

Analisa Keterkaitan (*Link Analysis*) Dengan Menggunakan *Sequential Pattern Discovery* Untuk Prediksi Cuaca

Wiwin Suwarningsih
P2 Informatika-LIPI
wiwin@informatika.lipi.go.id

Nuryani
P2 Informatika-LIPI
nuryani@informatika.lipi.go.id

Andria Arisal
P2 Informatika-LIPI
andria.arisal@informatika.lipi.go.id

Abstract

The purpose of this research was to analyze the relationship between attribute/item-set connection in the parameters used in weather prediction to generate a rule for proving a condition whether rain or no rain. The method will be used for the analysis inter-attribute is sequential pattern mining, where his work is to analyze an attribute linking patterns will affect other attributes and how dependency of an attribute with other attributes.. The result of this research is a pattern of knowledge hidden in data. That pattern is a rule that can be used to determine whether today are rain or not.

Keywords: Sequential pattern mining, Weather prediction, Link analysis, Association Rule Mining (ARM).

Abstrak

Tujuan penelitian ini adalah menganalisa keterkaitan antar atribut/itemset pada parameter yang digunakan dalam prediksi cuaca untuk menghasilkan suatu aturan(rule) yang dapat membuktikan suatu kondisi apakah hujan atau tidak hujan. Metoda yang akan digunakan untuk analisa antar atribut ini adalah Sequential Pattern Mining, dimana pola kerjanya adalah menganalisa keterkaitan suatu atribut akan mempengaruhi atribut lainnya dan bagaimana ketergantungan atribut yang satu dengan atribut lainnya. Hasil akhir dari penelitian ini adalah menemukan pola-pola pengetahuan yang tersembunyi di dalam data. Pola tersebut berbentuk aturan (rule) yang dapat membantu menentukan kondisi cuaca apakah hari ini hujan atau tidak.

Kata Kunci : Sequential pattern mining, Prediksi cuaca, Link analisa, Association Rule Mining (ARM).

1. Pendahuluan

Pemanasan oleh sinar matahari menyebabkan dinamika di atmosfer bumi. Akibat faktor perubahan lingkungan serta gejala meteorologi dan geofisika lainnya dapat menimbulkan kondisi anomali cuaca yang terkadang ekstrim[1]. Seperti kondisi cuaca di bulan Juni yang biasanya sudah menginjak ke musim panas/kemarau, namun ternyata hujan masih tetap ada dan terkadang tidak tentu datangnya. Curah hujan yang cukup lebat disertai dengan angin kencang

muncul dengan intensitas yang tak terbatas. Beberapa faktor penyebab terjadinya hujan adalah kecepatan angin, pembentukan awan-awan hujan disuatu kawasan, suhu dan kelembaban udara. Dari hasil penelitian ini diharapkan memperoleh pola-pola pengetahuan yang tersembunyi dalam data untuk menghasilkan sebuah informasi yang dibutuhkan yaitu prediksi cuaca terkini untuk hari esok berdasarkan data yang ada apakah akan terjadi hujan atau tidak hujan.

Tahapan yang dilakukan pada penelitian ini adalah analisa keterkaitan antar parameter-parameter penyebab terjadinya

hujan dengan menggunakan metoda *Sequential Pattern Mining* untuk menemukan pola-pola pengetahuan yang tersembunyi di dalam data cuaca.

1.1 Link analysis

Tujuan *link analysis* adalah mencari hubungan antara atribut-atribut suatu basis data[4]. Dibandingkan dengan operasi-operasi data mining lainnya, link analysis lebih sering dipakai karena informasi yang dihasilkan lebih sederhana dan dapat dirapakan langsung untuk pengambilan keputusan dibanding dengan operasi data mining lain yang cenderung membutuhkan analisa lanjut[2].

