

Penggunaan Metode Algoritma Apriori Dalam Menentukan Pola Pesebaran Penyakit Menular Di Kabupaten Kampar

Fenty Kurnia Oktorina ¹⁾

¹⁾ Teknik Informatika Politeknik Kampar
Jl. Tengku Muhammad Km 12 Bangkinang, Kampar, Riau-Indonesia
email : fenty@poltek-kampar.ac.id

Intisari— Dinas Kesehatan di Kabupaten Kampar memiliki banyak data tentang kesehatan masyarakat, terutama untuk penyakit menular. Diperlukan upaya dalam menganalisa data agar bisa menjadi informasi yang bermanfaat seperti adanya pola pesebaran penyakit menular. Tujuan dari penelitian ini adalah menganalisa pola pesebaran penyakit menular di Kabupaten Kampar menggunakan metode algoritma apriori. Metode ini memiliki 2 tahap penyelesaian, dimana tahap pertama mencari pola frekuensi tertinggi, dengan cara mencari kombinasi item yang memenuhi syarat minimum dari nilai support yang telah ditentukan, yaitu 20%. Tahap kedua adalah pembentukan aturan asosiatif dari pola frekuensi tinggi yang didapat pada tahap 1, dengan nilai minimum untuk confidence adalah 70%. Adapun pola pesebaran penyakit menular yang telah ditemukan sebanyak 3 pola dengan melibatkan penyakit TB Paru, Pneumonia dan diare. Confidence yang dihasilkan ketiga pola tersebut berada di rentang 80-100%. Berdasarkan hal tersebut, maka dapat disimpulkan bahwa algoritma apriori dapat digunakan untuk menentukan pola pesebaran penyakit menular.

Kata Kunci: algoritma apriori, penyakit menular, Kampar, aturan asosiatif, data mining

Abstract— Health Service in Kampar regency has lot of data on public health, especially for infectious diseases. It takes effort in analyzing the data to be useful information such as the pattern of infectious disease spread. The aim of this study is to analyze the pattern of infectious disease spread in Kampar regency using apriori algorithm method. This method has two stages of completion, where the first stage is to determine frequent itemset by finding all the itemsets that meet the minimum requirement of the support predetermined which is 20%. The next stage is to generate association rules from frequent itemsets found in the previous stage, which is minimum confidence value is 70%. As a result, 3 patterns have been found involving pulmonary tuberculosis, pneumonia and diarrhea. The pattern formed has a range of confidence between 80-100%. Therefore, this study shows that apriori algorithm method can be used to determine the pattern of infectious disease spread.

Keywords— infectious disease, apriori algorithm, kampar, data mining, association rule

I. PENDAHULUAN

Masalah kesehatan dan bahaya kematian di Indonesia, sebagai negara berkembang menjadi sangat kompleks, tidak hanya di kota-kota besar namun juga di pelosok daerah. Faktor kebiasaan, lingkungan, sosial, adat, budaya dan teknologi mejadi faktor utama penyebab penyakit. Pertumbuhan populasi yang pesat, peningkatan angka kemiskinan dan terjadinya urbanisasi telah memfasilitasi muncul dan berkembangnya penyakit menular baik penyakit baru, maupun penyakit lama yang muncul kembali (Arias, 2010).

Pengendalian penyakit menular umumnya dilakukan dengan cara menghilangkan sumber penyakit. Strateginya secara umum adalah kasus penyakit menular ditemukan secara proaktif untuk kemudian dilakukan pengobatan hingga sembuh. Namun implementasi di lapangan selalu menghadapi berbagai kendala, yang terkait dengan pelaksanaan surveilans dan upaya penanggulangan. Surveilans adalah proses pengumpulan, pengolahan, analisis dan interpretasi data secara sistemik dan berkelanjutan serta penyebaran informasi

kepada unit yang membutuhkan untuk mengambil tindakan. Kinerja surveilans terkait pada faktor tenaga, pengetahuan, dana, SOP, Sarana dan data. Beberapa dari faktor tersebut menjadi penyebab rendahnya kinerja para surveilans. Masalah ketenagaan dalam kegiatan surveilans dialami oleh semua level baik di Provinsi, kabupaten / kota hingga puskesmas. Minimnya jumlah SDM dan beban kerja yang tinggi merupakan permasalahan yang ditemukan disemua tingkatan. Begitu juga dengan ketersediaan data yang belum dianalisa sehingga bisa menjadi informasi yang berguna dalam merumuskan kegiatan penanggulangan penyakit menular secara optimal. (Atmawikarta, Hadiat, Ratman, & Tresna, 2006).

