

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Aplikasi Data Mining

*Data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar (Turban et al, 2005).

Dalam aplikasinya, *data mining* merupakan bagian dalam proses *Knowledge Discovery in Database* (KDD) yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik. *Data mining* merupakan suatu bagian langkah yang penting dalam proses *knowledge discovery in database* terutama berkaitan dengan ekstraksi dan penghitungan data-data yang ditelaah. Adapun proses KDD adalah suatu proses menggunakan database dalam jumlah besar yang meliputi penyeleksian data, *preprocessing* (subsampling), *transformasi* data, menerapkan metode *data mining* (algoritma) untuk menghitung pola data, dan mengevaluasi pola-pola yang terbentuk dari proses *data mining* sehingga dapat menjadi informasi / pengetahuan baru yang bermanfaat (Fayyad, 1996).

1. Menurut Yanchang Zhao dan Yonghua Cen ini mempersembahkan buku *Data Mining* dengan menggunakan Program R, harapannya dengan buku ini pembaca dapat mendemonstrasikan data mining dengan menggunakan program open source. Selain *open source*, program R ini dapat digunakan untuk analisis statistika dan data mining. Selain itu, buku ini tidak hanya menyajikan teori, namun juga dilengkapi dengan praktik yang dilengkapi dengan code atau sintaks-sintaks dalam analisis data mining.

2. Menurut Mohammed J. Zaki and Wagner Meira JR., 2014) *Data mining* adalah proses menemukan wawasan, pola menarik, serta deskriptif, dimengerti, dan prediksi dari model data dari data berskala besar. Dengan melihat sifat dasar data yang dimodelkan sebagai matriks data, yang menekankan pandangan geometris dan aljabar, serta interpretasi probabilistik dari data
3. Menurut Ledolter, Johannes, 2001 menyatakan dengan adanya database dalam jumlah besar, maka dengan teknik data mining akan didapatkan pola dan model tertentu dari database tersebut. Penerapan data mining dan analisis statistika merupakan alat yang tepat menggali informasi yang bermanfaat dari data yang besar, sehingga diperoleh model sederhana yang mendeskripsikan data tersebut
4. Menurut Ullah, 2010 Jumlah luas data yang disimpan dalam *database* medis membutuhkan pengembangan alat khusus untuk mengakses data, analisis data, penemuan pengetahuan. Penggunaan system informasi medis dan pembesaran ledakan *database* medis memerlukan analisis data manual konvensional akan digabungkan dengan metode analisis dengan bantuan komputer yang kompeten. Dalam tulisan ini, penulis menggunakan teknik *data mining* untuk analisis data, akses data dan prosedur penemuan pengetahuan melalui eksperimen dan praktik konsisten, mampu dan cepat untuk studi bidang tertentu. Hasil yang diperoleh akan diuji dengan menerapkan pendekatan ke database, gudang data dan semua ukuran yang berbeda dengan nilai-nilai entri yang berbeda. Hasil yang dibentuk dalam penggunaan yang berbeda misalnya diagnose pasien, frekuensi penyakit yang berbeda

### **2.3. Aplikasi *Sequential Pattern Mining***

*Sequential pattern mining* pertama kali diperkenalkan oleh Agrwal dan Srikant. Proses *sequential pattern mining* dapat digambarkan misalnya, diberikan sejumlah *sequence*. Setiap *sequence* terdiri atas sederetan elemen, dan setiap elemen terdiri atas sejumlah item, serta diberikan nilai minimum support. Penggalian pola *sequential* adalah pencarian semua subsequence berulang, yaitu subsequence yang

frekuensi kejadiannya lebih besar dari minimum-support.(Agrawal, 1995). Untuk menyelesaikan permasalahan sekuensial ini dapat dilakukan dengan beberapa metode antara lain *Generalizes Sequential Pattern* atau disebut *GSP*, *FreeSpan*, dan *PrefixSpan*. Sebagai contoh proses *sequential pattern mining*, terdapat table transaksi penjualan yang berisikan customer, tanggal dan item. Dari table transaksi tersebut kemudian dibentuk sequence transaksi tersebut kemudian dibentuk sequence transaksi berdasarkan customer dan diurutkan berdasar tanggal sehingga membentuk beberapa sequence. (Ardiansyah ,2013). Adapun beberapa penelitian dari aplikasi *sequential pattern mining* dengan *Sequential Pattern Discovery using Equivalent classes (SPADE)* antara lain adalah :

*Sequential Pattern Mining* pada Data Transaksi Penjualan Menggunakan Algoritma *Sequential Pattern Mining Discovery using Equivalent classes (SPADE)* (Ardiansyah, 2013) Penelitian ini mencoba menggunakan algoritma *Sequential Pattern Mining Discovery using Equivalent classes (SPADE)* untuk mencari *Sequential Pattern* pada data transaksi penjualan. Algoritma *sequential pattern discovery using equivalent classes (SPADE)* dapat diterapkan pada data transaksi penjualan. Dari data transaksi yang digunakan, dicari *frequent sequence* (pola urutan pembelian barang yang sering muncul) yang dapat dibentuk. Hasil *frequent sequence* menunjukkan pola belanja konsumen yang sering muncul. Pola belanja konsumen tersebut kemudian dibentuk rule yang diuji menggunakan nilai lift rasio, apabila nilai lift rasio lebih dari 1, maka rule dianggap kuat. Hanya ditemukan 11 rule dari total 6097 rule yang dihasilkan memiliki nilai lift rasio dibawah 1. Sehingga disimpulkan bahwa algoritma SPADE akurat untuk mencari pola sekuensial dari data transaksi penjualan.

