

IMPLEMENTASI ALGORITMA *K-MEANS CLUSTERING* UNTUK PENGELOMPOKAN PENYEBARAN *PNEUMONIA* PADA BALITA DI KOTA BANDUNG

Nelia Vitalaya¹, Rizki Tri Prasetyo²

¹Universitas Adhirajasa Reswara Sanjaya
Jl. Sekolah Internasional No. 1-2 Antapani, Bandung, 022-7100124
e-mail: neliavitalaya778@gmail.com

²Universitas Adhirajasa Reswara Sanjaya
Jl. Sekolah Internasional No. 1-2 Antapani, Bandung, 022-7100124
e-mail: rizki@ars.ac.id

Abstrak

Pneumonia adalah bentuk infeksi pernapasan akut yang menyerang paru-paru. *Pneumonia* saat ini masih tercatat sebagai masalah kesehatan utama pada anak di negara berkembang. Penyakit *Pneumonia* merupakan penyebab utama kematian balita baik di Indonesia maupun di dunia, namun tidak banyak perhatian terhadap penyakit ini, karena gejalanya yang hampir mirip dengan flu biasa. Kota Bandung merupakan salah satu kota penyumbang *Pneumonia* tertinggi pada balita di Jawa Barat dengan jumlah temuan kasus sebanyak 10.735 kasus pada tahun 2017. Tujuan dilakukannya penelitian ini adalah untuk mengelompokkan wilayah penyebaran penyakit pneumonia pada balita di Kota Bandung, sehingga dinas kesehatan kota Bandung dapat melakukan pencegahan dan penanggulangan penyebaran penyakit pneumonia pada balita dengan efektif. Pada penelitian ini penulis menggunakan algoritma *K-Means clustering* untuk mengelompokkan wilayah penyebaran penyakit *pneumonia* pada balita di Kota Bandung dari tahun 2015-2018. Hasil dari penelitian ini dibentuk 3 *cluster* yang telah di evaluasi menggunakan *Davies Bouldin Index* dengan nilai sebesar 0,571. Hasil *clustering* dari jumlah keseluruhan 30 Kecamatan didapatkan *cluster* tinggi, sedang dan rendah, yaitu *cluster* 0 (C0) rendah dengan jumlah 19 Kecamatan, *cluster* 1 (C1) sedang dengan jumlah 9 Kecamatan dan *cluster* 2 (C2) tinggi dengan jumlah 2 Kecamatan, dengan didominasi oleh penderita laki-laki.

Kata Kunci: *Pneumonia*, Balita, *Clustering*, *K-Means*

Abstract

Pneumonia is a form of acute respiratory infection that attacks the lungs. *Pneumonia* is currently still listed as a major health problem in children in developing countries. *Pneumonia* is the main cause of under-five mortality both in Indonesia and in the world, but there is not much attention to this disease, because the symptoms are almost similar to the common cold. Bandung City is one of the cities with the highest pneumonia contributor to children under five in West Java with 10,735 cases in 2017. The purpose of this research is to classify the area of the spread of pneumonia in children under five in the city of Bandung, so that the city health office of Bandung can effectively prevent and control the spread of pneumonia in children under five. In this study the authors used the *K-Means clustering* algorithm to classify the area of the spread of pneumonia among toddlers in Bandung City from 2015-2018. The results of this study formed 3 clusters which were evaluated using the *Davies Bouldin Index* with a value of 0.571. The results of clustering from a total of 30 Districts obtained high, medium and low clusters, namely cluster 0 (C0) low with 19 Districts, cluster 1 (C1) medium with 9 Districts and cluster 2 (C2) high with 2 Districts, with dominated by male sufferers.

Keywords: *Pneumonia*, Toddlers, *Clustering*, *K-Means*

1. Pendahuluan

Kesehatan merupakan hal yang terpenting bagi kehidupan manusia, semua orang tentu ingin memiliki kehidupan yang sehat, tetapi terkadang seseorang dapat mengalami gangguan kesehatan seperti *Pneumonia*. *Pneumonia* adalah bentuk infeksi pernapasan akut yang menyerang paru-paru. Paru-paru terdiri dari kantung-kantung kecil yang disebut *alveoli*, ketika bernafas *alveoli* akan terisi udara, tapi ketika seseorang mengidap *pneumonia*, *alveoli* dipenuhi nanah dan cairan, sehingga ketika bernafas akan terasa menyakitkan dan membatasi asupan oksigen (WHO, 2019).

Pneumonia saat ini masih tercatat sebagai masalah kesehatan utama pada anak di negara berkembang. Penyakit *pneumonia* merupakan penyebab utama kematian balita baik di Indonesia maupun di dunia, namun tidak banyak perhatian terhadap penyakit ini, karena gejalanya yang hampir mirip dengan flu biasa. Oleh sebab itu penyakit ini sering disebut sebagai pembunuh balita yang terlupakan (*The Forgotten Killer of Children*) (Rahayu, 2016). *Pneumonia* merupakan pembunuh utama anak dibawah usia lima tahun (Balita), dibandingkan dengan penyakit lain seperti AIDS, Malaria dan Campak (KeMenKes, 2017a). Menurut hasil analisis yang dilakukan *United Nations Children's Fund* (UNICEF) tahun 2018 menunjukkan bahwa, *pneumonia* merenggut nyawa anak balita lebih dari 800.000 di seluruh dunia, atau 39 anak per detik (UNICEF, 2019).

