

BAB IV

HASIL DAN PEMBAHASAN

Pada penulisan ini data diagnosa terdiri dari 18.718 peserta JKN-KIS BPJS Kesehatan Kantor Cabang Prabumulih, data peserta diproses dengan mengelompokkan seluruh kebutuhan peserta yang telah berobat pada FKRTL di lingkungan kerja BPJS Kesehatan Kantor Cabang Prabumulih. Membuang data yang tidak diperlukan yang bukan merupakan diagnosa penyakit, menghitung jumlah data diagnosa penyakit dan membuang data duplikat dari data diagnosa penyakit. Kemudian data ditransformasikan untuk mengubah data dari bentuk asalnya ke dalam bentuk yang cocok untuk dikelompokkan.

4.1. Pengumpulan Data

Sumber data utama yang digunakan dalam penulisan ini adalah *database* JKN-KIS BPJS Kesehatan Kantor Cabang Prabumulih dengan jumlah keseluruhan data ada 18.718 *record*. Data yang diterima dalam bentuk tabel pada *microsoft excel* sehingga mempermudah untuk dilakukan pembersihan data atau memfilter data, seperti pada gambar 4.1. Data ini di dapatkan melalui perijinan kepada pihak JKN-KIS BPJS Kesehatan Kantor Cabang Prabumulih.

Diagekunder	Ikpst	Kddati2TerdFTAR	Kddiagprimer	Kdinachgs	Klsrawat	Nmdati2TerdFTAR	Nmdiagprimer	Nminachgs	Nmtkp	Umur
N40 - Hyperplasi	Laki-laki	0085	Z098	N-3-14-0	Kelas III	KAB. MUARA ENIM	Follow-up examination afte	PROSEDUR KECIL ENDOSKOPI PADA SALURAN I	RJTL	77
	Perempu	0085	M791	Q-5-44-0	Kelas III	KAB. MUARA ENIM	Myalgia	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	84
M4786 - Other sp	Perempu	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	78
M545 - Low back	Perempu	0091	Z501	M-3-16-0	Kelas III	KOTA PRABUMULIH	Other physical therapy	PROSEDUR THERAPI FISIK DAN PROSEDUR KECI	RJTL	78
M545 - Low back	Perempu	0091	Z501	M-3-16-0	Kelas III	KOTA PRABUMULIH	Other physical therapy	PROSEDUR THERAPI FISIK DAN PROSEDUR KECI	RJTL	78
M1996 - Arthrosi	Laki-laki	0091	Z501	M-3-16-0	Kelas III	KOTA PRABUMULIH	Other physical therapy	PROSEDUR THERAPI FISIK DAN PROSEDUR KECI	RJTL	76
M179 - narthrosi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	76
	Laki-laki	0085	N40	Q-5-44-0	Kelas III	KAB. MUARA ENIM	Hyperplasia of prostate	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	80
	Laki-laki	0085	N390	Q-5-40-0	Kelas III	KAB. MUARA ENIM	Urinary tract infection, site r	INFEKSI SALURAN KEMIH AKUT	RJTL	80
N40 - Hyperplasi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	84
N40 - Hyperplasi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	84
	Laki-laki	0091	N40	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Hyperplasia of prostate	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	84
N40 - Hyperplasi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	84
N40 - Hyperplasi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	84
N40 - Hyperplasi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	85
N40 - Hyperplasi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	85
N40 - Hyperplasi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	84
	Laki-laki	0091	N40	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Hyperplasia of prostate	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	84
N40 - Hyperplasi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	85
N40 - Hyperplasi	Laki-laki	0091	Z098	Q-5-44-0	Kelas III	KOTA PRABUMULIH	Follow-up examination afte	PENYAKIT KRONIS KECIL LAIN-LAIN	RJTL	84

Gambar 4. 1 Data set peserta

missing value dan memperbaiki kesalahan-kesalahan yang ada pada *data set* baru dalam format *excel*. Pada tahapan ini data akan dilakukan *cleaning* atau pembersihan data, terlihat seperti gambar 4.3 sehingga data tersebut dapat diolah dan dilakukan proses *data mining*. Data yang telah melewati tahap *pre-processing* akan disimpan dalam *data set* yang baru menggunakan *Microsoft Office Excel*. Perangkat lunak *Rapid Miner Studio* akan membuka file tersebut dan mengolahnya menggunakan metode *clustering* dengan algoritma *k-means*.

