

## **BAB IV**

### **METEDOLOGI PENELITIAN**

#### **4.1. Desain Penelitian**

Desain penelitian yang digunakan adalah pendekatan *cross sectional*. Penelitian *cross sectional* dicirikan dengan satu pengukuran atau observasi untuk satu unit. Pendekatan *cross sectional* digunakan dengan cara pendekatan observasi atau pengumpulan data sekaligus pada satu waktu (Notoatmodjo, 2002).

Pendekatan *cross sectional* yang digunakan adalah pengumpulan (*entry*) data rekam medik data kunjungan pasien rawat inap Rumah Sakit Jiwa Grhasia Jl Kaliurang Km 17 Pakem Sleman D.I Yogyakarta. Pengumpulan (*entry*) data yang dilakukan sebanyak 940 data transaksi data kunjungan pasien rawat inap Rumah Sakit Jiwa Grhasia.

Menurut Notodiputro (2015) dalam (Septiani, 2015) pengambilan sampel pada *data mining* berbeda halnya dengan pengambilan sampel pada statistika secara umum, karena statistika memerlukan contoh (sampel) untuk memperoleh pengetahuan baru, sedangkan *data mining* memanfaatkan *database* untuk menemukan pengetahuan baru. Analisis yang digunakan dalam *data mining* dengan menggunakan algoritma *data mining* untuk menemukan struktur dan pola, dengan penarikan kesimpulan deskriptif berupa pengetahuan apa yang telah ditemukan di dalam data besar yang dianalisis.

#### **4.2. Jenis dan Sumber Data**

Jenis data yang digunakan dalam penelitian ini yaitu data sekunder. Adapun data bersumber dari data rekam medik pasien tahun 2014-2015 data kunjungan pasien jiwa rawat inap Rumah Sakit Jiwa Grhasia.

#### **4.3. Definisi Operasional Variabel**

Menurut Chourmain (2008) dalam (Septiani, 2015) definisi operasional variabel merupakan penarikan batasan yang lebih menjelaskan ciri-ciri spesifik yang lebih substantive dari suatu konsep. Tujuannya agar peneliti dapat mencapai

suatu alat ukur yang sesuai dengan hakikat variabel yang sudah didefinisikan konsepnya, maka harus memasukkan proses atau operasional alat ukur yang akan digunakan untuk kuantifikasi gejala atau variabel yang ditelitinya. Dalam penelitian ini, variabel yang digunakan yaitu diagnosa. Yang dimaksud dengan diagnosis dalam penelitian ini yaitu identifikasi sifat-sifat penyakit atau kondisi atau membedakan satu penyakit atau kondisi dari yang lainnya. Penilaian dapat dilakukan melalui pemeriksaan fisik, tes laboratorium, atau sejenisnya, dan dapat dibantu oleh program komputer yang dirancang untuk memperbaiki proses pengambilan keputusan.

#### **4.4. Populasi dan Sampel**

Konsep statistik, menerapkan adanya pengambilan sampel dalam suatu data, yang dikenal dengan istilah sampling. Sedangkan data mining menerapkan beberapa algoritma *mechine learning*, baik dalam analisis suatu data mentah yang ada maupun dalam proses pembentukan model dalam proses klasifikasi data (Han, 2006). Secara formal, populasi adalah sekumpulan objek (orang, benda, gejala alami dan gejala insani) yang memiliki sedikitnya satu karakteristik umum yang sama. Sampel adalah sebagian dari anggota populasi yang memiliki karakteristik yang sama dengan populasi. Perbedaan statistika dan Data Mining terletak pada pangkal permasalahannya. Di Statistika data (sampel) biasanya diambil dari populasi, lalu diolah dengan metode statistika yang bersesuaian, kemudian menyimpulkan hasilnya dari model yang dibuat (interpretasi).

#### **4.5. Metode Analisis Data**

Tujuan dari penelitian ini adalah untuk mengetahui mengetahui mengetahui untuk karakteristik pasien dan pola asosiasi yang terbentuk dari diagnosis setiap pasien rawat inap Rumah Sakit Jiwa Grhasia sehingga dapat memberikan informasi khususnya pihak Rumah Sakit Jiwa atau instansi terkait.

