

PENGELOMPOKAN PENYEBARAN PENYAKIT ISPA DI WILAYAH KOTA SEKAYU MENGUNAKAN ALGORITMA *K-MEANS CLUSTERING* (STUDI KASUS: RSUD SEKAYU)

Tria Anjarsari¹, Megawaty², Ade Putra³

Fakultas Teknik Ilmu Komputer, Universitas Bina Darma
yayak.januari96@gmail.com¹, megawaty@binadarma.ac.id², ade.putra@binadarma.ac.id³

ABSTRACT

The general hospital of Sekayu is one of the major hospital in the city of Sekayu. Every day the number of patients who come and do medical treatment at the hospital is relatively large. The high number of patient visits causes the number of medical records to be very large as well. So far, data that contains information about patients in hospitals has not been used properly. This information can actually be used as knowledge for hospitals, one of which is to classify the spread of ARI disease by region. In this study the Data Mining method is used to find out which clusters are formed. Using the K-Means algorithm is expected to produce a grouping of the spread of ARI disease based on the area of its spread, which can then be used as a consideration to overcome problems in the supply of drugs in the nearest pharmacy and in the prevention of similar cases and to provide more effective future services.

Keywords: *Data mining, Clustering, K-Means*

ABSTRAK

Rumah sakit umum daerah Sekayu adalah salah satu rumah sakit besar yang ada di daerah kota Sekayu. Setiap harinya jumlah pasien yang datang dan melakukan perawatan medis di rumah sakit relatif banyak. Tingginya jumlah kunjungan pasien ini menyebabkan jumlah data rekam medis menjadi sangat banyak pula. Selama ini data yang berisi informasi mengenai pasien di rumah sakit belum dimanfaatkan dengan baik. Informasi tersebut sebenarnya dapat dijadikan suatu pengetahuan bagi rumah sakit, salah satunya untuk mengelompokkan penyebaran penyakit ISPA berdasarkan wilayahnya. Pada penelitian ini digunakan metode *Data Mining* untuk mengetahui *cluster* yang terbentuk. Dengan menggunakan algoritma *K-Means* diharapkan dapat menghasilkan pengelompokan penyebaran penyakit ISPA berdasarkan wilayah penyebarannya, yang kemudian dapat dijadikan bahan pertimbangan untuk mengatasi masalah dalam persediaan obat di apotik terdekat dan dalam pencegahan kasus yang hampir sama serta untuk memberikan pelayanan ke depannya yang lebih efektif.

Kata Kunci : *Data Mining, Clustering, K-Means*

1. PENDAHULUAN

Teknologi di jaman sekarang semakin pesat dan berkembang terutama kemajuan teknologi di bidang kesehatan. Saat ini instansi tersebut memiliki media penyimpanan yang memungkinkan untuk menyimpan data dalam jumlah besar. Hal tersebut mengakibatkan jumlah data yang semakin banyak, sehingga dalam kondisi ini memiliki tuntutan agar data yang ada tidak hanya disimpan, melainkan dilakukan analisis lebih lanjut agar bisa menghasilkan sebuah informasi atau pengetahuan yang berguna bagi kehidupan manusia. Metode yang digunakan untuk memenuhi kebutuhan tersebut adalah metode *data mining*.

Data mining adalah suatu konsep yang digunakan untuk menguraikan penemuan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam *database* besar [2].

Rumah Sakit Umum Daerah Sekayu adalah salah satu rumah sakit yang memiliki jumlah pasien dalam skala besar. Setiap harinya jumlah pasien yang datang dan melakukan perawatan medis di rumah sakit terus bertambah. Bertambahnya pasien yang berobat menyebabkan penyimpanan data rekam medis yang lebih banyak lagi.

Data rekam medis dapat dimanfaatkan untuk meningkatkan kepuasan pasien terhadap rumah sakit. Dengan menggunakan metode *data mining*, data rekam medis yang pada awalnya hanya berupa tumpukan data biasa dapat dijadikan solusi untuk mencari informasi yang terkandung dalam aktifitas rekam medis tersebut. Banyak informasi yang bisa diambil dari data di rumah sakit. Seperti informasi yang dapat kita ketahui mengenai pola dalam berbagai macam penyakit. Salah satu contohnya yaitu penyakit ISPA.

Infeksi Saluran Pernapasan Akut (ISPA) adalah infeksi pada saluran pernapasan baik saluran pernapasan atas atau bawah, dan dapat menyebabkan berbagai spektrum penyakit dari infeksi ringan sampai penyakit yang parah dan mematikan, yang dipengaruhi oleh patogen penyebab, faktor lingkungan, dan faktor pejamu. Banyak sekali faktor yang dapat mempengaruhi penyakit ISPA baik itu secara langsung ataupun tidak langsung. Faktor resiko yang paling banyak menyebabkan penyakit ISPA adalah sosio-ekonomi (pendapatan keluarga, perumahan, sanitasi dan pendidikan orang tua), status gizi, tingkat pengetahuan sebuah keluarga dan faktor lingkungan (kualitas udara). Selain itu penyebab ISPA yang paling utama yaitu adanya seorang perokok dan udara yang tercemar. Daerah Sekayu sendiri menjadi salah satu kabupaten yang mengalami kabut asap pada tahun 2014 dan mengalami peningkatan bagi penderita ISPA. Banyaknya penderita ini mengakibatkan data pasien yang berobat menjadi lebih banyak [3].

