

# **ANALISIS POLA ASOSIASI DAN SEKUENSIAL DATA REKAM MEDIS RSUD DR. H. SLAMET MARTODIRDJO PAMEKASAN DENGAN TEKNIK DATA MINING MENGGUNAKAN ALGORITMA APRIORI**

**Nilam Ramadhani<sup>1)</sup>, Badar Said<sup>2)</sup>**

<sup>12</sup>Jurusan Teknik Informatika Fakultas Teknik, Universitas Madura

Jl.Raya Panglegur km 3.5 Pamekasan

Telp. (0324) 322231 – 325786 Fax. 327418

E-mail : nilam\_ramadhani@yahoo.com

---

## ***Abstract***

*Public health service levels in Dr. H. Slamet Martodirdjo hospital increasing with the number of population, weather changes, and living habits of Maduranese people. It causes an increasing the number of patients that can affect to the volume of service provision such as medical personel, medicines, and other facilities. To cope with these impacts, it is necessary to estimate by using the patient's medical record data periodically. Patient medical record data model needed to get a representation of the disease pattern. The linkage pattern can be determined by applying data mining using Apriori-based algorithms. Input attributes used in both methods are Patient\_ID, Check\_Date, and Disease\_Diagnosys. Medical record data were used in 2010, 2011 and 2012. The experiments result showed patient disease patterns can be known in a 3 years period. It also can calculate the range percentage for each cases of the disease occurs. The mining results are expected to be utilized by the hospital management as a study to analyze the problems and needs of patients seeking treatment.*

**Keywords :** *data mining, medical record, association rule, sequential pattern, apriori.*

## ***Abstrak***

*Tingkat kebutuhan layanan kesehatan masyarakat di RSUD Dr. H. Slamet Martodirdjo Pamekasan semakin hari kian meningkat seiring bertambahnya jumlah penduduk, perubahan cuaca yang tidak menentu, dan pola hidup masyarakat Madura. Hal ini akan berdampak pada peningkatan jumlah pasien yang dapat memengaruhi volume penyediaan pelayanan seperti tenaga medis, obat-obatan, fasilitas dan lain-lain. Untuk menanggulangi dampak tersebut, perlu perhitungan/estimasi yang lebih spesifik dengan memanfaatkan data rekam medis pasien secara berkala. Pemodelan data rekam medis pasien diperlukan untuk mendapatkan gambaran pola penyakit pasien. Keterkaitan pola penyakit yang diderita oleh pasien dapat diketahui dengan menerapkan algoritma data mining berbasis Apriori. Pada kedua metode atribut input yang digunakan adalah Pasien\_ID, Tanggal\_Periksa, dan Diagnosa\_Penyakit. Data rekam medis yang dipakai adalah data pada tahun 2010, 2011, dan 2012. Dari serangkaian percobaan dengan penentuan jumlah minimum support yang digunakan, hasil penelitian menunjukkan bahwa dalam kurun waktu 3 tahun dapat diketahui pola penyakit yang diderita oleh pasien. Selain itu, dapat diketahui juga kisaran prosentase untuk masing-masing kasus penyakit yang terjadi. Hasil mining diharapkan dapat dimanfaatkan oleh pihak manajemen rumah sakit sebagai bahan kajian untuk menganalisis permasalahan dan kebutuhan terkait pasien yang berobat.*

**Kata kunci :** *penggalan data, data rekam medis, association rule, sequential pattern, apriori.*

## **1. PENDAHULUAN**

Menyambut datangnya era reformasi, teknologi dan otonomi daerah (otoda) yang mengutamakan profesionalisme, keterbukaan dan peningkatan kualitas, maka peningkatan pelayanan bidang kesehatan khususnya di Rumah Sakit Umum Daerah (RSUD) Dr.H.Slamet Martodirdjo di Kabupaten Pamekasan sudah seharusnya menggunakan strategi dan manajemen berbasis teknologi informasi.

