penelitian ini menggunakan sumber data yang terekam di situs departemen kesehatan dengan alamat url <https://www.depkes.go.id/>. Data yang digunakan adalah (2014-2016) yang terdiri dari 34 provinsi. Kriteria yang digunakan, yakni: 1) jumlah kabupaten/kota dan 2) kabupaten/kota yang terjangkit. Data diolah dengan menggunakan K-means yang dibagi dalam 3 cluster yaitu: tingkat cluster tinggi (C1), tingkat cluster sedang (C2) dan tingkat cluster rendah (C3). Proses iterasi berlangsung 6 kali sehingga diperoleh penilaian dalam mengelompokkan daerah yang terjangkit demam berdarah dengue (DBD) berdasarkan provinsi. Hasil yang diperoleh bahwa terdapat 4 provinsi dengan cluster tingkatan tinggi (C1), 13 provinsi dengan cluster tingkatan sedang (C2), dan 17 provinsi dengan cluster tingkatan rendah (C3). Hal ini dapat menjadi masukan kepada masyarakat untuk menjaga kesehatan dengan meningkatkan kewaspadaan terhadap penularan demam berdarah, sehingga diperlukan kepedulian peran serta aktif masyarakat untuk bergotong-royong melakukan langkah-langkah pencegahan penularan penyakit DBD, melalui kegiatan pemberantasan nyamuk dan jentik secara berkala.

**Kata Kunci**—Demam Berdarah Dengue, Data Mining, K-means, Clustering

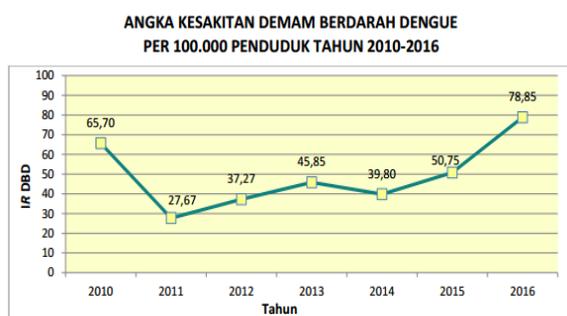
### I. PENDAHULUAN

Penyakit Demam Berdarah Dengue (DBD) adalah penyakit yang disebabkan oleh virus Dengue yang tergolong *Arthropod-Borne Virus*, genus *Flavivirus*, dan famili *Flaviviridae*. DBD ditularkan melalui gigitan nyamuk dari genus *Aedes*, terutama *Aedes aegypti* atau *Aedes albopictus*. Penyakit DBD dapat muncul sepanjang tahun dan dapat menyerang seluruh kelompok umur. Penyakit ini berkaitan dengan kondisi lingkungan dan perilaku masyarakat. Adanya peran lingkungan dalam terjadinya penyakit dan wabah penyakit terjadi karena adanya interaksi antara manusia dengan lingkungan tidak selalu menguntungkan, kadang-kadang manusia bahkan dirugikan seperti terjangkit penyakit demam berdarah. Unsur lingkungan memegang peranan penting dalam menentukan terjadinya proses interaksi antara manusia dan unsur penyebab dalam proses terjadinya penyakit, kondisi lingkungan yang buruk memberi keuntungan virus penyakit cepat berkembang biak. Pembawa dan penyebar penyakit DBD yaitu *Aedes aegypti* menyukai lingkungan yang kualitasnya buruk, yang ditandai dengan permukiman padat penduduk dengan lingkungan yang kurang cahaya matahari, lembab,

gelap, dekat dengan sungai dengan alirannya lambat karena adanya banyak sampah sehingga menimbulkan genangan sebagai tempat perkembangbiakan nyamuk *Aedes aegypti*. Ada dua faktor pembawa penyakit DBD yaitu *Aedes aegypti* dan *Aedes albopictus*, namun habitat kedua nyamuk ini sangatlah berbeda. *Aedes aegypti* yang lebih menyukai hidup di permukiman penduduk dan menyukai darah manusia, berbeda dengan *Aedes albopictus* yang habitatnya di kebun dan menghisap darah hewan. Nyamuk penular dengue ini terdapat hampir di seluruh pelosok Indonesia, kecuali di tempat yang memiliki ketinggian lebih dari 1000 meter di atas permukaan laut. Penyakit DBD banyak dijumpai terutama di daerah tropis dan sering menimbulkan kejadian luar biasa (KLB).

