

## PENERAPAN DATA MINING UNTUK PENGELOMPOKAN HASIL DIAGNOSA PENYAKIT PESERTA JKN-KIS BPJS KESEHATAN KANTOR CABANG PRABUMULIH

### APPLICATION OF DATA MINING FOR CLASSIFICATION OF DIAGNOSIS RESULTS OF JKN-KIS BPJS KESEHATAN PARTICIPANTS PRABUMULIH BRANCH OFFICE

Agrin Aulia<sup>1\*)</sup>

<sup>1</sup>Sistem Informasi, Universitas Sriwijaya

Email: <sup>1\*)</sup> [agrinauliaprakoso@yahoo.co.id](mailto:agrinauliaprakoso@yahoo.co.id)

#### ARTICLE HISTORY

Received [25 November 2022]

Revised [25 December 2022]

Accepted [17 January 2023]

#### KEYWORDS

Diagnosis, Clustering, Data Mining, K-Means

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



#### ABSTRAK

BPJS Kesehatan telah bekerjasama dengan beberapa fasilitas kesehatan di Indonesia untuk memberikan pelayanan kesehatan. Selama ini penggunaan pengelompokan hasil diagnosa penyakit peserta JKN-KIS yang digunakan oleh Mitra BPJS Kesehatan dan juga Dinas Kesehatan hanya menggunakan gambaran umum dari data pasien. Pemanfaatan data diagnosis penyakit dari pasien JKN-KIS secara optimal belum dilakukan. Oleh karena itu, Clustering juga merupakan metode pengelompokan beberapa objek data menjadi suatu kelompok informasi tertentu dengan tingkat kemiripan yang tinggi. Pengujian nilai terbaik dengan metode Elbow menghasilkan nilai K yang mendekati Bentuk Sudut 90 derajat yang berada pada cluster ke-4. Sehingga sistem pengelompokan data rekam medis menghasilkan cluster terbaik dengan menggunakan 8 cluster.

#### ABSTRACT

BPJS Kesehatan has collaborated with several health facilities in Indonesia to provide health services. So far, the use of grouping the results of disease diagnosis of JKN-KIS participants used by BPJS Health Partners and also the Health Office only uses a general description of the patient data. Optimal utilization of disease diagnosis data from JKN-KIS patients has not been done. Therefore, Clustering is also a method of grouping several data objects into a certain group of information with a high degree of similarity. Testing the best value with the Elbow method produces a K value that is close to The Shape of a 90 degree angle that is in the 4th cluster. So therefore, the medical record data grouping system produces the best cluster by using 8 clusters.

#### PENDAHULUAN

Data Mining adalah ekstraksi pengetahuan tersembunyi dari gudang data. Ini adalah teknologi yang kuat dengan ruang lingkup yang baik untuk menganalisis dan memprediksi informasi penting dari database. Data Meteorologis bersifat produktif, dinamis, kompleks dan dimensi tinggi. Beberapa teknik penambangan data sangat ideal untuk membuat prediksi cuaca. Banyak pekerjaan yang telah dilakukan dan masih berlangsung karena peneliti menganggapnya sebagai hal yang menjanjikan dan bermanfaat. (A Geetha, et al 2014)

BPJS Kesehatan telah bekerja sama dengan beberapa fasilitas kesehatan yang ada di Indonesia untuk memberikan pelayanan Kesehatan. Pelayanan kesehatan memiliki keunikan tersendiri, ada banyak aspek yang dapat mengukur kepuasan pasien tapi hanya sedikit yang dapat menggambarkan kepuasan pasien, salah satu dimensi mutu kepedulian petugas rumah sakit yang sangat mempengaruhi kepuasan para pengguna jasa pelayanan kesehatan dengan memahami kebutuhan mereka dan memberikan kemudahan untuk mendapatkan informasi saat mereka ingin memperoleh bantuannya (Widia et al, 2019). Pelayanan kesehatan yang dinilai dari sudut *assurance* (jaminan) adalah ketersediaan dokter di fasilitas kesehatan setiap saat dibutuhkan oleh pasien, perilaku dokter yang selalu memberikan rasa aman dan nyaman. Di samping itu juga ketersediaan perawat yang terdidik dan mampu melayani pasien dengan baik (Sri et al., 2018).

Berdasarkan data per 31 Desember 2020, jumlah penduduk Indonesia yang sudah terdaftar sebagai peserta JKN-KIS adalah sebanyak 222.461.906 jiwa. BPJS Kesehatan sebagai Badan Penyelenggara Jaminan Sosial



Kesehatan yang mana sebagai pihak yang memiliki dan bertanggung jawab atas sumber data peserta JKN-KIS tentunya dapat bersinergi dengan Dinas Kesehatan dan RSUD Kota Prabumulih yang mana sebagai mitra BPJS Kesehatan untuk dapat meningkatkan mutu pelayanan dan memberikan kenyamanan kepada peserta JKN-KIS.

Sejauh ini pemanfaatan pengelompokan hasil diagnosa penyakit peserta JKN-KIS yang digunakan oleh mitra BPJS Kesehatan dan juga Dinas Kesehatan hanya menggunakan gambaran umum dari data pasien tersebut. Belum dilakukannya pemanfaatan secara optimal terhadap data diagnosa penyakit dari pasien peserta JKN-KIS. Sehingga belum dilakukannya identifikasi dan analisa terhadap pola penyakit yang terjadi pada suatu daerah tertentu. Terciptanya suatu inovasi merupakan ukurannya dari keberhasilan dari otonomi daerah. (Cindy, 2013).

