

Model *Multinomial Bayesian Network* pada Data Simulasi Curah Hujan

NANDA ARISTA RIZKI, SYARIPUDDIN, DAN SRI WAHYUNINGSIH

Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman
Email: nanda89@rocketmail.com, syarif92@yahoo.co.id, dan swahyuningsih@gmail.com

ABSTRACT

Bayesian Networks is one of simple Probabilistic Graphical Models are built from theory of bayes probability and graph theory. Probability theory is directly related to data while graph theory directly related to the form representation to be obtained. Multinomial Bayesian Network method is one method that involves the influence of spatial linkages suggest a link between rainfall observation stations. The objective of this study was seek the result of the model probabilistic a graph Multinomial Bayesian Network and apply it in forecasting with Oldeman classification based on one or two rainfall stations are known. This research uses simulated data for 14 stations respectively each 300 sets of data. The data generated is normal distribution of data based on parameters that have been determined and classified using the classification Oldeman. Bayesian Network structure constructed using the K2 algorithm. Markov chain transition matrix is formed based on the Bayesian of the nodes are directional. Model of Multinomial Bayesian Network was established based on Markov transition matrices. The result of probability model can predict the probability of rainfall in some stations based on one or two rainfall stations are known, which is a model graph with 14 nodes and 13 arcs.

Keyword: Multinomial Bayesian Network, Direct Acyclic Graph, K2 algorithm, spatial effects.

1. PENDAHULUAN

Menurut Rahmat (2008), curah hujan merupakan variabel yang paling sering digunakan dalam penelitian mengenai dampak perubahan iklim. Berkaitan dengan iklim di Indonesia, proses pembentukan hujan di kawasan tropis merupakan proses yang paling sukar disimulasikan. Hingga saat ini belum ada suatu model iklim yang mampu mensimulasikan curah hujan di Indonesia dengan baik. Model-model iklim resolusi tinggi perlu dikembangkan dalam skala-skala lokal atau setara dengan skala provinsi dan kabupaten.

Informasi tentang banyaknya curah hujan sangat berguna bagi para petani dalam mengantisipasi kemungkinan terjadinya peristiwa-peristiwa ekstrim (kekeringan dan banjir) yang akan berakibat kegagalan dalam proses produksinya. Dengan demikian, ketersediaan informasi ini memerlukan suatu metode peramalan curah hujan yang akurat. Banyak metode yang dikembangkan untuk memprakirakan cuaca, khusus kejadian curah hujan. Beberapa penelitian yang sudah dilakukan belum memberikan hasil yang memuaskan.

Bayesian Networks merupakan salah satu model graf probabilistik sederhana yang dibangun dari teori peluang bayes dan teori graf. Teori peluang berhubungan langsung dengan data sedangkan teori graf berhubungan langsung dengan bentuk representasi yang ingin didapatkan.

Berdasarkan permasalahan tersebut, diperlukan suatu metode peramalan curah hujan yang mempertimbangkan pengaruh spasial antar stasiun. Metode *Multinomial Bayesian Network* merupakan salah satu metode yang melibatkan pengaruh spasial yang menunjukkan adanya hubungan keterkaitan antara stasiun pengamatan curah hujan. Peneliti ingin membuat model *Multinomial Bayesian Network* dan nilai peramalannya dengan klasifikasi Oldeman di beberapa stasiun berdasarkan satu atau dua curah hujan stasiun yang diketahui dari data simulasi curah hujan.

2. TINJAUAN PUSTAKA

2.1 Rantai Markov

Dalam rantai Markov yang dihasilkan adalah suatu informasi probabilistik yang dapat digunakan untuk membantu pembuatan keputusan, sehingga analisis ini bukan suatu teknik optimisasi melainkan suatu teknik deskriptif. Rantai Markov merupakan bentuk khusus dari model probabilistik yang lebih umum yang dinamakan proses stokastik, yaitu proses perubahan probabilistik yang terjadi terus-menerus.

Informasi yang dapat dihasilkan dari rantai Markov adalah peluang berada dalam suatu status pada satu periode di masa depan. Dalam banyak kasus, proses Markov akan menuju kepada kondisi *steady state* (keseimbangan) artinya setelah proses berjalan selama beberapa periode, peluang status akan bernilai tetap.

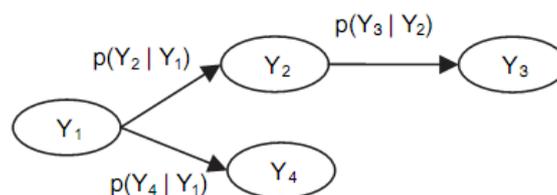
Menurut Karlin dan Taylor (1998), sebuah rantai Markov didefinisikan oleh satu langkah matriks transisi peluang dan spesifikasi dari sebuah distribusi peluang. Analisis rantai Markov biasanya mempunyai masalah terutama dalam perhitungan peluang realisasi yang mungkin dari suatu proses. Inti dalam perhitungan ini adalah n langkah transisi peluang dari matriks $\mathbf{P}^{(n)} = \|\|P_{ij}^{(n)}\|\|$. Di sini $P_{ij}^{(n)}$ menunjukkan peluang bahwa proses berjalan dari keadaan i ke keadaan j dalam transisi n . Secara umum,

$$P_{ij}^{(n)} = P\{Y_{m+n} = j | Y_m = i\}. \quad (1)$$

Perhatikan bahwa disini hanya mempersoalkan proses temporal homogen memiliki peluang transisi stasioner, karena jika tidak sisi kiri dari persamaan (1) juga akan tergantung pada m .

2.2 Bayesian Networks (BNs)

BNs merupakan salah satu *Probabilistic Graphical Model (PGM)* yang sederhana yang dibangun dari teori peluang bayes dan teori graf. Teori peluang berhubungan langsung dengan data sedangkan teori graf berhubungan langsung dengan bentuk representasi yang ingin didapatkan. *BNs* dapat memberikan informasi yang sederhana dan padat mengenai informasi peluang. Berdasarkan komponennya, *BNs* terdiri dari *Bayesian structure (Bs)* dan *Bayesian parameter (Bp)* (Cooper dan Herskovits, 1992). *Bs* merupakan sebuah *Direct Acyclic Graph (DAG)* yang menggambarkan ketergantungan antar setiap variabel, dan *Bp* merupakan himpunan dari parameter dari distribusi peluang bersyarat setiap variabel berdasarkan graf tersebut. *Bs* terdiri dari *node* yang merepresentasikan variabel-variabel dan *edge* yang merepresentasikan hubungan antar *node* seperti pada Gambar 1. Setiap *node* yang dihubungkan secara langsung, menunjukkan hubungan ketergantungan. Misalkan himpunan dari *node* dinyatakan dengan $\{Y_1, \dots, Y_n\}$, jika terdapat *edge* dari *node* Y_j ke *node* Y_k , dikatakan bahwa Y_j adalah *parent* dari Y_k , dan Y_k adalah *child* dari Y_j . Himpunan *parent* dari *node* Y_i dinotasikan sebagai π_i . Sebagai contoh, berdasarkan Gambar 1 *parent* dari Y_2 adalah Y_1 dan *child* untuk Y_2 adalah Y_3 .