Beberapa tahun terakhir, kasus Demam Berdarah Dengue (DBD) seringkali muncul di musim pancaroba, khususnya bulan Januari di awal tahun. Karena itu, masyarakat perlu mengetahui penyebab penyakit DBD, mengenali tanda dan gejalanya, sehingga mampu mencegah dan menanggulangi dengan baik. Beberapa faktor yang mempengaruhi munculnya DBD antara lain rendahnya status kekebalan kelompok masyarakat dan kepadatan populasi nyamuk penular karena

banyaknya tempat perindukan nyamuk yang biasanya terjadi pada musim penghujan. Pada tahun 2014, sampai pertengahan bulan Desember tercatat penderita DBD di 34 provinsi di Indonesia sebanyak 71.668 orang, dan 641 diantaranya meninggal dunia. Angka tersebut lebih rendah dibandingkan tahun sebelumnya, yakni tahun 2013 dengan jumlah penderita sebanyak 112.511 orang dan jumlah kasus meninggal sebanyak 871 penderita. Tahun 2016 terdapat jumlah kasus DBD sebanyak 204.171 kasus dengan jumlah kematian sebanyak 1.598 orang. Jumlah kasus DBD tahun 2016 meningkat dibandingkan jumlah kasus tahun 2015 (129.650 kasus). Jumlah kematian akibat DBD tahun 2016 juga meningkat dari tahun 2015 (1.071 kematian). *IR* atau angka kesakitan DBD tahun 2016 juga meningkat dari tahun 2015, yaitu 50,75 menjadi 78,85 per 100.000 penduduk. Namun, *Case Fatality Rate (CFR)* mengalami penurunan dari 0,83% pada tahun 2015 menjadi 0,78% pada tahun 2016. Berikut tren angka kesakitan DBD selama kurun waktu 2010-2016.



Sumber: Ditjen P2P, Kemenkes RI, 2017

Gbr 1. Angka Kesakitan Demam Berdarah Dengue Per 100.000 Penduduk Tahun 2010-2016

Berdasarkan Penelitian terdahulu yang dilakukan oleh Mahasiswa Departemen Statistika FSM Universitas Diponegoro, 2017 yang berjudul "Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue (DBD) menggunakan Support Vector machine (SVM) berbasis GUI Matlab. Penelitian selanjutnya dilakukan oleh Aryu Chandra, 2010 yang berjudul "Demam berdarah dengue: Epidemiologi, Patogenesis, dan Faktor Resiko Penularan. Penelitian ini menjadi salah satu acuan penulis dalam melakukan penelitian sehingga penulis dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan. Dari penelitian terdahulu, penulis tidak menemukan penelitian dengan judul yang sama seperti judul penelitian penulis. Namun penulis mengangkat beberapa penelitian sebagai referensi dalam memperkaya bahan kajian pada penelitian penulis.

Berdasarkan penelitian diatas, maka peneliti mengangkat judul pengelompokan jumlah daerah yang terjangkit Demam Berdarah Dengue (DBD) berdasarkan provinsi Dengan K-Means Clustering. Diharapkan nantinya hasil penelitian yang didapatkan adalah daerah provinsi mana yang terkena penyakit

demam berdarah dengue (DBD) paling tinggi. Ini dapat menjadi masukan bagi pemerintah untuk memberikan kebijakan dalam mengembangkan pola hidup sehat kepada masyarakat tentang jumlah daerah yang terjangkit Demam Berdarah Dengue (DBD).

