

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dengan berlimpahnya data yang merupakan akumulasi data yang tersimpan hingga bertahun – tahun pada suatu instansi atau media terutama di internet. Mengakibatkan penumpukan data yang tidak terpakai dalam jumlah besar.

Sebenarnya, data dalam jumlah besar tersebut dapat dimanfaatkan untuk mendapatkan informasi yang bernilai lebih. Maka, untuk mendapatkan manfaat tersebut dilakukan proses penggalian data (data mining). Dan model dari proses data mining sendiri ada banyak. Salah satunya adalah, model yang akan digunakan dalam penelitian ini, yaitu *sequential pattern mining*.

*Sequential Pattern Mining* adalah model dari data mining, dimana model tersebut akan mendapatkan informasi dari suatu peristiwa yang secara berurutan terjadi dalam beberapa waktu dan dengan jumlah data yang sangat banyak yang akan membentuk pola.

Agar dalam melakukan komputasi *sequential pattern mining* dapat dilakukan secara maksimal dan efisien, maka digunakan algoritma VMSP (*Vertical Mining of Maximal Sequential Pattern*) yang terbukti maksimal dan efisien dalam penggunaan ruang memori dan jangka waktu komputasi (Fornier et. al, 2014).

Namun pada penelitian tersebut, VMSP belum diketahui tingkat ketepatan dan presisi dalam menghasilkan informasi keluarannya dalam kinerjanya. Maka pada penelitian ini dilakukan analisis pada algoritma VMSP menggunakan analisis *Accuracy* dan *precision*. Dengan data yang digunakan dalam penelitian ini adalah data yang sama dengan penelitian sebelumnya. Yaitu, *web click stream* dari media FIFA, BMS, Kosarak, MSBC, dan Lethiavan (Fornier et. al, 2014). Diharapkan setelah dilakukan analisis, algoritma VMSP memiliki tingkat ketepatan dan presisi maksimal dalam kinerjanya.

## 1.2 Perumusan Masalah

Dari uraian latar belakang di atas, maka rumusan masalah dari penelitian ini sebagai berikut:

1. Bagaimana hasil analisis dari tingkat ketepatan (*accuracy*) dan ketelitian (*precision*) kinerja algoritma VMSP.
2. Apa saja kelebihan dari algoritma VMSP pada model *sequential pattern mining*.

## 1.3 Tujuan

Tujuan penelitian ini adalah, sebagai berikut:

1. Menganalisis tingkat ketepatan (*accuracy*) dan ketelitian (*precision*) kinerja algoritma VMSP.
2. Mengetahui kelebihan dari algoritma VMSP pada model *sequential pattern mining*.

## 1.4 Manfaat

Manfaat dari penelitian ini sebagai berikut :

1. Dapat mengetahui *sequential pattern* yang terbentuk dari dataset skala besar sehingga diperoleh informasi atau relasi-relasi penting.
2. Dapat diketahui hasil analisis ketepatan dan presisi kinerja dari algoritma.
3. kelebihan dari algoritma VMSP pada model *sequential pattern mining*.
4. Dapat dijadikan landasan dalam penggalian data yang berurutan (*sequential pattern mining*) dengan algoritma yang efisien.

## 1.5 Batasan Masalah

Berdasarkan masalah yang telah dikemukakan di atas, maka dibuat batasan masalah sebagai berikut :

1. *Dataset* yang akan digunakan dalam penelitian analisis dari algoritma VMSP ini terdiri dari 5 *dataset*, diantaranya adalah: *web click stream* Kosarak (portal berita dari hongaria) sebanyak 10. 000 *sequence*,

Leviathan sebanyak 5.834 sequence, FIFA sebanyak 95.370 sequence, BMS sebanyak 1.020.153 sequence, dan MSNBC (portal berita dari Amerika Serikat) sebanyak 31.790 sequence. Data – data ini berupa pola data yang besar dengan banyak item kejadian yang terjadi pada banyak waktu. Dan merupakan data yang telah dipakai pada penelitian sebelumnya (Fornier at. al, 2014).

2. Penggalan data untuk mendapatkan informasi secara berurutan (*sequential pattern*) menggunakan algoritma VMSP.
3. Untuk informasi keluaran berupa pola yang maksimal dari data yang ada pada *dataset* dan nilai *support* dari pola tersebut.
4. Minimal support yang digunakan 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%
5. Untuk mengetahui tingkat ketepatan dan presisi kinerja algoritma VMSP, maka digunakan analisis *Accuracy* dan *Precision*.