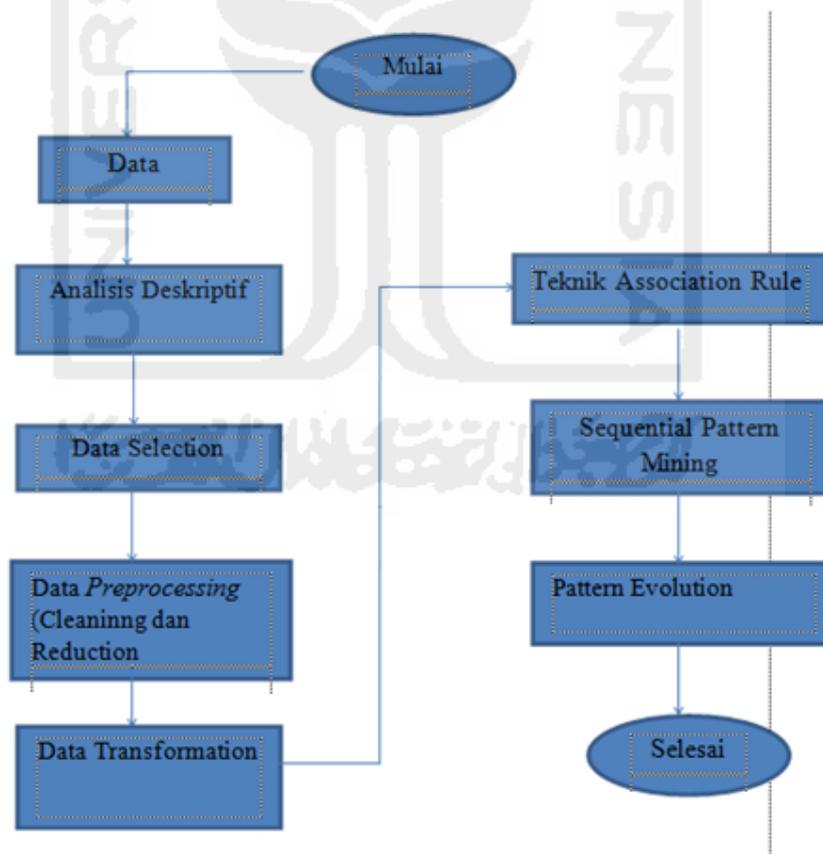
Salah satu metode dalam *data mining* yang dapat digunakan untuk mencari pola aturan adalah *Sequential Pattern Mining*. *Sequential Pattern Mining* sendiri adalah pola yang menggambarkan urutan waktu terjadinya peristiwa. Pola

tersebut dapat ditemukan apabila data yang disimpan relatif besar dan peristiwa yang berurutan terjadi beberapa kali (Agrawal, 1955).

Menurut Ardiansyah (2013) dalam (Septiani, 2015) *Sequential Pattern Mining* untuk menemukan pola agar mendapat informasi yang berguna harus dicari *frequent sequences* atau urutan peristiwa tertentu yang sering muncul. Sedangkan algoritma yang digunakan adalah algoritma SPADE. Algoritma SPADE menggunakan *id-list* vertikal untuk memudahkan pencarian dalam *database*. Algoritma SPADE dapat mencari *frequent sequence* dengan beberapa kali pencarian *database* saja.

#### 4.6. Langkah-langkah Penelitian

Dalam penelitian ini, eksplorasi data dilakukan sesuai dengan langkah-langkah dalam proses yang tergambar dalam gambar berikut:



Gambar 4.1. Alur Penelitian

Adapun penjelasan dari gambar diatas adalah:

### 1. Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari database rekam medik data kunjungan pasien jiwa rawat inap Rumah Sakit Jiwa Grhasia pada tahun 2014-2015

### 2. Analisis Deskriptif

Pada analisis ini bertujuan untuk mengetahui karakteristik dari setiap kunjungan pasien jiwa rawat inap Rumah Sakit Jiwa Grhasia pada tahun 2014-2015 berdasarkan atribut data yang ada yaitu nomor rekam medis, jenis diagnosis, nama diagnosis, umur, jenis kelamin, alamat, nama bangsal yang dipakai selama di rawat, jenis kelas, keadaan, dan cara keluar.

