Jurnal Sistem Informasi (JSI), VOL. 7, NO. 2, Oktober 2015,
ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index
email: jsi.fasilkom.unsri@gmail.com

Partisi Data Secara Vertikal Untuk Menentukan Aturan Asosiasi Item Set Data Cuaca

Wiwin Suwarningsih[1], Andria Arisal[2]
Pusat Penelitian Informatika, Komplek LIPI
Email: wiwin@informatika.lipi.go.id[1]; andria.arisal@informatika.lipi.go.id[2]

ABSTRACT

This paper discussed about association rule mining among item sets of weather records, where observation results are distributed from data source and partitioned in order to create an optimal rule pattern. We use decision tree classifiers as method for data partitioning, which each item set has several attributes and these item sets are used to identify the valid global association rule, but did not disclose the items set individual transaction data. The final results of this study was to partition the data to generate a frequency associated items set weather data with the minimal level of support without revealing the value of the item set of individuals. Frequency value associated items set the partition of this data can be used for weather prediction simulations whether there will be rain or no rain.

Keywords: association rule mining, item set, weather records, partition, decision tree classifiers

ABSTRAK

Makalah ini membahas aturan penambangan asosiasi (association rule mining) antar item set data cuaca dimana data hasil pemantauan didistribusikan dari sumber data dan dipartisi untuk memperoleh pola aturan yang optimal. Metoda yang akan digunakan untuk partisi data adalah pengklasifikasian pohon keputusan (decision tree classifiers) yaitu setiap item set memegang beberapa atribut dan item set tersebut mengidentifikasi aturan asosiasi global yang valid, namun item set tidak mengungkapkan data transaksi individu. Hasil akhir dari penelitian ini adalah partisi data untuk menghasilkan frekuensi asosiasi item set data cuaca dengan tingkat dukungan minimal tanpa mengungkapkan nilai item set individu. Nilai frekuensi asosiasi item set hasil partisi data ini dapat digunakan untuk simulasi prediksi cuaca apakah akan terjadi hujan atau tidak hujan.

Kata-kuci : aturan penambangan asosiasi, item set, data cuaca, partisi, pengklasifikasian pohon keputusan.

1. PENDAHULUAN

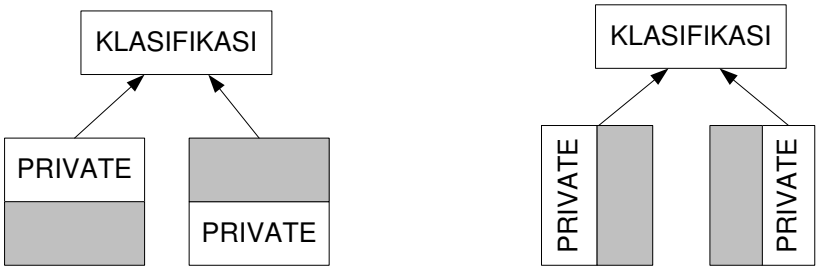
Penggunaan aplikasi untuk pendistribusian informasi atau pengetahuan yang berasal dari data harus disiasati agar privasi data tetap terjaga, karena pengumpulan data dan pengembangan pengetahuan dibutuhkan biaya yang tidak sedikit[1]. *Data mining* (penambangan data) digunakan sebagai sarana untuk menemukan pola-pola dan model kecenderungan dari data-data yang sangat banyak. Metode dasar yang digunakan pada penambangan data adalah klastering (*clustering*), klasifikasi (*classification*), penambangan kaidah asosiasi (*association rule mining*) dan deteksi urutan (*sequence detection*). Secara umum, semua metode tersebut dikembangkan sebagai model terpusat, dimana operasi dilakukan terhadap data-data yang sudah dikumpulkan pada suatu sistem (situs pusat)[2]. Dalam tulisan ini akan dibahas aturan asosiasi antar item set dari data cuaca. Data cuaca tersebut dipartisi dengan menggunakan metoda klasifikasi *'privacy preserving*', yaitu model klasifikasi yang menggabungkan antar bagian-bagian pembentuk model klasifikasi tanpa menghilangkan sifat privasi dari data-data tersebut. Klasifikasi antara dua kolaborasi yang akan digunakan adalah partisi data secara vertikal (*vertical partitioned data*), karena sangat sesuai untuk data-data cuaca[2].

Jurnal Sistem Informasi (JSI), VOL. 7, NO. 2, Oktober 2015,
ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index
email: jsi.fasilkom.unsri@gmail.com

2. Penambangan Data Yang Mempertahankan Privasi *(Privacy Preserving Data Mining)*

Privasi bertujuan untuk melindungi data individu. Banyak upaya hukum diarahkan untuk tujuan tersebut, sehingga banyak informasi yang tidak dapat dilacak secara spesifik dalam ruang lingkup privasi hukum. Hal ini mengarahkan solusi privasi dalam data mining pada penggunaan data tetapi bukan untuk mengidentifikasi data secara individu.