Didalam link analysis [3] terdapat tiga kategori yaitu :

a. Association Discovery:

menganalisa barang-barang yang dibeli secara bersamaan pada suatu transaksi, sehingga ditemukan hubungan yang mungkin tersembunyi di antara barang-barang tersebut

b. Sequential Pattern Discovery:

mengidentifikasi hubungan antara transaksi pembelian barang dari waktu ke waktu untuk mendapatkan informasi mengenai urutan barang yang dibeli pelanggan. Tujuannya adalah mengenali kebiasaan pelanggan dalam jangka waktu tertentu. Contoh: "*Pelanggan yang membeli pentium PC sembilan bulan yang lalu cenderung untuk memesan CPU baru dalam satu bulan ini*". Pola sekuensial mining berguna dalam analisa data untuk marketing atau ramalan cuaca.

c. Similar Time Sequence Discovery

menemukan hubungan antara dua kelompok data yang bergantung pada waktu, berdasarkan tingkat kemiripan pola yang ditunjukkan data tersebut. Jadi tujuannya mencari kemunculan semua sekeun yang mirip dengan sekuen yang diberikan, dalam suatu time-series basis data. Tidak seperti query basis data pada umumnya yang mencari data sama persis seperti yang diberikan dalam query. *Similarity search* adalah mencari data

sekuen yang mempunyai perbedaan kecil dari query yang diberikan.

1.2 Sequential pattern mining

Sebuah sekuen[4] adalah urutan dari elemen-elemen (transaksi), yang dapat dinotasikan sebagai berikut :

$$s = \langle e_1, e_2, e_3 \dots \rangle$$

Setiap elemen (e) terdiri dari kumpulan kejadian-kejadian (item=i).

$$e_i = \{i_1, i_2, \dots, i_k\}$$

Setiap elemen merupakan atribut yang dihubungkan dengan suatu lokasi atau waktu tertentu (spesifik). Panjang Sekuen, |s|, adalah banyaknya unsur-unsur sekuen yang diberikan. k-sekuens adalah sebuah sekuen yang terdiri dari k kejadian (item)

Sequential Pattern Mining adalah pencarian pola yang terjadi cukup sering sehubungan dengan waktu atau sekuen lainnya[3]. Pencarian pola sekuensial terdiri dari beberapa tahapan proses yaitu :

- Penentuan struktur basis data
Langkah ini dilakukan untuk menentukan ekstraksi data yang dibutuhkan.
- Menentukan nilai ambang batas (threshold)
Tahapan ini membentuk pola sekuensial dan menentukan urutan item besar yang terdiri dari dua buah item data.
- Pencarian satu set pola terpanjang
Yaitu mencari satu set pola terpanjang yang memenuhi nilai ambang batas yang telah ditentukan
- Pembuatan graf Asosiasi
Tahap ini dilakukan untuk mendapatkan satu set pola yang memenuhi nilai ambang batas.

1.3 Ekstraksi pola sekuensial

Cara mengekstraksi pola sekuensial [4] adalah :

- Terdapat n kejadian: $i_1, i_2, i_3, \dots, i_n$
- Tentukan kandidat 1-subsekuen:
 $\langle \{i_1\} \rangle, \langle \{i_2\} \rangle, \langle \{i_3\} \rangle, \dots, \langle \{i_n\} \rangle$
- Tentukan kandidat 2-subsekuen:
 $\langle \{i_1, i_2\} \rangle, \langle \{i_1, i_3\} \rangle, \dots, \langle \{i_1\} \{i_1\} \rangle, \langle \{i_1\} \{i_2\} \rangle, \dots, \langle \{i_{n-1}\} \{i_n\} \rangle$

- Tentukan kandidat 3-subsekuen:
 $\langle \{i_1, i_2, i_3\} \rangle, \langle \{i_1, i_2, i_4\} \rangle, \dots, \langle \{i_1, i_2\} \{i_1\} \rangle, \langle \{i_1, i_2\} \{i_2\} \rangle, \dots,$
 $\langle \{i_1\} \{i_1, i_2\} \rangle, \langle \{i_1\} \{i_1, i_3\} \rangle, \dots,$
 $\langle \{i_1\} \{i_1\} \{i_1\} \rangle, \langle \{i_1\} \{i_1\} \{i_2\} \rangle, \dots$

Setelah dihasilkan kandidat sekuensialnya maka dapat ditentukan sekuen besar dengan melakukan generalisasi kandidat yaitu dengan cara sebagai berikut :

- Kasus Dasar ($k=2$):
 Penggabungan dua frekuensi 1-sekuen $\langle \{i_1\} \rangle$ dan $\langle \{i_2\} \rangle$ akan menghasilkan dua kandidat 2-sekuen : $\langle \{i_1\} \{i_2\} \rangle$ dan $\langle \{i_1 i_2\} \rangle$
- Kasus Umum ($k>2$):
 Sebuah frekuensi ($k-1$) sekuen w_1 digabungkan dengan frekuensi yang lain ($k-1$) sekuen w_2 untuk menghasilkan sebuah kandidat k -sekuen jika subsekuen yang diperoleh dengan memindahkan kejadian pertama di w_1 adalah sama dengan subsekuen yang diperoleh dengan memindahkan kejadian terakhir di w_2 .

