

## MINING ASSOCIATION RULES DAN SEQUENTIAL PATTERNS DARI DATA REKAM MEDIS RUMAH SAKIT DR.H.SLAMET MARTODIRDJO PAMEKASAN

Nilam Ramadhani<sup>1</sup> dan Gunawan<sup>2</sup>

<sup>1</sup>Teknik Informatika Universitas Madura

<sup>2</sup>Teknik Informatika Sekolah Tinggi Teknik Surabaya  
nilam\_ramadhani@yahoo.com dan gunawan@stts.edu

### ABSTRAK

Pelayanan kesehatan yang baik untuk masyarakat di RSUD Dr.H.Slamet Martodirdjo Pamekasan menjadi sebuah keharusan di tengah meningkatnya jumlah penduduk, perubahan cuaca, dan pola hidup masyarakat Madura. Sehingga volume penyediaan layanan seperti tenaga medis, obat-obatan, dan fasilitas pendukung lainnya perlu ditingkatkan. Adanya data rekam medis pasien dapat dimanfaatkan untuk mendapatkan gambaran pola penyakit pasien yang berobat dengan teknik data mining. Pemanfaatan algoritma Association Rules Mining dapat mengetahui pola penyakit yang diderita oleh seluruh pasien pada waktu yang sama. Sedangkan pemanfaatan algoritma Sequential Patterns Mining dapat mengetahui pola penyakit yang diderita oleh seluruh pasien pada waktu yang berbeda. Input atribut yang dipakai adalah Pasien\_ID, Tanggal\_Periksa, dan Diagnosa\_Penyakit. Dataset yang digunakan tahun 2010 s/d tahun 2012. Hasil mining serta analisisnya menunjukkan pola penyakit yang diderita pasien dalam kurun waktu 3 tahun. Selain itu, hasil mining juga menunjukkan adanya kesesuaian dengan hasil diagnosa dan komentar oleh dokter di rumah sakit setempat.

*Kata kunci: data mining, data rekam medis rumah sakit, mining association rules, mining sequential patterns*

### ABSTRACT

*Good public health services at Dr.H.Slamet Martodirdjo hospital Pamekasan become a necessity amid increasing population, weather changes, and living habits of Madurese people. So that, volume of provision services such as medical personnels, medicines, and other supporting facilities should be increased. Availability of patients medical records data can be used to get a representation of disease pattern with data mining techniques. Utilization of Association Rules Mining algorithms can be used for finding the patterns of disease suffered by all patients at the same time. On the other hand, the utilization of Sequential Patterns Mining algorithms can be used for finding the patterns of disease suffered by all patients at different times. The input attributes used in both methods are Patient\_ID, Check\_Date, and Disease\_Diagnosys. The medical record data used are those of 2010, 2011 and 2012. The experiments result shows that patient's disease patterns can be known in a three-year period. Mining results indicate suitability with the doctor diagnosis and comments at the local hospital.*

*Keywords: data mining, hospital medical records data, mining association rules, mining sequential patterns*

## I. PENDAHULUAN

Profesionalisme di tengah otonomi daerah merupakan sebuah tuntutan yang harus dipenuhi. Termasuk pada instansi kesehatan seperti Rumah Sakit Umum Daerah (RSUD) Dr.H.Slamet Martodirdjo di Kabupaten Pamekasan. Permasalahan dalam mewujudkan pelayanan yang baik terkendala oleh volume fasilitas kesehatan seperti tenaga medis, obat, dan peralatan penunjangnya.

Dibutuhkan sebuah upaya sebagai cara untuk mengestimasi kebutuhan yang diperlukan oleh rumah sakit. Data rekam medis pasien dapat dimanfaatkan untuk pendekatan estimasi agar didapatkan informasi dan pengetahuan yang dapat membantu permasalahan yang ada.

Penerapan teknik data mining khususnya association rules dan sequential patterns akan menghasilkan luaran pengetahuan yang dapat membantu memberi solusi permasalahan yang muncul. Harapannya, pengetahuan yang diperoleh dapat dimanfaatkan lebih lanjut oleh pihak RSUD Dr.H.Slamet Martodirdjo Pamekasan.

