

Clustering Menggunakan Algoritma *K-Means* Pada Penyakit ISPA di Puskesmas Kabupaten Karawang

Vickry Ramadhan¹, Apriade Voutama²

^{1,2}Program, Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang

Email: vickry.ramadhan18183@student.unsika.ac.id², apriade.voutama@staff.unsika.ac.id²

Abstrak

ISPA merupakan penyakit yang sering terjadi di masyarakat. ISPA dapat menyerang berbagai kalangan terutama anak-anak baik pernapasan bawah maupun atas, dan dapat menyebabkan berbagai infeksi dari infeksi ringan hingga dapat menyebabkan kematian. Dalam kasus beberapa yang terjadi khususnya yang terjadi di beberapa puskesmas yang berada di Kabupaten Karawang itu tersendiri sudah sangat banyak. Oleh karena itu agar tidak terjadinya peningkatan penyakit ISPA di perlukan suatu metode yang efektif untuk membantu kegiatan puskesmas dalam pencegahan penyebaran penyakit ISPA. Tujuannya agar puskesmas bisa melihat desa mana saja yang membutuhkan perhatian lebih dalam menangani kasus penyebaran penyakit ISPA. Untuk itu diperlukan pengelompokan (*cluster*) terhadap desa mana saja yang membutuhkan perhatian dan penanganan yang efektif. Sehingga diperoleh puskesmas mana saja yang memiliki kasus terbanyak sehingga dapat dijadikan salah satu acuan untuk menentukan sebuah wilayah mana saja yang paling banyak sehingga wilayah tersebut bisa menjadi fokus utama. Pada hasil ini akan dijelaskan tentang yang sudah diteliti yaitu dengan sesuai alur yang telah direncanakan antara lain yaitu menggunakan algoritma *K-Means* dengan menggunakan metodologi *knowledge discovery in database* (KDD) yang terdiri beberapa tahapan yaitu *data selection*, *data cleaning*, *data integration*, *data transformation*, *data mining*, dan *knowledge* atau *evaluation*. Serta menggunakan tools RStudio dalam melakukan proses pengolahan data.

Kata Kunci: ISPA, *Clustering*, *K-Means*, *RStudio*

Abstract

ISPA is a disease that often occurs in the community. ISPA can attack various groups, especially children, both lower and upper respiratory, and can cause various infections from mild infections to death. In the case of several cases, especially those that occurred in several puskesmas in Karawang Regency, there were already too many. Therefore, in order to avoid an increase in ISPA disease, an effective method is needed to assist puskesmas activities in preventing the spread of ISPA disease. The goal is that the puskesmas can see which villages need more attention in handling cases of the spread of ISPA disease. For this reason, it is necessary to group (*cluster*) which villages require effective attention and handling. So that it is obtained which health centers have the most cases so that it can be used as a reference to determine which area has the most so that the area can be the main focus. This result will explain what has been researched, namely according to the planned flow, including using the *K-Means* algorithm using the *Knowledge Discovery in Database* (KDD) methodology which consists of several stages, namely *data selection*, *data cleaning*, *data integration*, *data transformation*, *data mining*, and *knowledge* or *evaluation*. And use RStudio tools in processing data.

Kata Kunci: ISPA, *Clustering*, *K-Means*, *RStudio*

PENDAHULUAN

Puskesmas merupakan fasilitas pelayanan kesehatan yang menyelenggarakan upaya kesehatan masyarakat dan upaya kesehatan perseorangan .tingkat pertama, dengan lebih mengutamakan upaya promotif dan preventif di wilayah kerjanya. Setiap bulannya pasien yang datang berobat semakin meningkat, kenaikan pasien yang berobat pun diiringi dengan meningkatnya berbagai macam penyakit juga. Salah satu dari beberapa macam kasus penyakit yang dialami oleh masyarakat yaitu ISPA (infeksi saluran pernapasan