Gambar 1. *Directed Acyclic Graph (DAG)*

Menurut Meigarani (2010), terdapat beberapa konsep dasar yang diperlukan dalam mengkonstruksi struktur *Bayesian Networks* dari data. Konsep dasar tersebut yaitu:

1. Struktur Graf
Struktur graf yang dimiliki oleh *BNs* adalah *Direct Acyclic Graph (DAG)*, yaitu graf berarah yang tidak memiliki siklus.
2. Kondisi Markov

Kondisi ini menunjukkan hubungan antara DAG dan distribusi peluang. *Bayesian Networks* memanfaatkan kondisi Markov untuk melakukan representasi distribusi peluang bersama secara efisien dan memperoleh adanya kebebasan kondisional antara variabel.

3. *D-separation*
Properti DAG yang menyatakan hubungan kebebasan yang terdapat pada DAG. Semua kebebasan kondisional yang diperoleh dari kondisi Markov akan diidentifikasi dengan properti ini.
4. Ekuivalensi
Konsep tentang adanya DAG-DAG yang memiliki *d-separation* yang sama, yang dapat direpresentasikan dalam sebuah DAG pattern. Dua DAG merupakan ekuivalen Markov jika dan hanya jika kedua DAG tersebut mempunyai *link-link* (*edge* tanpa memperhatikan arah) yang sama dan himpunan *uncoupled head-to-head meeting* yang sama.
5. Kondisi *Faithfulness*
Kondisi yang harus dipenuhi agar adanya *edge* diantara *node* pada DAG berarti ada ketergantungan langsung antar *node* tersebut.

Struktur ketergantungan atau kebebasan yang digambarkan dengan DAG dapat diterjemahkan ke dalam fungsi densitas bersama dari variabel-variabel dengan jalan mengalikan semua peluang berdasarkan *parent*-nya sebagai berikut:

$$P(Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n P(Y_i | \pi_i) \tag{2}$$

Berdasarkan persamaan (2) dapat diketahui bahwa DAG ini mendefinisikan dekomposisi dari fungsi peluang berdimensi besar ke dalam distribusi lokal berdimensi rendah.

2.3 Pembentukan BNs berdasarkan Data

Menurut Meigarani (2010), pembentukan BNs dibagi menjadi dua tahap, yaitu konstruksi struktur dan estimasi parameter. Struktur BNs adalah DAG yang merepresentasikan sebuah pola dari sekumpulan data. Perepresentasian dalam bentuk graf dapat dilakukan dengan mengidentifikasi konsep-konsep informasi yang relevan terhadap masalah. Selanjutnya konsep-konsep tersebut disebut himpunan variabel. Himpunan tersebut kemudian direpresentasikan menjadi *node-node* dalam graf. Pengaruh antara variabel dinyatakan eksplisit menggunakan *edge* pada graf.

Setelah struktur BNs terbentuk, parameter dan hubungan ketergantungan antar *node* ditentukan dengan menggunakan pengetahuan pakar. Informasi ini dibutuhkan agar dapat menghitung distribusi peluang bersama.

Permasalahan yang dihadapi adalah menentukan struktur yang terbaik dari semua struktur yang mungkin. Banyaknya struktur untuk *n node* diformulasikan sebagai berikut (Cooper dan Herskovits, 1992):

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i). \tag{3}$$

Pada persamaan (3) terdapat batasan untuk $f(0) = 1$, banyaknya struktur untuk beberapa *n* dapat dilihat pada Tabel 1.

Tabel 1. Banyaknya Struktur untuk *n Node*

<i>n</i>	Jumlah Struktur
2	3
3	25
5	29000
10	4,2 x 10 ¹⁸

Jika diasumsikan bahwa variabel tersebut diurut, dimana jika Y_i mendahului Y_j dalam urutan, maka tidak diperbolehkan terdapat tanda panah dari Y_j ke Y_i . Berdasarkan aturan pengurutan tersebut maka kemungkinan struktur BNs yang dapat terbentuk sebanyak $2^{\binom{n}{2}} = 2^{\binom{n(n-1)}{2}}$ kemungkinan. Sehingga perlu suatu algoritma yang memberikan struktur terbaik. Salah satu algoritma dalam mencari struktur BNs adalah algoritma K2 (Rachmat, 2008).

2.4 Algoritma K2

Algoritma K2 menentukan struktur jaringan Bs yang memaksimalkan peluang Bs dengan Db pengamatan yang dinotasikan dengan $P(Bs, Db)$ dengan mengasumsikan bahwa variabel tersebut terurut. Algoritma ini diawali dengan mengasumsikan bahwa setiap *node* tersebut tidak mempunyai *parent*, kemudian dilakukan penambahan *parent* dimana penambahan tersebut meningkatkan peluang dari hasil akhir struktur. Jika penambahan *parent* sudah tidak lagi meningkatkan peluang dari hasil akhir struktur, maka penambahan *parent* dihentikan. Adapun fungsi yang menjadi acuan peningkatan nilai dari peluang strukturnya adalah sebagai berikut:

$$g(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i-1)!}{(N_{ij}+r_i-1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (4)$$

N_{ijk} dihitung relatif terhadap π_i yang merupakan *parent* dari Y_i dan relatif terhadap himpunan pengamatan Db . Fungsi $Pred(Y_i)$ merupakan fungsi yang mengembalikan himpunan dari *node* yang mendahului Y_i dalam urutan *node*.

Adapun algoritma K2 adalah sebagai berikut:

```

Procedure K2
For i:=1 to n do
   $\pi_i = \phi$ ;
  Pold=  $g(i, \pi_i)$ ;
  OKToProceed:=true
  while OKToProceed and  $|\pi_i| < u$  do
    Let  $v$  be the node in  $Pred(Y_i) - \pi_i$  that maximizes  $g(i, \pi_i \cup \{v\})$ ;
    Pnew=  $g(i, \pi_i \cup \{v\})$ ;
    if Pnew>Pold then
      Pold:=Pnew;
       $\pi_i = \pi_i \cup \{v\}$ ;
    else OKToProceed:=false;
  end {while}
  write("node:", "parent of this nodes:",  $\pi_i$ );
end {for}
end {K2}

```

2.5 Multinomial Bayesian Network

Dalam *Multinomial Bayesian Network* diasumsikan bahwa semua variabel adalah diskrit, dimana setiap variabel memiliki himpunan nilai yang terbatas, dan bahwa peluang bersyarat untuk setiap variabel berdasarkan *parent*-nya menyebar *multinomial*.