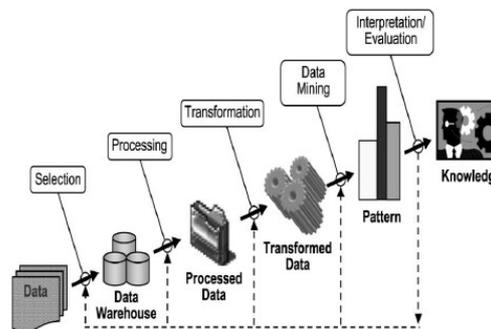
## II. LANDASAN TEORI

### A. Data Mining

Data mining merupakan proses menemukan korelasi baru yang bermanfaat, pola dan trend dengan menambang sejumlah repository data dalam jumlah besar, menggunakan teknologi pengenalan pola seperti statistik dan teknik matematika[1]. Data mining disebut juga dengan knowledge discovery in database (KDD) ataupun pattern recognition [2]. Data Mining dapat dibagi menjadi empat kelompok, yaitu model prediksi (*prediction modelling*), analisis kelompok (*Cluster analysis*), analisis asosiasi (association analysis) dan deteksi anomaly (*anomaly detection*)[3].

### B. Tahapan Data Mining

Tahapan yang dilakukan pada proses data mining diawali dari seleksi data dari data sumber ke data target, tahap pre-processing untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Secara detail dijelaskan sebagai berikut [4]



Gbr 2. Tahapan Data Mining

Berikut ini adalah penjelasan tahapan data mining berdasarkan gambar 2:

#### 1. Data selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

#### 2. Pre-processing /cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

#### 3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk

proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

#### 4. Data mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

#### 5. Interpretation/evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

#### C. Clustering

Clustering atau klusterisasi adalah metode pengelompokan data. Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum [5].

#### D. K-Means

K-means merupakan salah satu teknik pengelompokan yang bekerja berdasarkan partitioned clustering. Prinsip kerja dari pengelompokan *hierarchical clustering* dilakukan secara bertahap. Dan disetiap iterasi dari pengelompokan *hierarchical clustering* hanya ada satu pemilihan penggabungan suatu item terhadap item lainnya [6]. Langkah-langkah melakukan Clustering dengan metode K-Means adalah sebagai berikut [7]:

1. Pilih jumlah *cluster*  $k$ .
2. Inisialisasi  $k$  pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/ objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

dimana:

$D(I,j)$  = Jarak data ke  $i$  ke pusat *cluster*  $j$

$X_{ki}$  = Data ke  $i$  atribut data ke  $k$

$X_{kj}$  = Titik pusat ke  $j$  pada atribut ke  $k$

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.

$$R_k = \frac{1}{N_k}(X_{1k} + X_{2k} + \dots + X_{nk}) \quad (2)$$

dimana :

$R_k$  = Rata-rata baru.

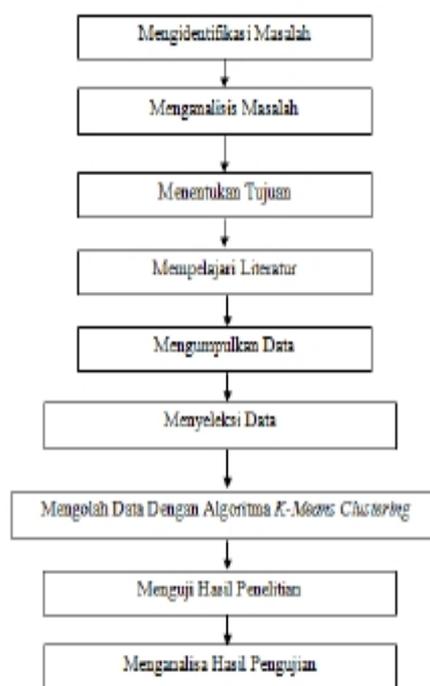
$N_k$  = Jumlah *training pattern* pada *cluster* ( $k$ ).

$X_{nk}$  = Pola ke ( $n$ ) yang menjadi bagian *cluster* ( $k$ ).

5. Tugaskan lagi tiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3

### III. METODE PENELITIAN

Dalam metodologi penelitian terdapat urutan kerangka kerja yang harus diikuti. Kerangka kerja ini merupakan langkah langkah yang dilakukan dalam penelitian. Adapun kerangka kerja yang digunakan dalam penelitian ini adalah seperti terlihat pada gambar dibawah ini.