Maka dari itu, *Data Mining* sangat diperlukan dalam penelitian ini agar mampu memberikan pola atau gambaran mengenai pengelompokan diagnosa penyakit yang ada dari para peserta JKN-KIS BPJS Kesehatan sehingga Dinas Kesehatan dan Pemerintah Daerah setempat dapat mengetahui dan mendapatkan gambaran pasti mengenai penyakit yang sering terjadi di masyarakat dan dapat ditindaklanjuti dengan memberikan edukasi mengenai pola hidup yang sehat.

Berdasarkan uraian di atas, penulis tertarik untuk membuat tugas akhir dengan judul “Penerapan *Data Mining* untuk Pengelompokan Hasil Diagnosa Penyakit Peserta JKN-KIS BPJS Kesehatan Kantor Cabang Prabumulih”.

## LANDASAN TEORI

### BPJS Kesehatan

BPJS Kesehatan merupakan Badan Hukum Publik yang menyelenggarakan Program Jaminan Kesehatan Nasional – Kartu Indonesia Sehat (JKN-KIS) berdasarkan Undang – Undang Nomor 24 tahun 2011 tentang Badan Penyelenggara Jaminan Sosial. Sejak bertransformasi dari PT Asuransi Kesehatan (Persero) atau PT Askes (Persero) pada 1 Januari 2014.

Jaminan pemeliharaan kesehatan di Indonesia sudah ada sejak zaman colonial Belanda, Setelah kemerdekaan, tepatnya pada tahun 1949, pasca pengakuan kedaulatan oleh Pemerintah Belanda, upaya untuk menjamin kebutuhan pelayanan kesehatan bagi masyarakat, khususnya bagi Pegawai Negeri Sipil (PNS) beserta keluarga, tetap dilanjutkan. Adalah Prof. Gerrut Augustinus Siwabessy atau yang lebih dikenal dengan sebutan Prof. Siwabessy, selaku Menteri Kesehatan yang menjabat pada saat itu, mengajukan sebuah gagasan untuk perlu segera menyelenggarakan program asuransi kesehatan semesta (Universal Health Insurance) yang pada saat itu mulai diterapkan di banyak negara maju dan tengah berkembang pesat.

Pada saat itu kepesertaan program tersebut baru mencakup Pegawai Negeri Sipil (PNS) beserta anggota keluarganya saja. Namun Siwabessy yakin suatu hari nanti, klimak dari pembangunan derajat kesehatan masyarakat Indonesia akan tercapai melalui system yang dapat menjamin kesehatan seluruh warga bangsa ini.

Pada 1968, pemerintah menerbitkan Peraturan Menteri Kesehatan Nomor 1 Tahun 1968 dengan membentuk Badan Penyelenggara Dana Pemeliharaan Kesehatan (BPDPK) yang mengatur pemeliharaan kesehatan bagi pegawai negara dan penerima pensiun beserta keluarganya.

Selang beberapa waktu kemudian, Pemerintah mengeluarkan Peraturan Pemerintah Nomor 22 dan 23 Tahun 1984. BPDPK pun berubah status dari sebuah badan di lingkungan Departemen Kesehatan menjadi BUMN, yaitu PERUM HUSADA BHAkti (PHB), yang melayani jaminan kesehatan bagi PNS, pensiunan PNS, veteran, perintis kemerdekaan, beserta anggota keluarganya. Pada tahun 1992, PHB berubah status menjadi PT Askes (Persero) melalui Peraturan Pemerintah Nomor 6 Tahun 1992. PT Askes (Persero) mulai menjangkau karyawan BUMN melalui program Askes Komersial. Selanjutnya pada Januari 2005, PT Askes (Persero) dipercaya pemerintah untuk melaksanakan program jaminan kesehatan bagi masyarakat miskin (PJKMM) yang kemudian dikenal menjadi program Askeskin dengan sasaran peserta masyarakat miskin dan tidak mampu sebanyak 60 juta jiwa yang iurannya dibayarkan oleh Pemerintah Pusat.

PT Askes (Persero) juga menciptakan Program Jaminan Kesehatan Masyarakat Umum (PJKMU), yang ditujukan bagi masyarakat yang belum tercover oleh Jamkesmas, Askes Sosial, maupun asuransi swasta. Hingga saat itu, ada lebih dari 200 kabupaten/kota atau 6,4 juta jiwa yang telah menjadi peserta PJKMU. PJKMU adalah Jaminan Kesehatan Daerah (Jamkesda) yang pengelolaannya diserahkan kepada PT Askes (Persero). Langkah menuju cakupan kesehatan semesta pun semakin nyata dengan resmi beroperasinya BPJS Kesehatan pada 1 Januari 2014, sebagai transformasi dari PT Askes (Persero). Hal ini bermula sejak tahun 2004 saat pemerintah mengeluarkan UU Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional (SJSN) dan kemudian pada tahun 2011 pemerintah juga menetapkan UU Nomor 24 Tahun 2011 tentang Badan Penyelenggara Jaminan Sosial (BPJS) serta menunjuk PT Askes (Persero) sebagai penyelenggara program jaminan sosial di bidang kesehatan, sehingga PT Askes (Persero) pun berubah menjadi BPJS Kesehatan. Melalui Program Jaminan Kesehatan Nasional-Kartu Indonesia Sehat (JKN-KIS) yang diselenggarakan oleh BPJS Kesehatan, negara hadir di tengah kita untuk memastikan seluruh penduduk Indonesia terlindungi oleh jaminan kesehatan yang komprehensif, adil, dan merata.