Menurut data Profil Kesehatan Indonesia 2017 Kementerian Kesehatan Republik Indonesia penemuan jumlah kasus penyakit *pneumonia* pada balita diperkirakan sebanyak 511.434 kasus, provinsi Jawa Barat merupakan provinsi tertinggi dalam penemuan kasus *pneumonia* dengan jumlah sebanyak 145.135 kasus (KeMenKes, 2017b). Kota Bandung merupakan salah satu kota penyumbang *pneumonia* tertinggi pada balita di Jawa Barat dengan jumlah temuan kasus sebanyak 10.735 kasus (Dinas Kesehatan Provinsi Jawa Barat, 2017).

Sehingga diperlukan pemetaan serta pengelompokan penyebaran penderita penyakit *pneumonia* pada balita di Kota Bandung. Karena merupakan salah satu kota yang memiliki angka penderita penyakit *pneumonia* pada balita yang cukup tinggi. Pengelompokan penyebaran penderita

penyakit *pneumonia* dapat dilakukan dengan mengolah data penderita *pneumonia* pada balita di Kota Bandung, selama ini terkadang data tersebut hanya digunakan sebagai bahan untuk menampilkan kondisi yang telah terjadi. Padahal sebenarnya ada potensi tersembunyi yang bisa ditemukan, yaitu bagaimana data tersebut bisa diolah untuk mendukung penyelesaian kasus (Toresa, 2020). Dengan adanya dukungan perkembangan teknologi, semakin berkembang pula kemampuan dalam mengumpulkan dan mengolah data. *Data mining* dapat digunakan untuk mencari informasi dari data tersebut sehingga didapatkan informasi yang bermanfaat (Purnia & Warnilah, 2017).

Salah satu teknik analisa *data mining* adalah *clustering*, *clustering* merupakan salah satu metode dalam *data mining* yang digunakan untuk mencari dan mengelompokkan data yang memiliki kesamaan karakteristik dengan lainnya metode *clustering* yang ada pada saat ini sangat banyak, salah satunya adalah *K-Means* (Wardhani, 2016).

Algoritma *K-Means* merupakan salah satu metode *clustering nonhierarki* yang mempartisi data ke dalam bentuk satu atau lebih *cluster*, sehingga data dengan karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama juga. Data dengan karakteristik yang berbeda dikelompokkan dalam kelompok yang lain. algoritma yang termasuk dalam metode *clustering* berbasis jarak yang membagi data kedalam sejumlah *cluster* serta algoritma ini hanya bekerja pada atribut *numeric* (Talakua, Leleury, & Talluta, 2017).

Pada penelitian kali ini penulis melakukan pengelompokan penyebaran penyakit *pneumonia* pada balita menggunakan algoritma *K-Means* didasarkan dari beberapa penelitian terdahulu, penelitian pertama adalah penelitian yang dilakukan oleh (Nasari & Sianturi, 2016) dengan judul " Penerapan Algoritma *K-Means Clustering* Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat ". Tujuan dilakukannya penelitian ini untuk mengelompokkan kawasan penyebaran diare dan mengetahui pusat penyebarannya. Hasil Penelitian menunjukkan bahwa algoritma *K-Means* dapat diimplementasikan dengan baik.

2. Metode Penelitian

2.1. Desain Penelitian

Penelitian yang digunakan adalah penelitian eksperimen. Penelitian eksperimen melibatkan penyelidikan perlakuan parameter/variabel tergantung pada penelitiannya dan menggunakan tes yang dikendalikan oleh peneliti itu sendiri. Pada penelitian eksperimen ini, digunakan model CRISP-DM (*Cross-Standard Industry Process for Data Mining*) yang terdiri dari 6 fase atau tahapan .

2.2. Fase Pemahaman Bisnis

Penyakit *Pneumonia* merupakan penyakit yang menyerang paru-paru, penyakit ini menjadi masalah kesehatan utama bagi balita di Indonesia maupun di dunia. Di Kota Bandung penyakit ini masih memiliki angka yang cukup tinggi sehingga di butuhkan suatu teknik pengelompokan daerah penyebaran penyakit *pneumonia* pada balita di Kota Bandung, agar pemerintah daerah khususnya Dinas Kesehatan Kota Bandung dapat meningkatkan pencegahan dan penanggulangan penyakit pneumonia dengan efektif, berdasarkan tingkat penyebaran daerah tinggi, sedang dan rendah.

2.3. Fase Pemahaman Data

Data penderita *pneumonia* pada balita diperoleh dari situs <http://data.bandung.go.id>. Data terdiri dari jumlah penderita *pneumonia* pada balita di 30 Kecamatan di Kota Bandung dari tahun 2015 sampai 2018. Dari data penderita *pneumonia* terdapat atribut Kecamatan, puskesmas dan jumlah penderita *pneumonia* laki-laki dan perempuan.

