

# Sequential Pattern Mining Data Pemeriksaan Pasien Menggunakan Algoritma SPADE (*Sequential Pattern Using Discovery Equivalent Classes*) Studi Kasus BPI Albasrah Wajak

Silmi Khashaish Fillah<sup>1</sup>, Dedy Rahman Prehanto<sup>2</sup>

<sup>1,2</sup> Program Studi S1 Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya

<sup>1</sup>[silmi.17051214051@unesa.ac.id](mailto:silmi.17051214051@unesa.ac.id)

<sup>2</sup>[dedyrahman@unesa.ac.id](mailto:dedyrahman@unesa.ac.id)

**Abstrak**— Tidak bisa dipungkiri bahwa kemajuan teknologi saat ini terus mengalami peningkatan. Dengan adanya kemajuan teknologi, perusahaan atau organisasi manapun dapat mengumpulkan data dari berbagai sumber dengan cepat dan mudah. Semakin besar dan luasnya data yang dimiliki oleh suatu perusahaan atau organisasi tersebut, maka semakin besar peluang masalah yang akan muncul. Begitu juga dengan sebuah tempat pelayanan kesehatan masyarakat seperti klinik. Semakin banyak data pemeriksaan medis, semakin banyak pula sesuatu yang bisa diolah sehingga dapat menghasilkan suatu ramalan ataupun prediksi. Pada kasus ini yaitu data pemeriksaan medis pasien di BPI Albasrah Wajak. Salah satu metode data mining yang dapat diterapkan yaitu *sequential pattern mining* dengan menggunakan algoritma SPADE (*Sequential Pattern Using Discovery Equivalent Classes*). Metode ini menggambarkan pola urutan kedatangan pasien dalam waktu yang relatif cepat. Algoritma SPADE akan mencari *frequent sequences* dari data pemeriksaan pasien dengan menggunakan database vertikal. Hasil dari penelitian ini menghasilkan beberapa *rule* yang ditentukan oleh nilai *minimum support*, *minimum confidence*, dan nilai *lift ratio*. Sehingga dari penelitian ini dapat diambil kesimpulan bahwa algoritma SPADE terbukti secara akurat dengan presentase sebesar 71,43% untuk mencari pola sekuensial dari data pemeriksaan pasien untuk mengetahui kedatangan kembali pasien ke BPI Albasrah Wajak pada waktu yang akan datang.

**Kata Kunci**— data mining, spade, spade algorithm, sequential pattern mining, lift ratio

## I. PENDAHULUAN

Tidak bisa dipungkiri bahwa kemajuan teknologi saat ini terus mengalami peningkatan. Dengan adanya kemajuan teknologi, suatu perusahaan atau organisasi dapat mengumpulkan data dari berbagai sumber dengan cepat dan mudah. Semakin besar dan luasnya data yang dimiliki oleh suatu perusahaan atau organisasi tersebut, maka semakin besar peluang untuk membuat data tersebut menjadi suatu informasi yang bermanfaat. Oleh karena itu, untuk memanfaatkan peluang yang ada, data tersebut perlu diolah menjadi suatu pengetahuan yang dapat digunakan untuk memenuhi kebutuhan suatu perusahaan kedepannya.

Salah satu caranya yaitu dengan melakukan *data mining*. Di era ini, *data mining* sudah menjadi hal yang penting bagi suatu teknologi. Teknologi yang berkembang selama ini tidak akan

ada artinya jika sistem tersebut dibuat tanpa *data mining*. *Data mining* sendiri merupakan sebuah ilmu yang mempelajari tentang pola kejadian yang akan dimasukkan ke dalam suatu sistem, sehingga sistem itu dapat dikatakan sistem yang cerdas<sup>[1]</sup>. Begitu pula dengan suatu rumah sakit maupun klinik yang menyimpan berbagai data pasien untuk keperluan operasional. *Data mining* dapat menjadi solusi bagi suatu organisasi atau perusahaan tersebut untuk mengolah data menjadi informasi dan atau pengetahuan yang bermanfaat bagi kelangsungan operasional organisasi atau perusahaan.

*Sequential pattern mining* merupakan salah satu metode dalam *data mining*. *Sequential patterns* merupakan pola yang menggambarkan urutan waktu terjadinya peristiwa<sup>[1]</sup>. *Sequential pattern mining* dapat dibantu oleh beberapa algoritma. Beberapa algoritma yang dapat dipakai antara lain AprioriAll dan GSP. *Sequential Pattern Mining* dengan algoritma AprioriAll merupakan perluasan dari *association rule mining*<sup>[2]</sup>. AprioriAll dan algoritma GSP terinspirasi dari algoritma Apriori dalam menemukan *frequent itemset mining*<sup>[3]</sup>. Selain itu, ada algoritma SPADE yang merupakan pengembangan dari metode sebelumnya yaitu GSP (*Generalized Sequential Pattern*) dikarenakan algoritma SPADE memiliki proses pengolahan data yang lebih cepat dibandingkan dengan algoritma GSP<sup>[2]</sup>.

Algoritma *Generalized Sequence Pattern* (GSP) dan *Sequential Pattern Discovery Using Equivalent Classes* (SPADE) memiliki persamaan yaitu *apriori-based*. Meskipun sama-sama *apriori-based*, algoritma SPADE lebih cepat pengolahan datanya dibandingkan dengan algoritma GSP. Algoritma SPADE merupakan algoritma alternatif yang memanfaatkan pencarian kedalaman pertama (*depth-first search*) dan menghindari beberapa kelemahan dari algoritma GSP. Algoritma SPADE terinspirasi oleh algoritma Eclat dalam *frequent itemset mining* dan merepresentasikan *vertical database* dibandingkan horizontal<sup>[4]</sup>. Algoritma SPADE menguraikan masalah menjadi beberapa submasalah yang lebih kecil, yang masing-masing dapat diselesaikan secara mandiri<sup>[5]</sup>. Algoritma SPADE memiliki skala yang linear dengan jumlah pelanggan dan juga memiliki properti *scaleup* yang baik dalam hal jumlah transaksi tiap pelanggan, ukuran transaksi, dan ukuran *frequent itemsets* serta *frequent sequences* yang terbentuk<sup>[5]</sup>.