## Diagnosa

Menurut (Retno et al, 2012) Diagnosa sering digunakan oleh dokter untuk menyebutkan suatu penyakit yang diderita oleh seorang pasien atau keadaan yang dapat menyebabkan seorang pasien memerlukan atau mencari atau menerima asuhan medis guna memperoleh pelayanan pengobatan, pencegahan memburuknya masalah Kesehatan atau juga untuk peningkatan Kesehatan. Sedangkan diagnosis utama adalah penyakit, cacat, luka atau keadaan sakit yang utama dari pasien yang dirawat di rumah sakit.

Batasan diagnosis utama adalah :

1. Diagnosis yang ditentukan setelah cermat dikaji
2. Menjadi alasan untuk dirawat
3. Menjadi fakta arahan atau pengobatan

Beberapa macam diagnosis menurut WHO adalah sebagai berikut :

1. *Principal Diagnosis*
2. *Other Diagnosis*
3. *Complication*

## Data Mining

Data Mining yang juga disebut sebagai *Knowledge Discovery in Databases* berarti proses ekstraksi nontrivial implisit, informasi yang sebelumnya tidak diketahui dan berpotensi berguna (seperti aturan pengetahuan, kendala, keteraturan) dari data dalam database. Ada juga banyak istilah lain, muncul di beberapa artikel dan dokumen, membawa arti yang serupa atau sedikit berbeda seperti penambangan pengetahuan dari database, ekstraksi pengetahuan, arkeologi data, pengerukan data, analisi data dan sebagainya. Dengan pertemuan pengetahuan dalam database, pengetahuan yang menarik, keteraturan atau informasi tingkat tinggi dapat diekstraksi dari kumpulan data yang relevan dalam database dan diselidiki dari sudut yang berbeda, dan dengan demikian database yang besar berfungsi sebagai sumber yang kaya dan dapat diandalkan untuk generasi pengetahuan dan verifikasi. Informasi dan pengetahuan pertambangan dari database yang besar telah diakui oleh banyak peneliti sebagai topik penelitian utama dalam system database dan pembelajaran mesin dan oleh banyak perusahaan industri sebagai area penting dengan peluang pendapatan besar.

Pengetahuan yang ditemukan dapat diterapkan pada manajemen informasi, pemrosesan query, pengambilan keputusan, *process control* dan banyak aplikasi lainnya. Para peneliti di banyak bidang yang berbeda, termasuk system database, system basis pengetahuan, kecerdasan buatan, pembelajaran mesin, akuisisi pengetahuan, statistic, database spasial dan visualisasi data telah menunjukkan minat yang besar dalam penambangan data (*data mining*). Selain itu, muncul beberapa aplikasi yang dalam penyediaan informasi yang menyediakan layanan seperti layanan *online* dan World Wide Web juga menyerukan berbagai teknik pengembangan data (*data mining*) untuk lebih memahami perilaku pengguna, memperbaiki layanan yang diberikan dan untuk meningkatkan peluang bisnis. (Ming et al, 1997).

Dalam sebuah perusahaan, *Data Mining* dapat digunakan untuk mencapai tujuan ataupun target tertentu. Telah banyak penelitian dari berbagai bidang yang menggunakan *Data Mining* untuk mencapai tujuannya. Sebuah kepastian bahwa semakin bertambah usia sebuah perusahaan, maka semakin bertambah pula jumlah data yang tersimpan dalam *database* perusahaan tersebut. Hal ini tentunya menjadi salah satu dasar diperlukannya penerapan *Data Mining* pada perusahaan tersebut.

Jika tambang data perusahaan yang tersimpan diibaratkan sebagai tumpukan dalam tambang bebatuan maka peranan *Data Mining* akan membantu perusahaan tersebut untuk memilah, mengekstraksi dan menghasilkan batubara yang pada hal ini merujuk pada informasi yang berguna. Batubara yang merupakan hasil dari proses *Data Mining* ini tentunya akan menjadi peluang baru bagi perusahaan tersebut untuk dapat meningkatkan kinerja pada perusahaan tersebut.

## METODE PENELITIAN

### Objek Penelitian

Untuk mencapai tujuan penelitian, peneliti menggunakan data riwayat pelayanan kesehatan peserta JKN-KIS BPJS Kesehatan Kantor Cabang Prabumulih di RSUD Kota Prabumulih pada triwulan 1 tahun 2020.

### Jenis Data

Jenis data yang dikumpulkan dalam tugas akhir ini adalah data primer dan data sekunder. Data pada penelitian tugas akhir ini merupakan yang dikumpulkan secara langsung dari objek penelitian tugas akhir.

### Sumber Data

Data yang digunakan dalam penelitian ini merupakan data riwayat pelayanan kesehatan peserta JKN-KIS BPJS Kesehatan Kantor Cabang Prabumulih di RSUD Kota Prabumulih pada triwulan pertama tahun 2020.



## Perangkat Lunak

Rapid Miner adalah tools commercial yang populer digunakan untuk Data Mining yang dikembangkan oleh Rapid-I, GmbH dari Dortmund Jerman. RapidMiner sebelumnya dikenal sebagai YALE (*Yet Another Learning Environment*), dulu dikembangkan pada tahun 2001. RapidMiner adalah platform perangkat lunak yang menyediakan lingkungan terintegrasi untuk mesin pembelajaran, penggalian data, penambangan teks, analitik prediktif dan analitik bisnis. Ini digunakan juga untuk bisnis dan aplikasi industry serta untuk penelitian, Pendidikan, pelatihan, pembuatan *prototype* dan pengembangan aplikasi serta mengintegrasikan semua Langkah proses Data Mining termasuk visualisasi hasil, validasi dan optimasi. RapidMiner juga adalah sebuah *interface* penggunaan grafis yang cukup kuat dan intuitif untuk desain proses analisis. Ini juga dapat dengan mudah mengintegrasikan algoritma khusus dengan memanfaatkan API ekstensi yang kuat dan terbuka, grafik 3D, matriks pencar dan peta yang dapat diolah sendiri. RapidMiner memungkinkan untuk mengubah data menjadi bagan yang dapat diolah dengan dukungan untuk memperbesar, menggeser dan mengubah skala untuk dampak visual yang maksimal. (A Geetha, et al 2014)

## Mengumpulkan Data

Pada tahap ini penulis melakukan permohonan pengumpulan data secara langsung melalui BPJS Kesehatan Kantor Cabang Prabumulih secara resmi dan sesuai prosedur yang telah ditetapkan.