akut). ISPA merupakan penyakit yang sering terjadi di masyarakat. ISPA dapat menyerang berbagai kalangan terutama anak-anak baik pernapasan bawah maupun atas, dan dapat menyebabkan berbagai infeksi dari infeksi ringan hingga dapat menyebabkan kematian. Dalam kasus beberapa yang terjadi khususnya yang terjadi di beberapa puskesmas yang berada di Kabupaten Karawang itu tersendiri sudah sangat banyak. Oleh karena itu agar tidak terjadinya peningkatan penyakit ISPA diperlukan suatu metode yang efektif untuk membantu kegiatan puskesmas dalam pencegahan penyebaran penyakit ISPA. Tujuannya agar Dinas Kesehatan bisa melihat puskesmas mana saja yang membutuhkan perhatian lebih dalam menangani kasus penyebaran penyakit ISPA. Untuk itu diperlukan pengelompokan (*cluster*) terhadap desa mana saja yang membutuhkan perhatian dan penanganan yang efektif. Sehingga diperoleh puskesmas mana saja yang memiliki kasus terbanyak sehingga dapat dijadikan salah satu acuan untuk menentukan sebuah wilayah mana saja yang paling banyak sehingga wilayah tersebut bisa menjadi fokus utama. Dari permasalahan ini maka akan dilakukan penelitian Clustering penyakit ISPA menggunakan algoritma K-Means dengan tahap Knowledge Discovery in Database (KDD) menggunakan tools RStudio. Penelitian tersebut dilakukan diharapkan untuk mempermudah Puskesmas yang ada di Kabupaten Karawang dalam menemukan penyebaran penyakit ISPA.

METODE

Metodologi yang digunakan pada penelitian ini adalah metodologi *Knowledge Discovery in Database* (KDD).

HASIL DAN PEMBAHASAN

Dataset diperoleh dari Dinas Kesehatan Kabupaten Karawang. Berikut tabel kasus penyakit ISPA yang ada di Kabupaten Karawang.

Tabel 4. 1 penyakit ispa yang ada di Kabupaten Karawang tahun 2019

No	Nama Puskesmas	Jumlah Penduduk tahun 2019	ISPA 2019
1	Adiarsa	65,962	103
2	Anggadita	24,394	90
3	Balongsari	19,543	78
4	BatuJaya	81,806	33
5	BayurLor	24,847	0
6	Ciampel	38,390	38
7	Cibuaya	53,035	81
8	Cicinde	30,589	54
9	Cikampek	111,415	134
10	Cikampek Utara	42,689	300
11	Cilamaya	51,295	25
12	Curug	29,746	7
13	Gempol	24,497	15
14	Jatisari	53,696	51
15	Jyakerta	35,953	12
16	Jomin	29,814	53
17	Kalangsari	30,355	0
18	Karawang	23,981	33
19	KrawangKulon	23,359	99
20	Kertamukti	38,510	780
21	Klari	91,290	2195
22	KotaBaru	51,936	114
23	Kutamukti	32,806	0
24	Kutawaluya	33,679	18

25	Lemah Duhur	23,116	3
26	Lemah Abang	70,855	286

Tabel 4.1 Penyakit ispa yang ada di Kabupaten Karawang tahun 2019 (Lanjutan)

27	Loji	36,079	48
28	Majalaya	43,890	0
29	Medang Asem	31,822	39
30	Nagasari	36,401	5
31	Pacing	19,733	40
32	PakisJaya	40,981	2
33	Pangkalan	36,821	12
34	Pasirukem	47,574	31
35	Pedes	57,666	2
36	Plawad	54,883	66
37	Purwasari	61,821	11
38	Rawamerta	33,901	77
39	Rengasdengklok	30,355	38
40	Sukatani	31,270	2
41	SungaiBuntu	24,192	34
42	TanjungPura	37,006	13
43	Telagasari	67,735	16
44	TelukJambe	65,727	32
45	Tempuran	43,419	1
46	TirtaJaya	70,443	272
47	TirtaMulya	48,787	156
48	TunggakJati	28,519	0
49	Wadas	50,476	100
50	Wanakerta	49,042	0

Tabel 4. 2 penyakit ispa yang ada di Kabupaten Karawang tahun 2020

no	Nama Puskesmas	Jumlah Penduduk tahun 2020	ISPA 2020
1	Adiarsa	66423	215
2	Anggadita	24567	46
3	Balongsari	20886	77
4	BatuJaya	93401	79
5	BayurLor	27053	1
6	Ciampel	40025	14
7	Cibuaya	61792	41
8	Cicinde	31892	67
9	Cikampek	114013	149
10	Cikampek Utara	37854	211
11	Cilamaya	57492	42
12	Curug	29080	8
13	Gempol	30701	5
14	Jatisari	61597	125
15	Jayakarta	38399	9
16	Jomin	30256	30