Pengolahan data rekam medis di rumah sakit ini memang sudah menggunakan komputer akan tetapi data yang ada hanya disimpan dan menjadi tidak berguna. Data tersebut selalu bertambah setiap tahunnya dan tersimpan secara keseluruhan, tidak adanya pengelompokan data dalam menentukan penyebaran pada suatu jenis penyakit berdasarkan wilayahnya. Sehingga pihak rumah sakit kesulitan jika ingin mencari sebuah data mengenai wilayah penyebaran penyakit tersebut. Dengan adanya proses ini akan membantu dalam pengelompokan dan penelusuran pada data rekam medis untuk mengidentifikasi pola dalam jumlah data yang besar. Banyaknya jumlah data tersebut memungkinkan perlunya sebuah pengelompokan penyebaran jenis penyakit ISPA berdasarkan wilayahnya, dari pengelompokan yang dilakukan dapat mengetahui wilayah mana saja yang terinfeksi salah satu jenis penyakit ISPA paling tinggi. Sehingga pihak rumah sakit dapat mengatasi dan menindaklanjuti penyakit tersebut dan menambahkan persediaan obat-obatan di rumah sakit itu sendiri maupun di apotik sekitar masyarakat. Hal ini dilakukan sebagai suatu sarana untuk mengetahui, mengawasi dan menyiapkan strategi penyembuhan penyakit di wilayah tersebut serta sebagai acuan untuk tindakan selanjutnya dalam upaya pencegahan kasus yang sama agar dapat memberikan pelayanan kedepannya yang lebih efektif lagi. Kemudian hasil tersebut dapat dijadikan bahan atau dasar penyuluhan kesehatan oleh Dinas Kesehatan setempat.

Adapun dalam penelitian ini akan dibahas bagaimana cara mengimplementasikan salah satu algoritma dalam data mining, yaitu algoritma k-means clustering. *Clustering* juga disebut sebagai *segmentation*. Metode ini digunakan untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada sebuah kelompok atribut, mengelompokkan data yang memiliki kemiripan atribut. *Clustering* adalah metode data mining yang *unsupervised*, karena tidak ada satu atribut pun yang digunakan untuk memandu proses pembelajaran, jadi seluruh atribut input diperlakukan sama. Kebanyakan algoritma *clustering* membangun sebuah model melalui serangkaian pengulangan dan berhenti ketika model tersebut telah memusat atau berkumpul (batasan dari segmentasi ini telah stabil) [5]. Proses yang akan dilakukan dengan menggunakan *tools rapidminer, RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining, text mining* dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data *preprocessing* dan *visualisasi*. *RapidMiner* merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. *RapidMiner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi [1].

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Dalam melakukan penelitian ini, peneliti menggunakan metode *deskriptif*. Metode *deskriptif* adalah suatu metode yang meneliti status kelompok manusia, suatu objek, suatu kondisi, suatu pemikiran ataupun suatu kelas peristiwa pada masa sekarang. Maka dari itu peneliti menggunakan metode *deskriptif* dikarenakan permasalahan yang sedang dilakukan peneliti sekarang berdasarkan data yang sebenarnya atau yang bersifat fakta yaitu mengenai data rekam medis pada Rumah Sakit Umum Daerah (RSUD) Kota Sekayu.

2.2 Metode Pengumpulan Data

Pengumpulan data merupakan tahap penting dalam penulisan karya ilmiah. Metode pengumpulan data yang tepat yaitu dengan mempertimbangkan penggunaannya berdasarkan jenis data dan sumbernya. Data objektif dan relevan dengan pokok permasalahan penelitian adalah indikator keberhasilan dari suatu penelitian. Adapun teknik atau metode pengumpulan data dalam penelitian antara lain :

1. Studi Pustaka
Metode ini dilakukan dengan cara mempelajari buku, jurnal ilmiah, laporan penelitian, yang berkaitan dengan pemodelan yang menggunakan algoritma *k-means* yang diharapkan dapat memberikan referensi yang sesuai beserta sumber-sumber dari berbagai situs yang dapat mendukung Tugas Akhir.
2. Wawancara (*interview*)
Wawancara atau *interview* merupakan suatu pengumpulan data yang dilakukan dengan cara tanya jawab atau dialog secara langsung dengan Kepala Ruangan penyimpanan data rekam medis dan perawat pada Rumah Sakit Umum Daerah (RSUD) Kota Sekayu.
3. Pengamatan (*Observasi*)

Metode pengumpulan data melalui pengamatan secara langsung kepada objek penelitian mengenai data rekam medis yang dibutuhkan untuk penelitian langsung ke Rumah Sakit Umum Daerah (RSUD) Kota Sekayu.

2.3 Metode Analisis

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan tahapan *Knowledge Discovery in Database (KDD)* yang terdiri dari beberapa tahapan [4], yaitu sebagai berikut :

1. Data selection

Sebelum melanjutkan ke dalam data mining terlebih dahulu kita harus melakukan langkah-langkah yang berguna untuk meningkatkan kualitas hasil dari algoritma selanjutnya. Proses data mining yang pertama adalah *selection* data berfungsi untuk memilih *field* yang sesuai kebutuhan untuk mendapatkan *knowledge*, pada penelitian ini *field* data yang sesuai akan digunakan untuk mendapatkan hasil yang diinginkan. Berikut adalah data awal yang didapat dari pengambilan data pada rumah sakit umum daerah Sekayu dalam bentuk *excel* sebelum digunakan untuk melakukan seleksi data.