Penyakit	Jumlah
PROSEDUR KECIL ENDOSKOPI PADA SALURAN KEMIH	38
PENYAKIT KRONIS KECIL LAIN-LAIN	4615
PROSEDUR THERAPI FISIK DAN PROSEDUR KECIL MUSKULOSKLETAL	3411
INFEKSI SALURAN KEMIH AKUT	12
PROSEDUR KANDUNG KEMIH DAN SALURAN URIN BAWAH RINGAN	9
PERAWATAN LUKA	747
PROSEDUR ULTRASOUND LAIN-LAIN	200
PROSEDUR KECIL ENDOSKOPI PADA SALURAN KEMIH	38
NEONATAL, BBL GROUP-5 DENGAN KONGENITAL/INFEKSI PERINATAL RINGAN	33
PENYAKIT AKUT BESAR LAIN-LAIN	91
ASTHMA & BRONKIOLITIS RINGAN	27
DIAGNOSIS SISTEM REPRODUKSI LAKI-LAKI SELAIN TUMOR (RINGAN)	22
NEONATAL, BBL GROUP-3 TANPA PROSEDUR MAYOR RINGAN	20
INFEKSI NON BAKTERI RINGAN	290
GANGGUAN SEL DARAH MERAH SELAIN KRISIS ANEMIA SEL SICKLE RINGAN	29
DIAGNOSIS SISTEM PENCERNAAN LAIN-LAIN (RINGAN)	120
GANGGUAN HATI SELAIN TUMOR, SIROSIS ATAU HEPATITIS ALKOHOLIK RINGAN	17
DIAGNOSIS SISTEM PENCERNAAN LAIN-LAIN (SEDANG)	25
HIPERTENSI RINGAN	84
NYERI ABDOMEN & GASTROENTERITIS LAIN-LAIN (RINGAN)	131

The image shows a filter dialog box overlaid on the spreadsheet. The dialog box has a search bar and a list of items with checkboxes. The items in the list are: (Select All), ABORSI, ABORSI MENGANCAM, ABORTUS MENGANCAM RINGAN, ABORTUS RINGAN, ANGINA PEKTORIS DAN NYERI DAE, ANGINA PEKTORIS DAN NYERI DAE, ANGINA PEKTORIS DAN NYERI DAE, and ARTRITIS SEPSIS RINGAN. All checkboxes are checked. The dialog box also has 'OK' and 'Cancel' buttons at the bottom.

Gambar 4.3 Filter data menggunakan *sort & filter*

4.4. Transformasi Data

Pada tahap ini dilakukan proses perubahan data, agar data dapat diolah dengan menggunakan algoritma *k-means clustering*. Data yang *non-numeric* yang merupakan data diagnosa penyakit akan dihitung jumlah peserta JKN-KIS nya yang terkena penyakit sehingga dilakukan

proses inisiasi ke dalam bentuk *numeric* yaitu jumlah peserta JKN-KIS yang terjangkit penyakit. Atribut yang digunakan pada penulisan ini adalah diagnosa penyakit dan jumlah peserta JKN-KIS nya. Proses inisiasi sebagai berikut:

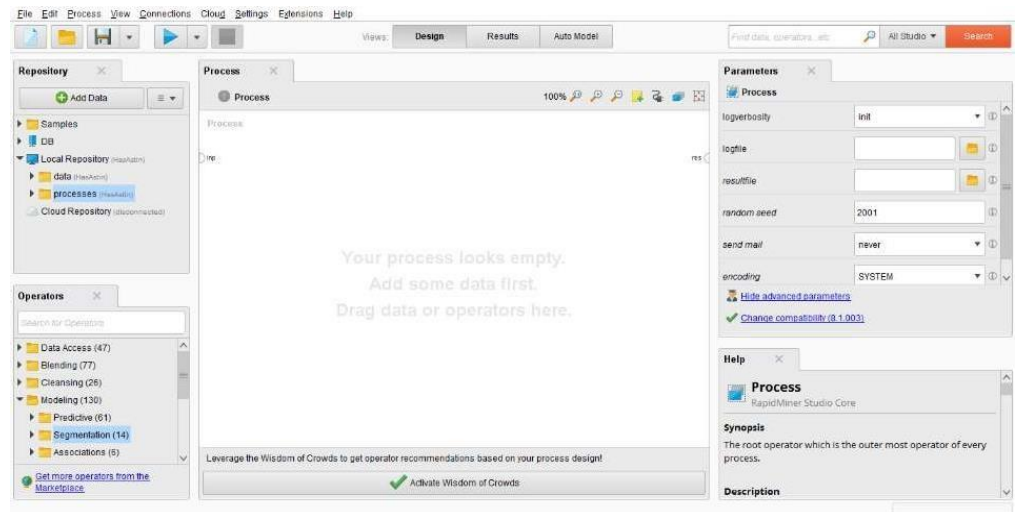
Penyakit	Jumlah
PENYAKIT KRONIS KECIL LAIN-LAIN	4615
INFEKSI SALURAN KEMIH AKUT	12
NEONATAL, BBL GROUP-5 DENGAN KONGENITAL/INFEKSI PERINATAL RINGAN	33
PENYAKIT AKUT BESAR LAIN-LAIN	91
ASTHMA & BRONKIOLITIS RINGAN	27
NEONATAL, BBL GROUP-3 TANPA PROSEDUR MAYOR RINGAN	20
INFEKSI NON BAKTERI RINGAN	290
GANGGUAN SEL DARAH MERAH SELAIN KRISIS ANEMIA SEL SICKLE RINGAN	29
GANGGUAN HATI SELAIN TUMOR, SIROSIS ATAU HEPATITIS ALKOHOLIK RINGAN	17
HIPERTENSI RINGAN	84
NYERI ABDOMEN & GASTROENTERITIS LAIN-LAIN (RINGAN)	131
PENYAKIT AKUT KECIL LAIN-LAIN	520
PENYAKIT INFEKSI BAKTERI DAN PARASIT LAIN-LAIN RINGAN	328
NEONATAL, BBL GROUP-4 TANPA PROSEDUR MAYOR RINGAN	7
PENYAKIT INFEKSI BAKTERI DAN PARASIT LAIN-LAIN RINGAN	161
PERADANGAN DAN INFEKSI PERNAFASAN RINGAN	73
GANGGUAN HATI SELAIN TUMOR, SIROSIS ATAU HEPATITIS ALKOHOLIK RINGAN	12
PENYAKIT KENCING MANIS & GANGGUAN NUTRISI/ METABOLIK BERAT	8
KEGAGALAN JANTUNG RINGAN	88
KECEDERAAN PEMBULUH DARAH OTAK DENGAN INFARK RINGAN	1
EPISODIKAL DAN BENGANGKATAN ALAT EKSKLATORIAL SEDANG	2

Gambar 4. 4 Hasil transformasi data

4.5. Pengolahan Data

Pada tahapan ini dilakukan pemodelan data menggunakan hasil dari transformasi data seperti gambar 4.4, adapun metode yang dipakai pada penulisan ini adalah metode *clustering* dengan menggunakan algoritma *k-means*. Data yang sudah dikumpulkan, diseleksi dan ditransformasi akan dilakukan pengolahan dari data tersebut dengan menggunakan metode *clustering*. Metode ini bekerja dengan melakukan pengelompokan data yang memiliki kesamaan karakteristik pada setiap data.