Tabel 4.2 Penyakit ispa yang ada di Kabupaten Karawang tahun 2020 (Lanjutan)

17	Kalangsari	39009	0
18	Karawang	24325	41
19	KrawangKulon	23905	70
20	Kertamukti	50983	809
21	Klari	93994	1505
22	KotaBaru	51081	66
23	Kutamukti	33289	0
24	Kutawaluya	34177	23
25	Lemah Duhur	24827	4
26	Lemah Abang	72507	2
27	Loji	42349	41
28	Majalaya	45759	0
29	Medang Asem	35645	57
30	Nagasari	37709	0
31	Pacing	19871	14
32	PakisJaya	45539	0
33	Pangkalan	40432	4
34	Pasirukem	48276	53
35	Pedes	64696	8
36	Plawad	45680	13
37	Purwasari	62734	0
38	Rawamerta	35347	52
39	Rengasdengklok	83223	39
40	Sukatani	35831	1
41	SungaiBuntu	24548	14
42	TanjungPura	37553	369
43	Telagasari	68212	12
44	TelukJambe	67199	101
45	Tempuran	48327	0
46	TirtaJaya	78808	340
47	TirtaMulya	53665	161
48	TunggakJati	29809	9
49	Wadas	50624	97
50	Wanakerta	55124	0

Tabel 4. 3 penyakit ispa yang ada di Kabupaten Karawang tahun 2021

no	Nama Puskesmas	Jumlah Penduduk tahun 2021	ISPA 2021
1	Adiarsa	66103	154
2	Anggadita	24449	23
3	Balongsari	20785	68
4	BatuJaya	92951	71
5	BayurLor	26923	7
6	Ciampel	39832	20

7	Cibuaya	61495	0
8	Cicinde	31738	95
9	Cikampek	113464	125
10	Cikampek Utara	37672	136
11	Cilamaya	57215	133
12	Curug	28940	1
13	Gempol	30553	10
14	Jatisari	61301	1
15	Jayakarta	38214	10
16	Jomin	30110	6
17	Kalangsari	38821	3
18	Karawang	24209	44
19	KrawangKulon	23790	77
20	Kertamukti	50738	2
21	Klari	93542	108
22	KotaBaru	50835	40
23	Kutamukti	33129	1
24	Kutawaluya	34012	17
25	Lemah Duhur	24707	6
26	Lemah Abang	72158	20
27	Loji	42145	39
28	Majalaya	45539	0
29	Medang Asem	35473	45
30	Nagasari	37527	2
31	Pacing	19775	0
32	PakisJaya	45320	0
33	Pangkalan	40237	7
34	Pasirukem	48044	84
35	Pedes	64385	1
36	Plawad	45460	13
37	Purwasari	62432	0
38	Rawamerta	35177	38
39	Rengasdengklok	82822	7
40	Sukatani	35659	2
41	SungaiBuntu	24430	8
42	TanjungPura	37372	129

Tabel 4.3 Penyakit ispa yang ada di Kabupaten Karawang tahun 2021 (Lanjutan)

43	Telagasari	67884	14
44	TelukJambe	66876	13
45	Tempuran	48094	0
46	TirtaJaya	78429	56
47	TirtaMulya	53407	128
48	TunggakJati	29666	0
49	Wadas	50380	74
50	Wanakerta	54859	0

4.1.1 Data Integration

Setelah melakukan *data selection* dari *dataset* penyakit ispa yang ada di puskesmas yang berada di

Kabupaten Karawang sehingga dapat melakukan sebuah penggabungan data. Berikut tabel kasus penyakit ispa yang ada di Kabupaten Karawang yang sudah dilakukan penggabungan data.

Tabel 4. 4 Dataset penyakit ispa Kabupaten Karawang tahun 2019-2021

no	Nama Puskesmas	Jumlah Penduduk tahun 2021	ISPA 2019-2021
1	Adiarsa	66103	472
2	Anggadita	24449	159
3	Balongsari	20785	223
4	BatuJaya	92951	183
5	BayurLor	26923	8
6	Ciampel	39832	72
7	Cibuaya	61495	122
8	Cicinde	31738	216
9	Cikampek	113464	408
10	Cikampek Utara	37672	647
11	Cilamaya	57215	200
12	Curug	28940	16
13	Gempol	30553	30
14	Jatisari	61301	177
15	Jayakarta	38214	31
16	Jomin	30110	89
17	Kalangsari	38821	3
18	Karawang	24209	118
19	KrawangKulon	23790	246
20	Kertamukti	50738	1591
21	Klari	93542	3808
22	KotaBaru	50835	220
24	Kutawaluya	34012	58