Menurut Cooper dan Herskovits (1992), nilai rata-rata dari peluang bersyarat dalam jaringan didefinisikan sebagai berikut: misal θ_{ijk} dinotasikan sebagai peluang bersyarat $P(Y_i = m_{ik} | \pi_i = w_{ij})$, yang merupakan peluang bahwa Y_i memiliki nilai m_{ik} , $k = 1, 2, \dots, r_i$, dengan syarat *parent* dari Y_i , yang dinyatakan dengan π_i memiliki nilai w_{ij} . Bila θ_{ijk} sebagai peluang bersyarat dari jaringan (*network conditional probability*), N_{ijk} adalah banyaknya jumlah pengamatan pada *node* ke- i yang memiliki *node parent* ke- j untuk kategori ke- k , $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$ dan misalkan δ dinotasikan sebagai asumsi bahwa:

1. Semua variabel merupakan variabel diskrit
2. Setiap observasi saling bebas
3. Tidak ada data yang hilang dari setiap variabel
4. Fungsi kepadatan peluang $f(Bp | Bs)$ adalah *uniform*,

maka nilai $E(\theta_{ijk} | Db, Bs, \delta)$, yang merupakan nilai rata-rata dari θ_{ijk} diketahui himpunan pengamatan Db , struktur jaringan Bs , dan asumsi δ dinyatakan sebagai berikut:

$$E(\theta_{ijk} | Db, Bs, \delta) = \frac{N_{ijk} + 1}{N_{ij} + r_i} \quad (5)$$

sedangkan variansinya adalah:

$$\text{Var}(\theta_{ijk}|Db, B_s, \delta) = \frac{(N_{ijk}+1)(N_{ij+r_i}-N_{ijk}-1)}{(N_{ij+r_i})^2(N_{ij+r_i}+1)}. \quad (6)$$

2.6 Akurasi dari Model Peramalan

Akurasi selalu dikaitkan dengan tingkat kesesuaian model (*Goodness of fit*), yang berarti seberapa baik model merepresentasikan pola data, sehingga dapat digunakan sebagai alat peramalan. Pengukuran akurasi didasarkan pada peramalan dari data yang digunakan untuk estimasi parameter model (Y_1, Y_2, \dots, Y_n) , yang berarti dibandingkan dengan nilai sebenarnya $(\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n)$.

Ukuran akurasi yang digunakan dalam skripsi ini adalah *Mean Square of Error* (MSE) yaitu nilai rata-rata dari jumlah kuadrat kesalahan, yang didefinisikan sebagai berikut:

$$MSE = \sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n} \quad (7)$$

dengan Y_i menyatakan nilai yang sebenarnya, \hat{Y}_i menyatakan hasil ramalan dari model yang digunakan, dan n adalah jumlah data yang diramalkan.

2.7 Curah Hujan

Menurut Tjasyono (2004), endapan (*presipitasi*) didefinisikan sebagai bentuk air cair dan padat (es) yang jatuh ke permukaan bumi. Meskipun kabut, embun, dan embun beku (*frost*) dapat berperan dalam alih kebasahan (*moisture*) dari atmosfer ke permukaan bumi, unsure tersebut tidak ditinjau sebagai endapan. Bentuk endapan adalah hujan, gerimis, salju, dan batu es hujan (*hail*). Hujan adalah bentuk endapan yang sering dijumpai, dan di Indonesia yang dimaksud dengan endapan adalah curah hujan.

Curah hujan adalah jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi (mm) di atas permukaan horizontal bila tidak terjadi *evaporasi*, *runoff* dan *infiltrasi*. Satuan curah hujan adalah mm, inch. Curah hujan 1 mm artinya air hujan setelah 1 mm tidak mengalir, tidak meresap ke dalam tanah dan tidak menguap ke atmosfer. Hari hujan artinya suatu hari dimana curah hujan kurang dari 0,5 mm per hari, jumlah ini tidak berarti bagi tanaman, karena akan habis menguap apabila ada angin. Hari hujan tanaman artinya suatu hari dimana curah hujannya kurang dari 2,5 mm dan dapat dimanfaatkan oleh tanaman.

Curah hujan ditentukan berdasarkan banyaknya curah hujan (dalam kurun waktu, misalnya selama sehari), intensitas hujan (banyaknya curah hujan per satuan waktu selama hujan berlangsung), dan hari hujan (banyaknya hari ada hujan dalam suatu kurun waktu, misalnya dalam sebulan). Banyaknya hujan 1 mm setara dengan 1 liter air hujan yang jatuh pada permukaan bumi seluas 1 m².

3. METODOLOGI PENELITIAN

Penelitian dilaksanakan pada bulan Oktober sampai November 2011 dengan tempat pembuatan data simulasi dan tempat penelitian di Laboratorium Statistika Komputasi Kampus Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman.

Penelitian ini menggunakan rancangan penelitian deskriptif karena secara sistematis hanya menggambarkan fakta dan karakteristik objek yang diteliti secara tepat. Data yang digunakan merupakan data simulasi curah hujan yang dibangkitkan dari Minitab.

Penelitian ini menggunakan data simulasi komputer. Data simulasi dibangkitkan secara acak dengan bantuan perangkat lunak Minitab 16. Data yang dibangkitkan adalah data berdistribusi normal berdasarkan parameter-parameter yang telah ditentukan yaitu rata-rata (μ) dan standar deviasi (σ). Untuk keperluan analisis, dibangkitkan masing-masing 300 set data untuk 14 stasiun dengan mengikuti distribusi normal berdasarkan parameter-parameter yang telah ditentukan. Masing-masing set data dipandang sebagai sampel dari populasi yang telah ditetapkan.

3.1 Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

- Y_1 = stasiun curah hujan di kota A
- Y_2 = stasiun curah hujan di kota B
- Y_3 = stasiun curah hujan di kota C
- Y_4 = stasiun curah hujan di kota D
- Y_5 = stasiun curah hujan di kota E
- Y_6 = stasiun curah hujan di kota F
- Y_7 = stasiun curah hujan di kota G
- Y_8 = stasiun curah hujan di kota H
- Y_9 = stasiun curah hujan di kota I
- Y_{10} = stasiun curah hujan di kota J
- Y_{11} = stasiun curah hujan di kota K
- Y_{12} = stasiun curah hujan di kota L
- Y_{13} = stasiun curah hujan di kota M
- Y_{14} = stasiun curah hujan di kota N

3.2 Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini adalah menerapkan model *Multinomial Bayesian Network* pada data simulasi curah hujan untuk 14 stasiun dan pada penelitian ini peneliti menggunakan bantuan perangkat lunak Minitab untuk membuat deskriptif dari data, pembentukan struktur menggunakan algoritma K2 dengan Matlab, dan membuat model *Multinomial Bayesian Network* dengan Netica. Adapun langkah-langkah yang dilakukan dalam analisis ini, yaitu sebagai berikut:

1. Analisis deskriptif
2. Pengklasifikasian data
Klasifikasi data curah hujan yang digunakan adalah klasifikasi Oldeman. Data curah hujan bulan kering diberi kode 3, curah hujan bulan lembab diberi kode 2, dan curah hujan bulan basah diberi kode 1.
3. Pembentukan struktur jaringan
Pembentukan jaringan *Bayesian* menggunakan algoritma K2 dimana peubah atau stasiun harus diurutkan terlebih dahulu. Pengurutan ini didasarkan pada posisi stasiun dan pertimbangan letak ketinggian.
4. Pembentukan matriks transisi Markov pada *node-node* yang berarah yang diperoleh dengan menghitung fungsi sebaran peluang lokal berdasarkan peluang bersyarat.
5. Pembentukan model *Multinomial Bayesian Network* menggunakan bantuan perangkat lunak Netica 4.16.
6. Peramalan peluang curah hujan di beberapa stasiun berdasarkan satu atau dua curah hujan stasiun yang diketahui.