4.5.1. Pengujian Software RapidManer



Gambar 4.5 RapidMiner Studio explorer

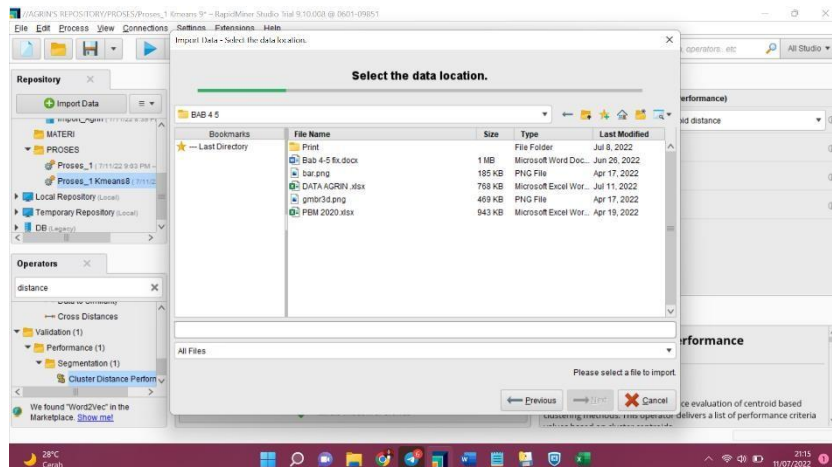
Atribut yang digunakan sebagai *class* adalah jumlah peserta JKN-KIS yang terdiagnosa penyakit. Data akan dianalisis berdasarkan jumlah peserta JKN-KIS yang paling banyak diantara peserta JKN-KIS lainnya. Data diagnosa penyakit dibuat dalam *data set* baru dengan format *.xlsx* (*excel*) untuk bisa dianalisis *software* yang dipakai adalah *RapidMiner Studio* seperti pada gambar 4.5.

Setelah menentukan operator yang akan digunakan lakukan *drag and drop operators* seperti gambar 4.6 dibawah, kemudian *open file data set* menggunakan operator *Read Excel* seperti pada gambar 4.7

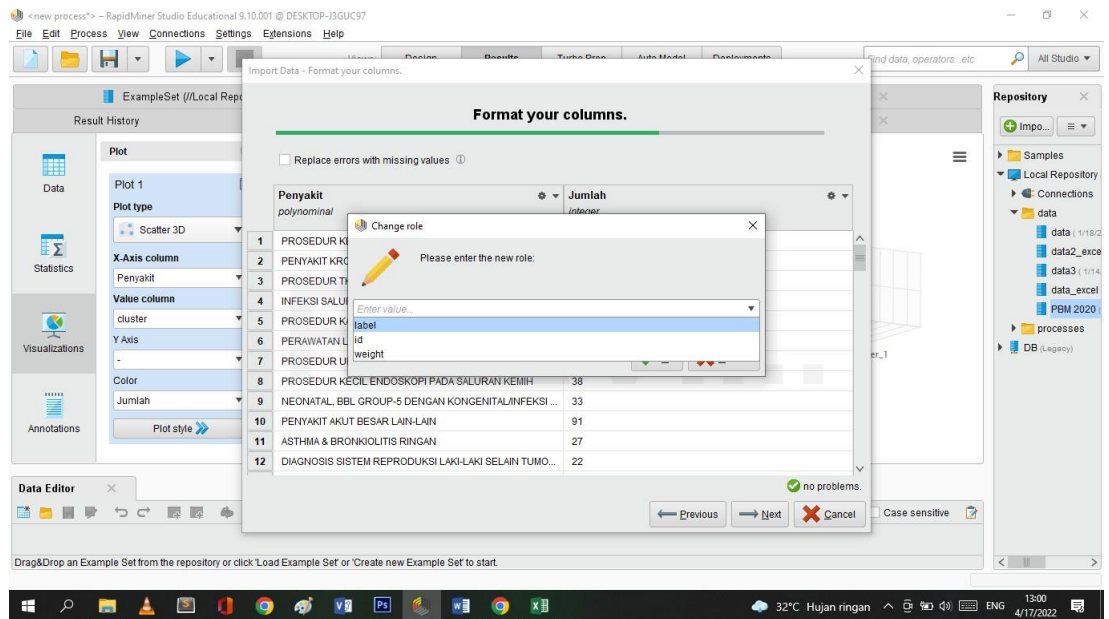


Gambar 4.6 Operator *read excel* dan *clustering*

Proses *open file* ini dilakukan oleh operator *read excel* dan melalui parameter *Import Configuration Wizard* untuk menemukan *file* yang sudah dibuat dan pilih *sheet* tempat data tersebut, agar tidak terjadi kesalahan *run* data seperti gambar 4.7 untuk memastikan datanya sudah benar sesuai dengan type datanya, jika *type* data tidak sesuai maka operator tidak akan melakukan proses *run*.