No. RM	Nama Pasien	Jenis Kelamin	Alamat	Desa	Tanggal Masuk	Jenis Ispa
010101	JAHMAD WIRI BUCAK	L	Desa BAILANGU MIRA SEKAYU 9071	Balango	19/01/2019	Asma
010420	PUJITO SUPRIAN	L	Desa SOAK BARU KIRI KIRI MIRA SEKAYU 9071	Soak Baru	19/01/2019	Pneumonia
010421	M. RAJIB. PRADANA SIBI PRANWAN	L	PRADIP Lais	Lais	19/01/2019	Asma
010424	M. RIFAL. HARASINDRAWA	L	Desa SUPRI PENGALAN	Serasan Jaya	19/01/2019	Asma
010441	LINTARO GABRIEL	L	TERMINAL BANGKOR Kayuara RT11 KIRI MIRA SEKAYU 9071	Kayuara	19/01/2019	TBC
010442	M. LOOQI ALFAN BIN ANKARI	L	Desa PELU TIDUR ULAN Serasan Jaya MIRA SEKAYU 9071	Serasan Jaya	19/01/2019	Asma
041484	MARISA DANIS PARADINA	L	Desa BALANG ASONO RT08 KIRI PLAKAT TENGGI	Sialang Agung	19/01/2019	Asma
010130	SEPTIYANSA RIFI ISHAN BIN BOMBEROHIME	P	Desa DAMAI E 11 Kayuara SEKAYU	Kayuara	19/01/2019	TBC
010443	SEPTIYANSA RIFI ISHAN BIN BOMBEROHIME	L	Desa SOAK BARU KIRI KIRI MIRA SEKAYU 9071	Soak Baru	19/01/2019	Asma
010438	M. FALDI MORGANA BIN IRMAH	L	Lumpatan MIRA SEKAYU 9071	Lumpatan	19/01/2019	Asma
010439	JALILA HISA ANDRIANI	P	Desa MERDEKA LK 7 Serasan Jaya RT 10 KIRI MIRA SEKAYU 9071	Serasan Jaya	19/01/2019	Asma
010437	PUTRI AUSTAR DEITA LESTARI	P	Desa 1 BALANG MARGA MIRA LAK BONGAI KIRI 9071	Lais	19/01/2019	Pneumonia
010436	BONGA ISMANTATI BT TALARI SO	P	Desa MERDEKA Serasan Jaya RT08 KIRI MIRA SEKAYU 9071	Serasan Jaya	19/01/2019	Pneumonia
010434	JAHMAD RIZKI AKBAR	L	Desa LK 7 Serasan Jaya MIRA SEKAYU 9071	Serasan Jaya	19/01/2019	TBC
010432	ADITYA AZHA RAMADHAN	L	Desa KOL WAHID UDUN LK 8 BALANG ASONO MIRA SEKAYU 9071	Balai Agung	19/01/2019	Asma
010431	M. KHARIS ARIF	L	Kayuara MIRA SEKAYU 9071	Kayuara	19/01/2019	Asma
010430	JAYU UTONGA BIN BUDALIRUDANI	L	Desa BANGKOR KIRI Kayuara Sekayu	Kayuara	19/01/2019	Pneumonia
010429	ACHSAR RAJAB SONDURI BANGSI ATAN	L	Desa LK 7 Kayuara SEKAYU	Kayuara	19/01/2019	Pneumonia
010428	RIZKI YUDITHA PRATAMA BIN SORICAL	L	Desa KARANG TANDING MIRA SEKAYU 9071	Karang Tanding	19/01/2019	Asma
010427	RAKA RIZQITA BIN HERDISEKARNO	L	Desa Pagar Kaya KATA Sanga Semb	Pagar Kaya	19/01/2019	Pneumonia
010426	JALILA HINDA KAMILAH	P	Desa 11 Lumpatan MIRA SEKAYU 9071	Lumpatan	19/01/2019	Pneumonia
010128	RENOLFA FERDINANDUS SIN STAFIUSINDARLAN	P	KOMP PERUMAHAN KEL BALANG ASONO Sekayu	Balai Agung	19/01/2019	Pneumonia
010427	MITHA WAHABANI HUSADA BT M. IRMAH	P	Desa KOL WAHID UDUN LK 8 BALANG ASONO MIRA SEKAYU 9071	Balai Agung	19/01/2019	Asma
010148	SHAFIA NELLY APRILIASTO	P	KOMP 03 BALANG ASONO Sekayu	Serasan Jaya	19/01/2019	TBC

Gambar 1. Data awal

Selanjutnya dari data awal tersebut akan di seleksi data yang hanya di perlukan sesuai dengan tujuan analisis. Adapun atribut yang akan digunakan yaitu jenis kelamin, nama desa dan jenis dari penyakit ispa yang terlihat pada gambar 2.

	A	B	C
1	Jenis Kelamin	Desa	Jenis Ispa
2	L	Balangu	Asma
3	L	Soak Baru	Pneumonia
4	L	Lais	Asma
5	L	Serasan Jaya	Asma
6	L	Kayuara	TBC
7	L	Serasan Jaya	Rinitis
8	L	Sialang Agung	Asma
9	P	Kayuara	TBC
10	L	Soak Baru	Bronkitis
11	L	Lumpatan	Bronkitis
12	P	Serasan Jaya	Asma
13	P	Lais	Pneumonia
14	P	Serasan Jaya	Pneumonia
15	L	Serasan Jaya	Faringitis
16	L	Balai Agung	Asma
17	L	Kayuara	Asma
18	L	Kayuara	Pneumonia
19	L	Kayuara	Faringitis
20	L	Karang Tanding	Asma
21	L	Pagar Kaya	Bronkitis
22	P	Lumpatan	Pneumonia
23	P	Balai Agung	Bronkitis
24	P	Balai Agung	Asma

2. Data Cleaning

Proses data mining yang selanjutnya adalah *cleaning* data yaitu proses pembersihan data yang diawali dari *procedure* pengambilan data hingga alur kerja dan proses pembersihan data yang dilakukan menggunakan *excel*. Data *cleaning*, juga dikenal sebagai data *cleansing*. Ini adalah fase dimana data-data yang tidak lengkap, mengandung error, data kosong, dan data yang tidak konsisten dibuang untuk diproses ulang untuk penggalan pengetahuan (*Discovery Knowledge*). Pada gambar 3 berikut terdapat data yang tidak lengkap atau kosong, data yang kosong tersebut akan dihapuskan.