Beberapa permasalahan yang muncul adalah penentuan pembelian fasilitas medis, obat-obatan dan sumber daya manusia. Hal ini disebabkan oleh minimnya informasi yang dapat digali dari data pasien dan penyakit dalam kurun waktu tertentu yang terekam di *database* rumah sakit. Penerapan teknik *data mining* khususnya *Association Rules*

*Mining* dan *Sequential Pattern Mining* dapat membantu menemukan pengetahuan berupa aturan dan pola sekuensial yang dapat digali dari data rekam medis pasien. Aturan dapat mengetahui hubungan sebab akibat pada penyakit, sedangkan pola sekuensial dapat mengetahui kecenderungan sekelompok kasus penyakit yang terjadi dan prosentasenya.

*Association Rule* dengan algoritma *Apriori* [1][2] merupakan salah satu bentuk terapan *data mining* yang menghasilkan model pengetahuan berupa aturan dengan nilai *confidence*. Model pengetahuan tersebut dapat digunakan untuk memprediksi kecenderungan data yang akan datang. *Association Rule* dengan algoritma *Apriori* memiliki dua tahap utama, yaitu menemukan semua *large itemset* dan membentuk aturan yang memiliki nilai *confidence*.

*Sequential Pattern Mining* dengan algoritma *AprioriAll* merupakan perluasan dari algoritma *association rule*. Karakteristik algoritma *Apriori All* secara umum sama dengan algoritma *association rule* karena keduanya berbasis *apriori*. *Sequential Pattern Mining* menghasilkan model pengetahuan berupa aturan dengan bentuk *maximal sequence*. Model pengetahuan tersebut dapat digunakan untuk mengetahui pola data pada periode yang berbeda. *Sequential Pattern Mining* dengan algoritma *AprioriAll* memiliki lima fase/ tahap utama, yaitu [3] 1. *Sort Phase*, 2. *Large Itemset Phase*, 3. *Transformation Phase*, 4. *Sequence Phase*, dan 5. *Maximal Phase*.

Penelitian sebelumnya pada bidang kesehatan yang menerapkan teknik *data mining* pada umumnya hanya membahas pencarian pola pada penyakit-penyakit tertentu. Misalnya penelitian yang dilakukan oleh Ruijuan Hu, China tahun 2010 [5] yang menggunakan algoritma *Apriori* untuk menemukan pola penyakit kanker payudara. Sedangkan penyakit di wilayah yang berbeda atau berjauhan memiliki perbedaan pola keterkaitan. Hal ini disebabkan beberapa faktor diantaranya keadaan geografis, perubahan cuaca, jumlah penduduk, dan pola hidup atau kebiasaan masyarakat setempat.

Oleh karena itu, sangat penting dilakukan penelitian untuk mengetahui pola penyebaran dan keterkaitan penyakit di wilayah yang lebih spesifik dalam rangka mencari solusi yang memanfaatkan data rekam medis pasien rumah sakit di Kabupaten Pamekasan. Data yang akan diolah adalah data rekam medis dari RSUD dr.H.Slamet Martodirdjo Pamekasan. Pola atau tren data yang sangat banyak pada kurun waktu beberapa tahun terakhir yang akan dijadikan informasi bahan pertimbangan kebijakan dan prediksi bidang kesehatan lainnya.

Tujuan dari penelitian ini adalah membuat pemodelan data rekam medis yang telah diolah menggunakan kedua algoritma berbasis *Apriori*, dan mengetahui pola penyakit yang diderita oleh seluruh pasien.

Hipotesa awal dalam penelitian ini adalah penerapan algoritma *apriori* pada *Association Rules Mining* dan *Sequential Pattern Mining* dapat menemukan pola keterkaitan penyakit pada waktu yang bersamaan dan pada waktu yang berbeda. Dari aturan dan pola yang dihasilkan, dapat membantu estimasi kebutuhan rumah sakit.