2.4. Fase Pengolahan Data

Data yang telah diperoleh akan diolah terlebih dahulu agar dapat dilakukan proses *clustering*. Pada tahap ini dilakukan tahap data *integration* dan data *cleaning*, proses pengolahan data tersebut dilakukan pada *Microsoft Excel*.

A. Integrasi Data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Dalam penelitian ini data penderita pneumonia pada balita dari tahun 2015 sampai 2018 digabungkan ke dalam satu *database*.

B. Pembersihan Data (*Data Cleaning*)

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang akan diolah. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, menghapus atau

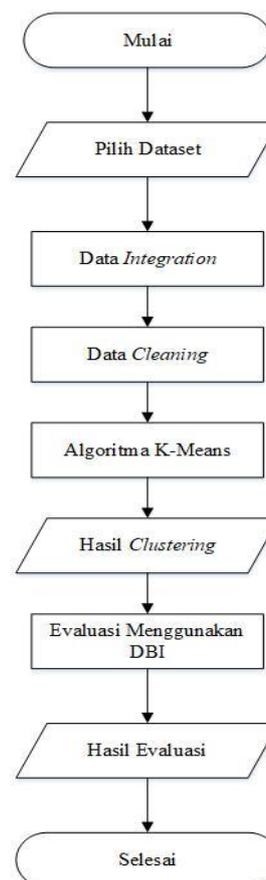
memperbaiki kesalahan pada data, seperti kesalahan penulisan. Pada tahap ini dilakukan pemilihan atribut yang tidak digunakan, maka atribut yang tidak digunakan yaitu UPT Puskesmas akan dihilangkan.

C. Populasi

Dalam penelitian ini, penulis mengambil populasi data dari situs <http://data.bandung.go.id>, dengan data sebanyak 30 Kecamatan di Kota Bandung. Data tersebut berisi jumlah penderita pneumonia pada balita laki-laki dan perempuan di setiap kecamatan yang ada di Kota Bandung.

2.5. Fase Pemodelan (*Modeling Phase*)

Pada tahap ini, akan dilakukan pemodelan menggunakan Algoritma *K-Means Clustering*. Berikut *flowchart* model yang di usulkan pada Gambar 1.



Gambar 1. Model Yang Diusulkan

2.6. Fase Evaluasi (*Evaluation Phase*)

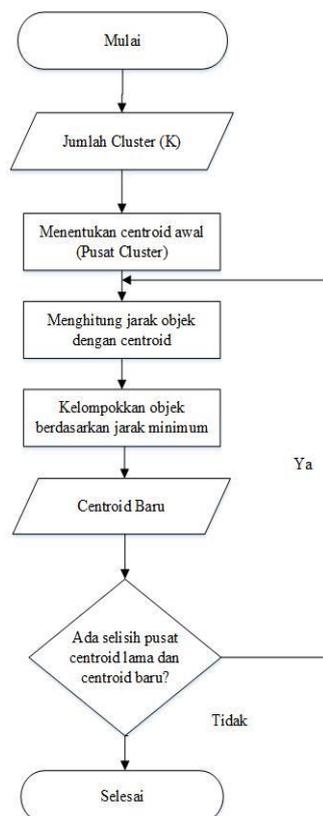
Dalam tahap ini akan dilakukan tahap evaluasi cluster menggunakan model *davies-bouldin index* (DBI). Pengukuran dengan *Davies-Bouldin Index* ini

memaksimalkan jarak *inter-cluster* dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah *cluster*. Semakin kecil nilai DBI yang diperoleh atau semakin mendekati nilai 0, maka semakin baik *cluster* yang diperoleh dari pengelompokan algoritma yang digunakan.

3. Hasil dan Pembahasan

3.1. Perhitungan Manual Algoritma K-Means

Dalam tahap ini akan dijelaskan langkah-langkah pengoperasian algoritma K-Means. Berikut merupakan Flowchart Algoritma K-Means:



Gambar 2. Flowchart algoritma K-means

Berikut merupakan tahapan perhitungan manual algoritma K-Means:

1. Tentukan Jumlah Cluster K

Tahap pertama dalam algoritma K-Means adalah menentukan jumlah *cluster*. Pada penelitian ini dibentuk 3 *cluster*.

2. Tentukan centroid awal (Pusat cluster)

Penentuan *centroid* awal dipilih secara *random* atau acak dari data yang telah diolah. Berikut *centroid* awal yang telah dipilih:

Tabel 1. Centroid Awal

Centroid Awal								
C0	226	118	212	112	184	108	340	204
C1	413	403	161	214	209	236	151	206
C2	430	358	399	341	254	208	346	363

3. Menghitung Jarak Objek Dengan Centroid

Setelah menentukan *centroid* awal, maka tahap selanjutnya adalah menghitung jarak setiap data ke *centroid* terdekat menggunakan rumus *Euclidean distance* untuk menentukan *cluster* yang diikuti oleh data tersebut. Persamaan yang digunakan adalah persamaan 2.1.