## Mempelajari Literatur

Pada tahap ini penulis melakukan pengumpulan informasi dan pengetahuan melalui berbagai sumber literatur baik berupa jurnal, karya ilmiah dan seluruh informasi yang berkaitan dengan judul dan topik yang sudah dipilih.

## Menganalisa Data

Pada tahap ini, dataset disiapkan untuk proses *Data Mining*. Dalam penelitian ini digunakan data dengan total 11 atribut data dengan deskripsi sebagai berikut :

## Perancangan *K-Means Clustering Structure*

Alasan dipilihnya metode *Clustering* dalam penelitian ini adalah karena terdapatnya beberapa kemiripan dari data yang tersedia. Metode *Clustering* merupakan metode pengolahan data dengan berdasarkan tingkat kemiripan / *similaritas* tertinggi hingga terendah.

## Melakukan Evaluasi Data

Tahap Evaluasi Data dilakukan untuk melakukan pengujian apakah data yang digunakan dalam penelitian ini sesuai dengan kebutuhan dari penelitian ini.

## Pengujian Hasil

Pengujian Hasil dalam tahap ini bertujuan untuk melakukan pengujian apakah metode *Clustering* yang digunakan oleh penulis sudah tepat dan mampu memberikan hasil dan solusi yang baik kepada RSUD Kota Prabumulih, Dinas Kesehatan dan BPJS Kesehatan Kantor Cabang Prabumulih. Pengujian dilakukan menggunakan aplikasi RapidMiner versi 5.3

## Penarikan Kesimpulan

Pada tahap penarikan kesimpulan akan ditentukan apakah informasi / pengetahuan yang dihasilkan sudah sesuai dengan tujuan penelitian.

## Menerapkan Pengetahuan

Dalam tahap penerapan pengetahuan ini, diharapkan informasi dan pengetahuan yang diperoleh dari penelitian ini dapat memberikan solusi yang efektif kepada RSUD Kota Prabumulih, Dinas Kesehatan dan BPJS Kesehatan Kantor Cabang Prabumulih.

## 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 asanya ke daam bentuk yang cocok untuk dikelompokkan.

## Pengumpulan Data

Sumber data utama yang digunakan daam penulisan ini adaah *database* JKN-KIS BPJS Kesehatan Kantor Cabang Prabumulih dengan jumlah keseluruhan data ada 18.718 *record*. Data yang diterima daam bentuk tabel pada *microsoft excel* sehingga mempermudah untuk dilakukan pembersihan data atau memfilter data, seperti pada gambar 1. Data ini di dapatkan melaui perijinan kepada pihak JKN-KIS BPJS Kesehatan Kantor Cabang Prabumulih.



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)	17
GANGGUAN HATI SELAIN TUMOR, SIROSIS ATAU HEPATITIS ALKOHOLIK RINGAN	25
DIAGNOSIS SISTEM PENCERNAAN LAIN-LAIN (SEDANG)	84
HIPERTENSI RINGAN	131
NYERI ABDOMEN & GASTROENTERITIS LAIN-LAIN (RINGAN)	131

Gambar 3. Filter data menggunakan sort & filter

### 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 daam bentuk *numeric* yaitu jumlah peserta JKN-KIS yang terjangkau penyakit. Atribut yang digunakan pada penulisan ini adaah diagnosa penyakit dan jumlah peserta JKN-KIS nya. Proses inisiasi sebagai berikut:

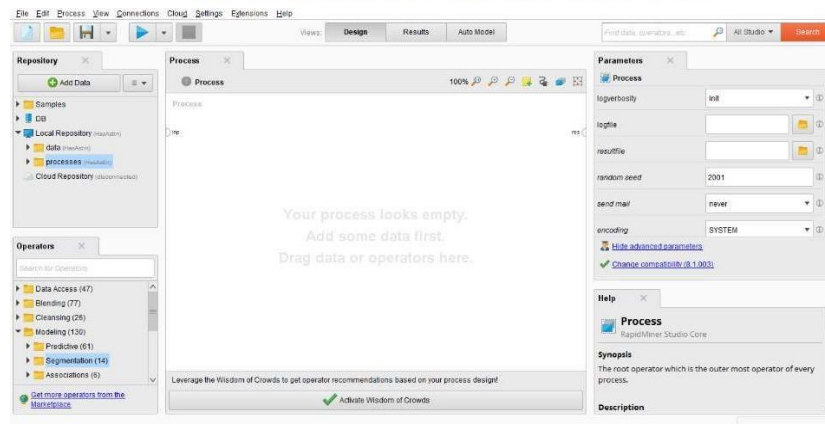
Penyakit	Jumlah
PENYAKIT KRONIS KECIL LAIN-LAIN	4615
INFEKSI SALURAN KEMIH AKUT	12
NEONATAL, BBL GROUP-5 DENGAN KONGENITAL/INFEKSI PERINATAL RINGAN	33
PENYAKIT AKUT BESAR LAIN-LAIN	91
ASTHMA & BRONKIOLITIS RINGAN	27
NEONATAL, BBL GROUP-3 TANPA PROSEDUR MAYOR RINGAN	20
INFEKSI NON BAKTERI RINGAN	290
GANGGUAN 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
FISILOGIKAL DAN BENGANGKATAN ALAT UKI HASIL INTERNAL SEDANG	2

Gambar 4. Hasil transformasi data

### Pengolahan Data

Pada tahapan ini dilakukan pemodelan data menggunakan hasil dari transformasi data seperti gambar 4, adapun metode yang dipakai pada penulisan ini adaah 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.