Penelitian-penelitian tentang *sequential pattern mining* dengan algoritma SPADE telah banyak dilakukan sebelumnya salah satunya oleh Dini dkk (2017) dalam jurnalnya yang berjudul “Sequential Pattern Mining of Rawi Hadis (Case Study: Shahih Hadits of Imam Bukhari from Software Ensiklopedi Hadis Kitab 9 Imam)”<sup>[6]</sup>. Penelitian ini bertujuan untuk mengetahui pola sekuensial yang terbentuk antara *hadits rawi* dalam kitab *shahih bukhari* dan proses algoritma SPADE akan menghasilkan *frequent sequence* yang kemudian digunakan untuk membentuk aturan-aturan *hadits rawi*. Adapun penelitian lainnya tentang algoritma SPADE yaitu jurnal yang ditulis oleh Saputra dkk pada tahun 2019 yang berjudul “Penentuan Pola Sekuensial Data Transaksi Penjualan Menggunakan Algoritma Sequential Pattern Discovery Using Equivalent Classes (SPADE)”<sup>[7]</sup>.

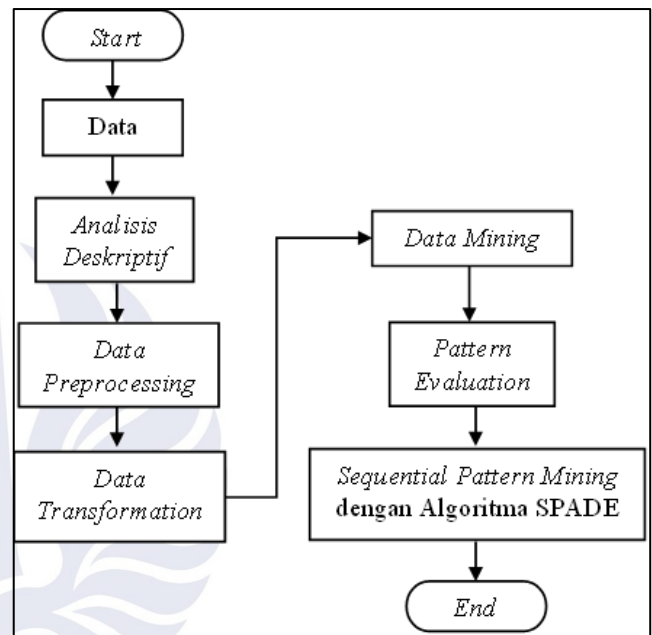
Balai Pengobatan Islam “BPI” Albasrah merupakan sebuah balai pengobatan islam yang berlokasi di Kecamatan Wajak, Kabupaten Malang. Balai Pengobatan Islam ini sama halnya dengan klinik yang merupakan fasilitas pelayanan kesehatan yang menyelenggarakan pelayanan kesehatan perorangan yang menyediakan pelayanan medis dasar dan/atau spesialis<sup>[8]</sup>. Balai pengobatan ini selalu ramai pasien karena letaknya yang strategis dan terkenal dengan pelayanannya yang baik. Data yang dimiliki oleh BPI Albasrah Wajak juga banyak, mulai dari data pasien, rekam medis, farmasi, hingga biaya pemeriksaan. Penelitian ini akan menggunakan data pemeriksaan medis pasien dari BPI Albasrah. Dalam penelitian ini, karakteristik pasien akan dianalisis secara deskriptif dari data yang diperoleh sebagai informasi dasar pihak balai pengobatan. Selain itu, suatu *data mining* dapat membantu proses pengambilan keputusan di BPI Albasrah. Salah satunya dalam mengetahui pola sekuensial pasien dari diagnosis penyakitnya sehingga dapat diketahui pola kedatangan kembali pasien. Dalam penelitian ini akan dilakukan *data mining* dari data pemeriksaan pasien milik BPI Albasrah dengan metode *sequential pattern mining*. Hasil dari proses *sequential pattern mining* dapat digunakan pihak balai pengobatan untuk menentukan strategi dan atau antisipasi di waktu yang akan datang dalam hal ketersediaan obat, ketersediaan peralatan pengobatan, serta jumlah dokter yang menangan.

Balai Pengobatan Islam Albasrah Wajak memiliki data pemeriksaan pasien yang relatif besar dalam kurun waktu 1 bulan. Terdapat kurang lebih 1.700 data pemeriksaan yang tersimpan dan belum dimanfaatkan dengan sebagaimana mestinya. Agar tidak menyia-nyiakan data yang ada, data tersebut dapat diolah untuk menghasilkan suatu informasi. Data tersebut dapat dilakukan proses *data mining* untuk menghasilkan suatu informasi sebagai bahan pertimbangan balai pengobatan dalam menentukan suatu keputusan ataupun solusi. Diagnosis penyakit pada data pemeriksaan pasien BPI Albasrah belum dilengkapi dengan kode ICD-10, yang merupakan standar dari klasifikasi penyakit, cedera, dan sebab kematian untuk tujuan statistik<sup>[9]</sup>. Dalam sidang World Health Assembly ke-43, WHO telah menetapkan ICD-10 sebagai pedoman klasifikasi internasional tentang penyakit edisi terbaru yang harus dipakai oleh seluruh negara anggotanya termasuk negara Indonesia<sup>[9]</sup>. Pada penelitian ini, diagnosis

penyakit akan diklasifikasikan berdasarkan kode penyakit sesuai dengan standar ICD-10 agar mudah dalam pengolahan data. Untuk membantu proses penelitian, proses *sequential pattern mining* akan dibantu oleh algoritma SPADE (*Sequential Pattern Discovery using Equivalent Classes*) untuk mendapatkan pola dengan mengidentifikasi diagnosis penyakit setiap pasien di BPI Albasrah. Penelitian ini dalam olah datanya akan dibantu dengan *software* R versi 4.0.4 yang sebagian besar digunakan untuk analisis data statistik dan grafis.

## II. METODE PENELITIAN

Pada penelitian ini, metode *sequential pattern mining* dengan algoritma SPADE dilakukan dengan melakukan langkah-langkah yang tertera pada gambar berikut ini.



Gbr. 1 Kerangka Penelitian

### A. Data

Data yang digunakan dalam penelitian yaitu data sekunder dari data yang ada di BPI Albasrah Wajak. Data tersebut berupa data biaya pemeriksaan pasien yang terdiri dari beberapa atribut antara lain kode pasien, nama pasien, jenis kelamin, alamat, usia, biaya pemeriksaan, dan diagnosis penyakit. Diagnosis penyakit nantinya akan diberi kode berdasarkan aturan ICD-10.

### B. Metode Analisis Deskriptif

Metode analisis deskriptif adalah statistik yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa bermaksud membuat kesimpulan yang berlaku untuk umum atau generalisasi<sup>[10]</sup>. Pada penelitian ini, analisis deskriptif digunakan untuk menentukan karakteristik pasien yang berobat di BPI Albasrah Wajak yang terdiri dari umur, jenis kelamin, diagnosis penyakit, alamat, dan rata-rata biaya pengobatan. Hasil dari metode ini berupa presentase dari beberapa karakteristik pasien.