Solusi pengamanan data yang dapat digunakan adalah dengan memecah (mempartisi) data secara horizontal atau vertikal (lihat gambar 1).



a) Partisi Horizontal **b) Partisi Vertikal**

Gambar. 1 Partisi Data [3]

2.1. Partisi Data Secara Horizontal

Proses pemisahan data secara horisontal yaitu setiap kelompok (*party*) akan menjadi milik beberapa bagian (sejumlah *record*) dari database[4]. Pada proses ini digunakan algoritma data mining dengan cara menggabungkan kebutuhan data dari kelompok yang satu ke semua kelompok yang lain atau semua kelompok mengirimkan datanya ke suatu tempat. Proses penerimaan data dapat diatur sehingga menghasilkan gabungan data yang sifatnya lebih besar dan menjadi database global.

2.2. Partisi Data Secara Vertikal

Didefinisikan D adalah kumpulan tuple, D ={x_1, x_2, ..., x_n, C} dan I={1, 2, 3, ..., n} adalah himpunan atribut untuk elemen di D. Dan C diindikasikan sebagai label atribut kelas. Dalam partisi secara vertikal dataset P_i akan berisi subset dari D termasuk C, dimana C={c_1,c_2,c_3, ...c_k} dan k adalah nilai dari atribut label kelas.

Banyak pengukuran yang dapat digunakan untuk menghasilkan cara pembagian data yang terbaik. Tahap awalnya adalah mendefinisikan item ke dalam distribusi kelas dari record data. Untuk menghitung nilai efisiensi model yang bisa digunakan adalah model Gini (lihat rumus 1) [3].

$$Gini(t) = 1 - \sum_{i=0}^{k-1} \left[p\left(\frac{i}{t}\right) \right]^2 \qquad \ldots\ldots\ldots \quad (1)$$

Dimana $p\left(\frac{i}{t}\right)$ dinotasikan sebagai record yang memiliki kelas *i* yang diberikan ke tuple *t*. Kami memilih atribut yang memiliki indeks paling rendah sebagai bentuk pembagian atribut terbaik.

2.3. Pembangunan Pohon Keputusan

Tahapan pembangunan pohon keputusan dari kelompok data yang sudah dipartisi adalah :

1. Setiap kelompok (*party*) dihitung dengan menggunakan indeks Gini untuk setiap atribut dan mengirimkan ke kelompok yang dianggap sebagai server.
2. Server menginisialisasi atribut yang akan dijadikan sebagai simpul akar (*root*) dengan indeks Gini yang minimum
3. Inisialisasi antrian Q yang berisi akar.
4. Cek kondisi Q yang kosong, jika tidak kosong maka dibuat antrian dengan simpul akar adalah R^Q.
5. Setiap atribut di S[i] dimana i = (1,2,..,m) dibagi menjadi kelompok tertentu, bila dimana setiap kelompok mengirimkan nilai i pada server jika nilai i adalah vektor.
6. Ditentukan pembagian yang terbaik dari setiap himpunan atribut.
7. Digunakan pembagian $R^$ kedalam $R_1^$ dan $R_2^$
8. Ditambahkan $R_1^$ dan $R_2^$ untuk Q jika semua nilai kelas memiliki kelas yang sama.
9. Diulangi langkah 4 sampai memperoleh pembagian yang terbaik.

Sedangkan tahapan proses perhitungan untuk menentukan partisi terbaik sehingga ditemukan simpul akar pada pohon keputusan[3] adalah:

1. Diperiksa atribut dalam R dan S[i] apakah berada dikelompok yang sama. Jika ya, maka kelompok harus membentuk **'gain'** untuk S[i]. Jika tidak, maka dihitung *Gini*(*A*) dan *Gini*(*A,S[i]*) untuk setiap atribut S[i].
2. *Gini*(*A*) dibagi dengan R menjadi m bagian, R_1, R_2, R_3, ..., R_m dimana m adalah jumlah kelompok (*party*).
3. Dipilih dua nilai indeks terkecil sebagai acuan untuk mempartisi item set dari kelompok (*party*).
4. Ditentukan item set yang akan dijadikan simpul akar di pohon keputusan dengan nilai indeks Gini yang terkecil.