Tabel 4.4 Dataset penyakit ispa Kabupaten Karawang tahun 2019-2021(lanjutan)

25	Lemah Duhur	24707	13
26	Lemah Abang	72158	308
27	Loji	42145	128
28	Majalaya	45539	0
29	Medang Asem	35473	141
30	Nagasari	37527	7
31	Pacing	19775	54
32	PakisJaya	45320	2
33	Pangkalan	40237	23
34	Pasirukem	48044	168
35	Pedes	64385	11
36	Plawad	45460	92
37	Purwasari	62432	11
38	Rawamerta	35177	167
39	Rengasdengklok	82822	84
40	Sukatani	35659	5
41	SungaiBuntu	24430	56
42	TanjungPura	37372	511
43	Telagasari	67884	42
44	TelukJambe	66876	146

45	Tempuran	48094	1
46	TirtaJaya	78429	668
47	TirtaMulya	53407	445
48	TunggakJati	29666	9
49	Wadas	50380	271
50	Wanakerta	54859	0

Dari tabel 4.2 kasus penyakit ispa Kabupaten Karawang tahun 2019-2021 terdiri dari 7 variabel dan 50 objek data.

Dari *dataset* diatas dilakukan proses seleksi data atau pemilihan data, sehingga hanya data-data tertentu yang akan digunakan pada proses *data mining*.

4.1.2 Data Cleaning

Setelah dilakukan *data selection* dan *data itergration* kemudia dilakukan penanganan *missing value*, duplikasi data dan data yang tidak sesuai atau inkonsisten data. Pada *dataset* penyakit ispa yang ada di beberapa puskesmas di Kabupaten Karawang tahun 2019-2021 tidak terdapat adanya data duplikasi dan data yang digunakan sudah konsisten, pada *dataset* tersebut tidak terdapat sebuah *missing value*.

no	Nama.Puskesmas
Min. : 1.00	Length:50
1st Qu.:13.25	Class :character
Median :25.50	Mode :character
Mean :25.50	
3rd Qu.:37.75	
Max. :50.00	
Jumlah.Penduduk.tahun.2019	Jumlah.Penduduk.tahun.2020
Min. : 19543	Min. : 19871
1st Qu.: 30355	1st Qu.: 32241
Median : 37698	Median : 41391
Mean : 43722	Mean : 47410
3rd Qu.: 52760	3rd Qu.: 60571
Max. :111415	Max. :114013
Jumlah.Penduduk.tahun.2021	ISPA.2019
Min. : 19775	Min. : 0.00
1st Qu.: 32086	1st Qu.: 8.00
Median : 41191	Median : 33.50
Mean : 47182	Mean : 111.98
3rd Qu.: 60280	3rd Qu.: 80.25
Max. :113464	Max. :2195.00
ISPA.2020	ISPA.2021
Min. : 0.00	Min. : 0.00
1st Qu.: 4.25	1st Qu.: 2.00
Median : 34.50	Median : 13.00
Mean : 100.48	Mean : 36.76
3rd Qu.: 75.25	3rd Qu.: 65.00
Max. :1505.00	Max. :154.00

Gambar 4. 1 Hasil dari *data cleaning*

Pada gambar 4.1 Setelah melakukan *data cleaning* pada *dataset* yang telah dilakukan tahap penggabungan maka mendapatkan hasil tidak adanya *missing value* dari *dataset* tersebut.

4.1.3 Data Transformation

Pada tahap ini data yang masih berjenis kategorik di transformasikan ke dalam bentuk numerik, dalam proses perubahan data kategorik ke numerik adalah dengan memberikan sebuah angka yang berbeda ke setiap kategori yang berbeda untuk setiap variable pada *dataset* atau juga bisa dengan melakukan tahapan transformasi dari bentuk kategorik biner ke numerik. Pada proses transformasi dengan metode skala menggunakan bantuan *tools RStudio* sehingga mendapatkan hasil seperti pada tabel dibawah ini. Berikut merupakan hasil tranformasi skala pada *dataset* penyakit ispa yang ada di beberapa puskesmas di Kabupaten