### C. Data Preprocessing

*Data preprocessing* merupakan proses mengubah data mentah menjadi format yang dapat dipahami. Kualitas data harus diperiksa agar hasil yang didapatkan juga sesuai. *Data preprocessing* pada penelitian ini terdiri dari *data cleaning* dan *data selection*.

*Data cleaning* digunakan untuk menghilangkan serta meminimalisir *noise* atau kesalahan dan data yang tidak konsisten. *Noise* yang dimaksud berupa data yang rangkap dan atau data yang tidak lengkap.

*Data selection* digunakan untuk memilah dan memilih data yang diperoleh dari data milik BPI Albasrah Wajak, sehingga lebih terperinci untuk dilakukan proses *sequential pattern mining*.

### D. Data Transformation

*Data transformation* digunakan untuk mengubah data atau mentransformasi data ke dalam bentuk yang disesuaikan dengan kebutuhan proses *data mining*, yaitu proses *sequential pattern mining*. Dalam hal ini, atribut data yang didapat dari database BPI Albasrah Wajak ditransformasi ke dalam format atribut yang dibutuhkan untuk proses *sequential pattern mining*.

### E. Data Mining

*Data mining* merupakan suatu ilmu yang mempelajari suatu pola dari data yang akan diteliti. Pada tahap ini, data yang telah ditransformasi akan diolah lebih lanjut dengan menggunakan salah satu metode data mining untuk mendapatkan pola-pola agar menjadi suatu pengetahuan baru. Metode *data mining* yang digunakan dalam penelitian ini yaitu *sequential pattern mining* yang sesuai tujuan penelitiannya yaitu untuk mengetahui pola sekuensial dari kedatangan pasien.

### F. Pattern Evaluation

*Pattern evaluation* digunakan untuk mengidentifikasi pola-pola yang menarik berdasarkan data pemeriksaan pasien. Pencarian pola sekuensial dilakukan dengan menggunakan *software R* versi 4.0.4. Sehingga akan diperoleh suatu pengetahuan atau informasi yang bermanfaat berupa rules atau aturan yang terbentuk. Kekuatan rules dapat ditentukan oleh nilai *minimum support*, *minimum confidence*, dan nilai *lift ratio*.

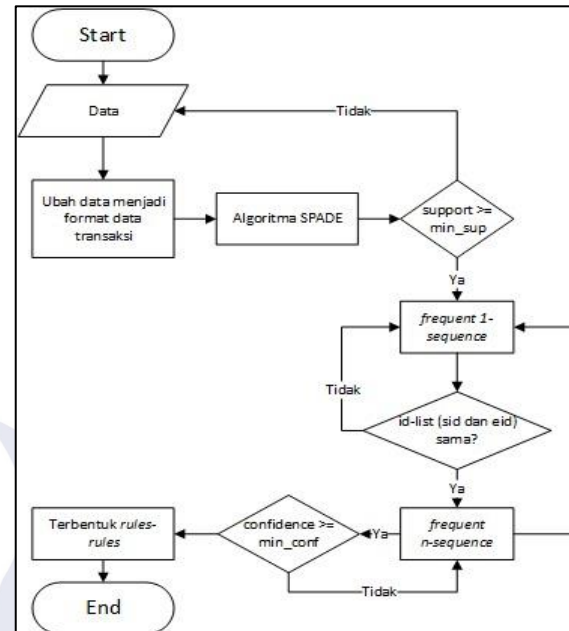
### G. Implementasi Sequential Pattern Mining dengan Algoritma SPADE

*Sequential pattern mining* ialah suatu metode dalam data mining yang digunakan untuk mengidentifikasi pola urutan peristiwa yang ada di dalam database. *Sequential Pattern Mining* pertama kali diperkenalkan oleh Agrawal dan Srikant pada tahun 1995. *Sequential patterns* merupakan pola yang menggambarkan urutan waktu terjadinya peristiwa<sup>[1]</sup>. Pola tersebut bisa ditemukan apabila data yang disimpan relatif besar dan memiliki objek yang sama dalam jumlah yang relatif besar dalam melakukan beberapa transaksi yang berulang kali.

Algoritma ini menggunakan format database vertikal, di mana perlu mengasosiasikan setiap urutan kejadian terjadi. Algoritma SPADE akan dibagi menjadi 4 variabel antara lain *sequenceID*, *eventID*, *size*, dan *items*. Data yang diperoleh dari

data BPI Albasrah Wajak akan disesuaikan dengan format data dalam algoritma SPADE. Algoritma SPADE memiliki skala yang linear dengan jumlah pelanggan dan juga memiliki properti *scaleup* yang baik dalam hal jumlah transaksi tiap pelanggan, ukuran transaksi, dan ukuran *frequent itemsets* serta *frequent sequences* yang terbentuk<sup>[5]</sup>.

Gambaran proses pada algoritma SPADE dapat dilihat pada gambar berikut ini.



Gbr. 2 Flowchart Algoritma SPADE

Referensi [2] menjelaskan tentang proses *sequential pattern* dengan algoritma SPADE yang telah dirangkum sesuai penjelasan di bawah ini.

#### a. Mencari *frequent 1-sequence*

Langkah pertama untuk mencari *frequent 1-sequence* yaitu dengan memasukkan data yang sudah disusun sesuai *sequenceID*, *eventID*, *size*, serta *items*. Kemudian program akan melakukan pemindaian untuk setiap *itemset* di dalam *sequence database* dan dilakukan pengecekan apakah *itemset* muncul berbentuk pasangan *id-list* (*sid* dan *eid*). Apabila muncul *id-list*, maka akan terbentuk *frequent 1-sequence*. *Sequence* yang dimasukkan dalam *frequent 1-sequence* ialah *sequence* yang memiliki nilai *support* lebih dari atau sama dengan nilai *min\_sup* yang telah ditentukan.