## 2. METODOLOGI

Data rekam medis yang akan diolah adalah data rekam medis pasien Jamkesmas RSUD Dr.H. Slamet Martodirdjo pada kurun tahun 2010, 2011, dan 2012. Sebelum dilakukan proses data mining, data rekam medis yang diperoleh dari basis data rumah sakit dilakukan tahap *preprocessing*. Mekanisme *preprocessing* pada tahap *integration* dan *transformation* dilakukan pengumpulan data dari tabel-tabel yang ada pada aplikasi resmi di rumah sakit menjadi tabel induk. Berdasarkan *data selection* dalam *preprocessing* maka atribut-atribut yang dipakai untuk melakukan data mining adalah: *PASIEN\_ID*, *TANGGAL\_PERIKSA*, dan *DIAGNOSA\_PENYAKIT* sebagai *input* atributnya.

Skenario penelitian dibagi menjadi 3 tahap. Tahap pertama menggunakan *dataset* rekam medis tahun 2010 untuk dicari pola asosiasi dan sekuensialnya. Tahap kedua dan ketiga masing-masing mengambil *dataset* rekam medis tahun 2011 dan tahun 2012 untuk juga dicari pola asosiasi dan sekuensialnya. *Minimum support count* yang digunakan untuk algoritma *association rule* untuk masing-masing tahun 2010 adalah 8, tahun 2011 adalah 10 dan tahun 2012 adalah 9. Nilai prosentase yang diambil sebagai *confidence* pada *rule* sebesar  $\geq 75\%$ . Untuk *minimum support count* yang digunakan untuk algoritma *sequential pattern* untuk masing-masing tahun 2010 adalah 10, 2011 adalah 15 dan tahun 2012 adalah 7. Penentuan angka *minimum support* mempertimbangkan pada jumlah *dataset* yang diperoleh (populasi) dan efisiensi pada proses *generate item*.

Semua hasil akhir dari penelitian akan dikonsultasikan dengan dokter/tenaga medis pada rumah sakit tempat penelitian. Hal ini dilakukan untuk mengetahui dan membandingkan kesesuaian antara tujuan, hipotesa dan hasil akhir penelitian dengan fakta kasus penyakit yang terjadi / ditangani di rumah sakit setempat.

## 2.1 Data Cleaning

Pada tahap ini dilakukan proses menghilangkan dan meminimalisasi *noise* atau mungkin kesalahan yang terdapat pada data. *Noise* yang ditemukan berupa data tidak lengkap/kosong, data rangkap, dan data tidak konsisten. Proses untuk masing-masing *data cleaning* adalah sebagai berikut:

- **Nilai yang Tidak Lengkap/Kosong**

Seringkali ditemui sebuah objek/*record* kehilangan salah satu atau lebih nilai atributnya. Hal tersebut terjadi bisa dikarenakan pada saat proses *entry* yang tidak lengkap atau tidak cermat. Pada *dataset* ditemukan cukup banyak *record* yang memiliki salah satu atau lebih atribut yang kosong dan tidak lengkap. Atribut pada *dataset* yang dipakai untuk melakukan *data mining* hanya berjumlah 3, maka nilai atribut yang kosong ditemui disalah satu atributnya. Misalnya PASIEN\_ID kosong sedangkan TANGGAL\_PERIKSA dan DIAGNOSA\_PENYAKIT ada nilainya. Tabel 1 memperlihatkan contoh *record* yang memiliki nilai kosong.