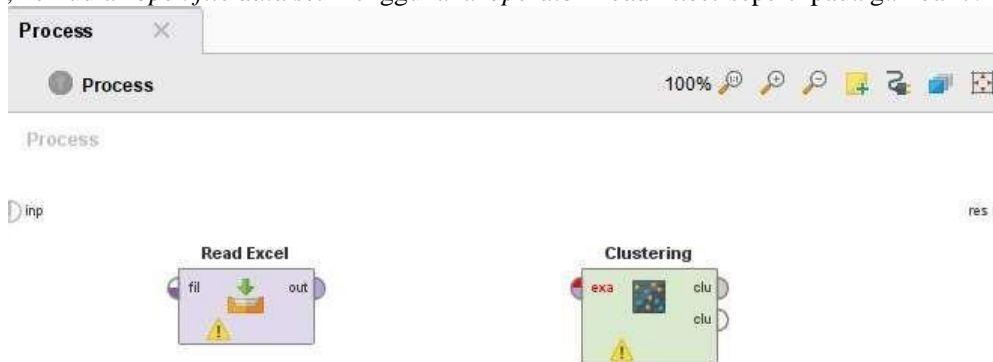
Pengujian Software RapidManer



Gambar 5. RapidMiner Studio explorer

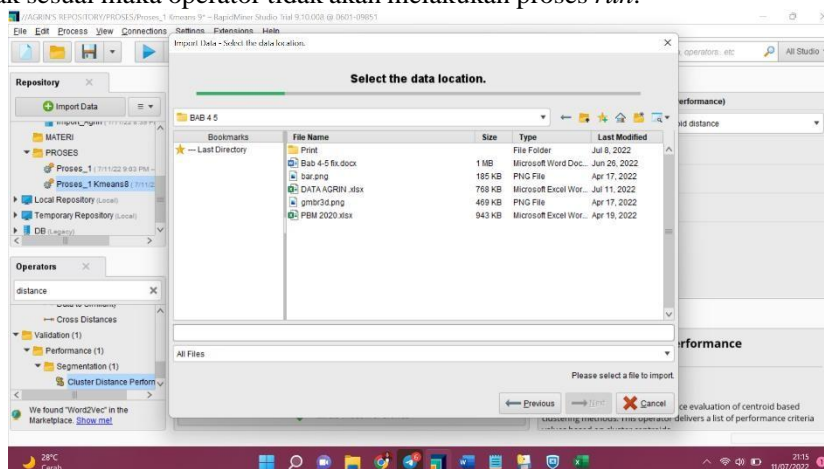
Atribut yang digunakan sebagai *class* adalah jumlah peserta JKN-KIS yang terdiagnosa penyakit. Data akan dianalisis berdasarkan jumlah peserta JKN-KIS yang paling banyak di antara peserta JKN-KIS lainnya. Data diagnosa penyakit dibuat dalam *data set* baru dengan format *.xlsx (excel)* untuk bisa dianalisis *software* yang dipakai adalah *RapidMiner Studio* seperti pada gambar 5.

Setelah menentukan operator yang akan digunakan lakukan *drag and drop operators* seperti gambar 6 dibawah, kemudian *open file data set* menggunakan operator *Read Excel* seperti pada gambar 7.

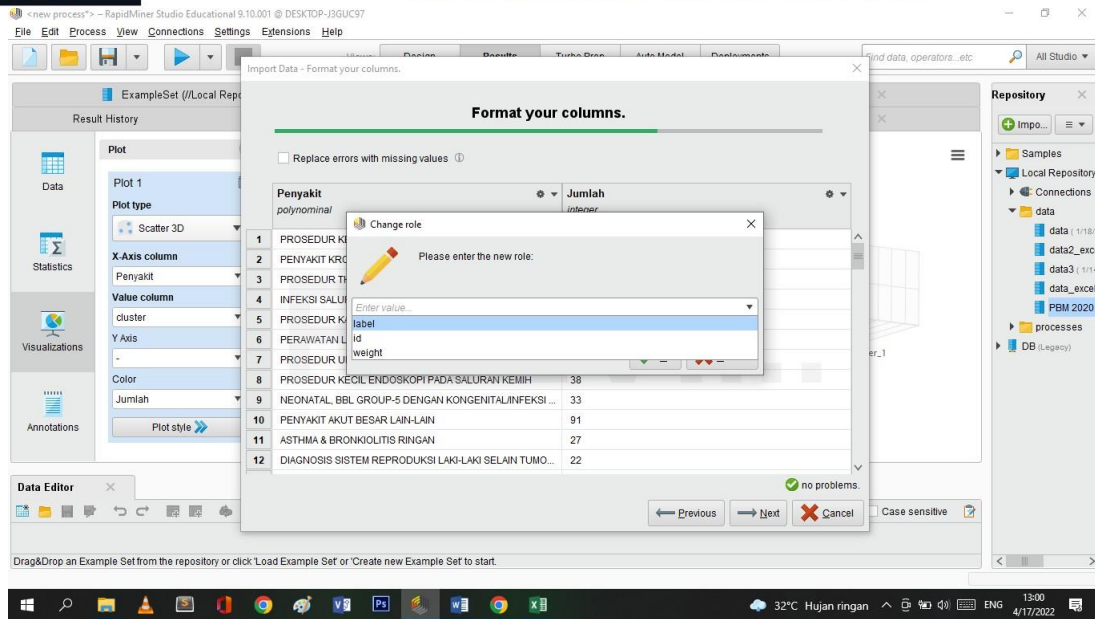


Gambar 6. Operator *read excel* dan *clustering*

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 7 untuk memastikan datanya sudah benar sesuai dengan type datanya, jika *type* data tidak sesuai maka operator tidak akan melakukan proses *run*.

