METODE KLASIFIKASI BERSTRUKTUR POHON BINER (STUDI KASUS PADA PRAKIRAAN SIFAT HUJAN BULANAN DI BOGOR)[1)]*T*

Aan Kardiana[2)], Aunuddin[3)], Aji Hamim Wigena[3)], Hari Wijayanto[3)]
[2)]*Jurusan Teknik Informatika, FTI, Universitas YARSI, Jakarta*
E-mail: kardiana@yarsi.ac.id
[3)]*Program Pascasarjana, Institut Pertanian Bogor*

ABSTRAKSI

Makalah ini menguraikan tentang aplikasi pohon biner berupa pembentukan Pohon Klasifikasi dengan menggunakan algoritma pemilahan secara rekursif dilanjutkan dengan kombinasi strategi pemangkasan pohon ber-cost-complexity terkecil dan penggunaan contoh validasi silang, untuk mendapatkan pohon yang berukuran tepat. Metode ini diterapkan dalam prakiraan sifat hujan bulanan menggunakan data klimatologi rata-rata bulanan di stasiun Klimatologi Kelas I Darmaga Bogor. Peubah rata-rata suhu pada jam 07.00 wib. merupakan peubah dominan. Hasil evaluasi prakiraan selama periode pengujian, tingkat ketepatan metode: Regresi Linier Osilasi Selatan; Probabilitas; Regresi Linier Curah Hujan dan Pohon Klasifikasi masing-masing bernilai: 42,86 %; 14,29 %; 14,29 %; dan 71,43%.

***Kata kunci**: Pohon Klasifikasi.*

1. PENDAHULUAN

Salah satu aplikasi dari pohon biner adalah metode Pohon Klasifikasi. Penggunaan model berbasis pohon ini masih sedikit, meskipun berbagai penelitian telah menunjukkan bahwa pohon ini merupakan salah satu cara menarik dalam melakukan eksplorasi data dan pengambilan kesimpulan. Apalagi saat ini telah didukung oleh teknologi komputer yang berkembang semakin pesat sehingga proses komputasi yang melibatkan perhitungan yang cukup rumit dan mengandung banyak iterasi dapat dilakukan dengan semakin mudah dan cepat sehingga memberikan hasil yang lebih: cepat; tepat dan akurat.

Perancangan pengambilan keputusan berbasis pohon klasifikasi bermula dari studi dalam bidang ilmu sosial pada tahun 1973 oleh Morgan & Messenger dengan menggunakan program *THAID* (*Theta Automatic Interaction Detection*). Dalam bidang statistika, Breiman *et al*. telah merancang pengambilan keputusan berbasis pohon dengan mengembangkan konsep-konsep sehingga berpengaruh pada dua hal yaitu membawa masalah untuk diperhatikan oleh statistikawan dan mengusulkan algoritma untuk membentuk pohon **[3]**. Pengantar dalam S tentang model berbasis pohon telah mengembangkan metode dengan fungsi-fungsi untuk memeriksa dan mengevaluasi pohon serta mempermudah penggunaannya sehingga banyak memberikan dukungan untuk eksplorasi dan analisis data **[4]**.

Penyusunan pohon klasifikasi dapat juga dilihat sebagai salah satu jenis pemilihan peubah dan untuk ukuran yang besar merupakan transformasi monotonik dari peubah penjelas $\boldsymbol{x}$ dan peubah respon $\boldsymbol{y}$. Kita akan dapat mengetahui manakah peubah yang dominan di antara sederet peubah yang ada serta dapat pula mengidentifikasi peubah-peubah yang hanya berpengaruh secara lokal dalam kelompok tertentu.

Metode ini telah diaplikasikan dalam bidang: kedokteran; pertahanan dan keamanan; fisika; dan lain-lain. Salah satu aplikasi di bidang teknologi informasi adalah dalam masalah pengenalan digit mesin hitung (kalkulator) elektrik.