Permasalahan tersebut juga dialami Kabupaten Kampar yang ada di Provinsi Riau. Data yang didapat dari hasil surveilans belum dikemas menjadi sebuah informasi yang berguna dalam menanggulangi penyakit menular. Padahal, jika data yang ada diolah atau dianalisa, maka pola pesebaran penyakit menular dapat diketahui. Dengan diketahuinya pola pesebaran penyakit menular, maka peluang untuk melakukan pencegahan dapat segera dilakukan. Kewaspadaan dini juga

dapat ditingkatkan, sehingga penyebaran penyakit menular tidak menyebar ke daerah yang lebih luas lagi. Diharapkan hasil penelitian ini dapat membantu dinas kesehatan kabupaten kampar dalam melakukan upaya pencegahan penyakit menular.

Adapun tujuan dari penelitian ini adalah untuk menganalisa pola persebaran penyakit menular di Kabupaten Kampar dengan menggunakan algoritma apriori. Metode ini telah banyak digunakan dalam penelitian karena metode ini berfungsi untuk mencari pola hubungan antar satu atau lebih item dalam suatu dataset. Sebelumnya, penulis melakukan penelitian tentang pengolahan data mining menggunakan algoritma apriori untuk menemukan hubungan antara data induk mahasiswa dengan prestasi akademik mahasiswa politeknik kampar.

Algoritma apriori juga ini pernah digunakan untuk mendapatkan pola kecelakaan lalu lintas (Ariwibowo, 201), berdasarkan jenis kendaraan dan merek. Selain itu, algoritma apriori juga diterapkan untuk memprediksi persediaan barang di waktu yang akan datang di (Kennedy Tampubolon, 2013). Pada tahun 2012, algoritma Apriori juga telah diterapkan untuk pengembangan sistem rekomendasi penelusuran buku sehingga menghasilkan rekomendasi buku berdasarkan transaksi peminjaman buku yang ada di Badan Perpustakaan dan Kearsipan Provinsi Jawa Timur (Nugroho Wandu, 2012).

1. Penyakit Menular

Penyakit menular menjadi permasalahan serius bagi negara berkembang, termasuk Indonesia. Suatu penyakit bersifat multifaktorial dimana terdapat keterkaitan yang kompleks antara berbagai macam agen, penjamu dan lingkungan atau dikenal juga dengan istilah segitiga epidemiologik (Arias, 2010).

Menurut Menteri Kesehatan Indonesia (Indonesia, 2014), berdasarkan cara penularannya, penyakit menular ini dikelompokkan menjadi dua jenis, yaitu penyakit menular langsung dan penyakit tular vektor dan binatang pembawa penyakit. Penyakit menular langsung terdiri dari Difteri; Pertusis; Tetanus; Polio; Campak; Typhoid; Kolera; Rubella; Yellow Fever; Influenza; Meningitis; Tuberkulosis; Hepatitis; penyakit akibat Pneumokokus; penyakit akibat Rotavirus; penyakit akibat Human Papiloma Virus (HPV); penyakit virus ebola; MERS-CoV; Infeksi Saluran Pencernaan; Infeksi Menular Seksual; Infeksi Human Immunodeficiency Virus (HIV); Infeksi Saluran Pernafasan; Kusta; dan Frambusia.

Jenis penyakit tular vektor dan binatang pembawa penyakit terdiri atas: Malaria; Demam Berdarah; Chikungunya; Filariasis dan Kecacingan; Schistosomiasis; Japanese Encephalitis; Rabies; Antraks; Pes; Toxoplasma; Leptospirosis; Flu Burung (Avian Influenza); dan West Nile (Indonesia, 2014).