3. METODA PENELITIAN

Penelitian ini menggunakan metoda klasifikasi pohon keputusan (*decision tree classifiers*), yang terdiri dari; inisialisasi atribut data yang menjadi akar, inisialisasi antrian, pembagian antribut berdasarkan klasifikasi data, dan penentuan rule. Sedangkan data yang akan digunakan untuk analisa masalah ini adalah data hasil pemantauan BKMG (Badan Klimatologi, Meteorologi dan Geofisika)[5], dimana data dibagi menjadi dua bagian yaitu *training set* (data dari bulan Januari sampai dengan Desember 2008), dan *test set* yang akan digunakan untuk menguji *rule* terhadap *training set* (data dari bulan Januari sampai Desember 2009). Perhitungan nilai pengujian (R) dapat dilihat pada rumus 2 dan 3, sedangkan metoda yang digunakan untuk menentukan nilai pengujian mendekati kondisi nyata adalah metoda anlisis regresi dan korelasi[7].

$$\%NP\ \text{`Hujan'} = \frac{\sum \text{kondisi cuaca hujan hasil partisi}}{\sum \text{data}} \times \% \quad \text{..... (2)}$$

Jurnal Sistem Informasi (JSI), VOL. 7, NO. 2, Oktober 2015,
ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index
email: jsi.fasilkom.unsri@gmail.com

$$\%NP\ \text{'Tidak Hujan'} = \frac{\sum \text{kondisi cuaca tidak hujan hasil partisi}}{\sum \text{data}} \times \% \quad \text{..... (3)}$$

4. HASIL DAN PEMBAHASAN

Pembentukan data set (lihat Tabel 1, 2, 3 dan 4) sebagai data utama yang akan digunakan untuk inisialisasi atribut data, ini berarti ada 4 kelompok (m = 4) yang akan dijadikan sebagai data klasifikasi dengan jumlah data 365. Himpunan data (dataset) dibentuk berdasarkan kelompok (*party*) dengan klasifikasi atribut ke dalam kategori klasifikasi penggolongan dalam rentang nilai dari setiap parameter[1]. Klasifikasi ini berguna untuk menghitung nilai efisiensi dalam proses pembagian data secara vertikal.

Berdasarkan Tabel 1 untuk parameter suhu (*party* 1) memiliki 5 nilai dengan kategori yaitu: {{ suhu ≤ 22}, {suhu > 22 && suhu ≤ 23}, {suhu > 23 && suhu ≤ 24},{suhu > 24 && suhu ≤ 25}, {suhu > 25}}.

Dari nilai kategori tersebut diproses sebagai berikut :

Gini (S, Suhu) = 9/365 * Gini(S, suhu ≤ 22) + 90/365 * Gini(S, suhu > 22 && suhu ≤ 23) + 175/365 * Gini(S, suhu > 23 && suhu ≤ 24) + 66/365 * Gini(S, suhu > 24 && suhu ≤ 25) + 1/365 * Gini(S, suhu > 25)

dimana :

Gini(S, suhu ≤ 22)	$= 1-(4/9)^2 - (5/9)^2$	= 0,49
Gini(S, suhu > 22 && suhu ≤ 23)	$= 1 - (67/90)^2 - (33/90)^2$	= 0,31
Gini(S, suhu > 23 && suhu ≤ 24)	$= 1 - (110/175)^2 - (65/175)^2$	= 0,47
Gini(S, suhu > 24 && suhu ≤ 25)	$= 1 - (26/66)^2 - (40/66)^2$	= 0,39
Gini(S, suhu > 25)	$= 1 - (1/6)^2 - (5/6)^2$	= 0,28

Berdasarkan hasil perhitungan indeks Gini untuk parameter suhu, nilai indeks terkecil yaitu Gini(S, suhu > 25) diabaikan.

Tabel. 1 Data set untuk Parameter Suhu (*party* 1)

Suhu	Kondisi	Suhu	Kondisi	Suhu	Kondisi
23.7	hujan	22.6	hujan	25.6	tidak hujan
24.7	tidak hujan	23.7	hujan	24.3	tidak hujan
24	tidak hujan	23.8	hujan	23.8	Hujan
24.2	tidak hujan	23.4	hujan	24.2	tidak hujan
23.5	hujan	23.5	hujan	24.1	Hujan
23.8	tidak hujan	22.7	hujan	24.5	Hujan
22.6	hujan	23.1	hujan	24.6	Hujan
23/01/12	hujan	24.1	hujan	24.1	tidak hujan
23.2	hujan	24.7	hujan	23.5	Hujan
23.3	hujan	24.1	tidak hujan	23.7	tidak hujan
23.9	tidak hujan	24.1	tidak hujan	23.7	Hujan
24	hujan	25.3	tidak hujan	23.5	tidak hujan
24.1	hujan	24.9	tidak hujan	23.6	Hujan

23.9	tidak hujan	23.6	hujan	23.8	Hujan
…..	…..	…..	…..	…..	…..
24,1	hujan	24.6	tidak hujan	23.1	Hujan
23.1	hujan	25	tidak hujan	23.7	Hujan