Row No.	Jenis Kelamin	Nama Pasien	Desa	Jenis Ipa
1	L	ARWID NUR.	Balang	Alma
2	L	RICKO OPTI.	Soak Baru	Pneumonia
3	L	M. FAREL F.E.	Lato	Alma
4	L	M REVAL RA.	Serasan Jaya	Alma
5	L	LINTANG GA.	Kawara	TBC
6	L	MWESA GA.	Balang Agung	Alma
7	P	SEPTI ARANI.	Kawara	TBC
8	L	REWMANDA R.	Soak Baru	Shenitis
9	L	M FAZ NUOR.	Lumpang	Shenitis
10	P	AULIA RISLA.	Serasan Jaya	Alma
11	P	BUNDA ISLA.	Serasan Jaya	Pneumonia
12	L	ARWID RADI.	Serasan Jaya	Faringitis
13	L	ADITYA ADGA.	Balai Agung	Alma
14	L	M. KHAMRAN.	Kawara	Alma
15	L	AZWAR FALA.	Kawara	Faringitis
16	L	RUKA REGIS.	Pagar Hara	Shenitis

Gambar 3. Contoh data kosong

Setelah dilakukan pembersihan data dengan menghapus *record* atau baris yang kosong maka didapatkan hasil sebelum dilakukan pembersihan data sebanyak 1173 data bahwa terdapat 28 data yang kosong atau tidak lengkap sehingga data berubah menjadi sebanyak 1145 data setelah dilakukan pembersihan data, hasil dari tahap pembersihan data dapat dilihat pada gambar 4.

Row No.	Jenis Kelamin	Nama Pasien	Desa	Jenis Ipa
1	L	ARWID NUR.	Balang	Alma
2	L	RICKO OPTI.	Soak Baru	Pneumonia
3	L	M. FAREL F.E.	Lato	Alma
4	L	M REVAL RA.	Serasan Jaya	Alma
5	L	LINTANG GA.	Kawara	TBC
6	L	MWESA GA.	Balang Agung	Alma
7	P	SEPTI ARANI.	Kawara	TBC
8	L	REWMANDA R.	Soak Baru	Shenitis
9	L	M FAZ NUOR.	Lumpang	Shenitis
10	P	AULIA RISLA.	Serasan Jaya	Alma
11	P	BUNDA ISLA.	Serasan Jaya	Pneumonia
12	L	ARWID RADI.	Serasan Jaya	Faringitis
13	L	ADITYA ADGA.	Balai Agung	Alma
14	L	M. KHAMRAN.	Kawara	Alma
15	L	AZWAR FALA.	Kawara	Faringitis
16	L	RUKA REGIS.	Pagar Hara	Shenitis

Gambar 4. Hasil pembersihan data kosong

3. Data Transformation

Pada tahap ini data diubah atau digabung ke dalam format yang diperlukan atau data yang telah terpilih ditransformasikan ke dalam bentuk-bentuk yang sesuai untuk proses *data mining* dengan cara memberikan inisialisasi pada masing-masing data yang memiliki tipe huruf agar bisa ditransformasikan ke dalam bentuk angka. karena pada proses *clustering* hanya bisa menerima input data berupa angka numerik. Dari ketiga atribut yang digunakan untuk transformasi data yang pertama yaitu jenis kelamin, kemudian jenis kelamin tersebut diurutkan dari yang terbesar berdasarkan frekuensi pasien yang berobat. Hasil dari inisialisasi jenis kelamin dapat dilihat pada tabel 1.

Tabel 1. Transformasi data pada atribut jenis kelamin

Jenis Kelamin	Inisialisasi	Frekuensi
L	1	613
P	2	532

Pada Tabel 1 jenis kelamin yang memiliki frekuensi terbesar diberi inisial dengan angka 1 dimana jenis kelamin terbesar pada laki-laki yang memiliki jumlah 613, dan jenis kelamin dengan frekuensi paling sedikit diberi inisial angka 2 dimana jenis kelamin paling sedikit yaitu perempuan dengan jumlah 532. Selain jenis kelamin, nama desa juga termasuk ke dalam jenis data nominal sehingga perlu di inisialisasikan ke dalam bentuk angka. Seperti pada jenis kelamin, pada desa juga diberikan inisialisasi berdasarkan frekuensi terbesar. Dapat dilihat pada tabel 2.

Tabel 2. Transformasi data pada atribut desa

Desa	Frekuensi	Inisialisasi
Kayuara	174	1
Balai Agung	172	2
Serasan Jaya	93	3
Lumpatan	84	4
Lais	40	5
Epil	38	6
Soak Baru	33	7
Gajah Mati	29	8
Bailangu Timur	21	9
Toman	20	10
Rantau Panjang	19	11
Keluang	17	12
Kertayu	17	13
Petaling	17	14
Lumpatan II	17	15
Sukarami	16	16
Ulak Paceh Jaya	15	17

Pada tabel 2 hanya sebagian data yang ditampilkan dari 103 data yang ada, dalam atribut desa dimana untuk desa yang pertama berdasarkan frekuensi terbesar yaitu desa kayuara ditransformasikan ke dalam angka 1 yang memiliki jumlah data sebanyak 174, selanjutnya untuk desa yang mempunyai frekuensi terbesar kedua yaitu desa balai agung ditransformasikan ke dalam angka 2 yang memiliki jumlah data sebanyak 172. Dan seterusnya hingga desa yang memiliki frekuensi paling sedikit. Selanjutnya pada atribut ketiga yaitu jenis ispa dapat dilihat pada tabel 3.