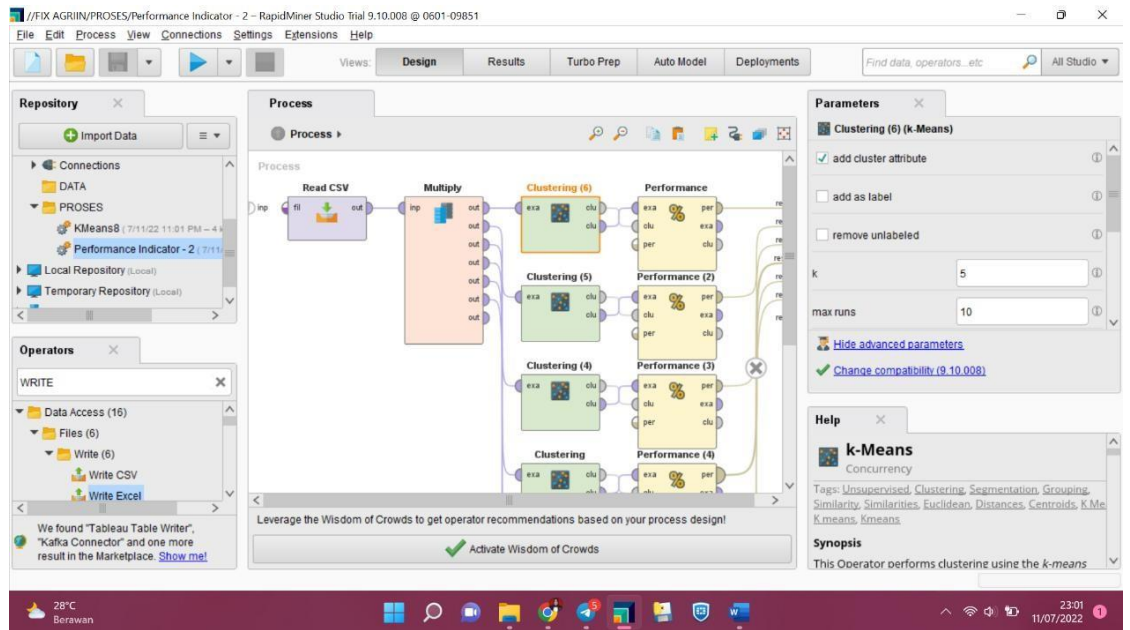
Gambar 4.7 *Open file excel*



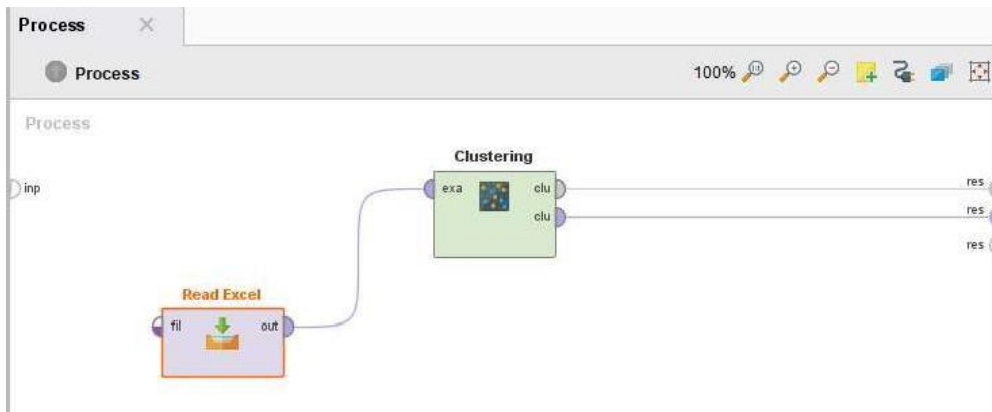
Gambar 4.8 Edit parameter list

Pada dasarnya *rapidminer studio* dapat membaca type data setiap atribut seperti pada gambar 4.8, namun untuk mengurangi kesalahan pada proses *run* ada baiknya dipastikan kembali. Setelah *open file excel* langkah selanjutnya adalah menghubungkan operator *read excel* dengan operator *clstering k-means* dan mengatur melalui parameternya, sekaligus menentukan jumlah