

Implementasi Algoritma K-Means untuk Pengelompokan Data Imunisasi Balita dengan Metode CRISP-DM

¹Ririn Restu Aria
Universitas Bina Sarana Informatika
Jakarta, Indonesia

ririn.rra@bsi.ac.id

*Penulis Korespondensi

Diajukan : 29/12/2024

Diterima : 10/01/2025

Dipublikasi : 12/01/2025

ABSTRAK

Imunisasi adalah serangkaian vaksin yang harus diberikan kepada balita usia 0 – 59 bulan untuk mencegah penyakit dan kecacatan serta mampu membantu agar balita tidak mudah sakit. Terdapat beberapa imunisasi dasar yang akan diberikan kepada balita yaitu BCG, DPT, Polio, Campak – Rubela, MMR dan Hepatitis B. Pada penelitian ini dilakukan analisa *cluster* pada dataset yang diperoleh dari statistik kesejahteraan rakyat tahun 2024 yang dipublikasikan oleh Badan Pusat Statistik berjumlah 38 provinsi dengan metode CRISP-DM dan algoritma K-Means dengan menggunakan aplikasi Rapidminer sebagai aplikasi untuk proses eksekusi data yang akan digunakan. Dari hasil pemodelan yang telah dilakukan terdapat 4 *cluster* yang memiliki nilai yang berbeda dan anggota dari setiap *cluster* akan dikelompokkan sesuai dengan kriteria yang telah ditentukan. Untuk *cluster* 0 memiliki nilai rata – rata 57.368 dengan kriteria sedang memiliki 14 anggota, untuk *cluster* 1 memiliki nilai rata – rata 28.381 dengan kriteria rendah memiliki 2 anggota, untuk *cluster* 2 memiliki nilai rata – rata 43.781 dengan kriteria tinggi memiliki 21 anggota dan untuk *cluster* 4 memiliki nilai rata – rata 0.00 dengan kriteria sangat rendah memiliki 1 anggota. Berdasarkan pengelompokan yang telah dilakukan diharapkan dapat membantu pemerintah pusat terutama pemerintah daerah untuk lebih meningkatkan lagi pelayanan kesehatan dalam pemberian imunisasi secara lengkap bagi balita terutama untuk provinsi Papua pengunungan yang mungkin terkendala kondisi alam saat petugas kesehatan memberikan fasilitas kesehatan kepada para balita disana.

Kata Kunci: CRISP-DM, Imunisasi, K-Means, Rapidminer

I. PENDAHULUAN

Imunisasi merupakan salah satu upaya untuk meningkatkan kekebalan seseorang secara aktif terhadap suatu penyakit sehingga apabila terkena penyakit tertentu tidak akan sakit atau hanya mengalami sakit ringan (Kemenkes, 2024). Dalam proses pemberian imunisasi yang diberikan kepada balita mulai usia 0 – 59 bulan secara lengkap terdiri dari BCG, DPT, Polio, Campak-Rubela (MR), MMR dan Hepatitis B. Pada tingkat kota atau kabupaten dari sebuah provinsi dilakukan oleh rumah sakit daerah atau pusat kesehatan masyarakat dalam melaksanakan pelayanan kesehatan terhadap masyarakat semaksimal mungkin (Chusyairi & Saputra, 2019). Imunisasi mempunyai peranan penting dalam meningkatkan kualitas Kesehatan pada anak sehingga orang tua harus memperhatikan kebutuhan imunisasi anak berdasarkan usianya (Wibowo et al., 2020). Menurut data statistik kesejahteraan rakyat 2024 berdasarkan 39 provinsi di Indonesia memiliki 6 jenis imunisasi yang diberikan yang terdapat pada halaman website <https://bps.go.id> (BPS, 2024). Salah satu teknik analisa data yang digunakan untuk data imunisasi