#### b. Mencari *frequent 2-sequence*

Untuk mencari *frequent 2-sequence*, dilakukan dengan menentukan *min\_sup* dan data *frequent 1-sequence*. Kemudian dicek apakah setiap *frequent 1-sequence* memenuhi *min\_sup*. Setiap *frequent 1-sequence* digabungkan dengan semua *frequent 1-sequence* lainnya. Setiap penggabungan *frequent 1-sequence* dilakukan pengecekan apakah dalam *id-list* memiliki *sid* yang sama. Jika memiliki *sid* yang sama, maka dicek lagi apakah memiliki *eid* yang sama. Apabila *id-list* sesuai, maka *id-list* dimasukkan dalam *frequent 2-sequence*. Perlu dipastikan



kembali bahwa nilai *support* sudah lebih dari atau sama dengan *min\_sup*.

c. Mencari *frequent k-sequence*

Untuk mencari *sequence-sequence* berikutnya dilakukan dengan proses yang sama. Untuk mencari *frequent k-sequence*, dilakukan dengan melakukan *join* pada *frequent (k-1)-sequence* yang memiliki *prefix* yang sama. Nilai *support* harus memenuhi *min\_sup* untuk dimasukkan ke dalam *frequent k-sequence*. Pencarian *frequent sequence* akan berhenti apabila tidak ada *frequent (k-1)-sequence* yang bisa dilakukan *join* lagi dan sudah tidak ditemukan *frequent k-sequence* lagi.

### III. HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Data yang digunakan dalam penelitian yaitu data sekunder dari suatu balai pengobatan. Data tersebut berupa data biaya pemeriksaan pasien dari bulan Desember 2020 hingga bulan Januari 2021 yang terdiri dari beberapa atribut antara lain kode pasien, nama pasien, jenis kelamin, alamat, usia, biaya pemeriksaan, dan diagnosis penyakit.

Untuk merealisasikan tujuan penelitian, seluruh atribut di dalam data biaya pemeriksaan pasien tidak semua digunakan. Atribut yang digunakan hanya terdiri dari kode pasien, waktu pemeriksaan, jumlah penyakit yang terdiagnosis, dan kode diagnosis penyakit. Sedangkan atribut lainnya dapat digunakan untuk mencari karakteristik dari keseluruhan pasien yang berobat di Balai Pengobatan Islam Albasrah Wajak.

Atribut kode penyakit pasien pada data pemeriksaan medis masih berupa nama penyakit dan belum diklasifikasi berdasarkan kode-kode. Oleh karena itu, untuk mempermudah proses *data mining*, atribut kode penyakit akan diklasifikasi berdasarkan ICD-10. ICD-10 merupakan singkatan dari *The International Statistical Classification of Diseases and Related Health Problems-10th Revision*. Fungsi dasar dari ICD (*International Classification of Disease*) ialah sebagai klasifikasi penyakit, cedera, dan sebab kematian untuk tujuan statistik (perekaman yang sistematis, analisis, interpretasi, dan komparasi data morbiditas maupun mortalitas). Hal-hal tentang ICD-10 ini tertulis pada Bahan Ajar Rekam Medis dan Informasi Kesehatan (RMIK) milik Kemenkes RI tentang Klasifikasi, Kodifikasi Penyakit, dan Masalah Terkait I tahun 2017<sup>[6]</sup>.

Berdasarkan data yang telah didapat dari BPI Albasrah Wajak, sebanyak 157 diagnosis penyakit akan diberi kode penyakit sesuai dengan aturan ICD-10. Berikut ini beberapa contoh penyakit dengan kodenya berdasarkan ICD-10.

TABEL I  
KODE PENYAKIT BERDASARKAN ICD-10

Kode ICD-10	Nama Penyakit
J06.9	ISPA
R50.9	Observasi Febris
K30	<i>Dyspepsia Syndrome</i>
I10	Hipertensi
R50.84	<i>Acute Febrile Illness</i>
J00	<i>Common Cold</i>
Z48.00	<i>Wound Toilet</i>

Kode ICD-10	Nama Penyakit
Z48.02	<i>Wound Toilet (hecting)</i>
K21	GERD
T14	<i>Vulnus Abrasio</i>
K29.70	<i>Gastritis</i>
L02.91	<i>Abscess</i>
M79.1	<i>Myalgia</i>
K52.9	<i>Gastroenteritis Acute</i>
M62.83	<i>Muscle Spasme</i>
Z41.2	<i>Circumcision</i>

#### B. Metode Analisis Deskriptif

Analisis ini digunakan untuk mengetahui beberapa karakteristik pasien di BPI Albasrah Wajak melalui data biaya pemeriksaan medis pada bulan Desember 2020 hingga Januari 2021. Data yang digunakan antara lain usia, jenis kelamin, dan kode diagnosis penyakit yang diderita pasien.



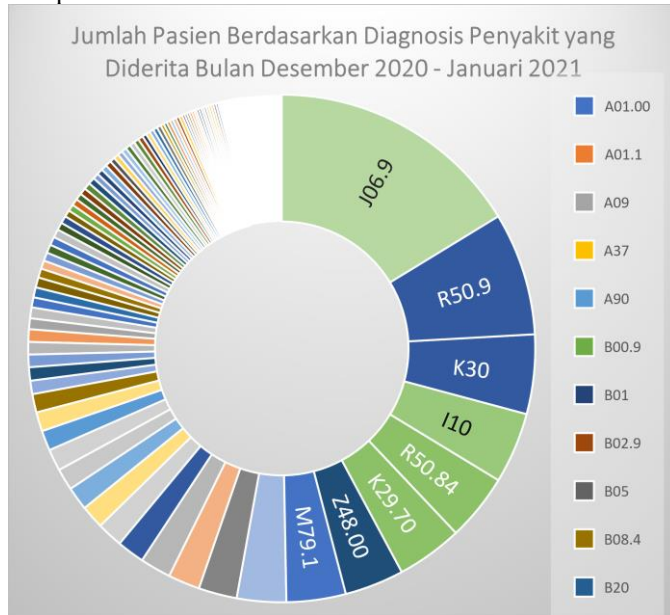
Gbr. 3 Diagram Jumlah Pasien Berdasarkan Jenis Kelamin

Berdasarkan gambar 3, kunjungan pasien BPI Albasrah Wajak pada bulan Desember 2020 hingga Januari 2021 didominasi oleh pasien dengan jenis kelamin laki-laki sebanyak 877 pasien dan perempuan sebanyak 835 pasien.



Gbr. 4 Diagram Jumlah Pasien Berdasarkan Rentang Usia

Pasien BPI Albasrah Wajak yang berkunjung terdiri dari berbagai rentang usia. Berdasarkan gambar 4, sebanyak 372 pasien berusia di bawah 15 tahun (anak-anak, sebanyak 164 pasien berusia 15-24 tahun, sebanyak 415 pasien berusia 25-44 tahun, sebanyak 532 pasien berusia 45-64 tahun, dan sebanyak 229 pasien berusia di atas 65 tahun.