Makalah ini bertujuan untuk menguraikan metode pohon klasifikasi dan aplikasinya dalam pembuatan prakiraan sifat hujan bulanan. Hasil yang diperoleh akan dibandingkan dengan hasil keluaran menurut metode yang sekarang digunakan oleh BMG yaitu: metode Probabilitas; metode Regresi Linier Osilasi Selatan; dan metode Regresi Linier Curah Hujan **[6]**.

2. TEORI PENUNJANG

2.1 Pohon Klasifikasi

Misalkan $(\boldsymbol{X},\boldsymbol{Y})$ merupakan peubah acak dengan $\boldsymbol{X}=(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_q)\subset\Re^q$ tersusun dari $\boldsymbol{q}$ vektor penjelas dan $\boldsymbol{Y}\in\zeta=\{1, 2, \ldots, \boldsymbol{J}\}$ vektor label kelas yang berkaitan, sehingga terdapat himpunan data **[5]** atau *Learning sample* **[3]** $\boldsymbol{L}^{(0)} = \{(x_n, j_n), n = 1, 2, \ldots, \boldsymbol{N}^{(0)}\}$. Tujuan dari metode klasifikasi adalah ingin menaksir $\boldsymbol{Y}$ jika diberikan $\boldsymbol{X}$.

Pohon klasifikasi terdiri atas $\boldsymbol{T}=\{1,2,3, \ldots\}$ dan fungsi $l(t)$ dan $r(t)$ dari $\boldsymbol{T}$ ke $\boldsymbol{T}+\{0\}$ yang memenuhi ($l(t)=0$ dan $r(t)=0$) atau ($l(t)>t$ dan $r(t)>t$) serta $\forall t\in\boldsymbol{T}-\{1\}, \exists! s\in\boldsymbol{T} \ni t=l(s)$ atau $t=r(s)$ **[5]**. $\boldsymbol{T}$ dinamakan pohon klasifikasi; $\forall t\in\boldsymbol{T}$ dinamakan simpul pohon; Elemen minimum dari $\boldsymbol{T}$ dinamakan *root*($\boldsymbol{T}$); Jika $s,t\in\boldsymbol{T}$ dan $t=l(s)$ atau $t=r(s)$, maka s dinamakan *parent(t)* dan t dinamakan *sons(s)*, *parent(root(**T**))=0*; $t\in\boldsymbol{T}\ni$ *sons*(t)=0, dinamakan simpul terminal sedangkan $t\neq 1\in\boldsymbol{T}-\{sons(t)=0\}$ dinamakan simpul internal; Jika $s=parent(t)$ atau $s=parent(parent(t))\ldots$, t dinamakan *descendant(s)* dan s dinamakan *ancestor(t)*;

$\forall t \in \boldsymbol{T}$,cabang dari $\boldsymbol{T}$ dengan *root* $t \in \boldsymbol{T}$ ditulis $\boldsymbol{T}_t$, adalah himpunan bagian dari $\boldsymbol{T}$ yang memuat t dan seluruh *descendant* t; Pemangkasan cabang $\boldsymbol{T}_t$ dari $\boldsymbol{T}$ dengan cara menghilangkan seluruh *descendant(t)* dari $\boldsymbol{T}$ yaitu potong seluruh $\boldsymbol{T}_t$ kecuali *root*($\boldsymbol{T}_t$). Pohon terpangkas dinotasikan dengan $\boldsymbol{T}$-$\boldsymbol{T}_t$; Cabang dari $\boldsymbol{T}_1$ dari $\boldsymbol{T}$ dinotasikan dengan $\boldsymbol{T}_{1t}$; $\boldsymbol{T}_1$ pohon bagian terpangkas dari $\boldsymbol{T}$, dinotasikan dengan $\boldsymbol{T}_1 \leq \boldsymbol{T}$ jika $\boldsymbol{T}_1$ pohon bagian dari $\boldsymbol{T}$ dan *root*($\boldsymbol{T}_1$)=*root*($\boldsymbol{T}$). $\boldsymbol{T}_1 < \boldsymbol{T}$ berarti $\boldsymbol{T}_1 \leq \boldsymbol{T}$ dan $\boldsymbol{T}_1 \neq \boldsymbol{T}$; $\tilde{T}$ adalah himpunan terminal dari $\boldsymbol{T}$ **[7]**.