balita adalah clustering atau pengelompokan. *Clustering* merupakan proses untuk pengelompokan elemen – elemen dalam *entitas* yang memiliki kesamaan (Nurhayah et al., 2024). Masalah yang terkait dengan pemberian imunisasi adalah keterbatasan fasilitas kesehatan dan tenaga medis dalam memberikan pelayanan kesehatan kepada balita. Oleh karena itu pemerintah di setiap daerah harus memastikan bahwa setiap balita mendapatkan pelayanan pemberian imunisasi dasar lengkap maka dalam penelitian ini dilakukan proses *clustering* imunisasi balita berdasarkan provinsi yang ada di Indonesia menjadi beberapa cluster yaitu cluster provinsi berdasarkan pemberian imunisasi tinggi, sedang, rendah dan sangat rendah dari 38 provinsi. Tujuan penelitian ini adalah memperoleh pemahaman mengenai hasil pengelompokan tingkat pemberian imunisasi dari berbagai provinsi di Indonesia dengan menerapkan teknik data mining dengan analisis clustering dengan menerapkan algoritma K-Means. Dengan demikian diharapkan hasil penelitian ini dilakukan untuk pengelompokan terhadap balita yang berusia dari 0 – 59 bulan menurut provinsi, klasifikasi desa, dan jenis imunisasi yang pernah didapatkan di tahun 2024. Pada proses penerapan data mining akan pemberian imunisasi terhadap balita akan menjadi lebih efektif dan efisien sehingga imunisasi bisa lebih ditingkatkan kepada provinsi dengan tingkat imunisasi rendah.

II. STUDI LITERATUR

2.1 Penelitian Terdahulu

Menurut (Abdullah et al., 2020) dalam penelitian yang dilakukan menggunakan data persentase bayi yang diimunisasi yang merupakan layanan dari puskesmas meliputi BCG, DPT/HBI-3, Polio 1-4 dan campak berdasarkan presentase bayi yang belum dikelompokkan dengan membandingkan dengan *output* metode clustering dasar k-means, hasil clustering akan diukur menggunakan metode silhouette index. Berdasarkan hasil evaluasi, diperoleh nilai silhouette index 0,798102756 untuk K-Means dan 0,789901932 untuk boost K-Means, dengan ini algoritma yang diusulkan memiliki kualitas hasil clustering minimal sama atau lebih baik dari *single clustering* k-means dengan jumlah iterasi yang lebih sedikit. Menurut (Tiyar & Samsudin, 2024) melakukan penelitian untuk memperoleh sasaran Imunisasi Dasar Lengkap (IDL), puskesmas (pusat kesehatan masyarakat) di kecamatan duren sawit perlu meningkatkan pelayanan dan pendataan imunisasi bagi bayi dan baduta (bayi dua tahun) karena terdapat kendala dalam melakukan proses pendataan dan pengelompokan data imunisasi pada bayi dan baduta, dikarenakan data yang terlalu banyak untuk diproses sehingga memakan waktu yang lama dalam pengelompokan puskesmas mana yang sudah mencapai target IDL dan yang belum mencapai target IDL, hal ini membuat proses menjadi tidak efektif dan efisien. Tujuh Puskesmas yang terletak di kecamatan Duren Sawit akan dibagi menjadi tiga kelompok, terdiri dari Puskesmas yang memperoleh sasaran imunisasi dengan status tinggi, sedang, dan rendah. Hasil penelitian ini mendapatkan *Clustering* sasaran imunisasi pada baduta terdapat enam puskesmas dengan predikat tinggi dan hanya satu puskesmas dengan predikat sedang. Dengan adanya sistem informasi klasifikasi data sasaran imunisasi bayi dan baduta pada puskesmas berbasis Web dapat mempermudah dan mempercepat pengolahan data dan pengelompokan data sehingga menjadi efektif dan efisien. Menurut (Manurung et al., 2021) memberikan data dari situs resmi Pemerintah diperoleh bahwa persentase anak umur dari 12 sampai 23 bulan yang menerima imunisasi tingkat kesejahteraan 40% terendah ditingkat provinsi di Indonesia. Penelitian dilakukan dengan pengelompokan terhadap minimnya tingkat pemberian imunisasi mulai dari tahun 2015 sampai 2018 dengan menggunakan metode K-means. Data-data yang dikelompokkan menjadi 2 Cluster mulai dari terendah dan tertinggi sehingga menyimpulkan masih ada minimnya pemberian imunisasi anak di Indonesia, sehingga ada diharapkan Pemerintah disetiap daerah di Indonesia agar lebih giat lagi untuk mensejahterahkan kesehatan anak. Penelitian selanjutnya dilakukan oleh Alia Fadilah, Mohammad Nurfaizy P, Supriyanto Lumbanbatu, dan Sofi Defiyanti (Fadilah et al., 2022) yang melakukan penelitian dengan hasil Susenas Maret 2019 dengan Studi Status Gizi Balita Indonesia (SSGBI) tahun 2019 menunjukkan kasus balita stunting di Indonesia adalah sebesar 27.7%, angka tersebut masih belum mencapai standar yang ditetapkan WHO yaitu sebesar 20%. Maka dibutuhkan suatu metode untuk mengelompokkan Kabupaten/Kota di