Tabel 1. Contoh record nilai kosong

Pasien_Id	Tgl Periksa	Diagnosa Penyakit
205	23 Agustus 2011	H16.0
205	23 Agustus 2011	Z20.1
212	23 April 2012	
212	24 April 2012	
215	31 Agustus 2010	N21.0
215	31 Agustus 2010	Z43.6

- **Nilai yang Rangkap**

Seringkali juga ditemui sebuah *record* memiliki nilai ganda. Misalnya PASIEN\_ID : 18 memiliki *record* rangkap sebanyak 2. Hal ini bisa diketahui dengan melihat nomor PASIEN\_ID, TANGGAL\_PERIKSA, dan DIAGNOSA\_PENYAKIT yang sama. Hal ini bisa terjadi karena proses relasi tabel pada saat proses memperoleh data dari *database* asal. Data rangkap bisa juga terjadi karena kurang cermatnya operator aplikasi dalam melakukan *entry* data kedalam komputer. Untuk mengatasi data yang rangkap maka dilakukan penghapusan data dengan proses *query* untuk mencari nilai yang unik. Tabel 2 menggambarkan contoh *record dataset* yang memiliki data rangkap/ganda.

Tabel 2. Contoh record rangkap

Pasien_Id	Tgl Periksa	Diagnosa Penyakit
18	28 Nopember 2011	A16.9
18	28 Nopember 2011	Z20.1
18	28 Nopember 2011	A16.9
18	28 Nopember 2011	Z20.1

- **Nilai yang Tidak Konsisten**

Pada *dataset* juga ditemui sejumlah *record* yang tidak konsisten nilai atributnya. Misalnya pada DIAGNOSA\_PENYAKIT dengan tipe data *text*, memiliki isian atribut dengan tipe data *number*. Misalnya pada contoh PASIEN\_ID 1572, Nilai "0" pada DIAGNOSA\_PENYAKIT bukanlah merupakan kode standar penyakit yang dipakai. Tabel 3 menggambarkan contoh *record dataset* yang memiliki nilai atribut tidak konsisten.

Tabel 3. Contoh record yang tidak konsisten

Pasien_Id	Tgl Periksa	Diagnosa Penyakit
1572	16 Januari 2012	E10.9
1572	16 Januari 2012	O03.0
1572	16 Januari 2012	0
1572	03 Maret 2012	Z11.9

## 2.2 Pembentukan Dataset

Setelah seluruh data dirasa cukup baik dalam hal inkonsistensi data, maka tahap selanjutnya adalah membuat *dataset* untuk proses *data mining*. Atribut yang dipakai untuk melakukan *data mining* adalah PASIEN\_ID, TANGGAL\_PERIKSA, dan DIAGNOSA\_PENYAKIT.

Dataset dibagi pertahun yaitu tahun 2010, 2011, dan 2012. Dataset yang dipakai disimpan dengan menggunakan format *file.xls* (Microsoft Excel) dengan alasan untuk kemudahan pembacaan dan proses *mining* pada aplikasi. Pemisahan *dataset* pertahun dilakukan dengan melakukan proses *query* untuk masing-masing tahun. Dari hasil proses *query* tersebut dihasilkan *dataset* pertahun yang akan dijadikan atribut *input* untuk proses *mining association rule* dan *sequential pattern*. Informasi detail jumlah *dataset* pertahun ditunjukkan pada Tabel 5.

Tabel 5. Hasil pembentukan dataset

Tahun	2010	2011	2012
Jumlah Dataset	5.037	19.353	6.076

### 2.3 Tahap Penggalan Aturan Asosiasi dan Pola Sekuensial

Pada proses penggalan aturan asosiasi terdapat dua proses utama yaitu proses penggalan/pencarian semua *large itemset* yang memenuhi *minimum support* dan proses membangkitkan aturan dari *large itemset* akhir yang telah ditemukan. Aturan yang dapat dibentuk sebanyak  $2^k - 2$ ,  $k$  merupakan bilangan bulat positif. Aturan yang dipakai yang memiliki nilai *confidence*  $\geq 75\%$ .