Pohon klasifikasi dapat juga dipandang sebagai suatu aturan keputusan yang merupakan fungsi $d(.)$ dari $\boldsymbol{R}^q$ ke $\boldsymbol{\zeta}$ **[5]** atau $\forall \boldsymbol{x} \in \boldsymbol{X}$, $d(\boldsymbol{x}) = \boldsymbol{y} \in \boldsymbol{\zeta}$ **[3]**. Jadi $d(.)$ mempartisi ruang $\boldsymbol{X}$ menjadi $\boldsymbol{J}$ himpunan saling lepas $\boldsymbol{A}(1), \boldsymbol{A}(2), \ldots, \boldsymbol{A}(j)$ sehingga kaitan antara aturan keputusan dengan pohon klasifikasi dapat dinyatakan sebagai:

$$\forall \boldsymbol{x} \in \boldsymbol{X} \quad \forall t \in \tilde{T} \quad d(\boldsymbol{x}) = j(t) \Leftrightarrow \boldsymbol{x} \in \boldsymbol{A}(t) \qquad (1)$$

2.2 Pembentukan Pohon Klasifikasi

Pohon klasifikasi dibentuk melalui pemilahan secara iteratif terhadap $\boldsymbol{X}$=*root(t)*=$\boldsymbol{A}$(1) kemudian dilanjutkan terhadap masing-masing $\boldsymbol{A}(s)$, $s \in sons(t)$, $t \in \boldsymbol{T}$, dan seterusnya.

Algoritma pembentukan pohon klasifikasi terdiri dari empat tahapan, yaitu: pemilihan pemilah; penentuan simpul terminal; penandaan label kelas; dan penentuan pohon dengan ukuran tepat.

2.2.1 Pemilihan Pemilah

Pada tahap ini dicari pemilah dari setiap simpul yang menghasilkan penurunan tingkat keheterogenan paling tinggi.

Keheterogenan suatu simpul diukur berdasarkan nilai *impurity*-nya. Nilai ini akan maksimum jika seluruh kelas secara seimbang bercampur di dalamnya dan akan minimum ketika hanya memuat satu kelas. Beberapa fungsi *impurity* yang dapat dipergunakan adalah indeks *Gini* : $f(p) = p(1\text{-}p)$ **[3]**, indeks informasi: $f(p) = -p\ln(p)$ dan indeks *twoing*: $I(t) = \min_{C_1 C_2} [f(p_{C1}) + f(p_{C2})]$ C_1, C_2 merupakan beberapa partisi dari J kelas menjadi dua himpunan bagian saling lepas. Jika J=2, *twoing* ekivalen dengan indeks *impurity* biasa **[9]**, atau indeks *entropi* $f(p) = N_j(t)\ln(p)$ **[11]**.

Pemilahan dimulai dari ruang umum dengan cara memeriksa nilai-nilai dari setiap peubah penjelas. Peubah kontinu atau ordinal x_j memilah berbentuk $x_j \leq t$ lawan $x_j > t$, $t \in \Re$ sedangkan untuk peubah nominal, $\boldsymbol{L}$ taraf dibagi menjadi dua himpunan saling lepas (akan terdapat sebanyak 2^{L-1}-1 kemungkinan pemilah).