Hal ini menghasilkan kandidat setelah penggabungan diberi oleh sekuen w_1 diperluas sampai dengan kejadian yang terakhir dari w_2 . Cara lainnya, kejadian akhir di dalam w_2 menjadi suatu element terpisah dengan menambahkan pada ujung w_1 dengan elemen yang sama, kemudian kejadian yang terakhir di dalam w_2 menjadi bagian dari element yang terakhir di dalam w_1 .

2. Metoda penelitian

Penelitian ini menggunakan metoda *sequential pattern mining*, yang terdiri dari penentuan stuktur data, pembentukan pola sekuensial, pencarian satu set pola terpanjang dan pembuatan graf asosiasi. Berdasarkan graf asosiasi maka pada tahap akhir dari penelitian ini akan dibuat *rule* (aturan) untuk menentukan kondisi cuaca hujan atau tidak hujan.

3. Hasil dan pembahasan

Struktur data yang terbentuk adalah tabel itemset berdasarkan data sampling yang diambil dari hasil pemantauan cuaca BMKG Bandung dari tahun 2005 sampai dengan 2009. (lihat tabel 1)

Tabel 1 Struktur data hasil pemantauan (Transaction database)

Tgl	Item set				Kondisi cuaca
	suhu	Kelembaban	Kec. Angin	Arah Angin	
1/1/05	23.7	80	6	W	H
1/1/06	23.6	78	6	W	H
1/1/07	23.1	83	6	W	H
1/1/08	22.8	88	4	W	H
1/1/09	23.6	79	3	NW	TH
1/2/05	24.7	74	7	W	TH
1/2/06	23.6	83	0	C	H
1/2/07	23.7	75	6	W	H
1/2/08	21.7	92	4	W	H
1/2/09	25	66	4	NW	TH
3/1/05	24,0	79	4	W	TH
3/1/06	22,7	89	9	W	H
3/1/07	23,5	74	5	W	TH
3/1/08	21,7	89	4	W	H
3/1/09	24,7	67	3	W	TH
4/1/05	24,2	82	6	W	TH
4/1/06	22,3	78	6	W	H
4/1/07	23,6	68	4	W	TH
4/1/08	21,5	87	5	W	H
4/1/09	23,6	72	2	NW	TH

Berdasarkan tabel 1, data di analisa dengan melihat pola dari masing-masing parameter pembentuk kondisi cuaca hujan (H) atau tidak hujan (TH). Data tersebut kemudian diekstrak guna memudahkan penentuan nilai ambang batas. Nilai ambang batas yang terbentuk adalah urutan item besar (lihat tabel.1) dimana pengelompokan item berdasarkan urutan yang dibuat dengan melihat kondisi hari hujan atau tidak hujan. Selanjutnya adalah pencarian pola sekuensial yang diawali dengan upaya untuk membentuk urutan besar (*large sequence*) yang terdiri dari dua sampai empat parameter yang memenuhi nilai ambang batas yang

ditentukan. Pada penelitian ini nilai ambang batasnya adalah 60% dimana setiap parameter harus di dukung minimal 4 data hasil pemantauan. Pencarian pola dilanjutkan dengan berupaya untuk menemukan sebanyak mungkin *sekuen item k-2*, untuk menggali semua urutan item data dengan panjang atau lebih besar sama

dengan 4. Pola sekuen yang dihasilkan (lihat tabel 2) merupakan pola yang diamati dari data berdasarkan tanggal dan bulan yang sama selama 5 tahun dengan mengelompokkan berdasarkan parameter suhu, kelembaban, kecepatan angin dan arah angin.