Tabel 3. Transformasi data pada atribut jenis penyakit ispa

Jenis Ispa	Frekuensi	Inisialisasi
Nasofaringitis	192	1
Pneumonia	165	2
Asma	147	3
TBC	120	4
Bronkitis	115	5
Faringitis	99	6
Influenza	82	7
Rinitis	81	8
Tonsilitis	45	9
Laringitis	34	10
Sinusitis	32	11
Otitis	30	12
Bronkiolitis	3	13

Pada tabel 3 dalam atribut jenis ispa dimana untuk jenis ispa yang memiliki frekuensi terbesar yaitu nasofaringitis ditransformasikan ke dalam angka 1 dengan jumlah sebanyak 192 selanjutnya untuk jenis ispa yang memiliki frekuensi terbesar kedua yaitu pneumonia ditransformasikan ke dalam angka 2 dengan jumlah sebanyak 165, dan seterusnya hingga jenis ispa yang memiliki frekuensi paling sedikit.

3. HASIL DAN PEMBAHASAN

3.1 Penerapan *K-Means Clustering* Pada *RapidMiner*

Pada hasil pengujian yang telah dilakukan terdapat beberapa *output* yang dihasilkan oleh *rapidminer* yaitu:

1. *Example Set*

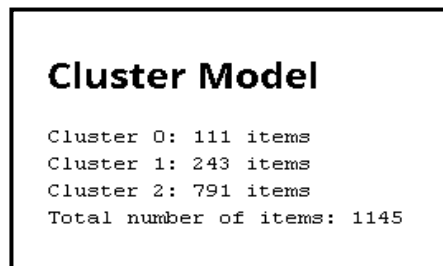
Pada *Example Set* dapat dilihat beberapa tampilan dari hasil *cluster*, yang terdapat pada *Data View*. *Data view* merupakan tampilan dari hasil *cluster* data secara keseluruhan yang sesuai dengan data yang sudah di input.

Row No.	Nama Desa	Jenis Ispa	cluster	Jenis Kotamin	Desa
1	Bailangu	3	cluster_1	1	23
2	Soak Baru	2	cluster_2	1	7
3	Lais	3	cluster_2	1	5
4	Serasan Jaya	3	cluster_2	1	3
5	Lumpatan	3	cluster_2	1	4
6	Kertayu	17	cluster_2	1	13
7	Petaling	17	cluster_2	1	14
8	Lumpatan II	17	cluster_2	1	15
9	Sukarami	16	cluster_2	1	16
10	Ulak Paceh Jaya	15	cluster_2	1	17

Gambar 5. Tampilan Hasil *Cluster* Pada *Data View*

2. *Cluster Model (Clustering)*

Pada *Cluster Model (Clustering)* dapat dilihat beberapa tampilan dari hasil *cluster*, yaitu pada *Text View* yang merupakan tampilan hasil pengelompokkan berdasarkan *cluster* dari jumlah anggotanya. Pada *cluster_0* yang memiliki jumlah sebanyak 111 *items*, *cluster_1* yang memiliki jumlah sebanyak 243 *items* dan pada *cluster_2* yang memiliki jumlah sebanyak 791 *items*. Dengan total jumlah data 1145 *items*. Dapat dilihat pada gambar 6 dibawah ini.



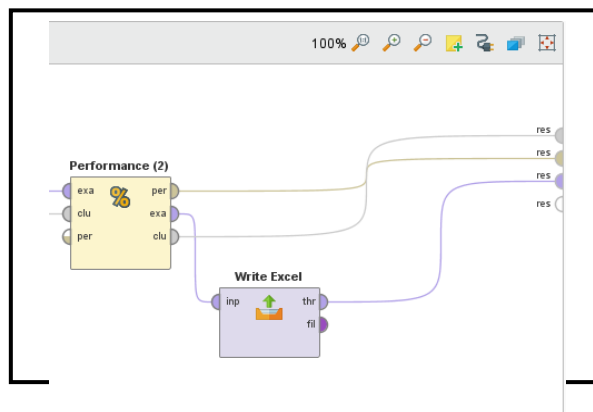
Gambar 6. Tampilan *Text View*



Gambar 7. *Result Overview*

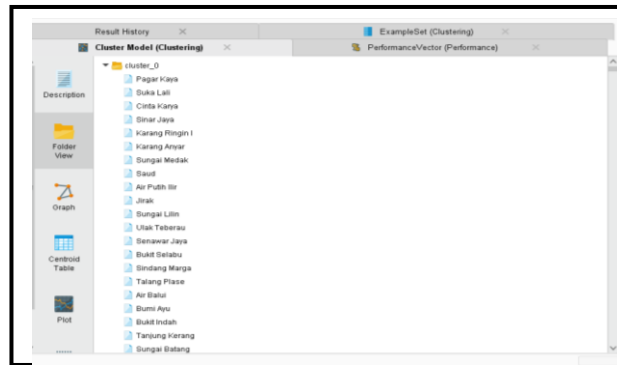
3. *Write Excel*

Write Excel disini merupakan proses melakukan *export* data ke *excel* untuk mempermudah pada saat melakukan pengolahan data dan evaluasi terhadap hasil dari *cluster* yang terbentuk saat melakukan implementasi *RapidMiner*. Dapat dilihat pada gambar 8 dibawah ini.



4. *Folder View*

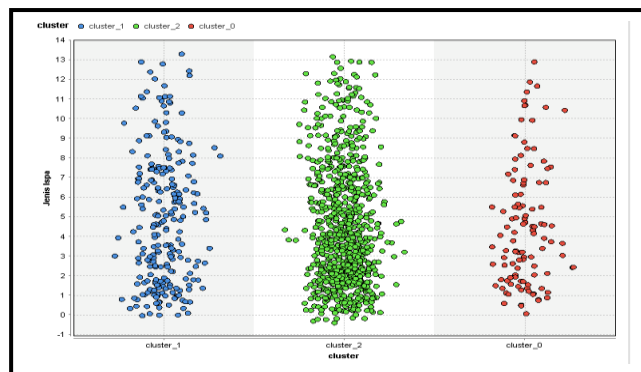
Folder view merupakan tampilan data bagian-bagian *cluster* secara keseluruhan, dimana keseluruhan masing – masing *cluster* menampilkan anggota nama desa. Dapat dilihat pada gambar 9 dibawah ini.