Untuk suatu t, misalkan terdapat calon pemilah s yang memilah t menjadi t_L (dengan proporsi p_L) dan menjadi t_R (dengan proporsi p_R), maka kebaikan dari s didefinisikan sebagai penurunan *impurity*:

$$\Delta i(s,t) = i(t) - p_L\, i(t_L) - p_R\, i(t_R)$$

Pemilah s mengirimkan semua $\boldsymbol{x}_n$ di dalam t yang menjawab "ya" ke t_L dan semua $\boldsymbol{x}_n$ di dalam t yang menjawab "tidak" ke t_R.

Pohon dikembangkan dengan cara: pada simpul t_1, carilah s* yang memberikan nilai penurunan *impurity* tertinggi yaitu:

$$\Delta i(s^*,t_1) = \max_{s \in S} \Delta i(s, t_1)$$

maka t_1 dipilah menjadi t_2 dan t_3 menggunakan s*. Dengan cara yang sama dilakukan juga pencarian pemilah terbaik pada t_2 dan t_3 secara terpisah, dan seterusnya.

2.2.2 Penentuan Simpul Terminal

Ketika suatu simpul t dicapai sehingga tidak terdapat penurunan *impurity* secara berarti (devian terminal<1% devian *root*) **[10]** atau banyaknya objek cukup kecil (n_i<5) **[4]** maka selanjutnya t tidak dipilah lagi tetapi dijadikan simpul terminal dan hentikan pembentukan pohon.

2.2.3 Penandaan Label Kelas

Misalkan bahwa $\boldsymbol{T}$ telah dikonstruksi menggunakan data dalam $\boldsymbol{L}^{(0)}$. Pada masing-masing $t \in \boldsymbol{T}$, nyatakanlah $p(t) = \boldsymbol{N}(t)/\boldsymbol{N}(0)$ [$\boldsymbol{N}(t)$ = jumlah data anggota $\boldsymbol{L}^{(0)}$ yang memenuhi $\boldsymbol{x}_n \in \boldsymbol{A}(t)$]; $p(t)$ merupakan taksiran dari peluang $\boldsymbol{P}(\boldsymbol{x}_n \in \boldsymbol{A}(t))$ berdasar $\boldsymbol{L}^{(0)}$. $\boldsymbol{P}(\boldsymbol{Y}=j | \boldsymbol{X} \in \boldsymbol{A}(t))$ berdasar $\boldsymbol{L}^{(0)}$ ditaksir oleh $p(j|t) = \boldsymbol{N}_j(t)/\boldsymbol{N}(t)$, j = 1, 2, …, J, [$\boldsymbol{N}_j(t)$ adalah jumlah data anggota $\boldsymbol{L}^{(0)}$ yang memenuhi $\boldsymbol{x}_n \in \boldsymbol{A}(t)$ dan $y_n = j$].

Label kelas dari simpul terminal ditentukan berdasarkan aturan jumlah terbanyak, yaitu jika $p(j_0|t) = \max_j p(j \mid t) = \max_j \frac{N_j(t)}{N(t)}$, maka label kelas untuk terminal t adalah j_0.

2.3 Penentuan Pohon optimum

Pohon klasifikasi hasil proses pembentukan tidak dibatasi ukurannya. Pohon klasifikasi terbesar memberikan nilai *misclassification error* paling kecil sehingga kita akan selalu cenderung memilih pohon ini untuk prakiraan. Tetapi, pohon ini cukup kompleks dalam menggambarkan struktur data.

Salah satu cara penentuan pohon klasifikasi optimum --tanpa mengorbankan kebaikan kecocokan melalui pengurangan simpul pohon (sehingga dicapai penghematan gambaran)-- dikenal dengan istilah pemangkasan (*pruning*). Pemangkas berturut-turut memangkas pohon bagian yang kurang penting. Tingkat kepentingan sebuah pohon bagian diukur berdasar ukuran *cost-complexity* **[10]** adalah:

$$D_\alpha(\boldsymbol{T}_k) = D(\boldsymbol{T}_k) + \alpha |\tilde{T}_k|$$

dimana:

$D(\boldsymbol{T}_k)$ = devian dari pohon bagian $\boldsymbol{T}_k$,
$\tilde{T}_k$ = himpunan simpul terminal pada $\boldsymbol{T}_k$,
$|\tilde{T}_k|$ = banyak simpul terminal pada $\tilde{T}_k$
α = parameter *cost-complexity*.