Indonesia berdasarkan faktor penyebab stunting pada balita, yaitu menggunakan metode *clustering* dengan algoritma K-Means. Tujuan penelitian dapat membantu pemerintah dalam proses mengambil kebijakan yang sesuai terkait penurunan prevalensi stunting pada balita berdasarkan karakteristik dan permasalahan masing-masing cluster. Hasil penelitian menunjukkan dengan bantuan metode elbow menghasilkan 2 cluster sebagai cluster terbaik dengan nilai selisih *Sum of Square Error* (SSE) sebesar 1401.5156, dimana *cluster* 1 merupakan *cluster* dengan faktor penyebab stunting tinggi yang terdiri dari 324 kabupaten/kota, dan *cluster* 2 merupakan *cluster* dengan faktor penyebab stunting rendah yang terdiri dari 49 kabupaten/kota.

2.2 K-Means

K-Means adalah algoritma *clustering* yang berulang – ulang untuk menetapkan nilai – nilai *cluster* (k) secara acak dimana nilai tersebut akan menjadi pusat dari *cluster* atau *centroid*, *mean*. Algoritma K-Means sangat mudah diaplikasikan dalam perhitungannya (Nasir, 2020)

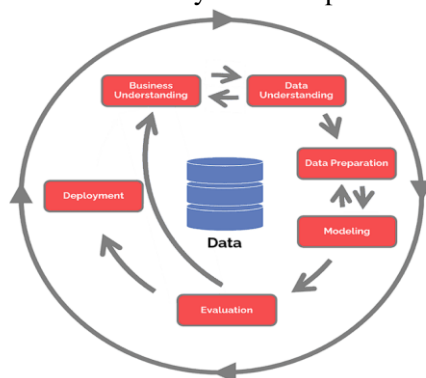
2.3 Rapidminer

Rapidminer adalah perangkat lunak yang digunakan untuk menganalisa data mandiri yang dapat diintegrasikan ke dalam berbagai bahasa pemrograman yang akan digunakan (Fauziah & Ramadhantya, 2024). RapidMiner menyediakan tampilan *user interface* yang ramah pengguna, sehingga memudahkan pengguna saat mengaplikasikannya. Tampilan yang terdapat pada RapidMiner disebut *Perspective* yaitu *Welcome Perspective*, *Design Perspective* dan *Result Perspective* (Prasetyo et al., 2021)

III. METODE

3.1 Metode Penelitian

Data yang digunakan dalam penelitian ini menggunakan data sekunder. Dataset yang diterapkan pada penelitian ini bersumber dari Badan Pusat Statistik yang berada dalam halaman website <https://www.bps.go.id/id/publication/2024/11/06/3ef10d3d82ed93f616ba9113/indikator-kesejahteraan-rakyat-2024.html>. Penelitian ini dilakukan dengan metode *Cross Industry Standard Process for Data Mining* untuk analisa data preesntase imunisasi balita berdasarkan provinsi. Langkah - langkah yang digunakan dalam menyelesaikan penelitian ini sebagai berikut :



Gambar 1. Tahapan CRISP-DM

Berikut adalah deskripsi dari penjelasan tahapan penelitian yang dilakukan yang terdapat pada gambar 1 diatas :

1. *Business Understanding*

Langkah yang digunakan untuk fokus terhadap pemahaman tujuan bisnis dan persyaratan suatu penelitian yang dilakukan.

2. *Data Understanding*

Langkah kedua ini melibatkan pengumpulan data dan familiarisasi dengan data yang digunakan dari sumber sekunder untuk dianalisa, memahami strukturnya, kualitas dan pola yang terdapat dalam data tersebut serta mampu mengidentifikasi masalah kualitas data. Contoh nilai data yang hilang atau nilai data yang tidak konsisten.