Gambar 9. Tampilan *Folder View*

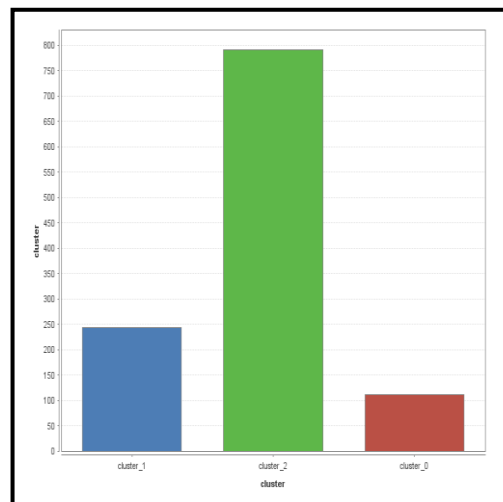
3.2 Hasil Pola *Rapidminer*

Dalam melihat pola *cluster* keluaran *rapidminer* terdapat pada tampilan *Chart* dalam *rapidminer*. *Chart* merupakan tampilan grafik hasil pengelompokkan atau *cluster* sampel data jenis kelamin, desa, dan jenis penyakit ispa yang terbentuk 3 *cluster*. Pada gambar 10 dibawah ini merupakan tampilan *Scatter* pada *Chart*.



Gambar 10. Pola *cluster* keluaran *rapidminer*

1. Pola *cluster* berwarna merah menghasilkan keluaran *rapidminer* adalah *cluster* 0 dengan penyebaran jenis penyakit ispa untuk tingkat terendah atau paling sedikit.
2. Pola *cluster* berwarna biru menghasilkan keluaran *rapidminer* adalah *cluster* 1 dengan penyebaran jenis penyakit ispa untuk tingkat sedang.
3. Pola *cluster* berwarna hijau menghasilkan keluaran *rapidminer* adalah *cluster* 2 dengan penyebaran jenis penyakit ispa tertinggi.



Gambar 11. Hasil *clustering k-means* dalam diagram batang

Pada diagram batang diatas yang memiliki hasil yang sama dimana pada diagram batang yang berwarna merah yaitu *cluster* 0 merupakan *cluster* tingkat terendah untuk penyebaran jenis penyakit ispa, dan pada

diagram batang yang berwarna biru yaitu *cluster* 1 merupakan *cluster* tingkat sedang, sedangkan pada diagram batang yang berwarna hijau yaitu *cluster* 2 merupakan *cluster* tingkat tertinggi untuk penyebaran jenis penyakit ispa.

3.3 Hasil Analisis

Berdasarkan penelitian yang telah dilakukan, telah didapatkan suatu pola, informasi dan pengetahuan baru dari proses *data mining* dalam menentukan dan pengelompokan jenis penyakit ispa berdasarkan wilayah penyebaran pada kota Sekayu yang diolah menggunakan data dari tahun 2017 sampai dengan tahun 2018 dengan melalui tahapan *knowledge discovery in database* (KDD). Dari penelitian ini telah didapat hasil *record* 1145 seperti yang telah dijelaskan pada proses sebelumnya. Telah didapat *cluster* atau pengelompokan sebanyak 3 *cluster* dimana *cluster* pertama merupakan *cluster* untuk tingkat terendah, *cluster* kedua merupakan *cluster* untuk tingkat sedang, dan *cluster* 3 merupakan *cluster* untuk tingkat tertinggi. Adapun pola persebaran dari masing-masing *cluster*, yaitu:

1. Cluster 0

Hasil pengelompokan untuk *cluster* 0 dilihat berdasarkan jenis ispa pada setiap atribut yang sudah ditransformasi kembali dari bentuk angka menjadi huruf yang sudah diurutkan berdasarkan frekuensi tertinggi. Adapun atribut yang pertama dapat dilihat pada gambar 12.

Jenis Kelamin	Inisialisasi	frekuensi
L	1	59
P	2	52
	Total	111

Gambar 12. Cluster 0 pada atribut jenis kelamin

Desa	Inisialisasi	Frekuensi
Saud	85	7
Sinar Jaya	49	3
Jirak	50	3
Sungai Lilin	51	3
Sindang Marga	52	3
Talang Plase	53	3
Air Balui	54	3
Bumi Ayu	55	3
Sungai Batang	56	3
Supat	57	3
Berlian Makmur	58	3
Tanjung Agung	59	3
Galih Sari	60	3
Ulak Embacang	61	3
Nganti	62	3
Kembang Umur	63	3
Mekar Jaya	64	3
Karang Rejo	65	3
Pagar Kaya	66	2
Karang Anyar	67	2
Sungai Medak	68	2
Ulak Teberau	69	2
Bukit Indah	70	2
Tanjung Kerang	71	2
Bayung Lencir	75	2
Lubuk Bintlalo	76	2
Purwa Agung	77	2
Peninggalan	78	2
Sungai Keruh	79	2
Batanghari Leko	80	2
Bandar Agung	81	2
Suka Lali	82	1
Cinta Karya	83	1
Karang Ringin I	84	1
Air Putih Ilir	86	1
Senawar Jaya	87	1
Bukit Selabu	88	1
Meranti	89	1
Bangun Sari	90	1
Rantai Kasih	91	1
Beji Mulyo	92	1
Keban II	93	1
Sungai Lilin Jaya	94	1
Nusa Serasan	95	1
Karya Maju	96	1
Sumber Harum	97	1
Sukamaju	98	1
Terusan	99	1
Ngulak II	100	1
Tangkal Jaya	101	1
Sri Gunung	102	1
Babat	103	1
	Total	111