## II. METODOLOGI

### 2.1 Data yang Dipakai

Data yang dipakai pada penelitian ini adalah data yang berasal dari database data rekam medis RSUD Dr. H. Slamet Martodirdjo Pamekasan. Adapun atribut yang dipakai adalah PASIEN\_ID, TANGGAL\_PERIKSA, dan DIAGNOSA\_PENYAKIT. Tabel 2.1 adalah hasil pemilihan atribut yang dibutuhkan untuk proses *mining association rules* dan *sequential patterns* menggunakan algoritma apriori.

**Tabel 2.1 Spesifikasi Atribut Dataset**

Atribut	Tipe Data	Keterangan
PASIEN_ID	Number	Kode register pasien
TANGGAL_PERIKSA	Date/Time	Tanggal pasien diperiksa oleh dokter
DIAGNOSA_PENYAKIT	Text	Hasil diagnosa penyakit yang direkam

Atribut tersebut dipilih berdasarkan analisis kebutuhan terhadap output yang diinginkan. Data yang diperoleh adalah data rekam medis tahun 2010 s/d 2012. Selain itu, pada parameter minimum support menggunakan input bilangan bulat. Hal ini bertujuan untuk memudahkan pemahaman terhadap jumlah frekuensi data/item dan parameter ini digunakan secara bersamaan untuk kedua metode mining.

Berikut diberikan spesifikasi dataset yang didapat dari pangkalan data rekam medis RSUD Dr.H.Slamet Martodirdjo Pamekasan yang disajikan pada Tabel 2.2.

**Tabel 2.2 Spesifikasi Jumlah Dataset**

Tahun	2010	2011	2012
Jumlah Pasien Unik	2936	3690	3424
Jumlah Diagnosa Penyakit	764	895	826
Jumlah Dataset	5.037	19.353	6.076

## 2.2 Keadaan Sistem Yang Dipakai

Aplikasi yang akan digunakan untuk proses data mining pada penelitian ini dibangun menggunakan *tool* bahasa pemrograman Visual Basic 6.0. Input untuk kedua algoritma menggunakan *file* input yang sama dengan format file .xls (Microsoft Office Excel). Selain itu pada parameter input yang lain yang juga digunakan untuk kedua algoritma adalah isian *minimum support count*. Rentang isian *minimum support count* dibuat dari angka 2 s/d 20.

Tabel 2.3 menunjukkan spesifikasi dari parameter yang digunakan untuk *association rules* dan *sequential patterns*.

**Tabel 2.3 Spesifikasi Parameter yang Dibuat**

Nama Parameter	Nilai Parameter	Keterangan
Support_Count	$\geq 2$	Jumlah kemunculan minimum itemset
Confidence	0-100	Nilai kepastian hubungan antar item
Max_ID	1-20000	Max id pasien yang akan diproses
Tahun	2010-2012	Tahun data yang akan diproses

Nilai yang diberikan untuk parameter tersebut dapat disesuaikan dengan jumlah dataset untuk pemrosesan data mining. Berikut beberapa ketentuan yang harus diperhatikan dalam menentukan nilai parameter *support\_count*:

1. Semakin kecil nilai *support*, semakin lama waktu pemrosesan karena data yang akan dibangkitkan untuk pembentukan pola semakin banyak. Banyaknya pola yang dapat dibangkitkan mengikuti rumus :  $(n \times (n-1))/2$ . Dengan semakin banyak pola yang dapat dibangkitkan akan menyebabkan kapasitas memori yang dibutuhkan menjadi besar.
2. Semakin besar nilai *support*, semakin cepat pemrosesan, namun lebih sedikit pola aturan yang teridentifikasi karena data yang terlibat dalam pencarian pola semakin kecil, namun memiliki nilai kepercayaan yang lebih baik. Kapasitas memori yang dibutuhkan semakin sedikit.

Pada penelitian ini digunakan kaidah penentuan *support count* nomor 2. Nilai untuk parameter *minimum\_support\_count* dapat dimasukkan dengan ketentuan  $\geq 2$ . Nilai 2 merupakan batasan terbawah dari *support\_count* yang akan diterapkan pada aplikasi. Untuk nilai *confidence*, ditetapkan diatas 75%. *Confidence* merupakan output sehingga pemilihan *rule* yang memiliki nilai *interestingness* dilakukan setelah *rule* diketahui melalui proses mining pada aplikasi.