3. *Data Preparation*

Langkah ketiga akan dilakukan pembersihan dan transformasi data sehingga data siap untuk digunakan dalam pemodelan dengan memastikan apakah ada nilai dalam dataset yang

“kosong” atau “NaN”. Mengisi nilai pada data *missing values*, memilih fitur dan target serta membagi dataset menjadi data pelatihan dan pengujian.

4. Modeling

Langkah keempat ini akan digunakan untuk membangun model menggunakan Teknik data mining dengan memilih teknik pemodelan yang sesuai dengan masalah bisnis dan jenis data serta melatih model menggunakan data pelatihan.

5. Evaluation

Tahap ini akan melibatkan evaluasi model untuk memastikan model memenuhi tujuan bisnis dan persyaratan penelitian. Ada beberapa evaluasi model *clustering* yang bisa digunakan untuk evaluasi tersebut yaitu SilhouetteScore, Davies-BouldinIndex, Calinski-HarabaszIndex, AdjustedRandIndex, ScatterPlot(Visualisasi), NormalizedMutualInformation.

6. Deployment

Pada tahap ini, model data mining yang telah dikembangkan diimplementasikan ke dalam lingkungan produksi untuk digunakan dalam proses bisnis.

IV. HASIL DAN PEMBAHASAN

4.1 Business Understanding

Tujuan bisnis berdasarkan pengolahan data presentase imunisasi usia 0 – 59 bulan dari 38 provinsi adalah untuk dapat melakukan pengelompokan provinsi dalam kategori sangat rendah agar bisa meningkatkan pelayanan terhadap pemberian imunisasi dari provinsi tersebut untuk mencegah terjangkitnya suatu penyakit. Untuk dapat menerapkan tujuan yang telah ditentukan maka dibutuhkan suatu dataset yang bersumber dari Badan Pusat Statistika didalam statistika kesejahteraan rakyat 2024 menggunakan metode K-Means.

4.2 Data Understanding

Pada langkah ini data diperoleh dari statistik kesejahteraan rakyat 2024 yang dipublikasikan oleh Badan Pusat Statistik dengan jumlah 38 *record* dengan 7 *atribut* yaitu Provinsi, BCG, DPT, Polio, Campak-Rubela, MMR dan Hepatitis B berbentuk excel dan data perlu dilakukan proses *cleaning* dan pemilihan data dengan atributnya.

4.3 Data Preparation

Dalam penelitian ini untuk proses pengolahan data dilakukan dengan mengimplementasikan pada aplikasi Rapidminer dengan melakukan beberapa tahapan yaitu:

1. Data Selection

Dataset presentase penduduk berumur 0-59 Bulan (Balita) menurut provinsi, klasifikasi desa dan jenis imunisasi yang pernah didapatkan 2024 dengan banyak 38 data dengan 7 atribut dipilih secara keseluruhan karena semua data tersebut akan digunakan dalam pengolahan proses algoritma ini. Data hasil seleksi dapat dilihat pada tabel 1 dan gambar 2, selanjutnya peneliti akan melakukan analisa data imunisasi dari setiap provinsi dengan menggunakan excel.

Tabel 1. Hasil data *selection* Imunisasi Batita

Provinsi	BCG	DPT	Polio	Campak-Rubela	MMR	Hepatitis B
Aceh	57.06	48.44	61.31	41.44	34.19	53.13
Sumatera Utara	84.76	79.34	84.48	66.26	55.47	78.85
Sumatera Barat	85.38	76.40	83.72	61.21	50.45	81.29
Riau	84.81	78.89	83.40	68.57	57.32	78.65
Jambi	82.99	78.83	82.89	63.59	53.52	80.63
Sumatera Selatan	92.58	88.52	91.62	75.71	66.07	89.35
Bengkulu	93.63	90.50	92.96	78.06	69.36	93.54
Lampung	94.46	90.90	93.45	78.46	71.45	94.71
Kep.Bangka Belitung	89.32	85.00	89.16	75.06	70.71	90.92
Kep.Riau	94.13	89.52	95.06	75.19	66.68	93.21

