

GROUPING THE PREVALENCE OF DISEASE CASES BY AGE IN BANDUNG CITY HOSPITALS USING K-MEANS

Irfan Soliani¹, Safitri Juanita^{*2}

^{1,2}Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, DKI Jakarta, Indonesia
Email: ¹1812510418@student.budiluhur.ac.id, ²safitri.juanita@budiluhur.ac.id

(Naskah masuk: 29 Juni 2022, Revisi: 3 Juli 2022, diterbitkan: 26 Desember 2022)

Abstract

In 2019, the World Health Organization (WHO) stated that the top 10 types of diseases accounted for 55% of the 55.4 million deaths in the world. Meanwhile, in Indonesia, the province of West Java has the largest population, with the capital city of Bandung. Based on the health profile of the Bandung City Hospital, there were the ten highest diseases based on 18,147 cases. However, the data has not been processed into helpful information for the health department, especially the city of Bandung, to help determine disease cases by age group. So that the contribution of this study is to classify the prevalence of disease cases by age in Bandung City Hospital; this study aims to help the Bandung City Health Office take preventive, treatment and counselling actions against diseases that have a prevalence of disease cases based on age. This study uses the CRISP-DM methodology, with the K-Means clustering method and the testing method using the elbow method and the Davies-Bouldin Index (DBI). Data processing using rapid miner software and python programming. This study concludes that the optimal cluster value is K=2. The value of cluster 0 consists of the type of disease with the lowest case, and cluster 1 consists of the kind of disease with the highest case. Cluster 1 is the elderly and adult age group, while the age group in cluster 0 is the infant age group, the toddler age group, and the child age group.

Keywords: CRISP-DM, Davies-Bouldin Index, K-Means Algorithm, Prevalence of Disease.

PENGELOMPOKAN PREVALENSI KASUS PENYAKIT BERDASARKAN USIA DI RSUD KOTA BANDUNG MENGGUNAKAN K-MEANS

Abstrak

Pada tahun 2019, Organisasi Kesehatan Dunia (WHO) menyatakan bahwa 10 jenis penyakit teratas menyumbang 55% dari 55,4 juta kematian di dunia. Sementara itu, di Indonesia, provinsi Jawa Barat memiliki jumlah penduduk terbesar, dengan ibu kota Bandung. Berdasarkan profil kesehatan RSUD Kota Bandung, terdapat sepuluh penyakit tertinggi berdasarkan 18.147 kasus. Namun data tersebut belum diolah menjadi informasi yang bermanfaat bagi dinas kesehatan khususnya Kota Bandung untuk membantu menentukan kasus penyakit berdasarkan kelompok umur. Sehingga kontribusi penelitian ini adalah mengklasifikasikan prevalensi kasus penyakit menurut umur di RSUD Kota Bandung; penelitian ini bertujuan untuk membantu Dinas Kesehatan Kota Bandung melakukan tindakan preventif, pengobatan dan penyuluhan terhadap penyakit yang memiliki prevalensi kasus penyakit berdasarkan umur. Penelitian ini menggunakan metodologi CRISP-DM, dengan metode clustering K-Means dan metode pengujian menggunakan metode elbow dan Davies-Bouldin Index (DBI). Pengolahan data menggunakan software rapid miner dan pemrograman python. Penelitian ini menyimpulkan bahwa nilai cluster yang optimal adalah K=2. Nilai cluster 0 terdiri dari jenis penyakit dengan kasus terendah, dan cluster 1 terdiri dari jenis penyakit dengan kasus tertinggi. Klaster 1 adalah kelompok umur lanjut usia dan dewasa, sedangkan kelompok umur pada klaster 0 adalah kelompok umur bayi, kelompok umur balita, dan kelompok umur anak.

Kata kunci: Algoritma K-Means, CRISP-DM, Davies-Bouldin Index (DBI), Prevalensi Penyakit.

1. PENDAHULUAN

Pada tahun 2019, *World Health Organization* (WHO) menyebutkan bahwa terdapat 10 jenis penyakit teratas menyumbang penyebab kematian 55% dari 55,4 juta kematian yang ada di Dunia [1].

Di Indonesia, penyakit teratas berdasarkan informasi Kementerian Kesehatan Republik Indonesia (Kemenkes RI) didominasi oleh penyakit tidak menular, seperti Jantung, Kanker, Gagal ginjal dan Stroke yang sebagian besar dialami oleh usia dewasa dan usia manula [2][3]. Salah satu provinsi yang

memiliki penduduk terbanyak di Indonesia, yaitu provinsi Jawa Barat atau tepatnya di Kota Bandung, berdasarkan dari data profil kesehatan tahun 2019 [4] terdapat 21 kasus penyakit terbesar di Kota Bandung, beberapa penyakit tersebut diantaranya yaitu nasopharyngitis akut, ispa, hipertensi, pencernaan yang terganggu, mialgia, *laryngopharyngitis* akut, diare dan lainnya.