clusternya. Untuk menemukan jumlah yaitu 8 *cluster* seperti pada gambar 4.9. Setelah operator *read excel* dan operator *clustering k-means* diatur langkah berikutnya adalah menghubungkan kedua operator tersebut, dan hubungkan juga pada proses *out put* untuk mendapatkan hasil *run* operatornya, seperti gambar 4.10. Pada gambar 4.9 Penulis menguji dengan melakukan clustering secara 6 kali, untuk menemukan jumlah cluster yang paling cocok untuk data

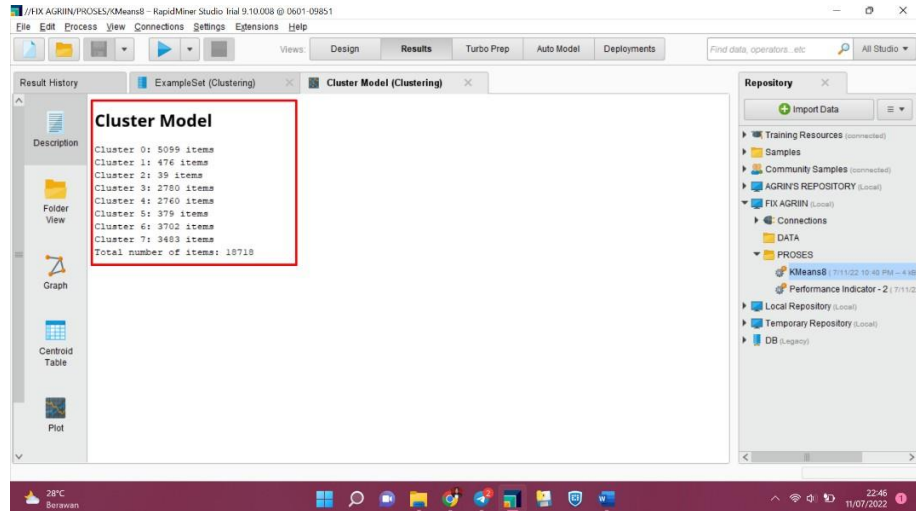
tersebut.