Sebagai contoh berikut akan dilakukan beberapa perhitungan:

Data pertama ke pusat *cluster* 0

$$\begin{aligned}
 d(x_1, c_0) &= \sqrt{(323 - 226)^2 + (318 - 118)^2 + (189 - 212)^2 + (185 - 112)^2 + (280 - 184)^2 + (219 - 108)^2 + (311 - 340)^2 + (252 - 204)^2} \\
 &= \sqrt{(97)^2 + (200)^2 + (-23)^2 + (73)^2 + (96)^2 + (111)^2 + (-29)^2 + (48)^2} \\
 &= \sqrt{9409 + 40000 + 529 + 5329 + 9216 + 12,321 + 841 + 2304} \\
 &= \sqrt{79949} \\
 &= 282,753
 \end{aligned}$$

Dari hasil perhitungan di atas didapat hasil bahwa jarak data pertama dengan *cluster* 0 adalah 282,753

Data pertama ke pusat *cluster* 1

$$\begin{aligned}
 d(x_1, c_1) &= \sqrt{(323 - 413)^2 + (318 - 403)^2 + (189 - 161)^2 + (185 - 214)^2 + (280 - 209)^2 + (219 - 236)^2 + (311 - 151)^2 + (252 - 206)^2} \\
 &= \sqrt{(-90)^2 + (-85)^2 + (28)^2 + (-29)^2 + (71)^2 + (-17)^2 + (160)^2 + (46)^2} \\
 &= \sqrt{8100 + 7225 + 784 + 841 + 5041 + 289 + 25600 + 2116} \\
 &= \sqrt{49,996} \\
 &= 223,598
 \end{aligned}$$

Dari hasil perhitungan di atas didapat hasil bahwa jarak data pertama dengan *cluster* 0 adalah 223,598

Data pertama ke pusat *cluster* 2

$$\begin{aligned}
 d(x_1, c_2) &= \sqrt{(323 - 430)^2 + (318 - 358)^2 + (189 - 399)^2 + (185 - 341)^2 +} \\
 &\quad \sqrt{(280 - 254)^2 + (219 - 208)^2 + (311 - 346)^2 + (252 - 363)^2} \\
 &= \sqrt{(-107)^2 + (-40)^2 + (-210)^2 + (-156)^2 + (26)^2 + (11)^2 + (-35)^2 + (-111)^2} \\
 &= \sqrt{11449 + 1600 + 44100 + 24336 + 676 + 121 + 1225 + 12321} \\
 &= \sqrt{95,828} \\
 &= 309,560
 \end{aligned}$$

Dari hasil perhitungan di atas didapat hasil bahwa jarak data pertama dengan *cluster* 0 adalah 309,560.

Kemudian lakukan perhitungan yang sama sampai data terakhir. Berikut perhitungan lengkapnya pada tabel 2 :

Tabel 2.
Hasil Perhitungan Pada Iterasi-1

No	Kecamatan	C0	C1	C2	Jarak Terdekat
1	Sukasari	282,753	223,598	309,561	223,598
2	Sukajadi	321,347	536,635	722,211	321,347
3	Cicendo	1444,552	1233,837	1049,716	1049,716
4	Andir	1449,062	1243,690	998,942	998,942
5	Cidadap	365,366	503,344	680,985	365,366
....
29	Buahbahu	423,132	597,870	817,152	423,132
30	Bandung Kidul	679,663	463,809	315,314	315,314

4. Kelompokkan Objek Berdasarkan Jarak Minimum

Setelah menghitung jarak tiap *cluster*, kemudian cari data yang memiliki jarak terdekat dengan *centroid*. Selanjutnya kelompokkan data sesuai dengan *cluster* terdekat.

Tabel 3.
Jarak Terdekat Hasil Iterasi-1

No	Kecamatan	C0	C1	C2	Jarak Terdekat	Cluster
1	Sukasari	282,753	223,598	309,561	223,598	C1
2	Sukajadi	321,347	536,635	722,211	321,347	C0
3	Cicendo	1444,552	1233,837	1049,716	1049,716	C2
4	Andir	1449,062	1243,690	998,942	998,942	C2
5	Cidadap	365,366	503,344	680,985	365,366	C0
....
29	Buahbahu	423,132	597,870	817,152	423,132	C0
30	Bandung Kidul	679,663	463,809	315,314	315,314	C2

Tabel 4.
Anggota *Cluster* 0 Pada Iterasi 1

Cluster 0								
No.	2015 Laki-laki	2015 Perempuan	2016 Laki-laki	2016 Perempuan	2017 Laki-laki	2017 Perempuan	2018 Laki-laki	2018 Perempuan
2	70	79	94	65	130	85	117	114
5	86	103	188	171	151	114	51	46
7	75	67	24	27	17	16	32	43
8	169	173	119	128	76	69	149	128
9	97	114	87	111	51	36	57	58
12	234	153	146	133	212	207	170	126
13	191	194	138	125	85	78	89	92
14	145	124	167	216	163	128	104	109
16	182	177	72	59	69	76	130	142
17	226	118	212	112	184	108	340	204
20	63	57	45	43	54	41	90	67
21	125	117	98	88	73	46	76	79
22	152	132	116	101	97	87	115	117
23	175	135	127	120	75	67	106	77
24	101	114	105	116	100	101	83	69
26	188	195	95	115	151	93	107	83
27	260	223	101	102	117	83	98	77
28	236	224	185	139	58	38	51	29
29	79	70	48	35	61	51	59	46
Jumlah Anggota: 19								
Rata-rata	150,211	135,211	114,053	105,579	101,263	80,211	106,526	89,789