Hasil proses pemangkasan berupa sederet pohon klasifikasi $\boldsymbol{T}_k$ dan dengan menggunakan contoh validasi silang (*cross-validation sample*) dapat ditentukan pohon optimum $\boldsymbol{T}_{k0}$ dengan $\boldsymbol{R}^{cv}(\boldsymbol{T}_{k0}) = \min_k R^{CV}(T_k)$.

2.4 Metode prakiraan BMG

Metode prakiraan sifat hujan bulanan yang sekarang digunakan oleh BMG adalah:

2.4.1 Metode Probabilitas

Analisis yang dihasilkan dari metode Probabilitas berupa nilai prosentase dari tiga kategori sifat hujan yaitu di bawah normal (BN), normal (N), dan di atas normal (AN). Nilai prosentase terbesar dari ketiga sifat tersebut merupakan hasil yang akan dipakai dalam suatu prakiraan **[6]**.

2.4.2 Metode Regresi Linier Osilasi Selatan

Peubah respon pada metode Regresi Linier Osilasi Selatan adalah jumlah curah hujan tiga bulan terakhir, dan indeks osilasi selatan pada tiga bulan sebelumnya sebagai peubah prediktor **[6]**.

2.4.3 Metode Regresi Linier Curah Hujan

Metode Regresi Linier Curah Hujan melibatkan data curah hujan bulan yang akan diprakirakan sebagai peubah respon dan jumlah curah hujan dua bulan sebelumnya sebagai prediktor **[6]**.

3. DATA DAN METODE

Data yang digunakan dalam penelitian ini adalah data klimatologi rata-rata bulanan yang diamati di stasiun Klimatologi Kelas I Darmaga Bogor, dari bulan Februari tahun 1980 sampai bulan April 1999 **[1]**.

Peubah responnya adalah sifat hujan bulanan (**s**) yang berjenis kategorik. Nilainya adalah **BN** jika jumlah curah hujan bulanan<0.85*normal curah hujan; **N**; jika 0,85*normal curah hujan≤jumlah curah hujan bulanan≤1,15*normal curah hujan **AN** jika jumlah curah hujan bulanan≥1,15*normal curah hujan. Sedangkan peubah penjelasnya terdiri dari peubah rata-rata : suhu (^{0}C) pada jam : 07.^{00} wib (**t07**); 13.^{00} wib (**t13**); 18.^{00} wib (**t18**), suhu terbesar (**tmax**); suhu terkecil (**tmin**), tingkat penyinaran matahari dalam % (**rm**), tekanan udara dalam mb (**tu**), kelembaban nisbi dalam % pada jam : 07.^{00} wib (**k07**); 13.^{00} wib (**k13**); 18.^{00} wib (**k18**); sifat hujan bulan sebelumnya (**s_0**); dan indeks osilasi selatan (**osi**). Selain itu digunakan juga data-data prakiraan dan evaluasi sifat hujan bulanan yang telah diterbitkan oleh BMG **[2]** untuk membandingkan tingkat ketepatan metode.

4. HASIL DAN PEMBAHASAN

Hasil prakiraan berdasarkan metode : Probabilitas; Regresi Linier Osilasi Selatan; dan Regresi Linier Curah Hujan masing-masing memberikan tingkat ketepatan sebesar: 14,29%; 42,86%, dan 14,29%. Hasil selengkapnya dapat dilihat pada Tabel 1, Tabel 2, dan Tabel 3.

Tabel 1. Prakiraan Sifat Hujan Bulanan menurut metode Probabilitas.