Pada proses penggalan pola sekuensial terdapat lima proses utama yaitu yaitu 1. *Sort Phase*, 2. *Large Itemset Phase*, 3. *Transformation Phase*, 4. *Sequence Phase*, dan 5. *Maximal Phase*. Hasil akhir dari *mining sequential pattern* ini adalah pola *maximal sequence*. *Sort* fase bertujuan untuk mengelompokkan *dataset* sesuai ID yang sama. *Large Itemset* fase bertujuan untuk menemukan semua *large itemset* yang memenuhi nilai *minimum support*. Setelah semua *large itemset* ditemukan, maka masing-masing  $L$  yang ditemukan tadi diubah ke dalam bentuk bilangan integer. Proses ini disebut *mapping*. Selanjutnya fase *transformation* yang bertujuan untuk merubah *dataset* pada *sort* fase kedalam bentuk *mapping*. Jika dalam *sort* fase ditemukan *dataset* yang tidak memenuhi *minimum support*, maka *item* pada *dataset* tersebut dihapus. Setelah itu menuju *sequence* fase, yang bertujuan menemukan *large itemset mapping mode* pada *dataset*. Tahap terakhir adalah *Maximal* fase, bertujuan untuk mencari semua *maximal sequence* yang terdapat pada *dataset*. Hasil dari *Maximal* fase merupakan *output* akhir dari proses *mining sequential pattern* berupa *maximal sequence*.

## 3. HASIL DAN PEMBAHASAN

Dari *dataset* yang terbentuk dari tahun 2010 sampai dengan 2012, penelitian dilakukan untuk mengetahui pola aturan asosiasi dan pola sekuensial yang terbentuk melalui proses penggalan data yang bertahap. Tahapan-tahapan percobaan ditunjukkan pada tabel 6 yang masing-masing algoritma terdiri dari 3 kali percobaan dari tahun 2010 sampai dengan tahun 2012. Total percobaan untuk kedua algoritma sebanyak 6 kali.

Tabel 6. Tahapan percobaan

Informasi Data Rekam Medis					Algoritma Data Mining	
					Association Rule	Sequential Pattern
Percob	Tahun	$\Sigma$ Record	$\Sigma$ Pasien ID	$\Sigma$ Penyakit	Min Sup	Min Sup
I	2010	5.037	2.936	764	8	10
II	2011	19.353	3.690	895	10	15
III	2012	6.076	3.424	826	9	7

### 3.1 Percobaan Pertama

Proses algoritma *apriori* yang diterapkan pada *dataset* tahun 2010 menghasilkan *large item* terbesar adalah  $L_4$  dengan jumlah 1 *itemset*. Dari  $L_4$  yang ditemukan, *rule* yang dapat dibangkitkan adalah sebanyak  $2^4 - 2 = 14$  *rule*. Dari total *rule* yang dihasilkan, terdapat 2 *rule* yang memiliki nilai *confidence*  $\geq 75\%$  seperti yang disajikan pada tabel 7.

Tabel 7. Strong rule dataset tahun 2010

NO	Strong Rule	Confidence
1	JIKA Senile cataract, morgagnian type DAN Senile cataract, unspecified DAN Dislocation of lens MAKA Disorder of lens, unspecified	80%
2	JIKA Senile cataract, unspecified DAN Dislocation of lens DAN Disorder of lens, unspecified MAKA Senile cataract, morgagnian type	80%

Dari penerapan algoritma *apriori* pada *dataset* tahun 2010 menghasilkan 2 *strong rule* dengan nilai *confidence* masing-masing 80%. Hasil *rule* menunjukkan bahwa tahun 2010 kasus terbanyak adalah penyakit mata yaitu katarak tipe morgagnian yang disertai dengan kerusakan pada lensa mata.

### 3.2 Percobaan Kedua

Proses algoritma *apriori* yang diterapkan pada *dataset* tahun 2011 menghasilkan *large item* terbesar adalah  $L_4$  dengan jumlah 2 *itemset*. Dari  $L_4$  yang ditemukan tersebut, *rule* yang dapat dibangkitkan adalah sebanyak  $2 \times (2^4 - 2) = 28$  *rule*. Dari total *rule* yang dihasilkan, terdapat 4 *rule* yang memiliki nilai *confidence*  $\geq 75\%$  seperti yang diperlihatkan pada tabel 8.