Berikut merupakan anggota pada *cluster* 1:

Tabel 5.
Anggota *Cluster* 1 Pada Iterasi 1

Cluster 1								
No.	2015 Laki-laki	2015 Perempuan	2016 Laki-laki	2016 Perempuan	2017 Laki-laki	2017 Perempuan	2018 Laki-laki	2018 Perempuan
1	323	318	189	185	280	219	311	252
15	413	403	161	214	209	236	151	206
25	404	371	281	193	180	154	140	123
Jumlah Anggota: 3								
Rata-rata	380,000	364,000	210,333	197,333	223,000	203,000	200,667	193,667

Berikut merupakan anggota pada *cluster* 2:

Tabel 6.
Anggota *Cluster* 2 Pada Iterasi 1

Cluster 2								
No.	2015 Laki-laki	2015 Perempuan	2016 Laki-laki	2016 Perempuan	2017 Laki-laki	2017 Perempuan	2018 Laki-laki	2018 Perempuan
3	819	729	711	581	882	701	488	363
4	993	835	781	654	488	425	633	522
6	564	476	397	326	288	228	245	180
10	430	358	399	341	254	208	346	363
11	499	443	415	320	166	179	297	363
18	393	331	347	271	356	310	252	190
19	324	277	469	359	361	303	432	480
30	489	415	417	395	430	430	263	315
Jumlah Anggota: 8								
Rata-rata	563,875	483,000	492,000	405,875	403,125	348,000	369,500	332,000

Rata-rata yang didapat dari 3 *cluster* di atas adalah *centroid* baru yang akan digunakan untuk menghitung jarak terdekat pada iterasi ke-2. Berikut merupakan rata-

rata dari setiap *cluster* yang terdapat pada iterasi ke-1.

Tabel 7.
Centroid Baru Hasil Iterasi 1

Centroid Baru								
C0	150,211	135,211	114,053	105,579	101,263	80,211	106,526	89,789
C1	380,000	364,000	210,333	197,333	223,000	203,000	200,667	193,667
C2	563,875	483,000	492,000	405,875	403,125	348,000	369,500	332,000

Setelah didapatkan titik pusat dari setiap *cluster*, lakukan kembali dari langkah di atas hingga titik pusat *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain. Pada penelitian ini percobaan perhitungan berhenti pada iterasi ke 5. Berikut merupakan hasil perhitungan pada iterasi ke-5:

Tabel 8.
Hasil Perhitungan Pada Iterasi ke-5

No	Kecamatan	C0	C1	C2	Jarak Terdekat	Cluster yang diikuti
1	Sukasari	441,049	226,129	1198,875	226,129	C1
2	Sukajadi	114,833	650,172	1664,707	114,833	C0
3	Cicendo	1615,135	1060,931	284,385	284,385	C2
4	Andir	1636,462	1063,054	284,385	284,385	C2
5	Cidadap	153,378	598,412	1598,673	153,378	C0
...
29	Buahbati	159,117	738,024	1756,601	159,117	C0
30	Bandung Kidul	821,113	281,899	810,217	281,899	C1

Berikut merupakan anggota *cluster* 0 pada iterasi-5

Tabel 9.
Anggota Cluster 0 Pada Iterasi-5

Cluster 0								
No.	2015 Laki-laki	2015 Perempuan	2016 Laki-laki	2016 Perempuan	2017 Laki-laki	2017 Perempuan	2018 Laki-laki	2018 Perempuan
2	70	79	94	65	130	85	117	114
5	86	103	188	171	151	114	51	46
7	75	67	24	27	17	16	32	43
8	169	173	119	128	76	69	149	121
9	97	114	87	111	51	36	57	58
12	234	153	146	133	212	207	170	126
13	191	194	138	125	85	78	89	92
14	145	124	167	216	163	128	104	106
16	182	177	72	59	69	76	130	141
17	226	118	212	112	184	108	340	204
20	63	57	45	43	54	41	90	67
21	125	117	98	88	73	46	76	79
22	152	132	116	101	97	87	115	111
23	175	135	127	120	75	67	106	77
24	101	114	105	116	100	101	83	69
26	188	195	95	115	151	93	107	83
27	260	223	101	102	117	83	98	77
28	236	224	185	139	58	38	51	29
29	79	70	48	35	61	51	59	46
Jumlah Anggota: 19								
Rata-rata	150,211	135,211	114,053	105,579	101,263	80,211	106,526	89,789