2. Algoritma Apriori

Algoritma apriori merupakan algoritma dasar yang digunakan untuk menentukan frequent itemsets dengan menggunakan teknik association rule yang merupakan teknik dari data mining. Teknik tersebut berfungsi untuk menemukan

aturan asosiatif antara suatu kombinasi item. Teknik analisa asosiasi yang menghasilkan algoritma yang efisien adalah analisa pola prekuensi tinggi atau Frequent Pattern Mining (Pratama, Bettiza, & Matulatan).

Algoritma apriori dirancang untuk mengoperasikan database yang terdiri dari banyak transaksi yang memiliki beberapa item atau disebut itemset. Algoritma apriori mengidentifikasi beberapa subset dari beberapa itemset dalam database tersebut. Apriori menerapkan pendekatan bottom-up dimana subset yang sering muncul akan menghasilkan satu item pada satu waktu, atau yang dikenal dengan istilah candidate generation. Kelompok dari kandidat tersebut diuji terhadap data (Tanna & Ghodasara, 2013).

Adapun langkah-langkah dalam mengerjakan algoritma apriori adalah sebagai berikut (Nofriansyah, 2014):

a. Analisis pola frekuensi tinggi

Pencarian kombinasi item yang memenuhi syarat minimum dari nilai support dalam database.

$$Support = (A \cap B) = \frac{\sum \text{Transaksi yang mengandung A dan B}}{\sum \text{Transaksi}} * 100\%$$

b. Pembentukan aturan asosiasi

Setelah semua pola frekuensi tinggi ditemukan, aturan asosiatif yang memenuhi syarat minimum untuk confidence.

$$Confidence = \frac{\sum \text{Transaksi yang mengandung A dan B}}{\sum \text{Transaksi mengandung A}} * 100\%$$

II. METODOLOGI PENELITIAN

Metode yang digunakan pada penelitian ini dibagi dalam 3 tahap, yaitu : pengumpulan data, persiapan analisa data, serta analisa data. Adapun uraiannya adalah sebagai berikut:

A. Tahap Pertama : Pengumpulan data

Tahap ini dimulai dengan mendatangi Dinas Kesehatan Kabupaten Kampar untuk berdiskusi dan melihat kondisi data yang tersedia. Hasil diskusi menunjukkan bahwa tidak semua data yang dibutuhkan tersedia di dinas tersebut. Untuk mendapatkan data yang lebih lengkap, tim pengumpulan data harus mendatangi puskesmas-puskesmas yang ada di seluruh kecamatan yang ada di Kabupaten. Data yang didapat dari lapangan berupa jenis penyakit menular dimulai dari tahun 2014 hingga 2016, untuk TB Paru, Filarias, Penumonia Bayi, Kusta, DBD, Malaria untuk masing-masing kecamatan yang ada di Kabupaten Kampar.

B. Tahap Kedua : Persiapan Analisa Data

Pada tahap ini, data cleaning akan diterapkan terhadap data yang diperoleh dari pengumpulan data dari Dinas Kesehatan Kabupaten Kampar serta puskesmas-puskesmas diseluruh kecamatan yang ada di kabupaten kampar. Data cleaning adalah proses menghilangkan dan meminimalisasi

kesalahan pada data. Kesalahan yang terjadi antara lain terdapatnya nilai yang tidak lengkap / kosong, nilai yang rangkap atau nilai yang tidak konsisten. setelah data cleaning selesai dilakukan, data tersebut selanjutnya akan dikelompokkan sehingga terbentuklah kandidat itemset yang siap untuk diolah.

C. Tahap Ketiga : Analisa Data

Analisa data dilakukan dengan menggunakan algoritma apriori. Melalui algoritma apriori, pola hubungan antar satu atau lebih item penyakit dalam suatu dataset dapat dicari. Berdasarkan data yang diperoleh dari hasil data cleaning, adapun jenis penyakit menular yang dapat dianalisa adalah sebanyak 4 item yaitu TB Paru, Pneumonia Bayi, DBD, Diare. Sementara itu, himpunan dibentuk berdasarkan item penyakit untuk 3 tahun terakhir, dimana transaksi dilakukan untuk seluruh kecamatan di Kabupaten Kampar dengan total 21 transaksi di setiap tahunnya.