### III. HASIL DAN PEMBAHASAN

#### 3.1 Mining Association Rules Dataset Semua Tahun

*Association Rules* dengan algoritma *Apriori* 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 Rules* dengan algoritma *Apriori* memiliki dua tahap utama, yaitu menemukan semua *large itemset* dan membentuk aturan yang memiliki nilai *confidence*.

Pada penggalian data asosiasi dataset semua tahun ini menggunakan nilai minimum support = 15. Adapun pada proses penggalian semua item yang dicari dari dataset semua tahun tersebut dihasilkan *largeitemset* seperti pada tabel 3.1.

**Tabel 3.1 Hasil Penggalian Large ItemSet Semua Tahun**

Generate ItemSet	Large Item-1		Large Item-2		Large Item-3		Large Item-4	
	CL-1	L-1	CL-2	L-2	CL-3	L-3	CL-4	L-4
<b>Jumlah</b>	1.257	215	23.005	73	48	10	2	0

Keterangan : CL: Candidate Large; L: Large

Dari hasil penggalian data didapatkan L1 sejumlah 215 item, L2 sejumlah 73 item, dan large item terbesar yang ditemukan adalah L3 dengan jumlah = 10. Dari L3 yang ditemukan tersebut, *rule* yang dapat dibangkitkan adalah sebanyak  $10 \times (2^3 - 2) = 60$  rule. Berikut disajikan rule yang dihasilkan dari program aplikasi dengan nilai *confidence* diatas 75%:

RULE#1

IF Insulin-dependent diabetes mellitus with other specified complications AND Non-insulin-dependent diabetes mellitus without complications THEN Insulin-dependent diabetes mellitus without complications. CONFIDENCE : 100%

RULE#2

IF Disorder of lens, unspecified AND Hypertensive heart disease without (congestive) heart failure THEN Senile cataract, unspecified. CONFIDENCE: 94.1%

RULE#3

IF Nontoxic single thyroid nodule AND Attention to unspecified artificial opening THEN Nontoxic goitre, unspecified. CONFIDENCE : 88.9%

RULE#4

IF Dislocation of lens AND Disorder of lens, unspecified THEN Senile cataract, morgagnian type. CONFIDENCE : 84.6%

RULE#5

IF Disorder of lens, unspecified AND Essential (primary) hypertension THEN Senile cataract, unspecified. CONFIDENCE : 78.9%

RULE#6

IF Senile cataract, unspecified AND Dislocation of lens THEN Senile cataract, morgagnian type. CONFIDENCE : 78.3%

RULE#7

IF Senile cataract, morgagnian type AND Disorder of lens, unspecified THEN Senile cataract, unspecified. CONFIDENCE : 75.7%

Adapun rule yang muncul berkenaan pada penyakit diabetes mellitus, penyakit katarak, dan tumor pada jaringan lunak. Level "unspecified" dicoret menunjukkan bahwa ada sebagian diagnosa penyakit yang masih belum dikodekan secara lebih spesifik, sehingga pada keterangan kode tersebut diberi label "unspecified".

### 3.2 Mining Sequential Patterns Dataset Semua Tahun

*Sequential Patterns Mining* dengan algoritma *AprioriAll* merupakan perluasan dari algoritma *association rules*. Karakteristik algoritma *AprioriAll* secara umum sama dengan algoritma *association rules* karena keduanya berbasis apriori. *Sequential Patterns Mining* menghasilkan model pengetahuan berupa aturan dengan bentuk maksimal sequence. Model pengetahuan tersebut dapat digunakan untuk mengetahui pola data pada periode yang berbeda. *Sequential Patterns Mining* dengan algoritma *AprioriAll* memiliki lima fase/tahap utama, yaitu *Sort Phase*, *Large Itemset Phase*, *Transformation Phase*, *Sequence Phase*, dan *Maximal Phase*.