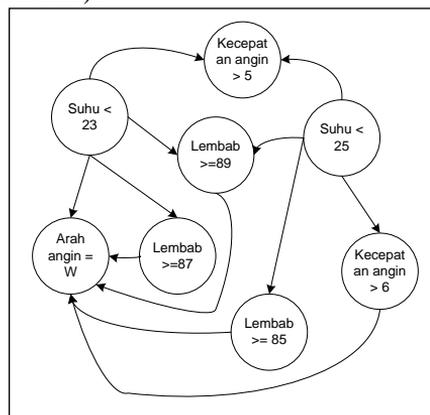
Tabel 2 Pola sekuensial

Kondisi cuaca	Urutan	Jumlah Dukungan	Tanggal pendukung
Hujan	{{(suhu <22), (lembab>=89), (kec.angin >=4), (arah angin = w)}}	9	13/1/09, 15/1/09, 4/1/08, 2/1/08, 3/1/08,12/1/09, 14/1/09, 29/1/05, 29/1/08
Hujan	{{(suhu <23), (lembab>=87), (kec.angin >=6), (arah angin = w)}}	14	25/1/06, 24/1/06, 23/1/06, 27/1/06, 30/1/09, 12/1/06, 13/1/06, 19/1/05, 7/1/05, 8/1/06, 3/1/06, 1/1/08, 9/1/06, 24/1/05
Hujan	{{(suhu <25), (lembab>=85), (kec.angin >=6), (arah angin = w)}}	26	28/1/06, 29/1/06, 28/1/07, 28/1/08, 31/1/07, 31/1/09, 9/1/05, 27/1/07, 14/1/06, 24/1/07, 25/1/09, 28/1/05, 12/1/06, 13/1/06, 19/1/05, 20/1/05, 7/1/05, 8/1/06, 26/1/06, 1/1/08, 9/1/06, 24/1/05, 17/1/05, 27/1/05
Hujan	{{(suhu >=25), (lembab>=85), (kec.angin >=5), (arah angin = w)}}	12	28/1/06, 29/1/07, 31/1/07, 8/1/05, 16/1/05, 9/1/05, 14/1/06, 28/1/05, 5/1/05, 31/1/05, 10/1/06, 30/1/06
Tidak Hujan	{{(suhu >=23.5), (lembab>88), (kec.angin <5), (arah angin = w)}}	11	17/1/06, 22/1/09, 5/1/09, 10/1/08, 19/1/09, 6/1/08 6/1/09, 9/1/08, 27/1/09, 29/1/09, 3/1/07
Tidak Hujan	{{(suhu >=26), (lembab>80), (kec.angin <4), (arah angin = w)}}	27	1/1/09, 4/1/07, 4/1/09, 7/1/09, 8/1/07, 20/1/08, 21/1/08, 8/1/09, 13/1/08, 7/1/08, 12/1/07, 19/1/06, 15/1/07, 3/1/05, 5/1/07, 21/1/06, 23/1/09, 20/1/06 3/1/09, 10/1/07, 18/1/08, 21/1/07, 9/1/07, 11/1/07 2/1/09, 13/1/07, 17/1/08, 6/1/07, 16/1/07

Pola sekuensial pada tabel 2 merupakan tabel set terpanjang setelah mengalami proses beberapa kali untuk pencarian urutan item besar. Sehingga pola sekuen pada tabel 2 ini dapat digunakan untuk menentukan pola asosiasi antar parameter. Yang digunakan untuk menentukan pola asosiasi ini dilihat dari jumlah dukungan yang paling besar karena data nyata yang mendukung kondisi pada pola asosiasi ini memenuhi ambang batas. Berdasarkan tabel 2 maka kecenderungan yang terjadi adalah suhu rata-rata kurang dari atau sama dengan 25°C dan kelembaban lebih besar atau sama dengan 85%, kecepatan angin kurang dari 5 dan arah angin kebanyakan ke arah W, maka akan terjadi 'hujan'. Kemudian suhu rata-rata lebih besar dari 25°C dan kelembaban diatas 80%, kecepatan angin kurang dari 4 dan arah angin kebanyakan ke arah W, maka akan terjadi 'tidak hujan'.

Tahap selanjutnya adalah pembentukan diagram atau graf asosiasi antar parameter -

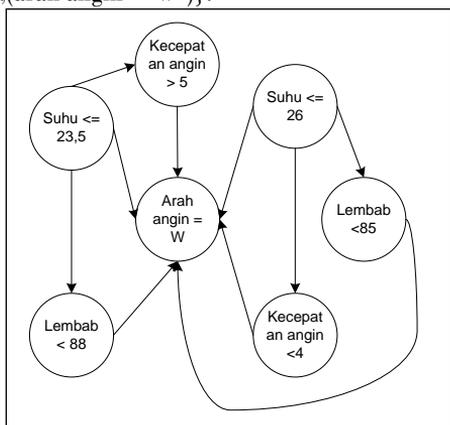
parameter pendukung prediksi cuaca, dimana graf yang dibangun untuk memperlihatkan parameter yang digunakan memenuhi ambang batas (lihat pada gambar 1 dan gambar 2).