Langkah-langkah penyelesaian yang dilakukan dalam analisa data menggunakan algoritma apriori adalah sebagai berikut:

1. Analisa Pola Frekuensi Tinggi

Analisa ini melakukan pencarian kombinasi *item* yang memenuhi syarat minimum dari nilai *support*. Hasil dari pencarian tersebut akan didapatkan *frequent itemset* atau pola frekuensi tinggi. Untuk dapat memulai analisa, langkah pertama yang harus dilakukan adalah menentukan *minimum support* yang merupakan nilai minimal dari jumlah / frekuensi kemunculan sebuah item dalam sebuah keranjang (*support*) yang harus dipenuhi. Dalam penelitian ini, ditetapkan *minimum support* adalah 20% untuk seluruh himpunan.

Langkah selanjutnya adalah mencari pola frekuensi tinggi. Berdasarkan hasil clearing data yang dilakukan pada tahap sebelumnya, maka dapat dilihat bahwa dari 21 transaksi (kecamatan), terdapat beberapa daerah yang terjangkit penyakit menular, dimana datanya dirangkum pada tabel 1

TABEL 1

JUMLAH PENGIDAP PENYAKIT MENULAR YANG AKAN DIANALISA

Nama Penyakit	Itemset	Jumlah Transaksi		
		2014	2015	2016
TB Paru	1	15	17	17
Pneumonia	2	15	11	13
Demam Berdarah	3	2	5	7
Diare	4	21	21	21

Berdasarkan data pada tabel tersebut diatas, maka akan dilakukan analisa untuk mendapatkan pola frekuensi tinggi. Pencarian pola frekuensi tinggi akan dilakukan untuk 1-itemset hingga tidak ditemukannya lagi pola frekuensi tinggi, dimana dalam penelitian ini terjadi iterasi hingga 3-itemset. Sementara itu, Berdasarkan data yang terdapat pada tabel 1, maka dilakukan iterasi pertama untuk kombinasi 1-itemset, sehingga ditemukan kandidat frekuensi tinggi seperti yang terlihat pada tabel 2

TABEL 1

KANDIDAT FREKUENSI TINGGI DARI KOMBINASI 1-ITEMSET

Items et	2014		2015		2016	
	Count	Support	Count	Support	Count	Support
{1}	15	71%	17	81%	17	81%
{2}	15	71%	11	52%	13	62%
{3}	2	10%	5	24%	7	33%
{4}	21	100%	21	100%	21	100%

Kandidat frekuensi tinggi yang tidak memenuhi minimum support yang telah ditetapkan, yaitu 20% tidak akan disertakan pada iterasi kedua untuk pencarian kombinasi 2-itemset. Tabel 3 menampilkan data kandidat frekuensi tinggi yang didapat dari iterasi kedua, yang merupakan hasil pencarian terhadap kombinasi 2-itemset.

TABEL 2

KANDIDAT FREKUENSI TINGGI DARI KOMBINASI 2-ITEMSET.

Itemset	2014		2015		2016	
	Count	Support	Count	Support	Count	Support
{1, 2}	12	57,14%	11	52,38%	10	47,62%
{1, 3}	-	-	5	23,81%	5	23,81%
{1, 4}	15	71,43%	17	80,95%	17	80,95%
{2, 3}	-	-	3	14,29%	4	19,05%
{2, 4}	15	71,43%	11	52,38%	13	61,90%
{3, 4}	-	-	5	23,81%	7	33,33%

Berdasarkan data yang terdapat pada tabel 4.3, maka dapat ditetapkan frekuensi tinggi dengan mencari itemset yang memenuhi minimum support untuk pembentukan kombinasi 3-itemset. Hasil kombinasi 3-itemset yang terbentuk dapat dilihat pada tabel 4