Gambar 4. 9 Menentukan cluster

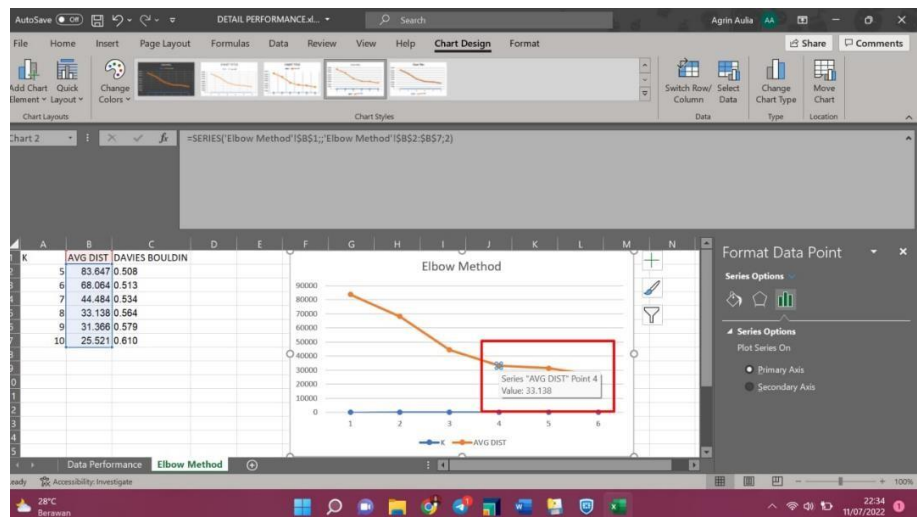


Gambar 4. 10 Menghubungkan operator

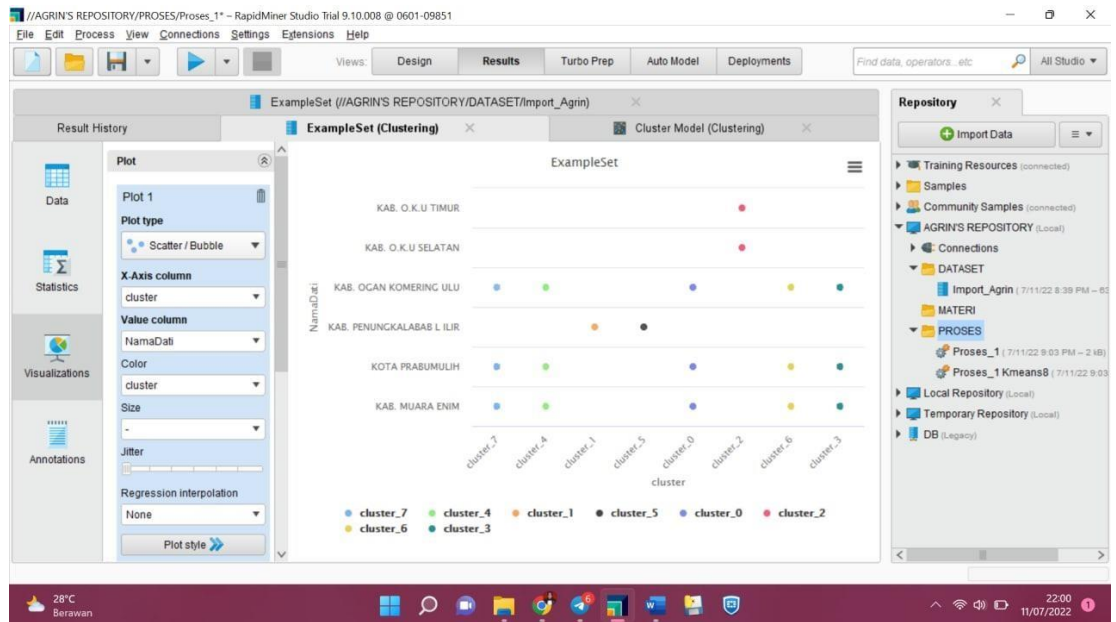


Gambar 4. 11 Hasil *cluster*

Telah dilakukan uji coba menggunakan metode KMeans dari 5,6,7,8,9 hingga 10 dan ditemukan metode dengan *performance* terbaik adalah cluster dengan k8. *Data Avg within centroid distance* tersebut disusun dan dipilih menggunakan Elbow Method untuk mencari angka cluster yang paling terbaik.

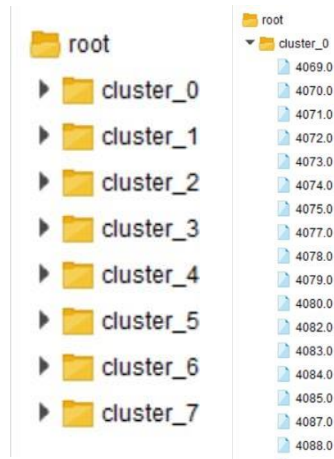


Gambar 4. 12 Hasil data *cluster K-Means* dalam implementasi *RapidMiner*



Gambar 4. 13 Visualize data berdasarkan Nama Dati Peserta JKN-KIS.

Diagnosa penyakit terdiri dari 200 diagnosa penyakit bisa dilihat pada gambar 4.13 dan tabel 1 sampai tabel 9. Pada data set yang diberikan oleh JKN- KIS BPJS kesehatan kantor cabang prabumulih ada keterangan yang bukan diagnosa penyakit yaitu ada prosedur-prosedur dalam menangani penyakit yang mana ini tidak teridentifikasi, sehingga keterangan tersebut tidak digunakan.



Gambar 4. 14 Isi folder *cluster*

Data yang di *cluster* akan masuk pada folder yang dikelompokkan berdasarkan kesamaan karakter seperti pada gambar 4.14 ketika salah satu data tersebut dibuka maka akan menunjukkan detailnya.