RSUD Kota Bandung melalui laman <http://data.bandung.go.id> mencatat 10 kasus penyakit tertinggi dengan total 18.146 records dari tahun 2016-2019. 10 penyakit tertinggi tersebut dialami oleh berbagai kelompok usia. Kumpulan data tersebut belum menjadi informasi yang berguna bagi pemerintahan kota Bandung khususnya bagi RSUD Kota Bandung untuk dapat mengetahui klaster jenis penyakit berdasarkan kelompok usia, sehingga kontribusi pada penelitian ini adalah melakukan proses pengelompokan kasus penyakit tertinggi berdasarkan usia berdasarkan data pada RSUD Kota Bandung menggunakan algoritma *K-Means*[5].

Algoritma *K-Means* merupakan salah satu metode dalam teknik klustering yang populer[6]. Algoritma ini sudah banyak digunakan pada penelitian sebelumnya, seperti yang dilakukan oleh Bastian et.al [7] pada penelitian penerapan algoritma *K-Means* terhadap penyakit menular manusia di kabupaten majalengka. Hasil penelitian tersebut didapatkan 6 klaster penyakit menular berdasarkan jumlah pasien di setiap puskesmas. Penelitian ini juga mengambil kesimpulan bahwa kelebihan algoritma *K-Means* yaitu mampu mengelompokkan objek dalam jumlah besar dengan pengelompokan yang cepat, dan kekurangan dari algoritma *K-Means* yaitu hasil pengelompokannya yang mudah berubah-ubah.

Penelitian lain yang menerapkan algoritma *K-Means* adalah penelitian yang dilakukan oleh Purba et.al [8]. Pada penelitiannya dilakukan penerapan *K-Means* terhadap penyebaran penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau. Dari penelitian tersebut didapatkan bahwa algoritma *K-Means* mampu mengelompokkan wilayah penyebaran ISPA di Provinsi Riau ke dalam dua klaster dengan klaster 1 merupakan wilayah klaster rekomendasi rendah dan klaster 2 merupakan wilayah klaster tinggi dengan tujuan dapat membantu Dinas Kesehatan Provinsi Riau untuk pencegahan, pengobatan dan juga penyuluhan.

Penelitian lainnya yang menggunakan algoritma *K-Means* adalah penelitian pengelompokan penyakit Demam berdarah dengue di Rumah Sakit Dharma Kerti yang dilakukan oleh Adiputra [9]. Penelitian tersebut bertujuan untuk mengetahui klaster data penyakit Demam berdarah dengue berdasarkan kelompok usia dari tahun 2014-2018. Dari hasil penelitian didapatkan ada 3 klaster kelompok usia dimana klaster 1 merupakan usia 0-

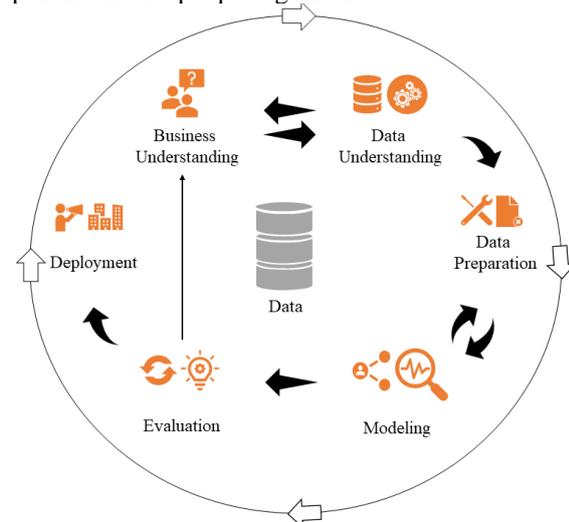
25 tahun, klaster 2 merupakan usia 26-45 tahun dan klaster 3 merupakan usia 46-80 tahun.

Penelitian dengan metode klustering *K-Means* juga pernah dilakukan oleh Saputri et.al [10] untuk mengelompokkan desa tervaksinasi covid-19 di kecamatan ujung pandang. Pada penelitian ini jumlah cluster ditentukan sebanyak 3 cluster. Hasil pengolahan data menunjukkan bahwa klaster C1 merupakan desa tervaksinasi tinggi, klaster C2 merupakan desa tervaksinasi sedang, dan klaster C3 merupakan desa tervaksinasi rendah, dimana klaster C1 terdiri dari 2 desa, klaster C2 terdiri dari 12 desa dan klaster C3 terdiri dari 6 desa. Dari hasil penelitian tersebut diharapkan dapat membantu untuk memudahkan pendataan, mempersingkat waktu dan dapat dijadikan acuan dalam program vaksinasi covid-19 di Indonesia.