3.3 Kerangka Penelitian

Penelitian diawali dengan merumuskan masalah yang akan diteliti dalam penelitian. Data simulasi curah hujan dibangkitkan menggunakan perangkat lunak Minitab. Untuk keperluan analisis, dibangkitkan masing-masing 300 set data untuk setiap stasiun. Dilakukan analisis deskriptif pada data untuk mengetahui gambaran umum curah hujan di setiap stasiun. Selanjutnya adalah mengklasifikasikan data simulasi curah hujan menggunakan klasifikasi Oldeman. Struktur jaringan *Bayesian* dibentuk dengan menggunakan algoritma K2. Matriks transisi rantai Markov dibentuk berdasarkan pada sifat Bayesian dari *node-node* (stasiun-stasiun) yang berarah. Model *Multinomial Bayesian Network* dibentuk berdasarkan matriks transisi Markov dan menggunakan bantuan perangkat lunak Netica 4.16. Model peluang ini dapat meramalkan peluang curah hujan di beberapa stasiun berdasarkan satu atau dua curah hujan stasiun yang diketahui.

4. HASIL DAN PEMBAHASAN

Salah satu manfaat curah hujan yaitu untuk mengetahui prakiraan hujan turun akan turun tidaknya di suatu wilayah, misalnya di Samarinda. Manfaat lainnya yaitu untuk mengetahui cocok tidaknya mengembangkan suatu jenis tanaman berdasarkan curah hujan.

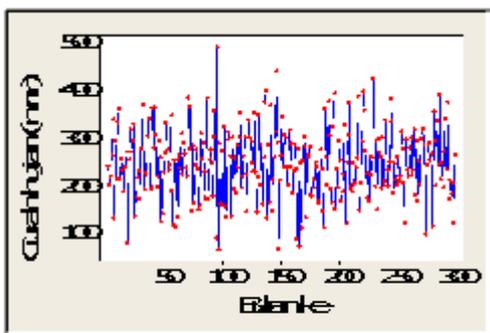
Penelitian ini menggunakan data simulasi komputer. Data simulasi dibangkitkan secara acak dengan bantuan perangkat lunak Minitab 16. Data yang dibangkitkan adalah data dengan distribusi normal berdasarkan parameter-parameter yang telah ditentukan. Nilai ketinggian diukur dari ketinggian stasiun di atas laut. Letak spasial yang digunakan mengikuti letak spasial di provinsi Kalimantan Timur, Indonesia. Adapun asumsi yang digunakan untuk penelitian ini adalah semua variabel merupakan variabel diskrit, setiap observasi saling bebas, dan tidak ada data yang hilang dari setiap variabel.

4.1 Analisis Deskriptif

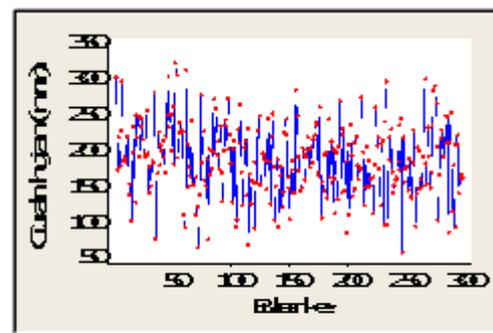
Analisis deskriptif dari data curah hujan di setiap stasiun masing-masing dapat dilihat pada Tabel 2.

Tabel 2. Rata-rata dan Standar Deviasi dari Data Simulasi

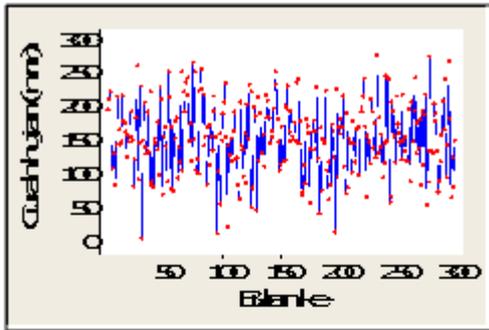
Variabel	Rata-rata (mm)	Standar Deviasi (mm)	Variabel	Rata-rata (mm)	Standar Deviasi (mm)
A	248,57	73,40	H	258,44	76,34
B	187,13	49,90	I	269,83	75,69
C	156,63	52,78	J	220,83	60,18
D	153,63	51,62	K	203,38	55,80
E	275,20	86,90	L	177,56	47,83
F	207,73	61,04	M	237,85	59,35
G	169,17	47,96	N	229,51	59,16



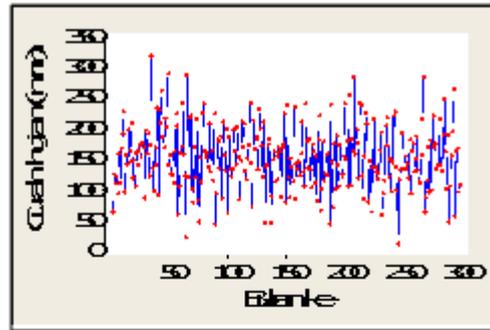
Gambar 2. Pola curah hujan di stasiun kota A



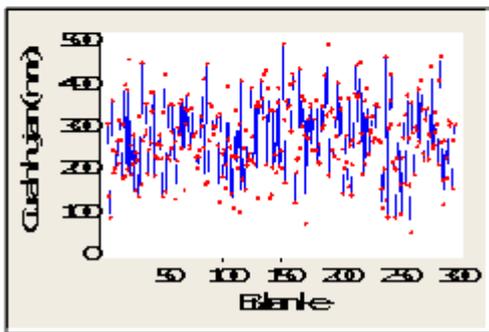
Gambar 3. Pola curah hujan di stasiun kota B



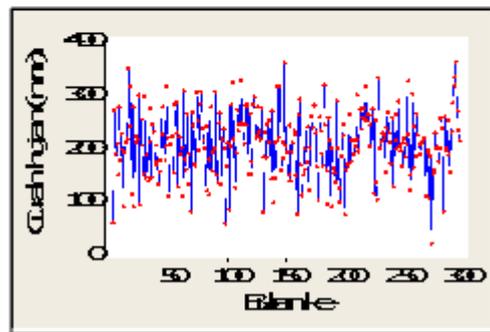
Gambar 4.
Pola curah hujan di stasiun kota C



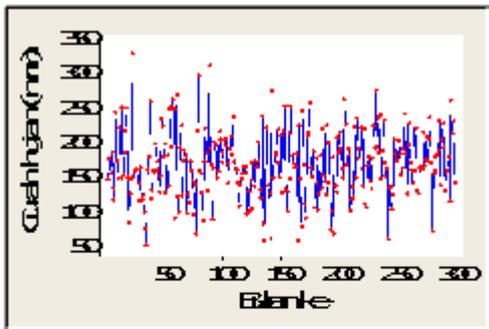
Gambar 5.
Pola curah hujan di stasiun kota D