Gambar 1 Graf asosiasi prediksi cuaca Hujan

Berdasarkan graf asosiasi prediksi cuaca Hujan (gambar 1) dapat ditelusuri penurunan busur berdasarkan irisan himpunan urutannya yaitu $P\{(suhu <22), (lembab>=89),$

$(\text{kec.angin} \geq 4), (\text{arah angin} = w) \} \cap P\{(\text{suhu} < 23), (\text{lembab} \geq 87), (\text{kec.angin} \geq 6), (\text{arah angin} = w) \} \cap P\{(\text{suhu} < 24), (\text{lembab} \geq 85), (\text{kec.angin} \geq 6), (\text{arah angin} = w) \} \cap P\{(\text{suhu} \leq 25), (\text{lembab} \geq 85), (\text{kec.angin} \geq 5), (\text{arah angin} = w) \}$, maka urutan item besar-2 yang dapat dikembangkan adalah $\{(\text{suhu} \leq 25), (\text{kelembaban} \geq 85), (\text{kecepatan angin} < 5), (\text{arah angin} = 'w') \}$.



Gambar 2 Graf asosiasi prediksi cuaca Tidak Hujan

```

Rule 1 :
if suhu <= 25 and
  kelembaban >= 85 and
  kecepatan angin < 5 and
  arah angin = 'w'
  then
    kondisi cuaca = 'hujan'

Rule 2 :
if suhu > 25 and
  kelembaban < 85 and
  kecepatan angin >= 5 and
  arah angin = 'w'
  then
    kondisi cuaca =
      'tidak hujan'
  
```

Gambar 3 Aturan keterkaitan item-set

Berdasarkan graf asosiasi prediksi cuaca Tidak Hujan dapat ditelusuri penurunan busur berdasarkan irisan himpunan urutannya yaitu $P\{(\text{suhu} \leq 23.5), (\text{lembab} < 88), (\text{kec.angin} < 5), (\text{arah angin} = w) \} \cap P\{(\text{suhu} \leq 26), (\text{lembab} < 80), (\text{kec.angin} < 4), (\text{arah angin} = w) \}$, maka urutan item besar-2 yang dapat dikembangkan adalah

$\{(\text{suhu} > 25), (\text{kelembaban} < 85), (\text{kecepatan angin} \geq 5), (\text{arah angin} = 'w') \}$.

Dari hasil analisa asosiasi itemset, pola sekuensial dan graf asosiasi, maka *rule* (aturan) yang diperoleh dapat dilihat pada gambar 3.

Berdasarkan aturan (*Rule*) yang terbentuk pada gambar 3, maka tahap pembangunan aplikasi untuk prediksi cuaca dapat menggunakan *rule* tersebut sebagai aturan yang dapat menyeleksi data masukan (berupa suhu, kelembaban, kecepatan angin, dan arah angin) sehingga hasil keluarannya adalah suatu prediksi cuaca apakah hujan atau tidak hujan.

4. Kesimpulan

Dari hasil analisa data dan melihat pola sekuensial, maka dapat disimpulkan bahwa metoda *sequence pattern mining* memudahkan untuk mencari pola-pola pengetahuan dari data pemantauan cuaca BMKG selama 5 tahun dan dapat membantu pula untuk menentukan aturan (*rule*) dalam memprediksi cuaca.

5. Daftar pustaka

- [1] Haris Syahbuddin dan Tri Nandar Wihendar, *Anomali Curah Hujan Periode 2010 – 2040 di Indonesia*, Balai Penelitian Agroklimat dan Hidrologi, Bogor, 2008.
- [2] Lise Getoor, *Link Mining: A New Data Mining Challenge*, University of Maryland. Diakses dari: <http://www.sigkdd.org/explorations/issue5-1/Getoor.pdf> (14 Juli 2010).
- [3] Ramakrishnan Srikant and Rakesh Agrawal, *Mining Sequential Patterns: Generalizations and Performance Improvements*. Diakses dari: <http://rakesh.agrawal-family.com/papers/edbt96seq.pdf> (15 juni 2010).
- [4] Tan P. N., Steinbach, M., Kumar, V. *Introduction to Data Mining*, Addison Wesley, 2006.