Berikut merupakan anggota *cluster* 1 pada iterasi-5

Tabel 10.
Anggota Cluster 1 Pada Iterasi-5

Cluster 1								
No.	2015 Laki-laki	2015 Perempuan	2016 Laki-laki	2016 Perempuan	2017 Laki-laki	2017 Perempuan	2018 Laki-laki	2018 Perempuan
1	323	318	189	185	280	219	311	252
6	564	476	397	326	288	228	245	180
10	430	358	399	341	254	208	346	363
11	499	443	415	320	166	179	297	303
15	413	403	161	214	209	236	151	206
18	393	331	347	271	356	310	252	190
19	324	277	469	359	361	303	432	400
25	404	371	281	193	180	154	140	123
30	489	415	417	395	430	430	263	315
Jumlah Anggota: 9								
Rata-rata	426,556	376,889	341,667	289,333	280,444	251,889	270,778	259,111

Berikut merupakan anggota *cluster* 2 pada iterasi-5

Tabel 11.
Anggota Cluster 2 Pada Iterasi-5

Cluster 2								
No.	2015 Laki-laki	2015 Perempuan	2016 Laki-laki	2016 Perempuan	2017 Laki-laki	2017 Perempuan	2018 Laki-laki	2018 Perempuan
3	819	729	711	581	882	701	488	383
4	993	835	781	654	488	425	633	522
Jumlah Anggota: 2								
Rata-rata	906,000	782,000	746,000	617,500	685,000	563,000	560,500	452,800

Berikut ini merupakan tabel *centroid* baru pada iterasi ke-5 yang merupakan hasil perhitungan akhir:

Tabel 12.
Centroid Akhir Hasil Iterasi ke-5

Centroid Akhir								
C0	150,211	135,211	114,053	105,579	101,263	80,211	106,526	89,789
C1	426,556	376,889	341,667	289,333	280,444	251,889	270,778	259,111
C2	906,000	782,000	746,000	617,500	685,000	563,000	560,500	452,800

3.2. Perhitungan Algoritma K-Means dengan Rapidminer

Pada tahap ini pemodelan dilakukan dengan menggunakan *software rapidminer* versi 9.4 dengan langkah sebagai berikut:

- Langkah pertama buka *software rapidminer* versi 9.4, pilih menu file kemudian *new proses*.
- Setelah memilih *new process* langkah selanjutnya pilih *Read Excel* pada panel *operators* kemudian di *drag* atau *double click*.
- Lalu *Import* dataset yang akan digunakan dengan klik *import configuration wizard* pada panel *parameter* kemudian pilih file *excel* yang akan digunakan.
- Kemudian masukan model yang akan digunakan, pada panel *operators* pilih *K-Means* lalu di *drag* atau *double click*.
- Langkah selanjutnya tentukan terlebih dahulu jumlah *cluster* yang akan

digunakan menjadi $K=3$ pada panel *parameters*

6. Langkah selanjutnya adalah menghubungkan *connector* pada masing-masing *process*. Kemudian klik *running*.

Setelah proses *running* selesai maka akan tampil hasil *clustering* dan *cluster model* yang dihasilkan dari *dataset* dan metode yang digunakan. Berikut merupakan rekapitulasi hasil *clustering*:

Tabel 13.
Rekapitulasi Hasil Rapidminer

Cluster	Jumlah Anggota	Anggota
C0	19	2
		5
		7
		8
		9
		12
		13
		14
		16
		17
		20
		21
		22
		23
C1	9	24
		26
		27
		28
		29
		1
		6
C2	2	10
		11
		15
		18
		19
		25
		30
		3
		4

3.3. Evaluasi dan Validasi Hasil Cluster

Untuk menentukan kualitas hasil *cluster* dari proses *clustering* menggunakan metode *K-Means* yang telah di bentuk maka dilakukan evaluasi hasil *cluster*. Pada penelitian ini menggunakan metode evaluasi *davies bouldin index* (DBI). Berikut merupakan langkah-langkah evaluasi pada aplikasi *rapidminer* dan perhitungan manual *davies bouldin index*:

1. Evaluasi DBI pada *rapidminer*.

Tahap pertama masukkan *cluster distance performance* pada panel *operator* lalu hubungkan *connector* pada masing-masing *process*, kemudian klik *button running*. Setelah proses *running* maka akan muncul hasil evaluasi *davies-bouldin index* pada *performance vector*, hasil evaluasi *cluster* yang di bentuk sebanyak 3 *cluster* dengan nilai sebesar 0,571.

2. Perhitungan Manual *Davies-Bouldin Index*

a. *Sum of Square Within-Cluster* (SSW)

Tahap pertama pada evaluasi menggunakan *davies bouldin index* adalah menghitung nilai *Sum of Square Within-Cluster* (SSW). Untuk menghitung nilai SSW diperlukan jarak terdekat data dengan *centroid cluster* yang didapat pada iterasi terakhir.