Tabel 4. 5 Hasil Transformasi Skala pada Dataset

no	jumlah penduduk 2020	jumlah penduduk 2021	ispa 2019	ispa 2020	ispa 2021
1	0.91136597	0.91135102	-0.0276031	0.471686593	2.54331565
2	-1.09492722	-1.09491359	-0.067563	-0.224392993	-0.298499
3	-1.27136944	-1.27139012	-0.104449	-0.096709755	0.67769687
4	2.2045086	2.20448474	-0.2427717	-0.088472127	0.74277659
5	-0.97576522	-0.97575339	-0.3442084	-0.409739628	-0.64559087
6	-0.35397539	-0.3539915	-0.2274025	-0.356195045	-0.36357873
7	0.68938719	0.68940673	-0.0952275	-0.244987064	-0.79744356
8	-0.74381633	-0.74383895	-0.1782211	-0.137897897	1.26341439
9	3.19250829	3.19249324	0.06768592	0.199844861	1.91421164
10	-0.45803842	-0.45802789	0.57794304	0.455211336	2.1528373
11	0.48327431	0.48326056	-0.2673624	-0.24086825	2.08775757
12	-0.87860457	-0.8786046	-0.3226915	-0.38090793	-0.77575032
13	-0.8009048	-0.80091447	-0.2981008	-0.393264372	-0.58051115
14	0.68004021	0.68006272	-0.1874426	0.100993322	-0.77575032
15	-0.43191481	-0.43192246	-0.3073223	-0.376789115	-0.58051115
16	-0.82223509	-0.82225156	-0.181295	-0.290294019	-0.66728411
17	-0.40267555	-0.40268631	-0.3442084	-0.413858442	-0.73236384
18	-1.10652707	-1.10647318	-0.2427717	-0.244987064	0.15705907
19	-1.12665902	-1.12665432	-0.0398984	-0.125541454	0.87293604
20	0.17127693	0.17129591	2.05338533	2.918262178	-0.75405708
21	2.232933	2.23295025	6.40286624	5.784956804	1.54542653
22	0.17597438	0.17596791	0.00620915	-0.142016711	0.0702861
23	-0.67685361	-0.67684145	-0.3442084	-0.413858442	-0.77575032
24	-0.6342889	-0.63431176	-0.2888793	-0.319125718	-0.42865845
25	-1.08246459	-1.08248702	-0.3349869	-0.397383186	-0.66728411
26	1.20299173	1.20299005	0.53490931	-0.405620814	-0.36357873

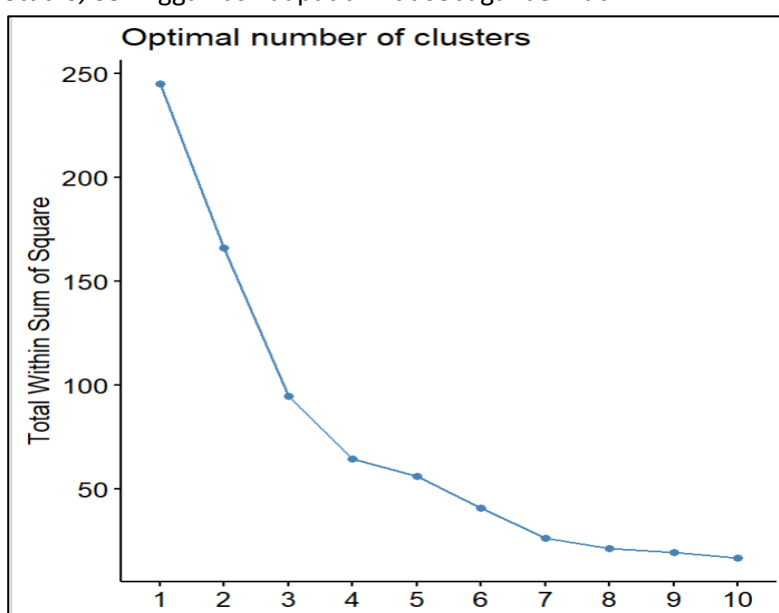
Tabel 4.5 Hasil Transformasi Skala pada *Dataset* (lanjutnya)