DKI Jakarta	95.63	93.35	94.89	79.74	71.37	94.04
Jawa Barat	90.60	85.18	90.39	70.70	60.83	89.12
Jawa Tengah	95.09	92.00	95.74	77.54	74.15	95.96
DI Yogyakarta	96.47	94.95	95.53	77.86	76.80	98.12
Jawa Timur	93.46	88.65	93.51	74.47	70.55	92.55
Banten	88.71	83.04	87.99	69.32	59.85	85.32
Bali	97.24	95.28	97.10	82.93	76.44	97.52
Nusa Tenggara Barat	94.49	91.20	93.94	78.44	70.37	94.19
Nusa Tenggara Timur	95.02	92.57	95.29	80.97	73.50	94.82
Kalimantan Barat	87.49	81.35	86.74	67.58	59.58	85.19
Kalimantan Tengah	92.22	86.70	91.33	73.67	65.79	89.16
Kalimantan Selatan	91.32	86.58	90.55	75.07	69.14	91.25
Kalimantan Timur	93.42	90.46	93.15	74.73	66.01	93.71
Kalimantan Utara	94.14	89.98	93.10	77.32	69.66	91.28
Sulawesi Utara	94.99	90.74	93.94	76.94	70.31	93.30
Sulawesi Tengah	89.55	84.99	89.84	71.77	63.11	87.56
Sulawesi Selatan	92.29	88.60	92.46	76.16	67.57	91.60
Sulawesi Tenggara	91.18	86.52	90.36	73.61	63.06	89.11
Gorontalo	92.89	86.76	91.42	72.18	69.01	91.96
Sulawesi Barat	89.57	82.92	88.34	66.64	60.95	84.57
Maluku	85.50	79.63	85.03	69.87	61.67	81.82
Maluku Utara	86.12	82.07	86.53	70.68	60.02	80.15
Papua Barat	84.31	78.09	83.00	67.39	61.36	77.98
Papua Barat Daya	89.67	87.14	89.74	76.68	62.57	86.74
Papua	90.08	82.87	89.58	77.54	62.82	81.96
Papua Selatan	89.49	81.30	91.09	71.12	58.93	82.71
Papua Tengah	55.86	46.54	54.76	38.44	34.64	45.63
Papua Pengunungan	71.58	68.84	71.37	58.54	47.98	65.72

Select the cells to import.

Sheet: Sheet1 Cell range: A:H Select All Define header row: 1

	A	B	C	D	E	F	G
1	Provinsi	BCG	DPT	Polio	Campak-R...	MMR	Hepatitis B
2	Aceh	57.06	48.44	61.31	41.44	34.19	53.13
3	Sumatera U...	84.76	79.34	84.48	66.26	55.47	78.85
4	Sumatera B...	85.38	76.40	83.72	61.21	50.45	81.29
5	Riau	84.81	78.89	83.40	68.57	57.32	78.65
6	Jambi	82.99	78.83	82.89	63.59	53.52	80.63
7	Sumatera S...	92.58	88.52	91.62	75.71	66.07	89.35
8	Bengkulu	93.63	90.50	92.96	78.06	69.36	93.54
9	Lampung	94.46	90.90	93.45	78.46	71.45	94.71
10	Kep.Bangk...	89.32	85.00	89.16	75.06	70.71	90.92
11	Kep.Riau	94.13	89.52	95.06	75.19	66.68	93.21
12	DKI Jakarta	95.63	93.35	94.89	79.74	71.37	94.04
13	Jawa Barat	90.60	85.18	90.39	70.70	60.83	89.12

Gambar 2. Data Selection Imunisasi berdasarkan provinsi

2. Data Preprocessing

Tahap *preprocessing* biasanya akan dilakukan pembersihan data dengan mengecek apakah terdapat *missing value* maupun data inkonsisten. Dalam data yang digunakan sudah lengkap dan tidak terdapat *missing value* dan inkonsisten data.

The screenshot shows the 'Results' tab in Altair AI Studio. It displays a table of statistics for the dataset 'ExampleSet (/Local Repository/data/Data Imunisasi Berdasarkan Provinsi)'. The table lists attributes: Provinsi (Nominal), BCG, DPT, Polio, Campak-Rubela, MMR, and Hepatitis B (all Real). For each attribute, it shows the number of missing values (0 for all), minimum and maximum values, and the average value.