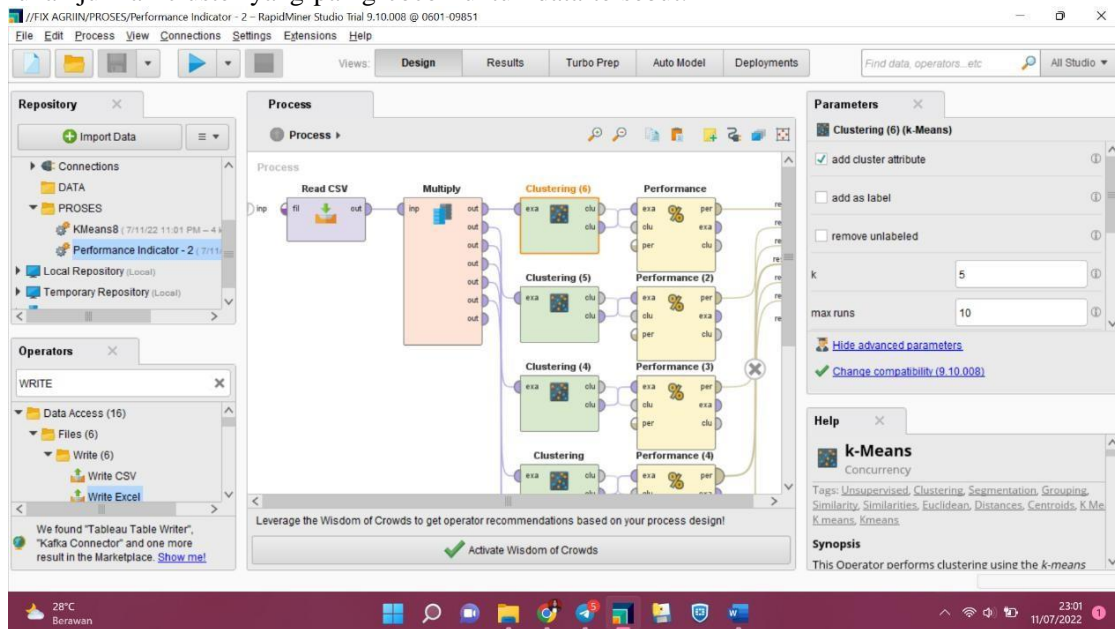


Gambar 7. *Open file excel*

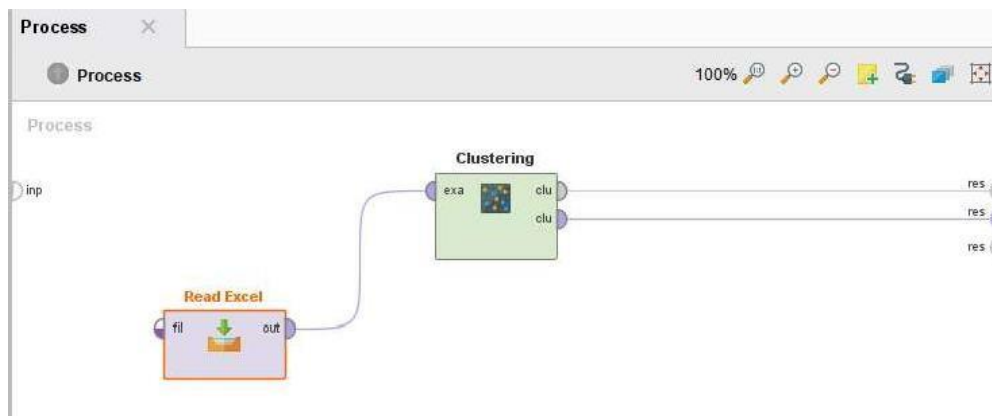


Gambar 8. Edit parameter list

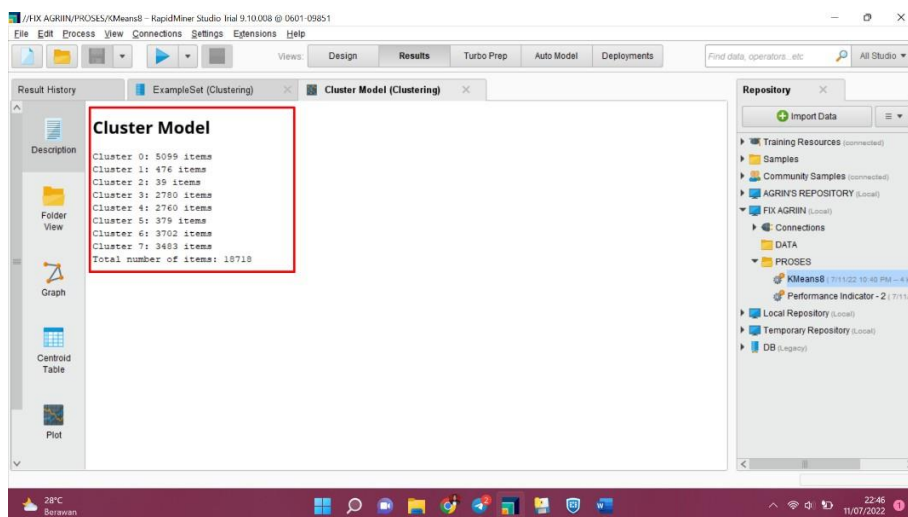
Pada dasarnya *rapidminer studio* dapat membaca type data setiap atribut seperti pada gambar 8, namun untuk mengurangi kesaaahan pada proses *run* adabainya dipastikan kembai. Setelah *open file excel* langkah selanjutnya adaah menghubungkan operator *read excel* dengan operator *clstering k-means* danmengatur melauai parameternya, sekaigus menentukan jumlah clusternya. Untuk menemukan jumlah yaitu 8 *cluster* seperti pada gambar 9. Setelah operator *read excel* dan operator *clustering k-means* diatur langkah berikutnya adaah menghubungkan kedua operator tersebut, dan hubungkan juga pada proses *out put* untuk mendapatkan hasil *run* operatornya, seperti gambar 10. Pada gambar 9 Penulis menguji dengan melaukan clustering secara 6 kai, untuk menemukan jumlah cluster yang paing cocok untuk data tersebut.



Gambar 9. Menentukan cluster

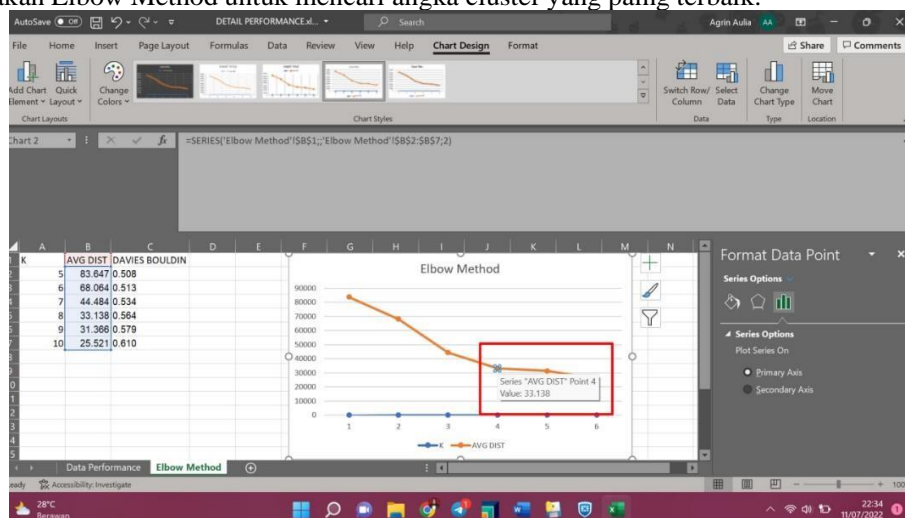


**Gambar 10** Menghubungkan operator

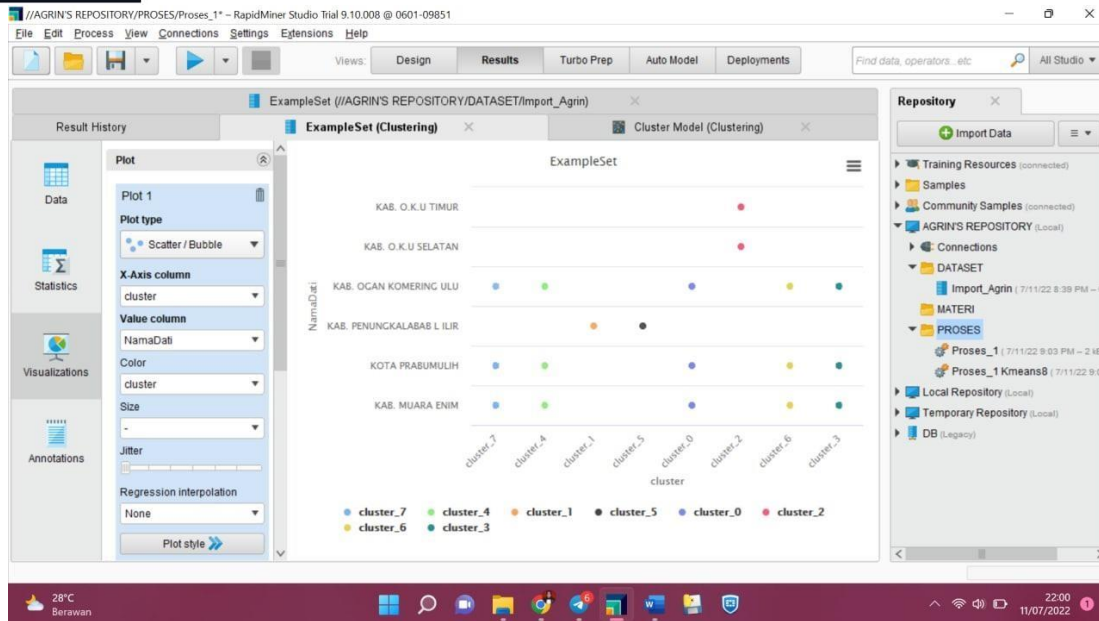


**Gambar 11.** Hasil cluster

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

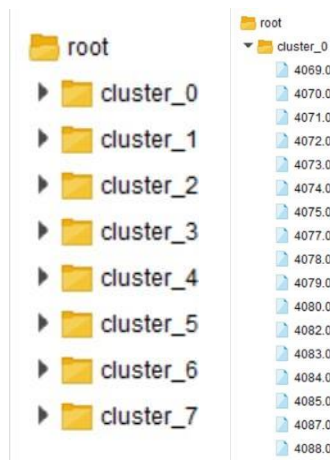


**Gambar 12.** Hasil data cluster K-Means daam implementasi RapidMiner



**Gambar 13.** Visualize data berdasarkan NamaDati Peserta JKN-KIS.

Diagnosa penyakit terdiri dari 200 diagnosa penyakit bisa dilihat pada gambar 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 daam menangani penyakit yang mana ini tidak teridentifikasi, sehingga keterangan tersebut tidak digunakan.