Tabel 8. Strong rule dataset tahun 2011

No	Strong Rule	Confidence
----	-------------	------------

1	JIKA Insulin-dependent diabetes mellitus with other specified complications DAN Non-insulin-dependent diabetes mellitus without complications DAN Special screening examination for infectious and parasitic diseases, <del>unspecified</del> MAKAs Insulin-dependent diabetes mellitus without complications	100%
2	JIKA Insulin-dependent diabetes mellitus with other specified complications DAN Special screening examination for infectious and parasitic diseases, <del>unspecified</del> MAKAs Insulin-dependent diabetes mellitus without complications DAN Non-insulin-dependent diabetes mellitus without complications	92.9%
3	JIKA Insulin-dependent diabetes mellitus with other specified complications DAN Insulin-dependent diabetes mellitus without complications DAN Special screening examination for infectious and parasitic diseases, <del>unspecified</del> MAKAs Non-insulin-dependent diabetes mellitus without complications	92.9%
4	JIKA Insulin-dependent diabetes mellitus with other specified complications DAN Insulin-dependent diabetes mellitus without complications MAKAs Special screening examination for infectious and parasitic diseases, <del>unspecified</del>	76.5%

Dari penerapan algoritma *apriori* pada *dataset* tahun 2011 menghasilkan 4 *strong rule* dengan nilai *confidence* masing-masing 100%, 92.9%, 92.9%, dan 76.5%. Dari hasil *rule* menunjukkan bahwa pada tahun 2011 banyak ditemukan kasus penyakit *diabetes mellitus* yang disertai penyakit infeksi parasit.

### 3.3 Percobaan Ketiga

Proses algoritma *apriori* yang diterapkan pada *dataset* tahun 2012 menghasilkan *large item* terbesar adalah L4 dengan jumlah 1 *itemset*. Dari L4 yang ditemukan tersebut, *rule* yang dapat dibangkitkan adalah sebanyak  $2^4 - 2 = 14$  *rule*. Dari total *rule* yang dihasilkan, terdapat 5 *rule* yang memiliki nilai *confidence*  $\geq 75\%$  seperti yang disajikan pada tabel 9.

Tabel 9. Strong rule dataset tahun 2012

No	Strong Rule	Confidence
1	JIKA Insulin-dependent diabetes mellitus with other specified complications DAN Special screening examination for infectious DAN parasitic diseases, <del>unspecified</del> MAKAs Insulin-dependent diabetes mellitus without complications DAN Non-insulin-dependent diabetes mellitus without complications	100%
2	JIKA Insulin-dependent diabetes mellitus with other specified complications DAN Insulin-dependent diabetes mellitus without complications DAN Special screening examination for infectious DAN parasitic diseases, <del>unspecified</del> MAKAs Non-insulin-dependent diabetes mellitus without complications	100%
3	JIKA Insulin-dependent diabetes mellitus with other specified complications DAN Non-insulin-dependent diabetes mellitus without complications DAN Special screening examination for infectious DAN parasitic diseases, <del>unspecified</del> MAKAs Insulin-dependent diabetes mellitus without complications	100%
4	JIKA Insulin-dependent diabetes mellitus with other specified complications DAN Non-insulin-dependent diabetes mellitus without complications MAKAs Insulin-dependent diabetes mellitus without complications DAN Special screening examination for infectious DAN parasitic diseases, <del>unspecified</del>	76.9%
5	JIKA Insulin-dependent diabetes mellitus with other specified complications DAN Insulin-dependent diabetes mellitus without complications DAN Non-insulin-dependent diabetes mellitus without complications MAKAs Special screening examination for infectious DAN parasitic diseases, <del>unspecified</del>	76.9%

Dari penerapan algoritma *apriori* pada *dataset* tahun 2012 menghasilkan 5 *strong rule* dengan nilai *confidence* masing-masing 100%, 100%, 100%, 76.9%, dan 76.9%. Dari hasil *rule* tersebut menunjukkan bahwa pada tahun 2012 banyak terjadi kasus penyakit kencing manis/*diabetes mellitus* yang disertai penyakit komplikasi dan juga mengalami penyakit infeksi yang disebabkan oleh parasit.