### 3. *Data Selection*

Dalam pemilihan data, data yang digunakan adalah hasil rekam medik pasien jiwa di inap Rumah Sakit Jiwa Grhasia pada tahun 2014-2015. Data asli yang didapat dari database sumber memiliki cukup banyak atribut / *field* yang tidak dibutuhkan sebagai atribut input dalam melakukan data mining. Oleh karena itu dilakukan seleksi atribut yang akan dipakai sebagai input pada proses data mining. Adapun atribut yang dipilih adalah Kode\_Pasien, waktu\_periksa (event), dan Diagnosa\_Penyakit. Keseluruhan data ini digunakan untuk penggalian informasi.

### 4. *Data Preprocessing*

Tahap ini dilakukan dua hal yaitu *data cleaning* dan *data reduction*

#### a. *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 (Ramadhani, 2015). Karena dalam penelitian, atribut ini tidak digunakan dalam proses pembentukan aturan. Setelah data dibersihkan, kemudian dilakukan perbaikan data dengan cara dihaluskan (*smoothing*). Untuk data hasil diagnosis

menghaluskannya yaitu dengan melakukan penyingkatan atau pengkodean sesuai dengan ICD (*Internasional Classificatoin Diseas revisi*) beberapa nama penyakit atau gejala yang didiagnosa agar mempermudah dalam menganalisis data..

#### b. *Data reduction*

Bertujuan untuk mengurangi ukuran data yang besar. Misalnya melakukan pengkategorian data. Dalam penelitian ini tidak terjadi proses pengkategorian data karena tiap hasil diagnosis sudah mempunyai kode masing-masing dan sesuai tujuan yang akan dicapai, tiap *item* (diagnosis) tidak dapat disamakan atau digabungkan dalam satu kategori dengan *item* (diagnosis) lainnya.

##### 1. *Data transformation*

Pada tahap ini dilakukan tranformasi yang meliputi organisasi data yang telah dipilih pada tahapan sebelumnya, melakukan konversi dari satu tipe data ke lainnya dengan membuat tabel *co-occurrence* yang bernilai 1 dan 0. Proses transformasi ini dilakukan dengan bantuan program R 3.0.3.

##### 2. *Data mining*

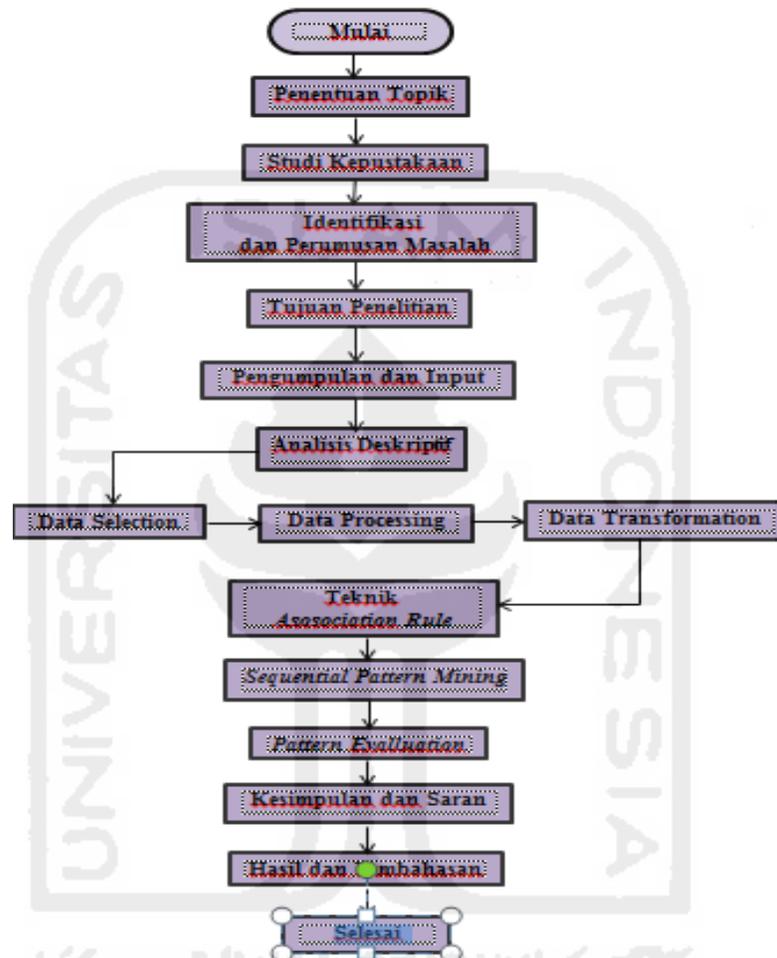
Pada tahap ini data yang telah ditransformasi pada tahapan sebelumnya digali menggunakan satu atau lebih teknik untuk mendapatkan pola-pola yang menarik perhatian dan menjadi pengetahuan baru. Teknik *data mining* yang digunakan dalam penelitian ini yaitu *sequential pattern mining*, dengan langkah-langkah yang perlu dilakukan yaitu: Menentukan *support* dan *minimum support* kemudian menentukan *confidence* dan *minimum confidence*