Berbeda dari penelitian yang dikaji sebelumnya, penelitian ini melakukan klustering[11] menggunakan data penyakit tertinggi di RSUD Kota Bandung dengan berbagai kelompok usia yang kemudian diolah menggunakan algoritma *K-Means*. Pengolahan data menggunakan tahapan standard CRISP-DM (*Cross-Industry Standard Process for Data mining*) yang dilakukan dengan menggunakan perangkat lunak RapidMiner 9.10[12], Microsoft Excel dan Bahasa pemrograman python dengan library Scikit-Learn [13] dalam pengolahannya.

2. METODE PENELITIAN

Penelitian ini menggunakan metodologi penelitian CRISP-DM (*Cross-Industry Standard Process for Data mining*) [14]. 6 (enam) tahapan metodologi CRISP-DM yang digunakan pada penelitian terdapat pada gambar 6.



Gambar 1. Tahapan Proses Metode CRISP-DM

1. *Business Understanding*

Pemahaman bisnis pada penelitian ini berfokus terhadap rumusan masalah dari data 10 kasus penyakit tertinggi di RSUD Kota Bandung berdasarkan kelompok usia. Dari rumusan masalah, arah tujuan penelitian dapat ditentukan untuk

membantu menangani permasalahan yang ada dan mendeskripsikan manfaat yang dapat diberikan.

2. Data Understanding

Pada tahap ini mengumpulkan data yang diperlukan dalam penelitian yaitu data 10 kasus penyakit tertinggi di RSUD Kota Bandung. Kemudian mendeskripsikan pemahaman tiap atribut yang terdapat pada data tersebut.

3. Data Preparation

Pada tahap ini data mentah dari laman <http://data.bandung.go.id/> mengenai 10 Kasus Penyakit Tertinggi di RSUD Kota Bandung dilakukan proses pembersihan data untuk membentuk dataset akhir yang optimal agar dapat menghasilkan pola pemodelan yang bagus. Proses tersebut diantaranya integrasi data, reduksi data, pembersihan data, seleksi data, dan transformasi data.

4. Modeling

Pada tahap ini dilakukan 2 proses, yang pertama melakukan analisis jumlah kluster yang optimal menggunakan metode *Elbow*[15][16] dengan bahasa pemrograman Python menggunakan *library* Scikit-Learn dan yang kedua melakukan proses pemodelan dengan teknik klustering menggunakan algoritma *K-Means* dengan bantuan perangkat lunak RapidMiner 9.10.

5. Evaluation

Pada tahap ini dilakukan evaluasi untuk melihat tingkat akurasi dan performa dari hasil modeling yang dilakukan menggunakan *Davies-Bouldin Index* (DBI)[15][17].

6. Deployment

Pada tahap ini dilakukan dengan menyusun laporan dan membuat hasil visualisasi dari hasil pengolahan data yang telah dilakukan.

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Pemahaman masalah pada penelitian ini mengacu pada kasus jenis penyakit yang selalu ditemukan dengan jumlah yang tinggi di RSUD Kota Bandung setiap tahunnya. Dibutuhkan sosialisasi dan penyuluhan yang tepat untuk meminimalisir jumlah kasus penyakit yang tinggi. Pada penelitian ini analisis kasus penyakit tertinggi akan dilakukan untuk menghasilkan informasi yang dapat membantu Dinas Kesehatan melakukan sosialisasi dan penyuluhan secara efektif dan efisien.

3.2 Data Understanding

Data sekunder yang digunakan dalam penelitian ini diperoleh dari RSUD Kota Bandung yang dipublikasikan di Portal Data Kota Bandung

dengan situs <http://data.bandung.go.id/> mengenai 10 Kasus Penyakit Tertinggi di RSUD Kota Bandung Berdasarkan Kelompok Usia tahun 2016-2019. Data pada penelitian ini berisi atribut yang dapat dilihat pada tabel 1.

Tabel 1. Keterangan Atribut Dataset

Atribut	Tipe	Keterangan
Bulan	Nominal	Menunjukkan waktu bulan
Tahun	Numerik	Menunjukkan waktu tahun
Jenis Kamar	Nominal	Tempat dimana pasien mendapat pelayanan atau perawatan medis
Kode ICD	Nominal	Kode klasifikasi yang mewakili nama jenis penyakit berdasarkan diagnosisnya untuk memudahkan dalam dokumentasi rekam medis
Nama Penyakit	Nominal	Nama penyakit lengkap beserta diagnosis
Kelompok Usia Bayi	Numerik	Jumlah pasien bayi
Kelompok Usia Balita	Numerik	Jumlah pasien balita
Kelompok Usia Anak	Numerik	Jumlah pasien anak
Kelompok Usia Dewasa	Numerik	Jumlah pasien dewasa
Kelompok Usia Manula	Numerik	Jumlah pasien manula

3.3 Data Preparation

Pada tahap ini peneliti melakukan proses untuk membentuk dataset akhir yang terbaik sebelum dilakukan pemodelan, hal itu dilakukan agar pemodelan dapat menghasilkan pola atau model yang akurat. Terdapat 4 langkah yaitu:

1. Integration Data

Langkah awal pada data preparation yaitu *Integration Data*, yaitu menggabungkan dataset mentah 10 Kasus Penyakit Tertinggi di RSUD Kota Bandung dari tahun 2016 hingga tahun 2019 menjadi satu dataset dengan menggunakan perangkat lunak Microsoft Excel. Menggabungkan data dilakukan karena dataset 10 Kasus Penyakit Tertinggi di RSUD Kota Bandung dipublikasikan berupa laporan data per bulan di sumber situs web <http://data.bandung.go.id/> dan didapatkan jumlah total dataset 18.147 baris.