27	-0.24257856	-0.24258588	-0.1966642	-0.244987064	0.04859286
28	-0.07912626	-0.07911389	-0.3442084	-0.413858442	-0.79744356
29	-0.56392292	-0.56394271	-0.2243287	-0.179086038	0.17875231
30	-0.46498874	-0.46501181	-0.3288392	-0.413858442	-0.75405708
31	-1.32002166	-1.32003676	-0.2212549	-0.356195045	-0.79744356
32	-0.08967157	-0.08966202	-0.3380607	-0.413858442	-0.79744356
33	-0.33446656	-0.33448468	-0.3073223	-0.397383186	-0.64559087
34	0.04152168	0.04153942	-0.2489194	-0.195561294	1.02478873
35	0.82858528	0.82860356	-0.3380607	-0.38090793	-0.77575032
36	-0.08291298	-0.08291892	-0.1413351	-0.360313859	-0.51543142
37	0.73454029	0.73453733	-0.3103962	-0.413858442	-0.79744356
38	-0.57820703	-0.57819954	-0.1075229	-0.199680108	0.02689962
39	1.7166442	1.71662153	-0.2274025	-0.253224692	-0.64559087
40	-0.55500734	-0.55498402	-0.3380607	-0.409739628	-0.75405708
41	-1.09583796	-1.09582872	-0.2396979	-0.356195045	-0.62389763

42	-0.47246633	-0.47247739	-0.3042485	1.105983967	2.00098461
43	0.99711851	0.99713287	-0.295027	-0.364432673	-0.49373818
44	0.94856215	0.94858256	-0.2458456	0.002141783	-0.51543142
45	0.04396627	0.04394767	-0.3411346	-0.413858442	-0.79744356
46	1.505019	1.50503272	0.49187557	0.986538357	0.41737797
47	0.29983385	0.29984827	0.13531035	0.24927063	1.97929137
48	-0.84366124	-0.84363682	-0.3442084	-0.376789115	-0.79744356
49	0.1540689	0.15405284	-0.0368246	-0.014333473	0.80785632
50	0.36976843	0.36978384	-0.3442084	-0.413858442	-0.79744356

4.1.4 Data Mining

Pada tahap *data mining*, algoritma yang digunakan yaitu algoritma *K-means*. Untuk pemilihan jumlah *cluster* yang optimal menggunakan metode *elbow*. Dalam pemilihan jumlah *cluster* yang optimal menggunakan tools *RStudio*, sehingga hasil dapat dilihat sebagai berikut.



Gambar 4. 2 Hasil Metode Elbow

Dari hasil metode *elbow* dapat diketahui nilai hasil jumlah *cluster* optimal pada *dataset* penyakit ispa yang ada di beberapa puskesmas di Kabupaten Karawang tahun 2019-2021 yaitu 3 *cluster*.

Adapun hasil *dataset* dari penyakit ispa yang ada di beberapa puskesmas di Kabupaten Karawang tahun 2019-2021 menggunakan algoritma *K-Means*. Dan menghasilkan 3 *cluster* yaitu *cluster* 1, *cluster* 2 dan *cluster* 3. Berikut tampilan hasil visual menggunakan tools *RStudio*.



Gambar 4. 3 Hasil Cluster

Adapun hasil *cluster* dari *dataset* penyakit ispa yang ada di beberapa puskesmas di Kabupaten Karawang tahun 2019-2021. Berikut dapat dilihat dari tabel berikut.

Tabel 4. 6 Hasil puskesmas dari cluster

Cluster	Nama puskesmas	Jumlah	Kriteria
1	Anggadita, Balongsari, Bayur Lor, Ciampel, Ciampel Ciciinde, Curug, Gempol, JayaKerta, Jomin, Kalangsari, Karawang, KarawangKulon, KutaMukti, Kutawaluya, Lemah Duhur, Loji, Majalaya, Medang Asem, Nagasari, Pacing, PakisJaya, Pangkalan, Plawad, Rawamerta, Sukatani, SungaiBuntu, Tempuran, TunggakJati, Wanakerta	30	Rendah
2	Adiarsa, BatuJaya, Cibuaya, Cibuaya, Cikampek, Cikampek Utara, Cilamaya, Jatisari, Kertamukti, Kotabaru, Lemah Abang, Pasirukem, Pedes, Purwasari, Rengasdengklok, TanjungPura, Telagasari, Telukjambe, Tirtajaya, Tirtamulya, Wadas	19	Sedang

Tabel 4.6 hasil puskesmas dari *cluster*

<i>cluster</i>	Nama puskesmas	jumlah	Kriteria
3	Klari	1	Tinggi

4.1.5 Knowledge/Evaluation

Setelah dilakukannya analisis pada cluster dengan menggunakan algoritma *K-means* tahap selanjutnya dilakukan evaluasi terhadap model yang telah dibuat dengan menggunakan *Silhouette Coefficient* yang bertujuan untuk mengetahui kualitas dari suatu objek pada suatu *cluster* menurut Handoyo, et al. (2014), dalam Wira, Budianto, & Wiguna, (2019). Dalam melakukan pengukuran pada kualitas *cluster* dengan menggunakan pengujian yang berpengaruh pada jumlah nilai *cluster*.