Gambar 13. Cluster 0 pada atribut desa

Jenis Ispa	Inisialisasi	Frekuensi
Nasofarin	1	21
Pneumoni	2	19
Asma	3	8
TBC	4	13
Bronkitis	5	13
Faringitis	6	9
Influenza	7	6
Rinitis	8	7
Tonsilitis	9	3
Laringitis	10	6
Sinusitis	11	3
Otitis	12	3
	Total	111

Gambar 14. Cluster 0 pada atribut jenis penyakit ispa

Berdasarkan data diatas yang sudah ditransformasi dari bentuk angka menjadi huruf dan sudah diurutkan berdasarkan frekuensi tertinggi. Dalam *cluster* 0 yang merupakan *cluster* terendah yang memiliki jumlah data

sebanyak 111 data. Didapatkan hasil karakteristik pada atribut jenis kelamin laki-laki berjumlah 59 orang dan jenis kelamin perempuan berjumlah 52 orang. sehingga dapat diketahui bahwa pada *cluster* 0 penyakit ispa lebih didominasi oleh jenis kelamin laki-laki. Selain itu pada atribut desa penyakit ispa lebih mendominasi wilayah desa Saud dengan jumlah 7 orang dan pada atribut jenis penyakit ispa sendiri didominasi oleh jenis penyakit *nasofaringitis* yang berjumlah 21 orang.

2. Cluster 1

Hasil pengelompokkan untuk *cluster* 1 dilihat berdasarkan jenis ispa pada setiap atribut yang sudah ditransformasi kembali dari bentuk angka menjadi huruf yang sudah diurutkan berdasarkan frekuensi tertinggi. Adapun atribut yang pertama dapat dilihat pada gambar 15.

Jenis Kelamin	Inisialisa	frekuensi
L	1	121
P	2	122
	Total	243

Gambar 15. Cluster 1 pada atribut jenis kelamin

Desa	Inisialisasi	Frekuensi			
Sukarami	16	16	Tanah Abang	29	6
Ulak Paceh	17	15	Simpang Sari	30	6
Teluk Kijing II	18	14	Bandar Jaya	31	6
Muara Teladan	19	14	Kemang	32	6
Ngulak I	20	13	Sido Mukti	33	6
Ngulak	21	12	Rantau Kasih	34	0
Rimba Ukur	22	11	Karang Tanding	35	5
Bailangu	23	11	Danau Cala	36	5
Tebing Bulang	24	10	Babat Toman	37	5
Mangun Jaya	25	10	Macang Sakti	38	5
Sialang Agung	26	9	Sido Rahayu	39	5
Kasmaran	27	9	Rantau Sialang	40	5
Sungai Napal	28	8	Kerta Jaya	41	5
Tanah Abang	29	6	Bukit Sejahtera	42	4
Simpang Sari	30	6	Karang Tirta	43	4
Bandar Jaya	31	6	Bero Jaya	44	4
Kemang	32	6	Srimulyo	45	4
Sido Mukti	33	6	Sialang Marga	46	4
Rantau Kasih	34	0	Tanjung	47	3
Karang	35	5	Tanjung Bali	48	3
Danau Cala	36	5			
Babat Toman	37	5			
Macang Sakti	38	5			
Sido Rahayu	39	5			
Rantau Sialang	40	5			
Kerta Jaya	41	5			
Bukit Sejahtera	42	4			
				Total	243

Gambar 16. Cluster 1 pada atribut desa

Jenis Ispa	Inisialisasi	Frekuensi
Nasofaringitis	1	42
Pneumonia	2	34
Asma	3	27
TBC	4	18
Bronkitis	5	24
Faringitis	6	23
Influenza	7	20
Rinitis	8	20
Tonsilitis	9	9
Laringitis	10	9
Sinusitis	11	7
Otitis	12	9
Bronkiolitis	13	1
	Total	243

Gambar 17. Cluster 1 pada atribut jenis penyakit ispa

Berdasarkan *cluster* 1 yang merupakan *cluster* tingkat sedang memiliki jumlah data sebanyak 243 data didapatkan hasil karakteristik pada atribut jenis kelamin laki-laki berjumlah 121 orang dan jenis kelamin perempuan berjumlah 122 orang. sehingga dapat diketahui bahwa pada *cluster* 1 penyakit ispa lebih didominasi oleh jenis kelamin perempuan. Selain itu pada atribut desa penyakit ispa lebih mendominasi wilayah desa

Sukarami dengan jumlah 16 orang dan pada atribut jenis penyakit ispa sendiri didominasi oleh jenis penyakit *nasofaringitis* yang berjumlah 42 orang.

3. Cluster 2

Hasil pengelompokkan untuk *cluster* 1 dilihat berdasarkan jenis ispa pada setiap atribut yang sudah ditransformasi kembali dari bentuk angka menjadi huruf yang sudah diurutkan berdasarkan frekuensi tertinggi. Adapun atribut yang pertama dapat dilihat pada gambar 18.