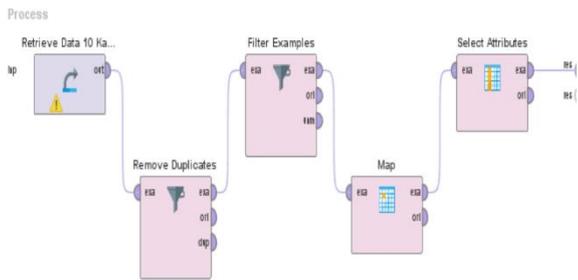
2. Data Cleaning

Langkah kedua yaitu Pembersihan data atau *Data Cleaning*. Proses ini dilakukan dengan menggunakan bantuan perangkat lunak RapidMiner 9.10. Beberapa fitur yang dipakai adalah: (a) menghilangkan duplikasi data dengan fitur *remove duplicates*. (b) Setelah menghilangkan duplikasi data, berikutnya adalah menghilangkan data *missing* dengan menggunakan fitur *filter examples*. (c) Kemudian yang terakhir adalah menghilangkan inkonsistensi data dengan menggunakan fitur *Map*.

3. *Selection Data*

Langkah selanjutnya melakukan *selection data* yaitu menentukan atribut yang digunakan. Pada proses ini masih menggunakan perangkat lunak RapidMiner 9.10 dengan fitur *Select Attributes*. Atribut yang digunakan pada penelitian ini yaitu Tahun, Kode ICD, Kelompok Usia Bayi, Kelompok Usia Balita, Kelompok Usia Anak, Kelompok Usia Dewasa dan Kelompok Usia Manula.

Proses langkah kedua dan ketiga terdapat pada gambar 2.



Gambar 2. Proses Data Preparation pada Dataset Penyakit Tertinggi di RSUD Bandung

4. *Data Reduction*

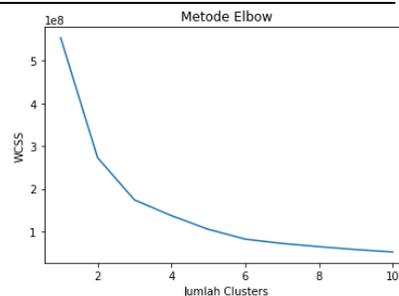
Kemudian langkah terakhir pada *data preparation* adalah melakukan reduksi data *records* dengan menggunakan perangkat lunak Microsoft Excel. Reduksi di sini bukan untuk menghilangkan data, tetapi menggabungkan jumlah data kasus masing-masing kelompok usia dari data per bulan menjadi data per tahun berdasarkan kode ICD. Proses dilakukan secara manual menggunakan fitur *Group by* dalam *Power Query Editor* yang ada di Microsoft Excel. Jumlah total dataset setelah dilakukan *data preparation* yaitu 1.443 baris dapat dilihat pada tabel 2.

Tabel 2. Dataset RSUD Bandung Setelah Data *Preparation*

No	Tahun	Kode ICD	Kelompok Usia				
			BY	BLT	ANK	DWS	MNL
1	2016	A16.9	541	1967	1456	8	0
2	2016	J39.9	201	525	271	1	1
3	2016	G40.9	56	446	881	3175	555
4	2016	A09	263	332	81	76	43
5	2016	J45.9	152	346	302	591	536
6	2016	P59.9	596	10	0	0	0
7	2016	Z24.0	292	3	1	0	0
8	2016	Z24.6	314	4	1	1	0
9	2016	J18.0	180	170	26	0	1
...
1441	2019	M62.6	0	0	0	5	1
1442	2019	H62.2	0	0	0	8	1
1443	2019	D34	0	0	0	3	1

3.4 *Modeling*

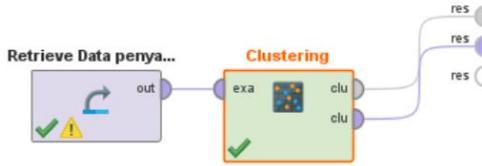
Teknik yang digunakan pada penelitian ini adalah teknik klastering dengan algoritma *K-Means*. Sebelum melakukan pemodelan, analisis menentukan jumlah klaster perlu dilakukan untuk mengetahui jumlah klaster yang optimal, hal tersebut dilakukan dengan pendekatan metode *Elbow*. Pada penelitian ini, proses pendekatan metode *Elbow* dilakukan dengan bahasa pemrograman python menggunakan *library Scikit-Learn* dan hasilnya dapat dilihat pada gambar 3.