### 3.4 Percobaan Keempat

Percobaan keempat, kelima dan keenam, merupakan percobaan pada skenario *mining sequential pattern*. Proses algoritma *aprioriAll* yang diterapkan pada *dataset* tahun 2010 menghasilkan maksimal *sequence pattern* sebagai berikut: 22% penyakit yang diderita berkaitan dengan pernafasan/paru-paru/*tuberculosis*, 15% penyakit yang berkaitan dengan anemia, 15% penyakit yang berkenaan dengan kencing manis/*f*, 7% penyakit yang berkenaan dengan pendengaran/telinga bernanah, 7% penyakit yang berkenaan dengan darah tinggi dan gagal jantung, 4% penyakit yang berkenaan dengan asma dan hernia, 11% penyakit yang berkenaan dengan panggul, dan 15% penyakit yang berkenaan dengan katarak/penglihatan mata.

### 3.5 Percobaan Kelima

Proses algoritma *aprioriAll* yang diterapkan pada *dataset* tahun 2011 menghasilkan maksimal *sequence pattern* sebagai berikut : 1% berkenaan dengan penyakit tipus, 11% penyakit berkenaan dengan pernafasan seperti TBC, 1% kasus kanker *ovarium*, 1% kasus penyakit kanker payudara, 9% kasus penyakit yang berkenaan dengan tumor dibawah kulit, 11% kasus pada penyakit anemia, 1% kasus penyakit gumpalan pada kelenjar tiroid, 11% kasus penyakit *diabetes mellitus* tipe I dan tipe II, 14% kasus penyakit katarak dan kerusakan pada lensa mata, 3% kasus penyakit berkaitan pada pendengaran/telinga, 8% kasus penyakit berkaitan dengan hipertensi/darah tinggi, 5%

kasus pada penyakit pernafasan disebabkan alergi seperti asma, 1% kasus penyakit saluran prostat, dan 19% kasus penyakit yang berkenaan dengan gejala pada kandungan dan organ genital.

### 3.6 Percobaan Keenam

Untuk percobaan keenam atau percobaan terakhir, proses algoritma *aprioriAll* yang diterapkan pada *dataset* tahun 2012 menghasilkan maksimal *sequence pattern* sebagai berikut : 14% kasus terjadi penyakit pada tonjolan tumor dibawah kulit, 9% kasus untuk penyakit pada kelenjar tiroid, 9% kasus pada penyakit terkait pendengaran/telinga, 5% kasus terkait penyakit radang saluran nafas hidung, 9% kasus penyakit hernia, 41% kasus terkait gangguan pada kandungan dan alat genital, 14% kasus penyakit terkait penglihatan/mata seperti katarak dan gangguan fungsi lensa mata.