Pada penggalian data pada dataset semua tahun ini menggunakan nilai minimum support = 20. Maksimal sequence yang dihasilkan sebanyak 47 pattern. Berikut akan disajikan 7 dari 47 pattern hasil maksimal sequence dari hasil penggalian dataset semua tahun:

PATTERN#1

1. Tuberculosis of lung, bacteriologically and histologically negative
2. Contact with and exposure to tuberculosis

PATTERN#2

1. Tuberculosis of lung, bacteriological and histological examination not done
2. Contact with and exposure to tuberculosis

PATTERN#3

1. Tuberculosis of intrathoracic lymph nodes, without mention of bacteriological or histological confirmation
2. Contact with and exposure to tuberculosis

PATTERN#4

1. Respiratory tuberculosis unspecified, without mention of bacteriological or histological confirmation
2. Contact with and exposure to tuberculosis

PATTERN#5

1. Tuberculosis of other specified organs
2. Contact with and exposure to tuberculosis

PATTERN#6

1. Carcinoma in situ of breast, unspecified
2. Surgical follow-up care, unspecified

PATTERN#7

1. Benign lipomatous neoplasm of skin and subcutaneous tissue of trunk
2. Benign lipomatous neoplasm, unspecified

### 3.3 Analisis Dokter Pada Hasil Mining Association Rules

Analisis oleh dokter yang akan dilakukan adalah dengan memberikan komentar terhadap *rule* yang dihasilkan dari hasil percobaan. *Rule* yang akan dikomentari adalah *rule* yang memiliki nilai confidence  $\geq 75\%$ . Komentar dokter terhadap hasil pada *rule* tersebut adalah:

- Diabetes mellitus adalah penyakit yang berkaitan dengan gula darah, yang disebabkan kadar gula darah glukosa dalam darah terlalu tinggi.

- Diabetes mellitus disebabkan oleh kurangnya insulin, terjangkit virus, factor gizi, genetic/keturunan, obesitas/kegemukan, dan tingginya kadar kortikosteroid.
- Senile cataract atau katarak senilis adalah gangguan pada lensa yang terjadi pada orang usia lanjut >50 tahun.
- Diagnosa tersebut sudah mencakup segala kemungkinan sebab dan akibat dari munculnya penyakit katarak. Penyebab katarak adalah: penuaan, diabetes, dan pasien dengan steroid dosis tinggi.
- Diperlukan pemeriksaan lebih lanjut untuk mencari kemungkinan yang lebih spesifik.
- *Rule* yang dihasilkan menunjukkan bahwa antar gejala saling berhubungan sehingga *rule* tersebut dapat diterima secara medis dan memang menunjukkan keterkaitan dan kasus yang terjadi di lapangan.

Analisis dan komentar dokter menunjukkan bahwa *rule* yang dihasilkan dari proses mining association rules berkesesuaian terhadap hubungan sebab akibat dari terjadinya penyakit diabetes mellitus dan komplikasinya, juga penyakit mata dan kerusakan pada lensa mata. Penyebab dan akibat dari penyakit yang telah didiagnosa oleh dokter saling berhubungan dan merupakan gejala yang dirasakan oleh pasien. Dokter mendefinisikan gejala tersebut sesuai apa yang diderita oleh pasien. Sehingga hasil diagnosa dan gejala komplikasinya sudah sesuai dengan hasil diagnosa di lapangan sehingga antara *rule* dengan komentar dari dokter saling berkaitan dan masuk akal.

Semua *rule* yang dihasilkan tidak bertentangan dengan hasil medis meskipun tingkat gejala yang dirasakan oleh pasien berbeda, namun hal tersebut diasumsikan sebagai gejala yang diderita oleh pasien dan menyertai gejala yang lain.

### 3.4 Analisis Dokter Pada Hasil Mining Sequential Patterns

Analisis oleh dokter yang akan dilakukan pada hasil *mining sequential patterns* ini adalah dengan memberikan komentar, keterangan dan analisis terhadap *pattern* yang dihasilkan dari hasil percobaan. *Pattern* dikelompokkan berdasarkan kasus penyakit yang sama atau mirip. Hal ini bertujuan memudahkan dokter dalam melakukan analisis pada kasus penyakit. Tidak semua *pattern* dikomentari, hanya *pattern* yang diketahui saja oleh dokter yang bersangkutan. Hal ini karena disesuaikan dengan pengalaman dokter di lapangan dalam menangani kasus penyakit tertentu.