Gambar 3. Hasil Metode *Elbow*

Pada gambar 3 terlihat hasil kurva metode *Elbow* yang membentuk siku berada pada angka 2 dan 3, tetapi karena perbedaan garis kurva yang ekstrim ada pada angka 2 maka jumlah klaster yang diambil adalah k=2. Setelah mendapatkan nilai k,

pemodelan dilakukan menggunakan RapidMiner 9.10 seperti terlihat pada gambar 4.



Gambar 4. Pemodelan *K-Means Clustering*

Hasil *Cluster Model* dari proses gambar 4 dapat dilihat pada gambar 5.

Cluster Model

Cluster 0: 1403 items
Cluster 1: 40 items
Total number of items: 1443

Gambar 5. Hasil *Cluster Model*

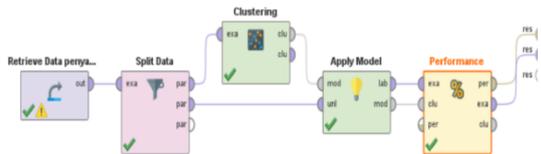
Pada gambar 5 dapat dilihat bahwa jumlah jenis penyakit pada klaster 0 terdiri dari 1.403 items dan jenis penyakit pada klaster 1 terdiri dari 40 items. Pemetaan hasil klaster model berdasarkan kode ICD dapat dilihat pada tabel 3.

Tabel 3. Pemetaan Hasil Klaster Model

No	Tahun	Kode ICD	Nama Penyakit	Cluster
1	2016	A16.9	Respiratory tuberculosis unspecified, without mention of bacteriological or histological confirmation	cluster_0
2	2016	J39.9	Disease of upper respiratory tract, unspecified	cluster_0
3	2016	G40.9	Epilepsy, unspecified	cluster_1
4	2016	A09	Diarrhoea and gastroenteritis of presumed infectious origin	cluster_0
...
99	2016	I69.4	Sequelae of stroke, not specified as haemorrhage or infarction	cluster_1
100	2016	M54.16	Radiculopathy, lumbar region	cluster_1
...
444	2017	E11.9	Non-insulin-dependent diabetes mellitus without complications	cluster_1
445	2017	I10	Essential (primary) hypertension	cluster_1
...
1441	2019	M62.6	Muscle strain	cluster_0
1442	2019	H62.2	Otitis externa in mycoses	cluster_0
1443	2019	D34	Benign neoplasm of thyroid gland	cluster_0

3.5 Evaluation

Setelah dilakukan pemodelan, tahap selanjutnya yaitu evaluasi menggunakan *Davies-Bouldin Index* (DBI) dengan data uji yang diambil dari data *training* menggunakan fitur *Split data* pada RapidMiner 9.10. Proses dapat dilihat pada gambar 6.



Gambar 6. Proses Evaluasi *K-Means Clustering* menggunakan *Davies-Bouldin Index* (DBI) pada Data Penyakit Kasus Tertinggi di RSUD Kota Bandung

Seperti terlihat pada gambar 6 fitur yang digunakan untuk evaluasi di RapidMiner 9.10 yaitu *Split Data* yang berfungsi untuk membuat data

ExampleSet dari data training dengan menggunakan *ratio*. Kemudian untuk uji performance dilakukan menggunakan fitur *Cluster Distance Performance*. Hasil evaluasi dapat dilihat pada tabel 4.

Tabel 4. Hasil Evaluasi menggunakan *Davies-Bouldin Index* (DBI) pada Data Penyakit Kasus Tertinggi di RSUD Kota Bandung

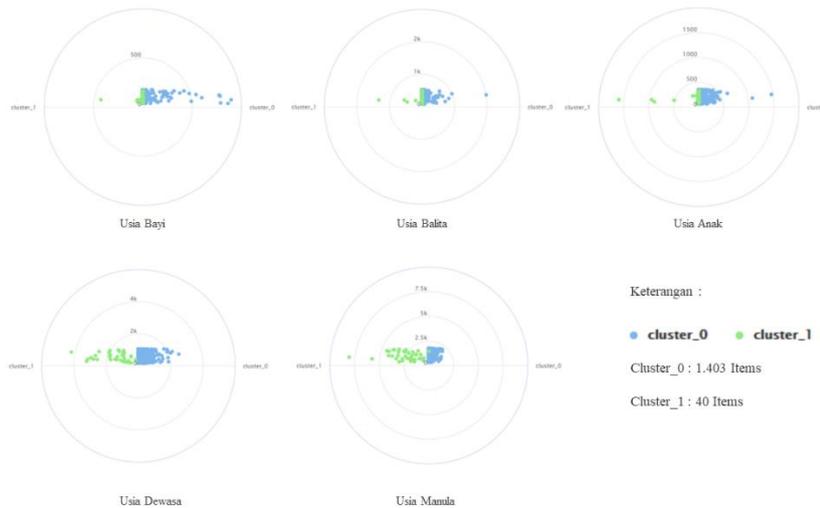
Nilai DBI			
Jumlah Cluster (k)	Ratio 0.6/0.4	Ratio 0.7/0.3	Ratio 0.8/0.2
2	-0,642	-0,642	-0,690
3	-0,840	-0,872	-0,819
4	-0,874	-0,896	-0,881
5	-0,999	-1,073	-1,015
6	-0,857	-0,883	-0,903
7	-0,898	-0,990	-0,903
8	-0,921	-1,001	-0,906
9	-0,988	-1,009	-0,870