Tahun	No	Bulan	S	$\hat{S}_p$	Hasil
1998	1	Oktober	AN	BN	Salah
	2	Nopember	BN	AN	Salah
	3	Desember	BN	AN	Salah
1999	4	Januari	BN	AN	Salah
	5	Februari	BN	N	Salah
	6	Maret	**BN**	**BN**	**Benar**
	7	April	N	BN	Salah

dimana:

S ≡ sifat hujan sebenarnya

$\hat{S}_p$ ≡ sifat hujan menurut metode probabilitas

Tabel 2. Prakiraan Sifat Hujan Bulanan menurut metode Regresi Linier Osilasi Selatan

Tahun	No	Bulan	S	$\hat{S}_{os}$	Hasil
1998	1	Oktober	**AN**	**AN**	**Benar**
	2	Nopember	BN	AN	Salah
	3	Desember	**BN**	**BN**	**Benar**
1999	4	Januari	**BN**	**BN**	**Benar**
	5	Februari	BN	AN	Salah
	6	Maret	BN	AN	Salah
	7	April	N	AN	Salah

dimana:

$\hat{S}_{os}$ ≡ sifat hujan menurut metode Regresi Linier Osilasi Selatan

Tabel 3. Prakiraan Sifat Hujan Bulanan menurut metode Regresi Linier Curah Hujan

Tahun	No	Bulan	S	$\hat{S}_{ch}$	Hasil
1998	1	Oktober	AN	N	Salah
	2	Nopember	BN	N	Salah
	3	Desember	BN	N	Salah
1999	4	Januari	BN	N	Salah
	5	Februari	BN	AN	Salah
	6	Maret	BN	N	Salah
	7	April	**N**	**N**	**Benar**

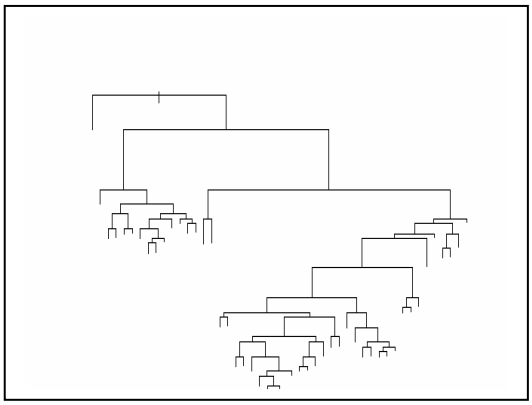
dimana:

$\hat{S}_{ch}$ ≡ sifat hujan menurut metode Regresi Linier Curah Hujan

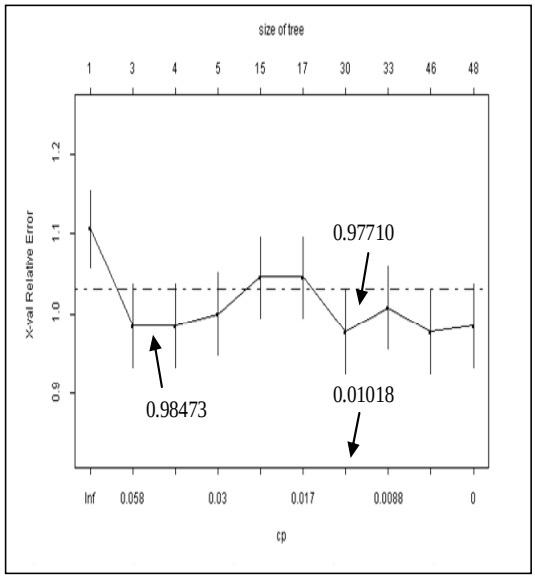
Penerapan metode Pohon klasifikasi pada data klimatologi rata-rata bulanan menghasilkan Pohon klasifikasi terbesar seperti terlihat pada Gambar 1.

Pohon klasifikasi yang telah diperoleh selanjutnya secara iteratif dipangkas menjadi deretan pohon bagian yang makin kecil dan tersarang.