**Gambar 14.** Isi folder *cluster*

Data yang di *cluster* akan masuk pada folder yang dikelompokkan berdasarkan kesamaan karakter seperti pada gambar 14 ketika saah satu data tersebut dibuka maka akan menunjukkan detailnya.

### Agoritma K-Means

Daam menggunakan algoritma *k-means* akan melakukan pengulangan tahapan hingga terjadi kestabilan. Penulis melakukan enam kai iterasi daam 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*. Tota 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 daam 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* adaah 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 adaah 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.

## KESIMPULAN DAN SARAN

### KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan terhadap sistem pengelompokan data diagnosa penyakit, diperoleh kesimpulan sebagai berikut:

1. Hasil dari *clustering* topik skripsi pada pemodelan *rapidminer* menggunakan algoritma *k-means* menghasilkan banyak data yang masuk pada setiap *cluster* berbeda-beda.
2. Sistem pengelompokan data diagnosa penyakit yang dibangun menghasilkan 200 diagnosa penyakit yaitu penyakit kronis, Infeksi sauran kemih akut, Penyakit akut besar, Athma & bronkiolitis ringan, Infeksi non bakteri ringan, Hipertensi ringan, Penyakit akut kecil, Kegagalan jantung ringan, Peradangan kulit dan lain-lain di daerah JKN-KIS BPJS Kesehatan Kantor Cabang Prabumulih.
3. Pengujian nilai k terbaik dengan *Sum of Squared Errors* menghasilkan *error* terendah pada *cluster* ke-4 yaitu sebesar 3,624. Dengan demikian, sistem pengelompokan data rekam medis menghasilkan *cluster* terbaik dengan menggunakan 8 *cluster*.

### SARAN

Penulisan ini tidak lepas dari batasan yang ada, berikut adaah beberapa saran yang di berikan untuk pengembangan penulisan selanjutnya yang diharapkan mampumengembangkan penulisan ini yaitu:

1. Dapat menggunakan data yang jumlahnya lebih banyak.
2. Dapat berkoordinasi agar memiliki dan mendapatkan detail penyakit yang lebih lengkap.



## UCAPAN TERIMA KASIH

Puji syukur penulis ucapkan kepada Allah SWT. atas kuasa dan karunia-Nya, sehingga penulis dapat menyelesaikan penelitian ini. Penulis ucapkan terimakasih atas bantuan dan dukungannya kepada pihak-pihak yang telah membantu selama penelitian dilakukan.

## DAFTAR PUSTAKA

- Anggraeny, C. (2013). *Imovasi Pelayanan Kesehatan. 1*, 85–93.
- Ayu, R. D. V., & Ernawati, D. (2012). Tinjauan Penulisan Diagnosis Utama dan Ketepatan Kode ICD-10 pada Pasien Umum di RSUD Kota Semarang Triwulan I. *UDiNus Repository*, 14.
- Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866–883. <https://doi.org/10.1109/69.553155>
- Ediyanto, Mara, N., & Intisari, N. S. (2013). Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, 02(2), 133–136.
- Endartiwi, S. S., & Setianingrum, P. D. (2019). Health care quality has correlation with participant satisfaction of NHI in the primary health facilities in the Province of Yogyakarta. *Buletin Penelitian Sistem Kesehatan*, 22(3), 158–166.
- Geetha, A., & Nasira, G. (2014). Artificial neural networks' application in weatherforecasting – Using RapidMiner. *International Journal of Computational Intelligence and Informatics*, 4(3), 177–182. <http://www.periyaruniversity.ac.in/ijcii/issue/Vol4No3December2014/IJCII 4-1-152.pdf>
- Hayman, L. L., & Worel, J. N. (2014). Healthy lifestyle behaviors the importance of individual and population approaches. *Journal of Cardiovascular Nursing*, 29(6), 477–478. <https://doi.org/10.1097/JCN.0000000000000199>
- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6), 2321–7782.
- Ongko, E. (2014). Perancangan Sistem Pakar Diagnosa Penyakit Pada Balita. *Jurnal Time*, II(1), 1–5. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Parteek, B. (2019). *Data Mining and Data Warehousing*. Cambridge University Press.
- Wati, W., Machmud, R., & Yurniwati, Y. (2019). Analisis Kualitas Pelayanan Program Jaminan Kesehatan Nasional di Ruang Rawat Inap Rumah Sakit X Kabupaten Kerinci. *Jurnal Kesehatan Andalas*, 8(2), 366. <https://doi.org/10.25077/jka.v8i2.1014>
- Xu, J., & Lange, K. (2019). Power k-means clustering. *36th International Conference on Machine Learning, ICML 2019, 2019-June*, 11977–11991.
- Yudi Agusta. (2007). K-Means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem Dan Informatika*, 3(Februari), 47–60.
- Zheng, X., Lei, Q., Yao, R., Gong, Y., & Yin, Q. (2018). Image segmentation based on adaptive K-means algorithm. *Eurasip Journal on Image and Video Processing*, 2018(1). <https://doi.org/10.1186/s13640-018-0309-3>