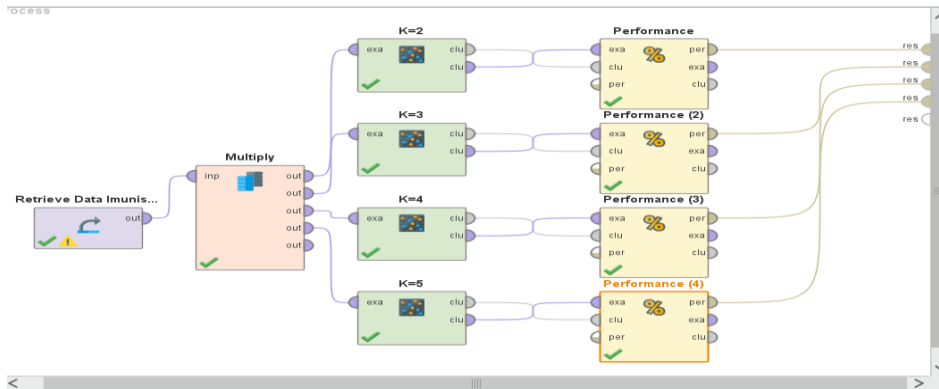
Name	Type	Missing	Statistics
Provinsi	Nominal	0	Level: Sumatera Utara (1), Aceh (1); Values: Aceh (1), Bali (1), ...[36 more]
BCG	Real	0	Min: 55.860, Max: 97.240, Average: 88.618
DPT	Real	0	Min: 46.540, Max: 95.280, Average: 83.806
Polio	Real	0	Min: 54.760, Max: 97.100, Average: 88.283
Campak-Rubela	Real	0	Min: 38.440, Max: 82.930, Average: 71.354
MMR	Real	0	Min: 34.190, Max: 76.800, Average: 63.244
Hepatitis B	Real	0	Min: 45.630, Max: 98.120, Average: 85.982

Gambar 3. Data Preprocessing

4.4 Modeling

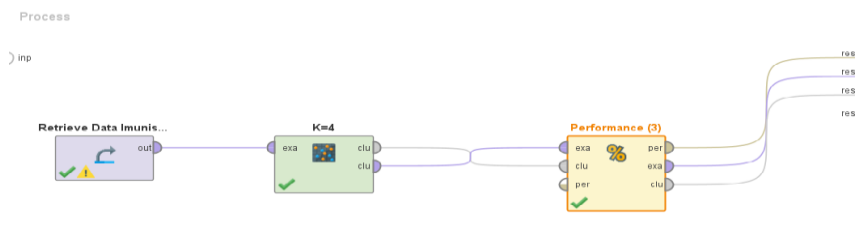
Tahapan dalam langkah ini akan dilakukan menganalisa dan mengeksekusi data dengan algoritma K-Means menggunakan aplikasi Rapidminer. Hal yang akan dilakukan dengan menentukan nilai dari *centroid* awal dari data yang dimiliki secara acak dan jumlah *cluster* yang akan sesuai dalam penelitian ini. Berikut adalah pemodelan dengan Rapidminer itu menentukan terlebih dahulu

jumlah *cluster* yang paling optimal.

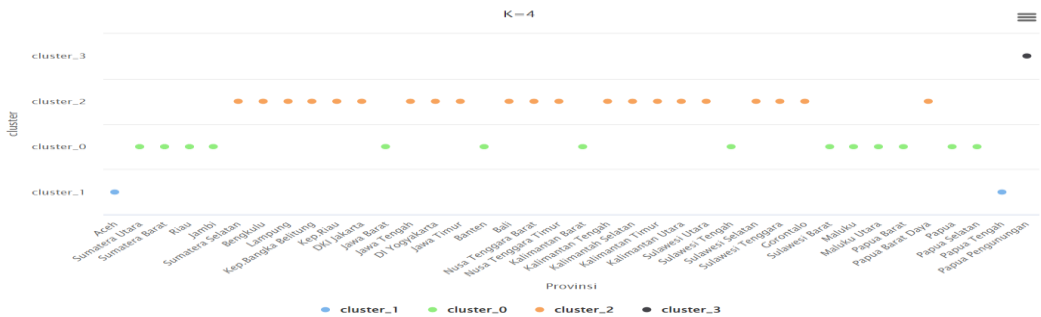


Gambar 4. Pemodelan *Clustering* dengan Rapidminer

Berdasarkan pemodelan yang dilakukan diatas, diperoleh bahwa jumlah *cluster* K = 4 adalah jumlah yang paling optimal. Untuk proses pemodelan digambarkan pada gambar 5 dibawah ini.