### 3.7 Pembuktian Hipotesa Terhadap Percobaan dan Diagnosis Dokter

Pada awal penelitian telah dijelaskan hipotesa awal bahwa penerapan algoritma *Association Rule* dan *Sequential Pattern* dapat mengetahui keterkaitan penyakit dan mengetahui prosentase penyakit berdasarkan kategori dari jenis penyakit yang diderita oleh pasien. Setelah dilakukan penelitian terhadap *database* rekam medis pasien dari tahun 2010 s/d 2012 didapatkan hasil bahwa kasus penanganan penyakit yang terjadi di RS Dr.H. Martodirdjo Pamekasan selama kurun waktu tahun 2010 adalah penyakit yang berkenaan dengan pernafasan baik yang disebabkan oleh bakteri dan non-bakteri, penyakit berkenaan dengan mata yaitu katarak dan kerusakan lensa mata, dan kasus penyakit berkenaan dengan kencing manis. Setelah dilakukan diagnosis dan komentar untuk tiap-tiap kasus penyakit pada *rule* dan *pattern maximal sequence* yang dihasilkan pada proses *mining sequential pattern*, didapatkan bahwa *rule* dan *pattern* tersebut masuk akal dan memang merupakan hubungan sebab akibat dari terjadinya suatu penyakit. Hal tersebut dikarenakan diagnosis pada status pasien hanya dokter yang berhak memberikan pengkodean. Adapun mekanisme pemeriksaan pasien oleh dokter adalah pertama dilakukan proses anamnesi, yaitu proses tanya jawab dengan dokter untuk mengetahui keluhan pasien. Setelah itu dilakukan pemeriksaan fisik kemudian dokter dapat menentukan diagnosis terhadap pasien. Setelah diagnosis diketahui, langkah selanjutnya dokter menentukan memberi tindakan atau cukup memberi resep obat. Komentar dari dokter menguatkan bahwa hasil dari percobaan tidak bertentangan dengan realitas pada kasus penyakit yang menjadi hubungan sebab akibat yang terjadi di lokasi penelitian. Keterangan dari dokter yang bertanggungjawab pada Unit Gawat Darurat (UGD) menyatakan bahwa *pattern* yang dihasilkan memberikan gambaran secara umum kasus penyakit yang banyak ditangani di RSUD Dr.H.Slamet Martodirdjo Pamekasan.

## 4. SIMPULAN DAN SARAN

Berikut akan diberikan simpulan dan saran dari penelitian ini.

### 4.1 Simpulan

Simpulan dari penelitian ini adalah sebagai berikut:

1. Hasil penggalian data memperlihatkan pola penyakit yang diderita oleh pasien di rumah sakit dan bisa diketahui kisaran prosentasenya untuk masing-masing kasus dan masing-masing tahun.
2. Informasi yang dihasilkan dari penggalian data berkesesuaian/tidak bertentangan dan masuk akal dengan hasil komentar/keterangan dari tenaga medis/dokter di rumah sakit tempat penelitian.

### 4.2 Saran

Adapun saran-saran yang bisa disampaikan adalah sebagai berikut:

1. Inefisiensi proses *generate* dan *join item* pada algoritma *Apriori* dan *AprioriAll* yang lambat dan menghabiskan ruang memori dapat dioptimasi dengan menggunakan algoritma lain yang tidak berbasis proses *join item*.
2. Untuk hasil analisis yang lebih optimal, perlu penambahan atribut *input* yang akan digali seperti misalnya Jenis\_Kelamin, Usia, dan Daerah\_Asal pada data rekam medis. Pada atribut *input* yang lebih banyak dapat menghasilkan multidimensi *association rule* yang cenderung lebih baik dan fleksibel terhadap analisis *interesting rule* yang diperoleh.

p

## 5. DAFTAR RUJUKAN

- [1] Rakesh dan Imieliski, Tomas, dan Swami, Arun. 1993. *Mining Association Rules between Sets of Items in Large Databases*: Proceedings of the 1993 ACM SIGMOD Conference Washington DC.
- [2] Agrawal, Rakesh dan Srikant, Ramakrishnan. 1994. *Fast Algorithms for Mining Association Rules*: Proceedings of the 20<sup>th</sup> VLDB Conference Santiago, Sept. 1994. Chile.
- [3] Agrawal, Rakesh dan Srikant, Ramakrishnan. 1995. *Mining Sequential Patterns*, Proceedings of International Conference Data Engineering Taipei, March 1995. Taiwan.

- [4] Han, Jiawei, dan Kamber, Micheline, 2006, *Data Mining : Concepts and Techniques Second Edition*, Morgan Kaufmann.
- [5] Hu, Ruijuan. 2010. *Medical Data Mining Based on Association Rules*. Computer and Information Science, 3(4).