Gbr. 5 Diagram Jumlah Pasien Berdasarkan Penyakit yang Diderita

Pada bulan Desember 2020 hingga Januari 2021, pasien BPI Albasrah Wajak terdiagnosis 147 jenis penyakit yang berbeda. Berdasarkan gambar di atas, sebagian besar penyakit yang diderita oleh pasien yaitu ISPA dengan kode penyakit J06.9 sebanyak 306 pasien atau 16,28% dan Observasi Febris dengan kode penyakit R50.9 sebanyak 147 pasien atau 7,8%.

### C. Data Preprocessing

Dalam penelitian ini, data yang digunakan yaitu data pemeriksaan pasien BPI Albasrah Wajak. Dari 1.796 data yang ada, setelah dilakukan *cleaning* dengan menghilangkan *noise*, redudansi, data kosong, dan data tidak sesuai sehingga menghasilkan data sebanyak 1.712 data.

Berikut ini adalah sebagian gambaran data yang diperoleh dari data BPI Albasrah Wajak yang dirangkai dalam format *excel* (.xlsx) dan diurutkan berdasarkan *id\_pasien*.

TABEL II  
FORMAT DATA PEMERIKSAAN PASIEN BPI ALBASRAH WAJAK

id	tgl_pemr	nama_pasien	jk	usia	alamat
1	18/12/20	Tn. Sunarto	L	32	Wonoayu
2	18/12/20	Tn. Denes	L	20	Kidangbang
3	18/12/20	Tn. Asrofi	L	40	Wajak
4	18/12/20	Tn. Siswanto	L	48	Ngawonggo
5	18/12/20	An. Gilang Wiji	L	1	P. Mulyo
6	18/12/20	Tn. Ahmad Muda'i	L	58	Jaruman
6	23/12/20	Tn. Ahmad Muda'i	L	58	Jaruman
7	18/12/20	Ny. Ginten	P	50	Dadapan
8	18/12/20	Tn. Mukid	L	67	K. Anyar
9	18/12/20	Tn. Jani	L	51	P. Mulyo

TABEL III  
FORMAT DATA PEMERIKSAAN PASIEN BPI ALBASRAH WAJAK (LANJUTAN)

kode_peny	diagnosis_peny	total_pemb
R11.10	Observasi Vomiting	Rp58,000.00
K21	GERD	Rp55,000.00
R50.84	Acute Febrile Illness	Rp37,000.00
K30, J06.9	Dyspepsia Syndrome + ISPA	Rp67,000.00
J06.9	ISPA	Rp48,000.00
R33.9	Retensi Urine	Rp83,000.00
N39.0	Aff. Kateter Urine (ISK)	Rp52,000.00
K52.9	Gastroenteritis Acute	Rp70,000.00
Z48.00	Wound Toilet	Rp65,000.00
S09.91	Laserasi MAE (Meatus Akustikus Eksternus)	Rp45,000.00

### D. Data Transformation

Untuk mengetahui perulangan dari suatu transaksi perlu diketahui id transaksinya, dalam hal ini yaitu *id\_pasien* sebagai *sequenceID*. *Event Identifier* atau *eventID* digunakan sebagai *timestamp* atau penanda waktu dari *itemset*, dalam hal ini yaitu tanggal pemeriksaan pasien (*tgl\_periksa*). Atribut *size* didapat dari jumlah diagnosis penyakit pasien pada setiap kedatangan. Sedangkan atribut *items* digunakan untuk mendapatkan pola asosiasi kedatangan batasan pasien ke BPI Albasrah Wajak. Atribut *items* akan diisi dengan kode penyakit pasien yang telah disesuaikan dengan aturan ICD X.

Dalam proses pembentukan aturan atau *rules* pada *sequential pattern mining* menggunakan algoritma SPADE, perlu menghilangkan atribut-atribut yang tidak diperlukan seperti nama, alamat, jenis kelamin, usia, dan total pembayaran. Atribut yang dibutuhkan yaitu *id\_pasien* sebagai *sequenceID*, *tgl\_periksa* sebagai *eventID*, kode penyakit sebagai *items*, dan jumlah penyakit yang terdiagnosa sebagai *size*. Hal tersebut dikarenakan dalam algoritma SPADE perlu keempat atribut tersebut untuk mencari data yang memiliki urutan kejadian yang berulang kali terjadi pada waktu yang berbeda.

Dari data pemeriksaan pasien tersebut, dapat dibentuk ke dalam tabel *sequence* atau urutan vertikalnya. Tabel *sequence* dibuat dengan cara mengelompokkan masing-masing *id\_pasien* ke dalam label *sequenceID*. Kemudian diberi nomor urutan dalam bentuk label *eventID*. Pada hal ini, *eventID* diasumsikan dalam waktu 1 bulan yaitu 30 hari. Sehingga penulisan *eventID* pada kedatangan pasien pertama kalinya ditulis dengan tanggal pertama dia datang. Pada kedatangan selanjutnya, *eventID* ditulis dengan menjumlahkan tanggal pertama dan selisih hari kedatangan kedua. Begitu pula untuk penulisan *eventID* selanjutnya. Kemudian mengelompokkan *items* dengan menuliskan kode diagnosis penyakit, serta *size* yang diisi dengan jumlah penyakit yang terdiagnosis pada sekali kedatangan.

TABEL IV  
DATA FORMAT VERTICAL DATABASE ALGORITMA SPADE

sequenceID	eventID	size	items
1	18	1	{R11.10}
2	18	1	{K21}
3	18	1	{R50.84}
4	18	2	{J06.9, K30}

sequenceID	eventID	size	items
5	18	1	{J06.9}
6	18	1	{R33.9}
6	23	1	{N39.0}
7	18	1	{K52.9}
8	18	1	{Z48.00}
9	18	1	{S09.91}
10	18	1	{Z48.00}
10	26	1	{Z48.00}

#### E. Pattern Evaluation

Pada tahap ini, *pattern evaluation* dilakukan dengan menuliskan baris kode dengan *software R* versi 4.0.4. Data yang diperoleh dari database BPI Albasrah diubah dulu ke dalam bentuk transaksi. Hal ini dikarenakan proses sekuensial mendeteksi suatu transaksi.

#### F. Implementasi Sequential Pattern Mining dengan Algoritma SPADE

Proses menentukan *sequential pattern mining* menggunakan algoritma SPADE dilakukan dengan menggunakan *software R* versi 4.0.4. Data yang telah diubah sesuai dengan format *sequenceID*, *eventID*, *size*, dan *items* di-export ke dalam *software R* dalam format teks (.txt).

Data diubah dulu menjadi model transaksi. Kemudian menuliskan fungsi *cspade()* dan parameter-parameter yang diperlukan untuk mendapatkan hasilnya. Dari hasil yang didapatkan, nilai *support* yang muncul tergolong rendah. Oleh karena itu, perlu ditambahkan nilai *confidence* untuk mengetahui apakah *rule* tersebut kuat dan dapat direkomendasikan.