Jenis Kelamin	Inisialisas	frekuensi
L	1	433
P	2	358
	Total	791

Gambar 18. *Cluster* 2 pada atribut jenis kelamin

Desa	Inialisasi	Frekuensi
Kayuara	1	175
Balai	2	172
Serasan	3	93
Lumpatan	4	83
Lais	5	40
Epil	6	38
Soak	7	33
Gajah	8	29
Bailangu	9	21
Toman	10	20
Rantau	11	19
Keluang	12	17
Kertayu	13	17
Petaling	14	17
Lumpatan	15	17
	Total	791

Gambar 19. *Cluster* 2 pada atribut desa

Jenis Ispa	Inialisasi	Frekuensi
Nasofaringitis	1	129
Pneumonia	2	112
Asma	3	112
TBC	4	89
Bronkitis	5	78
Faringitis	6	67
Influenza	7	56
Rinitis	8	54
Tonsilitis	9	33
Laringitis	10	19
Sinusitis	11	22
Otitis	12	18
Bronkiolitis	13	2
	Total	791

Gambar 20. *Cluster* 2 pada atribut jenis penyakit ispa

Berdasarkan *cluster* 2 yang merupakan *cluster* tertinggi memiliki jumlah data sebanyak 791 data didapatkan hasil karakteristik pada atribut jenis kelamin laki-laki berjumlah 433 orang dan jenis kelamin perempuan berjumlah 358 orang. Sehingga dapat diketahui bahwa pada *cluster* 2 penyakit ispa lebih didominasi oleh jenis kelamin laki-laki. Selain itu pada atribut desa penyakit ispa lebih mendominasi wilayah desa Kayuara dengan jumlah 175 orang dan pada atribut jenis penyakit ispa sendiri didominasi oleh jenis penyakit *nasofaringitis* yang berjumlah 129 orang.

Pada masing-masing *cluster* dapat diketahui bahwa jenis penyakit ispa yang paling banyak diderita di kota sekayu yaitu jenis penyakit *nasofaringitis* karena disetiap kelompok *cluster* penyakit *nasofaringitis* selalu menjadi jenis penyakit ispa tertinggi yang di alami oleh masyarakat pada masing-masing kelompok *cluster*. Selain itu, berdasarkan seluruh *cluster* diketahui bahwa *cluster* 2 merupakan *cluster* dengan tingkat paling tinggi terkena jenis penyakit ispa pada desa kayuara dan pada jenis kelamin laki-laki sehingga wilayah tersebut dapat direkomendasikan sebagai wilayah yang diutamakan dalam penyediaan stok obat penyakit ispa di apotik terdekat dan dapat menjadi fokus utama pihak dinas kesehatan untuk melakukan penyuluhan secara berkala dalam upaya mencegah dan mengurangi penyebaran penyakit ispa khususnya di wilayah kota sekayu.

4. KESIMPULAN

Pada hasil penelitian ini dapat disimpulkan bahwa terdapat beberapa bahan evaluasi yang dapat dijadikan sebagai salah satu referensi yaitu:

1. Pada penelitian menggunakan algoritma *k-means* dan berhasil membagi data menjadi tiga *cluster* yaitu *cluster 0*, *cluster 1*, dan *cluster 2*. Dimana pada *cluster 0* merupakan *cluster* dengan tingkat paling rendah atau paling sedikit penyebaran jenis penyakit ispanya, pada *cluster 1* merupakan *cluster* dengan tingkat sedang penyebaran jenis penyakit ispanya, dan pada *cluster 2* merupakan *cluster* dengan tingkat penyebaran jenis penyakit ispa tertinggi.
2. Berdasarkan hasil dari masing-masing *cluster* dapat diketahui bahwa jenis penyakit ispa yang paling banyak diderita oleh masyarakat yang tinggal dikota sekayu merupakan penyakit nasofaringitis dan desa dengan kelompok terbanyak yang terkena penyakit ispa yaitu desa kayuara untuk yang berjenis kelamin laki-laki.
3. Berdasarkan analisis tersebut maka data yang dihasilkan dari proses *clustering* diharapkan dapat membantu menanggulangi penyebaran penyakit ispa pada kota Sekayu.
4. Berdasarkan *cluster 2* sebagai *cluster* tertinggi penyebaran jenis penyakit ispa bahwa wilayah yang diutamakan dalam penyediaan stok obat penyakit ispa adalah wilayah desa kayuara dan dapat menjadi fokus utama pihak dinas kesehatan dikota Sekayu dalam melakukan penyuluhan dan konseling secara berkala dalam upaya pencegahan lebih dini dan mengurangi penyebaran penyakit ispa khususnya di wilayah kota sekayu serta dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan penyediaan stok obat.
5. Dari penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma *k-means* dapat diterima untuk pengklasteran data tersebut.

DAFTAR PUSTAKA

- [1] Ependi, U., & Putra, A. (2019). Solusi prediksi persediaan barang dengan menggunakan algoritma apriori (studi kasus: regional part depo auto 2000 Palembang). *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 5(2), 139-145.
- [2] Ambarwati, L dkk. 2013. Belajar *Data Mining* Dengan *RapidMiner*. Diakses dari https://www.academia.edu/7712860/Belajar_Data_Mining_dengan_RapidMiner, tanggal 15 Januari 2018.
- [3] Kusriani dan Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*. Diakses dari https://books.google.co.id/books?id=Ojclag73O8C&pg=PP1&lpg=PA3&focus=viewport&dq=data+mining+adalah&hl=id&output=html_text, tanggal 21 Februari 2018.
- [4] Lebuhan, A.W & Somia, A. 2017. *Faktor Yang Berhubungan Dengan Infeksi Saluran Pernapasan Akut Pada Siswa Taman Kanak-Kanak Di Kelurahan Dangin Puri Kecamatan Denpasar Timur Tahun 2014*. *Jurnal Medika*, Vol. 6, No. 6 (Juni 2017).
- [5] Sigit, P. 2013. Tahapan KDD (Knowledge Discovery in Database). Diakses dari http://www.sigitprabowo.id/2013/04/data-mining-tahap-tahapan_knowledge.html, pada tanggal 2 maret, 2018.
- [6] Vlandari, R.T. 2017. *Data Mining Teori dan Aplikasi RapidMiner*. Yogyakarta:Gava Media