Gbr 3. Kerangka Kerja Penelitian

IV. HASIL DAN PEMBAHASAN

A. Sumber Data

Dalam penelitian ini, sumber data diambil dari Departemen Kesehatan dimana data tersebut merupakan data yang terjangkit demam berdarah dengue (DBD) berdasarkan provinsi tahun 2014 – 2016. Dalam melakukan *clustering*, data yang diperoleh akan dihitung terlebih dahulu berdasarkan jumlah Kasus Penyakit DBD berdasarkan Provinsi. Hasil rata-rata berdasarkan 2 kriteria penilaian yakni 1). Rata-rata jumlah Kab/Kota 2). Rata-rata jumlah Kab/Kota yang terjangkit seperti yang ditunjukkan pada table berikut:

TABEL I  
JUMLAH KABUPATEN/KOTA YANG TERJANGKIT DEMAM BERDARAH DENGUE BERDASARKAN PROVINSI (2014-2016)

Provinsi	Jumlah Kab/Kota			Kabupaten/Kota Terjangkit			Rata-rata
	2014	2015	2016	2014 Jumlah	2015 Jumlah	2016 Jumlah	
Aceh	23	23	23	21	21	21	21
Sumatera Utara	33	33	33	30	31	30	30,333
Sumatera Barat	19	19	19	18	18	19	18,333
Riau	12	12	12	12	10	12	11,333
Jambi	11	11	11	10	11	11	10,667
Sumatera Selatan	17	17	17	16	17	17	16,667
Bengkulu	10	10	10	10	10	10	10
Lampung	15	15	15	15	15	15	15
Kep. Bangka Belitung	7	7	7	7	7	7	7
Kepulauan Riau	7	7	7	5	7	6	6
DKI Jakarta	6	6	6	5	6	6	5,667
Jawa Barat	27	27	27	27	27	27	27
Jawa Tengah	35	35	35	35	35	35	35
DI Yogyakarta	5	5	5	5	5	5	5
Jawa Timur	38	38	38	38	38	38	38
Banten	8	8	8	8	8	8	8
Bali	9	9	9	9	9	9	9
NTB	10	10	10	10	10	10	10
NTT	22	22	22	6	11	11	9,333
KalBar	14	14	14	14	14	14	14
KalTeng	14	14	14	14	14	14	14
KalSel	13	13	13	13	13	13	13
KalTim	10	10	10	10	10	10	10
KalUt	5	5	5	5	5	5	5
Sulawesi Utara	15	15	15	14	14	15	14,333
Sulawesi Tengah	13	13	13	13	12	13	12,667
Sulawesi Selatan	24	24	24	22	23	24	23
Sulawesi Tenggara	14	17	17	9	12	16	12,333
Gorontalo	6	6	6	6	6	6	6
Sulawesi Barat	6	6	6	6	5	5	5,333
Maluku	11	11	11	4	6	8	6
Maluku Utara	10	10	10	6	6	8	6,667
Papua Barat	13	13	13	3	5	5	4,333
Papua	29	29	29	7	5	10	7,333

Sumber: DepKes <https://www.depkes.go.id/>

Data tersebut kemudian diakumulasikan dan diambil nilai rata-rata dari setiap kriteria seperti yang ditunjukkan pada tabel 2 berikut:

TABEL II  
DATA RATA-RATA JUMLAH KABUPATEN/KOTA YANG TERJANGKIT DEMAM BERDARAH DENGUE BERDASARKAN PROVINSI TAHUN 2014 – 2016