##### 3. *Pattern evaluation*

Dalam proses ini, pencarian *rules* dilakukan dengan bantuan program R 3.0.3 yang dapat mengidentifikasi pola yang benar-benar menarik berdasarkan data rawi hadis. Dengan demikian akan diperoleh suatu informasi yang bermanfaat dari hasil aturan atau asosiasinya yang terbentuk.

#### 4.7. Struktur Penelitian

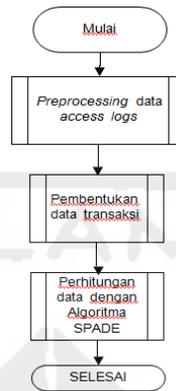
Adapun tahapan penelitian ini secara keseluruhan, dapat dilihat dari gambar berikut:



Gambar 4.2. Struktur Penelitian

## 4.8. Sequential Pattern Mining Menggunakan Algoritma SPADE

### 4.8.1. Tahap proses algoritma SPADE adalah sebagai berikut:



**Gambar 4.3.** Alur Algoritma SPADE

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

Langkah dalam mencari *frequent 1-sequence* adalah sebagai berikut:

1. Dimasukkan data yang diperoleh dari *database sequence* yang terdiri dari *sid*, *eid*, *id-list* (pasangan *sid*, *eid*) dan *itemset*.
2. Dilakukan perulangan untuk setiap *itemset*, dicek pada setiap *sid* dan *eid*.
3. Dilakukan pengecekan apakah *itemset* muncul dalam pasangan *id-list*.
4. Apabila muncul maka *id-list* tersebut dimasukkan ke dalam *id-list itemset* tersebut dan membentuk *1-sequence*.
5. Dilakukan pengecekan apakah *sid* yang dimasukkan sudah muncul sebelumnya, apabila belum maka nilai *support* ditambah 1. Apabila sudah ada maka *id-list* hanya dimasukkan saja, tapi nilai *support* tidak ditambahkan.
6. Dilakukan pengecekan jumlah *support* untuk masing – masing *1-sequence*, apabila lebih dari *min\_sup*, maka diterima sebagai *frequent 1-sequence*.

#### b. Mencari *frequent 2-sequence* adalah sebagai berikut:

Tahap mencari *frequent 2-sequence* adalah sebagai berikut:

1. Dilakukan input  $min\_sup$  dan data  $2-sequence$ .
2. Dilakukan pengecekan untuk setiap anggota  $2-sequence$  apakah  $supportnya$  memenuhi syarat lebih dari atau sama dengan  $min\_sup$ . Apabila memenuhi maka dimasukkan dalam data  $frequent\ 2-sequence$ .

b. Mencari  $k-sequence$

Tahap mencari  $k-sequence$  adalah sebagai berikut:

1. Diinputkan  $min\_sup$  dan  $frequent\ (k-1)-sequence$ .
2. Untuk setiap anggota  $(k-1)-sequence$  dilakukan pengecekan apakah  $prefix\ item$  ke- $i$  (a) dan  $prefix\ item$  setelah  $i$  atau ke- $j$  (b) sama atau tidak.
3. Jika memiliki  $prefix$  yang sama, kemudian dicek hubungan item  $i$  dan  $j$ :  
a. Jika  $a_i$  dan  $a_j$  maka dicek apakah  $eid\ i = eid\ j$ , jika memenuhi maka terbentuk  $k-sequence\ a_i, j$ .