

1. PENDAHULUAN

1.1 Latar Belakang

Kegiatan manusia semakin berkembang dan memunculkan kebutuhan akan data semakin berkembang pula. Penggunaan data tumbuh begitu cepat, mengakibatkan semakin besarnya jumlah data yang harus dikumpulkan dan disimpan. Basis data yang digunakan juga semakin berkembang dari segi ukuran yang akan semakin besar dan semakin beragam jenisnya. Dengan tersediannya data yang sangat melimpah tersebut ternyata informasi yang dapat digali terbatas karena keterbatasan manusia untuk melihat seluruh data dan menjadikannya informasi yang lebih bermanfaat secara langsung. Diperlukan sebuah aplikasi yang mampu mengubah data yang begitu besar menjadi informasi yang bernilai, yaitu melalui *data mining*.

Data mining dideskripsikan sebagai proses untuk melakukan ekstraksi pengetahuan dari data yang jumlahnya besar. Banyak pihak yang menggunakan data mining sebagai suatu langkah *knowledge Discovery in Database* atau KDD. *Data mining* adalah sebuah proses esensial dimana metode-metode intellegent digunakan untuk mengekstraksi pola-pola(pattern) yang terbentuk dari kata.

Terdapat beberapa jenis pola hasil proses *data mining*, diantaranya *sequential pattern*. *Sequential pattern* menggambarkan keterurutan suatu peristiwa yang terjadi beberapa kali dalam kurun waktu tertentu. Contoh dari penggunaan *sequential pattern* adalah pada bisnis retail. Apabila terdapat suatu pola, yaitu 'seorang pelanggan yang membeli raket tenis bulan ini akan membeli bola tenis bulan depan', pola tersebut adalah *Sequential Pattern*. *Sequential Pattern* dapat digunakan untuk memprediksi kejadian yang dapat terjadi berikutnya ataupun mengidentifikasi pengulangan kejadian dari data yang dinamis.

Dengan struktur data multidimensional, pengguna dapat menampilkan data dengan dimensi yang lebih luas, contohnya data pembelian berdasar jumlah penjualan per-counter (lokasi toko) karakteristik customer dan juga, data dapat ditampilkan per harian, bulanan, tahunan atau tingkatan waktu yang lain yang dibutuhkan pengguna.

Pada kondisi tertentu, data multidimensional akan memunculkan juga data *Sekuensial Multidimensional*. Sekuensial ini muncul ketika dimensi yang digunakan dalam menampilkan data adalah beragam segi dan kriteria. Misalkan setiap atribut yang digunakan diasumsikan memiliki 1 dimensi, dan di dalam atribut tersebut memiliki dimensi lagi sesuai dengan apa yang kita inginkan. Dari contoh kejadian inilah, diperlukan suatu usaha tersendiri untuk dapat melihat keterurutan berdasarkan kriteria yang berbeda.

Oleh karena itu dibutuhkan sebuah program yang dapat digunakan pada kasus data *sequential multidimensional* sehingga dapat dilihat keterurutan berdasarkan kriteria yang berbeda. Pada kasus ini kriteria yang ingin dibahas adalah mengenai perbedaan dimensi waktu, sehingga didapatkan pengetahuan dari setiap dimensi waktu tersebut.

Terdapat beberapa algoritma untuk melakukan *sequential pattern mining*, antara lain *Apriori-All*, *Apriori-Some*, *FreeSpan*, *PrefixSpan* dan lain-lain. Namun yang digunakan dalam tugas akhir ini adalah *PrefixMDSpan* (*Prefix-projected Sequential pattern mining*), karena terbukti bahwa algoritma *PrefixMDSpan* cenderung lebih baik dari algoritma lainnya [10].

1.2 Perumusan masalah

Terdapat banyak literatur yang membahas mengenai pengembangan *sequential pattern*. Namun beberapa hanya membahas teori dan algoritma umumnya saja. Selain itu, karena data yang digunakan multidimensional, diperlukan strategi implementasi agar waktu proses *sequential pattern* menjadi lebih singkat dan *sequential pattern* yang dihasilkan menjadi sebuah informasi yang bermanfaat.