Setelah mengetahui jarak terdekat setiap data dengan *centroid cluster* yang didapat dari iterasi terakhir, langkah selanjutnya menghitung nilai SSW.

$$SSW_0 = \frac{114,833 + 153,378 + 208,711 + 79,600 + \dots + 90,212 + 142,953 + 178,398 + 159,117}{19} = 129,073$$

$$SSW_1 = \frac{226,129 + 201,671 + 159,380 + 192,253 + 248,406 + 133,457 + 310,655 + 262,306 + 281,899}{9} = 224,017$$

$$SSW_2 = \frac{284,385 + 284,385}{2} = 284,385$$

Dari hasil perhitungan didapatkan rata-rata jarak setiap data di *cluster 0* (SSW_0) sebesar 129,073. Untuk rata-rata jarak setiap data di *cluster 1* (SSW_1) sebesar 224,017. Sedangkan rata-rata jarak setiap data di *cluster 2* (SSW_2) sebesar 284,385.

b. *Sum of Square Between-Cluster* (SSB)

Untuk menghitung nilai SSB maka diperlukan *centroid* pada iterasi terakhir.

Nilai SSB didapatkan dengan menghitung jarak antar *centroid* :

$$SSB01 = \sqrt{\frac{(150,211 - 426,556)^2 + (135,211 - 376,889)^2 + (114,053 - 341,667)^2 + (105,579 - 289,333)^2 + (101,263 - 280,444)^2 + (80,211 - 251,889)^2 + (106,526 - 270,778)^2 + (89,789 - 259,111)^2}{8}}$$

$$= \sqrt{76,367 + 58,408 + 51,808 + 33,766 + 32,106 + 29,473 + 26,979 + 28,670}$$

$$= \sqrt{337,577}$$

$$= 581,013$$

$$SSB02 = \sqrt{\frac{(150,211 - 906,000)^2 + (135,211 - 782,000)^2 + (114,053 - 746,000)^2 + (105,579 - 617,500)^2 + (101,263 - 685,000)^2 + (80,211 - 563,000)^2 + (106,526 - 560,500)^2 + (89,789 - 452,500)^2}{8}}$$

$$= \sqrt{571,217 + 418,336 + 399,357 + 262,063 + 340,749 + 233,085 + 206,092 + 131,559}$$

$$= \sqrt{2,562,458}$$

$$= 1600,769$$

$$SSB12 = \sqrt{\frac{(426,556 - 906,000)^2 + (376,889 - 782,000)^2 + (341,667 - 746,000)^2 + (289,333 - 617,500)^2 + (280,444 - 685,000)^2 + (251,889 - 563,000)^2 + (270,778 - 560,500)^2 + (259,111 - 452,500)^2}{8}}$$

$$= \sqrt{229,867 + 164,115 + 163,485 + 107,694 + 163,666 + 96,790 + 83,939 + 37,399}$$

$$= \sqrt{1,046,953}$$

$$= 1023,208$$

c. Rasio

Setelah menghitung nilai SSW dan SSB, kemudian akan dilakukan perhitungan rasio antar *cluster* satu dengan *cluster* yang lain. Untuk menghitung nilai rasio antar *cluster*.

$$R_{01} = \frac{SSW_0 + SSW_1}{SSB_{0,1}} = \frac{129,073 + 224,017}{581,013} = 0,608$$

$$R_{02} = \frac{SSW_0 + SSW_2}{SSB_{0,2}} = \frac{129,073 + 284,385}{1600,769} = 0,258$$

$$R_{12} = \frac{SSW_1 + SSW_2}{SSB_{1,2}} = \frac{224,017 + 284,385}{1023,208} = 0,497$$

d. *Davies Bouldin Index*

Tahap terakhir adalah mencari nilai DBI, nilai DBI. Hasilnya ditampilkan pada tabel dibawah ini:

Tabel 14. Nilai DBI

R	0	1	2	R MAX	DBI
0	0,000	0,608	0,258	0,608	0,571
1	0,608	0,000	0,497	0,608	
2	0,258	0,497	0,000	0,497	

Untuk mengetahui apakah hasil *cluster* yang dibentuk tergolong dalam kategori baik, maka dilakukan percobaan evaluasi dari pembentukan jumlah k=2 sampai k=5. Berikut hasil percobaan evaluasi *cluster* dengan menggunakan DBI pada tabel berikut.:

Tabel 15. Hasil Uji DBI

Jumlah Cluster	Nilai DBI
2	0,608
3	0,571
4	0,861
5	0,686

Berdasarkan hasil evaluasi pada aplikasi *rapidminer* dan perhitungan manual *davies bouldin index* k=3 menghasilkan nilai DBI yang sama yaitu 0.571. Berdasarkan nilai tersebut maka pengelompokan data dapat dikatakan baik, karena dengan percobaan evaluasi dari pembentukan jumlah k=2 sampai k=5, k=3 merupakan nilai yang paling mendekati 0.