Provinsi	Jumlah Kab/Kota			Rata-rata	Kabupaten/Kota Terjangkit			Rata-rata
	2014	2015	2016		2014 Jumlah	2015 Jumlah	2016 Jumlah	
	Aceh	23	23		23	23	21	
Sumatera Utara	33	33	33	33	30	31	30	30,333
Sumatera Barat	19	19	19	19	18	18	19	18,333
Riau	12	12	12	12	12	10	12	11,333
Jambi	11	11	11	11	10	11	11	10,667
Sumatera Selatan	17	17	17	17	16	17	17	16,667
Bengkulu	10	10	10	10	10	10	10	10
Lampung	15	15	15	15	15	15	15	15
Kep. Bangka Belitung	7	7	7	7	7	7	7	7
Kepulauan Riau	7	7	7	7	5	7	6	6
DKI Jakarta	6	6	6	6	5	6	6	5,667
Jawa Barat	27	27	27	27	27	27	27	27
Jawa Tengah	35	35	35	35	35	35	35	35
DI Yogyakarta	5	5	5	5	5	5	5	5
Jawa Timur	38	38	38	38	38	38	38	38
Banten	8	8	8	8	8	8	8	8
Bali	9	9	9	9	9	9	9	9
NTB	10	10	10	10	10	10	10	10
NTT	22	22	22	22	6	11	11	9,333
KalBar	14	14	14	14	14	14	14	14
KalTeng	14	14	14	14	14	14	14	14
KalSel	13	13	13	13	13	13	13	13
KalTim	10	10	10	10	10	10	10	10
KalUt	5	5	5	5	5	5	5	5
Sulawesi Utara	15	15	15	15	14	14	15	14,333
Sulawesi Tengah	13	13	13	13	13	12	13	12,667
Sulawesi Selatan	24	24	24	24	22	23	24	23
Sulawesi Tenggara	14	17	17	16	9	12	16	12,333
Gorontalo	6	6	6	6	6	6	6	6
Sulawesi Barat	6	6	6	6	6	5	5	5,333
Maluku	11	11	11	11	4	6	8	6
Maluku Utara	10	10	10	10	6	6	8	6,667
Papua Barat	13	13	13	13	3	5	5	4,333
Papua	29	29	29	29	7	5	10	7,333

Setelah diakumulasikan dan dicari nilai rata-rata maka akan didapatkan nilai dari setiap variabel. Data tersebut masuk ke tahapan *clustering* dan diolah dengan *K-means* berdasarkan *cluster* yang sudah ditentukan.

B. Cendroid Data

Dalam penerapan algoritma *K-means* dihasilkan nilai titik tengah atau *centroid* dari data. Proses pencarian nilai titik tengah dilakukan dengan mengambil nilai terbesar (*maksimum*) untuk *cluster* tingkat tinggi (C1),

nilai rata-rata (*average*) untuk *cluster* tingkat sedang (C2) dan nilai terkecil (*minimum*) untuk *cluster* tingkat rendah (C3). Berikut ini nilai centroid awal untuk iterasi 1 dapat diketahui pada tabel berikut:

TABEL III  
CENTROID DATA AWAL (ITERASI 1)

Centroid		
Max (C1)	38	38
Average (C2)	15,09	13,16
Min (C3)	5	4,33

C. Clustering Data

Proses clustering dengan menggunakan nilai centroid awal yang terdapat pada tabel 3, akan memperoleh hasil pengelompokan pada iterasi 1 yang dapat dilihat pada tabel dibawah ini:

TABEL IV  
PENGELOMPOKAN DATA ITERASI 1

Provinsi	C1	C2	C3
Aceh		1	
Sumatera Utara	1		
Sumatera Barat		1	
Riau		1	
Jambi		1	
Sumatera Selatan		1	
Bengkulu		1	
Lampung		1	
Kepulauan Bangka Belitung			1
Kepulauan Riau			1
DKI Jakarta			1
Jawa Barat	1		
Jawa Tengah	1		
DI Yogyakarta			1
Jawa Timur	1		
Banten			1
Bali			1
Nusa Tenggara Barat		1	
Nusa Tenggara Timur		1	
Kalimantan Barat		1	
Kalimantan Tengah		1	
Kalimantan Selatan		1	
Kalimantan Timur		1	
Kalimantan Utara			1
Sulawesi Utara		1	
Sulawesi Tengah		1	
Sulawesi Selatan		1	
Sulawesi Tenggara		1	
Gorontalo			1
Sulawesi Barat			1
Maluku			1
Maluku Utara			1
Papua Barat			1
Papua		1	
<b>Hasil</b>	<b>4</b>	<b>18</b>	<b>12</b>