10 -1,014 -1,040 -0,929

Pada Tabel 4 kita dapat melihat bahwa nilai *Davies-Bouldin Index* (DBI) yang paling rendah adalah -0,642 pada *ratio* 0.6/0.4 dan 0.7/0.3 dengan jumlah klaster $k=2$. Dapat disimpulkan bahwa jumlah klaster $k=2$ adalah jumlah yang paling optimal dan jumlah ini sama dengan hasil pendekatan metode *Elbow* yang telah dilakukan sebelumnya.

3.6 Deployment

Pada tahap ini hasil penelitian dibuatkan laporan dalam bentuk deskripsi serta dalam bentuk visualisasi data agar informasi mudah dipahami dan dipresentasikan. Pada gambar 7 adalah bentuk visualisasi hasil pengelompokan prevalensi penyakit berdasarkan usia di RSUD Kota Bandung menggunakan K-Means, yang ditampilkan menggunakan *Polar Chart Scatter*.



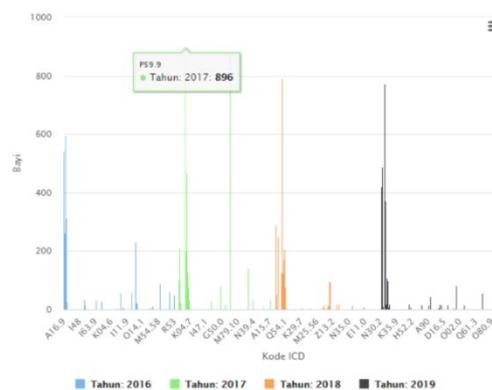
Gambar 7. Visualisasi *Polar Chart Scatter* Hasil Pengelompokan Klaster Prevalensi Kasus Penyakit Berdasarkan Kelompok Usia di RSUD Kota Bandung menggunakan Algoritma K-Means

Pada gambar 7 terlihat visualisasi *Polar Chart Scatter* menunjukkan bahwa kelompok usia bayi, usia balita dan usia anak dominan ke klaster 0 sedangkan pada kelompok usia dewasa dan usia manula dominan ke klaster 1. Pembahasan lebih detil hasil visualisasi data terdapat pada sub-bab diskusi.

4. DISKUSI

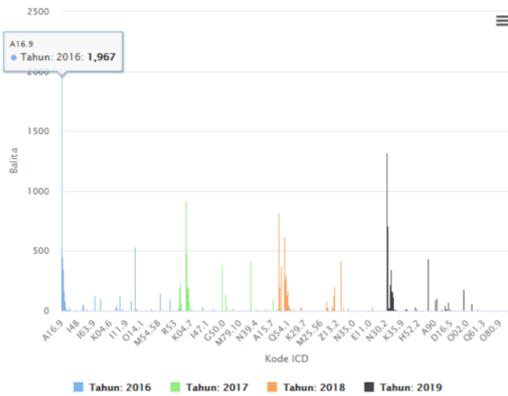
Berdasarkan hasil *Cluster Model* yang telah dilakukan pada gambar 7, hasil eksperimen juga menunjukkan jenis penyakit tertinggi berdasarkan kelompok usia.

Pada kelompok usia bayi, jenis penyakit yang memiliki jumlah kasus tertinggi terdapat di tahun 2017 dengan total jumlah 896 kasus, nama penyakit tersebut adalah *Neonatal jaundice* atau penyakit kuning dengan kode ICD P59.9 seperti terlihat pada gambar 8.