Proses menentukan *sequential pattern mining* menggunakan algoritma SPADE dilakukan dengan menggunakan *software R* versi 4.0.4. Data yang telah diubah sesuai dengan format *sequenceID*, *eventID*, *size*, dan *items* di-import ke dalam *software R* dalam format teks (.txt).

Data diubah dulu menjadi model transaksi. Kemudian menuliskan fungsi *cspade()* dan parameter-parameter yang diperlukan untuk mendapatkan hasilnya. Dari hasil yang didapatkan, nilai *support* yang muncul tergolong rendah. Oleh karena itu, perlu ditambahkan nilai *confidence* untuk mengetahui apakah *rule* tersebut kuat dan dapat direkomendasikan.

#### G. Perhitungan Manual Sequential Pattern dengan Algoritma SPADE

Perhitungan manual untuk mencari *sequential pattern* dengan algoritma SPADE menggunakan beberapa data transaksi dari data pemeriksaan medis pasien BPI Albasrah Wajak. Proses pada algoritma SPADE dapat dilihat pada gambar berikut ini.

##### 1) Menghitung Frequent 1-Sequence Data Pemeriksaan Medis

Pada *vertical database*, atribut yang menjadi acuan yaitu atribut *sequenceID* dan *eventID*. Berikut adalah tabel *vertical database* data pemeriksaan medis pasien untuk menghitung *frequent 1-sequence* dengan memberi inputan nilai *minimum support* sebanyak 0.001.

TABEL V  
DAFTAR FREQUENT 1-SEQUENCE

Z48.00		Z48.02		Z41.2		Z48.816		L02.91	
sid	eid	sid	eid	sid	eid	sid	eid	sid	eid
157	25	157	34	195	25	195	30	157	20
157	28	157	38	671	29	671	31	157	22
568	31	607	39					568	28
607	31	1353	46						
607	33								
1400	48								

T14		H61.20	
sid	eid	sid	eid
607	28	731	30
607	36	731	32
1353	43		
1400	45		

##### 2) Menghitung Frequent 2-Sequence Data Pemeriksaan Medis

Hasil dari *frequent 1-sequence* tersebut digunakan untuk membentuk *frequent 2-sequence*. Caranya yaitu dengan melakukan *join* untuk masing-masing *frequent 2-sequence*. Salah satu contoh proses *join* yaitu apabila itemset Z48.00 di-join dengan itemset Z48.02, maka kemungkinan hasil yang didapat adalah *sequence* Z48.02.

TABEL VI  
DAFTAR FREQUENT 2-SEQUENCE

Z48.00 -> Z48.00			Z48.00 -> Z48.02		
sid	eid (Z48.00)	eid (Z48.00)	sid	eid (Z48.00)	eid (Z48.02)
157	22	25	157	28	34
157	25	28			
607	31	33			
Z48.02 -> Z48.02			Z41.2 -> Z48.816		
sid	eid (Z48.02)	eid (Z48.02)	sid	eid (Z41.2)	eid (Z48.816)
157	34	38	195	25	30
			671	29	31
L02.91 -> Z48.00			T14 -> Z48.02		
sid	eid (L02.91)	eid (Z48.00)	sid	eid (T14)	eid (Z48.02)
157	20	22	607	36	39
568	28	31	1353	43	46
H61.20 -> H61.20			T14 -> Z48.00		
sid	eid (H61.20)	Eid (H61.20)	sid	eid (T14)	eid (Z48.00)
731	30	32	607	28	31
			1400	45	48
			Z48.00 -> T14		
			sid	eid (Z48.00)	eid (T14)
			607	33	36

##### 3) Menghitung Frequent 3-Sequence Data Pemeriksaan Medis

Hasil dari *frequent 1-sequence* dan *frequent 2-sequence* digunakan untuk membentuk *frequent 3-sequence*. Caranya yaitu dengan melakukan *join* antar masing-masing *frequent sequence*. Hasil dari *frequent 3-sequence* dapat dilihat pada tabel di bawah ini.



TABEL VII  
DAFTAR FREQUENT 3-SEQUENCE

sid	eid1	eid2	Sequence
157	20	22	L02.91, Z48.00 -> Z48.00
157	22	25	Z48.00, Z48.00 -> Z48.02
157	25	28	Z48.00, Z48.00 -> Z48.02
157	28	34	Z48.00, Z48.02 -> Z48.02
607	28	31	T14, Z48.00 -> Z48.00
607	31	33	Z48.00, Z48.00 -> T14
607	33	36	Z48.00, T14 -> Z48.00

#### 4) Menghitung Frequent 4-Sequence Data Pemeriksaan Medis

Hasil dari beberapa frequent (k-1)-sequence digunakan untuk membentuk frequent 4-sequence. Caranya yaitu dengan melakukan join antar masing-masing frequent sequence. Hasil dari frequent 4-sequence dapat dilihat pada tabel di bawah ini.

TABEL VIII  
DAFTAR FREQUENT 4-SEQUENCE

sid	eid1	eid2	eid3	Sequence
157	20	22	25	L02.91, Z48.00, Z48.00 -> Z48.02
157	22	25	28	Z48.00, Z48.00, Z48.00 -> Z48.02
157	25	28	34	Z48.00, Z48.00, Z48.02 -> Z48.02
607	28	31	33	T14, Z48.00, Z48.00 -> T14
607	31	33	36	Z48.00, Z48.00, T14 -> Z48.02

#### 5) Menghitung Frequent 5-Sequence Data Pemeriksaan Medis

Cara yang sama dilakukan juga untuk menentukan frequent 5-sequence yaitu dengan melakukan join antar masing-masing frequent (k-1)-sequence. Hasil dari frequent 5-sequence dapat dilihat pada tabel di bawah ini.

TABEL IX  
DAFTAR FREQUENT 5-SEQUENCE

sid	eid1	eid2	eid3	eid4	Sequence
157	20	22	25	28	L02.91, Z48.00, Z48.00, Z48.00 -> Z48.02
157	22	25	28	34	Z48.00, Z48.00, Z48.00, Z48.02 -> Z48.02
607	28	31	33	36	T14, Z48.00, Z48.00, T14 -> Z48.02

#### 6) Rules yang Terbentuk

Hasil dari frequent k-sequence dengan nilai minimum support sebesar 0.001 dan nilai minimum confidence sebesar 0.2, sehingga didapatkan 16 rule yang dapat dilihat pada tabel berikut ini.