TABEL 3

KANDIDAT FREKUENSI TINGGI DARI KOMBINASI 3-ITEMSET

Itemset	2014		2015		2016	
	Count	Support	Count	Support	Count	Support
{1, 2, 3}	-	-	3	14,29%	2	9,52%
{1, 2, 4}	12	57,14%	11	52,38%	10	47,62%
{1, 3, 4}	-	-	-	-	-	-
{2, 3, 4}	-	-	3	14,29%	4	19,05%

Data yang terdapat pada tabel tersebut diatas menunjukkan bahwa kombinasi 3-itemset {1, 3, 4} tidak muncul di setiap tahun yang ada. Hal ini dikarenakan 2-itemset yang mengandung item {3} tidak memenuhi minimal support untuk dilakukan iterasi 3-itemset.

2. Pembentukan Aturan Asosiatif

Setelah semua pola frekuensi tinggi ditemukan, maka tahap selanjutnya adalah mencari aturan asosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiatif $AB \rightarrow C$. Dalam penelitian ini ditetapkan minimum confidence sebesar 70%.

Adapun aturan asosiatif yang terbentuk dapat dilihat pada tabel berikut ini:

TABEL 5

ATURAN ASOSIATIF DARI KOMBINASI 3-ITEMSET

Pola Kombinasi 3-itemset	2014		2015		2016	
	Support	Confidence	Support	Confidence	Support	Confidence
{1, 2} → {4}	57,14%	100,00%	52,38%	100,00%	47,62%	100,00%
{4, 2} → {1}	57,14%	80,00%	52,38%	100,00%	47,62%	76,92%
{4,1} → {2}	57,14%	80,00%	52,38%	64,71%	47,62%	58,82%

III. HASIL DAN PEMBAHASAN

Hasil percobaan sebaiknya ditampilkan dalam berupa grafik atau pun tabel. Untuk grafik dapat mengikuti format untuk diagram dan gambar.

Berdasarkan data yang terdapat pada tabel 2 hingga tabel 4 yang berisikan kandidat frekuensi tinggi dari hasil iterasi 1, iterasi 2 dan iterasi 3, maka didapatkan frekuensi tinggi untuk masing-masing iterasi sebagai berikut:

- c. Iterasi 1 (Kombinasi 1-itemset)
 - Tahun 2014 : I = {{1}, {2}, {4}}
 - Tahun 2015 : I = {{1}, {2}, {3}, {4}}
 - Tahun 2016 : I = {{1}, {2}, {3}, {4}}
- d. Iterasi 2 (Kombinasi 2-itemset)
 - Tahun 2014 : I = {{1, 2}, {1, 4}, {2, 4}}
 - Tahun 2015 : I = {{1, 2}, {1, 3}, {1, 4}, {2, 4}, {3, 4}}
 - Tahun 2016 : I = {{1, 2}, {1, 3}, {1, 4}, {2, 4}, {3, 4}}
- e. Iterasi 3 (Kombinasi 3-itemset)
 - Tahun 2014 : I = {1, 2, 4}
 - Tahun 2015 : I = {1, 2, 4}
 - Tahun 2016 : I = {1, 2, 4}

Dari uraian diatas, dapat dilihat dengan jelas bahwa pola frekuensi tinggi yang terbentuk pada iterasi 3 di setiap tahunnya adalah sama, yaitu {1, 2, 4}. Dari pola frekuensi tinggi ini, aturan asosiatif {A, B} → {C} dibentuk, sehingga menghasilkan 3 pola yaitu {1, 2} → {4}; {4, 2} → {1}; dan {4, 1} → {2}. Adapun minimum confidence yang telah ditetapkan adalah sebesar 70%.