Proses menentukan indeks Gini untuk parameter Kelembaban (*party* 2) dikelompokkan menjadi sub-himpunan sebagai berikut : {{kelembaban ≤ 74}, {kelembabab > 74 && kelembaban ≤ 76}, {kelembaban > 76 && kelembaban ≤ 78}, {kelembaban > 78 && kelembaban ≤ 80}, {kelembaban > 80 && kelembaban ≤ 82}, {kelembaban > 82 && kelembaban ≤ 84}, {kelembaban > 84 && kelembaban ≤ 86}, {kelembaban > 86}}

Tabel . 2 Data set untuk Parameter Kelembaban (*party* 2)

Kelembaban	Kondisi	Kelembaban	Kondisi	Kelembaban	Kondisi
80	hujan	87	hujan	80	tidak hujan
74	tidak hujan	84	hujan	81	tidak hujan
79	tidak hujan	84	hujan	89	Hujan
82	tidak hujan	87	hujan	83	tidak hujan
86	hujan	86	hujan	83	Hujan
82	tidak hujan	87	hujan	86	Hujan
89	hujan	84	hujan	84	Hujan
87	hujan	84	hujan	80	tidak hujan
87	hujan	78	hujan	78	Hujan
76	hujan	74	tidak hujan	75	tidak hujan
74	tidak hujan	79	tidak hujan	79	Hujan
76	hujan	74	tidak hujan	81	tidak hujan
84	hujan	74	tidak hujan	86	Hujan
83	tidak hujan	87	hujan	85	Hujan
84	hujan	79	tidak hujan	89	Hujan
…	…	…	…	…	…
86	hujan	82	tidak hujan	83	Hujan
76	hujan	74	tidak hujan	85	Hujan

Maka perhitungan indeks Gininya adalah

Gini(S, kelembaban) = 35/365 * Gini(S, kelembaban ≤ 74) + 24/365 * Gini(S, kelembaban > 74 && kelembaban ≤ 76) + 21/365 * Gini(S, kelembaban > 76 && kelembaban ≤ 78) + 39/365 * Gini(S, kelembaban > 78 && kelembaban ≤ 80) + 49/365 * Gini(S, kelembaban > 80 && kelembaban ≤ 82) + 52/365 * Gini(S, kelembaban > 82 && kelembaban ≤ 84) + 70/365 * Gini(S, kelembaban > 84 && kelembaban ≤ 86) + 75/365 * Gini(S, kelembaban > 86)

dimana :

Gini(S, kelembaban ≤ 74) $= 1 - (4/35)^2 - (31/35)^2 = 0{,}2$

Gini(S, kelembaban > 74 && kelembaban ≤ 76) $= 1 - (9/24)^2 - (15/24)^2 = 0{,}47$

Jurnal Sistem Informasi (JSI), VOL. 7, NO. 2, Oktober 2015,
ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index
email: jsi.fasilkom.unsri@gmail.com

Gini(S, kelembaban > 76 && kelembaban ≤ 78) = 1– $(4/21)^2$ – $(17/21)^2$ = 0,31
Gini(S, kelembaban > 78 && kelembaban ≤ 80) = 1– $(12/39)^2$ – $(27/39)^2$ = 0,43
Gini(S, kelembaban > 80 && kelembaban ≤ 82) = 1– $(26/49)^2$ – $(23/49)^2$ = 0,5
Gini(S, kelembaban > 82 && kelembaban ≤ 84) = 1– $(36/52)^2$ – $(16/52)^2$ = -0,17
Gini(S, kelembaban > 84 && kelembaban ≤ 86) = 1– $(67/70)^2$ – $(3/70)^2$ = 0,08
Gini(S, kelembaban > 86) = 1– $(71/75)^2$ – $(4/75)^2$ = 0,1

Tabel .3 Data set untuk Parameter Kelembaban (*party* 2)

Kec Angin	Kondisi	Kec Angin	Kondisi	Kec Angin	Kondisi
6	hujan	4	hujan	4	tidak hujan
7	tidak hujan	3	hujan	4	tidak hujan
4	tidak hujan	5	hujan	5	Hujan
6	tidak hujan	5	hujan	4	tidak hujan
8	hujan	4	hujan	2	Hujan
7	tidak hujan	6	hujan	3	Hujan
0	hujan	6	hujan	3	Hujan
6	hujan	4	hujan	4	tidak hujan
6	hujan	4	hujan	3	Hujan
5	hujan	4	tidak hujan	3	tidak hujan
0	tidak hujan	4	tidak hujan	3	Hujan
7	hujan	4	tidak hujan	4	tidak hujan
6	hujan	4	tidak hujan	3	Hujan
7	tidak hujan	3	hujan	3	Hujan
5	hujan	6	tidak hujan	0	Hujan
…	…	…	…	…	…
6	hujan	4	tidak hujan	4	Hujan
8	hujan	4	hujan	2	Hujan