TABEL X  
RULES DENGAN NILAI CONFIDENCE < 0.2

	rule	support	confidence	lift
11	<{Z48.02},{Z48.00}>=><{Z48.00}>	0.001295	1	27.08772
13	<{L02.91},{Z48.00}>=><{Z48.00}>	0.001295	1	27.08772
14	<{Z48.02},{Z48.00},{Z48.00}>=><{Z48.00}>	0.001295	1	27.08772
15	<{L02.91},{L02.91}>=><{Z48.00}>	0.001295	1	27.08772

	rule	support	confidence	lift
2	<{Z41.2}>=><{Z48.816}>	0.002591	0.5	110.2857
8	<{T14.1}>=><{Z48.00}>	0.001295	0.4	10.83509
20	<{N40.0}>=><{N40.0}>	0.001295	0.4	123.52
29	<{J06.9},{J06.9}>=><{J06.9}>	0.001295	0.4	2.072483
12	<{Z48.00},{Z48.00}>=><{Z48.00}>	0.001943	0.333333	9.02924
27	<{J44.9}>=><{J06.9}>	0.001295	0.333333	1.727069
6	<{L02.91}>=><{Z48.00}>	0.001295	0.285714	7.739348
21	<{L02.91}>=><{L02.91}>	0.001295	0.285714	63.02041
36	<{H66.0}>=><{H66.0}>	0.001295	0.285714	63.02041
37	<{H61.20}>=><{H61.20}>	0.001943	0.230769	27.40828
5	<{Z48.00},{Z48.00}>=><{Z48.02}>	0.001295	0.222222	38.12346
10	<{Z48.02}>=><{Z48.00}>	0.001295	0.222222	6.019493

Apabila ditambah rentang waktu atau maxgap sebesar 5 hari, maka menghasilkan 9 rule yang tertera pada tabel di bawah ini.

TABEL XI  
RULES DENGAN NILAI CONFIDENCE < 0.2 DAN NILAI MAXGAP = 5

	rule	support	confidence	lift
9	<{Z48.02},{Z48.00},{Z48.00}>=><{Z48.00}>	0.001295	1	27.08772
10	<{Z48.02},{Z48.00}>=><{Z48.00}>	0.001295	1	27.08772
1	<{Z41.2}>=><{Z48.816}>	0.002591	0.5	110.2857
5	<{T14.1}>=><{Z48.00}>	0.001295	0.4	10.83509
8	<{Z48.00},{Z48.00}>=><{Z48.00}>	0.001295	0.333333	9.02924
3	<{L02.91}>=><{Z48.00}>	0.001295	0.285714	7.739348
15	<{H66.0}>=><{H66.0}>	0.001295	0.285714	63.02041
16	<{H61.20}>=><{H61.20}>	0.001943	0.230769	27.40828
7	<{Z48.02}>=><{Z48.00}>	0.001295	0.222222	6.019493

Lift ratio merupakan salah satu cara untuk menentukan apakah aturan yang didapat dari proses sekuensial kuat atau tidak. Pada umumnya, nilai lift ratio berada di antara 0 sampai dengan tak terhingga. Semakin tinggi lift ratio, semakin besar pula aturan itu dapat direkomendasikan.

Dari aturan-aturan yang didapat, diambil 5 aturan teratas yang sudah diurutkan berdasarkan nilai confidence. Penjelasan dari aturan-aturan tersebut adalah sebagai berikut.

##### a. Rule <{Z48.02},{Z48.00},{Z48.00}>=><{Z48.00}>

Aturan ini memiliki nilai support sebesar 0.001295 dan nilai confidence sebesar 1. Nama penyakit untuk kode penyakit Z48.02 yaitu encounter of removal sutures, pertemuan pasien untuk pelepasan jahitan, dan kode penyakit Z48.00 yaitu wound toilet, pembersihan luka atau perawatan luka. Pada rule ini, diasumsikan bahwa pasien yang baru melepas jahitan pada lukanya akan melakukan beberapa kali perawatan luka (wound toilet) pada kunjungan berikutnya.

##### b. Rule <{Z48.02},{Z48.00}>=><{Z48.00}>

Aturan ini memiliki nilai support sebesar 0.001295 dan nilai confidence sebesar 1. Nama penyakit untuk kode penyakit Z48.02 yaitu encounter of removal sutures, pertemuan pasien untuk pelepasan jahitan, dan kode penyakit Z48.00 yaitu wound toilet, pembersihan luka atau perawatan luka. Pada rule ini, diasumsikan bahwa pasien yang baru melepas jahitan pada lukanya akan melakukan perawatan luka (wound toilet) pada kunjungan ke-3.

c. Rule  $\langle \{Z41.2\} \rangle \Rightarrow \langle \{Z48.816\} \rangle$

Aturan ini memiliki nilai *support* sebesar 0.002591 dan nilai *confidence* sebesar 0.5. Nama penyakit untuk kode penyakit Z41.2 yaitu *encounter for routine and ritual male circumcision*, pertemuan pasien laki-laki untuk melakukan sirkumsisi atau sunat, dan kode penyakit Z48.816 yaitu *encounter for surgical aftercare following surgery on the genitourinary system*, pertemuan untuk melakukan perawatan pada sistem genitourinari. Pada *rule* ini, diasumsikan bahwa pasien laki-laki yang baru melakukan sirkumsisi atau sunat akan melakukan perawatan *post-circumcision* pada kunjungan berikutnya.

d. Rule  $\langle \{T14.1\} \rangle \Rightarrow \langle \{Z48.00\} \rangle$

Aturan ini memiliki nilai *support* sebesar 0.001295 dan nilai *confidence* sebesar 0.4. Nama penyakit untuk kode penyakit T14.1 yaitu *vulnus laceratum or open wound of unspecified body region*, luka terbuka atau luka robek, dan kode penyakit Z48.00 yaitu *wound toilet*, pembersihan luka atau perawatan luka. Pada *rule* ini, diasumsikan bahwa pasien yang baru mengalami luka robek akan melakukan perawatan luka (*wound toilet*) pada kunjungan berikutnya.

e. Rule  $\langle \{Z48.00\}, \{Z48.00\} \rangle \Rightarrow \langle \{Z48.00\} \rangle$

Aturan ini memiliki nilai *support* sebesar 0.001295 dan nilai *confidence* sebesar 0.33. Nama penyakit untuk kode penyakit Z48.00 yaitu *wound toilet*, pembersihan luka atau perawatan luka. Pada *rule* ini, diasumsikan bahwa pasien yang melakukan perawatan pada lukanya akan melalui beberapa tahapan perawatan lagi hingga luka itu sembuh.