Dari 3 pola yang terbentuk pada masing-masing tahun, pola yang digunakan adalah yang memiliki nilai *support* dan *confidence* yang tertinggi, sehingga terbentuklah aturan sebagai berikut:

- {1, 2} → {4}; *support* 57,14%; *confidence* 100%; dimana artinya penyakit pada itemset {1,2} mendominasi sebesar 57,14% dari seluruh wilayah di kabupaten Kampar. Selain itu, penyakit pada itemset {4} memiliki kecenderungan untuk terjadi sebesar 100% pada wilayah yang telah terjangkit penyakit pada itemset {1,2}.
- {4, 2} → {1}; *support* 57,14%; *confidence* 100%; aturan ini memiliki arti bahwa penyakit pada itemset {4,2} menjangkiti 57,14% wilayah kabupaten kampar, dengan kecenderungan sebesar 100% terjangkitnya penyakit pada itemset {4} di suatu wilayah, jika penyakit pada itemset {4,2} telah terjangkit di wilayah tersebut.

• {4, 1} → {2}; *support* 57,14%; *confidence* 80%; terbentuk pola dimana penyakit pada itemset {4,1} tersebar di 57,14 % wilayah di kabupaten Kampar, dimana terdapat kecenderungan sebesar 80% untuk penyakit pada itemset {2} akan menjangkiti suatu wilayah yang terdapat penyebaran penyakit pada itemset {4,1} di wilayah tersebut.

Adapun itemset {1} adalah penyakit TB Paru, itemset {2} untuk penyakit Pneumonia dan itemset {4} untuk penyakit Diare.

IV. KESIMPULAN

Adapun kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut:

1. Analisa terhadap pola pesebaran penyakit telah selesai dilakukan dan mendapatkan hasil berupa adanya 3 (tiga) pola yang terbentuk berdasarkan data dari tahun 2014 hingga 2016.
2. Ketiga pola tersebut melibatkan penyakit TB Paru, Pneumonia, dan Diare, dimana ketiga penyakit ini menjangkiti sebesar 57,14% dari seluruh wilayah Kabupaten Kampar secara bersamaan.
3. Kemungkinan terjadinya ketiga penyakit tersebut menyerang sebuah wilayah di Kabupaten Kampar adalah sebesar 80-100%.

REFERENSI

Arias, K. M. (2010). *Investigasi dan Pengendalian Wabah di Fasilitas Pelayanan Kesehatan*. Jakarta: EGC.

Ariwibowo, A. S. (201). Analisa Asosiatif Data Mining Untuk Mengetahui Pola Kecelakaan Lalu Lintas. *Telematika* , 8 (2), 117.

Atmawikarta, A., Hadiat, Ratman, D. R., & Tresna, Y. D. (2006). *Kajian kebijakan Penanggulangan (Wabah) Penyakit Menular*. Jakarta: Direktorat Kesehatan dan GIzi Masyarakat.

Indonesia, P. M. (2014). SK Menteri Kesehatan No 82 Tahun 2008 tentan Penanggulangan Penyakit Menular. Pekanbaru: Peraturan Menteri Kesehtan.

Kennedi Tampubolon, H. S. (2013). Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan. *Informasi dan Teknologi Ilmiah* , 1 (1), 93.

Nofriansyah, D. (2014). *Konsep Data Mining Vs Sistem Pendukung Keputusan* (1 ed.). Yogyakarta: Deepublish.

Nugroho Wandu, R. A. (2012). Pengembangan Sistem Rekomendasi Penelusuran Buku dengan Penggalian Association Rule Menggunakan Algoritma Apriori (Studi

Kasus Badan Perpustakaan dan Kearsipan Provinsi Jawa Timur). *Jurnal Teknik ITS* , 1, A-445.

Pratama, H. C., Bettiza, M., & Matulatan, T. (n.d.). *Penerapan Algoritman Apriori Dalam menemukan Hubungan Data Awal Masuk Mahasiswa Dengan Prestasi Akademik*. Tanjung Pinang.

Tanna, P., & Ghodasara, T. (2013). Foundation for Frequent Pattern Mining Algorithms Impementation. *International Journal of Computer Trends and Technology* , 4 (7), 2159.

Fenty Kurnia Oktorina memperoleh Gelar ST dari STMIK Jabar pada tahun 2002, kemudian pada tahun 2011 memperoleh M.Sc dari Glasgow Caledonian University, Scotland, United Kingdom. Saat ini sebagai Staf Pengajar program studi Teknik Informatika di Politeknik Kampar