4. Kesimpulan

Berdasarkan hasil *cluster* 0, *cluster* 1 dan *cluster* 2, pada perhitungan manual maupun aplikasi *rapidminer*, maka dapat ditarik kesimpulan bahwa:

1. Kelompok wilayah penyebaran rendah pada *cluster* 0 terdiri dari 19 wilayah, yaitu kecamatan Sukajadi, Cidapad, Bandung Wetan, Sumur Bandung, Cibeunying Kaler, Batununggal, Lengkong, Regol, Bojongloa Kaler, Bojongloa Kidul, Antapani, Mandalajati, Arcamanik, Ujungberung, Cinambo dan Panyileukan dengan jumlah penderita pneumonia pada balita didominasi penderita laki-laki. Jumlah penderita laki-laki sebanyak 23.937 dan jumlah penderita perempuan 20.701. Rata-rata jumlah penderita perkecamatan sebanyak 110 balita. Kelompok wilayah penyebaran sedang pada *cluster* 1 terdiri dari 9 wilayah, yaitu kecamatan Sukasari, Cobleng, Cibeunying Kidul, Kiaracandong, Astanaanyar, Babakan Ciparay, Bandung Kulon, Cibiru dan Bandung Kidul dengan jumlah penderita pneumonia pada balita didominasi penderita laki-laki. Jumlah penderita laki-laki sebanyak 26.639 dan jumlah penderita perempuan 23.230. Rata-rata jumlah penderita perkecamatan sebanyak 313 balita. Kelompok wilayah penyebaran tinggi pada *cluster* 2 terdiri dari 2 wilayah, yaitu Kecamatan Andir dan Kecamatan Cicendo dengan jumlah penderita pneumonia pada balita didominasi penderita laki-laki. Jumlah penderita laki-laki berjumlah 5795 dan jumlah penderita perempuan berjumlah 4830. Rata-rata jumlah penderita perkecamatan sebanyak 664 balita. Hasil penelitian ini dapat dijadikan informasi bagi Dinas Kesehatan Kota Bandung untuk menunjang suatu keputusan dalam pencegahan atau penanggulangan penyebaran penyakit pneumonia pada balita di Kota Bandung.
2. Berdasarkan evaluasi *davies-bouldin index* (DBI) pembentukan 3 *cluster* telah dinilai baik, karena dengan percobaan evaluasi dari pembentukan jumlah k=2 sampai k=5, k=3 merupakan nilai yang paling mendekati 0 dengan nilai sebesar 0,571.

Referensi

- Dinas Kesehatan Provinsi Jawa Barat. (2017). Profil Kesehatan Dinas Kesehatan Provinsi Jawa Barat 2017. *Diskes Jabarprov*, 52. Retrieved from <http://diskes.jabarprov.go.id/dmdocuments/01b3018430a412a520e2b4a4b9d9864f.pdf> (17 Mei 2019)
- KeMenKes, R. I. (2017a). *Pneumonia Penyebab Kematian Utama Balita*. Retrieved from <https://www.kemkes.go.id/article/view/410/pneumonia-penyebab-kematian-utama-balita.html> (17 Mei 2020)
- KeMenKes, R. I. (2017b). *Profil Kesehatan Indonesia 2017*. Retrieved from <https://www.kemkes.go.id/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-Indonesia-tahun-2017.pdf> (17 Mei 2020)
- Nasari, F., & Sianturi, C. J. M. (2016). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat. *Cogito Smart Journal*, 2(2), 108. <https://doi.org/10.31154/cogito.v2i2.19108-119>
- Purnia, D. S., & Warnilah, A. I. (2017). Implementasi Data Mining Pada Penjualan Kacamata Menggunakan Algoritma Apriori. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 2(2), 31–39. <https://doi.org/https://doi.org/10.31294/ijcit.v2i2.2776>
- Rahayu, P. (2016). Gambaran Praktik/Kebiasaan Keluarga Terkait Dengan Kejadian Pneumonia Pada Balita Di Upt Puskesmas Sigaluh 2 Banjarnegara. *Jurnal Kesehatan Masyarakat (e-Jurnal)*, 4(3), 93–97. Retrieved from [http://ejournal-s1.undip.ac.id/index.php/jkm%0A\(21](http://ejournal-s1.undip.ac.id/index.php/jkm%0A(21) Juli 2020)
- Siyoto, S., & Sodik, M. A. (2015). *Dasar Metodologi Penelitian*. Yogyakarta: Literasi Media Publishing.
- Talakua, M. W., Leleury, Z. A., & Talluta, A. W. (2017). Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *Jurnal Ilmu Matematika Dan Terapan*, 11(2), 119–128. Retrieved from <https://ojs3.unpatti.ac.id/index.php/barekeng/article/view/353/308> (10 Juni 2020)
- Toresa, D. (2020). Implementasi k-means terhadap penyebaran penyakit. *JUTIM (Jurnal Teknik Informatika Musirawas)*, 05(01), 35–42. Retrieved from <http://jurnal.univbinainsan.ac.id/index.php/jutim/article/view/809/553> (18 Mei 2020)
- UNICEF. (2019). *Lembaga kesehatan dan anak memperingatkan satu anak meninggal akibat pneumonia setiap 39 detik*. Retrieved from <https://www.unicef.org/indonesia/id/press-releases/lembaga-kesehatan-dan-anak-memperingatkan-satu-anak-meninggal-akibat-pneumonia-setiap> (17 Mei 2020)
- Wardhani, A. K. (2016). Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*, 14(1), 30–37. Retrieved from <http://journals.usm.ac.id/index.php/transformatika/article/download/387/248> (15 Mei 2020)
- WHO. (2019). *Pneumonia*. Retrieved from <https://www.who.int/en/news-room/fact-sheets/detail/pneumonia> (17 Mei 2020)