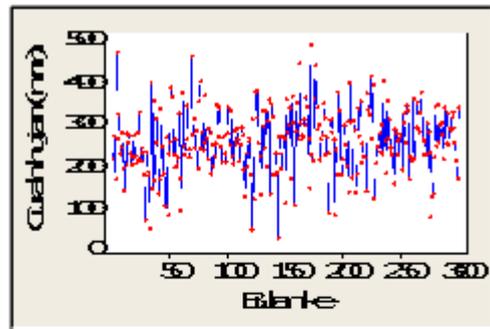
Gambar 6.
Pola curah hujan di stasiun kota E



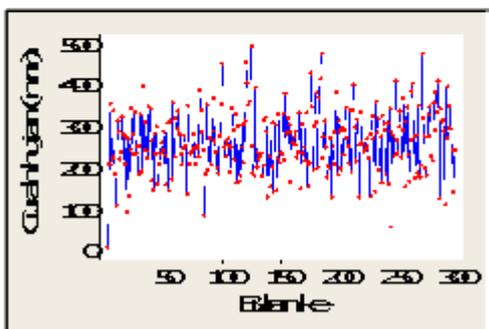
Gambar 7.
Pola curah hujan di stasiun kota F



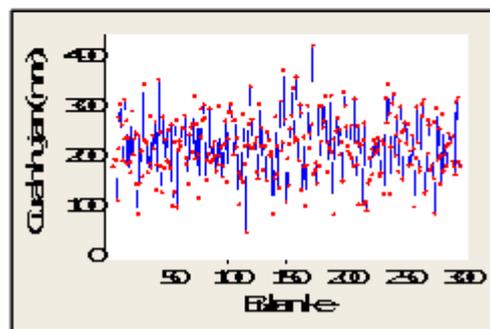
Gambar 8.
Pola curah hujan di stasiun kota G



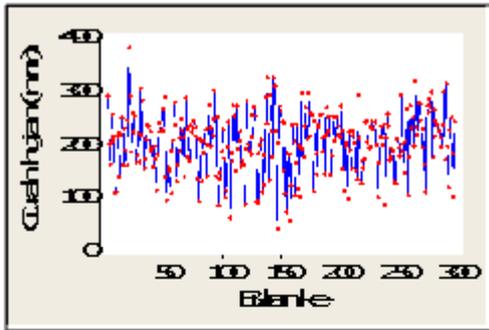
Gambar 9.
Pola curah hujan di stasiun kota H



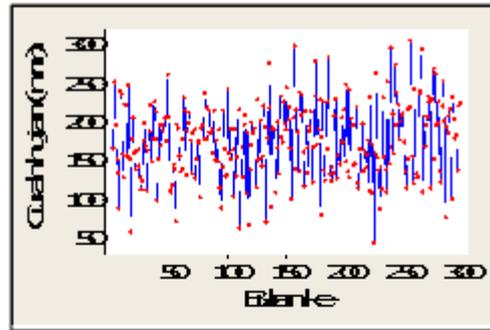
Gambar 10.
Pola curah hujan di stasiun kota I



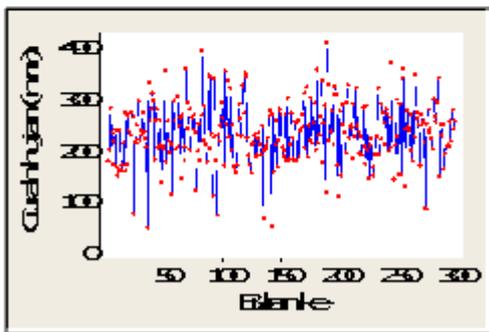
Gambar 11.
Pola curah hujan di stasiun kota J



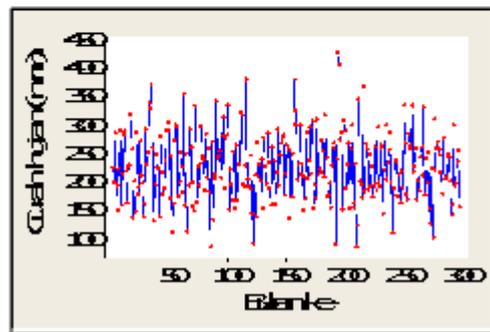
Gambar 12.
Pola curah hujan di stasiun kota K



Gambar 13.
Pola curah hujan di stasiun kota L



Gambar 14.
Pola curah hujan di stasiun kota M



Gambar 15.
Pola curah hujan di stasiun kota N

Gambar 2 sampai 15 menunjukkan bahwa pola data curah hujan dalam jangka panjang di stasiun kota A sampai kota N adalah stasioner dalam rata-rata.

4.2 Pengklasifikasian Data

Klasifikasi curah hujan menggunakan klasifikasi Oldeman dimana kode 3 menandakan curah hujan tersebut kering (antara 0-100 mm per bulan), kode 2 menandakan curah hujan tersebut lembab (antara 101-200 mm per bulan), dan kode 1 jika curah hujan basah (lebih dari 200 mm per bulan). Statistik deskriptif data klasifikasi dapat dilihat pada Tabel 3.

Tabel 3 menunjukkan bahwa peluang bulan kering untuk setiap stasiun kota lebih rendah dari peluang curah hujan bulanan lainnya, sedangkan stasiun kota yang mempunyai nilai peluang bulan basah lebih tinggi dari peluang curah hujan bulanan lainnya adalah kota A (75%), kota E (82%), kota F (57,3%), kota H (81,3%), kota I (83%), kota J (64%), kota K (57%), kota M (76,3%), dan kota N (67%). Nilai peluang curah hujan bulanan untuk setiap stasiun pada *state* tertentu berdasarkan klasifikasi Oldeman diperoleh dari jumlah curah hujan dengan *state* tersebut dibagi dengan jumlah set data.