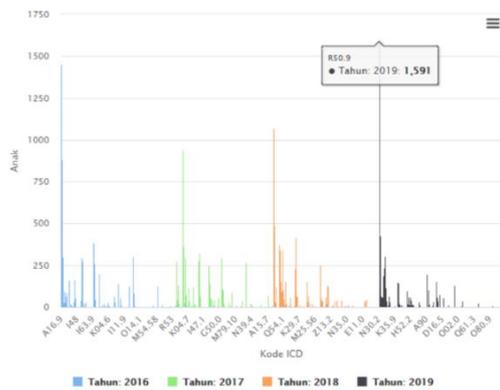
Gambar 8. Tampilan Jenis Penyakit dengan Kasus Tertinggi Pada Kelompok Usia Bayi Berdasarkan Data pada RSUD Bandung

Kemudian pada kelompok usia balita, jenis penyakit yang paling tinggi jumlah kasusnya ada di tahun 2016 dengan total jumlah 1.967 kasus, nama penyakit tersebut adalah Tuberkulosis pernapasan tidak ditentukan, tanpa menyebutkan konfirmasi *bakteriologis* atau *histologis* dengan kode ICD A16.9 seperti terlihat pada gambar 9.



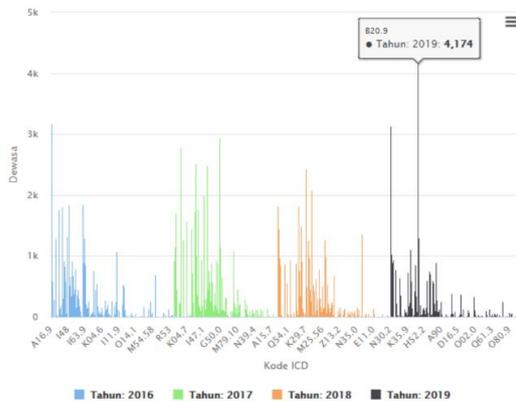
Gambar 9. Tampilan Jenis Penyakit dengan Kasus Tertinggi Pada Kelompok Usia Balita Berdasarkan Data pada RSUD Bandung

Sedangkan pada kelompok usia anak, jenis penyakit paling tinggi jumlahnya ada di tahun 2019 dengan total jumlah 1.591 kasus, nama penyakit tersebut adalah Demam, tidak ditentukan dengan kode ICD R50.9 seperti terlihat pada gambar 10.



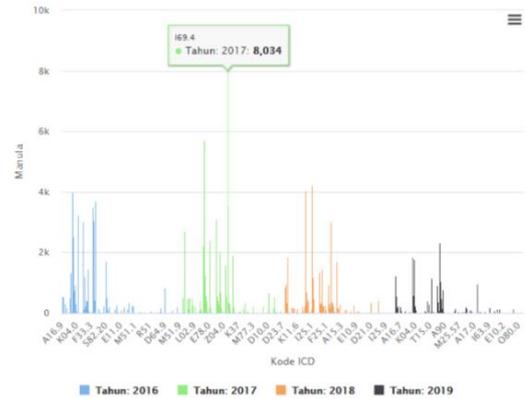
Gambar 10. Tampilan Jenis Penyakit dengan Kasus Tertinggi Pada Kelompok Usia Anak Berdasarkan Data pada RSUD Bandung

Selanjutnya pada kelompok usia dewasa, jenis penyakit paling tinggi jumlahnya ada di tahun 2019 dengan total jumlah 4.174 kasus, nama penyakit tersebut adalah HIV yang mengakibatkan penyakit infeksi atau parasit yang tidak spesifik dengan kode ICD B20.9 seperti terlihat pada gambar 11.



Gambar 11. Tampilan Jenis Penyakit dengan Kasus Tertinggi Pada Kelompok Usia Dewasa Berdasarkan Data pada RSUD Bandung

Dan terakhir pada kelompok usia manula, jenis penyakit yang paling tinggi jumlahnya ada di tahun 2017 dengan total jumlah 8.034 kasus, nama penyakit tersebut adalah Gejala sisa stroke (*Sequelae of stroke*), tidak ditetapkan sebagai pendarahan atau infark dengan kode ICD I69.4 seperti terlihat pada gambar 12.



Gambar 12. Tampilan Jenis Penyakit dengan Kasus Tertinggi Pada Kelompok Usia Manula Berdasarkan Data pada RSUD Bandung

Pada tabel 5, menampilkan ringkasan mengenai penyakit tertinggi pada masing-masing kelompok usia.

Tabel 5. Ringkasan Hasil Pengelompokan Prevalensi Kasus Penyakit Tertinggi Berdasarkan Usia di RSUD Kota Bandung

Kelompok Usia	Kode ICD	Nama Penyakit	Tahun
Bayi	P59.9	Penyakit Kuning	2017
Balita	A16.9	Tuberkulosis Pernapasan	2016
Anak	R50.9	Demam	2019
Dewasa	B20.9	HIV	2019
Manula	I69.4	Gejala sisa stroke	2017
Semua Usia	I69.4	Gejala sisa stroke	2017

5. KESIMPULAN

Berdasarkan hasil eksperimen, maka dapat disimpulkan bahwa berdasarkan metode *Elbow* dan pengujian *Davies-Bouldin Index (DBI)*, jenis penyakit tertinggi di RSUD Kota Bandung terbagi menjadi 2 kluster yaitu kluster 0 terdiri dari jenis penyakit dengan jumlah kasus rendah dan kluster 1 terdiri dari jenis penyakit dengan jumlah kasus tinggi. Hasil *Cluster Model* pada penelitian ini menunjukkan bahwa kluster 0 terdiri dari 1.403 items dan kluster 1 terdiri dari 40 items.