Tabel 4. 7 Hasil *Cluster*

Data Uji	Nilai K	Silhouette Coefficient
50	4	0.4526
	5	0.4960
	6	0.4866
	7	0.4504
	8	0.4051

Tabel di atas menunjukkan hasil dari pengujian pada kualitas *cluster* pada *dataset* Penyakit Diare Kabupaten Karawang Tahun 2017-2019 dengan data uji 50 data, pengujian kualitas *cluster* yang di uji yaitu nilai k=4 sampai k=8. Kualitas *cluster* paling baik diperoleh pada *cluster* 5 dengan nilai *Silhouette Coefficient* sebesar 0.4960.

Berdasarkan hasil tersebut dapat disimpulkan bahwa *cluster* 5 merupakan *cluster* yang kualitasnya baik karena mendekati angka 1 (*Medium Structure*).

4.1.6 Pembahasan

Dalam penelitian ini untuk mengelompokan puskesmas yang memiliki kasus penyakit ispa yang berada di Kabupaten Karawang menggunakan algoritma *K-Means* dengan menggunakan metodologi *knowledge discovery in database* (KDD) yang terdiri beberapa tahapan yaitu *data selection*, *data cleaning*, *data integration*, *data transformation*, *data mining*, dan *knowledge* atau *evaluation*. Serta menggunakan tools RStudio dalam melakukan proses pengolahan data.

Data selection merupakan proses pemilihan data yang akan dilakukan dalam proses *data mining*, kemudian dilanjutkan pada tahap *data integration* dimana data yang telah dipilih pada tahap *data selection* lalu melakukan penggabungan data, *data cleaning* dimana data dilakukan pengecekan apabila terdapat data yang hilang atau terdapat *typography* (kesalahan *input*) maka data harus ditangani terlebih dahulu, kemudian dilakukan *data transformation* dimana data yang telah dibersihkan melalui tahap *data preprocessing* diubah standarisasi datanya dengan menggunakan *transformasi scala* selain itu data yang masih berbentuk kategori juga ditransformasikan pada bentuk numerik, selanjutnya adalah proses *data mining* yaitu data tersebut diekstraksi bentuk dan polanya dengan menggunakan algoritma *K-Means* dan menggunakan tools RStudio sedangkan untuk melakukan proses *mining* menggunakan algoritma *K-Means* dapat menggunakan *package cluster* dan *Cvalid* untuk menampilkan visualisasi dari *cluster* menggunakan *package factoextra*. Selanjutnya dilakukannya evaluasi dari hasil model *data mining* yang telah didapatkan yaitu untuk mengukur optimalisasi dari *cluster* yang dihasilkan pada tools RStudio dengan menggunakan *package Cvalid* untuk mengetahui nilai evaluasi dari *internal measure* berupa nilai *silhouette*.

Adapun *cluster* optimal yang dihasilkan pada penelitian ini adalah pada *dataset* Penyakit Ispa Kabupaten Karawang Tahun 2019-2021 menghasilkan 3 *cluster* dan optimal *cluster* pada *cluster* 4. Parameter untuk mengetahui kualitas suatu *cluster* dapat dilakukan dengan menggunakan *Silhouette Coefficient* yang bertujuan untuk mengukur kualitas dari *cluster*. Nilai *silhouette* menghitung rata-rata nilai separation dan

compactness yang dibagi dengan maksimum antara keduanya. Jumlah cluster yang terbaik ditunjukkan dengan nilai Silhouette yang semakin mendekati 1.