Gambar 5. Pemodelan K = 4 dengan Rapidminer



Gambar 6. Anggota dari masing – masing *Cluster*

4.5 Evaluation

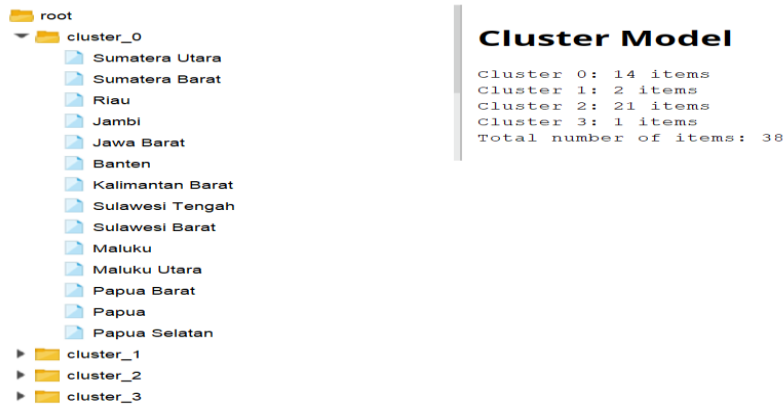
Untuk melakukan evaluasi terhadap *cluster* yang akan dipilih maka digunakan nilai dari David Bouldin (DBI) sebagai nilai sebuah perbandingan dari beberapa *cluster* yang dimodelkan. Berikut adalah nilai DBI dari beberapa *cluster* :

Tabel 2. Nilai dari DBI dari setiap jumlah *cluster*

Jumlah Cluster	Nilai DBI
2	-0.406
3	-0.480
4	-0.400
5	-0.531

Dari hasil perhitungan setiap *cluster* yang dilakukan, untuk dapat menentukan nilai yang terbaik dari beberapa *cluster* yang tersebut maka dipilih angka yang semakin kecil atau yang mendekati nol. Dari tabel 2 diatas dari 4 kategori *cluster* yang memiliki nilai DBI maka jumlah *cluster* yang

paling baik adalah *cluster* 4 dengan nilai DBI mendekati 0 yaitu -0.400. Oleh karena itu pengelompokan dilakukan dengan menggunakan 4 *cluster* dengan masing – masing anggota *cluster* dapat dilihat dari gambar 7.



Gambar 7. Hasil dari anggota masing-masing cluster

Dari hasil perhitungan yang telah dilakukan berdasarkan setiap *atribut* yang dimiliki dalam dataset terdapat nilai dari setiap *cluster* yang ada mulai dari *cluster* 0 sampai dengan *cluster* 3.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
BCG	87.097	56.460	93.507	71.580
DPT	81.064	47.490	89.806	68.840
Polio	86.644	58.035	93.062	71.370
Campak-Rubela	68.731	39.940	76.704	58.540
MMR	58.991	34.415	69.551	47.980
Hepatitis B	82.557	49.380	92.716	65.720