4.5.2. Algoritma K-Means

Dalam menggunakan algoritma *k-means* akan melakukan pengulangan tahapan hingga terjadi kestabilan. Penulis melakukan enam kali iterasi dalam melakukan pengujian menggunakan algoritma *k-means* pada data peserta JKN- KIS dengan tahapan sebagai berikut.

1. Menentukan jumlah *cluster* dan menentukan koordinat titik tengah *cluster*. Kelompok *cluster* yang dibuat adalah 8 kelompok agar terlihat jarak antara titik *centroidnya*. Total jumlah inisiasi data dari atribut tersebut untuk menentukan kelompok diambil dari frekuensi kurang, frekuensi sedang dan frekuensi padat secara acak.

2. Penentuan nilai *cluster* untuk dijadikan acuan dalam melakukan perhitungan jarak objek ke *centroid*, perhitungan jarak mangacu pada rumus *euclidean*.

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j - y_j)^2}$$

Keterangan:

d = data titik dokumen (*euclidean*) P = data *record*

Q = data *centroid*

Rumus *euclidean* merupakan perhitungan jarak antar *centroid*, perhitungan ini dilakukan di *excel* karena jumlah data yang sangat banyak.

3. Setelah jarak antar *centroid* dihitung dengan menggunakan rumus *euclidean distance*, maka dilakukan pengelompokan *centroid* sesuai dengan hasil dari jarak antar *centroid*. Hasil pada perhitungan jarak tersebut digunakan untuk penentuan kelompok *clustering*. Penentuan dalam pengelompokan *centroid* adalah sebagai berikut:
 - a. Jika jarak *centroid* 1 lebih kecil dari jarak *centroid* 2, *centroid* 3, *centroid* 4, *centroid* 5, *centroid* 6, *centroid* 7, dan *centroid* 8, maka termasuk kelompok *centroid* 1.
 - b. Jika jarak *centroid* 2 lebih kecil dari jarak *centroid* 1, *centroid* 3, *centroid* 4, *centroid* 5, *centroid* 6, *centroid* 7, dan *centroid* 8, maka termasuk kelompok *centroid* 2.
 - c. Jika jarak *centroid* 3 lebih kecil dari jarak *centroid* 1,

centroid 2, centroid 4, centroid 5, centroid 6, centroid 7,
dan *centroid 8*, maka termasuk kelompok *centroid 3*.

- d. Jika jarak *centroid 4* lebih kecil dari jarak *centroid 1, centroid 2, centroid 3, centroid 5, centroid 6, centroid 7,* dan *centroid 8*, maka termasuk kelompok *centroid 4*.
 - e. Jika jarak *centroid 5* lebih kecil dari jarak *centroid 1, centroid 2, centroid 3, centroid 4, centroid 6, centroid 7,* dan *centroid 8*, maka termasuk kelompok *centroid 5*.
 - f. Jika jarak *centroid 6* lebih kecil dari jarak *centroid 1, centroid 2, centroid 3, centroid 4, centroid 5, centroid 7,* dan *centroid 8*, maka termasuk kelompok *centroid 6*.
 - g. Jika jarak *centroid 7* lebih kecil dari jarak *centroid 1, centroid 2, centroid 3, centroid 4, centroid 5, centroid 6,* dan *centroid 8*, maka termasuk kelompok *centroid 7*.
 - h. Jika jarak *centroid 8* lebih kecil dari jarak *centroid 1, centroid 2, centroid 3, centroid 4, centroid 5, centroid 6,* dan *centroid 7*, maka termasuk kelompok *centroid 8*.
- Setelah semua diagnosa penyakit masuk dalam kelompoknya, langkah berikutnya adalah menghitung kembali *centroid* baru masing-masing cluster dengan cara menjumlahkan data-data yang ada di masing-masing *cluster* dan membaginya dengan jumlah data pada *cluster* tersebut. Pada iterasi kedua, melakukan

perhitungan jarak setiap diagnosa penyakit dengan masing-masing *cluster* dengan *centroid* yang baru. Proses iterasi bisa dihentikan jika pada suatu iterasi ke- n , tidak ada pertukaran anggota sama sekali antar *cluster* atau anggota masing-masing *cluster* sama dengan iterasi sebelumnya.