Dari dataset yang digunakan yaitu dataset Penyakit Ispa Kabupaten Karawang Tahun 2019-2021 dengan nilai *Silhouette Coefficient* 0.4960 menunjukkan bahwa kualitas *cluster* yang baik (*medium structure*) karena mendekati angka 1. Hal tersebut menunjukkan bahwa jumlah data mempengaruhi terhadap kualitas dari sebuah *cluster*, jika jumlah data yang digunakan semakin sedikit, nilai *Silhouette Coefficient* yang dihasilkan semakin rendah (Pramesti et al., 2017).

SIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan yang sudah dilakukan maka kesimpulan yang didapatkan adalah sebagai berikut:

Penelitian ini untuk mengelompokkan beberapa puskesmas yang memiliki kasus penyakit ispa yang ada di Kabupaten Karawang menggunakan algoritma *K-Means* dengan menggunakan metodologi *Knowledge Discovery in Database (KDD)* yang terdiri beberapa tahapan yaitu, *data selection, data integration, data cleaning, data transformation, data mining, dan knowledge* atau *evaluation*. Serta menggunakan tools *RStudio* dalam melakukan proses pengolahan data.

Dalam penelitian ini menggunakan dataset penyakit ispa yang ada di beberapapuskesmas di Kabupaten Karawang tahun 2019-2021 dengan menggunakan tools *RStudio* dan mendapatkan hasil perhitungan clustering yang memiliki hasil 3 cluster yaitu cluster 1 menghasilkan 30 puskesmas dikategorikan sebagai puskesmas yang memiliki kasus penyakit ispa dalam tingkatan rendah diantaranya Anggadita, Balongsari, Bayur Lor, Ciampel, Ciampel Cicide, Curug, Gempol, JayaKerta, Jomin, Kalangsari, Karawang, KarawangKulon, KutaMukti, Kutawaluya, Lemah Duhur, Loji, Majalaya, Medang Asem, Nagasari, Pacing, PakisJaya, Pangkalan, Plawad, Rawamerta, Sukatani, SungaiBuntu, Tempuran, TunggakJati, Wanakerta. Pada cluster 2 menghasilkan 19 puskesmas dikategorikan sebagai puskesmas yang memiliki kasus penyakit ispa dalam tingkatan sedang diantaranya Adiarsa, BatuJaya, Cibuyaya, Cibuyaya, Cikampek, Cikampek Utara, Cilamaya, Jatisari, Kertamukti, Kotabaru, Lemah Abang, Pasirukem, Pedes, Purwasari, Rengasdengklok, TanjungPura, Telagasari, Telukjambe, Tirtajaya, Tirtamulya, Wadas. Dan cluster 3 menghasilkan 1 puskesmas dikategorikan sebagai puskesmas yang memiliki kasus penyakit ispa dalam tingkatan paling tinggi yaitu Klari.

Hasil dari pengujian pada kualitas *cluster* pada *dataset* Penyakit Diare Kabupaten Karawang Tahun 2017-2019 dengan data uji 50 data, pengujian kualitas *cluster* yang di uji yaitu nilai $k=4$ sampai $k=8$. Pada tahap *evaluation* menggunakan *Silhouette Coefficient* dari hasil *cluster* terbaik 5 dengan nilai *Silhouette Coefficient* 0.4960.

DAFTAR PUSTAKA

- Alfannisa Annurullah Fajrin¹, Algifanri Maulana. 2018. "100-304-1-PB." *PENERAPAN DATA MINING UNTUK ANALISIS POLA PEMBELIAN KONSUMEN DENGAN ALGORITMA FP- GROWTH PADA DATA TRANSAKSI PENJUALAN SPARE PART MOTOR 05*.
- Ananda, Lidya Rizki. 2018. "Clustering Untuk Menentukan Calon Mahasiswa Berprestasi." *Jiti* 1(2):16–19.
- Anon. n.d. *SKRIPSI PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASIFIKASI PENYAKIT ISPA*.
- Bastian, Ade, Harun Sujadi, and Gigin Febrianto. n.d. *PENERAPAN ALGORITMA K-MEANS CLUSTERING ANALYSIS PADA PENYAKIT MENULAR MANUSIA (STUDI KASUS KABUPATEN MAJALENGKA)*.
- Chandra, Muhammad Dwi, Eka Irawan, Ilham Syahputra Saragih, Agus Perdana Windarto, and Dedi Suhendro. 2021. "Penerapan Algoritma K-Means Dalam Mengelompokkan Balita Yang Mengalami Gizi Buruk Menurut Provinsi." *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer* 2(1):30–38. doi: 10.37148/bios.v2i1.19