Selanjutnya untuk menentukan indeks Gini dari parameter kecepatan angin dan arah angin dilakukan hal berikut. Parameter kecepatan angin dibagi menjadi 3 nilai kategori (Tabel 3), yaitu : {{kec_angin ≤ 3}, {kec_angin > 3 && kec_angin ≤ 6}, {kec_angin > 6}}. Sehingga formula index Gini untuk parameter kecepatan angin (*party* 3) adalah Gini(S,kec_angin) = 38/365 * Gini(S, kec_angin ≤ 3) + 274/365 * Gini(S, kec_angin > 3 && kec_angin ≤ 6) + 53/365 * Gini (S, kec_angin > 6).
dimana :

Gini(S, kec_angin ≤ 3) = 1 – $(28/38)^2$ – $(10/38)^2$ =0,51
Gini(S, kec_angin > 3 && kec_angin ≤ 6) =1 – $(141/274)^2$ – $(133/274)^2$ =0,5
Gini (S, kecepatan angin > 6) = 1 – $(42/53)^2$ – $(11/53)^2$ = 0,24

Jurnal Sistem Informasi (JSI), VOL. 7, NO. 2, Oktober 2015,
ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index
email: jsi.fasilkom.unsri@gmail.com

Tabel. 4 Data set untuk Parameter Kelembaban (*party* 2)

Arah Angin	Kondisi	Arah Angin	Kondisi	Arah Angin	Kondisi
E	tidak hujan	V	hujan	W	Hujan
W	tidak hujan	V	hujan	W	tidak hujan
W	hujan	W	hujan	W	tidak hujan
WSW	tidak hujan	W	hujan	W	tidak hujan
E	hujan	W	hujan	W	Hujan
E	hujan	W	hujan	W	tidak hujan
E	hujan	W	hujan	C	Hujan
W	tidak hujan	W	hujan	W	Hujan
E	hujan	W	hujan	N	Hujan
NNE	tidak hujan	W	tidak hujan	W	Hujan
W	hujan	W	tidak hujan	C	tidak hujan
W	tidak hujan	W	tidak hujan	W	Hujan
W	hujan	W	tidak hujan	W	Hujan
V	hujan	W	hujan	W	tidak hujan
C	hujan	W	tidak hujan	W	Hujan
…	…	…	…	…	…
W	hujan	W	tidak hujan	W	Hujan
NE	hujan	W	tidak hujan	W	Hujan

Penentuan kelompok (*party* 4) untuk parameter Arah Angin dari Tabel.4 dibagi menjadi 3 kategori nilai (dimana simbol E=Timur, N=Utara, S=Selatan dan W=Barat). Sub-himpunan yang terbentuk dipartisi dengan pengelompokan sebagai berikut : {{{E,N},{W}}, {{W,N},{S}},{S,E}, {N}}}.

Perhitungan indeks Gini untuk parameter arah angin adalah :
Gini(S, arah_angin) = 99/365 * Gini(S, arah_angin{E,N},{W}) + 78/365 * Gini(S, arah_angin{W,N},{S}) + 62/365 * Gini(S, arah_angin{S,E},{N})
dimana :

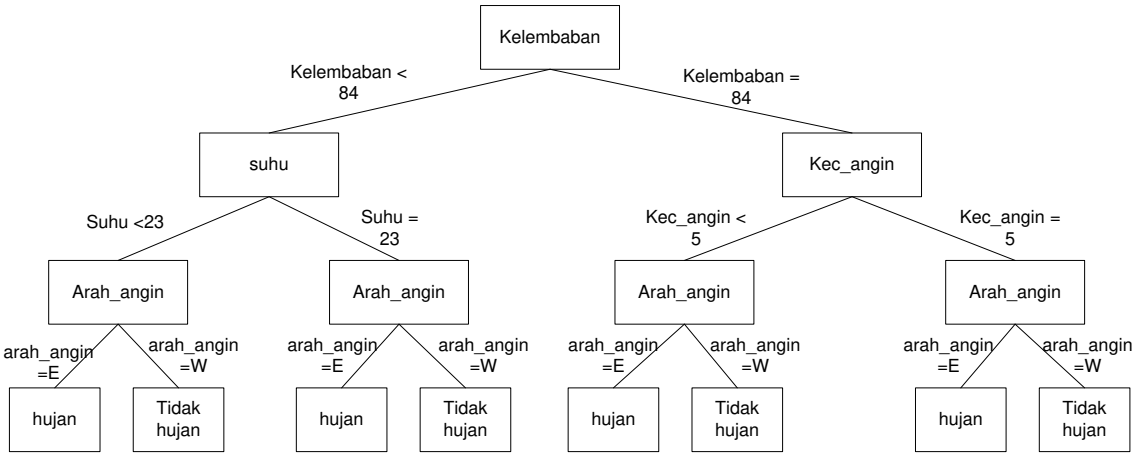
$$\text{Gini}(S, \text{arah\_angin}\{E,N\},\{W\}) = 1 - (45/99)^2 - (54/99)^2 = 0{,}5$$
$$\text{Gini}(S, \text{arah\_angin}\{W,N\},\{S\}) = 1 - (72/78)^2 - (6/78)^2 = 0{,}14$$
$$\text{Gini}(S, \text{arah\_angin}\{S,E\},\{N\}) = 1 - (48/62)^2 - (14/62)^2 = 0{,}35$$