#### H. Uji Coba

Proses pengujian dilakukan menggunakan data pemeriksaan medis pasien BPI Albasrah Wajak. Pada penelitian ini, data pemeriksaan medis menggunakan nilai *minimum support* 0.001 dan nilai *minimum confidence* 0.2 untuk mengetahui kedatangan kembali pasien ke BPI Albasrah Wajak dalam waktu 1 bulan. Sedangkan data pemeriksaan medis yang akan digunakan untuk uji coba yaitu pada bulan Januari hingga bulan Februari 2021, sehingga dapat membuktikan apakah *rule-rule* yang ada telah terbukti secara akurat. Sampel yang digunakan untuk uji coba yaitu 7 pasien dengan hasil *frequent sequence* yang berbeda – beda.

Dari hasil perhitungan *frequent k-sequence* pada tabel 3-7, dapat dilihat bahwa algoritma SPADE dapat menentukan pola kedatangan pasien ke BPI Albasrah Wajak dari data pemeriksaan medis pasien berdasarkan kode penyakit.

TABEL XII  
HASIL UJI COBA

id	Sequence	Uji Coba			Ket.
		tgl1	tgl2	tgl3	
568	Z48.00 -> Z48.00	21 Jan	27 Jan		Terbukti
568	Z48.02, Z48.00 -> Z48.00	14 Jan	21 Jan	27 Jan	Terbukti
607	Z48.02 -> Z48.00	23 Jan	28 Jan		Terbukti
568	L02.91 -> Z48.00	19 Jan	23 Jan		Terbukti

id	Sequence	Uji Coba			Ket.
		tgl1	tgl2	tgl3	
1621	Z41.2 -> Z48.816	7 Feb	11 Feb		Terbukti
700	T14.1 -> Z48.00				Tidak Terbukti
1353	T14, Z48.00 -> Z48.00				Tidak Terbukti
1045	H66.00 -> H66.00	14 Jan	19 Jan		Terbukti
1776	L02.91, Z48.00 -> Z48.00	26 Jan	31 Jan	5 Feb	Terbukti

Dari proses uji coba yang telah dilakukan, terdapat 5 dari 7 pasien yang memenuhi nilai *minimum support* pada hasil *frequent sequence* selama 1 bulan. Sehingga besar nilai akurasi yang dapat dihitung adalah sebagai berikut.

$$\text{Nilai Akurasi} = \frac{5}{7} \times 100\% = 71.43\% \quad (1)$$

Dari perhitungan akurasi yang telah dilakukan, diperoleh nilai akurasi sebesar 71.43%. Oleh karena itu, dapat disimpulkan bahwa pada penelitian ini algoritma SPADE terbukti secara akurat untuk mengetahui kedatangan kembali pasien ke BPI Albasrah Wajak.

#### IV. KESIMPULAN DAN SARAN

##### A. Kesimpulan

Kesimpulan dari penelitian ini adalah sebagai berikut.

1. Pola sekuensial dari penelitian ini mendapatkan hasil yang baik dan dapat dijadikan salah satu cara untuk mengetahui kedatangan kembali pasien ke BPI Albasrah Wajak.
2. Uji coba kekuatan *rule* dapat dilakukan dengan menggunakan nilai *support*, nilai *confidence*, dan *lift ratio*.
3. Semakin tinggi nilai *lift ratio* dan semakin nilai *confidence* mendekati angka 1.0, maka *rule* tersebut juga semakin kuat atau semakin direkomendasikan.
4. Berdasarkan uji coba yang telah dilakukan, kekuatan *rule* dipengaruhi oleh masukan nilai *minimum support* dan jumlah transaksi. Sehingga dapat diambil kesimpulan bahwa algoritma SPADE terbukti secara akurat dengan presentase sebesar 71,43% untuk mencari pola sekuensial dari data pemeriksaan medis untuk mengetahui kedatangan kembali pasien ke BPI Albasrah Wajak di waktu yang akan datang.

##### B. Saran

Saran dari penelitian ini adalah sebagai berikut.

1. Rentang waktu data dapat diambil lebih lama seperti dalam beberapa bulan atau 1 tahun agar *rule* yang terbentuk lebih bervariasi.
2. Nilai *minimum support* dapat dinaikkan agar *rule* yang dihasilkan memiliki keterkaitan yang kuat.

#### REFERENSI

- [1] Agrawal, R., & Srikant, R., "Proceedings of the Eleventh International Conference on Data Engineering - Mining Sequential Patterns", *IEEE Comput. Soc. Press Eleventh International Conference on Data*



- Engineering, (pp. 3-14). Taipei, Taiwan, 1995. doi:10.1109/icde.1995.380415
- [2] Zaki, M. J., "SPADE: An Efficient Algorithm for Mining Frequent Sequences", The Netherlands: Kluwer Academic Publishers, 2001.
- [3] Agrawal, R., & Srikant, R., "Fast Algorithm for Mining Association Rules", *The International Conference on Very Large Databases*, pp. 487 – 499, 1994.
- [4] Fournier-Viger, P. L., "A Survey of Sequential Pattern Mining. *Data Science and Pattern Recognition*", 54-77, 2017.
- [5] Zaki, M. J., "Fast Mining of Sequential Patterns in Very Large Databases", New York: The University of Rochester Computer Science Department Rochester, New York 14627, 1997.
- [6] Dini, S. K., Aidi, M. N., & Sartono, B., "Sequential Pattern Mining of Rawi Hadis (Case study: Shahih Hadis of Imam Bukhari from Software Ensiklopedi Hadis Kitab 9 Imam)", *AIP Conference Proceedings: Proceedings of the 2nd International Conference on Applied Statistics (ICAS II)*, Jawa Barat, 2017.
- [7] Saputra, D. A., Torida, E. T., & Fajar, D., "Penentuan Pola Sekuensial Data Transaksi Penjualan Menggunakan Algoritma Sequential Pattern Discovery Using Equivalent Classes (SPADE)", *Komputasi Jurnal Ilmiah Ilmu Komputer dan Matematika*, 2019.
- [8] Pemerintah Indonesia, "Peraturan Menteri Kesehatan Republik Indonesia Nomor 9 Tahun 2014 tentang Klinik", Jakarta: Sekretariat Negara, 2014.
- [9] Garmelia, E., Kresnowati, L., Irmawati, & Anggraini, M., "Bahan Ajar Rekam Medis dan Informasi Kesehatan (RMIK) tentang Klasifikasi, Kodifikasi Penyakit, dan Masalah Terkait I", Jakarta: Pusat Pendidikan Sumber Daya Manusia Kesehatan, 2017.
- [10] Sugiyono, D., "Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif dan R&D", 2014.