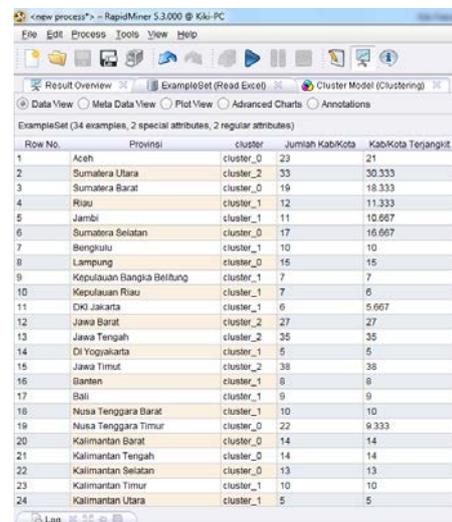
Berdasarkan tabel 4, diperoleh pengelompokan untuk iterasi 1: *Cluster* tingkat tinggi (C1) terdiri dari

4 provinsi yakni Sumatera Utara, Jawa Barat, Jawa Tengah, Jawa Timur, *cluster* tingkat sedang (C2) terdiri dari 18 provinsi yakni Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, NTB, NTT, KalBar, KalTeng, KalSel, KalTim, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Selatan, Papua dan *cluster* tingkat rendah (C3) terdiri dari 12 provinsi yakni Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Yogyakarta, Banten, Bali, KalUt, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Proses iterasi akan terus berlangsung sampai hasil iterasi terakhir sama dengan hasil iterasi sebelumnya. Proses nilai titik tengah atau *centroid* akan menyesuaikan sesuai dengan iterasi yang ada. Dalam hal ini proses akan dilanjutkan dengan menggunakan aplikasi *RapidMiner*.

D. Penerapan aplikasi *RapidMiner*

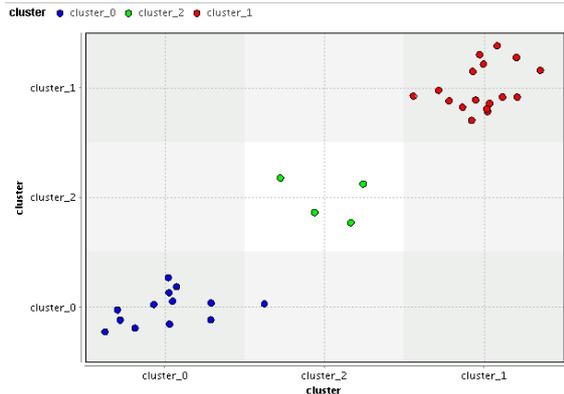
Hasil pengelompokan data daerah terjangkit Demam Berdarah Dengue (DBD) berdasarkan provinsi dengan *Rapid Miner* :

1. *Data view*: merupakan *sheet* di *Result Perspective* untuk menampilkan data yang telah diolah secara keseluruhan lengkap dengan klasternya dari *example set (read excel)*. *Data view* dapat lihat pada gambar.



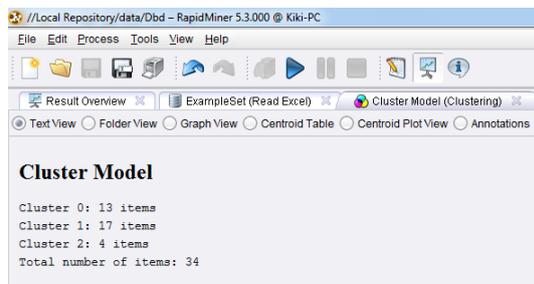
Gbr 4. *Data View*

2. *Flot view* : merupakan *sheet* di *Result Perspective* untuk menampilkan data yang telah diolah secara keseluruhan lengkap dengan klasternya dari *example set (read excel)* dalam bentuk diagram *Scatter*. Tampilannya dapat dilihat pada gambar berikut