Untuk menentukan kelompok besar (Gain) sebagai dasar pembentukan pohon keputusan adalah dengan menentukan gain terbesar yang diperoleh dari Gini terkecil, sehingga

$$\text{Gain}\ (S_{\text{kelembaban} < 84}, \text{suhu}) =$$
$$\text{Gini}(S_{\text{kelembaban} < 84}) - (S_{\text{suhu} < 23}, \text{arah\_angin}) \mid$$
$$\text{Gini}(S_{\text{kelembaban} < 84}) - (S_{\text{suhu} \geq 23}, \text{arah\_angin})$$
$$\text{Gain}(S_{\text{kelembaban} \geq 84}, \text{kec\_angin}) =$$
$$\text{Gini}(S_{\text{kelembaban} \geq 84}) - (S_{\text{kec\_angin} < 5}, \text{arah\_angin}) \mid$$
$$\text{Gini}(S_{\text{kelembaban} \geq 84}) - (S_{\text{kec\_angin} \geq 5}, \text{arah\_angin})$$

Jurnal Sistem Informasi (JSI), VOL. 7, NO. 2, Oktober 2015,
ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index
email: jsi.fasilkom.unsri@gmail.com

Dari hasil pembentukan kelompok besar dan perhitungan nilai indeks Gini terdapat nilai indeks kelembaban ada yang mencapai nilai -0,17. Ini berarti parameter kelembaban menjadi simpul akar pada pohon keputusan[3], dengan partisi parameter kelembaban dalam rentang {kelembaban < 84 dan kelembaban ≥ 84}. Kemudian simpul yang dijadikan anak dari simpul kelembaban berturut turut adalah suhu, kecepatan angin dan arah angin. Adapun bentuk pohon keputusan dapat dilihat pada Gambar 2.



Gambar. 2 Pohon Keputusan Data Cuaca

Berdasarkan Gambar 2 proses partisi dilakukan dari hasil perhitungan indeks Gini. Ada satu parameter menjadi simpul anak dari dua simpul yang berbeda, hal ini dibuat berdasarkan proses analisa pembentukan kelompok besar (Gain), dimana nilai dari parameter arah angin akan mempengaruhi dua parameter suhu dan kecepatan angin.

5. Pengujian Pohon Keputusan

Pengujian pohon keputusan ini dilakukan untuk melihat optimalisasi proses partisi terhadap data cuaca. Data yang digunakan untuk pengujian ini adalah data cuaca dari bulan Januari sampai dengan bulan Desember 2009 diambil secara acak yang digenerasi oleh komputer dengan menggunakan metoda *Random Number Variate* Generator[6]. Pengujian dilakukan pada partisi 1 yang merupakan dahan sebelah kiri pohon keputusan dan partisi 2 yang merupakan dahan sebelah kanan pohon keputusan. Pengujian partisi 1 (lihat Tabel 5) dilakukan untuk melihat pengaruh parameter suhu, kelembaban dan arah angin terhadap kondisi cuaca.

Tabel. 5 Pengujian Partisi 1 Pohon Keputusan

Tanggal	Kelembaban	Suhu	Arah Angin	Kondisi nyata	Kondisi Hasil Partisi
1/1/09	79	23,6	NW	TidakHujan	Tidak Hujan
16/1/09	84	22,4	W	Hujan	Tidak Hujan
23/1/09	79	24,3	W	Tidak hujan	Tidak Hujan
6/2/09	88	21,2	W	Hujan	Tidak Hujan
21/2/09	85	22,8	NW	Hujan	Hujan