Dalam jurnal yang berjudul “*Sequential Pattern Mining* Pada Pencarian Pola Perilaku Pengguna Internet Menggunakan Algoritma SPADE” menjelaskan tentang Perilaku penggunaan internet dapat dilihat melalui sebuah *file* yang mencatat segala informasi transaksi penggunaan internet yang disebut *access logs*. Nilai *minimum support* dan *minimum confidence* ditentukan untuk menemukan *rule-rule* yang

bermanfaat. Nilai kekuatan *rule* atau *lift ratio* ditentukan dengan besar minimum *support* sedangkan nilai minimum *confidence* tidak berpengaruh terhadap besar nilai *lift ratio*. Dilanjutkan dengan pembentukan data transaksi dan perhitungan SPADE dengan mengkombinasikan *itemset* dan menghitung *frequent*-nya untuk mendapatkan *rule* yang kemudian akan dicari kekuatan setiap *rule* dengan menghitung *lift ratio*-nya. Nilai *lift ratio* terbesar terdapat pada Fakultas Peternakan disaat jam kerja dengan besaran nilai 18,868 dengan *rule* jika mengakses mozilla.com maka mengakses bbc.co.uk (Yunianto, 2014; Dewi, 2014; yudistira, 2014).

*Sequential Pattern Mining Dengan Spade Untuk Prediksi Pembelian Spare Part dan Aksesoris Komputer Pada kedatangan Kembali Konsumen* (Juliastio & Gunawan, 2015), hasil frequent 4-sequence yang diperoleh dalam penelitian ini mendapatkan hasil yang baik dan dapat dijadikan sebuah prediksi pembelian spare part dan aksesoris komputer pada kedatangan kembali konsumen di toko *Seagate Computer* Situbondo. Berdasarkan uji coba yang dilakukan menggunakan parameter nilai minimum support, minimum confidence dan jumlah data, kekuatan rule yang dihasilkan terpengaruh pada input nilai minimum support dan jumlah data transaksi, sedangkan nilai minimum *confidence* tidak berpengaruh terhadap kekuatan rule. Sehingga disimpulkan bahwa algoritma SPADE akurat untuk mencari pola sekuensial dari data transaksi penjualan untuk memprediksi kedatangan kembali konsumen.

### 2.3. Data Rekam Medis

Dalam PERMENKES No: 269/MENKES/PER/III/2008 yang dimaksud rekam medis adalah berkas yang berisi catatan dan dokumen antara lain identitas pasien, hasil pemeriksaan, pengobatan yang telah diberikan, serta tindakan dan pelayanan lain yang telah diberikan kepada pasien. Catatan merupakan tulisan-tulisan yang dibuat oleh dokter mengenai tindakan-tindakan yang dilakukan kepada pasien dalam rangka pelayanan kesehatan. Bentuk Rekam Medis dalam berupa manual yaitu tertulis lengkap dan jelas dan dalam bentuk elektronik sesuai ketentuan. Berikut beberapa penelitian yang menggunakan data rekam medis :

*Mining Assosiation Rules dan Sequential Pattern* dari Data Rekam Medis Rumah Sakit DR. H Slamet Martodirdjo Pamekasan (Ramadhani, 2014) Penelitian dilakukan di Rumah Sakit DR. H Slamet Martodirdjo Pamekasan memanfaatkan algoritma *Assosiation Rules Mining* untuk dapat mengetahui pola penyakit yang diderita oleh seluruh pasien dalam waktu yang sama. Sedangkan pemanfaatan algoritma *Sequential Pattern Mining* dapat mengetahui pola-pola penyakit yang diderita oleh seluruh pasien pada waktu yang berbeda. Pada kedua metode atribut yang digunakan adalah pasien\_ID, tanggal\_periksa dan diagnose\_penyakit. Data yang dipakai adalah data rekam medis tahun 2010, 2011, 2012.

Penerapan algoritma *Sequential Patterns Discovery Using Equivalent Classes* (SPADE) untuk mencari *sequential pattern mining* Pada Data Rekam Medis dari Puskesmas Bolo di Kabupaten Bima Provinsi Nusa Tenggara Barat (rifani, 2016), penelitian ini dilakukan di Puskesmas Bolo di Kabupaten Bima Provinsi Nusa Tenggara Barat untuk mengetahui pola asosiasi dari diagnose setiap pasien Puskesmas Bolo sehingga dapat memberikan informasi khususnya pihak Puskesmas atau instansi terkait.

*Information Extraction* Data Rekam Medis Pasien Untuk Perolehan *Assosiation Rule* Sebagai Pendukung Pengambilan Keputusan (Mauladi, 2013), penelitian ini bertujuan untuk memperoleh *Assosiation Rule* sebagai pendukung pengambilan keputusan dengan menggunakan *Information Extraction* dari data rekam medis pasien. Informasi ekstraksi dari dokumen rekam medis pasien bias digunakan mengklarifikasikan jenis penyakit yang diderita berdasarkan keluhan yang disampaikan pasien, juga hasil proses mining akan menghasilkan keterkaitan antara jenis penyakit satu dengan penyakit yang lainnya yang sering muncul secara bersamaan.