Pembahasan masalahnya sebagai berikut :

1. Mengimplementasikan teknik *sequential pattern mining* pada data *sekuensial multidimensional* berdasarkan algoritma *PrefixMDSpan* dan mengimplementasikannya menjadi sebuah program yang dapat digunakan.
2. Mendapatkan informasi yang bermanfaat dan dapat digunakan dari *sequential pattern* yang dihasilkan oleh program.

Sebagai bagian dari teknik ekstraksi pengetahuan dari data, proses yang dilakukan dalam tugas akhir ini meliputi proses data mining dengan pendekatan *sequential pattern mining* pada kumpulan data. Batasan masalahnya antara lain :

1. Aplikasi ini menggunakan sumber data yang telah siap digunakan (data fix).
2. Implementasi fokus pada data sekuensial multidimensional.
3. Tidak membahas mengenai teknik-teknik optimasi program.

1.3 Tujuan

Tujuan yang ingin dicapai dalam tugas akhir ini adalah memperoleh pemahaman terhadap data *sekuensial multidimensional* dan teknik yang dapat digunakan untuk melakukan *sequential pattern mining* pada data *sekuensial multidimensional*.

Tujuan umum yang ingin dicapai adalah :

1. Melakukan pengujian dengan parameter waktu eksekusi algoritma yang dipengaruhi oleh beberapa karakteristik seperti jumlah *record*, jumlah *dimensi*, panjang *sequence* rata-rata yang terlibat dan *minimum support*.
2. Mendapatkan pengetahuan berupa informasi yang dari *sequential pattern* yang dihasilkan sehingga memberi suatu pengetahuan bermanfaat dan dapat diterapkan pada periode waktu selanjutnya ataupun yang akan datang.

1.4 Metodologi penyelesaian masalah

Untuk menyelesaikan pembuatan tugas akhir yang merupakan tujuan dari dibuatnya proposal ini, akan dilakukan beberapa langkah kerja sebagai berikut:

1. Eksplorasi dan Stuli Literatur
Tahap ini dilakukan dengan cara mempelajari cara kerja *PrefixSpan* dan *PrefixMDSpan* dari literatur-literatur seperti buku (textbook), paper dan sumber lain seperti website, artikel dokumen teks yang berhubungan.
2. Analisis
Dari analisis literatur, dilakukan analisis tentang pembentukan data sekuensial multidimensional. Analisis juga meliputi pemahaman mendetail tentang sekuensial *pattern mining* pada bentuk data sekuensial multidimensional yang diajukan, dan aspek-aspek yang meliputinya.
3. Perancangan Sistem
Dilakukan perancangan yang mampu mengakomodasi penerapan teknik *sequential pattern mining* pada data sekuensial multidimensional dalam sebuah aplikasi perangkat lunak.
4. Implementasi

Implementasi dilakukan untuk mewujudkan perangkat lunak yang telah dirancang untuk penerapan teknik sequential pattern mining pada data sekuensial multidimensional.

5. Pengujian

Pengujian dilakukan untuk memperoleh kesimpulan dari analisis dan sebagai bahan evaluasi terhadap penerapan teknik sequential pattern mining yang dilakukan.

1.5 Sistematika Pembahasan

Sistematika laporan tugas akhir ini adalah sebagai berikut :

- BAB 1** Bab pertama membahas latar belakang, rumusan masalah, tujuan, batasan, dan metodologi pelaksanaan tugas akhir.
- BAB 2** Bab kedua membahas dasar teori untuk sequence patter mining termasuk definisi multidimensional sequence pattern.
- BAB 3** Bab ketiga adalah analisis tentang PrefixSpan, analisis tentang multidimensional sequence pattern mining dan termasuk perancangan sistem yang akan di implementasikan.
- BAB 4** Bab keempat fokus pada implementasi dan pengujian, termasuk strategi pengembangan untuk menerapkan algoritma PrefixMDSpan dalam sebuah perangkat lunak dan dilakukan evaluasi terhadap hasil pengujian dan implementasi.
- BAB 5** Bab kelima akan menutup laporan tugas akhir ini dengan kesimpulan dan saran.