2/3/09	74	23,7	NW	Tidak Hujan	Hujan
18/3/09	70	24,4	NW	Tidak Hujan	Tidak Hujan
30/3/09	86	22,1	W	Hujan	Tidak Hujan
9/4/09	77	23,8	SE	Tidak Hujan	Tidak Hujan
22/4/09	84	23.7	N	Hujan	Hujan
2/5/09	71	25.3	SW	Tidak Hujan	Tidak Hujan
17/5/09	76	24,3	W	Tidak Hujan	Tidak Hujan
31/5/09	80	24,8	C	Hujan	Hujan
8/6/09	81	22,5	W	Hujan	Tidak Hujan
25/6/09	82	23,3	W	Hujan	Tidak Hujan
3/7/09	81	22.8	W	Hujan	Tidak Hujan
20/7/09	78	24	S	Tidak Hujan	Tidak Hujan
1/8/09	67	22,5	NE	TidakHujan	Tidak Hujan
13/8/09	76	24,2	NE	TidakHujan	Tidak Hujan
30/8/09	71	23.8	NE	TidakHujan	Tidak Hujan
4/9/09	69	24,6	E	TidakHujan	Hujan
15/10/09	81	23,4	W	Hujan	Tidak Hujan
31/10/09	76	23,7	SE	TidakHujan	Hujan
9/11/09	71	25,2	W	TidakHujan	Tidak Hujan
25/11/09	84	22,5	W	Hujan	Tidak Hujan
5/12/09	84	23,6	W	Hujan	Tidak Hujan
19/12/09	72	24,8	W	TidakHujan	Tidak Hujan
31/12/09	92	21,3	W	Hujan	Tidak Hujan

Tabel 5. menunjukan nilai hasil partisi (R) dengan parameter kelembaban, suhu, arah angin mempengaruhi kondisi cuaca tidak hujan sebesar $R_{(tidak\ hujan)}$ = 22/28 = 0,786 sedangkan kondisi cuaca hujan sebesar $R_{(hujan)}$ = 6/28 = 0,214. Salah satu metoda yang digunakan untuk membandingkan proses prediksi cuaca dengan cara partisi vertikal dengan kondisi nyata adalah analisis regresi dan korelasi. Untuk permasalahan ini sebagai variabel terikat (NA) adalah kondisi nyata dengan $NA_{(tidak\ hujan)}$ = 15/28 = 0,536 dan kondisi nyata dengan $NA_{(\ hujan)}$ = 15/28 = 0,464. Jika dicari nilai R kuadrat maka $R^2_{(tidak\ hujan)}$ =0,617 dan $R^2_{(hujan)}$=0,046, maka dapat disimpulkan bahwa $R^2_{(tidak\ hujan)}$ = 0,617 mengindikasinya besarnya hubungan antara $NA_{(tidak\ hujan)}$ yang berarti mendekati nilai nyata (53,6%) dengan nilai statistika sebesar 61,7% melebihi nilai nyata dengan selisih lebih 8, sedangkan $R^2_{(hujan)}$ = 0,046 dengan nilai statistik 4,6% hubungan antara $NA_{(\ hujan)}$ belum bisa dibuktikan mendekati nilai nyata 46,4% karena selisihnya cukup besar yaitu selisih kurang sebesar 41,8%.

Nilai persentase pengujian 61,7% ini digunakan untuk membuktikan prediksi kondisi cuaca tidak hujan dapat dipengaruhi oleh tiga parameter cuaca yaitu kelembaban, suhu dan arah angin akibat proses partisi vertikal terhadap parameter cuaca.

Pengujian berikutnya adalah pengujian partisi 2 yaitu melihat pengaruh parameter kelembaban, kecepatan angin dan arah angin terhadap kondisi cuaca.(lihat tabel. 6).

Tabel. 6 Pengujian Partisi 2 Pohon Keputusan

Tanggal	Kelembaban	Kecepatan Angin	Arah Angin	Kondisi Nyata	Kondisi Hasil Partisi
1/1/09	79	3	NW	TidakHujan	Tidak Hujan
16/1/09	84	3	W	hujan	Tidak Hujan
23/1/09	79	2	W	Tidak hujan	Tidak Hujan
6/2/09	88	4	W	Hujan	Tidak Hujan
21/2/09	85	1	NW	Hujan	Hujan
2/3/09	74	2	NW	Hujan	Hujan
18/3/09	70	2	NW	Tidak Hujan	Tidak Hujan
30/3/09	86	1	W	Hujan	Tidak Hujan
9/4/09	77	2	SE	Tidak Hujan	Hujan
22/4/09	84	2	N	Hujan	Hujan
2/5/09	71	3	SW	Tidak Hujan	Tidak Hujan
17/5/09	76	2	W	Tidak Hujan	Tidak Hujan
31/5/09	80	1	C	Hujan	Hujan
8/6/09	81	2	W	Hujan	Tidak Hujan
25/6/09	82	0	W	Tidak Hujan	Tidak Hujan
3/7/09	81	1	W	Tidak Hujan	Tidak Hujan
20/7/09	78	1	S	Tidak Hujan	Tidak Hujan
1/8/09	67	2	NE	TidakHujan	Tidak Hujan
13/8/09	76	2	NE	TidakHujan	Tidak Hujan
30/8/09	71	3	NE	TidakHujan	Tidak Hujan
4/9/09	69	1	E	TidakHujan	Hujan
19/9/09	76	1	E	TidakHujan	Hujan
2/10/09	75	3	E	Hujan	Hujan
15/10/09	81	2	W	Hujan	Tidak Hujan
31/10/09	76	1	SE	TidakHujan	Tidak Hujan
9/11/09	71	1	W	TidakHujan	Tidak Hujan
25/11/09	84	1	W	Hujan	Tidak Hujan
5/12/09	84	1	W	TidakHujan	Tidak Hujan
19/12/09	72	1	W	TidakHujan	Tidak Hujan
31/12/09	92	1	W	Hujan	Tidak Hujan