Kemudian dengan mengkombinasikan aturan *cost complexity minimum* dan penggunaan contoh validasi silang *10-fold* (*10-fold cross-validation sample*), diperoleh pohon berukuran tepat (optimum) yang dicapai pada parameter *complexity* bernilai *cp* = 0.01018 karena memberikan nilai $\boldsymbol{R}^{cv}$ (*x-val Relative Error*) terkecil yaitu $\boldsymbol{R}^{cv}$ = 0.9771 (Gambar 2).



Gambar 1. Pohon klasifikasi terbesar

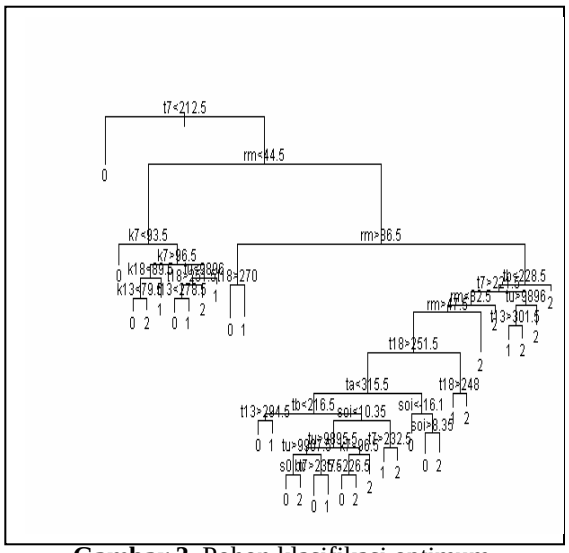


Gambar 2. Plot parameter *complexity*

Pohon optimum mengandung 30 buah simpul terminal dengan tingkat ketepatan pengelompokkan amatan sebesar 69.47% (Gambar 3).

Peubah penjelas yang pertama memilah amatan adalah peubah suhu pada jam 07.^{00} wib. sehingga peubah ini merupakan peubah dominan. Hasil ini sesuai dengan Mc. Hattie & Schnelle **[8]** bahwa yang berperanan penting dalam menentukan klasifikasi sifat hujan adalah curah hujan dan suhu udara.

Amatan sebanyak 205 buah, pada simpul pertama (*root*) dipilah menjadi kelompok kiri atau kelompok kanan berdasarkan nilai peubah rata-rata suhu pada jam 07.^{00} wib. Amatan yang memiliki nilai suhu pada jam 07.^{00} wib<21,25^{0}C sebanyak 7 buah mengelompok di simpul 2 (simpul terminal) sedangkan amatan yang memiliki nilai suhu pada jam 07.^{00} wib≥21,25^{0}C sebanyak 198 buah mengelompok di simpul 3. Kelompok amatan pada simpul 3 selanjutnya oleh peubah tingkat penyinaran matahari dipilah menjadi kelompok di simpul 6 yang terdiri atas 43 buah jika tingkat penyinaran matahari<44,5% dan amatan dengan tingkat penyinaran matahari≥44,5% sebanyak 155 buah mengelompok di simpul 7.



Gambar 3. Pohon klasifikasi optimum

Amatan pada simpul 6 dipilah lagi berdasar peubah kelembaban pada jam 07.^{00} wib. Sebanyak 5 buah amatan yang mempunyai nilai peubah kelembaban pada jam 07.^{00} wib<93.5% mengelompok di simpul 12 (simpul terminal) sedangkan amatan lainnya dengan kelembaban pada jam 07.^{00}wib≥93.5% sebanyak 38 buah mengelompok di simpul 13. Proses ini dilanjutkan hingga dicapai semua simpul terminal.