Berikut beberapa komentar dokter yang dirangkum terkait *pattern* yang dihasilkan:

A. *Pattern* nomor 1,2,3,4,5 yang berkenaan dengan penyakit pernafasan:

- Tuberculosis (TBC) disebabkan oleh bakteri tuberculosis yang menular dan bukan turunan.
- Penyakit ini menyerang sistem pernafasan, paru-paru, limfa, dan sekitar thorax.
- Kecenderungan di lapangan, pasien yang menderita penyakit TBC juga mengalami gejala penyakit lainnya seperti anemia dan adanya infeksi oleh bakteri pada sistem pernafasan.
- Dibutuhkan pemeriksaan laboratorium agar mengetahui pasien terpapar bakteri atau tidak.
- Secara keseluruhan, *pattern* yang dihasilkan memang kasus yang banyak dialami oleh pasien di RSUD.
- *Pattern* yang dihasilkan berkesesuaian dan masuk akal.

- B. Pattern nomor 7 yang berkaitan dengan penyakit benjolan tumor jinak
- Lipoma merupakan tumor jinak jaringan lemak. Penyebab secara pasti belum diketahui. Terdapat pada jaringan dibawah kulit yang terjadi di kepala, muka, dan leher.
  - Kasus pada pattern#7 termasuk kasus yang cukup banyak ditangani di RSUD dan berkesesuaian.

## IV. PENUTUP

### 4.1 Kesimpulan

Penelitian ini telah mengeksplorasi teknik *mining association rules* dan *sequential patterns* menggunakan algoritma *Apriori* dan *AprioriAll* terhadap data rekam medis Rumah Sakit Dr.H.Slamet Martodirdjo Pamekasan. Dari percobaan yang telah dilakukan menunjukkan hasil sebagai berikut:

1. Hasil model *rule* dan *pattern* yang ditemukan menggambarkan kasus penyakit yang terjadi di rumah sakit selama kurun waktu tahun 2010 sampai dengan 2012.
2. *Rule* dan *pattern* yang dihasilkan dari penggalian data berkesesuaian dan tidak bertentangan dengan hasil komentar/keterangan dari tenaga medis/dokter di rumah sakit tempat penelitian.
3. Algoritma *Apriori* dapat diimplementasikan terhadap data rekam medis pasien untuk mendapatkan gambaran kasus penyakit yang pernah ditangani di rumah sakit.

### 4.2 Saran

Melihat kinerja dari teknik *mining association rules* dan *sequential patterns* menggunakan algoritma *Apriori* dan *AprioriAll* dan implementasinya terhadap data rekam medis, saran-saran yang bisa disampaikan adalah sebagai berikut:

1. Pertimbangan inefisiensi proses *generate, join item* dan *prune* pada algoritma *Apriori* dan *AprioriAll*, dapat dioptimasi menggunakan algoritma lain yang tidak berbasis proses *join and prune item*.
2. Untuk hasil analisis yang lebih optimal, perlu penambahan atribut input yang akan dimining seperti misalnya Jenis\_Kelamin, Usia, dan Daerah\_Asal pada data rekam medis. Sehingga dapat membentuk multilevel *association rules* yang cenderung lebih baik terhadap analisis *interesting rules*.
3. Pihak RSUD Dr.H.Slamet Martodirdjo Pamekasan dapat mengoptimalkan penggunaan aplikasi data mining ini secara kontinyu dengan meng-entry secara lengkap dan benar data rekam medis pasien yang berobat, sehingga hasil mining lebih akurat. Selain itu perlu ada mekanisme yang secara berkala melakukan analisis hasil mining yang dibandingkan dan dikolaborasi dengan kondisi riil dilapangan oleh pihak internal manajemen rumah sakit.

## V. DAFTAR PUSTAKA

- [1] Agrawal, Rakesh, Imieliski, Tomas, dan Swami, Arun. *Mining Association Rules between Sets of Items in Large Databases*. Proceedings of the 1993 ACM SIGMOD Conference Washington DC, USA. May. 1993.

- [2] Agrawal, Rakesh dan Srikant, Ramakrishnan. *Fast Algorithms for Mining Association Rules*. Proceedings of the 20<sup>th</sup> VLDB Conference Santiago, Chile. 1994.
- [3] Han, Jiawei, dan Kamber, Micheline. *Data Mining: Concepts and Techniques Second Edition*. Morgan Kaufmann. 2006.
- [4] Hu, Ruijuan. *Medical Data Mining Based on Association Rules*. Computer and Information Science, Vol. 3 No. 4. November. 2010.
- [5] Rustiyanto, Ery. *Statistik Rumah Sakit Untuk Pengambilan Keputusan*. Graha Ilmu. Yogyakarta. 2010.