Tabel.6 menunjukan nilai hasil partisi (R) dengan parameter kelembaban, kecepatam angin, arah angin mempengaruhi kondisi cuaca tidak hujan sebesar $R_{(tidak\ hujan)}$ = 22/30 = 0,733 sedangkan kondisi cuaca hujan sebesar $R_{(hujan)}$ = 8/30 = 0,267 dengan sebagai variabel terikat (NA) adalah kondisi nyata dengan $NA_{(tidak\ hujan)}$ = 18/30 = 0,600 dan kondisi nyata dengan $NA_{(hujan)}$ = 12/30 = 0,400. Jika dicari nilai R kuadrat maka $R^2_{(tidak\ hujan)}$ = 0,537 dan $R^2_{(hujan)}$ = 0,071, maka dapat disimpulkan bahwa $R^2_{(tidak\ hujan)}$ = 0,537 mengindikasinya besarnya hubungan antara $NA_{(tidak\ hujan)}$ yang berarti mendekati nilai nyata dengan nilai statistika sebesar 53,7%, sedangkan $R^2_{(hujan)}$ = 0,071 dengan nilai statistik 7,1% hubungan antara $NA_{(hujan)}$ belum

bisa dibuktikan mendekati nilai nyata 40% karena selisihnya cukup besar yaitu selisih kurang sebesar 32,9%. Nilai persentase pengujian 53,7% ini membuktikan pula bahwa prediksi kondisi cuaca tidak hujan dapat dipengaruhi oleh parameter cuaca yaitu kelembaban, kecepatan angin dan arah angin.

Berdasarkan hasil analisis pada partisi 1 dan partisi 2 perbandingan nilai persentase pengujian dengan nilai nyata dapat digunakan untuk prediksi kondisi cuaca tidak hujan saja, sedangkan untuk kondisi cuaca hujan belum bisa terbukti prediksinua karena nilai persentase pengujian kondisi cuaca jauh dari nilai kondisi nyata.

6. Kesimpulan

Pada makalah ini telah dilakukan partisi data untuk menghasilkan frekuensi asosiasi item set data cuaca dengan tingkat dukungan minimal tanpa mengungkapkan nilai item set individu. Dimana metoda klasifikasi pohon keputusan (*decision tree classifiers*) membantu dalam mengoptimalkan frekuensi asosiasi item set data cuaca tersebut**.** Nilai frekuensi asosiasi item set hasil partisi data ini dapat digunakan untuk simulasi prediksi cuaca apakah akan terjadi hujan atau tidak hujan.

DAFTAR PUSTAKA

[1] Divanis,A.G. and Verykios. V.S., "An Overview of Privacy Preserving Data Mining", The ACM Students Journal Cross Roads, Summer 2009 / Vol. 15, No. 4, 2009.

[2] Amirbekyan, A., Estivill, V., Castro, "The privacy of k-NN retrieval for horizontal partitioned data new methods and applications", Eighteenth Australasian Database Conference (ADC2007), Ballarat, Victoria, Australia, 2007.

[3] Vaidya, J., "Privacy Preserving Association Rule Mining in Vertically Partitioned Data", ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002.

[4] Sumana and Hareesdh, "An Approach of Private Classification on Vertically Partitioned Data", International Conference and Workshop on Emerging Trends in Technology (ICWET 2010) – TCET, Mumbai, India, 2010.

[5] Data Klimatologi Station Geofisika Kelas I Bandung (Badan Meteorologi, Klimatologi dan Geofisika (BMKG) Bandung), Garis Lintang : 06^{o}55' S, Garis Bujur : 107^{o} 36'E, Tinggi DPL 791 M. Periode Tahun 2005 – 2009.

[6] Law. A.M., and Kelton, W.D., "Simulation Modeling and Analysis", Mc.Graw Hill International Editions, New York, 2009.