Tabel 4. Prakiraan Sifat Hujan Bulanan menurut metode Pohon Klasifikasi

Tahun	***No***	***Bulan***	***s***	$\hat{S}_o$	***Hasil***
1998	1	Oktober	**AN**	**AN**	**Benar**
	2	Nopember	BN	N	Salah
	3	Desember	**BN**	**BN**	**Benar**
1999	4	Januari	**BN**	**BN**	**Benar**
	5	Februari	BN	N	Salah
	6	Maret	**BN**	**BN**	**Benar**
	7	April	**N**	**N**	**Benar**

dimana:

$\hat{S}_o \equiv$ sifat hujan menurut pohon optimum

Selanjutnya pohon klasifikasi yang diperoleh digunakan untuk membuat prakiraan sifat hujan bulanan dengan cara menelusuri pohon klasifikasi optimum dari mulai simpul *root* hingga dicapai simpul terminal berdasarkan nilai peubah penjelas yang diperoleh.

Metode pohon klasifikasi optimum (yang mengandung 30 buah simpul terminal) memberikan tingkat ketepatan sebesar 71,43%. Hasil selengkapnya bisa dilihat pada Tabel 4.

KESIMPULAN

Kesimpulan yang diperoleh dalam penelitian ini adalah:

1. Peubah rata-rata suhu pada jam $07.^{00}$ wib. merupakan peubah dominan.
2. Dari hasil evaluasi prakiraan selama 7 (tujuh) bulan terakhir, metode pohon klasifikasi optimum memberikan tingkat ketepatan sebesar 71,43 %.
3. Selama periode pengujian, dibandingkan dengan metode prakiraan yang sekarang digunakan oleh BMG, tingkat ketepatan metode pohon klasifikasi nilainya 28,57 % di atas tingkat ketepatan metode Regresi Linier Osilasi Selatan (yang bernilai 42,86 %), dan 57,14 % di atas tingkat ketepatan metode Probabilitas dan Regresi Linier Curah Hujan (yang keduanya bernilai 14,29 %). Jadi, metode pohon klasifikasi mempunyai tingkat ketepatan yang lebih baik dibandingkan dengan metode yang sekarang digunakan oleh BMG.

DAFTAR PUSTAKA

[1] Badan Meteorologi dan Geofisika Balai Wilayah II. 1980-1999. *Data-data Klimatologi Bulanan.* Stasiun Darmaga, Bogor.

[2] Badan Meteorologi dan Geofisika Balai Wilayah II. 1998-1999. *Evaluasi dan Prakiraan Curah Hujan Propinsi Jawa Barat.* Stasiun Darmaga, Bogor.

[3] Breiman, L., J. H. Friedman, R. A. Olshen & C.J. Stone. 1984. *Classification and Regression Tree.* Chapman and Hall, New York.

[4] Clark, L.A. & D. Pregibon. 1992. *Tree-based models*. In Chambers, J.M. & T. J. Hastie (eds). 1992. *Statistical Models in S.* Chapman and Hall, New York.

[5] Gelfand, S.B., C. S. Ravishakar & E. J. Delp. 1991. An iterative growing and pruning algorithm for classification tree design. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 167-174.

[6] Hadiyanto, S. 1994. Metode Prakiraan. Badan Meteorologi Dan Geofisika Balai Wilayah II, Jakarta (tidak dipublikasikan).

[7] Keprta, S. 1996. Non-binary classification trees. *Statistics and Computing*. 06:231-243.

[8] Mc. Hattie, L.B. & F. Schnelle. 1974. *An introduction to Agrotopoclimatology*. WMO. TN. No. 133. Geneva.

[9] Therneau, T.M. & E. J. Atkinson. 1997. An Introduction to Recursive Partitioning using the RPART routines. *Technical Report*; *Mayo Foundation*.

[10] Venables, W. N. & B. D. Ripley. 1994. *Modern Applied Statistics with S-plus*. Springer-Verlag. New York. Inc.

[11] Zhang, Heping. 1998. Classification trees for multiple binary responses. *Journal of the American Statistical Association*. 93: 180-193.