Gambar 8. Nilai *cluster* dari setiap atribut

V. KESIMPULAN

Kesimpulan dengan melakukan evaluasi terhadap nilai dari DBI yang mengoptimalkan jumlah *cluster* yang akan digunakan. *Clustering* ini memberikan pengelompokan berdasarkan provinsi dengan pemberian imunisasi untuk *cluster* 0 dengan kriteria sedang memiliki 14 anggota yaitu Provinsi Sumatera Utara, Sumatera Barat, Riau, Jambi, Jawa Barat, Banten, Kalimantan Barat, Sulawesi Tengah, Maluku, Maluku Utara, Papua Barat, Papua, Papua Selatan. Untuk *cluster* 1 memiliki kriteria rendah memiliki 2 anggota yaitu Provinsi Aceh, Papua Tengah. Untuk *cluster* 2 memiliki kriteria tinggi memiliki 21 anggota yaitu Provinsi Sumatera Selatan, Bengkulu, Lampung, Kep.Bangka Belitung, Kep.Riau, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo dan Papua Barat Daya dan untuk *cluster* 4 memiliki kriteria sangat rendah memiliki 1 anggota yaitu Provinsi Papua Peggunungan. Hasil penelitian ini, peneliti mengharapkan dapat digunakan pada penelitian-penelitian selanjutnya, dan dengan pembagian data menggunakan clustering ini diharapkan dapat membantu pemerintah untuk mengoptimalkan pemberian imunisasi kepada batita dari setiap provinsi yang ada di Indonesia. Peneliti menyadari bahwa hasil penelitian ini masih jauh dari kesempurnaan. Oleh karena itu, dalam penelitian selanjutnya, dapat dilakukan perbandingan dengan metode clustering yang berbeda.

VII. REFERENSI

Abdullah, A. I., Winarko, E., & Musdholifah, A. (2020). Metode Boost-K-means untuk Clustering Puskesmas berdasarkan Persentase Bayi yang Diimunisasi. *Jurnal Riset Sains Dan Teknologi*, 4(2), 89–96.

BPS. (2024). *Statistik Kesejahteraan Rakyat 2024*. <https://bps.go.id>

Chusyairi, A., & Saputra, P. R. N. (2019). Pengelompokan Data Puskesmas Banyuwangi Dalam

- Pemberian Imunisasi Menggunakan Metode K-Means Clustering. *Telematika*, 2(2), 139–148. <http://ejournal.amikompurwokerto.ac.id/index.php/telematika/>
- Fadilah, A., P. M. N., Lumbanbatu, S., & Defiyanti, S. (2022). PENGELOMPOKAN KABUPATEN/KOTA DI INDONESIA BERDASARKAN FAKTOR PENYEBAB STUNTING PADA BALITA MENGGUNAKAN ALGORITMA K-MEANS. *JIKO (Jurnal Informatika Dan Komputer)*, 6(2), 223–230.
- Fauziah, & Ramadhantya, A. S. (2024). Penggunaan Rapidminer Untuk Memprediksi Kelulusan Mahasiswa Dengan Algorithm Naive Bayes. *Jurnal Teknologi Informasi*, 10(1), 52–60.
- Kemendes. (2024). *Seputar Imunisasi*. <https://ayosehat.kemkes.go.id/1000-hari-pertama-kehidupan/seputar-imunisasi>
- Manurung, R. H., Lubis, M. R., & Saifullah. (2021). Penerapan Algoritma K-Means Dalam Pengelompokan Penerimaan Imunisasi Dasar Lengkap Menurut Provinsi. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, 2(1), 24–30.
- Nasir, J. (2020). PENERAPAN DATA MINING CLUSTERING DALAM MENGELOMPOKAN BUKU DENGAN METODE K-MEAN. *Jurnal SIMETRIS*, 11(2).
- Nurhayah, Suarna, N., & Prihartono, W. (2024). DATA MINING MENGGUNAKAN ALGORITMA K-MEANS ANALISIS JUMLAH PENDUDUK MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN KABUPATEN/KOTA DI INDONESIA. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 894–901.
- Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier. *JURNAL NASIONAL TEKNOLOGI DAN SISTEM INFORMASI*, 7(1), 8–17. <https://doi.org/https://doi.org/10.25077/TEKNOSI.v7i1.2021.8-17>
- Tiyar, M. N., & Samsudin. (2024). Klasifikasi Data Sasaran Imunisasi Bayi dan Baduta pada Puskesmas Berbasis Web Metode Clustering Algoritma K-Means. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 1(7), 143–154.
- Wibowo, C. A., Ashila, U. S., Aditya, I. G. Y., Probo, A., Widya, S. K., Rino, S. A., Rosaningrum, J., Ni Wayan Krisnayanti, Nurullia Tanjung, M. H., & Sulistyarini, A. (2020). PENGETAHUAN DAN SIKAP IBU TENTANG IMUNISASI DASAR PADA BALITA. *Jurnal Farmasi Komunitas*, 7(1), 17–22.