Kelompok usia yang rentan terkena jenis penyakit tinggi atau kluster 1 adalah kelompok usia manula dan usia dewasa, sedangkan kelompok usia yang rentan terkena jenis penyakit rendah atau kluster 0 adalah kelompok usia bayi, usia balita, dan usia anak, namun tidak menutup kemungkinan untuk usia dewasa dan manula terkena jenis penyakit kluster 0.

Dari hasil penelitian yang telah dilakukan, diharapkan penelitian ini dapat membantu Dinas Kesehatan Kota Bandung untuk melakukan

sosialisasi dan penyuluhan khususnya terhadap jenis penyakit dengan jumlah kasus tinggi, sesuai dengan kelompok usia.

DAFTAR PUSTAKA

- [1] World Health Organization, "The top 10 causes of death," *www.who.int*, 2020. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (accessed Jun. 20, 2022).
- [2] Bambang Purwanto, "Masalah dan Tantangan Kesehatan Indonesia Saat Ini," <https://kesmas.kemkes.go.id/>, 2022. <https://kesmas.kemkes.go.id/konten/133/0/masalah-dan-tantangan-kesehatan-indonesia-saat-ini> (accessed Jun. 20, 2022).
- [3] E. Rahmawaty, S. Handayani, M. H. N. Sari, and I. Rahmawati, "Sosialisasi Dan Harmonisasi Gerakan Masyarakat Hidup Sehat (Germas) Dan Program Indonesia Sehat Dengan Pendekatan Keluarga (Pis-Pk) Di Kota Sukabumi," *Link*, vol. 15, no. 1, p. 27, 2019, doi: 10.31983/link.v15i1.4385.
- [4] Dinas Kesehatan Kota Bandung, *PROFIL KESEHATAN KOTA BANDUNG TAHUN 2019*. 2019. [Online]. Available: <https://dinkes.bandung.go.id/wp-content/uploads/2020/09/Profil-Kesehatan-Kota-Bandung-Tahun-2019.pdf>
- [5] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J Multidiscip. Sci. J.*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.
- [6] K. P. Sinaga and M. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 1–12, 2020.
- [7] A. Bastian, H. Sujadi, and G. Febrianto, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *J. Sist. Inf.*, vol. 14, no. 1, pp. 26–32, 2018.
- [8] N. Purba, P. Poningsih, and H. S. Tambunan, "Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau," *J. Inf. Syst. Res.*, vol. 2, no. 3, pp. 220–226, 2021, [Online]. Available: <http://ejurnal.seminar-id.com/index.php/josh/article/view/736>
- [9] I. N. M. Adiputra, "Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means," *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 2, no. 2, p. 99, 2022, doi: 10.23887/insert.v2i2.41673.
- [10] D. S. Saputri, G. M. Putra, and M. F. Larasati, "Implementation of the K-Means Clustering Algorithm for the Covid-19 Vaccinated Village in the Ujung Padang Sub-District Implementasi Algoritma K-Means Clustering Untuk Desa Tervaksinasi Covid-19 Pada Kecamatan Ujung Padang," *J. Tek. Inform.*, vol. 3, no. 2, pp. 261–267, 2022.
- [11] A. S. Osman, "Data mining techniques: Review," *Int. J. Data Sci. Res.*, vol. 2, no. 1, pp. 1–4, 2019.
- [12] L. Kovács and H. Ghous, "Efficiency comparison of Python and RapidMiner," *Multidiscip. Tudományok*, vol. 10, no. 3, pp. 212–220, 2020, doi: 10.35925/j.multi.2020.3.26.
- [13] E. Ramadhanta Machmud Razaq, J. Deden Witarsyah, and F. Hamami, "Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran Online Selama Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Perbandingan Algoritma Klasifikasi," in *e-Proceeding of Engineering*, 2021, vol. 8, no. 5, pp. 9000–9006.
- [14] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [15] R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering optimization in RFM analysis based on k-means," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 1, pp. 470–477, 2019, doi: 10.11591/ijeecs.v18.i1.pp470-477.
- [16] E. Umargono, J. E. Suseno, and V. G. S. K., "K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median," in *Proceedings of the International Conferences on Information System and Technology*, 2020, no. Conrist 2019, pp. 234–240. doi: 10.5220/0009908402340240.
- [17] A. K. Singh, S. Mittal, P. Malhotra, and Y. V. Srivastava, "Clustering Evaluation by Davies-Bouldin Index(DBI) in Cereal data using K-Means," in *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, 2020, no. Iccmc, pp. 306–310. doi: 10.1109/ICCMC48092.2020.ICCMC-00057.