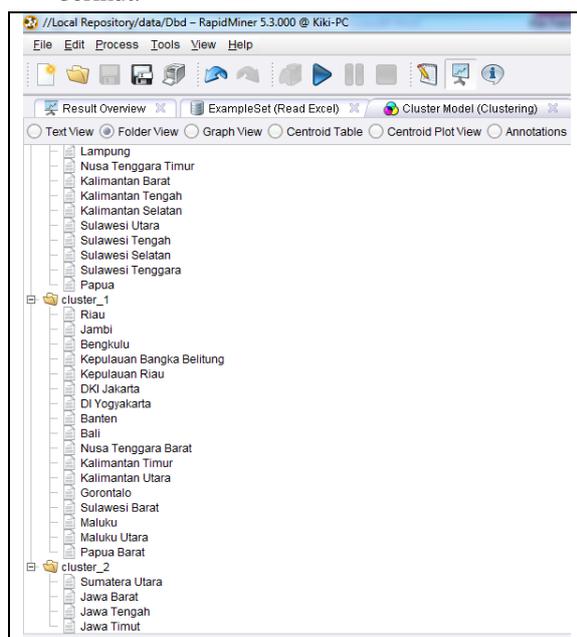
Gbr 5. Hasil Clustering

3. *Text view* : merupakan *sheet* untuk menampilkan *database* yang telah diolah secara keseluruhan lengkap dengan *clusternya*. Tampilan dari *cluster model (clustering)* dapat lihat pada gambar berikut.



Gbr 6. Cluster Model

4. *Folder view* : merupakan *sheet* untuk menampilkan *database* yang telah diolah secara keseluruhan lengkap dengan *clusternya*. Tampilan folder view dapat dilihat pada gambar berikut.



Gbr 7. Folder View

## V. KESIMPULAN

Dari hasil penelitian yang telah dilakukan, maka penulis dapat menarik beberapa kesimpulan, diantaranya sebagai berikut :

1. Pengujian yang dilakukan dalam penelitian ini, iterasi *clustering* pada data penyakit DBD (Demam Berdarah Dengue) terjadi sebanyak 6 kali iterasi.
2. Data tersebut diolah menggunakan *Rapidminer* untuk ditentukan nilai *centroid* dalam 3 *cluster* yaitu *cluster* tingkat tinggi (C1), *cluster* tingkat sedang (C2) dan *cluster* tingkat rendah (C3).
3. Dari 34 data jumlah kasus yang terjangkit Demam Berdarah Dengue (DBD) berdasarkan provinsi dapat diketahui, 4 provinsi *cluster* tingkat tinggi untuk penderita kasus DBD yakni *Sumatera Utara*, *Jawa Barat*, *Jawa Tengah* dan *Jawa Timur*, 13 provinsi *cluster* tingkat sedang dan 17 provinsi lainnya termasuk *cluster* tingkat rendah.

## UCAPAN TERIMA KASIH

Terima kasih kepada Bapak Agus Perdana Windarto yang membimbing dalam penulisan serta STIKOM Tunas Bangsa Pematangsiantar khususnya program studi Sistem Informasi yang telah menyediakan sarana dan prasarana selama penelitian ini berlangsung.

## REFERENSI

- [1] C. J. M. S. Fina Nasari, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," pp. 108-119.
- [2] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, pp. 26-33, 2017.
- [3] S. Mulyati, "Penerapan Data Mining Dengan Metode Clustering Untuk Pengelompokan Data Pengiriman Burung," vol. 1, no. Senatkom, 2015.
- [4] A. P. Windarto, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering," *Techno.COM*, vol. 16, no. 4, pp. 348-357, 2017.
- [5] N. Aggarwal, K. Aggarwal, and K. Gupta, "Comparative Analysis of K-means and Enhanced K-means Clustering Algorithm for Data Mining," *Int. Journal Sci. Eng. Res.*, vol. 3, no. 3, 2012.
- [6] Y. Siyamto and Program, "Pemanfaatan Data Mining Dengan Metode Clustering Untuk Evaluasi Biaya Dokumen Ekspor," vol. 1, no. 2, pp. 28-31, 2017.
- [7] R. Hidayat, R. Wasono, and M. Y. Darsyah, "Pengelompokan Kabupaten / Kota Di Jawa Tengah," pp. 240-250, 2017.