Tabel 3. Statistik Deskriptif Data Klasifikasi

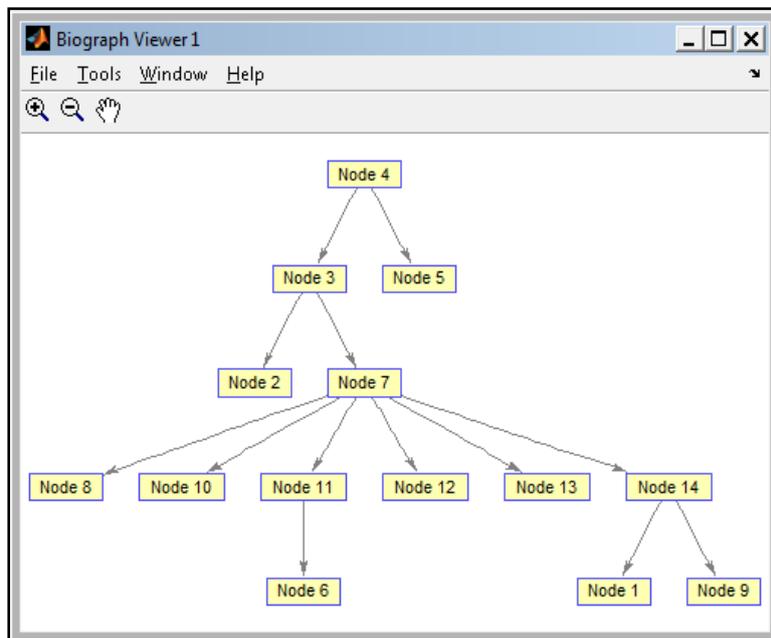
Variabel	Rata-rata Klasifikasi	Standar Deviasi Klasifikasi	Peluang		
			Bulan Kering	Bulan Lembab	Bulan Basah
A	1,273	0,496	0,023	0,227	0,750
B	1,660	0,570	0,050	0,560	0,390
C	1,950	0,607	0,160	0,630	0,210
D	1,953	0,576	0,143	0,667	0,190
E	1,203	0,458	0,023	0,157	0,820
F	1,480	0,598	0,053	0,373	0,573
G	1,823	0,535	0,070	0,683	0,247
H	1,217	0,480	0,030	0,157	0,813
I	1,183	0,421	0,013	0,157	0,830
J	1,387	0,540	0,027	0,333	0,640
K	1,477	0,586	0,047	0,383	0,570
L	1,713	0,534	0,040	0,633	0,327
M	1,257	0,481	0,020	0,217	0,763
N	1,343	0,503	0,013	0,317	0,670

4.3 Pembentukan Struktur Jaringan

Pembentukan struktur pada ilustrasi ini didasarkan pada data dengan menggunakan algoritma K2. Algoritma K2 adalah algoritma *Greedy Search* sederhana untuk mendapatkan struktur *Bayesian* berkualitas tinggi dengan waktu yang wajar. Iterasi dari algoritma K2 mengasumsikan bahwa *node* tersebut dalam keadaan terurut dan dimulai dari struktur yang tidak mempunyai jaringan (Rachmat, 2008). Dalam urutan tersebut ada aturan bahwa *node* pada urutan awal tidak mungkin menjadi *child* bagi *node* urutan berikutnya. Proses pengurutan tersebut didasarkan pada informasi bahwa pantai sebelah barat setiap pulau memperoleh curah hujan selalu lebih banyak daripada pantai sebelah timur dan curah hujan juga bertambah sesuai dengan ketinggian tempat. Curah hujan terbanyak umumnya berada ketinggian antara 600-900 m di atas permukaan laut (Pamungkas, 2006), informasi ini diperkuat dengan hasil penelitian dari Coffino *et al* (Rachmat, 2008). Berdasarkan informasi tersebut maka urutan stasiunnya adalah stasiun kota D, C, G, L, B, K, F, J, N, M, A, H, I, dan E. Pembentukan matriks *DAG* menggunakan algoritma K2 dengan bantuan perangkat lunak Matlab. Matriks *DAG* merupakan matriks yang merepresentasikan struktur *DAG*.

$$DAG = \begin{matrix} & \begin{matrix} A & B & C & D & E & F & G & H & I & J & K & L & M & N \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \\ I \\ J \\ K \\ L \\ M \\ N \end{matrix} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

Angka 1 menunjukkan adanya *node* ke- yang memiliki arah panah ke *node* ke-. *Node C* memiliki arah panah ke *node B* dan *node G*; *node D* memiliki arah panah ke *node C* dan *node E*; *node G* memiliki arah panah ke *node H*, *node J*, *node K*, *node L*, *node M*, dan *node N*; *node K* memiliki arah panah ke *node F*, dan *node N* memiliki arah panah ke *node A* dan *node I*, sehingga *DAG* yang terbentuk memiliki 14 *node* dan 13 *arc* seperti pada Gambar 16.



Gambar 16. Struktur *DAG* pada 14 stasiun

dengan *node* ke-1 adalah stasiun kota *A*, *node* ke-2 adalah stasiun kota *B*, *node* ke-3 adalah stasiun kota *C*, dan seterusnya hingga *node* ke-14.

4.4 Pembentukan Matriks Transisi Markov pada *Node-Node* yang Berarah

Langkah selanjutnya adalah membentuk matriks transisi Markov yang merupakan peluang bersyarat suatu *node* terhadap *parent*-nya. Dengan *state* 1 menyatakan curah hujan bulanan

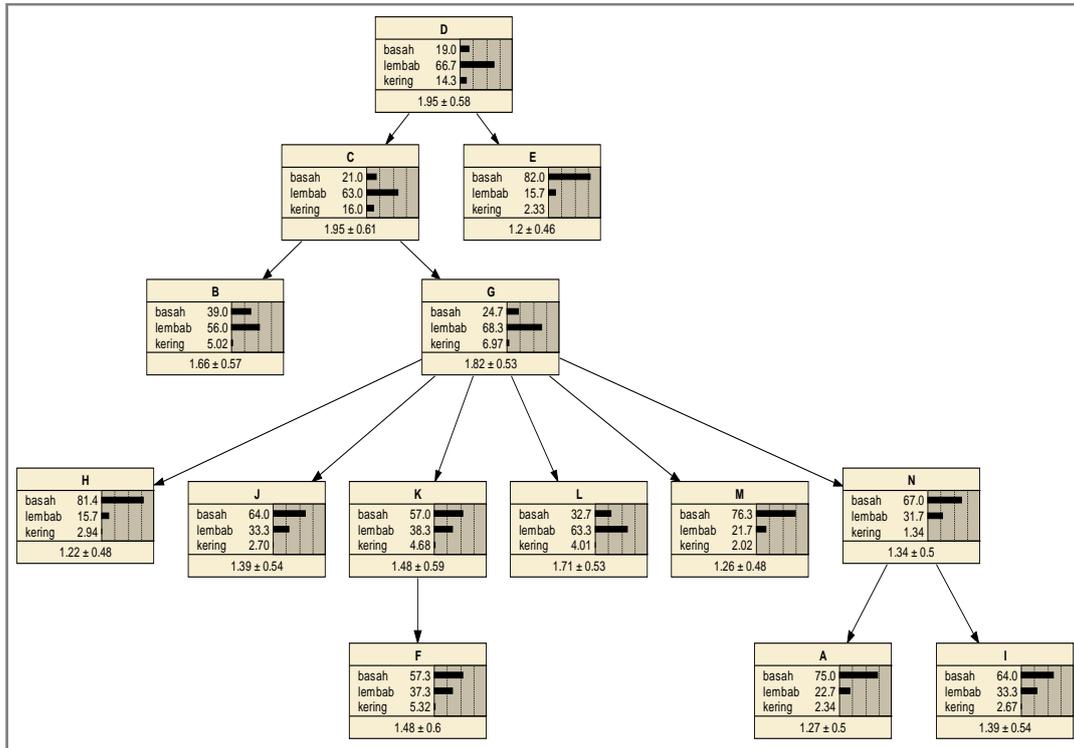
pada di suatu stasiun kota adalah basah, *state 2* menyatakan curah hujan bulanan pada di suatu stasiun kota adalah lembab, dan *state 3* menyatakan curah hujan bulanan pada di suatu stasiun kota adalah kering. Adapun matriks transisi Markov yang dibentuk adalah sebagai berikut:

$$\begin{aligned}
 P(C|D) &= \begin{pmatrix} 0,176 & 0,684 & 0,140 \\ 0,215 & 0,625 & 0,160 \\ 0,233 & 0,581 & 0,186 \end{pmatrix}; P(E|D) = \begin{pmatrix} 0,930 & 0,070 & 0,000 \\ 0,775 & 0,190 & 0,035 \\ 0,884 & 0,116 & 0,000 \end{pmatrix} \\
 P(B|C) &= \begin{pmatrix} 0,271 & 0,681 & 0,048 \\ 0,397 & 0,550 & 0,053 \\ 0,520 & 0,438 & 0,042 \end{pmatrix}; P(G|C) = \begin{pmatrix} 0,159 & 0,778 & 0,063 \\ 0,265 & 0,656 & 0,079 \\ 0,292 & 0,666 & 0,042 \end{pmatrix} \\
 P(H|G) &= \begin{pmatrix} 0,811 & 0,176 & 0,013 \\ 0,820 & 0,161 & 0,019 \\ 0,762 & 0,048 & 0,190 \end{pmatrix}; P(J|G) = \begin{pmatrix} 0,662 & 0,284 & 0,054 \\ 0,620 & 0,360 & 0,020 \\ 0,762 & 0,238 & 0,000 \end{pmatrix} \\
 P(K|G) &= \begin{pmatrix} 0,473 & 0,473 & 0,054 \\ 0,600 & 0,351 & 0,049 \\ 0,619 & 0,381 & 0,000 \end{pmatrix}; P(L|G) = \begin{pmatrix} 0,297 & 0,676 & 0,027 \\ 0,332 & 0,619 & 0,049 \\ 0,381 & 0,619 & 0,000 \end{pmatrix} \\
 P(M|G) &= \begin{pmatrix} 0,676 & 0,311 & 0,013 \\ 0,785 & 0,195 & 0,020 \\ 0,857 & 0,095 & 0,048 \end{pmatrix}; P(N|G) = \begin{pmatrix} 0,757 & 0,230 & 0,013 \\ 0,639 & 0,351 & 0,010 \\ 0,667 & 0,285 & 0,048 \end{pmatrix} \\
 P(F|K) &= \begin{pmatrix} 0,585 & 0,345 & 0,070 \\ 0,539 & 0,435 & 0,026 \\ 0,714 & 0,214 & 0,072 \end{pmatrix}; P(A|N) = \begin{pmatrix} 0,736 & 0,239 & 0,025 \\ 0,800 & 0,179 & 0,021 \\ 0,250 & 0,750 & 0,000 \end{pmatrix} \\
 P(I|N) &= \begin{pmatrix} 0,652 & 0,318 & 0,030 \\ 0,621 & 0,358 & 0,021 \\ 0,500 & 0,500 & 0,000 \end{pmatrix}
 \end{aligned}$$

4.5 Pemodelan Model Peluang Multinomial Bayesian Network

Pemodelan peluang diperoleh menggunakan *Multinomial Bayesian Networks*. Penentuan peluang berdasarkan jaringan diperoleh dari nilai peluang bersyarat dalam jaringan.

Pada Gambar 17 diperoleh informasi mengenai besaran peluang (dalam persen) munculnya bulan basah, lembab dan kering untuk setiap stasiun.

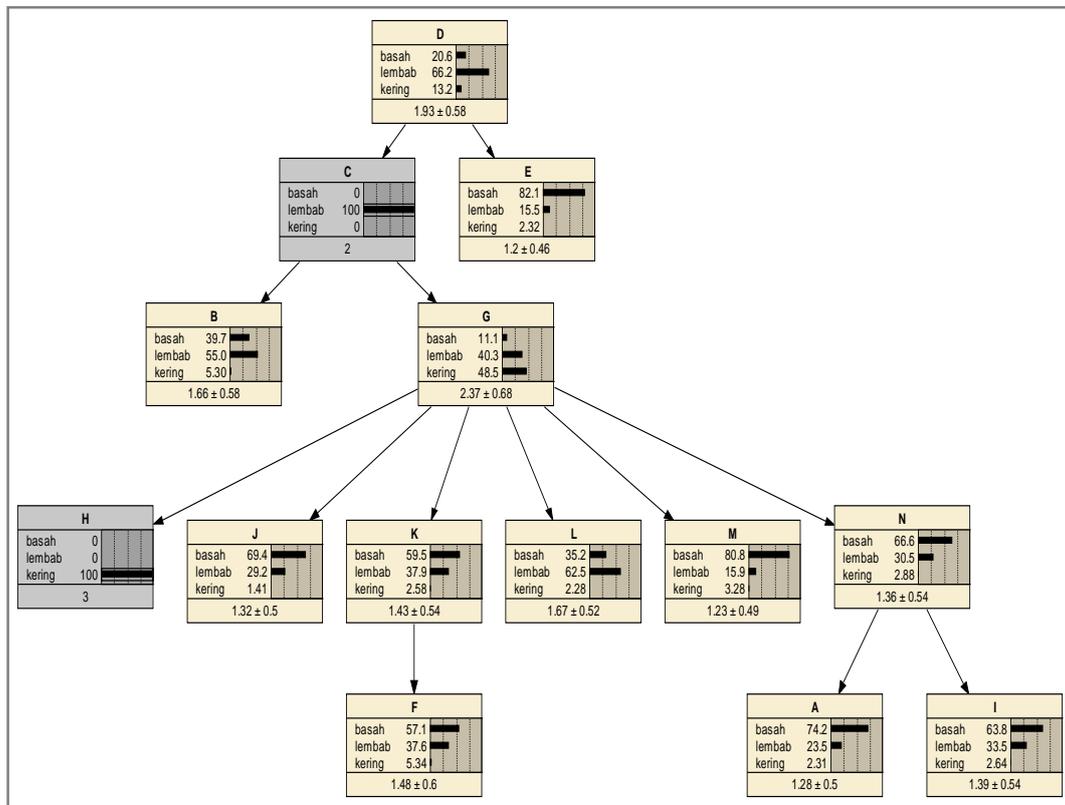


Gambar 17. Model *Multinomial Bayesian Network* berdasarkan data yang digunakan.

Kecenderungan terjadinya bulan kering untuk setiap stasiun lebih sedikit dibanding bulan yang lain. Kecenderungan terjadinya bulan basah terdapat pada stasiun kota E, H, J, K, M, N, F, A, dan I sedangkan kecenderungan terjadinya bulan lembab terdapat pada stasiun kota D, C, B, G, dan L. Nilai peluang bulan kering terkecil dan terbesar masing-masing adalah 1,3 % (stasiun kota M) dan 16 % (stasiun kota C), nilai peluang bulan lembab terkecil dan terbesar masing-masing adalah 15,7 % (stasiun kota E dan H) dan 68,3 % (stasiun kota G), dan nilai peluang bulan kering terkecil dan terbesar masing-masing adalah 19 % (stasiun kota D) dan 82 % (stasiun kota E). Informasi yang lain adalah rata-rata dan standar deviasi dari klasifikasi.

4.6 Peramalan Peluang Curah Hujan di Beberapa Stasiun

Peluang untuk setiap stasiun dapat berubah dengan ditentukannya nilai dari stasiun lain. Sebagai contoh jika diketahui di stasiun kota C terjadi bulan lembab dan stasiun kota H terjadi bulan kering maka perubahan per peluang dapat dilihat pada Gambar 18. Tampak bahwa peluang kecenderungan setiap stasiun kota adalah stabil, namun stasiun kota G cenderung menunjukkan curah hujan kering dibanding dengan stasiun kota G pada model graf pada Gambar 18 yang menunjukkan curah hujan lembab.



Gambar 18. Model *Multinomial Bayesian Network* jika di stasiun kota C terjadi bulan lembab dan stasiun kota H terjadi bulan kering.

Kecenderungan terjadinya bulan kering untuk setiap stasiun lebih sedikit dibanding bulan yang lain kecuali pada stasiun kota G yaitu 48,5% berpeluang bulan kering. Kecenderungan terjadinya bulan basah terdapat pada stasiun kota E, J, K, M, N, F, A, dan I sedangkan kecenderungan terjadinya bulan lembab terdapat pada stasiun kota D, B, dan L. Nilai ramalan peluang bulan kering terkecil dan terbesar masing-masing berdasarkan stasiun kota C terjadi bulan lembab dan stasiun kota H terjadi bulan kering adalah 1,4 % (stasiun kota J) dan 48,5 % (stasiun kota G), nilai ramalan peluang bulan lembab terkecil dan terbesar masing-masing berdasarkan stasiun kota C terjadi bulan lembab dan stasiun kota H terjadi bulan kering adalah 15,5 % (stasiun kota E) dan 66,2 % (stasiun kota D), dan nilai ramalan peluang bulan kering terkecil dan terbesar masing-masing berdasarkan stasiun kota C terjadi bulan lembab dan stasiun kota H terjadi bulan kering adalah 11,1 % (stasiun kota G) dan 82,1 % (stasiun kota E). Ukuran akurasi peramalan dengan menggunakan model graf *Multinomial Bayesian Network* seperti graf pada Gambar 18 dapat dinyatakan dengan rata-rata kuadrat kesalahan atau *Mean Square of Error* (MSE).

$$= \sum_{i=1}^n \frac{(-)^2}{12} = \frac{0,351501889}{12} \approx 0,0293$$

Model peluang ini sangat bermanfaat untuk menentukan peluang terjadinya bulan tertentu untuk stasiun yang diamati berdasarkan stasiun lain.

5. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan, maka diperoleh kesimpulan yaitu:

1. Model graf dari simulasi curah hujan dengan menggunakan model *Multinomial Bayesian Network* yang dihasilkan adalah model graf dengan 14 *node* dan 13 *arc*. Peluang curah

- hujan bulan kering untuk stasiun *D, C, E, B, G, H, J, K, L, M, N, F, A*, dan *I* masing-masing adalah 14,3%; 5%; 7%; 2,9%; 2,7%; 4,7%; 4%; 2%; 1,3%; 5,3%; 2,3%, dan 2,7%. Peluang curah hujan bulan lembab untuk stasiun *D, C, E, B, G, H, J, K, L, M, N, F, A*, dan *I* masing-masing adalah 66,7%; 56%; 68,3%; 15,7%; 33,3%; 38,3%; 63,3%; 21,7%; 31,7%; 37,3%; 22,7%; dan 33,3%. Peluang curah hujan bulan basah untuk stasiun *D, C, E, B, G, H, J, K, L, M, N, F, A*, dan *I* masing-masing adalah 19%; 39%; 24,7%; 81,4%; 64%; 57%; 32,7%; 76,3%; 67%; 57,3%; 75%; dan 64%.
2. Model graf dari simulasi curah hujan dalam peramalan berdasarkan klasifikasi *Oldeman* di beberapa stasiun berdasarkan curah hujan di stasiun kota *C* terjadi bulan lembab dan stasiun kota *H* terjadi bulan kering adalah model graf dengan *Bs* (*Bayesian Structure*) yang sama yaitu masing-masing 14 *node* dan 13 *arc*, namun dengan *Bp* (*Bayesian parameter*) yang berbeda. Peramalan nilai peluang curah hujan bulan kering untuk stasiun *D, E, B, G, J, K, L, M, N, F, A*, dan *I* masing-masing adalah 13,2%; 2,3%; 5,3%; 48,5%; 1,4%; 2,6%; 2,3%; 3,3%; 2,9%; 5,3%; 2,3%; dan 2,6%. Peramalan nilai peluang curah hujan bulan lembab untuk stasiun *D, E, B, G, J, K, L, M, N, F, A*, dan *I* masing-masing adalah 66,2%; 15,5%; 55%; 40,3%; 29,2%; 37,9%; 62,5%; 15,9%; 30,5%; 37,6%; 23,5%; dan 33,5%. Peramalan nilai peluang curah hujan bulan basah untuk stasiun *D, E, B, G, J, K, L, M, N, F, A*, dan *I* masing-masing adalah 20,6%; 82,1%; 39,7%; 11,1%; 69,4%; 59,5%; 35,2%; 80,8%; 66,6%; 57,1%; 74,2%; dan 63,8%.

DAFTAR PUSTAKA

- [1] Cooper G.F., and Herskovits E.(1992). *A Bayesian Method for The Induction of Probabilistic Networks from Data*. Machine Learning Journal 9:308-347.
- [2] Karlin, H.M. and Taylor, S.(1998). *An Introduction to Stochastic Modeling*. Third Edition. Academic Press: California.
- [3] Meigarani, I.(2010). *Penggunaan Metode Bayesian Network dalam Sistem Pakar untuk Diagnosis Penyakit Leukemia*. Skripsi yang dipublikasikan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Pendidikan Indonesia: Bandung.
- [4] Mulyono, S.(2004). *Riset Operasi*. Edisi Revisi. Fakultas Ekonomi Universitas Indonesia: Jakarta.
- [5] Pamungkas, P.(2006). Pola umum curah hujan di Indonesia.
- [6] <http://klastik.wordpress.com/2006/12/03/pola-umum-curah-hujan-di-indonesia/> diakses pada tanggal 2 Agustus 2011.
- [7] Rachmat, H.F.(2008). *Pendugaan Curah Hujan dengan Bayesian Networks – Studi Kasus Curah Hujan di Daerah Indramayu*. Tesis yang dipublikasikan oleh Sekolah Pascasarjana Institut Pertanian Bogor: Bogor.
- [8] Kurniawan, R.(2001). Bayesian Network.
- [9] <http://rahmadmartboy.blogspot.com/2011/07/kecerdasan-buatan-artificial.html> diakses pada tanggal 31 April 2011.
- [10] Sutanto, H.T.(2007). *Pendekatan Bayesian pada Model Regresi Polinomial*. Tesis yang dipublikasikan oleh Program Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh November: Surabaya.
- [11] Tjasyono, B.(2004). *Klimatologi*. Edisi Kedua. Penerbit ITB: Bandung.